Delft University of Technology

Large deviations for the empirical measure of the zig-zag process

Bierkens, Joris; Nyquist, Pierre; Schlottke, Mikola C.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Large deviations for the empirical measure of the zig-zag process

Joris Bierkens

Department of Mathematics
Vrije Universiteit
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands
e-mail: j.bierkens@vu.nl

Pierre Nyquist

Department of Mathematics
KTH
100 44
Stockholm, Sweden
e-mail: pierren@kth.se

and

Mikola C. Schlottke

Department of Mathematics and Computer Science
Technische Universiteit Eindhoven
PO Box 513
5600 MB Eindhoven
The Netherlands
e-mail: M.Schlottke@tue.nl

Abstract: The zig-zag process is a piecewise deterministic Markov process in position and velocity space. The process can be designed to have an arbitrary Gibbs type marginal probability density for its position coordinate, which makes it suitable for Monte Carlo simulation of continuous probability distributions. An important question in assessing the efficiency of this method is how fast the empirical measure converges to the stationary distribution of the process. In this paper we provide a partial answer to this question by characterizing the large deviations of the empirical measure from the stationary distribution. Based on the Feng-Kurtz approach, we develop an abstract framework aimed at encompassing piecewise deterministic Markov processes in position-velocity space. We derive explicit conditions for the zig-zag process to allow the Donsker-Varadhan variational formulation of the rate function, both for a compact setting (the torus) and one-dimensional Euclidean space. Finally we derive an explicit expression for the Donsker-Varadhan functional for the case of a compact state space and use this form of the rate function to address a key question concerning the optimal choice of the switching rate of the zig-zag process.

1

1. Introduction

The problem of sampling from a given high-dimensional probability distribution arises in a wide range of applications, for example chemistry, physics, statistics and engineering. Specifically, consider the task of sampling from a distribution π with density with respect to Lebesgue measure given by

$$\pi(y) \propto \exp\{-U(y)\}, \ y \in E,$$

for some potential function $U : E \to \mathbb{R}$ and state space E. The most common approach is to use Markov chain Monte Carlo (MCMC) methods, which are now essential tools in areas such as computational statistics, molecular dynamics and machine learning, see e.g. [RC04, AG07, AdFDJ03] and references therein.

The idea behind MCMC is to construct a Markov process Y_t with π as the invariant measure and use the corresponding empirical measure to obtain approximations. For example, under ergodicity, for any observable $f \in L^1(\pi)$ and $t > 0$ large, $\frac{1}{t} \int_0^t f(Y_s)ds$ can be used to approximate the expected value $\int_E f(y)\pi(dy)$.

Although many standard MCMC constructions, such as the Metropolis-Hastings algorithm [MRR$^+$53], can be used to sample from essentially any target distribution π, most suffer from slow convergence to the invariant distribution or heavy computational costs per iteration. Designing new, efficient dynamics has therefore become an important research direction within applied probability.

Over the last decade piecewise deterministic Markov processes (PDMPs) [Dav84] have emerged as a new tool for the numerical simulation of probability distributions, potentially mitigating both problems. The two main examples of such processes are the Bouncy Particle Sampler and the Zig-Zag Sampler [BCVD17, BFR19], after similar ideas appeared first in [PD12] and [Mon16]. The idea of using PDMPs extends the ubiquitous discrete time MCMC methodology towards a new continuous time approach, having several advantageous aspects. First, by construction PDMPs are non-reversible Markov processes, which typically results in a smaller asymptotic variance as compared to reversible methods (see e.g. [AL19] for a recent study). Furthermore, computation of PDMP trajectories on a computer may be carried out using a subsampling technique, reducing the computational effort required in each iteration; we refer to e.g. [BFR19, FBPR18] for details of the computational aspects.

In order to employ this new PDMP methodology a solid understanding of mathematical properties of these methods is necessary. Whereas the theoretical properties of PDMPs have been an active research area in recent years, our understanding of the performance of the corresponding MCMC methods is still incomplete. In particular knowledge of the speed of convergence of time averages is essential in choosing the most suitable sampling technology for a particular problem and in tuning the parameters of the chosen method. In the spirit of recent work on empirical measure large deviations in the MCMC context [DLPD12, RBS15], we propose the use of large deviation results for studying and comparing the performance of PDMPs.

In summary, the main contributions of this work are as follows:

- A semigroup approach to establish the large deviations principle for the empirical measures of a class of Markov processes satisfying assumptions aimed at position-velocity PDMPs.
- The large deviations principle for empirical measures of the zig-zag process in both a compact and non-compact setting.
- Derivation of an explicit form of the rate function associated with the zig-zag process.
- Evaluation of the zig-zag rate function as a function of the additional switching rate γ, providing an answer to a key question about the switching rate.

The study of large deviations for empirical measures dates back to the work by Donsker and Varadhan [DV75a, DV75b, DV76b]. In the simulation context it is well-known that for rare-event simulation, sample-path large deviations play an important rôle in evaluating and designing efficient algorithms ; see [AG07, Buc04, BD19] and references therein. In contrast, empirical measure large deviations are much less explored as a tool for analysing Monte Carlo methods. Standard measures for analysing the efficiency of methods based on ergodic Markov processes include the spectral gap of the associated semigroup and the asymptotic variance for given observables, see for example [Ros03, BR08, DHN00, FHPS10, FdSHS93, HHMS05, MT96, RR04]. However these measures are not necessarily appropriate for studying the rate of convergence, as they only link indirectly to the empirical measure, the quantity of interest in Monte Carlo methods. Empirical measure large deviations on the other hand connect explicitly to the relevant properties, such as the transient behaviour of the underlying process. In a similar spirit, [BRB19] recently used concentration inequalities to obtain non-asymptotic performance guarantees for PDMPs.

To the best of our knowledge, the first results using empirical measure large deviations for analysing MCMC methods were [PDD⁺11, DLPD12]. Therein empirical measure large deviations, specifically the associated rate function, was proposed as a tool for analysing parallel tempering, one of the computational workhorses of the physical sciences, leading to a new type of simulation method (infinite swapping). In the subsequent work [DDN18] empirical measure large deviations were again used, combined with associated stochastic control problems, to analyse the convergence properties of these algorithms. Similarly, in [RBS15] Rey-Bellet and Spiliopoulos use empirical measure large deviations to analyse the performance of certain non-reversible MCMC samplers.

The work by Donsker and Varadhan is the starting point for many results and application of empirical measure large deviations and their work has been extended in numerous directions, see e.g. [Dem94, FK06, BD19] for an overview and further references. However, naively applying the existing theory to PDMPs does not work since the transition probabilities are not sufficiently regular: for every $t > 0$ there is a positive probability that the process has not switched by time t, resulting in an atomic component of the Markov transition kernel. As a first step towards using empirical measure large deviations for analysing the performance of PDMPs we must therefore establish the relevant large deviations

principles.

The focus of this paper is to establish general large deviation results aimed at PDMPs and then specialize to the zig-zag process. In the process of proving the necessary large deviation results we consider a general class of Markov processes that can have position-velocity PDMPs, such as the bouncy particle and the zig-zag samplers, as special cases. In particular, this class includes processes that are not of diffusion type and non-reversible processes. When specialising to the zig-zag process, we derive an explicit form of the rate function, going beyond the variational form typical for results of Donsker-Varadhan-type. To the best of our knowledge this is the first instance where an explicit form of the rate function has been obtained for non-reversible processes that do not have a drift-diffusion character.

A key question for using the zig-zag process for MCMC is whether or not it is advantageous for convergence to use the minimal (canonical) switching rates, or if one should allow for additional switches according to a fixed refreshment rate $\gamma > 0$. Our analysis of the rate function associated with the zig-zag process allows us to give a partial answer to this question: in Section 3.3 we establish that the rate function is decreasing as a function of the additional rate γ, establishing that from a large deviations perspective it is optimal to use the smallest possible rates, i.e. set $\gamma = 0$. This goes in the opposite direction of the conclusion drawn from a spectral analysis (see [BV19, Section 7.3]), which shows at least a small benefit of increasing gamma beyond zero. This highlights the different nature of convergence of empirical averages (by studying large deviations or asymptotic variance, see e.g. [AL19, BD17]) and convergence to equilibrium, using e.g. the spectral gap to describe rate of convergence; see [Ros03] for more on this phenomenon. Our conclusion is in line with the earlier observation that having more non-reversibility increases the rate function [RBS15]: one can view increasing γ as decreasing the extent of non-reversibility inherent to the process.

Evaluation of the large deviation rate function for empirical measures, beyond the variational form given by Donsker and Varadhan, is typically a challenging task. For the diffusion setting, including both reversible and non-reversible processes, see [DL18] and the references therein. In [DL15] the authors consider reversible jump Markov processes and use stochatic control and weak convergence arguments to derive an explicit form of the rate function. Lastly, in the MCMC context, [RBS15] consider diffusion processes on a compact manifold where the drift can be decomposed into sufficiently smooth reversible and non-reversible parts. The rate function can then be expressed in terms of the rate function of a related reversible diffusion and the solution of an elliptic PDE associated with the non-reversible component of the drift.

The proofs of the large deviation results in this paper are based on the general Hamilton-Jacobi approach to empirical measures developed by Feng and Kurtz in [FK06, Chapter 12]. We describe this approach, in the context of this paper, in more detail in Section 4.1.

The remainder of the paper is organised as follows. In Section 2 we give the necessary preliminaries: notation and relevant definitions, background on the zig-zag process and empirical measure large deviations. In particular we

recall well-known large deviation results for empirical measures by Donsker and Varadhan. The main results are then presented in Section 3. The section is split into the main assumptions and general large deviation statements (Section 3.1), large deviation results for the zig-zag process (Section 3.2) and an explicit expression of the rate function associated with the zig-zag process (Section 3.3). All proofs are deferred to Section 4.

2. Preliminaries

2.1. Notation and definitions

Throughout the paper, E will denote a complete separable metric space and $\mathcal{B}(E)$ the relevant σ-algebra on E; unless otherwise stated this is taken to be the Borel σ-algebra. $C(E)$ and $C_b(E)$ are the spaces of functions $f : E \to \mathbb{R}$ that are continuous and bounded continuous, respectively. The space of continuous and right-continuous functions from $[0, \infty)$ to E is denoted by $C_E[0, \infty)$ and $D_E[0, \infty)$, respectively. A sequence of functions $\{f_n\}_n$ on E converges *boundedly and uniformly on compacts* to a function f if and only if $\sup_n \|f_n\| < \infty$ and for each compact $K \subseteq E$,

$$\lim_{n \to \infty} \sup_{x \in K} |f_n(x) - f(x)| = 0.$$

This is denoted as $f = buc - \lim_{n \to \infty} f_n$.

For a Markov process $Y = \{Y_t : \ t \geq 0\}$, we denote by $S = \{S(t) : t \geq 0\}$ the associated Markov semigroup. A semigroup $S(t)$ acting on $C(E)$ is *Feller continuous* if, for any t, $S(t) : C(E) \to C(E)$, *strongly continuous* if $S(t)f \to f$ as $t \to 0$ for any $f \in C(E)$ and *buc-continuous* if $buc - \lim_{t \to 0} S(t)f = f$ for $f \in C_b(E)$.

For an operator L, $\mathcal{D}(L)$ denotes the domain of L. For functions in $\mathcal{D}(L)$, $\mathcal{D}^+(L)$ denotes those that are strictly positive and $\mathcal{D}^{++}(L)$ those that are positive and uniformly bounded from below by a positive constant. For a given L we use B to denote the extended generator associated with L.

We use $\mathcal{P}(E)$ to denote the space of probability measures on E, and $\mathcal{P}_c(E)$ is the subset of probability measures with compact support. Throughout the paper we equip $\mathcal{P}(E)$ with the topology of weak convergence: $\rho_n \to \rho$ in this topology if

$$\int_E f(x)\rho_n(dx) \to \int_E f(x)\rho(dx), \ n \to \infty, \ \ \forall f \in C_b(E).$$

A special case that will be considered several times is $\mathcal{P}(D_E[0, \infty))$, which is also equipped with the weak topology. For a process $\{Y(t), t \geq 0\}$ taking values in E and $y \in E$, we denote by $\mathbb{P}_y \in \mathcal{P}(D_E[0, \infty))$ the distribution of the process $Y(t)|_{t \geq 0}$ starting at $y \in E$.

The set of positive Borel measures on E is denoted by $\mathcal{M}(E)$ and the set of finite Borel measures on E are denoted by $\mathcal{M}_f(E) \subset \mathcal{M}(E)$. We let $\mathcal{L}(E)$

denote the following subset of $\mathcal{M}(E \times [0, \infty))$:

$$\mathcal{L}(E) = \{z \in \mathcal{M}(E \times [0, \infty)) : \; z(E \times [0, t]) = t, \; t \geq 0\}.$$

The set $\mathcal{L}(E)$ is endowed with the topology of weak convergence on bounded time intervals: for $\{\rho_n\} \subset \mathcal{L}(E)$, $\rho_n \to \rho$ if for all $f \in C_b(E \times [0, \infty))$ and all $t \geq 0$,

$$\int_{E \times [0,t]} f(x, s) d\rho_n(x, s) \to \int_{E \times [0,t]} f(x, s) d\rho(x, s).$$

Then $\mathcal{L}(E)$ is the set of Borel-measures on $E \times [0, \infty)$ of the form

$$d\rho(x, t) = \mu_t(dx)dt,$$

for probability measures $\mu_t \in \mathcal{P}(E)$. That is, for every $\rho \in \mathcal{L}(E)$, there exists a measurable path $s \mapsto \mu_s \in \mathcal{P}(E)$ such that

$$\rho(A \times [0, t]) = \int_0^t \mu_s(A)ds, \quad \text{for any } A \in \mathcal{B}(E), \; t > 0.$$

For two real-valued functions f and g, $f \sim g$ as $x \to x_0$ means that they are asymptotically equivalent in the limit $x \to x_0$:

$$\frac{f(x)}{g(x)} \to 1, \quad \text{as } x \to x_0.$$

2.2. Large deviations for empirical measures

Consider a Markov process $Y = \{Y_t : t \geq 0\}$ taking values in a complete separable metric space E, with associated generator $L : \mathcal{D}(B) \subseteq C_b(E) \to C_b(E)$ and semigroup $S(t)$. The *empirical measure* η_t associated with Y_t is the stochastic process with values in $\mathcal{P}(E)$ defined by

$$\eta_t(A) = \frac{1}{t} \int_0^t \mathbf{1}_A(Y_s)ds, \quad A \in \mathcal{B}(E).$$

Empirical measures play an important rôle in, for example, the settings of MCMC methods and steady-state simulations, via the pairing of measures and observables: For a probability measure $\mu \in \mathcal{P}(E)$ and a function $V \in C_b(E)$, we write

$$\mu(V) = \int_E V(y)d\mu(y)$$

for the pairing of measures and observables. For the empirical measure η_t, this pairing corresponds to time averages,

$$\eta_t(V) = \frac{1}{t} \int_0^t V(Y_s)ds. \tag{1}$$

If there is an invariant measure $\pi \in \mathcal{P}(E)$ associated with the generator L, ergodicity of the process Y_t will ensure the convergence $\eta_t \to \pi$ as $t \to \infty$, w.p. 1 in $\mathcal{P}(E)$, from which it follows that for any $V \in C_b(E)$,

$$\eta_t(V) \to \pi(V) \quad \text{as } t \to \infty, \ \mathbb{P} - a.s.$$

Thus, time averages such as (1) are precisely what is used to form approximations in Monte Carlo methods and there is a direct link between the performance of such simulation methods and the properties of the empirical measure.

