Approximations of Piecewise Deterministic Markov Processes and their convergence properties

Bertazzi, Andrea; Bierkens, Joris; Dobson, Paul

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

Approximations of Piecewise Deterministic Markov Processes and their convergence properties

Andrea Bertazzi[*], Joris Bierkens, Paul Dobson

Delft Institute of Applied Mathematics, Delft University of Technology, Netherlands

Abstract

Piecewise deterministic Markov processes (PDMPs) are a class of stochastic processes with applications in several fields of applied mathematics spanning from mathematical modelling of physical phenomena to computational methods. A PDMP is specified by three characteristic quantities: the deterministic motion, the law of the random event times, and the jump kernels. The applicability of PDMPs to real world scenarios is currently limited by the fact that these processes can be simulated only when these three characteristics of the process can be simulated exactly. In order to overcome this problem, we introduce discretisation schemes for PDMPs which make their approximate simulation possible. In particular, we design both first order and higher order schemes that rely on approximations of one or more of the three characteristics. For the proposed approximation schemes we study both pathwise convergence to the continuous PDMP as the step size converges to zero and convergence in law to the invariant measure of the PDMP in the long time limit. Moreover, we apply our theoretical results to several PDMPs that arise from the computational statistics and mathematical biology literature.
© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Piecewise Deterministic Markov Processes (PDMPs) [21,22] are nowadays widely used in mathematical modelling in fields such as mathematical biology [5,14,42], biochemistry [45], insurance risk theory [20,29], materials science [1], neuroscience [40], and neutron transport [32].

---

 * Corresponding author.
   E-mail addresses: a.bertazzi@tudelft.nl (A. Bertazzi), joris.bierkens@tudelft.nl (J. Bierkens), p.dobson@tudelft.nl (P. Dobson).

Mathematical properties of PDMPs such as stability and stationarity have been extensively investigated in the mathematics community, see e.g. [4,16,26]. Moreover, in recent years these processes have also quickly gained in popularity for purposes of Monte Carlo computation in statistical physics [38,41,47] and in Bayesian statistics [30,48], for example in the form of the Bouncy Particle Sampler (BPS) and the Zig-Zag Sampler (ZZS) [7,13]. Several papers have further investigated the use of PDMPs in this area, e.g. [2,3,6,8,10,25,31,37].

PDMPs are continuous time Markov processes which move along deterministic trajectories (typically in Euclidean space) on a time interval of random length, after which a (possibly random) transition occurs to a new state, followed by another deterministic motion, etc. The deterministic motion is prescribed by the integral curves, φ_t, of a vector field $\Phi : \mathbb{R}^d \to \mathbb{R}^d$, the length of the random time intervals between transitions is governed by a transition rate $\lambda : \mathbb{R}^d \to [0, \infty)$, and the transitions are described by a Markov kernel $Q : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$. Together the vector field Φ, transition rate λ and transition kernel Q comprise the *characteristics* of the PDMP.

These processes are relatively easy to understand from a conceptual point of view and in some special cases their simulation can be performed exactly. In particular, if (i) the vector field Φ is explicitly integrable, (ii) it is possible to generate random times exactly as prescribed by λ, and (iii) it is possible to simulate from the transition kernel Q, then the iterative computation of trajectories of the associated PDMP is relatively straightforward.

Simulation of trajectories becomes problematic if one or more of these conditions are not met. Let us discuss possible problems that may arise. Concerning (i), the vector field Φ, as is well known in the field of differential equations, explicit solutions to the ODE $\dot{\varphi}_t = \Phi(\varphi_t)$ are only available in special cases, for example when Φ is affine, or when it is has some other special structure or symmetry. Concerning (ii), the transition rate λ, it is easy to simulate when the rate λ is constant or globally bounded. If λ is constant, then the random times between transitions are simply Exponential(λ)-distributed and thus easily simulated. If λ is globally bounded, say by a constant M, we may use a technique called *Poisson thinning* [36], which allows us to first simulate the random times according to an Exponential(M)-distribution and then accept a proposed transition time as a true transition with a probability governed by the ratio between $\lambda(\cdot)$ and M. The use of Poisson thinning may be extended to cases with non-constant bounds $M(s)$ along trajectories under the condition that it is simple to simulate from an inhomogeneous Poisson process with rate $M(s)$. However, finding a sharp bound $M(s)$ can be an extremely challenging problem in most practical settings. Moreover, the looser the bound the greater the computational cost of the simulation of the PDMP. For more extensive descriptions of Poisson thinning we refer to, e.g., [7,13]. Finally problems with (iii), the simulation of transitions according to Q, may arise in various ways. For instance it may be interesting to approximate the transition kernel of the BPS (see Section 2.4 of [44]).

In this paper we propose several schemes to approximate a PDMP in cases that are otherwise not straightforward to simulate, and we accompany these schemes by a detailed analysis of the convergence of the approximate process towards its exact, theoretical counterpart as the parameter governing the numerical precision, δ, converges to zero. Moreover, in the setting in which the PDMP is geometrically ergodic with a specified invariant measure, we investigate the theoretical convergence of the law of the approximate scheme to the invariant measure of the PDMP.

We introduce the *Fully Discrete PDMP* (FD-PDMP) Algorithm, the *Partially Discrete PDMP* (PD-PDMP) Algorithm and the Higher Order Partially Discrete PDMP Algorithm (Algorithms 2, 3 and 5, respectively). The FD-PDMP algorithm (Algorithm 2) defines a Markov

chain $\{\overline{Z}_{t_n}\}_{n \in \mathbb{N}}$ on a mesh $0 = t_0 < t_1 < t_2 < \ldots$ that moves deterministically between time steps, and a random event may occur at each of the mesh points with suitable probability. The PD-PDMP algorithm (Algorithm 3) defines a Markov chain that moves deterministically with exception of at most one random event in each interval of the form $[t_{n-1}, t_n]$. In contrast to the FD-PDMP the random event does not need to occur at mesh points. This difference motivates the choice of name of the two algorithms. By allowing at most p random events per time step, the higher order algorithm (Algorithm 5) constructs an approximation of the PDMP of order p.

Naturally these algorithms are designed to be straightforward to simulate. Both the FD-PDMP and the PD-PDMP algorithms rely on first order approximations of the characteristics of the PDMP. A wide range of approximations for φ_t, λ, Q is allowed, see Assumptions 4.4, 4.5, 4.6 for the formal requirements. As a simple yet important example, consider the case in which we are interested in simulating a PDMP for which the event times are hard to obtain. With an Euler-type approach, we can use an approximation of λ that is constant between mesh points, based on the state of the process at the initial point of each time step. For such approximation, the next event time in the case of PD-PDMP is simply exponentially distributed with constant rate, which is straightforward to simulate. Similarly, in the case of the FD-PDMP a random event takes place at the end of the time interval according to a Bernoulli distributed random variable. In comparison to the simulation of the continuous time PDMP, both algorithms do not require an upper bound to the switching rates, which is required to apply Poisson thinning. In a similar fashion, simple approximations of φ_t and Q can be employed. We refer to Section 3 for a detailed description of the algorithms.

We study convergence of these algorithms as a function of the step size and of the time horizon. Under very broad assumptions on the approximation, in particular allowing for approximations of all three λ, φ_t, and Q, in Theorem 4.9 we are able to show convergence in a Wasserstein distance to the PDMP as the step size tends to zero. In the case in which it is possible to simulate φ_t and Q exactly, we obtain convergence of the PD-PDMP algorithm in the stronger metric of total variation (see Theorem 4.17). In this setting weaker assumptions on the continuous time PDMP are required. For instance we show in Examples 5.2 and 5.5 that BPS satisfies the assumptions of Theorem 4.17 but not those of Theorem 4.9. Moreover, both Theorems establish convergence of order p as long as the approximations of φ_t, λ, and Q are of order p. The proofs of both these theorems rely on couplings of the continuous time PDMP with its approximation and are described respectively in Couplings 6.1 and 7.1.

In many areas it is important to understand the long time behaviour of the approximation schemes. In the field of Markov chain Monte Carlo (MCMC) algorithms the goal is to simulate a process that converges in law to the correct probability measure, which is the posterior distribution in Bayesian statistics and the Boltzmann–Gibbs distribution in statistical physics. In this context, such a probability measure is the invariant distribution of the PDMP. In Theorem 4.24 we prove uniform in time convergence of the weak error between the PDMP and the approximations given by the FD-PDMP or the PD-PDMP algorithms. In particular, we obtain convergence in law of the approximation and its time average to the invariant measure of the PDMP in the joint limit as time tends to infinity and step size tends to zero (see Corollary 4.27).

We confirm the applicability of our theorems on a variety of examples. ZZS and BPS are instances of PDMPs for which exact simulation of the random event times is not always possible. In Example 3.2 we discuss how to approximate the ZZS, and in Examples 5.1, 5.11 we show that our Theorems apply to the proposed approximation. An attractive feature of ZZS

is that it allows for exact *subsampling* (see [7]), which means that in a Bayesian statistics setting for each "iteration" of the algorithm only a subset of the data has to be accessed. In Example 5.7 we propose an approximation of ZZS with subsampling which also has the property of accessing a batch of the data over each time step and prove convergence in total variation as the step size tends to zero. For BPS we construct an approximation and prove convergence as step size tends to zero and time tends infinity in Examples 3.3, 5.5 and 5.17. In contrast to ZZS and BPS, randomised Hamiltonian Monte Carlo (RHMC) (see [12]) is an example of a PDMP in which it is typically not possible to simulate the flow φ_t exactly. In Examples 3.4, 5.3, and 5.8 we discuss approximations of RHMC and show convergence as the step size tends to zero and time tends to infinity. We also discuss continuous time approximation schemes of a PDMP in Examples 5.6 and 5.18.

Related works

Whereas discretisations of stochastic differential equations such as the Langevin equation have been studied extensively in the literature, see e.g. the book [34] or recent papers [27,28, 43], the same has not been done for PMDPs. Here we give a brief overview of works that are to some extent related to the present manuscript.

An approximation scheme for PDMPs suitable for a specific setting was proposed in [35]. The authors consider the case in which the ODE describing the deterministic motion can only be solved numerically, a global upper bound for the switching rates is available, and the kernels Q can be simulated exactly. Their proposal is to move deterministically according to a numerical integrator and to draw a proposal for the following event time according to the upper bound of the switching rates and then accept or reject it by Poisson thinning. The framework we propose in Algorithms 3 and 5 is more general as approximations of all characteristics are possible. Moreover, the approximations in this manuscript do not require existence or knowledge of an upper bound for the switching rates. As discussed in Example 5.4 it is possible to closely resemble the proposal of [35] using our framework. Moreover, we obtain similar finite time strong and weak error results as in [35] by applying Theorem 4.9.

In [48] the authors focus on how to design a discrete time PDMP with a specific invariant measure. This is a fundamentally different approach to the focus of this paper. A related work is [44], which defines a discrete time chain that resembles a BPS.

The book [15] discusses approximations of PDMP based on finite volume schemes for the Chapman–Kolmogorov type equations. Such schemes approximate the law of the process and are thus different in nature compared to this manuscript.

Finally, we discuss papers that deal with continuous time approximations of the ZZS. In [17] the authors propose an approximation that relies on an integrator and a root finding method to generate the random event times. The paper [33] discusses the effect of approximate switching rates $\{\tilde{\lambda}\}_{i=1}^d$ on the stationary measure of the ZZS. This approximation relies on the availability of suitable $\{\tilde{\lambda}\}_{i=1}^d$ for which it is possible to (efficiently) simulate the corresponding ZZS. In Examples 5.6 and 5.18 we discuss applications of our theory to these approximation schemes. A similar setting is considered in [26, Theorems 11 and 25], where the authors establish bounds in total variation distance between (the invariant measures of) two PDMPs with same deterministic dynamics, but different switching rates and jump kernels. The authors prove such bounds by a coupling of the two continuous time PDMPs that is similar in spirit to our Coupling 7.1. In this paper, in particular in Section 4.2, we bound the TV distance between a PDMP and a discrete time approximation. Thus the statements and proofs differ from [26] in this sense.

Organisation of the paper

The paper is organised as follows. In Section 2 we define notation that we use throughout the paper. In Section 3 we describe the setting and the proposed algorithms. In particular in Section 3.1 we discuss first order schemes and in Section 3.3 we consider higher order schemes. Section 4 contains the main results together with the required assumptions. This section is divided into three parts. Section 4.1 is devoted to convergence in Wasserstein distance, which is established in Theorem 4.9. Section 4.2 concerns convergence in total variation as stated in Theorem 4.17. Section 4.3 gives conditions for uniform in time convergence of the weak error, as expressed by Theorem 4.24. In Section 5 we gather examples to demonstrate the when the assumptions of the main theorems are satisfied. The proofs of the three main theorems can be found respectively in Section 6, Section 7 and Section 8. All other results as well as all auxiliary lemmas from Sections 4.1, 4.2, and 4.3 can be found respectively in Appendix A, Appendix B, and Appendix C.

2. Notation

We denote the semigroup of the continuous time PDMP, $\{Z_t\}_{t\geq0}$, as $\mathcal{P}_t$ which acts on suitable functions by

$$\mathcal{P}_t f(z) = \mathbb{E}_z[f(Z_t)].$$

Here the subscript z denotes that the process Z_t has initial position $Z_0 = z$. Note that the semigroup is related to the transition probability of Z_t, which is denoted by $\mathcal{P}_t(z, A)$. These concepts are related for any function f and measurable set $A \subseteq E$ by

$$\mathcal{P}_t f(z) = \int f(y)\mathcal{P}_t(z, \mathrm{d}y), \quad \mathcal{P}_t(z, A) = (\mathcal{P}_t \mathbb{1}_A)(z).$$

Similarly we denote the transition probability of the approximation processes $\{\overline{Z}_{t_n}\}_{n\in\mathbb{N}}$ as $\overline{\mathcal{P}}_{t_n}$.

Consider a metric $d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ and let P, Q be probability measures on $\mathbb{R}^d$. Then we define the Wasserstein distance of order 1 with respect to the metric d as

$$\mathcal{W}_1(P, Q) = \inf_{R\in\Pi(P,Q)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} d(x, y)R(\mathrm{d}x, \mathrm{d}y) \right\}, \tag{1}$$

where $\Pi(P, Q)$ is the set of couplings of the two probability measures P, Q, that is the set of probability measures R on $\mathbb{R}^d \times \mathbb{R}^d$ such that $R(A, \mathbb{R}^d) = P(A)$ and $R(\mathbb{R}^d, B) = Q(B)$.

We will denote a norm by $\|\cdot\|$. The maximum between $a \in \mathbb{R}$ and 0 is denoted by $(a)_+ = \max\{a, 0\}$.

Let us define the space $\mathcal{C}^k$ to be the set of functions $f : \mathbb{R}^d \to \mathbb{R}$ which are k times continuously differentiable. $\mathcal{C}_b^k$ ($\mathcal{C}_c^k$ respectively) denotes the subset of $\mathcal{C}^k$ to functions which are bounded (resp. have compact support) with bounded and continuous derivatives up to order k. We endow the space $\mathcal{C}_b$ with the supremum norm $\|\cdot\|_\infty$ and the space $\mathcal{C}_b^1$ is endowed with the norm

$$\|f\|_{\mathcal{C}_b^1} = \|f\|_\infty + \sum_{i=1}^{d} \|\partial_i f\|_\infty.$$

Consider a random variable I with values in $\{1, \ldots, m\}$ such that $\mathbb{P}(I = i) = w_i$ for $i = 1, \ldots, m$. Then we say I has a discrete distribution with probabilities w_i and we denote this as $I \sim \text{Discrete}(\{w_i\}_{i=1}^m)$.

Given a measure π we define for any $f \in L^1_\pi$

$$\pi(f) = \int f(z)\pi(\mathrm{d}z).$$

Similarly for a probability kernel $Q(x, \mathrm{d}y)$ we write

$$Qf(x) = \int f(y)Q(x, \mathrm{d}y) \tag{2}$$

for f measurable and integrable with respect to $Q(x, \mathrm{d}y)$ for all x. Note that (2) allows us to consider a probability kernel as a map from the space of bounded and measurable functions, B_b, to B_b.

Let us define the *total variation distance* between two probability measures μ, and ν as

$$\|\mu - \nu\|_{TV} = \sup_{f \in \mathcal{C}_b : \|f\|_\infty \leq 1} |\mu(f) - \nu(f)|.$$

Given a vector field Φ we can view this as a map, $\boldsymbol{\Phi}$, which acts on $\mathcal{C}^1$ functions as

$$\boldsymbol{\Phi}(f)(x) = \Phi(x)^T \nabla f(x), \quad \text{for } f \in \mathcal{C}^1.$$

Given two maps $X, Y : \mathcal{C}^\infty \to \mathcal{C}^\infty$ we shall define the commutator of X and Y, $[X, Y]$ to be the map $[X, Y] : \mathcal{C}^\infty \to \mathcal{C}^\infty$ by

$$[X, Y]f = XYf - YXf, \quad \text{for } f \in \mathcal{C}^\infty.$$

We will use this with the maps $\boldsymbol{\Phi}$ and Q. If we assume that Q preserves $\mathcal{C}^1$ and $\boldsymbol{\Phi}$ is bounded then we can view $[\boldsymbol{\Phi}, Q] : \mathcal{C}^1_b \to B_b$ defined by

$$[\boldsymbol{\Phi}, Q]f = \boldsymbol{\Phi}(Qf) - Q\boldsymbol{\Phi}(f), \quad \text{for } f \in \mathcal{C}^1_b.$$

Note that although the commutator was defined for smooth vector fields the above definition makes sense for all $\mathcal{C}^1_b$-functions since $\boldsymbol{\Phi}$ is a map from $\mathcal{C}^1_b$ to $\mathcal{C}_b$ and we have $Q(\mathcal{C}^1_b) \subseteq \mathcal{C}^1_b$, $Q(\mathcal{C}_b) \subseteq \mathcal{C}_b$ so both the operations $Q(\boldsymbol{\Phi}f)$ and $\boldsymbol{\Phi}(Qf)$ are well defined for $f \in \mathcal{C}^1_b$.

3. Algorithms

Consider a PDMP $(Z_t)_{t \geq 0}$ taking values on a state space E, which is a subset of a finite dimensional vector space. Examples are $E = \mathbb{R}^d \times \mathbb{R}^d$ or $E = \mathbb{R}^d \times \{-1, +1\}^d$. The dynamics of the process are described by the generator $\mathcal{L}$, which applied on a function in the domain of the extended generator gives

$$\mathcal{L}f(z) = \langle \Phi(z), \nabla_z f(z) \rangle + \sum_{i=1}^m \lambda_i(z) \int_E (f(y) - f(z))Q_i(z, \mathrm{d}y). \tag{3}$$

The generator here is understood to be the extended generator, see [22, Theorem 26.14] for the exact description of the domain of the extended generator. Note, in particular that functions that are differentiable in the direction Φ and bounded are included in the domain. Here Φ is a smooth and globally Lipschitz vector field, $\lambda_i : E \to [0, \infty)$ are continuous functions and Q_i are probability kernels. Let φ_t denote the integral curve of Φ, i.e. the solution to the following ordinary differential equation (ODE)

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t(z) = \Phi(\varphi_t(z)), \quad \varphi_0(z) = z, \quad \text{for all } t \geq 0, z \in E.$$

Algorithm 1: Pseudo-code for the simulation of a PDMP

Input : Time horizon T, initial condition z.

Set $t = 0$, $Z_0 = z$;

while $t < T$ **do**

> simulate next event time as
>
> $$\tau = \inf\left\{r > 0 : 1 - \exp\left(-\int_0^r \lambda(\varphi_s(Z_t))\mathrm{d}s\right) \geq U\right\}$$
>
> where $U \sim \mathrm{Unif}[0, 1]$;
> simulate $Z_{t+s} = \varphi_s(Z_t)$ for $s \in (0, \tau)$;
> draw $I \sim \mathrm{Discrete}\left(\left\{\frac{\lambda_i(Z_{t+\tau-})}{\lambda(Z_{t+\tau-})}\right\}_{i=1}^m\right)$;
> simulate $Z_{t+\tau} \sim Q_I(Z_{t+\tau-}, \cdot)$;
> set $t = t + \tau$;

end

Note that φ_t exists since Φ is globally Lipschitz. We assume that φ_t leaves E invariant. Define the total switching rate

$$\lambda(z) = \sum_{i=1}^m \lambda_i(z).$$

As shown in [22, Section 26] (3) corresponds to a PDMP where the next event time is distributed as

$$\mathbb{P}_z(\tau \leq t) = 1 - \exp\left(-\int_0^t \lambda(\varphi_s(z))\mathrm{d}s\right), \tag{4}$$

and that between two random events the process follows the flow-map φ_t, i.e. $Z_t = \varphi_t(z)$. At event time, τ, the process jumps according to probability kernel Q_I, where I is distributed according to the following discrete distribution

$$I \sim \mathrm{Discrete}\left(\left\{\frac{\lambda_i(\varphi_\tau(z))}{\lambda(\varphi_\tau(z))}\right\}_{i=1}^m\right).$$

Algorithm 1 describes the simulation procedure for a PDMP with generator (3).

Note 3.1. It is possible to rewrite (3) to the form

$$\mathcal{L}f(z) = \langle \Phi(z), \nabla_z f(z)\rangle + \lambda(z)\int_E (f(y) - f(z))Q(z, \mathrm{d}y) \tag{5}$$

for some continuous function $\lambda : E \to [0, \infty)$ and probability kernel Q. Indeed this can be achieved by setting

$$\lambda(z) = \sum_{i=1}^m \lambda_i(z), \quad Q(z, \mathrm{d}y) = \sum_{i=1}^m \frac{\lambda_i(z)}{\lambda(z)} Q_i(z, \mathrm{d}y). \tag{6}$$

Therefore there is no loss of generality for the PDMP to take $m = 1$. However we will see in Section 4.1 that allowing $m \geq 1$ leads to weaker assumptions for our convergence results, in particular we will see in Example 5.1 a case where the assumptions are satisfied with $m > 1$ but would not be satisfied when written in the form (5).