The theory of large deviations for empirical measures is concerned with deviations of η_t from π as t grows large. The gist of the so-called large deviations principle is that for any $\rho \in \mathcal{P}(E)$, for large t

$$\mathbb{P}_y(\eta_t \approx \rho) \simeq \exp\left\{-tI(\rho)\right\},$$

where the function $I : \mathcal{P}(E) \to [0, \infty]$ is the rate function associated with the process. The formal definition underlying all of large deviation theory, making the previous display rigorous, is as follows: The sequence $\{\eta_t\}_{t \geq 0}$ satisfies a *large deviations principle (LDP)* with speed t and *rate function I*, if I is lower semicontinuous, has compact sublevel-sets and for any measurable set $A \subseteq \mathcal{P}(E)$,

$$- \inf_{\mu \in A^\circ} I(\mu) \leq \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\eta_t \in A^\circ\right)$$

$$\leq \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\eta_t \in \bar{A}\right) \leq - \inf_{\mu \in \bar{A}} I(\mu),$$

where A° and $\bar{A}$ is the interior and closure of the set A.

Under relatively mild conditions on the dynamics of the process Y the rate function will be strictly convex and satisfy $I(\mu) = 0$ if and only if $\mu = \pi$. Thus, the rate function characterises the exponential rate of decay of probabilities of sets not including the invariant distribution π. Moreover the rate function can be used to characterise *how* events may occur - for sets A that do not include π, the minimisers of I over A represent the behaviour η_t is most likely to exhibit if A occurs.

For empirical measures of Markov processes, the rate function associated with an LDP can often be expressed using a variational form, obtained by Donsker & Varadhan [DV75a], involving the generator L of the underlying process. For the compact setting, they proved the following result in [DV75a].

Theorem 2.1 ([DV75a, Theorem 3])**.** *Take E to be a compact, complete separable metric space. Let $S(t)$ be a Markov semigroup acting on $C(E)$ equipped with the supremum norm, L is the generator associated to S. Assume the following:*

(DV.1) The semigroup is Feller continuous and strongly continuous.
(DV.2) There exists a probability measure $\lambda \in \mathcal{P}(E)$ such that for each $t > 0$ and $x \in E$, the transition probabilities $P(t, x, dy)$ are absolutely continuous with respect to λ, that is

$$P(t, x, dy) = p(t, x, y)\lambda(dy),$$

for some p with $0 < a(t) \le p(t,x,y) \le A(t) < \infty$.

Then the associated sequence $\{\eta_t\}_{t>0}$ satisfies a large deviation principle in $\mathcal{P}(E)$, with rate function $\mathcal{I} : \mathcal{P}(E) \to [0,\infty]$ given by

$$\mathcal{I}(\mu) = -\inf_{u \in \mathcal{D}^+(L)} \int_E \frac{Lu}{u} d\mu. \tag{2}$$

The theorem applies in particular to drift-diffusions taking values in a compact space. Roughly speaking, for such processes, with reasonable coefficients, the Feller-continuity is satisfied and the diffusive part ensures absolute continuity with respect to a volume measure dx. In [RBS15] Rey-Bellet and Spiliopoulos use this result to study performance of specific non-reversible MCMC methods based on drift-diffusions; their Assumption (H) allows for an application of Theorem 2.1.

Condition (DV.2) is a reasonable transitivity assumption for processes that involve a diffusive term. However, this condition excludes many interesting examples, such as continuous-time jump processes, see e.g. [DL15]. The issues highlighted therein are present also for the zig-zag process on $\mathbb{R} \times \{\pm 1\}$: in a sense, the absence of a diffusive operator excludes the possibility of finding a suitable reference measure.

If the process Y_t is reversible with respect to the reference measure, that is $p(t,x,y) = p(t,y,x)$, then the rate function takes a more explicit form, see e.g. Theorem 5 in [DV75a]. However, our interests are explicitly in non-reversible processes, such as the zig-zag process, and therefore such representations are not available.

In conclusion, while Theorem 2.1 can be a starting point for many drift-diffusion processes, it is not a sufficient tool for many other interesting processes, including the position-velocity PDMPs. In order to use large deviation results to study performance of such MCMC algorithms we must first overcome this obstacle and establish the relevant large deviations principles.

In [DL18], Dupuis and Lipshutz consider large deviations of empirical measures of $\mathbb{R}^d$-valued drift-diffusions. Their Condition 2.2 corresponds to a type of stability criterion in terms of a Lyapunov function. A transitivity property similar to Condition (DV.2) of Theorem 2.1 is satisfied due to the diffusive part, and they prove a different, explicit representation of the rate function, assuming only standard regularity conditions on the coefficients. In particular, this representation holds for non-reversible drift-diffusions.

2.3. The zig-zag process

In this section we will discuss very concisely the zig-zag process. As discussed in the introduction the zig-zag process is an example of a piecewise deterministic Markov process [Dav84]. As the name indicates, a piecewise deterministic Markov process is a Markov process with deterministic trajectories, in between event times at which the process makes a discontinuous change.

For the (one-dimensional) zig-zag process, the state space is either $E = \mathbb{R} \times \{\pm 1\}$ or $E = \mathbb{T} \times \{\pm 1\}$ and a typical state is denoted in this paper by (x, v). Here x represents a position and v a velocity. Starting from (x, v) at time $t = T_0 := 0$, the dynamics of a Markov process (X_t, V_t) are given, until the first (random) event time $T_1 > 0$, by

$$(X_t, V_t) = (x + tv, v), \quad 0 \leq t < T_1.$$

In other words, the position changes according to the constant velocity v, which itself does not change in between event times. The random time T_1 at which the first event happens is distributed according to

$$\mathbb{P}_{x,v}(T_1 \geq t) = \exp\left(-\int_0^t \lambda(X_s, V_s)\, ds\right) = \exp\left(-\int_0^t \lambda(x + vs, v)\, ds\right),$$

where $\lambda : E \to [0, \infty)$ is the *event rate*, which is in the case of the zig-zag process also known as the *switching rate*, which we will discuss in more detail below. At an event time T the velocity changes sign and the position remains unchanged:

$$V_{T_1} = -V_{T_1-} \quad \text{and} \quad X_{T_1} = X_{T_1-}.$$

From the time T_1 onward, the process repeats the dynamics described above: for $i = 1, 2, \ldots$

$$X_t = X_{T_{i-1}} + (t - T_{i-1})V_{T_{i-1}}, \quad V_t = V_{T_{i-1}}, \quad T_{i-1} \leq t < T_i,$$

$$\mathbb{P}(T_i \geq t \mid T_{i-1}, X_{T_{i-1}}, V_{T_{i-1}}) = \exp\left(-\int_{T_{i-1}}^t \lambda(X_{T_{i-1}} + sV_{T_{i-1}}, V_{T_{i-1}})\, ds\right),$$

$$X_{T_i} = X_{T_i-}, \quad V_{T_i} = -V_{T_i-}.$$

The switching rate $\lambda : E \to \mathbb{R}$ is assumed to be continuous. If λ satisfies

$$\lambda(x, 1) - \lambda(x, -1) = U'(x), \tag{3}$$

for a continuously differentiable function U, then the measure $\pi(dx, dv) = \exp(-U(x))\, dx \otimes \mathrm{Unif}_{\pm 1}(dv)$ is a stationary measure for (X_t, V_t). An equivalent condition to (3) is that, for some continuous non-negative function $\gamma(x)$ we have

$$\lambda(x, v) = \max(0, vU'(x)) + \gamma(x). \tag{4}$$

Here $(\max 0, vU'(x))$ is called the *canonical switching intensity*, and γ is called the *excess switching intensity* or *refreshment rate*. We study the dependence of the empirical measure of the process (X_t, V_t) on γ in Section 3.3.

The zig-zag process can be extended in a natural way to a multi-dimensional process in $\mathbb{R}^d \times \{\pm 1\}^d$; see [BFR19, BRZ19]. Since we focus in this paper on properties of the one-dimensional process we will not discuss this extension here. The ergodic properties of the zig-zag process are essential in order to establish a large deviations theory for the empirical measure. Under mild conditions it can

be shown that the zig-zag process is (exponentially) ergodic; see [BR17] for the one-dimensional case, and [BRZ19] for the multi-dimensional zig-zag process.

By [Dav93, Theorem 26.14], the extended generator of the zig-zag process is given by

$$Bf(x, v) = v\partial_x f(x, v) + \lambda(x, v)[f(x, -v) - f(x, v)], \quad (x, v) \in E,$$

with

$$\mathcal{D}(B) = \{f : E \to \mathbb{R} : f(\cdot, v) \text{ is absolutely continuous for } v = \pm 1\}.$$

3. Large deviations for empirical measures of PDMPs

In this section we present our main results: we establish a large deviations principle for the empirical measure of a Markov process under fairly general assumptions which include in particular examples of position-velocity PDMPs such as the zig-zag process. After obtaining these general results we focus for concreteness on the zig-zag process, for which we verify the stated assumptions. We also give an explicit characterisation of the corresponding rate function, a necessary step towards using the LDP for analysing the performance and properties of approximations based on the zig-zag process. To streamline the presentation we split the analysis according to whether we consider a compact or non-compact state space E.

To facilitate the proof of the LDP for the empirical measures, we first formulate in Section 3.1 two more general large deviations results (compact and non-compact setting) for empirical measures arising from certain continous-time stochastic processes. We then show that the zig-zag process is a special case in this class of processes in Section 3.2. It is worth to emphasise that we do not aim for greatest generality in the large deviations results Theorems 3.1 and 3.2. Rather, we settle for conditions that make the general conditions of Lemma 4.3 more transparent and concrete whilst still allowing us to prove the large deviations principle for the empirical measures of the zig-zag process.

3.1. Results in an abstract setting aimed at position-velocity PDMPs

Before we specialize to the zig-zag process, we consider the setting described in Section 2.1 to PDMPs: Y is a Markov process taking values in a locally compact complete separable metric space E, with associated semigroup $S(t)$ and infinitesimal generator L. We also make use of the extended generator B; see [EK09] and Section 2.3. Typically, $E = \mathbb{R}^d \times \mathcal{S}$ where $\mathbb{R}^d$ is the state space for a position variable X_t and $\mathcal{S}$ is a compact set that models the state space of the velocity variables V_t. For the zig-zag process, $\mathcal{S} = \{\pm 1\}^d$, and for the Bouncy Particle Sampler $\mathcal{S}$ can be taken to be the $(d-1)$-dimensional unit sphere. Note that for $d = 1$ these two choices coincide.

The following are the assumptions we will impose in order to establish an LDP for the empirical measures of the process Y. Not all conditions are required at the same time: we impose conditions (A.1), (A.2) and (A.3) for the compact case and (A.1), (A.2), (A.4) and (A.5) for the non-compact case.

(A.1) The semigroup $S(t)$ is a Feller semigroup.

(A.2) For any compact set $K \subseteq E$, the set of measures $\{\mathbb{P}_y : y \in K\}$ is tight in $\mathcal{P}(D_E[0,\infty))$.

(A.3) For any function $V \in C(E)$, there exists a function $u \in \mathcal{D}^+(L)$ and a real eigenvalue $\beta \in \mathbb{R}$ such that pointwise on E,

$$(V + L)u = \beta u.$$

(A.4) There exist two non-negative functions $g_1, g_2 \in C(E; [0,\infty))$ such that:

    (a) For any $\ell \geq 0$, the sublevel-sets $\{g_i \leq \ell\}$ are compact and $g_i(y) \to \infty$ as $|y| \to \infty$,

    (b) $g_1(y)/g_2(y) \to 0$ as $|y| \to \infty$,

    (c) $e^{g_i} \in \mathcal{D}(B)$, for any $c \in \mathbb{R}$ the superlevel-sets $\{y \in E : e^{-g_i(y)}B(e^{g_i})(y) \geq c\}$ are compact, and $e^{-g_1(y)}B(e^{g_1})(y) \to -\infty$ as $|y| \to \infty$,

    where we recall that B is the extended generator of Y. We write $|y_n| \to \infty$ if $d(y_n, z) \to \infty$ for any point $z \in E$.

(A.5) For any two compactly supported probability measures $\nu_1, \nu_2 \in \mathcal{P}_c(E)$, there exist constants $T, M > 0$ and measures $\rho_1, \rho_2 \in \mathcal{P}([0, T])$ such that for all Borel sets $A \subseteq E$,

$$\int_0^T \int_E P(t, y, A)\, d\nu_1(y) d\rho_1(t) \leq M \int_0^T \int_E P(t, z, A)\, d\nu_2(z) d\rho_2(t), \quad (5)$$

    where $P(t, y, dy')$ denotes the transition probabilities associated to Y_t.

Conditions (A.1)-(A.3) are enough to prove Theorem 3.2, the large deviations principle in a compact setting. In this setting conditions (A.1) and (A.2) replace Condition (DV.1) of Theorem 2.1; Condition (A.2) can also be weakened to $\mathbb{P}_{y_n} \to \mathbb{P}_y$ in $\mathcal{P}(D_E[0,\infty))$ whenever $y_n \to y$. Together Conditions (A.1) and (A.2) imply strong continuity of the semigroup S (see e.g. [FK06, Remark 11.22]).

As pointed out in Section 2.2, the processes we have in mind do not satisfy a transitivity condition similar to Condition (DV.2) of Theorem 2.1. In the compact setting this can be replaced by condition (A.3), which corresponds to a principal-eigenvalue problem for the operator $L + V$. In compact settings, such eigenvalue problems can usually be solved if the coefficients of the generator are regular enough. In Section 3.2 we show that this is the case for the zig-zag process taking values in the compact torus.

In the non-compact setting, the eigenvalue problem (A.3) is replaced by conditions (A.4) and (A.5). Condition (A.4) is closely related to the stability conditions assumed in [DV76a] and [DL18]. Because e^{g_1} is unbounded, formally

we have to use the extended generator B instead of the infinitesimal generator L to formulate Condition (A.4)c. The same problem occurs in Condition 2.2 of [DL18]: for a diffusion process Y_t in $\mathbb{R}^d$ satisfying $dY_t = -Y_t dt + dW_t$, the second-order differential operator

$$Bf(x) = \frac{1}{2}\Delta f(x) - x\nabla f(x)$$

acting on $C^2(\mathbb{R}^d)$ is well-defined and is equal to the infinitesimal generator L of the process when restricted to $C_b^2(\mathbb{R}^d)$. With $g_1(x) = \delta|x|^2/2$, the function $e^{-g_1(x)}Be^{g_1}(x)$ goes to minus infinity for δ small enough. A second Lyapunov function is $g_2(x) = \sqrt{1 + |x|^2}$. In the context of the zig-zag process, since E is of the form $E = \mathbb{R}^d \times \{\pm 1\}^d$, using continuous functions that grow to infinity when fixing the velocity variable is sufficient for obtaining compact level sets.

Condition (A.5) plays the role of a transitivity assumption in the non-compact case. While it is feasible to solve a principal-eigenvalue problem for a compact state space, this is much more difficult in the non-compact setting. It would require deriving not only the eigenvalue itself, but also the corresponding eigenfunction on a non-compact space, for which general existence results are not available. In this setting the transitivity condition (DV.2) of Theorem 2.1 is instead partly replaced by the mixing property (A.5). It is a slightly weakened version of Condition B.8 in [FK06], in that it requires the transition probabilities to be comparable only for compactly supported initial conditions $\nu_1, \nu_2 \in \mathcal{P}_c(E)$. This weakening is crucial for the results in this paper, because the stronger condition fails to be true for the zig-zag process if for instance $\nu_1 = \mathcal{N}(0,1)\otimes\text{Unif}_{\pm 1}$ and $\nu_2 = \delta_y$. In that example, while the left-hand side of (5) is in this case positive for any Borel set $A \subseteq E$, the right-hand side can become zero. This is because the zig-zag process has finite speed propagation, so that for arbitrary $T > 0$, if $\text{dist}(A, y) > T$, then the probability of transitioning from y into A is zero. However for compactly supported measures the condition is satisfied. We verify the conditions of Theorem 3.2 for the zig-zag process in Section 3.2.

We are now ready to state the two general large deviations results of this paper, which in Section 3.2 will be used to derive the large deviations principle for the empirical measures of the zig-zag process. We start with the compact setting.

Theorem 3.1. *Let E be compact, $S(t)$ a Markov semigroup acting on $C(E)$ equipped with the supremum norm, and Y_t the corresponding Markov process. Let L be the infinitesimal generator of Y_t, and assume that Y_t solves the associated martingale problem. Suppose Assumptions (A.1), (A.2) and (A.3) hold. Then the empirical measures $\{\eta_t\}_{t>0}$ associated to Y_t satisfy a large deviations principle in $\mathcal{P}(E)$ with rate function $\mathcal{I} : \mathcal{P}(E) \to [0, \infty]$ given by (2).*

Theorem 3.1 remains valid when replacing the eigenvalue-problem condition (A.3) by the mixing condition (A.5). This is because the latter is a weaker condition sufficient for verifying the inequality (14), upon which the proof of the theorem hinges.

The next theorem gives the corresponding large deviations result for the non-compact setting; this is the result we use for proving the large deviations principle for the zig-zag process on $\mathbb{R} \times \{\pm 1\}$ (Theorem 3.4).

Theorem 3.2. *Let $S(t)$ be a Markov semigroup acting on $C_b(E)$ and Y_t the corresponding Markov process. Let L be the infinitesimal generator of Y_t and assume that Y_t solves the associated martingale problem. Assume (A.1), (A.2), (A.4) and (A.5). Then, if $Y_0 \in K$ for some compact set K, the empirical measures $\{\eta_t\}_{t>0}$ associated to the Markov process Y_t satisfy a large deviations principle in $\mathcal{P}(E)$, with rate function $\mathcal{I} : \mathcal{P}(E) \to [0, \infty]$ given by*

$$\mathcal{I}(\mu) = - \inf_{u \in \mathcal{D}^{++}(L)} \int_E \frac{Lu}{u} d\mu.$$

The proofs of Theorems 3.1 and 3.2 are given in Sections 4.1.1 and 4.1.2, respectively.

3.2. The empirical measures of the zig-zag process

Having established the general large deviations results Theorems 3.1 and 3.2, we now specialize to the zig-zag process. Throughout the section, Y_t is used to denote the zig-zag process, $Y_t = (X_t, V_t)$ with X_t and V_t as in Section 2.3. However the state space E will change as we split the large deviations statements for the empirical measures of Y into compact (torus) and non-compact ($\mathbb{R}$) settings. Although 3.2 holds for arbitrary dimension $d \geq 1$, for the zig-zag process we limit ourselves to verifying the conditions for the case $d = 1$. Extending these results to $d > 1$ is substantially more difficult and remains a topic of further research. While conditions (A.1), (A.2), (A.4) hold true, the main challenge is verifying (A.5).

We begin by considering the compact state space $\mathbb{T} \times \{\pm 1\}$. In this case the infinitesemal generator L of the semigroup $S(t)$ is has domain $\mathcal{D}(L) = C^1(\mathbb{T} \times \{\pm 1\}) = \{f \in C(\mathbb{T} \times \{\pm 1\}) : f(\cdot, \pm 1) \in C^1(\mathbb{T})\}$, and takes the form

$$Lf(x, v) = v \partial_x f(x, v) + \lambda(x, v) \left[f(x, -v) - f(x, v) \right], \qquad (6)$$

with λ given by (4). The LDP for the empirical measures associated with Y and this state space is given in Theorem 3.3. We prove this result in Section 4.2.1 by verifying the conditions of Theorem 3.1, the large deviations principle for processes taking values in a compact state space.

Theorem 3.3. *Suppose that $U \in C^2(\mathbb{T})$. Then the family of empirical measures $\{\eta_t\}_{t>0}$ of the zig-zag process taking values in $\mathbb{T} \times \{\pm 1\}$ satisfies a large deviations principle in the limit $t \to \infty$, with rate function $\mathcal{I} : \mathcal{P}(\mathbb{T} \times \{\pm 1\}) \to [0, \infty]$ given by*

$$\mathcal{I}(\mu) = - \inf_{u \in \mathcal{D}^+(L)} \int_{\mathbb{T} \times \{\pm 1\}} \frac{Lu}{u} d\mu.$$

We now move to the setting of a non-compact state space. Specifically, we consider the zig-zag process $Y_t = (X_t, V_t)$ taking values in $\mathbb{R} \times \{\pm 1\}$. As before, L is the generator of this process, i.e. $L : \mathcal{D}(L) \subseteq C_b(\mathbb{R} \times \{\pm 1\}) \to C_b(\mathbb{R} \times \{\pm 1\})$ is a densely defined linear operator, on the set of functions $\{f(\cdot, \pm 1) \in C_b^1(\mathbb{R})\}$ we have the representation

$$Lf(x, v) = v\partial_x f(x, v) + \lambda(x, v) \left[f(x, -v) - f(x, v) \right], \quad f \in \mathcal{D}(L),$$

with $\lambda(x, v) = \max(0, vU'(x)) + \gamma(x)$. To prove the large deviations principle in this non-compact setting we need additional assumptions on the potential function U determining the jump rates.

(B.1) $U(x) \to \infty$ as $|x| \to \infty$ and $U'(x) \to \pm\infty$ as $x \to \pm\infty$,
(B.2) $U'(x)/U(x) \to 0$ as $|x| \to \infty$,
(B.3) $U''(x)/U'(x) \to 0$ as $|x| \to \infty$.

Furthermore, we will assume that there exists a second potential $V \in C^2(\mathbb{R})$ such that:

(C.1) $V(x) \to \infty$ as $|x| \to \infty$ and $V'(x) \to \pm\infty$ as $x \to \pm\infty$,
(C.2) $V(x)/U(x) \to 0$, $U'(x)/V(x) \to 0$ and $V'(x)/U'(x) \to 0$ as $|x| \to \infty$,
(C.3) $U''(x)/V'(x) \to 0$ as $|x| \to \infty$.

In Section 4.2.2, we prove the following Theorem.

Theorem 3.4. *Assume that $U \in C^3(\mathbb{R})$ satisfies (B.1) - (B.3), that there is a function $V \in C^2(\mathbb{R})$ satisfying (C.1) - (C.3) and the function γ in (4) is uniformly bounded by some $\bar{\gamma}$. Suppose that the initial condition Y_0 belongs to a compact set $K \subseteq \mathbb{R} \times \{\pm 1\}$. Then the empirical measures $\{\eta_t\}_{t>0}$ of Y satisfies a large deviations principle on $\mathcal{P}(\mathbb{R} \times \{\pm 1\})$ with speed t and rate function $\mathcal{I} : \mathcal{P}(\mathbb{R} \times \{\pm 1\}) \to [0, \infty]$ given by*

$$\mathcal{I}(\mu) = - \inf_{u \in \mathcal{D}^{++}(L)} \int_{\mathbb{R} \times \{\pm 1\}} \frac{Lu}{u} d\mu.$$

Some comments on the additional assumptions (B.1) - (B.3) and (C.1) - (C.3) are in place. The condition $U \in C^3(\mathbb{R})$ is imposed to allow for an application of Theorem 4 of [BRZ19], which is used to verify that (A.5) holds. The auxiliary potential V is used to find a second Lyapunov function for L that grows slower than U at infinity; roughly speaking, V behaves asymptotically in-between the potential U and its derivative U' as $|x|$ grows. As an example, in the Gaussian case, $U(x) = x^2/2$ satisfies Conditions (B.1) - (B.3), and for any $0 < \kappa < 1$, the potential $V(x) = |x|^{1+\kappa}/(1+\kappa)$ satisfies (C.1) - (C.3). In general, any potential U growing at infinity as $(1 + |x|^2)^{\beta/2}$ with $\beta > 1$ satisfies the conditions, with $0 < \kappa < 1$ such that $\beta - \kappa > 1$ and auxiliary potential $V(x) \sim (1 + |x|^2)^{(\beta-\kappa)/2}$.

3.3. Explicit expression for the rate function

In Theorems 3.3 and 3.4 we establish the LDP for the empirical measures of the zig-zag process taking values in $\mathbb{T} \times \{\pm 1\}$ and $\mathbb{R} \times \{\pm 1\}$, respectively. In

those results the rate function is given on the variational form of the results by Donsker and Varadhan, see Section 2.2. This form follows from the more general large deviations results in Section 3.1 and are not specific to the zig-zag process. Here, we use the properties of the latter to derive a more explicit form of the rate function for the case $E = \mathbb{T} \times \{\pm 1\}$, taking a first step towards using it as a tool for analysing the corresponding simulation algorithms.

We assume throughout that $E = \mathbb{T} \times \{\pm 1\}$ and the switching rate $\lambda(x, v)$ satisfies $\lambda(x, v) > 0$ for all $(x, v) \in E$. This does not include the canonical rates $\lambda(x, v) = \max(0, vU'(x))_+$; however at the end of this section we present a formal expression for this case.

Define a reference measure ν_0 on E by $\nu_0(dx, dv) = \text{Leb}(dx) \otimes \text{Unif}_{\pm 1}(dv)$. For any function $f : E \to \mathbb{R}$ we write $f^+(x) := f(x, +1)$ and $f^-(x) := f(x, -1)$. Recall the arcsinh function,

$$\text{arcsinh}(\xi) = \log\left(\xi + \sqrt{\xi^2 + 1}\right), \quad \xi \in \mathbb{R}.$$

The proofs of the following results are given in Section 4.3.

Proposition 3.5. *Suppose* $\mu(dx, dv) = \rho(x, v)\nu_0(dx, dv)$ *for a continuously differentiable function* $\rho : E \to [0, \infty)$. *If* $\frac{d\rho^+}{dx}(x) = \frac{d\rho^-}{dx}(x)$ *and* ρ^+, ρ^- *are strictly positive for all* $x \in \mathbb{T}$ *then the Donsker-Varadhan functional is given by*

$$\mathcal{I}(\mu) = \int_{\mathbb{T}} \left\{ \tfrac{1}{2}\rho' \log\left(\frac{\lambda^+ \rho^+}{\lambda^- \rho^-}\right) + \rho' \, \text{arcsinh}\left(\frac{\rho'}{2\sqrt{\lambda^+ \lambda^- \rho^+ \rho^-}}\right) \right.$$
$$\left. - \sqrt{4\lambda^+ \lambda^- \rho^+ \rho^- + (\rho')^2} + \lambda^+ \rho^+ + \lambda^- \rho^- \right\} \, dx. \qquad (7)$$

If $\rho^+ \geq 0$ *and* $\rho^- \geq 0$ *are constant, then*

$$\mathcal{I}(\mu) = \int_{\mathbb{T}} \left(\sqrt{\lambda^+ \rho^+} - \sqrt{\lambda^- \rho^-}\right)^2 \, dx.$$

If $\frac{d\rho^+(x)}{dx}(x) \neq \frac{d\rho^-(x)}{dx}$ *for some* $x \in \mathbb{T}$ *then* $\mathcal{I}(\mu) = \infty$.

Note that if $\mu(dx, dv) = \rho(x, v)\nu_0(dx, dv)$ and $\frac{d\rho^+}{dx} = \frac{d\rho^-}{dx}$ on $\mathbb{T}$, then for some constant $c \in \mathbb{R}$ and a probability density function ρ on $\mathbb{T}$ we have $\rho^+(x, v) = \rho(x) + c$ and $\rho^-(x, v) = \rho(x) - c$.

A useful application of the rate function $\mathcal{I}(\mu)$ is in estimating deviations of ergodic averages, which typically requires the computation of $\inf_{\mu \in \mathcal{P}} \left(\mathcal{I}(\mu) - \int_E V \, d\mu\right)$. If V does not depend on v, then by the following result we can safely assume $c = 0$ and thus restrict the minimization problem to minimization over probability densities on $\mathbb{T}$.

Proposition 3.6. *Let* $\rho \in C^1(\mathbb{T})$ *be a strictly positive probability density function on* $\mathbb{T}$. *Let* $|k| := \inf_x \rho(x)$. *Consider the one-parameter family of probability measures* $(\mu_c)_{c \in (-k, +k)} \in \mathcal{P}_{eq}$ *with probability density functions* $\rho_c : E \to (0, \infty)$ *given by*

$$\rho_c(x, +1) = \rho(x) + c, \quad \rho_c(x, -1) = \rho(x) - c, \quad c \in (-k, +k).$$

Then $c \mapsto \mathcal{I}(\mu_c)$ is minimized at $c = 0$. Furthermore, for $\mu = \mu_0$,

$$\mathcal{I}(\mu) = \int_{\mathbb{T}} \left\{ \tfrac{1}{2}\rho' \log\left(\frac{\lambda^+}{\lambda^-}\right) + \rho' \operatorname{arcsinh}\left(\frac{\rho'}{2\rho\sqrt{\lambda^+\lambda^-}}\right) \right.$$
$$\left. - \sqrt{4\lambda^+\lambda^-\rho^2 + (\rho')^2} + (\lambda^+ + \lambda^-)\rho \right\} dx. \tag{8}$$

We will specialize to the case in which $\rho = \rho^+ = \rho^-$ and use the representation $\rho = \exp(-W)$, where $W \in C^1(\mathbb{T})$. We then find

$$\mathcal{I}(\mu) = \int_{\mathbb{T}} \left\{ -\tfrac{1}{2}W' \log\left(\frac{\lambda^+}{\lambda^-}\right) + W' \operatorname{arcsinh}\left(\frac{W'}{2\sqrt{\lambda^+\lambda^-}}\right) \right.$$
$$\left. - \sqrt{4\lambda^+\lambda^- + (W')^2} + \lambda^+ + \lambda^- \right\} \exp(-W)\, dx. \tag{9}$$

Let

$$\lambda^+(x) = \gamma + \max(0, U'(x)) \quad \text{and} \quad \lambda^-(x) = \gamma + \max(0, -U'(x)), \tag{10}$$

where $\gamma > 0$ is constant, so that $\lambda^\pm$ satisfy (4) and hence the measure with ν_0-density $\exp(-U(x))$ is invariant. We call γ the *excessive switching intensity* or *refreshment rate*.

We can now investigate the dependence of the rate function $\mathcal{I}$, through the expression (9), on γ. The derivative of the integrand of (9) with respect to γ can be computed to be

$$\left(\frac{4\lambda^+\lambda^- + (\lambda^+ - \lambda^-)W' - (\lambda^+ + \lambda^-)\sqrt{4\lambda^-\lambda^+ + (W')^2}}{2\lambda^+\lambda^-} \right) \exp(-W), \tag{11}$$

which is non-positive, and zero only if $W' = \lambda^+ - \lambda^- = U'$ (which can be seen by maximizing with respect to W'). It follows that $\mathcal{I}(\mu)$ is strictly decreasing as a function of γ (for μ not equal to the stationary measure). In other words, for a smaller refreshment rate γ, the rate function increases, so that the convergence of empirical averages to equilibrium is faster.