The focus of this paper is to define and analyse approximations of PDMPs that can be employed in settings where their simulation cannot be performed exactly. As explained in the introduction, there are three quantities which characterise a PDMP and may be difficult to simulate. These are the flow map φ_t, the random event times with rates λ_i, and the Markov kernels Q_i. The idea is then to introduce pth order approximations of the three characteristics for some $p \geq 1$. Precise conditions on the approximations are given in Assumptions 4.4, 4.5, 4.6, but here we provide a heuristic description. The flow map $\varphi_t(z)$ can be approximated with a numerical integrator, which is denoted as $\overline{\varphi}_t(z; \delta, p)$. The parameters δ, p have the meaning that for $s \in [0, \delta]$ we have that $\overline{\varphi}_s(z; \delta, p)$ is an approximation of order δ^p of $\varphi_s(z)$. Classical examples of numerical integrators from the ODE literature include the Euler discretisation, the leap frog scheme, and higher order numerical schemes. Then we want to approximate the switching rates in such a way that the random times (4) can be simulated easily at the cost of a small error. This can be done by using order δ^p approximations of $\lambda(\varphi_s(z))$, i.e. the switching rate along the deterministic flow. We denote the corresponding approximation as $\overline{\lambda}(z, s; \delta, p) : E \times [0, \infty) \to [0, \infty)$. The motivation is to ensure the following as an approximation of order δ^p for $t \leq \delta$:

$$\mathbb{P}_z(\tau \leq t) \approx 1 - \exp\left(-\int_0^t \overline{\lambda}(z, s; \delta, p)\mathrm{d}s\right).$$

Here

$$\overline{\lambda}(z, s; \delta, p) = \sum_{i=1}^m \overline{\lambda}_i(z, s; \delta, p).$$

Let us give some examples with $p = 1$. A possible choice is to "freeze" the switching rate, thus taking $\overline{\lambda}_i(z, s; \delta, 1) = \lambda_i(z)$. This is supported by the intuition that $\lambda(\varphi_s(z)) \approx \lambda(z)$ for small s. In this case $\mathbb{P}_z(\tau \leq t)$ is approximately equal to $1 - \exp(-t\lambda(z))$, which is the cumulative distribution function of the exponential distribution with constant rate $\lambda(z)$. We refer to the $\overline{\lambda}_i$ as *frozen switching rates* and to the corresponding approximation process as *Euler approximation*. Alternatively one could take $\overline{\lambda}_i(z, s; \delta, 1) = \lambda_i(\varphi_\delta(z))$, or the switching rates along the trajectory given by the numerical integrator $\overline{\lambda}_i(z, s; \delta, 1) = \lambda_i(\overline{\varphi}_s(z; \delta, 1))$, or more generally $\overline{\lambda}_i(z, s; \delta, p) = \lambda_i(\overline{\varphi}_s(z; \delta, p))$. Finally, consider the Markov kernels Q_i. We define a function F_i which describes a choice of implementation of Q_i. Let $F_i : E \times \mathcal{U} \to E$ be a deterministic map such that $F_i(z, U)$ is distributed according to $Q_i(z, \cdot)$ when U is distributed according to a probability distribution $\nu_{\mathcal{U}}$. We can then approximate each map F_i by a map $\overline{F}_i(\cdot; \delta, p) : E \times \mathcal{U} \to E$, where once again δ, p denotes the order of accuracy of our estimate. To simplify the notation, when we consider first order schemes, i.e. $p = 1$, we shall suppress the p-dependence and write $\overline{\varphi}_s(z; \delta)$, $\overline{\lambda}_i(z, s; \delta)$, $\overline{F}_i(z, U; \delta)$.

Now that we have introduced the problem and the various approximations we wish to exploit, we illustrate how to design first order and higher order approximation schemes for PDMPs. By an *order p scheme* we mean an approximation process for which the *local error*, i.e. the error between the PDMP and the approximation over a step of size δ with identical initial conditions, is proportional to δ^{p+1}. Therefore after n steps of size δ the *global error* is proportional to $t_n \delta^p$ where $t_n = n\delta$, which motivates the term order p scheme.

3.1. First order schemes

Let us introduce a mesh $\{t_n\}_{n \in \mathbb{N}}$ for the time variable where $t_n = \sum_{\ell=1}^n \delta_\ell$, and δ_ℓ are step sizes. For example if the step size is constant $\delta_\ell = \delta$ then $t_n = n\delta$ for all $n \in \mathbb{N}$. In this section

we introduce two alternative first order schemes: the FD-PDMP algorithm and the PD-PDMP algorithm. We define the FD-PDMP approximation $\{\overline{Z}_{t_n}\}$ on the mesh $\{t_n\}_{n\in\mathbb{N}}$ by setting $\overline{Z}_0 = z$ and then following the procedure

$$\tilde{Z}_{t_{n+1}} = \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n}; \delta_{n+1}),$$
$$\overline{Z}_{t_{n+1}} = \alpha_{n+1}\overline{F}_{\bar{I}_{n+1}}(\tilde{Z}_{t_{n+1}}, U_{n+1}; \delta_{n+1}) + (1 - \alpha_{n+1})\tilde{Z}_{t_{n+1}}.$$

Here we have $U_{n+1} \sim \nu_{\mathcal{U}}$. The value of α_{n+1} is determined as follows. We simulate $\bar{\tau}$ which has distribution conditional on $\overline{Z}_{t_n}$ given by

$$\mathbb{P}_z(\bar{\tau} \leq t | \overline{Z}_{t_n}) = 1 - \exp\left(-\int_0^t \overline{\lambda}(\overline{Z}_{t_n}, s; \delta_{n+1})\mathrm{d}s\right).$$

Then $\alpha_{n+1} = 1$ if and only if $\bar{\tau} \leq \delta_{n+1}$, otherwise $\alpha_{n+1} = 0$. We then draw

$$\bar{I}_{n+1} \sim \mathrm{Discrete}\left(\left\{\frac{\overline{\lambda}_i(\overline{Z}_{t_n}, \bar{\tau}; \delta_{n+1})}{\overline{\lambda}(\overline{Z}_{t_n}, \bar{\tau}; \delta_{n+1})}\right\}_{i=1}^m\right). \tag{7}$$

The resulting Markov chain $\overline{Z}_{t_n}$ is thus updated by first following the approximate flow map and then establishing whether a random event takes place at the end of the current time interval. This procedure is written in pseudo-code form in Algorithm 2. Note that if $\overline{\lambda}(z, s; \delta_{n+1})$ is independent of s, i.e. $\overline{\lambda}(z, s; \delta_{n+1}) = \overline{\lambda}(z; \delta_{n+1})$, then we do not need to simulate $\bar{\tau}_{n+1}$ and we have that α_{n+1} is a Bernoulli random variable with success rate $1 - \exp(-\delta_{n+1}\overline{\lambda}(z; \delta_{n+1}))$ and $\bar{I}_{n+1}$ is distributed as

$$\bar{I}_{n+1} \sim \mathrm{Discrete}\left(\left\{\frac{\overline{\lambda}_i(\overline{Z}_{t_n}; \delta_{n+1})}{\overline{\lambda}(\overline{Z}_{t_n}; \delta_{n+1})}\right\}_{i=1}^m\right).$$

This is for instance the case of frozen switching rates.

A different approach is shown in Algorithm 3, which describes the PD-PDMP approximation. Here the idea is to simulate the switching time $\bar{\tau}$ with rate $\overline{\lambda}(\overline{Z}_{t_n}, s; \delta_{n+1})$, then if $\bar{\tau}$ is before the end of the current time step set $t = t_n + \bar{\tau}$, draw $\bar{I}_{n+1}$ as in (7), and follow the procedure below:

$$\overline{Z}_t = \overline{F}_{\bar{I}_{n+1}}(\tilde{Z}_t, U_{n+1}; \delta_{n+1}), \quad \text{where } \tilde{Z}_t = \overline{\varphi}_{\bar{\tau}}(\overline{Z}_{t_n}; \delta_{n+1}),$$
$$\overline{Z}_{t_{n+1}} = \overline{\varphi}_{t_{n+1}-t}(\overline{Z}_t; \delta_{n+1}).$$

On the other hand, when $\bar{\tau} > \delta_{n+1}$ the process is simply moving deterministically according to the approximate flow map, i.e. $\overline{Z}_{t_{n+1}} = \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n}; \delta_{n+1})$. Only one random jump per time step is allowed, and in this case it happens at time $\bar{\tau}$ instead of at the end of the time step. This choice comes with advantages and disadvantages if compared to Algorithm 2. As we shall see in Sections 4.2 and 4.3 it is possible to obtain stronger results under weaker assumptions on the PDMP in the setting of Algorithm 3 compared to Algorithm 2. However this may come at a larger computational cost (see e.g. Example 3.3).

3.2. Examples

In this section we introduce several examples, which will be revisited as illustrative applications of our results. In the first three examples, i.e. Examples 3.2, 3.3, 3.4, we discuss MCMC samplers which target a probability measure with density $\pi(x) \propto \exp(-\psi(x))$ for $x \in \mathbb{R}^d$.

Algorithm 2: Fully Discrete Approximation of a PDMP

Input : Number of iterations N, initial condition z, step sizes $(\delta_n)_{n=0}^N$.

Output: Chain $(\overline{Z}_{t_n})_{n=0}^N$.

Set $n = 0$, $\overline{Z}_0 = z$;

while $n < N$ **do**

    simulate $\tilde{Z} = \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n}; \delta_{n+1})$;

    simulate

$$\bar{\tau} = \inf\left\{ r > 0 : 1 - \exp\left(-\int_0^r \overline{\lambda}(\overline{Z}_{t_n}, s; \delta_{n+1})\mathrm{d}s\right) \geq U \right\}$$

    where $U \sim \mathrm{Unif}[0, 1]$;

    if $\bar{\tau} \leq \delta_{n+1}$ **then**

        draw $U_{n+1} \sim \nu_{\mathcal{U}}$ and $\bar{I}_{n+1} \sim \mathrm{Discrete}\left(\left\{\frac{\overline{\lambda}_i(\overline{Z}_{t_n}, \bar{\tau}; \delta_{n+1})}{\overline{\lambda}(\overline{Z}_{t_n}, \bar{\tau}; \delta_{n+1})}\right\}_{i=1}^m\right)$;

        set $\tilde{Z} = \overline{F}_{\bar{I}_{n+1}}(\tilde{Z}, U_{n+1}; \delta_{n+1})$;

    end

    set $\overline{Z}_{t_{n+1}} = \tilde{Z}$;

    set $n = n + 1$;

end

Algorithm 3: Partially Discrete Approximation of a PDMP

Input : Number of iterations N, initial condition z, step sizes $(\delta_n)_{n=0}^N$.

Output: Chain $(\overline{Z}_{t_n})_{n=0}^N$.

Set $n = 0$, $\overline{Z}_0 = z$;

while $n < N$ **do**

    simulate

$$\bar{\tau} = \inf\left\{ r > 0 : 1 - \exp\left(-\int_0^r \overline{\lambda}(Z_t, s; \delta_{n+1})\mathrm{d}s\right) \geq U \right\}$$

    where $U \sim \mathrm{Unif}[0, 1]$;

    if $\bar{\tau} < \delta_{n+1}$ **then**

        set $t = t_n + \bar{\tau}$;

        simulate $\tilde{Z}_t = \overline{\varphi}_{\bar{\tau}}(\overline{Z}_{t_n}; \delta_{n+1})$;

        draw $U_{n+1} \sim \nu_{\mathcal{U}}$ and $\bar{I}_{n+1} \sim \mathrm{Discrete}\left(\left\{\frac{\overline{\lambda}_i(\overline{Z}_{t_n}, \bar{\tau}; \delta_{n+1})}{\overline{\lambda}(\overline{Z}_{t_n}, \bar{\tau}; \delta_{n+1})}\right\}_{i=1}^m\right)$;

        set $\overline{Z}_t = \overline{F}_{\bar{I}_{n+1}}(\tilde{Z}_t, U_{n+1}; \delta_{n+1})$;

        simulate $\overline{Z}_{t_{n+1}} = \overline{\varphi}_{t_{n+1}-t}(\overline{Z}_t; \delta_{n+1})$;

    else

        simulate $\overline{Z}_{t_{n+1}} = \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n}; \delta_{n+1})$;

    end

    set $n = n + 1$;

end

Example 3.2 (*Zig-Zag Sampler [7]*). Let $E = \mathbb{R}^d \times \{+1, -1\}^d$, and for any $z \in E$ we write $z = (x, v)$ for $x \in \mathbb{R}^d$, $v \in \{+1, -1\}^d$. Set $\Phi(x, v) = (v, 0)^T$, $m = d$, $\lambda_i(x, v) = (v_i \partial_i \psi(x))_+ + \gamma_i(x, v)$, and $Q_i((x, v), (dy, dw)) = \delta_{(x, R_i v)}(dy, dw)$, where δ_z denotes the Dirac Delta measure and $R_i v = (v_1 \ldots, v_{i-1}, -v_i, v_{i+1}, \ldots, v_d)$. The ZZS is described by its generator

$$\mathcal{L} f(x, v) = \langle v, \nabla_x f(x) \rangle + \sum_{i=1}^d \lambda_i(x, v)[f(x, R_i v) - f(x, v)]. \tag{8}$$

Simulating the event times with rates of this form is in general a very challenging problem as the integral in (4) cannot be computed for general potentials ψ.

We can apply Algorithm 2 to the ZZS to obtain $(\overline{X}_{t_{n+1}}, \overline{V}_{t_{n+1}})$ given the previous state by first simulating the next switching time $\bar{\tau}$ with rate $\overline{\lambda}((\overline{X}_{t_n}, \overline{V}_{t_n}), s; \delta_{n+1})$ and then

$$\overline{X}_{t_{n+1}} := \overline{X}_{t_n} + \overline{V}_{t_n} \delta_{n+1}$$

$$\overline{V}_{t_{n+1}} := \begin{cases} R_{\bar{I}_{n+1}} \overline{V}_{t_n} & \text{if } \bar{\tau} \leq \delta_{n+1}, \\ \overline{V}_{t_n} & \text{if } \bar{\tau} > \delta_{n+1}, \end{cases}$$

where $\overline{\lambda}(z, s; \delta) = \sum_{i=1}^d \overline{\lambda}_i(z, s; \delta)$, and

$$\bar{I}_{n+1} \sim \text{Discrete}\left(\left\{ \frac{\overline{\lambda}_i((\overline{X}_{t_n}, \overline{V}_{t_n}), \bar{\tau}; \delta_{n+1})}{\overline{\lambda}((\overline{X}_{t_n}, \overline{V}_{t_n}), \bar{\tau}; \delta_{n+1})} \right\}_{i=1}^d \right).$$

The only approximation concerns the switching rates, whereas it is straightforward to simulate the linear dynamics and the jumps at event times. As mentioned above, a simple choice is to take $\overline{\lambda}_i((x, v), s; \delta) = \lambda_i(x, v)$, which results in an Euler approximation of the ZZS. An alternative choice is

$$\overline{\lambda}_i((x, v), s; \delta) = \frac{1}{\delta} (\psi(x + v_i e_i \delta) - \psi(x))_+ + \gamma_i(x, v), \tag{9}$$

which is obtained by a finite difference scheme approximation for $\partial_i \psi$. Here e_i is the ith vector of the canonical basis. Observe that with this choice of $\overline{\lambda}_i$ the approximation is gradient free, as it does not require computing $\nabla \psi$. An approximation given by Algorithm 3 may be introduced analogously.

Example 3.3 (*Bouncy Particle Sampler [13,41]*). Let $E = \mathbb{R}^d \times \mathbb{R}^d$, and for any $z \in E$ we write $z = (x, v)$ for $x \in \mathbb{R}^d$, $v \in \mathbb{R}^d$. Set $\Phi(x, v) = (v, 0)^T$, $m = 2$,

$$Q_1((x, v), (dy, dw)) = \delta_{(x, R(x)v)}(y, w), \quad Q_2((x, v), (dy, dw)) = \delta_x(dy)\nu(dw),$$

with $\lambda_1(x, v) = (v^T \nabla_x \psi(x))_+$, $\lambda_2 = \lambda_r$ for $\lambda_r > 0$, ν is the Gaussian measure, and finally

$$R(x)v = v - 2 \frac{\langle v, \nabla_x \psi(x) \rangle}{\|\nabla_x \psi(x)\|^2} \nabla_x \psi(x).$$

In this example $\| \cdot \|$ denotes the Euclidean norm. The BPS has generator

$$\mathcal{L} f(x, v) = \langle v, \nabla_x f(x) \rangle + \lambda_1(x, v)[f(x, R(x)v) - f(x, v)] + \lambda_2 \int \big(f(x, w) - f(x, v) \big) \nu(dw).$$

For the same reasons as for the ZZS, simulating the event times can be very challenging for the BPS.

For this process we introduce an approximation based on Algorithm 3. Let $U_{n+1} = (\mathscr{Z}_{n+1}, \mathscr{U}_{n+1})$ for $\mathscr{Z}_{n+1}$ distributed according to the standard Gaussian distribution ν and $\mathscr{U}_{n+1} \sim \text{Unif}([0, 1])$ is an independent uniform random variable. For $n \geq 0$ we define the next state of the approximation $(\overline{X}_{t_{n+1}}, \overline{V}_{t_{n+1}})$ given the previous state by first simulating $\bar{\tau}$ with distribution $\mathbb{P}_z(\bar{\tau} > t) = \exp(-\int_0^t \bar{\lambda}((\overline{X}_{t_n}, \overline{V}_{t_n}), s; \delta_{n+1})ds)$ and then

$$\overline{X}_{t_{n+1}} := \overline{X}_{t_n} + \bar{\tau}\overline{V}_{t_n} + (\delta_{n+1} - \bar{\tau})\overline{V}_{t_{n+1}},$$

$$\overline{V}_{t_{n+1}} := \begin{cases} F((\overline{X}_{t_n+\bar{\tau}}, \overline{V}_{t_n}), U_{n+1}) & \text{if } \bar{\tau} \leq \delta_{n+1}, \\ \overline{V}_{t_n} & \text{if } \bar{\tau} > \delta_{n+1}. \end{cases}$$

Here $\bar{\lambda}((x, v), t; \delta_{n+1}) = \bar{\lambda}_1((x, v), t; \delta_{n+1}) + \lambda_r$ where $\bar{\lambda}_1((x, v), t; \delta_{n+1})$ approximates $\lambda_1(x + vt, v)$ and

$$F((\overline{X}_{t_n+\bar{\tau}}, \overline{V}_{t_n}), U_{n+1}) = \begin{cases} R(\overline{X}_{t_n+\bar{\tau}})\overline{V}_{t_n} & \text{if } \mathscr{U}_{n+1} > \frac{\lambda_r}{\bar{\lambda}((\overline{X}_{t_n}, \overline{V}_{t_n}), \bar{\tau}; \delta_{n+1})}, \\ \mathscr{Z}_{n+1} & \text{if } \mathscr{U}_{n+1} \leq \frac{\lambda_r}{\bar{\lambda}((\overline{X}_{t_n}, \overline{V}_{t_n}), \bar{\tau}; \delta_{n+1})}. \end{cases}$$

It is thus clear that applying Algorithm 3 rather than Algorithm 2 can be more computationally expensive in the case of BPS, as when an event takes place $\nabla\psi$ has to be evaluated at some midpoint $\overline{X}_{t_n+\bar{\tau}}$ in order to compute the reflection operator. In contrast, $\nabla\psi$ has to be computed only at gridpoints in Algorithm 2. We shall see in Section 4 that our theoretical results can only be applied to approximations of the BPS based on Algorithm 3, motivating the need for that algorithm.

Similarly to the case of the ZZS described in Example 3.2, possible approximations of $\lambda_1(x + vt, v)$ are $\bar{\lambda}_1((x, v), t; \delta) = \lambda(x, v)$ or

$$\bar{\lambda}_1((x, v), t; \delta) = \frac{1}{\delta}(\psi(x + v\delta) - \psi(x))_+.$$

The latter choice is not enough to not make the simulation of $(\overline{X}_t, \overline{V}_t)$ gradient free because $\nabla\psi$ is needed in the computation of the reflection operator.

Example 3.4 (*Randomised Hamiltonian Monte Carlo Algorithm*). The randomised Hamiltonian Monte Carlo algorithm (see [12]) is defined on $E = \mathbb{R}^d \times \mathbb{R}^d$ by the generator

$$\mathcal{L}f(q, p) = \langle p, \nabla_q f(q, p)\rangle - \langle \nabla_q \psi(q), \nabla_p f(q, p)\rangle + \lambda_r \int (f(q, p') - f(q, p))\nu(dp'),$$

where ν is a Gaussian measure on $\mathbb{R}^d$. The Hamiltonian flow cannot be simulated exactly in most cases, and thus it becomes necessary to approximate it by a numerical integrator $\overline{\varphi}_s$. Then according to Algorithm 2 we obtain the next state by first denoting $(\tilde{Q}_{t_{n+1}}, \tilde{P}_{t_{n+1}}) = \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_{n+1}}; \delta_{n+1})$ and thus

$$(\overline{Q}_{t_{n+1}}, \overline{P}_{t_{n+1}}) = \begin{cases} (\tilde{Q}_{t_{n+1}}, \tilde{P}_{t_{n+1}}) & \text{with probability } \exp(-\lambda_r \delta_{n+1}), \\ (\tilde{Q}_{t_{n+1}}, \mathscr{Z}) & \text{with probability } 1 - \exp(-\lambda_r \delta_{n+1}), \end{cases}$$

where $\mathscr{Z} \sim \nu$. We remark that the most efficient implementation is to simulate the next refreshment time and then follow the numerical integrator until then, without drawing a new refreshment time at each iteration.

Example 3.5 (*Modelling the Size of a Cell*). Following Section 1.5 in [42], denote the size of a cell by $z \in \mathbb{R}$. The cell grows in time with deterministic flow φ_t, and splits into two daughter

cells with division rate $\lambda(z)$. Then denote as τ_n the time when a cell from the n-generation splits. The size of a daughter cell is half of the parent cell, and thus $Z_{\tau_n} = \frac{1}{2}Z_{\tau_n-}$. We can characterise the resulting process with its generator:

$$\mathcal{L}f(z) = \langle \Phi(z), \nabla f(z) \rangle + \lambda(z) \left(f\left(\frac{z}{2}\right) - f(z) \right).$$

Therefore it may not be possible to simulate such a process if the desired φ and λ are complicated functions. An approximation can be obtained applying the ideas above introducing a numerical integrator $\overline{\varphi}$ and approximate division rate $\overline{\lambda}$.