Suppose that $\nu_0(\{x \in \mathbb{T} : U'(x) = 0\}) = 0$, i.e. the set of points where the derivative of U vanishes is ν_0-negligible. In the (formal) limit as $\gamma \downarrow 0$ in (10), we obtain the following expression for $\mathcal{I}(\mu)$:

$$\mathcal{I}(\mu) = \begin{cases} \int_{\mathbb{T}} \left\{ |W'| \left(\log\left(\frac{W'}{U'}\right) - 1 \right) + |U'| \right\} \exp(-W)\, dx & \text{if } \operatorname{sign}(W') \equiv \operatorname{sign}(U'), \\ \infty & \text{otherwise.} \end{cases} \tag{12}$$

4. Proofs

4.1. Proofs of the general large deviations Theorems 3.1 and 3.2

In this section we give the proofs of Theorems 3.1 and 3.2, which are used to obtain the large deviations principle for the empirical measures of the zig-zag

process. The case of a compact state space is treated in Section 4.1.1 and the non-compact case in Section 4.1.2. Before we embark on these proofs we outline the overall strategy; a more detailed description can be found in the book by Feng and Kurtz in [FK06].

Consider the empirical measure

$$\eta_t(\cdot) = \frac{1}{t}\int_0^t \delta_{Y_s}(\cdot)ds.$$

With a change of variable $s \mapsto ts$ in the integral we can express this as

$$\eta_t(\cdot) = \int_0^1 \delta_{Y_{st}}(\cdot)ds,$$

the empirical measure for the sped-up process (we can think of $t > 1$) $Y_s^t = Y_{st}$ over the time interval $[0, 1]$; in fact we will use $t = n \in \mathbb{N}_+$ below. We can consider the empirical measure of this time-scaled process Y_t on time intervals of lengths other than unity: for $\tau > 0$ define η_t^τ as

$$\eta_t^\tau(\cdot) = \int_0^\tau \delta_{Y_{st}}ds.$$

This empirical measure is viewed as an element of $\mathcal{L}(E)$, the set of Borel measures on $E \times [0, \infty)$ of the form $d\rho(x, s) = \mu_s(dx)ds$, $\mu_s \in \mathcal{P}(E)$ (see Section 2.1). Any such $\rho \in \mathcal{L}(E)$ defines a continuous path $t \mapsto \rho_t = \rho(\cdot \times [0, t]) \in \mathcal{M}_f(E)$ and for $t = 1$ this is a probability measure.

The strategy for proving the large deviations principle for $\{\eta_t\}$ is to first show that $\{\eta_t^\tau\}$ satisfies a large deviations principle in $\mathcal{L}(E)$. We can then use the fact that projections are continuous maps on $\mathcal{L}(E)$ (Lemma 4.2) and an application of the contraction principle to obtain the sought-after large deviations principle on $\mathcal{P}(E)$. This is summarised in the following proposition.

Proposition 4.1. *Suppose that the family $\{\eta_t^\tau\}_{t>0}$ satisfies a large deviations principle in $\mathcal{L}(E)$ with rate function $\mathcal{J} : \mathcal{L}(E) \to [0, \infty]$ given by*

$$\mathcal{J}(\rho) = \int_0^\infty \mathcal{I}(\mu_s)ds, \quad for \ \rho_t = \int_0^t \mu_s ds,$$

where $\mathcal{I} : \mathcal{P}(E) \to [0, \infty]$ is the rate function appearing in the Donsker-Varadhan results,

$$\mathcal{I}(\mu) = -\inf_{u \in \mathcal{D}^{++}(L)} \int_E \frac{Lu}{u}d\mu.$$

Then η_t satisfies a large deviations principle in $\mathcal{P}(E)$ with rate function $\mathcal{I}$.

Proof. Because $\{\eta_t^\tau\}_{t>0}$ satisfies a large deviations principle on $\mathcal{L}(E)$ and the projection $\pi_1 : \mathcal{L}(E) \to \mathcal{P}(E)$ given by

$$\pi_1(\rho) = \rho_1,$$

is continuous (Lemma 4.2), by the contraction principle the sequence $\{\eta_t^1\}_{t>0}$ satisfies a large deviation principle on $\mathcal{P}(E)$ with rate function $\tilde{\mathcal{I}} : \mathcal{P}(E) \to [0, \infty]$ given by

$$\tilde{\mathcal{I}}(\nu) = \inf \left\{ \mathcal{J}(\rho) = \int_0^\infty \mathcal{I}(\mu_s)\, ds : \rho_t = \int_0^t \mu_s\, ds \in \mathcal{L}(E), \rho_1 = \nu \right\}.$$

It remains to show that $\tilde{\mathcal{I}}(\nu) = \mathcal{I}(\nu)$ for every $\nu \in \mathcal{P}(E)$. First, in the integral defining $\mathcal{J}(\rho)$, the integrand is always positive after time $t = 1$. It is therefore enough to consider only integrating to time $t = 1$ in the infimum, as we are free to chose the the form of ρ after that time. Thus,

$$\tilde{\mathcal{I}}(\nu) = \inf \left\{ \int_0^1 \mathcal{I}(\mu_t)dt : \rho_t = \int_0^t \mu_s ds \in \mathcal{L}(E),\ \rho_1 = \nu \right\}.$$

For a fixed $\nu \in \mathcal{P}(E)$, take any $\rho_t = \int_0^t \mu_s ds \in \mathcal{L}(E)$ such that $\rho_1 = \nu$. The rate function $\mathcal{I}$ is convex on $\mathcal{P}(E)$ and by Jensen's inequality we have

$$\mathcal{I}(\nu) = \mathcal{I}(\rho_1)$$
$$= \mathcal{I}\left(\int_0^1 \mu_s ds \right)$$
$$\leq \int_0^1 \mathcal{I}(\mu_s)\, ds.$$

Taking the infimum over all such $\rho \in \mathcal{L}(E)$ yields the inequality

$$\mathcal{I}(\nu) \leq \tilde{\mathcal{I}}(\nu),$$

The constant path $\mu_s = \nu$ gives equality and we have that $\tilde{\mathcal{I}} = \mathcal{I}$ as functionals on $\mathcal{P}(E)$. □

Lemma 4.2 (Projection is continuous). *Let $\mathcal{L}(E)$ be the above space with the topology of weak convergence on bounded time intervals. Let $\mathcal{P}(E)$ be equipped with the weak topology. Then the projection $\pi_1 : \mathcal{L}(E) \to \mathcal{P}(E)$ defined by $\pi_1(\rho) := \rho_1$ is a continuous map.*

Proof of Lemma 4.2. Let $\rho^n \to \rho$ in $\mathcal{L}(E)$. We need to prove that for any bounded and continuous function g on E, we have

$$\int_E g(u)d\rho_1^n(u) \to \int_E g(u)d\rho_1.$$

Since

$$\int_E g(u)d\rho_1^n(u) = \int_{E\times[0,1]} g(u)d\rho^n(u,s),$$

and $\varphi(u,s) = g(u)$ is continuous and bounded on $E \times [0, \infty)$, this is implied by $\rho^n \to \rho$. □

Armed with Proposition 4.1, one way to prove Theorems 3.1 and 3.2 is to prove the large deviations principle for the empirical measures of the associated sped-up versions of underlying processes and apply the proposition. This is the approach we take and we rely on the following result from [FK06] for proving the large deviations principles on $\mathcal{L}(E)$.

Lemma 4.3 (Theorem 12.7 of [FK06]). *Suppose that the following conditions hold:*

(FK.1) The martingale problem for L is well-posed.

(FK.2) The semigroup S is Feller-continuous.

(FK.3) The semigroup S is buc-continuous.

(FK.4) There is an index set $\mathcal{Q}$ and a family of subsets of E, $\{\tilde{K}_n^q \subset E : q \in \mathcal{Q}\}$, such that for $q_1, q_2 \in \mathcal{Q}$, there exists $q_3 \in \mathcal{Q}$ with $\tilde{K}_n^{q_1} \cup \tilde{K}_n^{q_2} \subset \tilde{K}_n^{q_3}$, and for every $y \in E$, there exists $q \in \mathcal{Q}$ such that $\lim_{n \to \infty} d(y, \tilde{K}_n^q) = 0$. Moreover, for each $q \in \mathcal{Q}$, $T > 0$ and $a > 0$, there exists a $\hat{q}(q, a, T) \in \mathcal{Q}$ satisfying

$$\limsup_{n \to \infty} \sup_{y \in \tilde{K}_n^q} \frac{1}{n} \log \mathbb{P}_y \left(Y_t \notin \tilde{K}_n^{\hat{q}(q,a,T)}, \text{ some } t \leq nT \right) \leq -a.$$

(FK.5) There exists an upper semicontinuous function Ψ on E, $\{\varphi_n\} \subset \mathcal{D}^{++}(B_0)$, and $q_0 \in \mathcal{Q}$ such that Ψ is bounded above, $\{y \in E : \Psi(y) \geq c\}$ is compact for each $c \in \mathbb{R}$, $0 < \inf_{y \in K_n^{q_0}} \varphi_n(y) < 2 \inf_{y \in E} \varphi_n(y)$, $\inf_{n,y \in E} \varphi_n(y) > 0$,

$$\lim_{n \to \infty} \frac{1}{n} \log \|\varphi_n\| = 0, \quad \sup_{n,y} \frac{L\varphi_n(y)}{\varphi_n(y)} < \infty,$$

and for each $q \in \mathcal{Q}$,

$$\lim_{n \to \infty} \sup_{y \in \tilde{K}_n^q} \left(\frac{L\varphi_n(y)}{\varphi_n(y)} - \Psi(y) \right) \leq 0, \quad q \in \mathcal{Q}.$$

In addition, for each n and $\beta \in (-\infty, 1]$,

$$\lim_{t \to 0} \|S(t)\varphi_n^\beta - \varphi_n^\beta\| = 0.$$

(FK.6) For each $a > 0$ there exists compact K and $q \in \mathcal{Q}$ such that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left(Y_0^n \notin K \cap \tilde{K}_n^q \right) \leq -a.$$

(FK.7) Take $\mathcal{C} \subset C_b(E)$ separating and define, with Ψ as in (FK.5),

$$H_1^{\beta,\Psi} = \inf_{0 < \kappa \leq 1} \inf_{f \in \mathcal{D}^{++}(L)} \sup_{y \in E} \left[\beta(y) \cdot p + (1 - \kappa) \frac{Lf(y)}{f(y)} + \kappa \Psi(y) \right],$$

$$H_2^{\beta,\Psi} = \sup_{\kappa > 0} \sup_{f \in \mathcal{D}^{++}(L)} \inf_{y \in E} \left[\beta(y) \cdot p + (1 + \kappa) \frac{Lf(y)}{f(y)} - \kappa \Psi(y) \right].$$

It holds that $H_1^{\beta,\Psi} \leq H_2^{\beta,\Psi}$, for $\beta \in \mathcal{C}^d$, $d = 1, 2. \ldots$.

Then $\{\eta_n^\tau\}_n$ satisfies the large deviations principle in $C_E[0,\infty)$ with rate function

$$\hat{\mathcal{J}}(\rho) = \int_0^\infty I^\Psi(\rho_s) ds, \ \rho \in \mathcal{L}(E),$$

where

$$I^\Psi(\mu) = -\min\left[\inf_{u \in \mathcal{D}^{++}(L)} \int_E \frac{Lu}{u} d\mu, \int_E \Psi d\mu\right].$$

4.1.1. Proof of Theorem 3.1—compact state-space E

As outlined in the previous section, we can prove Theorem 3.1 by first verifying the conditions of Lemma 4.3 under the given assumptions and then apply Proposition 4.1.

Proof of Theorem 3.1. First, Conditions (FK.1)-(FK.3) follow from the assumption of Feller continuity (A.1) and tightness; see e.g. Remark 11.22 in [FK06].

Next, Conditions (FK.4) and (FK.5) always hold for compact E: take $\varphi_n \equiv 1$, $\Psi \equiv 0$, $Q = \{q\}$ (singleton), and $K_n^q := E$ for every $n \in \mathbb{N}$. For this choice both conditions are met—it is only for non-compact spaces E that these conditions become non-trivial (see the proof of Theorem 3.2). Condition (FK.6) is trivially true for compact E.

Remains to verify the inequality $H_1^\beta \leq H_2^\beta$. Take $d \geq 1$ and $\beta \in \mathcal{C}^d$. With the choice $\Psi \equiv 0$ the definitions of $H_i^\beta : \mathbb{R}^d \to \mathbb{R}$, $i = 1, 2$, become

$$H_1^\beta = \inf_{f \in \mathcal{D}^{++}(L)} \sup_{y \in E}\left[\beta(y) \cdot p + \frac{Lf(y)}{f(y)}\right],$$

$$H_2^\beta = \sup_{f \in \mathcal{D}^{++}(L)} \inf_{y \in E}\left[\beta(y) \cdot p + \frac{Lf(y)}{f(y)}\right].$$

We now show that the required inequality follows from Assumption (A.3), solvability of the principal eigenvalue problem.

For any $\beta \in \mathcal{C}^d$ and $p \in \mathbb{R}^d$, define the map $V_p(y) : E \to \mathbb{R}$ as

$$V_p(y) = \beta(y) \cdot p$$

This is a continuous function on E and for every p there exists a function $f_p \in \mathcal{D}^+(L)$ and real eigenvalue λ_p such that

$$(L + \beta \cdot p)f_p = \lambda_p f_p.$$

It follows that, for any $p \in \mathbb{R}^d$, we have

$$\lambda_p = \sup_{y \in E}\left[\frac{Lf_p(y)}{f_p(y)} + \beta(y) \cdot p\right] = \inf_{y \in E}\left[\frac{Lf_p(y)}{f_p(y)} + \beta(y) \cdot p\right],$$

which leads to the upper bound

$$
\begin{aligned}
H_1^\beta(p) &= \inf_{f\in\mathcal{D}^+(L)} \sup_{y\in E} \left[\frac{Lf(y)}{f(y)} + \beta(y)\cdot p \right] \\
&\leq \sup_{y\in E} \left[\frac{Lf_p(y)}{f_p(y)} + \beta(y)\cdot p \right] \\
&= \inf_{y\in E} \left[\frac{Lf_p(y)}{f_p(y)} + \beta(y)\cdot p \right] \\
&\leq \sup_{f\in\mathcal{D}^+(L)} \inf_{y\in E} \left[\frac{Lf(y)}{f(y)} + \beta(y)\cdot p \right] \\
&= H_2^\beta(p).
\end{aligned}
$$

This shows that Condition (FK.7) of Lemma 4.3 follows from (A.3). As a result, in the setting of compact E, Assumptions (A.1) - (A.3) ensure that Lemma 4.3 is applicable. This gives the large deviations principle for the empirical measures associated with sped-up versions of the process Y and Proposition 4.1 transfers this to the empirical measures of the original process. This concludes the proof of the large deviations principle. The form of the rate function is trivially seen to be equal to the prescribed form because of the choice of $\Psi \equiv 0$. $\qquad\square$

4.1.2. Proof of Theorem 3.2—non-compact state space E

As in the proof of Theorem 3.1, we prove large deviations of the family of measures $\{\eta_t^\tau\}_{t>0}$ introduced at the beginning of Section 4.1 by verifying the assumptions of Lemma 4.3. Proposition 4.1 then implies the large deviations principle of the empirical measures $\{\eta_t\}_{t>0}$ with the prescribed rate function. Whereas the conditions of Lemma 4.3 where straightforward to verify in the compact setting of Theorem 3.1, the non-compact case requires more work. Specifically, because we can no longer assume that there is a solution to the principal eigenvalue problem - such an assumption would not allow us to prove the large deviations principle for the zig-zag process - and the state space is no longer compact, (FK.4)-(FK.7) are more difficult to verify. A crucial component of the proof of Theorem 3.2 is an inequality that is connected to the necessary comparison principle. To streamline the proof we now state this inequality as a separate result.