Example 3.6 (*Chemotaxis in Escherichia coli*)**.** It was shown in [5] that the bacteria Escherichia coli have two types of behaviour describing their motion, which are called "runs" and "twiddles". When the bacteria is "running" it moves with near uniform speed. However when "twiddling" the bacteria changes direction very abruptly. We will describe this using the stochastic model as given in [46]. We describe the bacteria by giving its position $x \in \mathbb{R}^3$ and velocity $v \in \mathbb{S}^2$ at each time, where $\mathbb{S}^2$ is the sphere in $\mathbb{R}^3$. Then there exists a function $\lambda : [0, \infty) \times \mathbb{R}^3 \times \mathbb{S}^2 \to (0, \infty)$ which describes the next time the bacteria twiddles; at such a twiddle the velocity changes according to some probability measure μ_v on $\mathbb{S}^2 \setminus \{v\}$ where v is the velocity before the twiddle. The dynamics of the bacteria are given as a PDMP described by the backward equation

$$\frac{\partial u}{\partial t}(t, x, v) + \langle v, \nabla_x u(t, x, v) \rangle + \lambda(t, x, v) \int_{\mathbb{S}^2} [u(t, x, \eta) - u(t, x, v)] \mu_v(\mathrm{d}\eta) = 0.$$

Note if λ is independent of t then we can describe this process by writing a generator in the form (3); otherwise we can extend the space to include a time variable and then write a corresponding generator in the form (3) which is given by

$$\mathcal{L}f(t, x, v) = \partial_t f(t, x, v) + \langle v, \nabla_x f(t, x, v) \rangle + \lambda(t, x, v) \int_{\mathbb{S}^2} [f(t, x, \eta) - f(t, x, v)] \mu_v(\mathrm{d}\eta).$$

We can introduce an approximation of this process by using frozen switching rates.

3.3. Higher order schemes

A natural question is how to obtain higher order schemes. The first important observation is that the probability that a PDMP has more than one jump in a time interval of length δ is of order δ^2. Therefore in order to construct higher order schemes it is natural to allow multiple jumps in the same time step. A detailed implementation of a higher order approximation scheme can be found in Algorithm 5. Let us first describe a second order algorithm. Starting at state $\overline{Z}_{t_n} = z$, the proposed time for the first event is given by $\bar{\tau}$ where

$$\mathbb{P}(\bar{\tau} > r) = \exp\left(- \int_0^r \overline{\lambda}(z, s; \delta_{n+1}, 2)\mathrm{d}s \right).$$

If $\bar{\tau} < \delta_{n+1}$, the process moves according to the numerical flow $\overline{\varphi}_s(z; \delta_{n+1}, 2)$ for time $\bar{\tau}$, and at time $t_n + \bar{\tau}$ the random event takes place according to $\overline{F}_{\bar{I}}(\overline{Z}_{t_n+\bar{\tau}}, \cdot; \delta_{n+1}, 2)$, where $\bar{I}$ has discrete distribution. In this case, a second jump is allowed in the current time step. The simulation of this event can be made using first order approximations $\overline{\lambda}(\cdot, \cdot; \delta_{n+1}, 1)$, $\overline{\varphi}_s(\cdot; \delta_{n+1}, 1)$, and $\overline{F}_i(\cdot, \cdot; \delta_{n+1}, 1)$.

Algorithm 4: Second order Partially Discrete Approximation of a PDMP

Input : Number of iterations N, initial condition z, step sizes $(\delta_n)_{n=0}^N$.
Output: Chain $(\overline{Z}_{t_n})_{n=0}^N$.
Set $n = 0$, $\overline{Z}_0 = z$;
while $n < N$ **do**

    set $\tilde{Z} = \overline{Z}_{t_n}$;
    draw $U^1 \sim \text{Unif}[0, 1]$ and simulate

$$\bar{\tau}_1 = \inf\left\{ r > 0 : 1 - \exp\left(-\int_0^r \overline{\lambda}(\tilde{Z}, s; \delta_{n+1}, 2)\mathrm{d}s \right) \geq U^1 \right\}$$

    if $\bar{\tau}_1 < \delta_{n+1}$ **then**

        draw $U^1_{n+1} \sim \nu_{\mathcal{U}}$ and $\bar{I}_1 \sim \text{Discrete}\left(\left\{ \frac{\overline{\lambda}_i(\tilde{Z}, \bar{\tau}_1; \delta_{n+1}, 2)}{\overline{\lambda}(\tilde{Z}, \bar{\tau}_1; \delta_{n+1}, 2)} \right\}_{i=1}^m \right)$;
        set $\tilde{Z} = \overline{\varphi}_{\bar{\tau}_1}(\tilde{Z}; \delta_{n+1}, 2)$;
        set $\tilde{Z} = \overline{F}_{\bar{I}_1}(\tilde{Z}, U^1_{n+1}; \delta_{n+1}, 2)$;
        draw $U^2 \sim \text{Unif}[0, 1]$ and simulate

$$\bar{\tau}_2 = \inf\left\{ r > 0 : 1 - \exp\left(-\int_0^r \overline{\lambda}(\tilde{Z}, s; \delta_{n+1}, 1)\mathrm{d}s \right) \geq U^2 \right\}$$

        if $\bar{\tau}_2 < t_{n+1} - \overline{\tau}_1$ **then**

            draw $U^2_{n+1} \sim \nu_{\mathcal{U}}$ and $\bar{I}_2 \sim \text{Discrete}\left(\left\{ \frac{\overline{\lambda}_i(\tilde{Z}, \bar{\tau}_2; \delta_{n+1}, 1)}{\overline{\lambda}(\tilde{Z}, \bar{\tau}_2; \delta_{n+1}, 1)} \right\}_{i=1}^m \right)$;
            set $\tilde{Z} = \overline{F}_{\bar{I}_2}(\tilde{Z}, U^2_{n+1}; \delta_{n+1}, 1)$;
            simulate $\overline{Z}_{t_{n+1}} = \overline{\varphi}_{t_{n+1} - \bar{\tau}_2 - \overline{\tau}_1}(\tilde{Z}; \delta_{n+1}, 1)$;
        else
            simulate $\overline{Z}_{t_{n+1}} = \overline{\varphi}_{t_{n+1} - \overline{\tau}_1}(\tilde{Z}; \delta_{n+1}, 1)$;
        end
    else
        set $\overline{Z}_{t_{n+1}} = \overline{\varphi}(\tilde{Z}; \delta_{n+1}, 2)$;
    end
end

Let us consider as an example how to obtain a second order approximation for smooth switching rates. For $s \leq \delta$ the first order Taylor approximation of $\lambda_i(\varphi_s(z))$ is given by

$$\overline{\lambda}_i(z, s; \delta_{n+1}, 2) = \lambda_i(z) + s\langle \Phi(z), \nabla\lambda_i(z)\rangle.$$

Because the integral in (4) is with respect to s, this choice of $\overline{\lambda}_i(z, s; \delta_{n+1}, 2)$ is such that computing the corresponding switching time is equivalent to computing the root of a second order polynomial. The downside is that an evaluation of the gradient of λ_i is needed and may be unavailable or expensive to compute. However, we can further approximate the product $\langle \Phi(z), \nabla\lambda_i(z)\rangle$ with a finite difference scheme to obtain for $s \leq \delta_{n+1}$ the expression

$$\overline{\lambda}_i(z, s; \delta_{n+1}, 2) = \lambda_i(z) + \frac{s}{\delta_{n+1}}(\lambda_i(\varphi_{\delta_{n+1}}(z)) - \lambda_i(z)), \tag{10}$$

Algorithm 5: Order p Partially Discrete Approximation of a PDMP

Input : Number of iterations N, initial condition z, step sizes $(\delta_n)_{n=0}^N$.
Output: Chain $(\overline{Z}_{t_n})_{n=0}^N$.
Set $n = 0$, $\overline{Z}_0 = z$;
while $n < N$ **do**
> set $q = p$, $\tilde{Z} = \overline{Z}_{t_n}$;
> set $t_{\text{left}} = \delta_{n+1}$;
> **while** $q > 0$ **do**
>> simulate
>> $$\bar{\tau} = \inf \left\{ r > 0 : 1 - \exp\left(- \int_0^r \overline{\lambda}(\tilde{Z}, s; \delta_{n+1}, q)\mathrm{d}s \right) \geq U \right\}$$
>> where $U \sim \text{Unif}[0, 1]$;
>> **if** $\bar{\tau} < t_{\text{left}}$ **then**
>>> draw $U_{n+1} \sim \nu_{\mathcal{U}}$ and $\overline{I} \sim \text{Discrete}\left(\left\{ \frac{\overline{\lambda}_i(\tilde{Z}, \bar{\tau}; \delta_{n+1}, q)}{\overline{\lambda}(\tilde{Z}, \bar{\tau}; \delta_{n+1}, q)} \right\}_{i=1}^m \right)$;
>>> set $\tilde{Z} = \overline{\varphi}_{\bar{\tau}}(\tilde{Z}; \delta_{n+1}, q)$;
>>> set $\tilde{Z} = \overline{F}_{\overline{I}}(\tilde{Z}, U_{n+1}; \delta_{n+1}, q)$;
>>> set $q = q - 1$ and $t_{\text{left}} = t_{\text{left}} - \bar{\tau}$;
>> **else**
>>> set $\tilde{Z} = \overline{\varphi}_{\delta_{n+1}}(\tilde{Z}; \delta_{n+1}, q)$;
>>> set $q = 0$;
>> **end**
> **end**
> set $\overline{Z}_{n+1} = \tilde{Z}$, $n = n + 1$;
end

---

which is a second order approximation provided λ is sufficiently smooth. The algorithm for $p = 2$ is given by Algorithm 4.

Similarly, it is possible to obtain an order $p > 2$ approximation. The simulation up to and counting the first event of each time step should be made according to approximations of order δ^p of the flow map, switching rates, and jump kernels. After the first event it is then possible to use approximations of order $p - 1$, then of order $p - 2$, and so on until one reaches the end of the current time interval, with the constraint that at most p events take place. Finally, it is clearly possible to use approximations of order δ^p for the simulation of all events in the same time step, although such approximations can be in general more expensive to compute.

4. Main results

4.1. Error bounds in Wasserstein distance

The main result of this section is Theorem 4.9, which shows convergence of the Wasserstein distance between the approximation and the continuous process as the step size goes to 0. We consider the Wasserstein distance of order 1 with respect to any normed distance, that is we

take $d(x, y) = \|x - y\|$ in Eq. (1) for any vector norm $\|\cdot\|$. For convenience we assume that for all $n \in \mathbb{N}$ we have an upper bound $\delta_n \leq \delta_0$.

Let us now state the assumptions on the process and on the various approximations that are required to show Theorem 4.9. We start with assumptions on the continuous time PDMP, and specifically from a condition on the deterministic dynamics. In particular, we require that Φ is Lipschitz.

Assumption 4.1. For the vector field Φ there exists a constant $C > 0$ such that for all $z, z' \in E$ it holds that

$$\|\Phi(z) - \Phi(z')\| \leq C\|z - z'\|.$$

We now shift our focus to the jump part of the process. In particular, we need the kernel $Q(z, \cdot)$ to satisfy the next conditions.

Assumption 4.2. There exist constants $D_1, D_2, D_3 > 0$ such that for $\tilde{U} \sim \nu_{\mathcal{U}}$ the following conditions hold for all $i \in \{1, \ldots, m\}$:

(a) For any $z \in E$

$$\mathbb{E}[\|z - F_i(z, \tilde{U})\|] \leq D_1.$$

(b) For all $z, z' \in E$

$$\mathbb{E}[\|F_i(z, \tilde{U}) - F_i(z', \tilde{U})\|] \leq D_2\|z - z'\|.$$

(c) For all $z \in E$ and all $s \leq \delta \leq \delta_0$

$$\mathbb{E}\left[\|\varphi_{\delta-s}(F_i(\varphi_s(z), \tilde{U})) - F_i(\varphi_\delta(z), \tilde{U})\|\right] \leq D_3\delta.$$

The first assumption asks that after a random jump the process is in expectation at bounded distance to its previous state, while condition (b) states that a Lipschitz condition with respect to the previous state holds for coupled jumps. Finally, condition (c) asks that the error committed by switching at the end of the time step or at an earlier time is of order δ if the two jumps are coupled. Moreover, the following Lipschitz condition for the switching rates is required.

Assumption 4.3. There exists $D_4 > 0$ such that for all $z, z' \in E$ and $i = 1, \ldots, m$

$$|\lambda_i(z) - \lambda_i(z')| \leq D_4\|z - z'\|.$$

Let us now focus on the required assumptions on the various approximations employed in the approximation process. We state the assumptions for a general order of accuracy $p \geq 1$, with $p \in \mathbb{N}$. Starting from the deterministic dynamics, we assume that the numerical integrator for the flow map is an approximation of order p.

Assumption 4.4. There exists $\tilde{C} \geq 0$ such that for any $z \in E$ and any $0 \leq s \leq \delta \leq \delta_0$

$$\|\varphi_s(z) - \overline{\varphi}_s(z; \delta, p)\| \leq \tilde{C}s^{p+1}.$$

In case the flow map can be simulated exactly, one can simply take $\overline{\varphi}_s = \varphi_s$ and $\tilde{C} = 0$. Next we focus on the approximate jump kernels $\overline{F}_i$.

Assumption 4.5. The approximate jump kernels $\overline{F}_i : E \times \mathcal{U} \times [0, \delta_0] \to E$, satisfy for any $z \in E$ and $\delta \in (0, \delta_0]$

$$\mathbb{E}_z[\|\overline{F}_i(z, \tilde{U}; \delta, p) - F_i(z, \tilde{U})\|] \leq M_1 \delta^p$$

for all $i = 1, \ldots, m$.

Let us now state the requirement on the approximate switching rates $\overline{\lambda}_i$.

Assumption 4.6. The following conditions hold:

(a) There exists $\overline{M}_2(z)$ such that for all $0 \leq s \leq \delta \leq \delta_0$ and $i \in \{1, \ldots, m\}$

$$|\overline{\lambda}_i(z, s; \delta, p) - \lambda_i(\varphi_s(z))| \leq \delta^p \overline{M}_2(z).$$

(b) For any $n \in \mathbb{N}$ there is a function $M_2(t, z)$ such that

$$\mathbb{E}_z\left[\overline{M}_2(\overline{Z}_{t_n})\right] \leq M_2(t_n, z) < \infty.$$

As a final assumption, we require that both the continuous time PDMP and the approximation process have almost surely bounded norm for a finite time horizon. This assumption is verified for instance if the state space is compact, or if the processes travel with bounded velocity.

Assumption 4.7. For any $t > 0$ there exists $B(t, z) > 0$ such that almost surely both $\|Z_t\| \leq B(t, z)$ and $\|\overline{Z}_t\| \leq B(t, z)$, where $Z_0 = \overline{Z}_0 = z$.

Note 4.8. Let us comment on these assumptions:

- It is worth observing that conditions such as Assumption 4.3 can be weakened to forms such as

$$|\lambda_i(z) - \lambda_i(z')| \leq D_4 \|z - z'\|(1 + \|z\|^q + \|z'\|^{q'}),$$

for some $q, q' \in \mathbb{N}$. This is because by Assumption 4.7 the norms at time t of the two processes are bounded almost surely and therefore for some $M(t)$ we have

$$(1 + \|Z_t\|^q + \|\overline{Z}_t\|^{q'}) \leq M(t) < \infty$$

almost surely. A similar reasoning can be applied to other assumptions that have this structure. For simplicity we will not consider this set of weakened assumptions in the proof of Theorem 4.9, but we remark that the extension is straightforward.
- In both Example 3.2 on the ZZS and Example 3.3 on the BPS we can write λ of the form

$$\lambda(x, v) = f(r)$$

where $r = \partial_i \psi(x) v_i$ for ZZS or $r = \langle \nabla \psi(x), v \rangle$ for BPS and $f(r) = r_+$. Note that it is possible to take a smooth function f for which the process still has the desired invariant measure, see [3]. We will demonstrate some choices of $\overline{\lambda}$ for ZZS which satisfy Assumption 4.6, and analogous choices hold for BPS. For smooth λ we can use (10) to obtain a second order approximation or similarly a pth order finite difference scheme to have an order p approximation. However if $\lambda(x, v) = (v_i \partial_i \psi(x))_+$ is only Lipschitz then this approximation is no longer valid; instead we can write

$$\overline{\lambda}_i((x, v), s; \delta, p) = (\overline{\partial_i \psi}((x, v), s; \delta, p) v_i)_+$$

where $\overline{\partial_i \psi}((x, v), s; \delta, p)$ is a pth order approximation in s of $\partial_i \psi(x + sv)$ and can be obtained either by a truncated Taylor expansion or using a finite difference scheme. Then using that $(\cdot)_+$ is 1-Lipschitz

$$|\overline{\lambda}_i((x, v), s; \delta, p) - \lambda_i(\varphi_s(x, v))| \leq |\overline{\partial_i \psi}((x, v), s; \delta, p) - \partial_i \psi(x + sv)| \leq \overline{M}_2 \delta^p.$$

For example, for ψ sufficiently smooth, we can take

$$\overline{\lambda}_i((x, v), s; \delta, p) = \left(\sum_{q=0}^{p-1} \frac{(sv_i)^q}{q!} \Delta_{i, \delta, p-q}^{q+1} \psi(x) \right)_+,$$

where $\Delta_{\delta, p-q}^q \psi$ denotes the δ^{p-q}-th order approximation of the qth derivative of ψ in the variable x_i.

We are ready to state the main result of this section.

Theorem 4.9. *Let $p \geq 1$. Denote by $\{\mathcal{P}_t\}_{t \geq 0}$ the semigroup of a PDMP with generator (3), which satisfies Assumptions 4.1–4.3. Denote by $\overline{\mathcal{P}}_t$ the transition probability of the Markov chain described by either Algorithms 2 or 3 in the case $p = 1$, or by Algorithm 5 for $p > 1$. Suppose that $\overline{\varphi}_t(\cdot; \delta, q), \overline{\lambda}(\cdot, \cdot; \delta, q), \overline{F}_i(\cdot; \delta, q)$ satisfy Assumptions 4.4–4.7 for some $\delta_0 > 0$ and for every $1 \leq q \leq p$ with $q \in \mathbb{N}$. Then for a fixed $T > 0$ there exist $K_1 = K_1(T)$, $K_2 = K_2(T)$ such that for any mesh $0 = t_0 < t_1 < \cdots < t_N = T$ with $\delta_n = t_n - t_{n-1}$ and $\delta_n \leq \delta_0$ for any $n \leq N$*

$$\mathcal{W}_1(\mathcal{P}_T(z, \cdot), \overline{\mathcal{P}}_T(z, \cdot)) \leq K_2 \sum_{k=1}^{N} \delta_k^{p+1} \left(\prod_{\ell=k}^{N} (1 + \delta_\ell K_1) \right).$$

If the step size is uniform, i.e. $\delta_n = \delta$ and $t_n = n\delta$, then

$$\mathcal{W}_1(\mathcal{P}_T(z, \cdot), \overline{\mathcal{P}}_T(z, \cdot)) \leq \delta^p \left(e^{TK_1} - 1 \right) \frac{K_2}{K_1}.$$

Proof of Theorem 4.9. The proof of Theorem 4.9 can be found in Section 6. $\quad \square$

We now give a setting in which Assumption 4.2 simplifies. This is motivated by and includes the ZZS. Let us now consider a PDMP $Z_t = (X_t, V_t) \in \mathbb{R}^n \times \mathcal{V}$, where X_t and V_t should be interpreted as the position and velocity at time t. Here $\mathcal{V}$ is some subset of Euclidean space. Consider the case in which the deterministic dynamics with initial condition (x, v) are of the form

$$\begin{cases} \dot{x} = \Phi(v), \\ \dot{v} = 0. \end{cases}$$

Therefore the deterministic motion is $X_t = \varphi_t(x, v)$ and $V_t = v$ if $(X_0, V_0) = (x, v)$. Then assume that the random events affect only the velocity, and leave the position unchanged, i.e. $F_i((x, v), U) = (x, F_i^v((x, v), U))$. This is the setting for example of the ZZS and BPS. Consider the following assumption.

Assumption 4.10. The space $\mathcal{V}$ is such that for all $v, w \in \mathcal{V}$ with $v \neq w$ it holds that

$$0 < V_{min} \leq \|v - w\| \leq V_{max} < \infty.$$

Assume also that there exists $D > 0$ such that for any $x, y \in \mathbb{R}^n$, $i \in \{1, \ldots, m\}$ and $v \in \mathcal{V}$

$$\mathbb{E}_{(x,v)}[\|F_i^v((x, v), U) - F_i^v((y, v), U)\|] \leq D\|x - y\|.$$

The next corollary states that in this setting Assumption 4.10 implies Assumption 4.2.

Corollary 4.11. *Consider a PDMP of the particular form described above. Suppose Assumptions 4.1, 4.3–4.6, as well as Assumption 4.10 hold. Then Theorem 4.9 applies.*

Proof. The proof can be found in Appendix A.2. □

Finally, we consider the setting in which we have a deterministic upper bound for the switching rates, but the process is not almost surely bounded as was required by Assumption 4.7. This is the case for instance of the Randomised HMC algorithm [12]. We shall show that in this case Theorem 4.9 holds as long as for a finite time horizon the processes are bounded in expectation. The formal condition is the following.

Assumption 4.12. There exists a constant $\lambda_{max} > 0$ such that $\lambda(z) \leq \lambda_{max}$ for all $z \in E$. Moreover there exists $L(t, z) < \infty$ such that

$$\max\{\mathbb{E}_z[\|Z_t\|], \mathbb{E}_z[\|\overline{Z}_t\|]\} \leq B(t, z).$$

Proposition 4.13. *Suppose Assumptions 4.1–4.6 and 4.12 hold. Then Theorem 4.9 applies.*

Proof. The proof is given in Appendix A.3. □

4.2. Error bounds in total variation distance

In this section we show that a bound of order δ^p on the total variation distance between the approximation and the PDMP can be derived for Algorithm 5 assuming it is possible to simulate exactly the flow φ_t and the Markov kernels Q_i. Interestingly, this result can be proved under considerably weaker assumptions on the PDMP compared to what is considered in Section 4.1. We remark in particular that no assumption on the maps F_i is needed, which was the case in Assumption 4.2. Moreover the process needs not be bounded almost surely for finite time horizons, as described in Assumption 4.7. The main result of this section is proved by coupling the event times of the PDMP and of the approximations via Poisson thinning. It follows that with a positive probability the processes, which are initialised at the same point, will remain together during a time step.

Let us state the required assumptions on the switching rates and on the continuous time process. Recall that for first order approximations of the characteristics we drop the specific order of accuracy, e.g. for switching rates we have $\overline{\lambda}_i(z, s; \delta) = \overline{\lambda}_i(z, s; \delta, 1)$ for $i = 1, \ldots, m$. We distinguish the assumptions between the setting $p = 1$ and $p > 1$. In the case $p = 1$ we impose the following assumption.

Assumption 4.14. Each of the approximate switching rates $\overline{\lambda}_i(\cdot; \delta)$ for $i = 1, \ldots, m$ satisfies Assumption 4.6(a) with $p = 1$ for some $\overline{M}_2(z)$. Furthermore for $z \in E$ and $s \geq 0$ define $\overline{\lambda}(z, s; \delta) = \sum_{i=1}^m \overline{\lambda}_i(z, s; \delta)$, and $\lambda_{tot}(z, s; \delta) = \lambda(z) + \overline{\lambda}(z, s; \delta) + m$. Let $T > 0$. Then there

exist $L_1(T, z)$, $L_2(T, z)$, $L_3(T, z) < \infty$ such that for any mesh $0 = t_0 < t_1 < \cdots < t_N = T$ with $t_{k+1} - t_k = \delta_{k+1}$ and $N \in \mathbb{N}$ the following conditions hold:

$$\sup_{n \leq N} \sup_{i=1,\ldots,m} \sup_{s \in [0,\delta_n]} \sup_{r \in [s,\delta_n]} \mathbb{E}_z \Big[\lambda(\varphi_s(F_i(\varphi_r(Z_{t_{n-1}}), \tilde{U}_n) \lambda_{tot}(Z_{t_{n-1}}, s; \delta)) \Big] \leq L_1(T, z),$$

$$\sup_{n \leq N} \sup_{s \in [0,\delta_n]} \mathbb{E}_z \Big[\overline{M}_2(Z_{t_{n-1}}) \lambda_{tot}(Z_{t_{n-1}}, s; \delta) \Big] \leq L_2(T, z),$$

$$\sup_{n \leq N} \sup_{s \in [0,\delta_n]} \sup_{r \in [s,\delta_n]} \mathbb{E}_z \Big[\big(\lambda(\varphi_r(Z_{t_{n-1}})) + \overline{\lambda}(Z_{t_{n-1}}, r; \delta) \big) \lambda_{tot}(Z_{t_{n-1}}, s; \delta) \Big] \leq L_3(T, z).$$

For the case $p > 1$ we make the following assumption. Recall in the case $p > 1$ if in a single time step there have been q jumps then we use $\overline{\lambda}_i(\cdot; \delta, p - q)$ to simulate the next jump time. As the probability of there having been q jumps in a time interval is order δ^q the conditions required on $\overline{\lambda}_i(\cdot; \delta, q)$ are lessened, for this reason there are different requirements for each q.

Assumption 4.15.   Each of the approximate switching rates $\overline{\lambda}_i(\cdot; \delta, q)$ for $i = 1, \ldots, m$ and $q = 1, \ldots, p$ satisfies Assumption 4.6(a) for some $\overline{M}_2(z)$. When $q = 1$ the approximate switching rates $\overline{\lambda}_i(\cdot; \delta, 1)$ for $i = 1, \ldots, m$ satisfy Assumption 4.14. We make the additional moment bound for any $1 \leq q \leq p$

$$\sup_{n \leq N} \sup_{s \in [0,\delta_n]} \mathbb{E}_z \Big[(1 + \overline{M}_2(Z_{t_{n-1}})) \lambda_{tot}(Z_{t_{n-1}}, s; \delta, q) + \lambda_{tot}(Z_{t_{n-1}}, s; \delta, q)^{q+1} \Big] \leq L_4(T, z).$$

Note 4.16.   The moment bounds in Assumption 4.14 are rather technical, but also general. For instance Assumption 4.14 holds if Assumptions 4.6 and 4.7 hold, i.e. when the process has bounded norm for any finite time horizon. Furthermore, as Assumption 4.14 does not depend on moment bounds for the approximate process $\{\overline{Z}_{t_n}\}_{n \geq 1}$, one can verify Assumption 4.14 by finding a suitable Lyapunov function for the PDMP. Indeed if there exists a Lyapunov function which bounds the functions appearing in Assumption 4.14 then Assumption 4.14 holds with L_1, L_2, L_3 independent of T. This is the case for instance of the ZZS and BPS, see Example 5.5. Alternatively, one can take advantage of Holder's inequality to reduce the problem to bounding polynomial moments of the various quantities. In Section 5.2 we show that the assumption holds for several examples. Finally we remark that in Assumption 4.14 it is possible to substitute $Z_{t_{n-1}}$ with $\overline{Z}_{t_{n-1}}$ and Theorem 4.17 still holds.

Theorem 4.17.   *Denote as $\overline{\mathcal{P}}_t(z, \cdot)$ the transition probability of the approximation process obtained by Algorithm 3 for $p = 1$ or by Algorithm 5 for $p > 1$. Denote by $\{\mathcal{P}_t\}_{t \geq 0}$ the semigroup of a PDMP with generator (3) satisfying Assumption 4.1. Let $p \geq 1$ and suppose the approximations $\overline{\lambda}_i(z, s; \delta, q)$ for $q \leq p$ satisfy Assumption 4.14 if $p = 1$ or Assumption 4.15 if $p > 1$. Suppose the mesh $t_n = \sum_{i=1}^n \delta_n$ is such that $\delta_n < \delta_0$ for δ_0 as in Assumption 4.6(a). Suppose that $\overline{\varphi}_s = \varphi_s$ and $\overline{F}_i = F_i$ for all $i = 1, \ldots, m$. Then for any $z \in E$ and any mesh $0 = t_0 < t_1 < \cdots < t_N = T$ with $\delta_n = t_n - t_{n-1}$ and $\delta_n \leq \delta_0$ for any $n \leq N$*

$$\|\mathcal{P}_T(z, \cdot) - \overline{\mathcal{P}}_T(z, \cdot)\|_{TV} \leq \sum_{i=1}^N \delta_i^{p+1} D(T, z) \prod_{\ell=i+1}^N (1 - D(T, z)\delta_\ell),$$

where $D(t, z)$ is a non-decreasing function of t. If $\delta_n = \delta$ for all $n \in \mathbb{N}$ then

$$\|\mathcal{P}_T(z, \cdot) - \overline{\mathcal{P}}_T(z, \cdot)\|_{TV} \leq 1 - e^{-D(T,z)T\delta^p}.$$

Proof. The proof can be found in Section 7.   □

Note 4.18. Let us for simplicity consider the constant step size case. If we fix a time horizon t, then the theorem shows that $\|\mathcal{P}_t(z, \cdot) - \overline{\mathcal{P}}_t(z, \cdot)\|_{TV} \to 0$ as $\delta \to 0$. On the other hand, the upper bound tends to 1 as $T \to \infty$ if the step size δ is fixed. Moreover, because $1 - \exp(-D(t_n, z)t_n\delta) \le D(t_n, z)t_n\delta$ we have

$$\|\mathcal{P}_{t_n}(z, \cdot) - \overline{\mathcal{P}}_{t_n}(z, \cdot)\|_{TV} \le D(t_n, z)\, t_n\, \delta^p$$

and therefore we have convergence of order δ^p as $\delta \to 0$.

Note 4.19. In a similar fashion to [26], it is possible to obtain a bound as that in Theorem 4.17 also when the jump kernel is approximated. To prove such result it is sufficient to define a coupled jump kernel that keeps the two processes together with strictly positive probability if they are together right before the jump.

4.3. Convergence to the invariant measure

In this section we give conditions for the approximation process $\{\overline{Z}_{t_n}\}_{n\ge 1}$ to converge to μ, the invariant measure of the PDMP, which we shall assume to exist and be unique. We do this by showing convergence in law to the PDMP uniformly in time and requiring that the PDMP converges to its invariant measure. In the following we consider the case of geometric convergence as it is verified for a range of PDMPs, however convergence with any rate $r(t)$ which is integrable over $[0, \infty)$ is sufficient.

The strategy of this proof is inspired by [18], which uses derivative estimates to obtain uniform in time convergence of an Euler Scheme for an SDE. In that case the authors rely on having exponential decay of the derivatives of the semigroup for the SDE of interest, for which conditions are given in [19].

Assumption 4.20.   Let $\{Z_t\}_{t\ge 0}$ be a PDMP with corresponding generator (5). Recall the definition of Q given by (6). We assume the following:

(a) There exists an invariant measure, μ, for the PDMP, $\{Z_t\}_{t\ge 0}$, and μ is invariant under Q, that is

$$\mu(Qf) = \mu(f)$$

for any f measurable and integrable.

(b) The Markov process $\{Z_t\}_{t\ge 0}$ is geometrically ergodic with invariant measure μ. Specifically fix $\overline{G} : E \to [1, \infty)$ and define $\mathcal{G} = \{$measurable $g : E \to \mathbb{R}, |g| \le \overline{G}\}$. Assume that $\overline{G}(Z_t)$ is integrable for all $t \ge 0$. For some $R_1 > 0$, $\omega > 0$

$$\sup_{g \in \mathcal{G}} |\mathbb{E}_z[g(Z_t)] - \mu(g)| \le R_1 e^{-\omega t}\overline{G}(z). \tag{11}$$

This Assumption has been shown in a variety of cases, for example for the 1-dimensional Zig-Zag process this was shown in [9, Theorem 5] and for higher dimensions in [10, Theorem 2]. For BPS this was shown in [23,25]. For RHMC see [12, Theorem 3.9].

The following assumption is required for Algorithm 2, but not for Algorithm 3, for the reasons explained in Note 4.25. In general, derivative estimates on the semigroup are useful for proving convergence of approximations as they control the effect of a small error in the initial

condition of a stochastic process. In this case we are not using explicitly a derivative estimate but instead the operator $[\Phi, Q]$. The role of this commutator is to describe the difference in the direction of the process over an infinitesimal time interval if the process first jumps then follows the flow map or first follows the flow map and then jumps.

Assumption 4.21. Let $\{\mathcal{P}_t\}_{t\geq 0}$ denote the semigroup corresponding to the generator $\mathcal{L}$ given by (5). Recall the notation $[\Phi, Q]$ defined in Section 2. Let $\overline{G}$ and $\mathcal{G}$ be given as in Assumption 4.20. There exist some $R_2 > 0$, $\omega > 0$ and set $\mathcal{G}_1 \subseteq \mathcal{G}$, such that for all $t \geq 0$ we have

$$\sup_{g\in\mathcal{G}_1} \sup_{\delta\in(0,\delta_0),s\in[0,\delta]} [\Phi, Q](\mathcal{P}_t g \circ \varphi_{\delta-s})(\varphi_s(z)) \leq R_2 e^{-\omega t}\overline{G}(z).$$

In Example 5.11 we show that this assumption is satisfied for ZZS with a non-trivial set $\mathcal{G}_1$. Finally, we require the following moment bounds.

Assumption 4.22. Let $\{\overline{Z}_t\}_{t\geq 0}$ be the process described by Algorithm i where i is either 2 or 3 and suppose Assumption 4.20 holds for the function $\overline{G}$. Recall the definition of λ and Q given by (6). Define for $i = 2$ or 3 (corresponding to Algorithm i)

$$\begin{aligned}
\overline{G}_i(z, r, s) = &K_i(z, r, s) + \lambda(\varphi_r(z))Q((Q\overline{G} + \overline{G})\lambda)(\varphi_{s-r}(z)) \\
&+ \lambda(\varphi_r(z))\lambda(\varphi_s(z))(Q\overline{G}(\varphi_s(z)) + \overline{G}(\varphi_s(z)))
\end{aligned} \tag{12}$$

where

$$\begin{aligned}
K_2(z, r, s) &= \left(\overline{G}(z)\overline{\lambda}(z, s; \delta) + K_3(z, r, s)\right), \\
K_3(z, r, s) &= \left((Q\overline{G}(\varphi_s(z)) + \overline{G}(\varphi_s(z)))(\overline{\lambda}(z, s; \delta)\overline{\lambda}(z, r; \delta) + \overline{M}_2(z))\right).
\end{aligned}$$

For $i = 2$ or 3 there exists a function $H_i(z)$ such that for any mesh $0 = t_0 \leq t_1 \leq \cdots$ with $\delta_k = t_k - t_{k-1} < \delta_0$ for any k

$$\mathbb{E}_z\left[\sup_{0\leq r<s\leq\delta_0} \overline{G}_i(\overline{Z}_{t_k}, r, s)\right] \leq CH_i(z).$$

Note 4.23. Observe that since $\overline{G}$ is a Lyapunov function for the PDMP $\{Z_t\}$ we have that $\mathbb{E}_z[\overline{G}(Z_t)]$ is bounded in t for any z. Since $\{\overline{Z}_{t_n}\}_{n\geq 0}$ is designed to be a good approximation of $\{Z_t\}_{t\geq 0}$ we may hope that $\mathbb{E}_z[\overline{G}(\overline{Z}_{t_n})]$ is also bounded in n. We confirm this for ZZS and BPS in 1 dimension in Lemma C.2 and test numerically in a higher dimensional setting.

In each of the references discussed after Assumption 4.20 there is some freedom in the choice of parameters in the Lyapunov function. By adjusting these parameters we can bound the terms in $\overline{G}_i(z)$ appearing in Assumption 4.22 by using a different choice of the parameters of the Lyapunov function. Confirming Assumption 4.22 then reduces to showing that, for a fixed Lyapunov function $\overline{G}$ for the PDMP, we have

$$\sup_n \mathbb{E}_z[\overline{G}(\overline{Z}_{t_n})] < \infty.$$

Theorem 4.24. *Let $\{Z_t\}_{t\geq 0}$ be the PDMP with generator given by (5). Let $\{\overline{Z}_t\}_{t\geq 0}$ be the process described by Algorithm 2 or 3, with $\overline{\varphi} = \varphi$ and $\overline{F} = F$. Suppose that Assumptions 4.6(a), 4.20, 4.22 hold and that if $\{\overline{Z}_t\}_{t\geq 0}$ is described by Algorithm 2 that Assumption 4.21 holds also. Let $\mathcal{G}_1 \subseteq \mathcal{G}$ be given as in Assumption 4.21 if this assumption is required and $\mathcal{G}_1 = \mathcal{G}$ otherwise.*

Then there exists $K > 0$ which depends only on R_1, R_2 and C such that for any $g \in \mathcal{G}_1$, $n \in \mathbb{N}$, $z \in E$ we have

$$\left| \mathbb{E}_z[g(Z_{t_n})] - \mathbb{E}_z[g(\overline{Z}_{t_n})] \right| \le K S_n H_i(z). \tag{13}$$

Here

$$S_n = \sum_{k=0}^{n-1} \delta_{k+1}^2 e^{-\omega(t_n - t_{k+1})}. \tag{14}$$

Proof of Theorem 4.24. The proof of this theorem is given in Section 8. □

The choice of the set $\mathcal{G}_1$ here determines the type of convergence that we obtain. For example if $\mathcal{G}_1$ contains the set of continuous functions with supremum norm bounded by 1 then this corresponds to convergence in the total variation distance. On the other hand, if $\mathcal{G}$ contains the set of functions with Lipschitz constant less than 1 then we have convergence in the Wasserstein distance of order 1. Since we do not require Assumption 4.21 to hold when we use Algorithm 3 we can typically take $\mathcal{G}_1 = \mathcal{G}$ in that case and hence we have convergence in a metric that is stronger than total variation. However for Algorithm 2 we need an additional bound on the derivatives of the function so we have convergence in a weaker metric, see Example 5.11.

Note 4.25. An important estimate in the proof of Theorem 4.24 will be obtaining a bound between the law of the first jump of the PDMP, τ, and of the approximation process, $\overline{\tau}$. This is done in Lemma C.1. In this lemma we need to treat Algorithm 2 differently to Algorithm 3. In particular, we show convergence as $\delta \to 0$ by considering $\mathbb{E}[h(\tau)] - \mathbb{E}[h(\overline{\tau})]$ for a class $\mathcal{C}$ of test functions h. In the case of Algorithm 3 we use the set $\mathcal{C} = \mathcal{C}_b([0, \delta])$ of test functions whereas in the case of Algorithm 2 we use the set $\mathcal{C} = \mathcal{C}_b^1([0, \delta])$. The result of using this weaker convergence is that we need a form of derivative estimate. The derivative estimate we require is given by Assumption 4.21 and is needed only if we are considering Algorithm 2.

Note 4.26. To simplify the exposition we have only considered the case when we can simulate the flow exactly. We can extend this proof to allow also for the use of a numerical integrator provided we have a suitable derivative bound. More precisely we require that for some $R_1 > 0$, $\omega > 0$ and some set $\mathcal{G}_1 \subseteq \mathcal{G}$ and for any $\delta \le \delta_0, t > 0, z \in E, i \in \{1, \dots, m\}$

$$\sup_{g \in \mathcal{G}_1} |\mathcal{P}_t g(\overline{\varphi}_\delta(z)) - \mathcal{P}_t g(\varphi_\delta(z))| \le \delta^2 R_3 e^{-\omega t} \overline{G}(z), \tag{15}$$

$$\sup_{g \in \mathcal{G}_1} |Q_i \mathcal{P}_t g(\overline{\varphi}_\delta(z)) - Q_i \mathcal{P}_t g(\varphi_\delta(z))| \le \delta^2 R_3 e^{-\omega t} \overline{G}(z).$$

Now using the uniform in time weak error estimate (13) and exponential ergodicity (11) we can show convergence to the invariant measure of the PDMP.

Corollary 4.27. *Suppose that the conclusion of Theorem 4.24 holds. Set $\delta_k = \delta$ for all $k \in \mathbb{N}$. Then for $g \in \mathcal{G}_1$ we have*

$$\left| \mathbb{E}_z[g(Z_{t_n})] - \mathbb{E}_z[g(\overline{Z}_{t_n})] \right| \le C \delta H(z), \tag{16}$$

$$\left| \mathbb{E}_z[g(\overline{Z}_{t_n})] - \mu(g) \right| \le C H(z)(\delta + e^{-\omega t_n}) \tag{17}$$

$$\left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_z[g(\overline{Z}_{t_n})] - \mu(g) \right| \le C H(z) \left(\delta + \frac{1}{t_N} \right). \tag{18}$$

Proof of Corollary 4.27. The proof of this corollary is given in Appendix C.1. □

Corollary 4.28. *Suppose that the assumptions of Theorem 4.24 hold. Assume that $\delta_k \to 0$ as $k \to \infty$ and $\sum_{k=1}^{\infty} \delta_k = \infty$. For any $g \in \mathcal{G}_1$ we have*

$$\lim_{n \to \infty} \left| \mu(g) - \mathbb{E}_{x,v}[g(\overline{X}_{t_n}, \overline{V}_{t_n})] \right| = 0.$$

Proof of Corollary 4.28. The proof of this corollary is given in Appendix C.1. □

5. Examples

5.1. Examples for Section 4.1

Example 5.1 (*Zig-Zag Sampler Continued*). We continue Example 3.2 checking that the conditions of the previous section are satisfied. Let us check that approximations of the ZZS based on Algorithm 2 or 3 satisfy Corollary 4.11. Assumption 4.1 clearly holds. Assumption 4.7 holds because the process travels with bounded velocity, so we can apply the reasoning in Note 4.8 to verify Assumption 4.3. In particular, Assumption 4.2 holds as long as $\psi \in \mathcal{C}^2$ and γ_i is locally Lipschitz for all $i \in \{1, \ldots, m\}$. Assumptions 4.4 and 4.5 follow from the fact that we can simulate exactly the flow and the kernels. Assumption 4.6(a) is satisfied for $p = 1$ both for $\overline{\lambda}_i(z, s; \delta) = \lambda_i(z)$ and (9) for $\psi \in \mathcal{C}^2$. Assumption 4.6(b) follows from Assumption 4.7. Finally Assumption 4.10 clearly holds for any $D > 0$.

Note that we could define the same algorithm with $m = 1$ according to Note 3.1. However, in this case neither Assumption 4.2 nor Assumption 4.10 holds as the function F is not Lipschitz.

In Fig. 1(a) we demonstrate numerically the difference between the ZZS with 50-dimensional Gaussian target and the approximation scheme corresponding to Algorithm 2 with constant step size δ and frozen switching rates.[1] In this plot the two processes have been coupled according to Coupling 6.1, which is a synchronous coupling that is used in Section 6 to prove Theorem 4.9. In the figure we see that as δ tends to zero that the distance between the two processes converges to zero. We also observe there is an upper bound on how large the error can get, which roughly corresponds to the velocities having the opposite sign, i.e. $\overline{V}_{t_n} = -V_{t_n}$.

Example 5.2 (*Bouncy Particle Sampler Continued*). We continue Example 3.3 and discuss the assumptions of this section in this context. We show that $x \mapsto R(x)v$ need not be Lipschitz. Indeed, for a Gaussian example with $d > 1$ fix $v \in \{1, -1\}^d$ and take $y \in \mathbb{R}^d$ orthogonal to v then for any $s > 0$ consider

$$\|R(sv)v - R(y)v\| = 2 \left\| \frac{\langle v, sv \rangle}{\|sv\|^2} sv \right\| = 2\|v\|.$$

Letting s tend to zero we see that $x \mapsto R(x)v$ is not Lipschitz at zero and hence Assumption 4.2 does not hold.

In Fig. 1(b) we demonstrate numerically the difference between the BPS with 50-dimensional Gaussian target, Gaussian refreshments with rate 1 and the approximation scheme corresponding to Algorithm 2 with constant step size δ, and frozen switching rates when coupled according to Coupling 6.1. We see that although the assumptions of the theory do

---

[1] The codes for all experiments in this paper can be found in a dedicated GitHub repository at https://github.com/andreabertazzi/Euler_PDMC_algorithms.
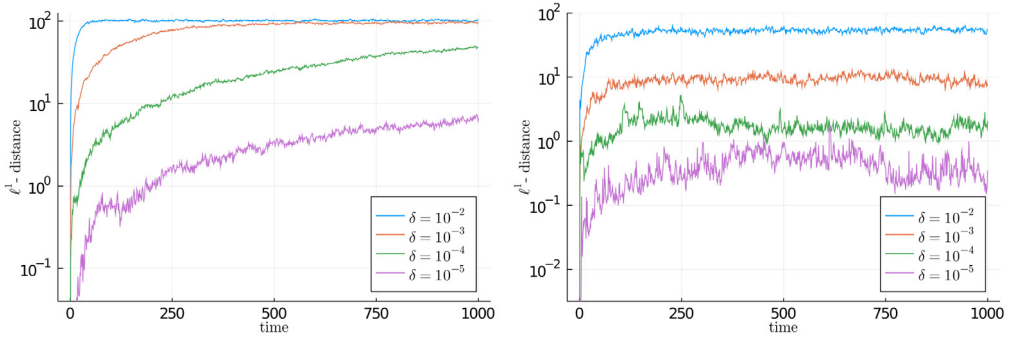
(a) Results for the ZZS with $\gamma(x, v) = 0$. (b) Results for the BPS with $\lambda_r = 1$.