For any $V \in C_b(E)$ and $\Psi : E \to \mathbb{R}$, define $H_1^\Psi, H_2^\Psi \in \mathbb{R}$ by

$$
\begin{aligned}
H_1^\Psi &= \inf_{0<\kappa\leq 1} \inf_{f\in\mathcal{D}^{++}(L)} \sup_{y\in E} \left[V(y) + (1-\kappa)\frac{Lf(y)}{f(y)} + \kappa\Psi(y) \right], \\
H_2^\Psi &= \sup_{\kappa>0} \sup_{f\in\mathcal{D}^{++}(L)} \inf_{y\in E} \left[V(y) + (1+\kappa)\frac{Lf(y)}{f(y)} - \kappa\Psi(y) \right].
\end{aligned}
\tag{13}
$$

Proposition 4.4. *Take any $V \in C_b(E)$ and suppose (A.5) holds and that for any $c \in \mathbb{R}$, the superlevel-set $\{\Psi \geq c\}$ is compact. Then*

$$
H_1^\Psi \leq H_2^\Psi.
\tag{14}
$$

We first complete the proof of Theorem 3.2.

Proof of Theorem 3.2. The proof amounts to showing that Conditions (FK.1)-(FK.7) of Lemma 4.3 hold. We start with the ones that are straightforward to obtain from the assumptions of the theorem.

Conditions (FK.1)-(FK.3) follow from (A.1) and (A.2). For condition (FK.6) the existence of such a compact set follows immediately from the assumption that the initial value $Y(0)$ belongs to a compact set $K \subseteq E$.

We now show that Conditions (FK.4) and (FK.5) follow from (A.4), the existence of Lyapunov functions g_1 and g_2 with certain growth properties. We start with (FK.4) and define the family of compact sets $K_n^q \subseteq E$ by

$$K_n^q = \{y \in E \,:\, g_2(y) \le qn\}, \quad q, n \in \mathbb{N}$$

For any q_1, q_2 and with $q_3 = \max(q_1, q_2)$, it then holds that

$$K_n^{q_1} \cup K_n^{q_2} \subseteq K_n^{q_3}, \quad \forall n \in \mathbb{N}.$$

Because $g_2(y)$ is finite for any $y \in E$, there exists $q, N \in \mathbb{N}$ such that $n \ge N$ implies that $y \in K_n^q$. In particular, $\mathrm{dist}(y, K_n^q) = 0$. For the last part of Condition (FK.4), take $q \in \mathbb{N}$ and $T, a > 0$. It remains to find a $\tilde{q}$ such that

$$\limsup_{n \to \infty} \sup_{y \in K_n^q} \frac{1}{n} \log \mathbb{P}_y \left(Y_t \notin K_n^{\tilde{q}}, \text{ some } t \le nT \right) \le -a.$$

By Lemma 4.20 in [FK06], for any open neighbourhood $\mathcal{O}$ of K_n^q,

$$\mathbb{P} \left(Y_t \notin \mathcal{O}, \text{ some } t \le nT \,|\, Y_0 \in K_n^q \right) \le \mathbb{P} \left(Y_0 \in K_n^q \right) e^{-\beta_q + nT\gamma(\mathcal{O})}, \qquad (15)$$

where the constants β_q and $\gamma(\mathcal{O})$ are given by

$$\beta_q = \inf_{E \setminus \mathcal{O}} g_2 - \sup_{K_n^q} g_2,$$

and

$$\gamma(\mathcal{O}) = \max \left(\sup_{\mathcal{O}} e^{-g_2} B e^{g_2}, 0 \right).$$

By the growth condition for g_2 (part (a) of (A.4)) for any $\tilde{q} > \hat{q} > q$ large enough, there exists an open set $\mathcal{O}$ such that

$$K_n^q \subseteq K_n^{\hat{q}} \subseteq \mathcal{O} \subseteq K_n^{\tilde{q}}.$$

By definition, $g_2 \le nq$ on K_n^q and because $K_n^{\hat{q}} \subseteq \mathcal{O}$, we have $g_2 \ge \hat{q}n$ on $E \setminus \mathcal{O}$. Combined with the upper bound $\gamma(\mathcal{O}) \le \gamma(E)$ this gives, starting from (15),

$$
\begin{aligned}
\frac{1}{n} \log \mathbb{P} \left(Y_t \notin K_n^{\tilde{q}}, \text{ some } t \le nT | Y_0 \in K_n^q \right) &\le T\gamma(\mathcal{O}) - \frac{1}{n}\beta_q \\
&\le T\gamma(E) + q - \hat{q}.
\end{aligned}
$$

The last part of Condition (FK.4) is now straightforward to obtain. First, take $\hat{q} = \hat{q}(q, a, T)$ large enough that the right-hand side of the last display is bounded by $-a$:

$$T\gamma(E) + q - \hat{q} \leq -a.$$

Next, choose $\tilde{q} = \tilde{q}(q, a, T)$ large enough that there is an open set $\mathcal{O}$ such that $K_n^{\hat{q}} \subseteq \mathcal{O} \subseteq K_n^{\tilde{q}}$. The asymptotic statement then follows, which concludes the verification of condition (FK.4) of Lemma 4.3.

To show that condition (FK.5) is fulfilled we generalize the arguments used in Example 11.24 in [FK06]. The functions φ_n are constructed from the Lyapunov functions g_1 and g_2. First, define

$$r_n := \sup\left\{g_1(y) \,:\, y \in E,\, g_1(y)g_2(y) \leq n^2\right\}. \tag{16}$$

Then $r_n \to \infty$ and $r_n/n \to 0$ as $n \to \infty$, since (A.4)b and the condition in the set imply

$$\frac{r_n}{n} = \frac{g_1(y_n)}{n} \leq \sqrt{\frac{g_1(y_n)}{g_2(y_n)}} \to 0.$$

Furthermore, for each q there exists n_q such that $n \geq n_q$ implies

$$K_n^q \subseteq \{y \,:\, g_1(y) \leq r_n\}.$$

For a smooth, non-decreasing and concave function $\rho : [0, \infty) \to [0, 2]$ satisfying $\rho(r) = r$ for $0 \leq r \leq 1$ and $\rho(r) = 2$ for $r \geq 3$, define the functions φ_n by cutting off g_1:

$$\varphi_n(y) := e^{r_n}\rho(e^{-r_n}e^{g_1(y)}). \tag{17}$$

We have $\varphi_n = e^{g_1}$ on the compact sets K_n^q. Setting $\Psi = e^{-g_1}Be^{g_1}$, we therefore obtain

$$\frac{L\varphi_n(y)}{\varphi_n(y)} = \Psi(y), \quad y \in K_n^q,\, n \geq n_q.$$

The fact that $r_n/n \to 0$ as $n \to \infty$ implies $n^{-1}\log\|\varphi_n\| \to 0$ as $n \to \infty$. For proving that

$$\sup_{n,y} \frac{L\varphi_n(y)}{\varphi_n(y)} < \infty,$$

it is sufficient to show that for any positive function $u : E \to (0, \infty)$ in the domain of B and for any $y_0 \in E$, we have

$$\frac{B(\rho(u))(y_0)}{\rho(u)(y_0)} \leq \frac{\max\left(Bu(y_0), 0\right)}{u(y_0)}. \tag{18}$$

Then with $u = e^{-r_n}e^{g_1}$ and noting that $L\varphi_n = B\varphi_n$, by linearity we obtain

$$\frac{L\varphi_n}{\varphi_n} \leq \frac{\max\left(Be^{g_1}, 0\right)}{e^{g_1}},$$

and the result follows since $\Psi(y) = e^{-g_1(y)} \left(Be^{g_1}\right)(y) \to -\infty$ as $|y| \to \infty$. Hence we are left with verifying (18).

If $u(y_0) \in (3, \infty)$, then $\rho(u)(y_0)$ is maximal. Hence by the positive maximum principle, $B\rho(u)(y_0) \leq 0$, and the inequality follows. If $u(y_0) \in (0, 3]$, then $a_0 := \rho'(u(y_0)) \in [0, 1]$, the region where ρ goes from slope one to slope zero. Consider the function $f_0 := \rho(u) - a_0 u$. Since $g_0(s) := \rho(s) - a_0 s$ is maximal for s_0 satisfying $\rho'(s_0) = a_0$, we obtain that y_0 is an optimizer, that is $f_0(y_0) = \sup_y f(y)$. Furthermore, $g_0(s_0) \geq 0$, so by the positive maximum principle $Bf_0(y_0) \leq 0$. By linearity of B and since $r \leq \max(r, 0)$, we obtain the inequality $B\rho(u)(y_0) \leq a_0 \cdot \max\left(Bu(y_0), 0\right)$. Hence

$$\frac{B\rho(u)(y_0)}{\rho(u)(y_0)} \leq \frac{a_0}{\rho(u)(y_0)} \max(Bu(y_0), 0) \leq \frac{1}{u(y_0)} \max(Bu(y_0), 0),$$

using that $0 \leq g_0(s_0) = \rho(u)(y_0) - a_0 u(y_0)$. This finishes the verification of (18).

It remains to show that condition (FK.7) is fulfilled. However, this is precisely the conclusion of Proposition 4.4 - the function $V(y) = \beta(y) \cdot p$ where β is as in condition (vii) is an element of $C_b(E)$, and by (A.4)c, the function Ψ has compact superlevel-sets.

We have shown that under the assumptions of the theorem, all conditions of Lemma 4.3 are fulfilled. The large deviations principle for the empirical measures of the sped-up versions thus holds and Proposition 4.1 then gives the large deviations principle for the empirical measures $\{\eta_t\}$ associated with Y.

We are left with showing that the rate function I^Ψ of Proposition 4.1 satisfies

$$I^\Psi(\mu) = - \inf_{u \in \mathcal{D}^{++}(L)} \int_E \frac{Lu}{u} d\mu.$$

Below, we prove that

$$\limsup_{n \to \infty} \int_E \frac{L\varphi_n}{\varphi_n} d\mu \leq \int_E \Psi d\mu.$$

Then

$$\inf_{u \in \mathcal{D}^{++}(L)} \int_E \frac{Lu}{u} d\mu \leq \inf_n \int_E \frac{L\varphi_n}{\varphi_n} d\mu \leq \int_E \Psi d\mu,$$

and hence the rate function is given by

$$I^\Psi(\mu) = - \min \left[\inf_{u \in \mathcal{D}^{++}(L)} \int_E \frac{Lu}{u} d\mu, \int_E \Psi d\mu \right] = - \inf_{u \in \mathcal{D}^{++}(L)} \int_E \frac{Lu}{u} d\mu.$$

To see that the functions φ_n satisfy the limsup inequality, note that Ψ has compact super-level sets and $\Psi(y) \to -\infty$ as $|y| \to \infty$. Since the compact sets K_n^q exhaust E in the sense that $E = \cup_n K_n^q$ and $K_n^q \subseteq K_{n+1}^q$, there exists a constant $C > 0$ such that

$$f_n = -\frac{L\varphi_n}{\varphi_n} + C \geq 0.$$

Pointwise, we have $f = -\Psi + C = \liminf_n f_n$. Therefore, by Fatou's lemma

$$\liminf_{n\to\infty} \int_E \left[-\frac{L\varphi_n}{\varphi_n} + C \right] d\mu \geq \int_E [-\Psi + C] \, d\mu,$$

and the required limsup inequality follows from reorganizing. $\qquad\square$

We now prove the important Proposition 4.4. The proof is essentially a combination of different arguments from Chapter 11 and Appendix B of [FK06] (see especially Lemmas 11.12, 11.37, B.9-B.11 for full details). We present the proof as to make the presentation self-contained and give a succinct derivation of the results for the setting considered in this paper. The main difference compared to the arguments in [FK06] is that we work with measures $\nu \in \mathcal{P}_c(E)$ rather than imposing the condition $\int_E \Psi\nu > -\infty$, and we must verify that we can indeed modify the latter.

Proof of Proposition 4.4. The strategy is to find two constants, depending on V, c_V^* and c_V^{**} such that $c_V^* \geq c_V^{**}$ and

$$H_1^\Psi \leq c_V^{**}, \text{ and } H_2^\psi \geq c_V^*, \tag{19}$$

To achieve this we study the following quantity: for $\nu \in \mathcal{P}_c(E)$, define

$$c_V(\nu) = \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E} \left[\exp\left\{ \int_0^t V(Y(s))ds \right\} \right].$$

It can be shown - see e.g. Lemma B.9 in [FK06] - that under (A.5), $c_V(\nu)$ exists for each $\nu \in \mathcal{P}_c(E)$ and the necessary inequalities for H_i^Ψ can be derived for

$$c_V^* = \inf_{\nu\in\mathcal{P}_c(E)} c_V(\nu) \quad \text{and} \quad c_V^{**} = \sup_{\nu\in\mathcal{P}_c(E)} c_V(\nu).$$

Cleary $c_V^* \leq c_V^{**}$. However it can be shown, again using (A.5), that the two quantities are in fact equal, that $c_V(\nu)$ is independent of ν on $\mathcal{P}_c(E)$. If we can prove (19) this would then yield the claim.

We start with the upper bound

$$H_1^\Psi \leq c_V^{**}.$$

An argument similar to what will follow is also used in [DV75c], in the proof of their Lemma 2.

Because Ψ has compact superlevel-sets $\{\Psi \geq c\}$, $c \in \mathbb{R}$, and $\Psi(y) \to -\infty$ as $\|y\| \to \infty$, it can be shown using the arguments of Lemma B.11 of [FK06] that

$$\sup_{\nu\in\mathcal{P}_c(E)} \inf_{f\in\mathcal{D}^{++}(L)} \int_E \left(V + \frac{Lf}{f} \right) d\nu \leq c_V^{**}.$$

It therefore suffices to show that

$$H_1^\Psi \leq \sup_{\nu\in\mathcal{P}_c(E)} \inf_{f\in\mathcal{D}^{++}(L)} \int_E \left(V + \frac{Lf}{f} \right) d\nu. \tag{20}$$

For any finite collection of functions $f_1, \ldots, f_m$ in $\mathcal{D}^{++}(L)$ and scalars $\alpha_i \geq 0$, $i = 1, \ldots, m$, $\sum \alpha_i = 1$, we have

$$H_1^{\Psi} \leq \inf_{0 < \kappa \leq 1} \inf_{\alpha_i} \inf_{f_1, \ldots, f_m} \sup_{y \in E} \left[V(y) + (1 - \kappa) \sum_{i=1}^m \alpha_i \frac{Lf_i(y)}{f_i(y)} + \kappa \Psi(y). \right]$$

Define

$$h_t = \frac{1}{t} \int_0^t S(\tau) \prod_{i=1}^m f_i^{\alpha_i} d\tau.$$

Then

$$\lim_{t \to 0} h_t = \prod_i f_i^{\alpha_i},$$

and we have the upper bound

$$\lim_{t \to 0} L h_t \leq \prod_i f_i^{\alpha_i} \sum_i \alpha_i \frac{Lf_i}{f_i},$$

where the convergence is uniform.