Fig. 1. Plots of the distance between the continuous time PDMPs and their approximations given by Algorithm 2 for several values of the step size. The x-axis shows continuous time units, i.e. the time of $\overline{Z}_{t_n}$ is $t_n = n\delta$. The distance is $\|x - y\|_1 = \sum_{i=1}^{d} |x_i - y_i|$. The plots show the average of 50 experiments. The processes are coupled according to Coupling 6.1. The continuous PDMPs have a 50-dimensional standard Gaussian as stationary measure. Here we choose $\overline{\lambda}((x, v), s; \delta) = \lambda(x, v)$.

not hold the error appears to tend to zero as $\delta \to 0$. Indeed in Section 4.2 we obtain theory supporting this observation. Moreover, from the plot it appears that the error converges as $\delta \to 0$ uniformly in time. We will investigate this property further in Section 4.3.

Example 5.3 (*Randomised Hamiltonian Monte Carlo Algorithm Continued*). We continue Example 3.4. As long as $\nabla \psi$ is Lipschitz, Proposition 4.13 can be applied to the approximations based on Algorithms 2 and 3.

Example 5.4 (*PDMP Two-Dimensional Morris–Lecar Model [35]*). Let us consider the PDMP defined on $E = \{0, \ldots, N_K\} \times \mathbb{R}$ whose characteristics are given by

$$\Phi(\theta, v) = \begin{pmatrix} 0 \\ \frac{1}{C}\left(1 - g_{\text{Leak}}(v - V_{\text{Leak}}) - g_{\text{Ca}}M_\infty(v)(v - V_{\text{Ca}}) - g_K \frac{\theta}{N_K}(v - V_K)\right) \end{pmatrix},$$

$$\lambda(\theta, v) = (N_K - \theta)\alpha_K(v) + \theta\beta_K(v),$$

$$Q((\theta, v), (\theta + 1)) = \frac{(N_K - \theta)\alpha_K(v)}{\lambda(\theta, v)}, \quad Q((\theta, v), \{\theta - 1\}) = \frac{\theta\beta_K(v)}{\lambda(\theta, v)},$$

$$M_\infty(v) = (1 + \tanh((v - V_1)/V_2))/2,$$

$$\alpha_K(v) = \lambda_K(v)N_\infty(v), \quad \beta_K(v) = \lambda_K(v)(1 - N_\infty(v)),$$

$$N_\infty(v) = (1 + \tanh((v - V_3)/4))/2, \quad \lambda_K(v) = \bar{\lambda}_K \cosh((v - V_3)/2V_4).$$

This model was given in [35] and is a PDMP version of the deterministic Morris–Lecar model introduced in [39] to explain the dynamics of the barnacle muscle fibre. Here v denotes the membrane potential, θ is number of open Potassium channels, $g_{\text{Leak}}, g_{\text{Ca}}, g_K$ is maximum conductance value for leak, Calcium, and Potassium respectively, C is the membrane capacitance, $V_{\text{leak}}, V_{\text{Ca}}, V_K$ is the equilibrium potential of relevant ion channels, $M_\infty(v)$ ($N_\infty(v)$ respectively) is the fraction of open Calcium (Potassium resp.) channels at steady state. V_1 (V_3 respectively) is the potential at which $M_\infty = 0.5$ ($N_\infty = 0.5$ resp.). V_2 (respectively V_4) is the reciprocal is the slope of the voltage dependence of M_∞ (N_∞ resp.).

We will consider a PD-PDMP approximation of this PDMP. Note that in this case the flow does not have an explicit solution so a numerical integrator is required. Therefore we will set $\overline{\varphi}_t$ to be an Euler approximation of φ_t and $\overline{\lambda}((\theta, v), t; \delta) = \lambda(\overline{\varphi}_t(\theta, v; \delta)) = \lambda((\theta, v) + t\,\Phi(\theta, v))$. Note we can simulate jump times with this approximate rate using Poisson thinning. Since the kernel Q can be simulated exactly we do not need to approximate this. This algorithm is very similar to the approximation proposed in [35] and we confirm their results in our framework. Indeed, one can verify that Assumptions 4.1, 4.3–4.6, as well as Assumption 4.10 hold so by Theorem 4.9 we have that the approximation converges as $\delta \to 0$ to the PDMP.

5.2. Examples for Section 4.2

It is straightforward to verify Assumption 4.14 for either ZZS or BPS. Below we give details for BPS.

Example 5.5 (*Bouncy Particle Sampler Continued*)**.** We continue Examples 3.3 and 5.2 and discuss when we may apply Theorem 4.17 in this setting. Recall in Example 5.2 we showed that we cannot expect BPS to satisfy the assumptions of Theorem 4.9 because the reflection operator is in general not Lipschitz. However, we do not need any assumption of this type for Theorem 4.17. Consider for instance a simple example in which $\overline{\lambda}_1((x, v), s) = \lambda_1(x, v)$. Then Assumption Assumption 4.6 (a) follows provided $\psi \in \mathcal{C}^2$. If in particular ψ has bounded Hessian, then $\overline{M}_2(x, v) \leq \|v\|\|\nabla^2 U\|_\infty$. It remains to verify the moment bounds in Assumption 4.14 hold. It is clear that these are satisfied if the velocities are bounded, as for instance when refreshments are from $\mathscr{Z}_{n+1} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ where $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$. On the other hand, for the BPS with Gaussian refreshments we observe that outside of a compact set one can bound the moments in Assumption 4.14 by the expectation of the Lyapunov functions derived in [23] or [25]. Therefore we can apply Theorem 4.17 to obtain convergence as $\delta \to 0$.

Let us derive a rough estimate on the dimensional dependence of Theorem 4.17 in the $p = 1$ case. In particular we focus on the dependence of $D(t_n, z)$ in the dimension. Observe from the proof of the theorem that $D(t, z)$ depends linearly on L_1, L_2, L_3, and thus it is sufficient to check the dimensional dependence of such constants. By applying Cauchy–Schwarz inequality the interesting terms are of the form $\mathbb{E}_z[(\lambda_1(X_t, V_t))^2]$. We approximate this expectation with its value in stationarity. Let us restrict to the case of Gaussian refreshments and a standard Gaussian invariant measure. Then in stationarity we obtain

$$\mathbb{E}_\pi[(\lambda_1(X, V))^2] = \mathbb{E}_\pi[\langle V, X\rangle_+^2] \leq d^2.$$

Therefore we expect $D(t, z)$ to have a quadratic dependence in the dimension of the PDMP. In order to obtain a fixed error in total variation distance one should then choose δ such that $D(t, z)\delta$ is constant, and thus δ of order d^{-2}. Observe that taking refreshments on the unit sphere the dependence would be linear in d.

Example 5.6 (*Continuous Time Approximations of PDMP*)**.** So far we have concentrated on discrete time approximations of PDMP, however it is also possible to apply our results to approximate a PDMP with a second continuous time PDMP. Similarly to the setting of [33] suppose we have an approximation $\tilde{\lambda}_i$ of λ_i, i.e. there exists $\varepsilon > 0$ such that for all $i \in \{1, \ldots, m\}$ we have

$$|\tilde{\lambda}_i(z) - \lambda_i(z)| \leq \tilde{M}(z)\varepsilon. \tag{19}$$

A possible motivation for this approach is the case of ZZS when we either cannot evaluate $\partial_i \psi$ exactly or it is too expensive to do so. Then we can use an approximation $\overline{\partial_i \psi}$ to obtain an approximation of λ, i.e. $\tilde{\lambda}_i = (v_i \overline{\partial_i \psi})_+$. Now we define a PDMP with approximated rates $\tilde{\lambda}_i$ which moves according to the generator $\tilde{\mathcal{L}}$ acting on sufficiently smooth functions by

$$\tilde{\mathcal{L}} f(z) = \langle \phi(z), \nabla_z f(z) \rangle + \sum_{i=1}^m \tilde{\lambda}_i(z) \int (f(z') - f(z)) Q_i(z, \mathrm{d}z'). \tag{20}$$

We are interested in comparing this process to the PDMP with generator $\mathcal{L}$ given by (3). In order to use Theorem 4.17 we introduce a discrete time process which we can use to compare to both the PDMPs corresponding to $\mathcal{L}$ and $\tilde{\mathcal{L}}$. Set $\delta = \varepsilon$ and $t_n = n\delta$, define $\{\overline{Z}_{t_n}\}_{n \in \mathbb{N}}$ to be given by Algorithm 3 with rates $\overline{\lambda}_i(z, t) = \tilde{\lambda}_i(\varphi_t(z))$, and according to the exact flow $\overline{\varphi}_t = \varphi_t$ and Markov kernels Q_i, i.e. $\overline{F}_i = F_i$. We assume that the moment bounds of Assumption 4.14 are satisfied which is clear for example if the processes move with bounded velocity. We may apply Theorem 4.17 both when the PDMP is given by $\mathcal{L}$ and by $\tilde{\mathcal{L}}$. Therefore for any $t > 0$ there exists a constant $D(t, z) > 0$ such that

$$\|\mathcal{P}_t(z, \cdot) - \tilde{\mathcal{P}}_t(z, \cdot)\|_{TV} \leq 2(1 - e^{-D(t,z) t \varepsilon}) \leq 2 D(t, z) t \varepsilon.$$

Here $\{\mathcal{P}_t\}_{t \geq 0}$ denotes the semigroup of a PDMP with generator $\mathcal{L}$ and $\{\tilde{\mathcal{P}}_t\}_{t \geq 0}$ denotes the semigroup of a PDMP with generator $\tilde{\mathcal{L}}$. We remark that the analysis above could be adapted to the setting of the Numerical Zig-zag algorithm introduced in [17].

Example 5.7 (*ZZS with Subsampling*). In Bayesian statistics the posterior distribution $\pi(x) \propto \exp(-\psi(x))$ is often of the form $\psi(x) = \sum_{j=1}^N \psi_j(x)$, where $\psi_j(x)$ depends only on a subset of the data. As described in [7], the ZZS allows for exact subsampling, which means that the simulation of each event time is calculated using ψ_J for some $J \sim \mathrm{Unif}\{1, \ldots, N\}$ instead of the full negative log-density ψ. This is achieved by defining switching rates $\lambda_i^j(x, v) = (v_i \partial_i \psi_j(x))_+$ and computational bounds $M_i(t)$ such that $\lambda_i^j(x + vt, v) \leq M_i(t)$. Then starting at state (x, v) at time t, a proposal for the next event time is found by taking $\tau_{i*} = \min \tau_i$, where τ_i has rate $M_i(t)$ for $i = 1, \ldots, d$. Then the proposal is accepted with probability $\lambda_{i*}^J(x + v\tau_{i*}, v)/M_{i*}(\tau_{i*})$ for $J \sim \mathrm{Unif}\{1, \ldots, N\}$ and in case of acceptance we set $V_{t+\tau_{i*}} = R_{i*} V_t$.

Motivated by the fact that the bounds $M_i(t)$ may be unavailable or hard to compute, we can approximate this process as follows. Here we restrict to the case of frozen switching rates, that is

$$\overline{\lambda}_i^j((x, v), s; \delta) = \lambda_i^j(x, v).$$

We apply the same idea behind Algorithm 3 to obtain $(\overline{X}_{t_{n+1}}, \overline{V}_{t_{n+1}})$ given the previous state by first drawing $J \sim \mathrm{Unif}\{1, \ldots, N\}$, and then simulating the next switching time $\overline{\tau} = \overline{\tau}_{i*} = \min \overline{\tau}_i$ with rates $\lambda_i^J(\overline{X}_{t_n}, \overline{V}_{t_n})$. Finally

$$\overline{X}_{t_{n+1}} := \overline{X}_{t_n} + (\delta_{n+1} - \overline{\tau}) \overline{V}_{t_n} + \overline{\tau} \overline{V}_{t_{n+1}}$$

$$\overline{V}_{t_{n+1}} := \begin{cases} R_{i*} \overline{V}_{t_n} & \text{if } \overline{\tau} \leq \delta_{n+1}, \\ \overline{V}_{t_n} & \text{if } \overline{\tau} > \delta_{n+1}, \end{cases}$$

where the operator R_i was defined in Example 3.2.

The fundamental difference with respect to the setting of Algorithm 3 lies in the additional level of randomness introduced by the random variable J. We can adapt the proof of Theorem 4.17 to also allow this additional randomness. Indeed in each step we use a synchronous

coupling of the random variable J, then conditional on J we can apply Coupling 7.1 with λ_i replaced by λ_i^J in (35)–(36), and setting

$$\lambda_{tot}^i((x, v), t; \delta) = \sum_{j=1}^{N} \lambda_i^j(x + vt, v) + \sum_{j=1}^{N} \lambda_i^j(x, v) + 1.$$

Thus provided $\psi \in \mathcal{C}^2$ for any $(x, v) \in E$ and $t > 0$ there exists $D = D(t, (x, v)) > 0$ such that

$$\|\mathcal{P}_t((x, v), \cdot) - \overline{\mathcal{P}}_t((x, v), \cdot)\|_{TV} \leq 1 - e^{-Dt\delta},$$

where $\{\mathcal{P}_t\}_{t \geq 0}$ is the semigroup of ZZS with subsampling and $\{\overline{\mathcal{P}}_t\}_{t \geq 0}$ is the transition probability of the approximation. We remark that a similar reasoning can be applied to the BPS with subsampling (see [13]).

5.3. Examples for Section 4.3

Example 5.8 (*Randomised Hamiltonian Monte Carlo Algorithm Continued*)**.** We continue Example 3.4 by verifying the various assumptions for Theorem 4.24. For this example we assume that $\psi \in \mathcal{C}^2$, is strongly convex and has bounded Hessian, i.e. for some $K, L > 0$

$$KI_d \preceq \nabla^2 \psi(q) \preceq LI_d. \tag{21}$$

When this holds we have that Assumption 4.20 is satisfied by [12, Theorem 3.9] with $\overline{G} = H$, where H is the Hamiltonian function

$$H(q, p) = \psi(q) + \frac{1}{2}\|p\|^2.$$

Since we consider the approximation based on Algorithm 3 we do not need Assumption 4.21. As λ is constant in this case Assumption 4.22 is satisfied provided

$$\sup_{n \in \mathbb{N}} \mathbb{E}_{(q, p)} \left[\psi(\overline{Q}_{t_n}) + \frac{1}{2}\|\overline{P}_{t_n}\|^2 \right] < \infty. \tag{22}$$

Because ψ has bounded second order derivative this reduces to showing that the second moment of the approximation is bounded uniformly in time. This condition depends on the choice of the numerical integrator and should be checked depending on the specific choice.

As mentioned in Note 4.26 in order to apply Theorem 4.24 with a numerical error we need to verify (15) holds. It is sufficient to show that the derivative of the semigroup decays exponentially. In order to prove this we shall rely on two Lipschitz conditions for the Hamiltonian flow φ_t: there exist $\nu, K_1, C \geq 1$, $\gamma \in (0, 1)$ such that for any $q, \bar{q}, p, \bar{p} \in \mathbb{R}^d$

$$\sup_{t > 0} \|\varphi_t(q, p) - \varphi_t(\bar{q}, \bar{p})\| \leq C\|(q, p) - (\bar{q}, \bar{p})\| \tag{23}$$

$$\|\varphi_t(q, p) - \varphi_t(\bar{q}, p)\| \leq \gamma\|q - \bar{q}\| \quad \text{for } \nu < t \leq K_1. \tag{24}$$

It is shown in [11] that under (21) the contraction (24) holds for some ν, γ, K_1. Indeed the authors prove a stronger result under which $\nu = 0$, but γ depends on t. There are also extensions to non-convex functions ψ, however here we only consider the convex setting. On the other hand, (23) is for instance satisfied for linear flows since the flow preserves the Hamiltonian and the Hamiltonian is equivalent to the norm. To simplify the exposition we will restrict to the case where (23) and (24) hold.

Proposition 5.9. *Let $\{\mathcal{P}_t\}_{t\geq 0}$ denote the semigroup of RHMC. Suppose that* (23) *and* (24) *hold. Moreover assume that*

$$0 < \kappa := C(1 - (e^{-\lambda v} - e^{-\lambda K_1})(1 - \gamma C^{-1})) < 1.$$

Then

$$\|\nabla_{q,p}\mathcal{P}_t f(q, p)\| \leq C^2 e^{-\kappa t}\|f\|_{\mathcal{C}_b^1}.$$

Proof. The proof is deferred to Appendix C.2. □

A case where it is easy to see that $\kappa < 1$ is the standard Gaussian case, i.e. $\psi(q) = \|q\|^2/2$, since in this case we have that $C = 1$.

Theorem 5.10. *Suppose that ψ satisfies* (21), (23), (24), *and the numerical integrator satisfies* (22). *Then the conclusions of Theorem* 4.24 *hold.*

Example 5.11 (*Zig Zag Sampler Continued*)**.** Recall the notation of Examples 3.2 and 5.1. Let us verify the assumptions of Theorem 4.24 for the ZZS.

We will make the following assumptions on ψ. Assume $\psi \in \mathcal{C}^2$ and

$$\lim_{\|x\|\to\infty} \frac{\max(1, \|\nabla_x^2\psi(x)\|)}{\|\nabla_x\psi(x)\|} = 0 \quad \text{and} \quad \lim_{\|x\|\to\infty} \frac{\|\nabla_x\psi(x)\|}{\psi(x)} = 0. \tag{25}$$

Verifying Assumption 4.20:
Geometric ergodicity of the ZZS was established in [10] under (25) with Lyapunov function

$$\overline{G}_{\alpha,\epsilon}(x, v) = \exp\left(\alpha\psi(x) + \sum_{i=1}^{d} \phi_\epsilon(v_i\partial_i\psi(x))\right). \tag{26}$$

Here $\phi_\epsilon(s) = \text{sign}(s)\log(1 + \epsilon|s|)/2$, $\alpha \in (0, 1)$, $\epsilon > 0$, $\alpha > \epsilon\overline{\gamma}$, where $\overline{\gamma}$ upper bounds the excess switching rate $\gamma : E \to \mathbb{R}_+$.

Verifying Assumption 4.21:
When dealing with this assumption it is convenient to use a smooth choice of λ_i, so for this section we will set $\phi(r) = r(1 + r)^{-1}$ and

$$\lambda_i(x, v) = -\log(\phi\exp(-v_i\partial_i\psi(x))) \tag{27}$$

which was shown in [3] to be a smooth choice of λ_i for which the ZZS has the correct invariant measure. Note that for this choice of λ_i the excess switching rate γ takes values between 0 and $\overline{\gamma} = \log(2)$.

Lemma 5.12. *Let $\{\mathcal{P}_t\}_{t\geq 0}$ denote the semigroup corresponding to the ZZS as described in Example* 3.2. *Assume $\psi \in \mathcal{C}^2$ and has bounded Hessian. For λ_i given by* (27) *there exists a constant C depending on the Hessian of ψ and on d such that for any $f \in \mathcal{C}^1$ we have*

$$[\Phi, Q](f \circ \varphi_{\delta-s})(x + sv, v) \leq C \sup_{i\in\{1,\dots,d\}} \{|f(x + sv + (\delta - s)F_iv, F_iv)|$$
$$+ |\partial_{x_i}f(x + sv + (\delta - s)F_iv, F_iv)|\}.$$

Proof of Lemma 5.12. The proof is deferred to Appendix C.3. □

We apply this Lemma with $f = \mathcal{P}_t g$ with $g \in \mathcal{G}_1$ where

$$\mathcal{G}_1 = \{g : E \to \mathbb{R} : x \mapsto g(x, v) \in \mathcal{C}^1, \mu(g) = 0, |g| \le \overline{G}_{\overline{\alpha}, \epsilon}, \|\nabla_x g\| \le \overline{G}_{\overline{\alpha}, \epsilon}\} \tag{28}$$

where $\overline{\gamma}\epsilon < \overline{\alpha} < \alpha < 1$. Such an $\overline{\alpha}$ can always be found by taking ϵ sufficiently small. It remains to show that $\nabla_x \mathcal{P}_t g$ converges to zero for $g \in \mathcal{G}_1$.

Theorem 5.13. *Let $\{\mathcal{P}_t\}_{t \ge 0}$ denote the semigroup of the ZZS with generator given by* (8) *and with λ_i such that $x \mapsto \lambda_i(x, v) \in \mathcal{C}^1$ for each v and has bounded derivative $\nabla_x \lambda_i(x, v)$. Fix $\epsilon \overline{\gamma} < \overline{\alpha} < \alpha$, and let $\mathcal{G}_1$ be given by* (28)*. Then there exists a constant C such that for any $g \in \mathcal{G}_1$*

$$\|\nabla_x \mathcal{P}_t g(x, v)\| \le C(1 + t)e^{-\omega t}\overline{G}_{\alpha, \epsilon}(x, v).$$

Proof of Theorem 5.13. The proof is deferred to Appendix C.3. □

Note that by adjusting C, κ and α we can show that

$$\sup_{\delta \in (0, \delta), s \in [0, \delta]} \|\nabla_x \mathcal{P}_t g \circ \varphi_{\delta - s}(x + sv, v)\| \le Ce^{-\omega t}\overline{G}_{\alpha, \epsilon}(x, v).$$

Therefore Assumption 4.21 holds by combining Lemma 5.12, Theorem 5.13 and Assumption 4.20.

Verifying Assumption 4.22:

Note that since λ grows at most linearly it is sufficient to show that there exists a function $H(x, v)$ for any mesh $0 = t_0 \le t_1 \le \cdots$ with $\delta_k = t_k - t_{k-1} < \delta_0$ for any k

$$\mathbb{E}_z\left[\|\overline{X}_{t_k}\|^2 \overline{G}_{\alpha, \epsilon}(\overline{X}_{t_k}, \overline{V}_{t_k})\right] \le H(x, v).$$

Note that $\overline{G}_{\alpha, \epsilon}$ is dominated by the $e^{\alpha \psi}$ term so we can bound $\|x\|^2 \overline{G}_{\alpha, \epsilon}(x, v)$ by $e^{\alpha_1 \psi}$ for any $\alpha_1 > \alpha$ so it remains to show there exists a function $H(x, v)$ for any mesh $0 = t_0 \le t_1 \le \cdots$ with $\delta_k = t_k - t_{k-1} < \delta_0$ for any k

$$\mathbb{E}_z\left[e^{\alpha_1 \psi(\overline{X}_{t_k})}\right] \le H(x, v). \tag{29}$$

Then in the 1-dimensional case we prove that the required bound holds.

Lemma 5.14. *Suppose* (25) *hold and $d = 1$. Then* (29) *holds for both Algorithms 2 and 3, hence Assumption 4.22 is satisfied.*

This follows from Lemmas C.2 and C.3 which can be found in Appendix C.3. The generalisation to the d-dimensional setting is challenging and is thus left as a conjecture, supported by the experiments in Fig. 2(a).

Conjecture 5.15. *Suppose ψ satisfies* (25)*. Then inequality* (29) *holds for Algorithms 2 and 3.*

Theorem 5.16. *Let $\{(X_t, V_t)\}_{t \ge 0}$ be the ZZS. Let $\{(\overline{X}_t, \overline{V}_t)\}_{t \ge 0}$ be the process described in Example 3.2. Assume that ψ satisfies* (25) *and that Conjecture 5.15 holds. Let $\mathcal{G}_1$ be given by* (28)*. Then the conclusions of Theorem 4.24 hold.*

In Fig. 3 we show some numerical results in the case of a Gaussian target. We observe that the error in the estimation of the first component of the mean of the approximations is similar to that of the continuous ZZS, while the error for the radius statistic, i.e. $t(x) = \sum_{i=1}^d x_i^2$, obtained with the approximations decreases to that of the ZZS as δ becomes smaller.
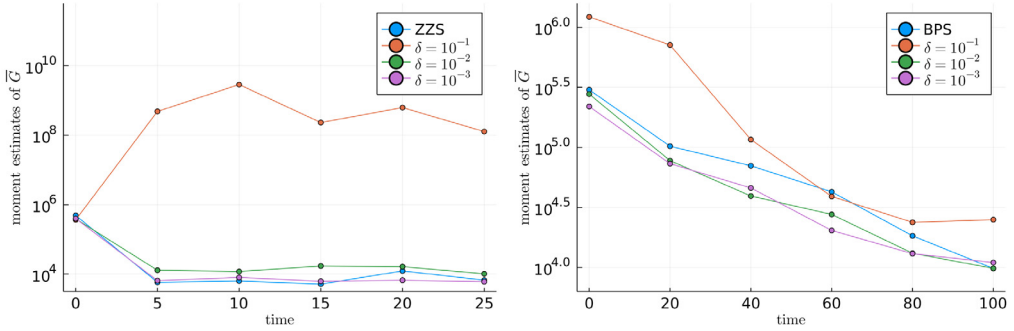
(a) Results for the ZZS and its approximations given by Algorithm 2. Here $\gamma(x, v) = 0$.

(b) Results for the BPS and its approximations given by Algorithm 3. Here $\lambda_r = 1$.

Fig. 2. Plots of the estimates of $\mathbb{E}[\overline{G}(X_t, V_t)]$ and $\mathbb{E}[\overline{G}(\overline{X}_t, \overline{V}_t)]$, which are respectively for the continuous time PDMPs and their approximations for several values of the step size. The plots show the average of 10^5 experiments. The continuous PDMPs have a 25-dimensional standard Gaussian as stationary measure. Here we choose $\overline{\lambda}((x, v), s; \delta) = \lambda(x, v)$. For each experiment $X_0, \overline{X}_0$ are given by an independent realisation of the sum of a 25-dimensional standard Gaussian and a uniform random variable on $[0, 1]^{25}$, while $V_0, \overline{V}_0$ are drawn from the stationary distribution of the continuous time PDMP.

Example 5.17 (*BPS Continued*). Let us discuss the assumptions of Theorem 4.24 for the approximation of BPS as given in Example 3.3, 5.2, and 5.5. Since this approximation is based on Algorithm 3 it is sufficient to check Assumptions 4.20 and 4.22. Conditions under which Assumption 4.20 holds are given in [23,25]. To be concrete we concentrate on [23], in which the Lyapunov function is given by

$$\overline{G}(x, v) = \frac{e^{\frac{1}{2}\psi(x)}}{\sqrt{\lambda(x, -v)}}.$$

Here at refreshment times a new velocity vector is drawn from the uniform distribution on the unit sphere. In Fig. 2(b) we estimate the moments of $\overline{G}$ for the continuous time BPS with a 25-dimensional standard Gaussian target and compare it to the approximations obtained by applying Algorithm 3 for several values of δ. We observe that the moments of the approximations resemble the continuous BPS and $\mathbb{E}[\overline{G}(\overline{X}_t, \overline{V}_t)]$ appears to be bounded uniformly in time. Therefore we conjecture that Theorem 4.24 holds under the assumptions of [23] for approximations of the BPS according to Algorithm 3.

In Fig. 3 we compare the errors of the BPS and its approximations given by Algorithm 3 in the case of a Gaussian target. We observe that the approximations perform similarly to the BPS. Note that for this target measure the BPS and the approximation are both rotationally invariant so they both have mean zero and hence in Fig. 3(c) we do not see the effect of the bias of the approximation.

Example 5.18 (*Continuous Time Approximation of a PDMP*). We continue our analysis of the setting introduced in Example 5.6. We wish to extend the conclusions of Theorem 4.24 to the continuous PDMP with generator $\tilde{\mathcal{L}}$ given by (20).
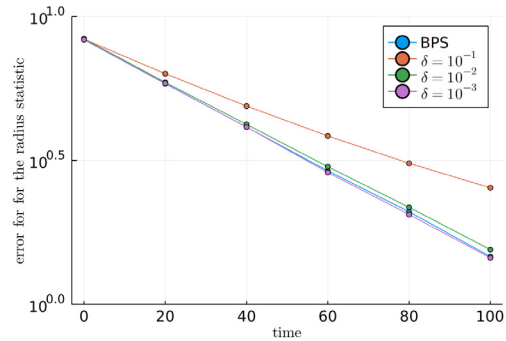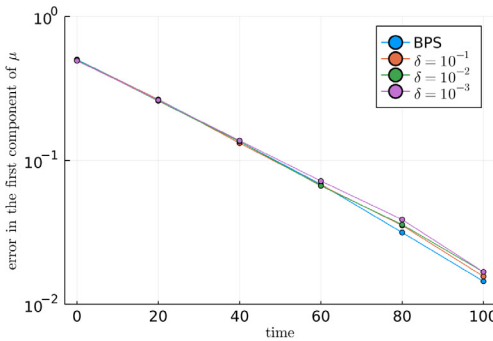
Theorem 5.19. *Suppose both the PDMPs with generators $\mathcal{L}$ and $\tilde{\mathcal{L}}$ satisfy Assumption 4.20 with Lyapunov functions $\overline{G}$ and $\tilde{G}$ respectively, and with invariant measures μ and $\tilde{\mu}$. Assume*
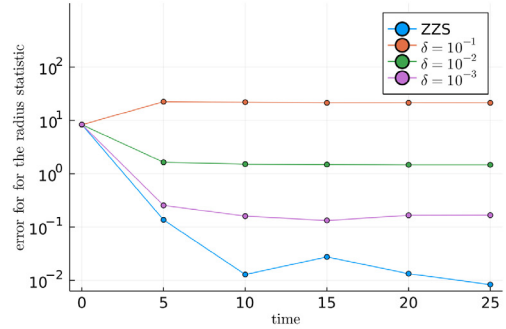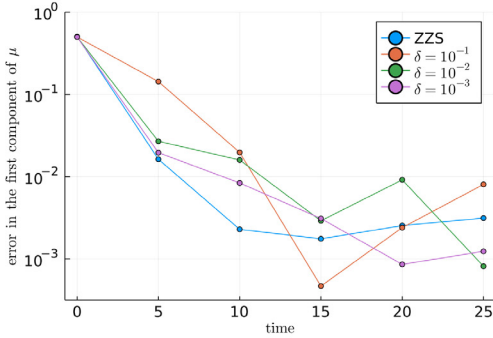
(a) Error for the mean for the ZZS and its approx- imations given by Algorithm 2.

(b) Error for the radius for the ZZS and its approx- imations given by Algorithm 2.

(c) Error for the mean for the BPS and its approx- imations given by Algorithm 3.

(d) Error for the radius for the BPS and its approx- imations given by Algorithm 3.

Fig. 3. Errors in the estimation of the first component of the mean and radius statistic in the context of Fig. 2. For the ZZS we take $\gamma(x, v) = 0$, while for the BPS we have $\lambda_r = 1$.

(19) *holds for some $\varepsilon > 0$. Moreover, suppose the approximation of the PDMP with generator $\mathcal{L}$ described in Example 5.6 satisfies Assumption 4.22 both for $\overline{G}$ and $\tilde{G}$. Set $\mathcal{G}_1 = \{g \in \mathcal{C}(E) : |g(x, v)| \leq \min\{\overline{G}(x, v), \tilde{G}(x, v)\}\}$. Then for all $g \in \mathcal{G}_1$*

$$\left| \mathbb{E}_z[g(Z_t)] - \mathbb{E}_z[g(\tilde{Z}_t)] \right| \leq C\varepsilon H(z).$$

Moreover, letting $t \to \infty$ we have

$$|\mu(g) - \tilde{\mu}(g)| \leq D\varepsilon.$$

Hence, in the case of ZZS we recover the result obtained in Theorem 6.2 of [33].

6. Proof of Theorem 4.9

We shall first prove the case of $p = 1$ in Section 6.1, and then in Section 6.2 we will use the $p = 1$ setting as a base case in a proof by induction to obtain the result for $p > 1$.

6.1. The case of $p = 1$

To prove Theorem 4.9 in this setting we define a coupling of Z_{t_n} and $\overline{Z}_{t_n}$ that satisfies the bounds in the statement. Then because the Wasserstein distance is defined as an infimum over all couplings we immediately obtain

$$\mathcal{W}_1(\mathcal{P}_T(z, \cdot), \overline{\mathcal{P}}_T(z, \cdot)) \leq \mathbb{E}_z[\|Z_T - \overline{Z}_T\|],$$

where the expectation on the right hand side is with respect to the specific coupling we consider.

Let us now introduce a general framework that contains both Algorithm 2 and Algorithm 3. Denote the approximation process as $\overline{Z}_{t_n}$ with initial state $\overline{Z}_0 = z$. Then given the previous state $\overline{Z}_{t_n}$ define

$$
\begin{aligned}
&\tilde{\tau}_{n+1}^i := \inf \left\{ t \geq 0 : 1 - \exp \left(-\int_0^t \overline{\lambda}_i(\overline{Z}_{t_n}, s; \delta_{n+1}) ds \right) \geq \tilde{U}_{n+1}^i \right\}, \\
&\tilde{\tau}_{n+1} := \min_{i=1,\ldots,m} \tilde{\tau}_{n+1}^i, \qquad I_{n+1} = \arg\min_{i=1,\ldots,m} \tilde{\tau}_{n+1}^i
\end{aligned}
\tag{30}
$$

where $\tilde{U}_{n+1}^1, \ldots, \tilde{U}_{n+1}^m \overset{iid}{\sim} \text{Unif}[0, 1]$. Then the switching time of the process is

$$\overline{\tau}_{n+1} = \overline{\tau}_{n+1}(\tilde{\tau}_{n+1}, \delta_{n+1}),$$

with the requirement that $\overline{\tau}_{n+1} \leq \delta_{n+1}$ if and only if $\tilde{\tau}_{n+1} \leq \delta_{n+1}$. In particular Algorithm 2 corresponds to the choice

$$\overline{\tau}_{n+1}(\tilde{\tau}_{n+1}, \delta_{n+1}) = \delta_{n+1} \mathbb{1}_{\{\tilde{\tau}_{n+1} \leq \delta_{n+1}\}} + \infty \mathbb{1}_{\{\tilde{\tau}_{n+1} > \delta_{n+1}\}},$$

while Algorithm 3 corresponds to

$$\overline{\tau}_{n+1}(\tilde{\tau}_{n+1}, \delta_{n+1}) = \tilde{\tau}_{n+1}.$$

The process can now be defined as follows:

$$
\overline{Z}_{t_{n+1}} = \begin{cases}
\overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n}; \delta_{n+1}) & \text{if } \overline{\tau}_{n+1} > \delta_{n+1}, \\
\overline{\varphi}_{\delta_{n+1} - \overline{\tau}_{n+1}}(\overline{F}_{I_{n+1}}(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}; \delta_{n+1}), U_{n+1}; \delta_{n+1}); \delta_{n+1}) & \text{if } \overline{\tau}_{n+1} \leq \delta_{n+1},
\end{cases}
$$

where $U_{n+1} \sim \nu_{\mathcal{U}}$ takes values in $\mathcal{U}$.

Let us now define the coupling of Z_{t_n} and $\overline{Z}_{t_n}$ that we will use to prove Theorem 4.9.

Coupling 6.1. Fix both processes up to time t_n and let $(Z_{t_{n+1}}, \overline{Z}_{t_{n+1}})$ evolve as follows. Let $U_{n+1} \sim \nu_{\mathcal{U}}$ and $\tilde{U}_{n+1}^1, \ldots, \tilde{U}_{n+1}^m \overset{iid}{\sim} \text{Unif}([0, 1])$ be independent of each other and of Z_{t_n} and $\overline{Z}_{t_n}$. The coupling evolves as follows:

- Define the next switching time of the continuous process as

$$
\begin{aligned}
&\tau_{n+1}^i := \inf \left\{ t \geq 0 : 1 - \exp \left(-\int_0^t \lambda_i(\varphi_s(Z_{t_n})) ds \right) \geq \tilde{U}_{n+1}^i \right\}, \\
&\tau_{n+1} := \min_{i=1,\ldots,m} \tau_{n+1}^i.
\end{aligned}
\tag{31}
$$

Then there are two cases:

  – if $\tau_{n+1} \leq \delta_{n+1}$, then set $Z_{t_n+s} = \varphi_s(Z_{t_n})$ for $s \in (0, \tau_{n+1})$ and

$$Z_{t_n + \tau_{n+1}} = F_{I_{n+1}}(\varphi_{\tau_{n+1}}(Z_{t_n}), U_{n+1})$$

where $I_{n+1} = \arg\min_i \tau^i_{n+1}$. Then simulate the process independently of the rest for the remaining time $\delta_{n+1} - \tau_{n+1}$.

- if $\tau_{n+1} > \delta_{n+1}$, then set $Z_{t_n+s} = \varphi_s(Z_{t_n})$ for $s \in (0, \delta_{n+1}]$.

- Define $\tilde{\tau}_{n+1}$ as in Eq. (30), where $\tilde{U}^1_{n+1}, \ldots, \tilde{U}^m_{n+1}$ is the same random variable used in (31). Compute $\overline{\tau}_{n+1} = \overline{\tau}_{n+1}(\tilde{\tau}_{n+1}, \delta_{n+1})$. Then the approximation process evolves as:

  - if $\overline{\tau}_{n+1} \leq \delta_{n+1}$, then set

  $$\overline{Z}_{t_{n+1}} = \overline{\varphi}_{\delta_{n+1}-\overline{\tau}_{n+1}}(\overline{F}_{\bar{I}_{n+1}}(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}; \delta_{n+1}), U_{n+1}; \delta_{n+1}); \delta_{n+1})$$

  where $\bar{I}_{n+1} = \arg\min_i \tilde{\tau}^i_{n+1}$.
  - if $\overline{\tau}_{n+1} > \delta_{n+1}$, then set $\overline{Z}_{t_{n+1}} = \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n}; \delta_{n+1})$.

Therefore the first switching times of the two processes are coupled, and so is the eventual random jump. Once $Z_{t_{n+1}}$ and $\overline{Z}_{t_{n+1}}$ have been obtained, repeat the same procedure to obtain $Z_{t_{n+2}}$ and $\overline{Z}_{t_{n+2}}$.

We remark that the marginal distributions of each process are the correct one, and thus this is indeed a valid coupling of the two processes.

In the proof that follows we simplify the notation denoting the approximations as $\overline{\varphi}_z(z)$, $\overline{\lambda}_i(z, s)$, and $\overline{F}_i(z, U)$, instead of $\overline{\varphi}_z(z; \delta_{n+1})$, $\overline{\lambda}_i(z, s; \delta_{n+1})$, and $\overline{F}_i(z, U; \delta_{n+1})$.

Proof of Theorem 4.9. We begin by partitioning the space as

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|] = \mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|(\mathbb{1}_{E_{00}} + \mathbb{1}_{E_{11}} + \mathbb{1}_{E_{10}} + \mathbb{1}_{E_{01}})],$$

where E_{ij} for $i, j = 0, 1$ denotes the event in which there are i random events for the approximation process, while $j = 0$ denotes that no events take place for the continuous process, and $j = 1$ that at least one event for the original process happens in the time interval $s \in [t_n, t_{n+1})$. The four events are considered respectively in Lemmas A.4, A.5, A.6, and A.7. Since the upper bounds in these results are non-decreasing functions of the time t_n, we combine the results of the Lemmas to obtain that there exist constants $K_1 = K_1(t_{n+1})$ and $K_2 = K_2(t_{n+1})$ such that

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|] \leq (1 + \delta_{n+1}K_1)\mathbb{E}_z[\|Z_{t_n} - \overline{Z}_{t_n}\|] + \delta^2_{n+1}K_2.$$

Since the two processes start at the same point this implies, by recursion,

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|] \leq \sum_{k=1}^{n+1} K_2\delta_k^2 \left(\prod_{\ell=k}^{n+1}(1 + \delta_\ell K_1)\right). \tag{32}$$

In the setting when $\delta_n = \delta$ the right hand side of (32) becomes a geometric series which leads to the estimate

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|] \leq \delta^2 \frac{(1 + \delta K_1)^{n+1} - 1}{(1 + \delta K_1) - 1} K_2$$
$$\leq \delta \left(e^{K_1(n+1)\delta} - 1\right) \frac{K_2}{K_1}. \quad \square \tag{33}$$

6.2. The case of $p > 1$

In order to simply the notation we shall restrict to the case $\delta_n = \delta$. To prove the result we reason by induction on p. In particular, we consider the following inductive hypothesis. Fix $p \geq 1$ and $n \geq 1$.

Inductive Hypothesis 6.2. Suppose the PDMP satisfies Assumptions 4.1–4.3. Moreover suppose the approximation given by Algorithm 5 satisfies Assumptions 4.4–4.7 hold for some $\delta_0 > 0$. Given $(Z_{t_n}, \overline{Z}_{t_n})$ there exists a coupling $(Z_{t_{n+1}}, \overline{Z}_{t_{n+1}})$ with respective marginals corresponding to $\mathcal{P}_\delta(Z_{t_n}, \cdot), \overline{\mathcal{P}}_\delta(\overline{Z}_{t_n}, \cdot; \delta, p)$ there exist $A = A(T)$, $B = B(T)$ independent of n such that for any $0 < \delta \leq \delta_0$

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|] \leq A\delta^{p+1} + (1 + B\delta)\mathbb{E}_z[\|Z_{t_n} - \overline{Z}_{t_n}\|]. \tag{34}$$

It is sufficient to show that the Inductive hypothesis holds and then the statement of the Theorem follows by recursion in n as done in (33). Observe that the case $p = 1$, which corresponds to Algorithm 3, holds by the proof of Section 6.1. Suppose the Inductive Hypothesis holds for some $p \geq 1$, let us consider the case of $p+1$. Let us define the following coupling of $(Z_{t_{n+1}}, \overline{Z}_{t_{n+1}})$ given $(Z_{t_n}, \overline{Z}_{t_n})$.

Coupling 6.3. Define for $0 \leq t \leq \delta$

$$\lambda_{tot}^i(z, \overline{z}, t; \delta, p + 1) = \overline{\lambda}_i(\overline{z}, t; \delta, p + 1) + \lambda_i(\varphi_t(z)) + 1.$$

Then for $i = 1, \ldots, m$ draw the proposed event times T_i with distribution given by

$$\mathbb{P}(T_i > t) = \exp\left(-\int_0^t \lambda_{tot}^i(Z_{t_n}, \overline{Z}_{t_n}, r; \delta, p + 1)\mathrm{d}r\right).$$

Let $T_{i*} = \min_{i=1,\ldots,m} T_i$ and let i^* be the argument that minimises T_i. If $T_{i*} \geq \delta$, then let $Z_{t_{n+1}} = \varphi_\delta(Z_{t_n})$, $\overline{Z}_{t_{n+1}} = \overline{\varphi}_\delta(\overline{Z}_{t_n}; \delta, p + 1)$.

Consider now the case in which $T_{i*} < \delta$. Let $U \sim \nu_\mathcal{U}$ and $\overline{U} \sim \text{Unif}([0, 1])$ independent of the T_i's and independent of each other. Then set

$$\tau_* = T_{i*} \quad \text{if } \overline{U} \leq \frac{\lambda_{i*}(\varphi_{T_{i*}}(Z_{t_n}))}{\lambda_{tot}^{i*}(Z_{t_n}, \overline{Z}_{t_n}, T_{i*}; \delta, p + 1)},$$

i.e. the proposed event time is accepted for the continuous time process. Alternatively set $\tau_*(z) = R$ for some constant $R > \delta$. Similarly let

$$\overline{\tau}_* = T_{i*} \quad \text{if } \overline{U} \leq \frac{\overline{\lambda}_{i*}(\overline{Z}_{t_n}, T_{i*}; \delta, p + 1)}{\lambda_{tot}^{i*}(Z_{t_n}, \overline{Z}_{t_n}, T_{i*}; \delta, p + 1)},$$

and thus conditional on acceptance $\overline{\tau}_*$ is the next event time for the approximation process. In case of rejection set $\overline{\tau}_* = R$ for some constant $R > \delta$ as done above. Set $Z_{t_n+s} = \varphi_s(z)$ and $\overline{Z}_{t_n+s} = \overline{\varphi}_s(\overline{Z}_{t_n}; \delta, p + 1)$ for $s \in [0, T_{i*})$. We distinguish three scenarios:

(1) The proposed switching time T_{i*} is accepted by both processes. Then set

$$Z_{t_n+T_{i*}} = F_{i*}(\varphi_{T_{i*}}(Z_{t_n}), U),$$
$$\overline{Z}_{t_n+T_{i*}} = \overline{F}_{i*}(\overline{\varphi}_{T_{i*}}(\overline{Z}_{t_n}; \delta, p + 1), U; \delta, p + 1).$$

To get from time $t_n + T_{i*}$ to t_{n+1} we apply the coupling given by the Inductive Hypothesis 6.2.