We can select a sequence of functions $\{f_i\}$ from $\mathcal{D}^{++}(L)$ such that for any $\mu \in \mathcal{P}(E)$,

$$\inf_i \int_E \frac{Lf_i}{f_i} d\mu = \inf_{f \in \mathcal{D}^{++}(L)} \int_E \frac{Lf}{f} d\mu.$$

Specialising to these functions, for any $m \in \mathbb{N}$ and $\kappa > 0$ we have the upper bound

$$H_1^{\Psi} \leq \inf_{\alpha_i} \sup_{y \in E} \left[V(y) + (1 - \kappa) \sum_{i=1}^m \alpha_i \frac{Lf_i(y)}{f_i(y)} + \kappa \Psi(y). \right]$$

The functions $V + Lf_i/f_i$ are bounded, but a priori there is no guarantee that the supremum is attained in a given compact set. However, because $\Psi(y) \to -\infty$ as $\|y\| \to \infty$, for any $m \in \mathbb{N}$ and $\kappa > 0$, there exists a constant $\ell = \ell(m, \kappa) > 0$ such that the supremum is attained in the compact set $K_\ell = \{\Psi \geq -\ell\}$. Therefore, if we define $\mathcal{K}_\ell = \{\nu \in \mathcal{P}(E) : \nu(K_\ell) = 1\}$, then

$$H_1^{\Psi} \leq \inf_{\alpha_i} \sup_{y \in K_\ell} \left[V(y) + (1 - \kappa) \sum_{i=1}^m \alpha_i \frac{Lf_i(y)}{f_i(y)} + \kappa \Psi(y) \right]$$

$$= \inf_{\alpha_i} \sup_{\nu \in \mathcal{K}_\ell} \left[\int_E \left(V(y) + (1 - \kappa) \sum_{i=1}^m \alpha_i \frac{Lf_i(y)}{f_i(y)} + \kappa \Psi(y) \right) d\nu(y) \right].$$

For any ℓ, we have $\mathcal{K}_\ell \subseteq \mathcal{P}_c(E)$, so that

$$H_1^{\Psi} \leq \inf_{\alpha_i} \sup_{\nu \in \mathcal{P}_c(E)} \left[\int_E \left(V + (1 - \kappa) \sum_{i=1}^m \alpha_i \frac{Lf_i}{f_i} + \kappa \Psi \right) d\nu \right].$$

For any $m \in \mathbb{N}$, the set $\{\alpha_i : \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1\}$ is compact and the infimum and supremum in the last display can be exchanged by Sion's Theorem. This yields

$$H_1^\Psi \leq \sup_{\nu \in \mathcal{P}_c(E)} \inf_{\alpha_i} \left[\int_E \left(V + (1 - \kappa) \sum_{i=1}^m \alpha_i \frac{Lf_i}{f_i} + \kappa \Psi \right) d\nu \right]$$

$$= \sup_{\nu \in \mathcal{P}_c(E)} \left[\int_E \left(V + (1 - \kappa) \min_{i \leq m} \frac{Lf_i}{f_i} + \kappa \Psi \right) d\nu \right],$$

where we have used that $\inf_{\alpha_i} \sum \alpha_i x_i = \min_i x_i$ for non-negative x and $f_i \in \mathcal{D}^{++}(L)$. Taking the infimum over κ and the limit $m \to \infty$,

$$H_1^\Psi \leq \lim_{m \to \infty} \inf_{0 < \kappa \leq 1} \sup_{\nu \in \mathcal{P}_c(E)} \left[\int_E V d\nu + (1 - \kappa) \min_{i \leq m} \int_E \frac{Lf_i}{f_i} d\nu + \kappa \int_E \Psi d\nu \right]$$

$$= \lim_{m \to \infty} \sup_{\nu \in \mathcal{P}_c(E)} \left[\int_E V d\nu + \min \left\{ \min_{i \leq m} \int_E \frac{Lf_i}{f_i} d\nu, \int_E \Psi d\nu \right\} \right].$$

The limit and supremum can be shown to commute similarly to the last part of the proof of Lemma 11.12 in [FK06], leading to

$$H_1^\Psi \leq \sup_{\nu \in \mathcal{P}_c(E)} \left[\int_E V d\nu + \min \left\{ \inf_{f \in \mathcal{D}^{++}(L)} \int_E \frac{Lf}{f} d\nu, \int_E \Psi d\nu \right\} \right]$$

$$\leq \sup_{\nu \in \mathcal{P}_c(E)} \left[\int_E V d\nu + \inf_{f \in \mathcal{D}^{++}(L)} \int_E \frac{Lf}{f} d\nu \right].$$

This completes the proof of the upper bound for H_1^Ψ.

Next, we move to the lower bound for H_2^Ψ. Take $\lambda < c_V^*$. We prove that for any $\varepsilon > 0$, we have $H_2^\Psi \geq \lambda - \varepsilon$. To this end, we define the new semigroup $\{T(t)\}$ by

$$(T(t)f)(y) = \mathbb{E} \left[f(Y_t) e^{\int_0^t V(Y_s) ds} | Y(0) = y \right],$$

set

$$R_\lambda^t f = \int_0^t e^{-\lambda s} T(s) g ds,$$

and take Γ to be the collection of functions f_γ of the form

$$f_\gamma = \int_0^\infty R_\lambda^t 1 \gamma(dt), \quad \gamma \in \mathcal{P}([0, \infty)).$$

Then $\Gamma \subseteq \mathcal{D}^{++}(L)$ and for any $f \in \Gamma$ we have the uniform lower bound

$$V(y) + (1 + \kappa) \frac{Lf(y)}{f(y)} \geq -(1 - 2\kappa) \|V\|, \quad y \in E. \tag{21}$$

Because $\Gamma \subseteq \mathcal{D}^{++}(L)$, for any $\kappa > 0$ we have the lower bound

$$H_2^{\Psi} \geq \sup_{f \in \Gamma} \inf_{y \in E} \left[V(y) + (1 + \kappa) \frac{Lf(y)}{f(y)} - \kappa \Psi(y) \right].$$

Due to the uniform lower bound (21) and the fact that $\Psi(y) \to -\infty$ as $\|y\| \to \infty$, for any κ there exists an $\ell = \ell(\kappa)$ such that the infimum over E is attained in the compact set $K_\ell = \{\Psi \geq -\ell\}$. Therefore,

$$
\begin{aligned}
H_2^{\Psi} &\geq \sup_{f \in \Gamma} \inf_{y \in K_\ell} \left[V(y) + (1 + \kappa) \frac{Lf(y)}{f(y)} - \kappa \Psi(y) \right] \\
&= \sup_{f \in \Gamma} \inf_{\nu \in \mathcal{K}_\ell} \left[\int_E \left(V + (1 + \kappa) \frac{Lf}{f} - \kappa \Psi \right) d\nu \right] \\
&= \sup_{f \in \Gamma} \inf_{\nu \in \mathcal{K}_\ell} \frac{1}{\int_E f d\nu} \left[-\kappa \int_E (V + \Psi) f d\nu + \int_E (1 + \kappa)(V + L) f d\nu \right],
\end{aligned}
$$

where $\mathcal{K}_\ell = \{\nu \in \mathcal{P}(E) : \nu(K_\ell) = 1\}$. The second equality follows from the fact that $\inf_\nu \int_E (a/b) d\nu = \inf_\nu (\int_E a d\nu)/(\int_E b d\nu)$ for $b > 0$. By compactness of $\mathcal{K}_\ell$ and the fact that both $\mathcal{K}_\ell$ and Γ are convex, the infimum and supremum are exchangable by Sion's Theorem. This gives the lower bound

$$
\begin{aligned}
H_2^{\Psi} &\geq \inf_{\nu \in \mathcal{K}_\ell} \sup_{f \in \Gamma} \frac{1}{\int_E f d\nu} \left[-\kappa \int_E (V + \Psi) f d\nu + \int_E (1 + \kappa)(V + L) f d\nu \right] \\
&\geq \inf_{\nu \in \mathcal{P}_c(E)} \sup_{f \in \Gamma} \frac{1}{\int_E f d\nu} \left[-\kappa \int_E (V + \Psi) f d\nu + \int_E (1 + \kappa)(V + L) f d\nu \right],
\end{aligned}
$$

The second estimate follows since $\mathcal{K}_\ell \subseteq \mathcal{P}_c(E)$ for any ℓ. The rest of the proof follows arguments similar to those used in [FK06]: taking the limit $\kappa \to 0$, and moving it inside the infimum and supremum, we obtain the lower bound

$$H_2^{\Psi} \geq \inf_{\nu \in \mathcal{P}_c(E)} \sup_{f \in \Gamma} \left[\frac{1}{\int_E f d\nu} \int_E (V + L) f d\nu \right].$$

Therefore, for any ε, there exists a $\nu_\varepsilon \in \mathcal{P}_c(E)$ such that

$$H_2^{\Psi} \geq \sup_{f \in \Gamma} \left[\frac{1}{\int_E f d\nu_\varepsilon} \int_E (V + L) f d\nu_\varepsilon \right] - \varepsilon.$$

There exist functions $f_t \in \Gamma$ satisfying

$$\frac{\int_E (V + L) f_t d\nu}{\int_E f_t d\nu} = \lambda + \frac{\int_E e^{-\lambda t} T(t) 1 d\nu - 1}{\int_E f_t d\nu},$$

for any $\nu \in \mathcal{P}_c(E)$. Specialising to such f_t, we obtain

$$H_2^{\Psi} \geq \lambda + \frac{\int_E e^{-\lambda t} T(t) 1 d\nu_\varepsilon - 1}{\int_E f_t d\nu_\varepsilon} - \varepsilon.$$

Since $\limsup_{t\to\infty} \int_E e^{-\lambda t} T(t) 1 d\nu_\varepsilon = \infty$, the second term is positive for t large enough, giving the bound

$$H_2^\Psi \geq \lambda - \varepsilon.$$

This completes the proof of the lower bound for H_2^Ψ, and thereby the lemma. $\quad\square$

4.2. Large deviations proofs for the empirical measure of the zig-zag process

In this section, we prove the large deviations theorems for the empirical measures of the zig-zag process.

4.2.1. Proof of Theorem 3.3—compact case

For the proof of Theorem 3.3, recall that the zig-zag generator takes the form

$$Lf(x,v) = v\partial_x f(x,v) + \lambda(x,v)\left[f(x,-v) - f(x,v)\right], \quad (x,v) \in E := \mathbb{T} \times \{\pm 1\}.$$

It is enough to show that assumptions (A.1)-(A.3) hold for the zig-zag process on E, the result then follows form Theorem 3.1.

We first verify that L is a closed operator that generates the zig-zag process. Note that L is a restriction of the extended generator (see Section 2.3). We verify that L is a closed operator. Let $\{f_n\}$ be a sequence in $\mathcal{D}(L)$ such that $f_n \to f$ and $Lf_n \to g$, some f, g, both uniformly on E. Then

$$\lim_{n\to\infty} v\partial_x f_n(x,v) = g(x,v) - \lambda(x,v)\left[f(x,-v) - f(x,v)\right].$$

We can represent $f_n(x,v)$ as

$$f_n(x,v) = f_n(0,v) + v\int_0^x v\partial_x f_n(\xi,v)\, d\xi,$$

and from the dominated convergence theorem we obtain that

$$f(x,v) = f(0,v) + v\int_0^x \left[g(\xi,v) - \lambda(\xi,v)\left(f(\xi,-v) - f(\xi,v)\right)\right] d\xi.$$

In particular, $f \in \mathcal{D}(L)$, and $Lf = g$ follows from taking derivative ∂_x and multiplying by v.

The Feller-continuity property (A.1) of the zig-zag semigroup S is proven in Proposition 4 of [BR17]. Since $\mathbb{T}$ is compact, this also follows from the boundedness of the continuous rates λ, see [Dav93, Theorem 27.6].

It remains to verify assumption (A.3), the principal-eigenvalue problem. Take $V \in C(E)$. We will show that for any constant $\gamma > \sup_E V$, as a map from $C(E)$ to $\mathcal{D}(L) \subseteq C(E)$, the resolvent

$$R_\gamma = (\gamma - (V + L))^{-1}, \tag{22}$$

is compact and strongly positive; here strongly positive means that if $f \geq 0$ and $f \neq 0$, then $R_\gamma f > 0$ on E. Given strong positivity and compactness, by the Krein-Rutman theorem there exists a strictly positive function $g \in C(E)$ and a real eigenvalue $\beta > 0$ such that

$$(\gamma - (V + L))^{-1} g = \beta g.$$

The resolvent maps into the domain of L, so that $g \in \mathcal{D}(L)$. An application of $\gamma - (V + L)$ in the eigenvalue equation gives

$$(V + L)g = \left(\gamma - \frac{1}{\beta}\right) g.$$

This is precisely (A.3) with function g and eigenvalue $(\gamma - 1/\beta)$.

We are left with verifying that the resolvents defined by (22) are strongly positive and compact. For strong positivity, because V is continuous on E, it is sufficient to prove strong positivity of $(\gamma - L)^{-1}$; see [AGG$^+$86, Proposition C-III-3.3]. The resolvent $(\gamma - L)^{-1}$ exists for any $\gamma > 0$, and is given by

$$(\gamma - L)^{-1} f = \int_0^\infty e^{-\gamma t} S(t) f \, dt. \tag{23}$$

The semigroup associated to the zig-zag process is irreducible in the following sense: for any $f \in C(E)$ such that $f \geq 0$ and $f \neq 0$,

$$\cup_{t \geq 0} \left\{ z \in E \,:\, S(t)f(z) > 0 \right\} = E.$$

Combined with (23) this implies strong positivity of $(\gamma - L)^{-1}$; see [AGG$^+$86, Definition C-III-3.1].

For compactness of R_γ, let $A \subseteq C(E)$ be bounded. We show that the image $B := R_\gamma(A) \subseteq C(E)$ is bounded and equi-continuous. Compactness of the resolvent then follows from an application of the Arzel-Ascoli theorem. To show boundedness, by dissipativity of L we obtain, for any $g \in B$,

$$\begin{aligned}
(\gamma - \|V\|_E)\|g\| &\leq \|(\gamma - (V + L))g\| \\
&\leq \sup_{f \in A} \|f\| \\
&< \infty.
\end{aligned}$$

Hence B is bounded by $C_A/(\gamma - \|V\|_E)$, where $C_A := \sup_{f \in A} \|f\|$, and we end the proof by showing that B is equi-continuous. For any $g \in B$ we have $(\gamma - (V + L))g = f$ for some $f \in A$, which implies

$$v\partial_x g(x, v) = f(x, v) + V(x, v)g(x, v) + \gamma g(x, v) - \lambda(x, v)(g(x, -v) - g(x, v)).$$

By boundedness of the functions $\lambda(x, v)$ and V on E and the sets A and B,

$$\sup_{g \in B} \|\partial_x g\| \leq C \sup_{g \in B} \|g\| + \sup_{f \in A} \|f\| < \infty.$$

Hence functions in B have uniformly bounded derivatives, and as a consequence, B is equi-continuous. It follows that R_γ in (22) is compact and strongly continuous. This finishes the verification of (A.3) and we have shown that assumptions (A.1)-(A.3) hold for the zig-zag process on the compact state space $\mathbb{T} \times \{\pm 1\}$. An application of Theorem 3.1 then proves the claimed large deviations principle.

4.2.2. *Proof of Theorem 3.4—non-compact case*

For notational simplicity we take $E = \mathbb{R} \times \{\pm 1\}$. Similar to the proof of Theorem 3.3, the strategy is to verify the conditions of the more general large deviations result Theorem 3.2, which covers the non-compact setting. That is, it suffices to verify (A.1), (A.2), (A.4) and (A.5).