(2) The proposed switching time T_{i*} is accepted for one process, but rejected for the other. To get from time t_n+T_{i*} to t_{n+1} we let the two processes evolve independently according to their marginal distributions.

(3) The proposed switching time T_{i*} is rejected by both processes. Then set $Z_{T_{i*}} = \varphi_{T_{i*}}(Z_{t_n})$ and $\overline{Z}_{T_{i*}} = \overline{\varphi}_{T_{i*}}(\overline{Z}_{t_n}; \delta, p+1)$. To get from time t_n+T_{i*} to t_{n+1} we repeat this procedure starting at time $t_n + T_{i*}$ and with δ replaced with $\delta - T_{i*}$.

Proof of Theorem 4.9. Assume $Z_0 = \overline{Z}_0 = z$. Suppose that (34) holds for some $p \geq 1$. We will show that (34) then follows for p replaced $p + 1$ by using Coupling 6.3.

Suppose first that $T_{i*} > \delta$. Then the two processes follow the deterministic flow and by Assumption 4.4 with order $p + 1$ and Lemma A.2 we have

$$
\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{T_{i*} > \delta}] = \mathbb{E}_z[\|\varphi_\delta(Z_{t_n}) - \overline{\varphi}_\delta(\overline{Z}_{t_n}; \delta, p + 1)\| \mathbb{1}_{T_{i*} > \delta}]
$$
$$
\leq (1 + CC'\delta)\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|] + \tilde{C}^{p+2}.
$$

Let us consider the case (1) in Coupling 6.3 and denote the corresponding event as E_1. Then using the Inductive Hypothesis 6.2

$$
\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_1}] = \mathbb{E}_z\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_1} | T_{i*}]
$$
$$
\leq \mathbb{E}_z[(A\delta^{p+1} + (1 + B\delta)\|Z_{t_n + T_{i*}} - \overline{Z}_{t_n + T_{i*}}\|)\mathbb{1}_{E_1}]
$$
$$
= \mathbb{E}_z[(A\delta^{p+1}
$$
$$
+ (1 + B\delta)\|F_{i*}(\varphi_{T_{i*}}(Z_{t_n}), U) - \overline{F}_{i*}(\overline{\varphi}_{T_{i*}}(\overline{Z}_{t_n}; \delta, p + 1), U; \delta, p + 1)\|)\mathbb{1}_{E_1}]
$$
$$
\leq \mathbb{E}_z[(A\delta^{p+1} + (1 + B\delta)(M_1\delta^{p+1} + D_2\|\varphi_{T_{i*}}(Z_{t_n}) - \overline{\varphi}_{T_{i*}}(\overline{Z}_{t_n}; \delta, p + 1)\|))\mathbb{1}_{E_1}]
$$
$$
\leq \mathbb{E}_z[(A\delta^{p+1} + (1 + B\delta)(M_1\delta^{p+1} + D_2(1 + CC'\delta)\|Z_{t_n} - \overline{Z}_{t_n}\| + D_2\tilde{C}\delta^{p+2}))\mathbb{1}_{E_1}].
$$

Here we used Assumption 4.2(b), and Lemma A.2. Then we take advantage of

$$
\mathbb{P}_z(T_{i*} < \delta) \leq 1 - \exp(-\delta(2L(t_{n+1}, z, p + 1) + m)) \leq \delta(2L(t_{n+1}, z, p + 1) + m)
$$

to get

$$
\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_1}] \leq \tilde{A}_1\delta^{p+2} + (1 + \tilde{B}_1\delta)\mathbb{E}_z[\|Z_{t_n} - \overline{Z}_{t_n}\|]
$$

for suitable constants $\tilde{A}_1, \tilde{B}_1$ and taking advantage of $\delta < \delta_0$.

Now consider the case (2) in Coupling 6.3 and denote the corresponding event as E_2. Note that

$$
\mathbb{P}_z(E_2 | Z_{t_n}, \overline{Z}_{t_n}) = \delta(2L(t_{n+1}, z, p + 1) + m)\frac{|\lambda_{i*}(\varphi_{T_{i*}}(Z_{t_n})) - \overline{\lambda}_{i*}(\overline{Z}_{t_n}, T_{i*}; \delta, p + 1)|}{\lambda_{tot}^{i*}(Z_{t_n}, \overline{Z}_{t_n}, T_{i*}; \delta, p + 1)}.
$$

Using Assumptions 4.1, 4.3, 4.6, and the triangle inequality we obtain

$$
\mathbb{P}_z(E_2 | Z_{t_n}, \overline{Z}_{t_n}) \leq \delta(2L(t_{n+1}, z, p + 1) + m)(D_4C'\|Z_{t_n} - \overline{Z}_{t_n}\| + \delta^{p+1}\overline{M}_2(\overline{Z}_{t_n})).
$$

Therefore using Assumption 4.7

$$
\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_2}] \leq \tilde{A}_2\delta^{p+2} + (1 + \tilde{B}_2\delta)\mathbb{E}_z[\|Z_{t_n} - \overline{Z}_{t_n}\|]
$$

for some constants $\tilde{A}_2, \tilde{B}_2$.

Let us consider the case (3) in Coupling 6.3 and denote the corresponding event as E_3. Note that since case (3) involves repeating the coupling we may have to repeat this step an arbitrary number of times. Let q denote the number of times we propose a candidate jumping time. If $q < p + 2$ then we must have reached case (1) or (2), so it is sufficient to use the respective estimate derived above to get the desired result. On the other hand, the probability that $q \geq p + 2$ is bounded by $(2L(t_{n+1}, z, p + 1) + m)^{p+2}\delta^{p+2}$, which gives us the correct order. $\square$

7. Proof of Theorem 4.17

7.1. The case of $p = 1$

To prove the result we define a coupling of the continuous process with the approximation process. The intuitive idea is that, assuming the two processes are equal at the beginning of the current time step, we can use Poisson thinning [24,36] to simulate a proposal for the next event time that is common to both processes. This is achieved by simulating a Poisson process with rate given by the sum of the rates of the two processes. The proposal is then accepted or rejected individually for each process based on the correct switching rates. For this acceptance–rejection step a common uniform random variable is used. If the proposal is accepted for both processes, then a coupled event takes place, thus ensuring that the processes are equal after the event has happened. If the thinning step is successful it follows that the processes are equal for all $s \in (t_n, t_{n+1}]$ unless a second event takes place for the continuous time process in the current time interval, which is an event with $\mathcal{O}(\delta_{n+1}^2)$ probability. Let us now give the formal definition of the coupling.

Coupling 7.1. Let t_n be the current time and assume $Z_{t_n} = \overline{Z}_{t_n} = z_n$. Define $\lambda_{tot}^i(z, t; \delta_{n+1}) = \overline{\lambda}_i(z, t; \delta_{n+1}) + \lambda_i(\varphi_t(z)) + 1$. Then for $i = 1, \ldots, m$ draw the proposed event times $T_i(z_n)$ with distribution

$$\mathbb{P}(T_i(z_n) \leq t) = 1 - \exp\left(-\int_0^t \lambda_{tot}^i(z_n, r; \delta_{n+1}) dr\right).$$

Let $T_{i*}(z) = \min_{i=1,\ldots,m} T_i(z)$. Now let $U_{n+1} \sim \nu_{\mathcal{U}}$ and $\overline{U} \sim \text{Unif}([0, 1])$ independent of the T_i's and of Z_{t_n}. Then set

$$\tau(z_n) = T_{i*}(z_n) \quad \text{if } \overline{U} \leq \frac{\lambda_{i*}(\varphi_{T_{i*}(z_n)}(z_n))}{\lambda_{tot}^{i*}(z_n, T_{i*}(z_n); \delta_{n+1})}, \tag{35}$$

hence upon acceptance the proposed event time is the next switching time for the continuous time process. Alternatively set $\tau(z_n) = R > \delta_{n+1}$ for some constant $R \neq T_{i*}(z_n)$. Similarly let

$$\overline{\tau}(z_n) = T_{i*}(z_n) \quad \text{if } \overline{U} \leq \frac{\overline{\lambda}_{i*}(z_n, T_{i*}(z_n); \delta_{n+1})}{\lambda_{tot}^{i*}(z_n, T_{i*}(z_n); \delta_{n+1})}, \tag{36}$$

and thus conditional on acceptance $\overline{\tau}(z_n)$ is the next event time for the approximation process. In case of rejection set $\overline{\tau}(z_n) = R > \delta_{n+1}$ for some constant $R \neq T_{i*}(z_n)$ as done above.

If $T_{i*} \geq \delta_{n+1}$, then let $Z_{t_n+s} = \overline{Z}_{t_n+s} = \varphi_s(z_n)$ for $s \in (0, \delta_{n+1}]$. In this case the two processes are equal at time t_{n+1}.

Alternatively, we have $T_{i*} < \delta_{n+1}$ and thus we set $Z_{t_n+s} = \overline{Z}_{t_n+s} = \varphi_s(z_n)$ for $s \in (0, T_{i*}(z_n))$. Then the continuous process evolves as follows:

- if $\tau(z_n) = T_{i*}(z_n)$, then set

$$Z_{t_n+\tau(z_n)} = F_{i*}(\varphi_{\tau(z_n)}(z_n), U_{n+1}).$$

  Then let the process evolve independently of the approximation until time t_{n+1}.
- if $\tau(z_n) \neq T_{i*}(z_n)$, the proposed event time is rejected and we let the process evolves independently of the approximation until time t_{n+1}.

On the other hand, the approximation process evolves as follows:

- if $\overline{\tau}(z_n) = T_{i*}(z)$, set

$$Z_{t_n + \overline{\tau}(z_n)} = F_{i*}(\varphi_{\overline{\tau}(z_n)}(z_n), U_{n+1}),$$

  and finally $\overline{Z}_{t_n + s} = \varphi_s(Z_{t_n + \overline{\tau}(z_n)})$ for $s \in (\overline{\tau}(z_n), \delta_{n+1}]$.
- if $\overline{\tau}(z_n) \neq T_{i*}(z_n)$, then repeat this procedure from the beginning starting at time $t_n + T_{i*}(z_n)$ and with step $\delta_{n+1} - T_{i*}(z_n)$.

Lemma 7.2. *Under Assumption 4.14, there exists $D(t_n, z) > 0$ such that*

$$\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n} | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}) \leq D(t_n, z)\delta_n^2,$$

for $D(t_n, z) = (L_1(t_n, z)/2 + L_2(t_n, z) + L_3(t_n, z)/2)$.

Proof. The proof is postponed to Appendix B.1. □

Proof of Theorem 4.17. By the coupling inequality we have

$$\| P_{t_n}(z, \cdot) - \overline{P}_{t_n}(z, \cdot) \|_{TV} \leq \mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n})$$

and thus it is sufficient to bound the right hand side. Apply Lemma 7.2 to obtain

$$
\begin{aligned}
\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}) &= \mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n} | Z_{t_{n-1}} \neq \overline{Z}_{t_{n-1}})\mathbb{P}_z(Z_{t_{n-1}} \neq \overline{Z}_{t_{n-1}}) \\
&\quad + \mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n} | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}})(1 - \mathbb{P}_z(Z_{t_{n-1}} \neq \overline{Z}_{t_{n-1}})) \\
&\leq \mathbb{P}_z(Z_{t_{n-1}} \neq \overline{Z}_{t_{n-1}}) + D(t_n, z)\delta_n^2(1 - \mathbb{P}_z(Z_{t_{n-1}} \neq \overline{Z}_{t_{n-1}})) \\
&= (1 - D(t_n, z)\delta_n^2)\mathbb{P}_z(Z_{t_{n-1}} \neq \overline{Z}_{t_{n-1}}) + D(t_n, z)\delta_n^2.
\end{aligned}
\tag{37}
$$

Thus by recursion and since $\overline{Z}_0 = Z_0 = z$ it follows that

$$\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}) \leq \sum_{i=1}^{n} D(t_n, z)\delta_i^2 \prod_{\ell=i+1}^{n} (1 - D(t_n, z)\delta_\ell^2).$$

In particular if $\delta_n = \delta$ for all $n \in \mathbb{N}$ we have that

$$
\begin{aligned}
\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}) &\leq D(t_n, z)\delta^2 \sum_{\ell=0}^{n-1}(1 - D(t_n, z)\delta^2)^\ell \\
&\leq 1 - (1 - D(t_n, z)\delta^2)^n \\
&\leq 1 - e^{-D(t_n, z)t_n\delta}. \quad \square
\end{aligned}
$$

7.2. The case of $p > 1$

In order to simplify the notation we shall restrict to the case $\delta_n = \delta$. To prove the result we reason by induction on p similarly to Section 6.2. In particular, we consider the following inductive hypothesis. Fix $p \geq 1$ and $n \geq 1$.

Inductive Hypothesis 7.3. Suppose $\overline{\lambda}$ satisfies Assumption 4.14 for some $\delta_0 > 0$. Given $Z_{t_n} = \overline{Z}_{t_n}$ there exist a coupling $(Z_{t_{n+1}}, \overline{Z}_{t_{n+1}})$ with respective marginals corresponding to $P_\delta(Z_{t_n}, \cdot)$, $\overline{P}_\delta(\overline{Z}_{t_n}, \cdot; \delta, p)$, and constants $A = A(T)$, $B = B(T)$ independent of n such that for any $0 < \delta \leq \delta_0$

$$\mathbb{P}_z(Z_{t_{n+1}} \neq \overline{Z}_{t_{n+1}} | Z_{t_n} = \overline{Z}_{t_n}) \leq A\delta^{p+1}.$$

It is sufficient to show that the Inductive hypothesis holds and the statement of the Theorem follows by recursion in n as done in (37). Observe that the case $p = 1$ holds by the proof of Section 7.1. To obtain the result we use Coupling 6.3 but with $\overline{\varphi} = \varphi$, $\overline{F}_i = F_i$, and replacing Inductive Hypothesis 6.2 with Inductive Hypothesis 7.3. Because the strategy is similar to that in Section 6.2 we postpone the formal proof to Appendix B.2.

8. Proof of Theorem 4.24

Recall in Section 6.1 we introduced a general framework which includes both Algorithms 2 and 3. We now introduce some further notation. Let $p_{\overline{\tau}}^{z,\delta,i}$ be a probability measure on $[0, \infty]$ which denotes the law of $\overline{\tau}$ for Algorithm i with initial condition at z for a time step of length δ. Note that for Algorithm 2 we have $p_{\overline{\tau}}^{z,\delta,i}$ is a point measure with

$$
\begin{aligned}
p_{\overline{\tau}}^{z,\delta,2}(\{\delta\}) \quad &= 1 - e^{-\int_0^\delta \overline{\lambda}(z,s;\delta)ds}, \\
p_{\overline{\tau}}^{z,\delta,2}(\{+\infty\}) &= e^{-\int_0^\delta \overline{\lambda}(z,s;\delta)ds}.
\end{aligned}
\tag{38}
$$

On the other hand, in the case of Algorithm 3 $p_{\overline{\tau}}^{z,\delta,3}$ admits a density which is given by

$$
p_{\overline{\tau}}^{z,\delta,3}(ds) = \overline{\lambda}(z,s;\delta) \exp\left(-\int_0^s \overline{\lambda}(z,r;\delta)dr\right)ds.
\tag{39}
$$

Proof of Theorem 4.24. Fix $g \in \mathcal{G}_1$. Then by a telescoping sum we have

$$
\mathbb{E}_z[g(\overline{Z}_{t_n})] - \mathbb{E}_z[g(Z_{t_n})] = \sum_{k=0}^{n-1}(\mathbb{E}_z[\mathcal{P}_{t_n-t_{k+1}}g(\overline{Z}_{t_{k+1}})] - \mathbb{E}_z[\mathcal{P}_{t_n-t_k}g(\overline{Z}_{t_k})]).
$$

For each $k \in \{0, \dots, n-1\}$, set $f_k(y, s) = \mathcal{P}_{t_n-t_k-s}g(y)$ then we have

$$
\mathbb{E}_z[g(\overline{Z}_{t_n})] - \mathbb{E}_z[g(Z_{t_n})] = \sum_{k=0}^{n-1}\mathbb{E}_z[f_k(\overline{Z}_{t_{k+1}}, \delta_{k+1}) - f_k(\overline{Z}_{t_k}, 0)].
$$

By conditioning on $\overline{Z}_{t_k}$ it is sufficient to prove that

$$
|\mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - f_k(z, 0)| \le Re^{-\omega(t_n-t_{k+1})}\overline{G}_i(z)\delta_{k+1}^2.
\tag{40}
$$

Here with an abuse of notation we have denoted as $\overline{Z}_{\delta_{k+1}}$ the approximation process with initial condition at z and step size δ_{k+1}. Indeed if we have that (40) holds then by Assumption 4.22 we have

$$
\begin{aligned}
|\mathbb{E}_z[g(Z_{t_n})] - \mathbb{E}_z[g(\overline{Z}_{t_n})]| &\le R\sum_{k=0}^{n-1}e^{-\omega(t_n-t_{k+1})}\delta_{k+1}^2\mathbb{E}_z[\overline{G}_i(\overline{Z}_{t_k})] \\
&\le RCS_nH_i(z).
\end{aligned}
$$

Which gives the desired result. It remains to show that (40) holds.

Using that the approximation process jumps according to Q at a time determined by $p_{\overline{\tau}}^{z,\delta_{k+1},i}$ we can evaluate the expectations

$$
\begin{aligned}
\mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - [f_k(z, 0)] &= \\
&= \mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) + f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) - f_k(z, 0)
\end{aligned}
$$

$$= \int_0^{\delta_{k+1}} Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)) - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) p_\tau^{z, \delta_{k+1}, i}(ds)$$
$$+ f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) - f_k(z, 0).$$

Recall $p_\tau^{z, \delta, 2}$ ($p_\tau^{z, \delta, 3}$ respectively) is defined in (38) (resp. (39)). Using the fundamental Theorem of calculus we can rewrite this as

$$\mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - f_k(z, 0) =$$
$$= \int_0^{\delta_{k+1}} Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)) - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) p_\tau^{z, \delta_{k+1}, i}(ds)$$
$$+ \int_0^{\delta_{k+1}} \frac{d}{dr} f_k(\varphi_r(z), r) dr$$
$$= \int_0^{\delta_{k+1}} Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\phi_s(z)) - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) p_\tau^{z, \delta_{k+1}, i}(ds)$$
$$+ \int_0^{\delta_{k+1}} \langle \Phi(\varphi_r(z)), \nabla f_k(\varphi_r(z), r) \rangle + (\partial_s f_k)(\varphi_r(z), r) dr.$$

Note that $\partial_s f_k(y, s) = -\mathcal{L} f_k(y, s)$

$$\mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - f_k(z, 0) =$$
$$= \int_0^{\delta_{k+1}} Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)) - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) p_\tau^{z, \delta_{k+1}, i}(ds)$$
$$+ \int_0^{\delta_{k+1}} \langle \Phi(\varphi_r(z)), \nabla f_k(\varphi_r(z), r) \rangle - \mathcal{L} f_k(\varphi_r(z), r) dr.$$

Recall $\mathcal{L}$ is given by (5) so we can write the above as

$$\mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - f_k(z, 0) =$$
$$= \int_0^{\delta_{k+1}} Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)) - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) p_\tau^{z, \delta_{k+1}, i}(ds)$$
$$+ \int_0^{\delta_{k+1}} -\lambda(\varphi_r(z))[Q(f_k(\cdot, r))(\varphi_r(z)) - f_k(\varphi_r(z), r)] dr.$$

We rewrite this as

$$\mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - f_k(z, 0) =$$
$$= \int_0^{\delta_{k+1}} \left(Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)) - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) \right) (p_\tau^{z, \delta_{k+1}, i}(ds) - \lambda(\varphi_s(z)) ds)$$
$$- \int_0^{\delta_{k+1}} \lambda(\varphi_r(z))[Q(f_k(\cdot, r))(\varphi_r(z)) - Q(f_k(\varphi_{\delta_{k+1}-r}(\cdot), \delta_{k+1}))(\varphi_r(z))] dr \qquad (41)$$
$$- \int_0^{\delta_{k+1}} \lambda(\varphi_r(z))[f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) - f_k(\varphi_r(z), r)] dr.$$

We will divide the remainder of the proof into 3 steps:

Step (i): For this step we distinguish between Algorithm 2 and 3. Let

$$h_s = Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)) - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}).$$

Then we will show that there exists a constant $R > 0$ such that for any $h \in C_b^1([0, \delta])$ (for Algorithm 3 we only need $h \in C_b([0, \delta])$) we have

$$\left| \int_0^{\delta_{k+1}} h_s(p_\tau^{z, \delta_{k+1}, i}(\mathrm{d}s) - \lambda(\varphi_s(z))\mathrm{d}s) \right| \leq Re^{-\omega(t_n - t_{k+1})}\delta_{k+1}^2 \sup_{s, r \in [0, \delta_0]} K_i(z, s, r) \tag{42}$$

where K_i is as in Assumption 4.22.