Condition (A.1), Feller-continuity of the Markov semigroup, is proven in Proposition 4 of [BR17].

Next, we use Theorem 7.2 of [EK09] to verify (A.2). Define the metric d on E as

$$d((x,v),(y,v')) = |x - y|_\mathbb{R} + |v - v'|,$$

and for any path $\gamma \in D_E[0,\infty)$ set

$$w'(\gamma, \delta, T) = \inf_{\{t_i\}} \max_i \sup_{s,t \in [t_i, t_{i+1})} d(\gamma(s), \gamma(t)),$$

where the infimum is taken over finite partitions $\{t_i\}$ of $[0, T]$ such that $\min_i |t_{i+1} - t_i| > \delta$. Theorem 7.2 of [EK09] states that tightness of $\{\mathbb{P}_y : y \in K\}$ is equivalent to the following two conditions:

(1) For any $\varepsilon > 0$ and rational $t > 0$, there exists a compact set $K_{\varepsilon,t} \subseteq E$ such that

$$\inf_{y \in K} \mathbb{P}_y [Y_t \in K_{\varepsilon,t}] \geq 1 - \varepsilon.$$

(2) For any $\varepsilon > 0$ and $T > 0$, there exists a $\delta > 0$ such that

$$\sup_{y \in K} \mathbb{P}_y [w'(Y, \delta, T) \geq \varepsilon] \leq \varepsilon.$$

The spatial component X_t of the zig-zag process propagates with finite speed. This implies that there exists a compact set $K_t \subseteq \mathbb{R}$ such that if $y \in K$, then

$$\mathbb{P}_y [X_t \in K_t] = 1.$$

For any $\varepsilon > 0$ and $t > 0$, taking $K_{\varepsilon,t} = K_t \times \{\pm 1\}$ gives (1).

For part (2), let $\varepsilon > 0$ and $T > 0$. For any realization $Y(\omega)$ of the zig-zag process on the time interval $[0, T]$, if the sojourn times τ_i satisfy $\min_i \tau_i > 2\delta$, then $w'(Y(\omega), \delta, T) \leq 2\delta$. In particular, for δ small enough, $w'(Y(\omega), \delta, T) < \varepsilon$. The probability of having at least one sojourn time that is less than 2δ can be

estimated uniformly over starting points $y \in K$. Let $K(T)$ denote the set of points that the zig-zag can reach in the time interval $[0, T]$ when starting in the set K and set $\lambda_K = \sup_{y \in K(T)} \lambda(y)$, a uniform upper bound on the jump rates $\lambda(x, v)$. An estimate for the probability of at least one sojourn time that is less than 2δ is then given by

$$\sup_{y \in K} \mathbb{P}_y \left[\min_i \tau_i \le 2\delta \right] \le 1 - e^{-\lambda_K 2\delta}.$$

For any $y \in K$ we obtain the bound

$$\begin{aligned}
\mathbb{P}_y \left[w'(Y, \delta, T) \ge \varepsilon \right] &= \mathbb{P}_y \left[\{w'(Y, \delta, T) \ge \varepsilon\} \cap \{\min_i \tau_i > 2\delta\} \right] \\
&\quad + \mathbb{P}_y \left[\{w'(Y, \delta, T) \ge \varepsilon\} \cap \{\min_i \tau_i \le 2\delta\} \right] \\
&\le 0 + \mathbb{P}_y \left[\{\min_i \tau_i \le 2\delta\} \right] \\
&\le 1 - e^{-\lambda_K 2\delta}.
\end{aligned}$$

It follows that, as $\delta \to 0$,

$$\sup_{y \in K} \mathbb{P}_y \left[\{\min_i \tau_i \le 2\delta\} \right] \le 1 - e^{-\lambda_K 2\delta} \to 0,$$

and (2) follows from taking δ small enough that $1 - e^{-\lambda_K 2\delta} < \epsilon$.

We now move to verifying Condition (A.4), by explicitly defining two Lyapunov functions $g_1, g_2 : E \to \mathbb{R}$ satisfying the condition. For brevity, we carry out the calculations for the case of $\gamma(x) \equiv 0$ in the switching rate λ (see (4)). Then we can use the following functions: for $\alpha_1, \alpha_2 \in (0, 1)$ and $\beta > 0$, let

$$\begin{aligned}
g_1(x, v) &= \alpha_1 V(x) + \beta v U'(x), \\
g_2(x, v) &= \alpha_2 U(x) + \beta v U'(x).
\end{aligned}$$

For non-constant γ that is uniformly bounded by some $\bar{\gamma}$, the following functions can instead be used:

$$\begin{aligned}
g_1(x, v) &= \alpha_1 V(x) + \phi(v U'(x)), \\
g_2(x, v) &= \alpha_2 U(x) + \phi(v U'(x)),
\end{aligned}$$

where $\phi(s) = \beta \frac{1}{2} \operatorname{sign}(s) \log(\bar{\gamma} + |s|)$ and $\beta \in (0, 1)$. For example, for the choice $\beta = 1/2$ calculations analogous to the ones below hold.

We now return to the case $\gamma \equiv 0$ and take g_1, g_2 accordingly. Without loss of generality we can assume $g_1, g_2 \ge 0$: we can take β small enough and if necessary add a constant to ensure that this holds. We show that for suitable α_i small enough, the functions g_1, g_2 satisfy (A.4).

By (C.1), $V(x) \to \infty$, and by (C.2), $U'(x)/V(x) \to 0$. It follows that g_1 grows to infinity as $|x| \to \infty$. Moreover, g_2 grows to infinity by the assumption (B.1) on U; since g_1 and g_2 are continuous, this settles part (a) of (A.4).

Part (b) of (A.4) requires that that g_2 grows faster than g_1 at infinity. This follows from Assumption $(C.2)$ on the potentials U and V: both dominate the derivative U', and U grows faster than V.

To show that (A.4) holds for the zig-zag process, we show that both g_1 and g_2 satisfy

$$e^{-g_i(x,v)}(Be^{g_i})(x,v) \to -\infty \quad |x| \to \infty.$$

Then since $e^{-g_i(x,v)}(Be^{g_i})(x,v)$ is continuous, the compactness of superlevel-sets follows.

By the definition of g_1, g_2 and the extended generator B,

$$e^{-g_1(x,v)}(Be^{g_1})(x,v) = \alpha_1 vV'(x) + \beta U''(x) + \max(vU'(x),0)\left[e^{-2v\beta U'(x)} - 1\right],$$

and

$$e^{-g_2(x,v)}(Be^{g_2})(x,v) = \alpha_2 vU'(x) + \beta U''(x) + \max(vU'(x),0)\left[e^{-2v\beta U'(x)} - 1\right].$$

We first verify the condition for g_2. For $v = +1$, we have

$$e^{-g_2(x,+1)}(Be^{g_2})(x,+1) = \alpha_2 U'(x) + \beta U''(x) + \max(U'(x),0)\left[e^{-2\beta U'(x)} - 1\right].$$

For $x \to +\infty$, we have $U'(x) \to +\infty$ by $(B.1)$, so that

$$e^{-g_2(x,+1)}(Be^{g_2})(x,+1) = U'(x)\left[\alpha_2 - 1 + \beta\frac{U''(x)}{U'(x)} + e^{-2\beta U'(x)}\right]$$
$$\sim U'(x)(\alpha_2 - 1) \to -\infty, \quad x \to +\infty,$$

since $U''/U' \to 0$ by $(B.3)$ and $\alpha_2 < 1$.

For $x \to -\infty$, we have $U'(x) \to -\infty$ by $(B.1)$, in particular $U'(x) < 0$ for large x. Hence,

$$e^{-g_2(x,+1)}(Be^{g_2})(x,+1) = U'(x)\left[\alpha_2 + \beta\frac{U''(x)}{U'(x)}\right]$$
$$\sim U'(x)\alpha_2 \to -\infty, \quad x \to -\infty.$$

For $v = -1$, the argument is analogous and we omit the details; this concludes the treatment of g_2.

We now consider g_1. For $v = +1$,

$$e^{-g_1(x,+1)}(Be^{g_1})(x,+1) = \alpha_1 V'(x) + \beta U''(x) + \max(U'(x),0)\left[e^{-2v\beta U'(x)} - 1\right].$$

In the limit $x \to +\infty$, $U'(x) \to +\infty$ and $V'(x)/U'(x) \to 0$ by $(C.2)$. It follows that

$$e^{-g_1(x,+1)}(Be^{g_1})(x,+1) = \alpha_1 V'(x) + \beta U''(x) + U'(x)\left[e^{-2v\beta U'(x)} - 1\right]$$
$$= U'(x)\left[\alpha_1\frac{V'(x)}{U'(x)} + \beta\frac{U''(x)}{U'(x)} + e^{-2\beta U'(x)} - 1\right]$$
$$\sim -U'(x) \to -\infty, \quad x \to +\infty.$$

For $x \to -\infty$, similar to the computations for g_2,

$$
\begin{aligned}
e^{-g_1(x,+1)}(Be^{g_1})(x,+1) &= \alpha_1 V'(x) + \beta U''(x) \\
&= V'(x) \left[\alpha_1 + \beta \frac{U''(x)}{V'(x)} \right] \to -\infty, \quad x \to +\infty,
\end{aligned}
$$

since $U''/V' \to 0$ by $(C.3)$ and $V' \to -\infty$ by $(C.1)$. The case $v = -1$ can be handled using similar arguments.

The preceding computations conclude the verification of Condition $(A.4)$. We are left with verifying the mixing property $(A.5)$.

Let $\nu_1, \nu_2 \in \mathcal{P}_c(E)$. Then there exists a compact set $K \subseteq E$ with $\nu_1(K) = \nu_2(K) = 1$. To show that $(A.5)$ holds, we must find $T, M > 0$ and $\rho_1, \rho_2 \in \mathcal{P}([0,T])$ such that for all $A \in \mathcal{B}(E)$,

$$
\int_0^T \int_E P(t, y, A)\, d\nu_1(y) d\rho_1(t) \leq M \int_0^T \int_E P(t, y, A)\, d\nu_2(y) d\rho_2(t).
$$

By Fubini's theorem, it is sufficient to prove that for any points $y_1 \in \mathrm{supp}(\nu_1)$ and $y_2 \in \mathrm{supp}(\nu_2)$,

$$
\int_0^T P(t, y_1, A)\, d\rho_1(t) \leq M \int_0^T P(t, y_2, A)\, d\rho_2(t), \tag{24}
$$

with ρ_1, ρ_2, T, M independent of y_1, y_2. To that end, let $K \subseteq E$ be a compact set containing the support of both ν_1 and ν_2. Without loss of generality, we can take K of the form $K_{\mathbb{R}} \times \{\pm 1\}$, where $K_{\mathbb{R}}$ is a closed interval. For $t_1 > 0$, let $K(t_1)$ be the set of points that the zig-zag process with speed one can reach in the time interval $[0, t_1]$ when starting in K:

$$
K(t_1) = \{y \in E : \mathrm{dist}_E(y, K) \leq t_1\}.
$$

We prove the inequality (24) for arbitrary points $y_1, y_2 \in K$, using the following two steps; in what follows we set $\mu = \mathrm{Leb} \otimes \mathrm{Unif}_{\pm 1}$.

(i) For any $t_1 > 0$ and with ρ_1 the uniform distribution over $[0, t_1]$, there is a positive constant C_{K,t_1} depending only on K and t_1 such that for any $T > t_1$, we have

$$
\int_0^T P(t, y_1, A)\, d\rho_1(t) \leq C_{K,t_1} \cdot \mu\left(A \cap K(t_1)\right), \quad \text{for all } A \in \mathcal{B}(E),
$$

with μ as the reference measure on $\mathcal{B}(E)$.

(ii) There exist positive constants $T > 0$ and $C'_{K,T}$ such that with ρ_2 the uniform distribution over $[0, T]$, we have

$$
\int_0^T P(t, y_2, A)\, d\rho_2(t) \geq C'_{K,T} \cdot \mu\left(A \cap K(t_1)\right), \quad \text{for all } A \in \mathcal{B}(E).
$$

Suppose (i) and (ii) hold. Then the estimate (24) also holds, with $M = C_{K,t_1}/C'_{K,T}$, some $t_1 < T$.

To verify (i), note that the measure

$$\mu_{y_1}(A) = \int_0^T P(t, y_1, A) \, d\rho_1(t)$$

is absolutely continuous with respect to $\mu = \text{Leb} \otimes \text{Unif}_{\pm 1}$ and its density is uniformly bounded in K. Now (i) follows since $P(t, K, A) = 0$ whenever $A \cap K(t_1) = \emptyset$ and $t \leq t_1$.

Next, we use Lemma 8 of [BRZ19] to show (ii). To that end, recall that a tuple (y, y') in $E \times E$ is called reachable if there exists an admissible path from y to y'. By Theorem 4 of [BRZ19], any two points are reachable as long as the potential U has at least one non-degenerate local minimum (which is trivially satisfied on $\mathbb{R}$ under our assumptions) and satisfies $U \in C^3(\mathbb{R})$.

By Lemma 8 in [BRZ19], for any two points $y_a = (x_a, v_a)$ and $y_b = (x_b, v_b)$ in E, there are open neighborhoods U_{y_a} of x_a and U_{y_b} of x_b, a time interval $(t_0, t_0 + \varepsilon]$ and a constant $c > 0$ such that for all $x'_a \in U_{y_a}$ and $t \in (t_0, t_0 + \varepsilon]$,

$$P\left(t, (x'_a, v_a), A_{\mathbb{R}} \times \{v_b\}\right) \geq c \, \text{Leb}(A \cap U_{z_b}), \quad \text{for all } A_{\mathbb{R}} \in \mathcal{B}(\mathbb{R}).$$

The spatial part of $K \times K(t_1)$ can be covered by open squares associated to all pairs of start and final points y_a and y_b, with $y_a \in K$ and $y_b \in K(t_1)$, as

$$K_{\mathbb{R}} \times K(t_1)_{\mathbb{R}} \subseteq \bigcup_{(y_a, y_b)} U_{y_a} \times U_{y_b},$$

where each U_y is an open interval in $\mathbb{R}$. By compactness, there exists a finite subcover by open squares $U_{y_a^i} \times U_{y_b^i}$ corresponding to pairs (y_a^i, y_b^i),

$$K_{\mathbb{R}} \times K(t_1)_{\mathbb{R}} \subseteq \bigcup_{i=1}^N U_{y_a^i} \times U_{y_b^i}.$$

Thereby, the set $K \times K(t_1) \subseteq E \times E$ is covered as

$$K \times K(t_1) \subseteq \bigcup_{i=1}^N \left[\left(U_{y_a^i} \times \{\pm 1\}\right) \times \left(U_{y_b^i} \times \{\pm 1\}\right)\right].$$

Hence for each $z = (x, v) \in K$, there are finitely many open sets $U_{y_b^i}$ covering $K(t_1)_{\mathbb{R}}$, with corresponding constants c_i, t_i, ε_i such that for all $t \in (t_i, t_i + \varepsilon_i]$,

$$P(t, y, A_{\mathbb{R}} \times \{v_b^i\}) \geq c_i \text{Leb}(A_{\mathbb{R}} \cap U_{z_b^i}), \quad \text{for all } A_{\mathbb{R}} \in \mathcal{B}(\mathbb{R}). \tag{25}$$

For any $A = A_{\mathbb{R}} \times A_{\pm} \in \mathcal{B}(E)$, write

$$A^+ := A \cap (\mathbb{R} \times \{+1\}),$$
$$A^- := A \cap (\mathbb{R} \times \{-1\}).$$

Then with $T > 0$ large enough for all intervals $(t_i, t_i + \varepsilon_i]$ to be contained in $[0, T]$, taking $\rho_2 = \text{Unif}([0, T])$, for any $z \in K$ it holds that

$$
\int_0^T P(t, y, A) \, d\rho_2(t) \geq \int_0^T P(t, y, A) \sum_i \mathbf{1}_{(t_i, t_i + \varepsilon_i]}(t) \, d\rho_2(t)
$$
$$
= \frac{1}{T} \sum_i \int_{t_i}^{t_i + \varepsilon_i} \left[P(t, y, A^+) + P(t, y, A^-) \right] \, dt.
$$

In each time interval $(t_i, t_i + \varepsilon_i]$, at least one transition probability is bounded from below as in (25), while the other one can be bounded from below by zero. Thereby,

$$
\int_0^T P(t, y, A) \, d\rho_2(t) \geq \frac{1}{T} \sum_i \varepsilon_i \cdot c_i \cdot \text{Leb}(A_{\mathbb{R}} \cap U_{y_b^i})
$$
$$
\geq \frac{1}{T} \min_i (\varepsilon_i c_i) \sum_i \mu \left[A \cap (U_{y_b^i} \times \{\pm 1\}) \right]
$$
$$
\geq \frac{1}{T} \min_i (\varepsilon_i c_i) \cdot \mu \left[A \cap \bigcup_i \left(U_{y_b^i} \times \{\pm 1\} \right) \right]
$$
$$
\geq \frac{1}{T} \min_i (\varepsilon_i c_i) \cdot \mu \left[A \cap K(t_1) \right],
$$

where the last inequality follows from $K(t_1)$ being covered by the $U_{y_b^i} \times \{\pm 1\}$. Hence (ii) follows with $C'_{K,T} = \min_i (\varepsilon_i c_i)/T$.

This finishes the verification of Condition (A.5), and thereby the proof of Theorem 3.4.

4.3. Derivation of the explicit form of the rate function

Here we prove the results described in Section 3.3. Recall that the state space is now taken as $E = \mathbb{T} \times \{\pm 1\}$.

Suppose μ is absolutely continuous with respect to ν_0 and write $\frac{d\mu}{d\nu_0}(x, v) = \rho(x, v)$ for the Radon-Nikodym density of μ with respect to ν_0, where ρ is assumed to be absolutely continuous. Define a mapping $H : \mathcal{D}^+(L) \to \mathbb{R}$ by

$$
H(u) = \int_E \frac{Lu}{u} \, d\mu. \tag{26}
$$

We compute

$$
\begin{aligned}
H(u) &= \sum_{v \in \{-1,+1\}} \int_{\mathbb{T}} \frac{Lu}{u}(x,v) \rho(x,v) \; dx \\
&= \int_{\mathbb{T}} \left\{ \frac{d \log u^+}{dx} + \lambda^+ \left(\frac{u^-}{u^+} - 1 \right) \right\} \rho^+ \; dx + \int_{\mathbb{T}} \left\{ -\frac{d \log u^-}{dx} + \lambda^- \left(\frac{u^+}{u^-} - 1 \right) \right\} \rho^- \; dx \\
&= \int_{\mathbb{T}} \left\{ -\log u^+ \frac{d\rho^+}{dx} + \lambda^+ \rho^+ \left(\frac{u^-}{u^+} - 1 \right) \right\} \; dx + \int_{\mathbb{T}} \left\{ \log u^- \frac{d\rho^-}{dx} + \lambda^- \rho^- \left(\frac{u^+}{u^-} - 1 \right) \right\} \; dx.
\end{aligned}
\tag{27}
$$

Lemma 4.5. *Suppose $\rho \in C(E)$ is absolutely continuous and satisfies*

$$
\nu_0 \left\{ \frac{d\rho^+}{dx} \neq \frac{d\rho^-}{dx} \right\} > 0.
$$

Then $\inf_{u \in \mathcal{D}^+(L)} H(u) = -\infty$.

Proof. Let $u_t^+(x) = u_t^-(x) = \exp\left(-t \left\{ \frac{d\rho^+}{dx} - \frac{d\rho^-}{dx} \right\} \right)$. From (27) it follows that

$$
H(u_t) = -t \int_{\mathbb{T}} \left(\frac{d\rho^+}{dx} - \frac{d\rho^-}{dx} \right)^2 \; dx.
$$

Now let $t \to \infty$. □

Lemma 4.6. *Suppose $\rho \in C(E)$ is absolutely continuous and $\frac{d\rho^+}{dx} = \frac{d\rho^-}{dx}$ for all $x \in \mathbb{T}$. Then $\mathcal{I}$ admits the representation*

$$
\mathcal{I}(\mu) = -\inf_{\eta \in C(\mathbb{T})} \int_{\mathbb{T}} \left\{ -\rho' \eta + \lambda^+ \rho^+ (\exp(-\eta) - 1) + \lambda^- \rho^- (\exp(\eta) - 1) \right\} \; dx.
\tag{28}
$$

Proof. Write $\rho' := \frac{d\rho^+}{dx}$, and note that by our assumption $\rho' = \frac{d\rho^-}{dx}$. By (27) we may write

$$
H(u) = \int_{\mathbb{T}} \left\{ -\log\left(u^+/u^-\right) \rho' + \lambda^+ \rho^+ (u^-/u^+ - 1) + \lambda^- \rho^- (u^+/u^- - 1) \right\} \; dx.
$$

We see that only the ratio u^+/u^- determines the value of $H(u)$. To any choice of $u \in \mathcal{D}^+(L)$ we may associate $\eta = \log u^+ - \log u^- \in C^1(\mathbb{T})$, and correspondingly, to any $\eta \in C^1(\mathbb{T})$ we can associate $u \in \mathcal{D}^+(L)$ by letting

$$
u^+(x) = \exp(\tfrac{1}{2}\eta(x)), \quad u^-(x) = \exp(-\tfrac{1}{2}\eta(x)), \quad x \in \mathbb{T}.
$$

By the continuous dependence of H on η, and the fact that $C^1(\mathbb{T})$ is dense in $C(\mathbb{T})$, we obtain the stated representation of $I(\mu)$. □

Lemma 4.7. *Suppose $\rho \in C^1(E)$ and $\frac{d\rho^+}{dx} = \frac{d\rho^-}{dx}$ for all $x \in \mathbb{T}$. Furthermore suppose $\lambda^- \lambda^+ \rho^- \rho^+ > 0$ on $\mathbb{T}$, and $\lambda^\pm$ are continuous. Then $\mathcal{I}$ is given by (7).*

Proof. Differentiating the integrand in (28) pointwise with respect to η gives the first order condition

$$-\rho' - \lambda^+ \rho^+ \exp(-\eta) + \lambda^- \rho^- \exp(\eta) = 0,$$

which is solved uniquely by

$$\eta = \tfrac{1}{2} \log \left(\frac{\lambda^+ \rho^+}{\lambda^- \rho^-} \right) + \operatorname{arcsinh} \left(\frac{\rho'}{2\sqrt{\lambda^+ \lambda^- \rho^+ \rho^-}} \right),$$

as long as $\lambda^- \lambda^+ \rho^- \rho^+ \neq 0$. Furthermore $\eta \in C(\mathbb{T})$ by the conditions on λ and ρ. The second order derivative with respect to η is given by

$$\lambda^+ \rho^+ \exp(-\eta) + \lambda^- \rho^- \exp(\eta) \geq 0,$$

which shows that the critical value of η corresponds to a pointwise global minimum of the integrand. $\qquad\square$

Proof of Proposition 3.5. The expression in case of equality follows from Lemma 4.7. The result for unequal derivatives follows from Lemma 4.5. $\qquad\square$

Proof of Proposition 3.6. We inspect the dependence of the various terms in the integrand of the expression (7) $I(\mu_c)$ on c. For the first term, interchanging integral and derivative,

$$\frac{d}{dc} \int_{\mathbb{T}} \rho' \log \left(\frac{\lambda^+ (\rho + c)}{\lambda^- (\rho - c)} \right) \, dx = \int_{\mathbb{T}} \rho' \left(\frac{1}{\rho + c} + \frac{1}{\rho - c} \right) \, dx$$
$$= \int_{\mathbb{T}} \frac{d}{dx} \left(\log(\rho + c) + \log(\rho - c) \right) \, dx = 0.$$

The following terms (i.e. the arcsinh and the square root) in the expression for $I(\mu_c)$ are decreasing with respect to the value of $\rho^+ \rho^- = \rho^2 - c^2$. It follows that the integrands are minimized at $c = 0$. Finally, we have that

$$\int_{\mathbb{T}} (\lambda^+ \rho^+ + \lambda^- \rho^-) \, dx = \int_{\mathbb{T}} \left\{ (\lambda^+ + \lambda^-)\rho + c(\lambda^+ - \lambda^-) \right\} \, dx.$$

The linear term in c vanishes since $\int_{\mathbb{T}} \{\lambda^+ - \lambda^-\} \, dx = \int_{\mathbb{T}} U' \, dx = 0$. It follows that $c = 0$ minimizes $c \mapsto I(\mu_c)$. The stated expression for $I(\mu_0)$ is a straightforward manipulation of (7). $\qquad\square$

Acknowledgment

References

[AdFDJ03] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1):5–43, 2003.

[AG07] S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

[AGG+86] W. Arendt, A. Grabosch, G. Greiner, U. Moustakas, R. Nagel, U. Schlotterbeck, U. Groh, H. P. Lotz, and F. Neubrander. *One-parameter semigroups of positive operators*, volume 1184. Springer, 1986.

[AL19] C. Andrieu and S. Livingstone. Peskun-Tierney ordering for Markov chain and process Monte Carlo: beyond the reversible scenario. pages 1–38, 2019.

[BCVD17] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The Bouncy Particle Sampler: A Non-Reversible Rejection-Free Markov Chain Monte Carlo Method. *Journal of the American Statistical Association*, 2017.

[BD17] J. Bierkens and A. Duncan. Limit theorems for the Zig-Zag process. *Advances in Applied Probability*, 49(3):791–825, jul 2017.

[BD19] A. Budhiraja and P. Dupuis. *Analysis and approximation of rare events*, volume 94 of *Probability Theory and Stochastic Modelling*. Springer, New York, 2019. Representations and weak convergence methods.

[BFR19] J. Bierkens, P. Fearnhead, and G. Roberts. The zig-zag process and super-efficient sampling for bayesian analysis of big data. *The Annals of Statistics*, 47(3):1288–1320, 2019.

[BR08] M. Bédard and J. S. Rosenthal. Optimal scaling of Metropolis algorithms: heading toward general target distributions. *Canad. J. Statist.*, 36(4):483–503, 2008.

[BR17] J. Bierkens and G. Roberts. A piecewise deterministic scaling limit of lifted metropolis–hastings in the curie–weiss model. *The Annals of Applied Probability*, 27(2):846–882, 2017.

[BRB19] J. Birrell and L. Rey-Bellet. Concentration inequalities and performance guarantees for hypocoercive mcmc samplers. 2019.

[BRZ19] J. Bierkens, G. O. Roberts, and P.-A. Zitt. Ergodicity of the zigzag process. *The Annals of Applied Probability*, 29(4):2266–2301, 2019.

[Buc04] J. A. Bucklew. *Introduction to rare event simulation.* Springer Series in Statistics. Springer-Verlag, New York, 2004.

[BV19] J. Bierkens and S. M. Verduyn Lunel. Spectral analysis of the zigzag process. *arXiv preprint arXiv: 1905.01691*, may 2019.

[Dav84] M. H. A. Davis. Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):353–376, 1984.

[Dav93] M. H. A. Davis. *Markov models and optimization*, volume 49 of

Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1993.

[DDN18] J. Doll, P. Dupuis, and P. Nyquist. A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms. *Appl. Math. Optim.*, 78(1):103–144, 2018.

[Dem94] A. Dembo. Zeitouni, 0.(1998). large deviations techniques and applications. *Applications of Mathematics (New York)*, 38, 1994.

[DHN00] P. Diaconis, S. Holmes, and R. M. Neal. Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, 10(3):726–752, 2000.

[DL15] P. Dupuis and Y. Liu. On the large deviation rate function for the empirical measures of reversible jump markov processes. *The Annals of Probability*, 43(3):1121–1156, 2015.

[DL18] P. Dupuis and D. Lipshutz. Large deviations for the empirical measure of a diffusion via weak convergence methods. *Stochastic Processes and their Applications*, 128(8):2581–2604, 2018.

[DLPD12] P. Dupuis, Y. Liu, N. Plattner, and J. D. Doll. On the infinite swapping limit for parallel tempering. *Multiscale Modeling & Simulation*, 10(3):986–1022, 2012.

[DV75a] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time, I. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.

[DV75b] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time, II. *Communications on Pure and Applied Mathematics*, 28(2):279–301, 1975.

[DV75c] M. D. Donsker and S. R. S. Varadhan. On a variational formula for the principal eigenvalue for operators with maximum principle. *Proceedings of the National Academy of Sciences*, 72(3):780–783, 1975.

[DV76a] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time, III. *Communications on pure and applied Mathematics*, 29(4):389–461, 1976.

[DV76b] M. D. Donsker and S. R. S. Varadhan. On the principal eigenvalue of secondorder elliptic differential operators. *Communications on Pure and Applied Mathematics*, 29(6):595–621, 1976.

[EK09] S. N Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, 2009.

[FBPR18] P. Fearnhead, J. Bierkens, M. Pollock, and G. O. Roberts. Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo. *Statistical Science*, nov 2018.

[FdSHS93] Arnoldo Frigessi, Patrizia di Stefano, Chii-Ruey Hwang, and Shuenn Jyi Sheu. Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *J. Roy. Statist. Soc. Ser. B*, 55(1):205–219, 1993.

[FHPS10] B. Franke, C.-R. Hwang, H.-M. Pai, and S.-J. Sheu. The behavior of the spectral gap under growing drift. *Trans. Amer. Math. Soc.*, 362(3):1325–1350, 2010.

[FK06]   J. Feng and T. G. Kurtz. *Large Deviations for Stochastic Processes*. Mathematical surveys and monographs. American Mathematical Society, 2006.

[HHMS05]   Chii-Ruey Hwang, Shu-Yin Hwang-Ma, and Shuenn-Jyi Sheu. Accelerating diffusions. *Ann. Appl. Probab.*, 15(2):1433–1444, 2005.

[Mon16]   P. Monmarché. Piecewise deterministic simulated annealing. *ALEA*, 13(1):357–398, 2016.

[MRR⁺53]   N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087, 1953.

[MT96]   K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.

[PD12]   E. A. J. F. Peters and G. De With. Rejection-free Monte Carlo sampling for general potentials. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 85(2):1–5, 2012.

[PDD⁺11]   N. Plattner, J. D. Doll, P. Dupuis, H. Wang, Y. Liu, and J. E. Gubernatis. An infinite swapping approach to the rare-event sampling problem. *The Journal of chemical physics*, 135(13):134111, 2011.

[RBS15]   L. Rey-Bellet and K. Spiliopoulos. Irreversible langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7):2081, 2015.

[RC04]   C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.

[Ros03]   J. S. Rosenthal. Asymptotic variance and convergence rates of nearly-periodic Markov chain Monte Carlo algorithms. *Journal of the American Statistical Association*, 98(461):169–177, 2003.

[RR04]   Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004.