Step (ii): For any $z \in E, r \in [0, \delta_{k+1}]$ we have

$$|f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) - f_k(\varphi_r(z), r)| \leq$$
$$\leq R_1 \int_r^{\delta_{k+1}} e^{-\omega(t_n - t_k - s)}\lambda(\varphi_s(z))(Q\overline{G}(\varphi_s(z)) + \overline{G}(\varphi_s(z)))\mathrm{d}s. \tag{43}$$

Step (iii): For any $z \in E, r \in [0, \delta_{k+1}]$ we have

$$|Q(f_k(\varphi_{\delta_{k+1}-r}(\cdot), \delta_{k+1}))(z) - Q(f_k(\cdot, r))(z)| \leq$$
$$\leq R_1 \int_r^{\delta_{k+1}} e^{-\omega(t_n - t_k - s)}Q((Q\overline{G} + \overline{G})\lambda)(\varphi_{s-r}(z))\mathrm{d}s. \tag{44}$$

Eq. (40) follows from Step (i), (ii), (iii) and (41), as this gives

$$\mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - f_k(z, 0) \leq Re^{-\omega(t_n - t_{k+1})}\delta_{k+1}^2 \sup_{s, r \in [0, \delta_0]} K_i(z, r, s)$$
$$+ \int_0^{\delta_{k+1}} \lambda(\varphi_r(z))R_1 \int_r^{\delta_{k+1}} e^{-\omega(t_n - t_k - s)}Q(\lambda[Q\overline{G} + \overline{G}])(\varphi_{s-r}(z))\mathrm{d}s\,\mathrm{d}r$$
$$+ \int_0^{\delta_{k+1}} \lambda(\varphi_r(z))R_1 \int_r^{\delta_{k+1}} e^{-\omega(t_n - t_k - s)}\lambda(\varphi_s(z))[Q\overline{G}(\varphi_s(z)) + \overline{G}(\varphi_s(z))]\mathrm{d}s\,\mathrm{d}r.$$

Recall that $\overline{G}_i(z, r, s)$ is given by (12), then we have

$$|\mathbb{E}_z[f_k(\overline{Z}_{\delta_{k+1}}, \delta_{k+1})] - f_k(z, 0)| \leq Re^{-\omega(t_n - t_{k+1})}\delta_{k+1}^2 \sup_{s, r \in [0, \delta_0]} \overline{G}_i(z, r, s).$$

Proof of Step (i): This step follows from Lemma C.1. It remains to find a bound for $|h_s|$ and $|\partial_s h_s|$ for the case of Algorithm 2. By Assumption 4.20

$$|h_s| = \left| Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)) - f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) \right|$$
$$\leq \left| Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)) - \mu(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1})) \right|$$
$$+ \left| \mu(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1})) - f_k \circ \varphi_{\delta_{k+1}-s}(\varphi_s(z), \delta_{k+1}) \right|$$
$$\leq R_1 e^{-\omega(t_n - t_{k+1})}[Q\overline{G}(\varphi_s(z)) + \overline{G}(\varphi_s(z))]. \tag{45}$$

For Algorithm 2 we also require to control $|\partial_s h_s|$ for which we require a bound on the derivative of f, for this case we use Assumption 4.21. Note that

$$\partial_s h_s = \left| \langle \Phi(\varphi_s(z)), \nabla_z(Q(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1})))(\varphi_s(z)) \rangle \right.$$
$$\left. - Q(\langle \Phi, \nabla_z(f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1})) \rangle)(\varphi_s(z)) \right|$$
$$= [\mathbf{\Phi}, Q](f_k(\varphi_{\delta_{k+1}-s}(\cdot), \delta_{k+1}))(\varphi_s(z)).$$

Recall here we have defined the commutator in Section 2 and we are denoting by $\mathbf{\Phi}$ the differential operator corresponding to Φ. This term is bounded by Assumption 4.21 and we

have

$$|\partial_s h_s| \le R_2 e^{-\omega(t_n - t_{k+1})} \overline{G}(z). \tag{46}$$

Combining Lemma C.1 with (45) and (46) we have that (42) holds.

Proof of Step (ii): Observe that since $\partial_s f_k(y, s) + \langle \Phi(y), \nabla f_k(y, s) \rangle = -\lambda(y)[Q f_k(y, s) - f_k(y, s)]$ we have

$$
\begin{aligned}
|f_k(\varphi_{\delta_{k+1}}(z), \delta_{k+1}) - f_k(\varphi_r(z), r)| &= \left| \int_r^{\delta_{k+1}} \frac{\mathrm{d}}{\mathrm{d}s} f_k(\varphi_s(z), s) \mathrm{d}s \right| \\
&= \left| \int_r^{\delta_{k+1}} \lambda(\varphi_s(z)) [Q f_k(\varphi_s(z), s) - f_k(\varphi_s(z), s)] \mathrm{d}s \right| \\
&\le \int_r^{\delta_{k+1}} \lambda(\varphi_s(z)) [|Q f_k(\varphi_s(z), s) - \mu(Q(f_k(\cdot, s)))| + |\mu(f_k(\cdot, s)) - f_k(\varphi_s(z), s)|] \mathrm{d}s.
\end{aligned}
$$

We can bound this using Assumption 4.20 we obtain (43).

Proof of Step (iii): Applying (43) with z replaced by $\varphi_{-r}(y)$ and applying Q we have (44). $\square$

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

Appendix A. Proofs of Section 4.1

A.1. Proof of Theorem 4.9

In this section we prove the lemmas that are used to prove Theorem 4.9 in the case $p = 1$. In the proofs that follow we simplify the notation denoting the approximations as $\overline{\varphi}_z(z)$, $\overline{\lambda}_i(z, s)$, and $\overline{F}_i(z, U)$, instead of $\overline{\varphi}_z(z; \delta_{n+1})$, $\overline{\lambda}_i(z, s; \delta_{n+1})$, and $\overline{F}_i(z, U; \delta_{n+1})$. Before proving bounds on the events E_{ij}, let us state three simple lemmas which will be used multiple times in the proof. The proofs are omitted as they are a straightforward consequence of the assumptions.

Lemma A.1. *Assumption* 4.1 *implies that for any $\delta_0 > 0$ there exists a constant $C' = C'(\delta_0) > 0$ such that for any $z, z' \in E$ and $t \in (0, \delta_0)$ it holds that*

$$\|\varphi_t(z) - \varphi_t(z')\| \le C' \|z - z'\|. \tag{47}$$

Moreover, for any $t \in (0, \delta_0)$ and any $z, z' \in E$ we have the alternative bound

$$\|\varphi_t(z) - \varphi_t(z')\| \le (1 + CC't) \|z - z'\|.$$

Lemma A.2. *Suppose Assumptions 4.1 and 4.4 hold. Then for any $p \geq 1$, $s \geq 0$ and $z, z' \in E$ it holds that*

$$\|\varphi_s(z) - \overline{\varphi}_s(z')\| \leq C'\|z - z'\| + \tilde{C}s^{p+1}.$$

Moreover, using Lemma A.1 we can replace C' with $1 + CC'\delta$.

Lemma A.3. *Under Assumptions 4.3, 4.6, and 4.7, for any $t \geq 0$, $p \geq 1$ there exists a positive constant $L(t, z, p)$ such that*

$$\sup_{r \in [0,t], s \in [0,\delta_0]} \max\{\lambda(Z_r), \overline{\lambda}(\overline{Z}_r, s; \delta_{n+1}, p)\} \leq L(t, z, p) \quad \text{a.s.}$$

where in particular $z = Z_0 = \overline{Z}_0$. Note if $p = 1$ we write $L(t, z, 1) = L(t, z)$.

We can now start showing a bound on event E_{00}, followed by the other events.

Lemma A.4. *Under Assumptions 4.1 and 4.4, it holds that*

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_{00}}] \leq (1 + \delta_{n+1}CC')\mathbb{E}_z[\|Z_{t_n} - \overline{Z}_{t_n}\|] + \tilde{C}\delta_{n+1}^2.$$

Proof. On E_{00} we are interested only in the error introduced by the integrator $\overline{\varphi}$. We have

$$\begin{aligned}
\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_{00}}] &= \mathbb{E}_z\left[\left\|\varphi_{\delta_{n+1}}(Z_{t_n}) - \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n})\right\| \mathbb{1}_{E_{00}}\right] \\
&\leq \mathbb{E}_z\left[\left\|\varphi_{\delta_{n+1}}(Z_{t_n}) - \varphi_{\delta_{n+1}}(\overline{Z}_{t_n})\right\|\right] \\
&\quad + \mathbb{E}_z\left[\left\|\varphi_{\delta_{n+1}}(\overline{Z}_{t_n}) - \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n})\right\|\right],
\end{aligned}$$

Then one can directly apply to the first term Assumption 4.1 and thus Lemma A.1, together with the assumption that $\delta_n \leq \delta_0$, and to the second term Assumption 4.4 to obtain the wanted result for $C' = C'(\delta_0)$. $\quad\square$

Lemma A.5. *Under Assumption 4.1, parts (b) and (c) of Assumption 4.2, as well as Assumptions 4.3–4.7, it holds that*

$$\begin{aligned}
\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_{11}}] &\leq \delta_{n+1}^2\left(mK_2 + m(m-1)\tilde{K}_2 + 2B(t_{n+1}, z)(L(t_{n+1}, z))^2\right) \\
&\quad + \delta_{n+1}(mK_1 + m(m-1)\tilde{K}_1)\mathbb{E}_z\left[\|Z_{t_n} - \overline{Z}_{t_n}\|\right],
\end{aligned}$$

where $L(t_{n+1}, z)$ was defined in Lemma A.3, while

$$\begin{aligned}
K_1 &= D_2(C')^2 L(t_{n+1}, z), \\
K_2 &= (D_2\tilde{C}C' + L(t_{n+1}, z)(2D_3 + M_1C')), \\
\tilde{K}_1 &= D_2(C')^2(L(t_{n+1}, z))^2, \\
\tilde{K}_2 &= (D_2\tilde{C}C' + (L(t_{n+1}, z))^2)(2D_3 + M_1C').
\end{aligned}$$

Proof. Let us first restrict to the event that Z_s has only one event for $s \in (t_n, t_{n+1}]$ and denote such event as $\overline{E}$. For any $i, j \in \{1, \ldots, m\}$, let A_{ij} be the event that Z_{t_n} jumps according to F_i and $\overline{Z}_{t_n}$ jumps according to $\overline{F}_j$ and no other jumps occur. Note that $\{A_{ij}\}_{i,j=1}^m$ is a partition of $E_{11} \cap \overline{E}$ so we may write

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_{11} \cap \overline{E}}] = \sum_{i,j=1}^m \mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{A_{ij}}].$$

Let us first consider event A_{ii}, i.e. the processes have a switch according to kernels F_i and $\overline{F}_i$. Considering Coupling 6.1, we first observe that A_{ii} is an order δ_{n+1} event. This follows from the fact that in this case we require

$$\tilde{U}_{n+1}^i \le \min\left\{ 1 - \exp\left(-\int_0^{\delta_{n+1}} \overline{\lambda}_i(\overline{Z}_{t_n}, s) \mathrm{d}s\right),\, 1 - \exp\left(-\int_0^{\delta_{n+1}} \lambda_i(\varphi_s(Z_{t_n})) \mathrm{d}s\right)\right\}.$$

Therefore, using that $1 - \exp(-z) \le z$ we obtain

$$\mathbb{E}_z\left[\mathbb{1}_{A_{ii}} | Z_{t_n}, \overline{Z}_{t_n}\right] \le \min\left\{\int_0^{\delta_{n+1}} \overline{\lambda}_i(\overline{Z}_{t_n}, s) \mathrm{d}s, \int_0^{\delta_{n+1}} \lambda_i(\varphi_s(Z_{t_n})) \mathrm{d}s\right\} \le \delta_{n+1} L(t_{n+1}, z),$$

(48)

where $L(t_{n+1}, z) < \infty$ was defined in Lemma A.3. We can then separate the effects of the different approximations by the triangle inequality:

$$\mathbb{E}_z\left[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{A_{ii}}\right] =$$
$$= \mathbb{E}_z[\|\varphi_{\delta_{n+1}-\tau_{n+1}}(F_i(\varphi_{\tau_{n+1}}(Z_{t_n}), U_{n+1})) - \overline{\varphi}_{\delta_{n+1}-\overline{\tau}_{n+1}}(\overline{F}_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1}))\| \mathbb{1}_{A_{ii}}]$$
$$\le \mathbb{E}_z\left[\|\varphi_{\delta_{n+1}-\tau_{n+1}}(F_i(\varphi_{\tau_{n+1}}(Z_{t_n}), U_{n+1})) - \varphi_{\delta_{n+1}-\overline{\tau}_{n+1}}(F_i(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1}))\| \mathbb{1}_{A_{ii}}\right] \qquad (*)$$
$$+ \mathbb{E}_z\left[\|\varphi_{\delta_{n+1}-\overline{\tau}_{n+1}}(F_i(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1})) - \varphi_{\delta_{n+1}-\overline{\tau}_{n+1}}(F_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1}))\| \mathbb{1}_{A_{ii}}\right] \qquad (**)$$
$$+ \mathbb{E}_z\left[\|\varphi_{\delta_{n+1}-\overline{\tau}_{n+1}}(F_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})) - \varphi_{\delta_{n+1}-\overline{\tau}_{n+1}}(\overline{F}_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1}))\| \mathbb{1}_{A_{ii}}\right]$$
$$(***)$$
$$+ \mathbb{E}_z\left[\|\varphi_{\delta_{n+1}-\overline{\tau}_{n+1}}(\overline{F}_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})) - \overline{\varphi}_{\delta_{n+1}-\overline{\tau}_{n+1}}(\overline{F}_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1}))\| \mathbb{1}_{A_{ii}}\right].$$
$$(****)$$

For term $(*)$ we first compare both terms to $F_i(\varphi_{\delta_{n+1}}(Z_{t_n}), U_{n+1})$, and then we condition on all random variables apart from U_{n+1} in order to apply Assumption 4.2(c):

$$(*) \le \mathbb{E}_z\big[\|\varphi_{\delta_{n+1}-\tau_{n+1}}(F_i(\varphi_{\tau_{n+1}}(Z_{t_n}), U_{n+1})) - F_i(\varphi_{\delta_{n+1}}(Z_{t_n}), U_{n+1})\| \mathbb{1}_{A_{ii}}\big]$$
$$+ \mathbb{E}_z\left[\|F_i(\varphi_{\delta_{n+1}}(Z_{t_n}), U_{n+1}) - \varphi_{\delta_{n+1}-\overline{\tau}_{n+1}} F_i(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1})\| \mathbb{1}_{A_{ii}}\right]$$
$$\le 2D_3 \delta_{n+1} \mathbb{P}_z(A_{ii})$$
$$\le \delta_{n+1}^2 2D_3 L(t_{n+1}, z).$$

In the last inequality we used the inequality derived in (48). Term $(**)$ can be bounded applying inequality (47), then conditioning on $Z_{t_n}, \overline{Z}_{t_n}, \overline{\tau}_{n+1}$ and using Assumption 4.2(b), and finally applying Lemma A.2 and (48):

$$(**) \le C' \mathbb{E}_z\left[\|F_i(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1}) - F_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})\| \mathbb{1}_{A_{ii}}\right]$$
$$\le C' D_2 \mathbb{E}_z\left[\|\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}) - \overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n})\| \mathbb{1}_{A_{ii}}\right]$$
$$\le (C')^2 D_2 \mathbb{E}_z[\|Z_{t_n} - \overline{Z}_{t_n}\| \mathbb{1}_{A_{ii}}] + \delta_{n+1}^2 C' D_2 \tilde{C}$$
$$\le (C')^2 D_2 L(t_{n+1}, z) \delta_{n+1} \mathbb{E}_z[\|Z_{t_n} - \overline{Z}_{t_n}\|] + \delta_{n+1}^2 C' D_2 \tilde{C}.$$

(49)

Term (***) is estimated by inequality (47), then again conditioning on $Z_{t_n}, \overline{Z}_{t_n}, \overline{\tau}_{n+1}$ and applying Assumption 4.5, and finally using (48):

$$
\begin{aligned}
(***) &\leq C' \mathbb{E}_z \left[\| F_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1}) - \overline{F}_i(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1}) \| \mathbb{1}_{A_{ii}} \right] \\
&\leq C' M_1 \delta_{n+1} \mathbb{P}_z(A_{ii}) \\
&\leq \delta_{n+1}^2 C' M_1 L(t_{n+1}, z).
\end{aligned} \tag{50}
$$

Term (****) is bounded using Assumption 4.4 and bounding by 1 the probability of A_{ii}:

$$
(****) \leq \tilde{C} \delta_{n+1}^2 \mathbb{P}_z(A_{ii}) \leq \tilde{C} \delta_{n+1}^2.
$$

Putting together terms (*), (**), (***), (****) we obtain the following bound on event A_{ii}:

$$
\mathbb{E}_z \left[\| Z_{t_{n+1}} - \overline{Z}_{t_{n+1}} \| \mathbb{1}_{A_{ii}} \right] \leq \delta_{n+1}^2 K_2 + \delta_{n+1} K_1 \mathbb{E}_z \left[\| Z_{t_n} - \overline{Z}_{t_n} \| \right] \tag{51}
$$

where K_1, K_2 are as in the statement of the lemma. Now consider event A_{ij} for $i \neq j$. In this case we take advantage of independence of $\tilde{U}_{n+1}^i$ and $\tilde{U}_{n+1}^j$ to conclude that

$$
\begin{aligned}
\mathbb{E}_z[\mathbb{1}_{A_{ij}} | Z_{t_n}, \overline{Z}_{t_n}] &\leq \\
&\leq \left(1 - \exp\left(- \int_0^{\delta_{n+1}} \lambda_i(\varphi_s(Z_{t_n})) \mathrm{d}s \right) \right) \left(1 - \exp\left(- \int_0^{\delta_{n+1}} \overline{\lambda}_j(\overline{Z}_{t_n}, s) \mathrm{d}s \right) \right) \\
&\leq \int_0^{\delta_{n+1}} \lambda_i(\varphi_s(Z_{t_n})) \mathrm{d}s \int_0^{\delta_{n+1}} \overline{\lambda}_j(\overline{Z}_{t_n}, s) \mathrm{d}s \\
&\leq \delta_{n+1}^2 (L(t_{n+1}, z))^2.
\end{aligned} \tag{52}
$$

Then we can use the decomposition

$$
\begin{aligned}
\mathbb{E}_z[\| Z_{t_{n+1}} - \overline{Z}_{t_{n+1}} \| \mathbb{1}_{A_{ij}}] &= \\
&= \mathbb{E}_z[\| \varphi_{\delta_{n+1} - \tau_{n+1}}(F_i(\varphi_{\tau_{n+1}}(Z_{t_n}), U_{n+1})) - \overline{\varphi}_{\delta_{n+1} - \overline{\tau}_{n+1}}(\overline{F}_j(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})) \| \mathbb{1}_{A_{ij}}] \\
&\leq \mathbb{E}_z[\| \varphi_{\delta_{n+1} - \tau_{n+1}}(F_i(\varphi_{\tau_{n+1}}(Z_{t_n}), U_{n+1})) - F_i(\varphi_{\delta_{n+1}}(Z_{t_n}), U_{n+1}) \| \mathbb{1}_{A_{ij}}] & (\dagger) \\
&+ \mathbb{E}_z[\| F_i(\varphi_{\delta_{n+1}}(Z_{t_n}), U_{n+1}) - F_j(\varphi_{\delta_{n+1}}(Z_{t_n}), U_{n+1}) \| \mathbb{1}_{A_{ij}}] & (\ddagger) \\
&+ \mathbb{E}_z[\| F_j(\varphi_{\delta_{n+1}}(Z_{t_n}), U_{n+1}) - \varphi_{\delta_{n+1} - \overline{\tau}_{n+1}}(F_j(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1})) \| \mathbb{1}_{A_{ij}}] & (\ddagger\dagger) \\
&+ \mathbb{E}_z[\| \varphi_{\delta_{n+1} - \overline{\tau}_{n+1}}(F_j(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1})) - \overline{\varphi}_{\delta_{n+1} - \overline{\tau}_{n+1}}(\overline{F}_j(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})) \| \mathbb{1}_{A_{ij}}] & (\ddagger\ddagger)
\end{aligned}
$$

To bound ($\dagger$) and ($\ddagger\dagger$) we use Assumption 4.2(c), while for ($\ddagger$) we add and subtract $\varphi_{\delta_{n+1}}(Z_{t_n})$ and use Assumption 4.2(a), and for ($\ddagger\ddagger$) we use a similar argument to the A_{ii} case. Combining this with the bound in (52) we obtain

$$
\begin{aligned}
\mathbb{E}_z[\| Z_{t_{n+1}} - \overline{Z}_{t_{n+1}} \| \mathbb{1}_{A_{ij}}] &\leq ((L(t_{n+1}, z))^2 (2D_3 + D_1) + \tilde{K}_2) \delta_{n+1}^2 \\
&\quad + \delta_{n+1} \tilde{K}_1 \mathbb{E}_z \left[\| Z_{t_n} - \overline{Z}_{t_n} \| \right]
\end{aligned}
$$

where $\tilde{K}_1, \tilde{K}_2$ are as in the statement of the lemma.

Let us finally consider $\overline{E}^c$, i.e. the case in which Z_t has two or more jumps. The probability this event is given by

$$
\begin{aligned}
\mathbb{P}_z(\overline{E}^c) &= \mathbb{E}_z\left[\int_0^{\delta_{n+1}}\left(1 - \exp\left(-\int_0^{\delta_{n+1}-t}\lambda(\varphi_s(F_{I_{n+1}}(\varphi_t(Z_{t_n}), U_{n+1})))\mathrm{d}s\right)\right)\right.\\
&\qquad\left.\lambda(\varphi_t(Z_{t_n}))\exp\left(-\int_0^t\lambda(\varphi_r(Z_{t_n}))\mathrm{d}r\right)\mathrm{d}t\right]\\
&\leq \mathbb{E}_z\left[\int_0^{\delta_{n+1}}\left(\int_0^{\delta_{n+1}-t}\lambda(\varphi_s(F_{I_{n+1}}(\varphi_t(z), U_{n+1})))\mathrm{d}s\right)\lambda(\varphi_t(z))\mathrm{d}t\right]\\
&\leq \delta_{n+1}^2(L(t_{n+1}, z))^2.
\end{aligned}
\tag{53}
$$

Then we can bound the norms $\|Z_{t_{n+1}}\|$ and $\|\overline{Z}_{t_{n+1}}\|$ by Assumption 4.7 to obtain

$$
\begin{aligned}
\mathbb{E}_z\left[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|\mathbb{1}_{E_{11}\cap\overline{E}^c}\right] &\leq 2B(t_{n+1}, z)\mathbb{E}_z\left[\mathbb{1}_{\overline{E}^c}\right]\\
&\leq 2\delta_{n+1}^2 B(t_{n+1}, z)(L(t_{n+1}, z))^2.
\end{aligned}
\tag{54}
$$

Combining the bounds on $\overline{E}$ and $\overline{E}^c$ we obtain the statement of the lemma. □

Lemma A.6. *Under Assumptions 4.1, 4.2(a), 4.3–4.7, it holds that*

$$
\begin{aligned}
\mathbb{E}_z\left[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|\mathbb{1}_{E_{10}}\right] &\leq \delta_{n+1}\, m(C')^2(D_1 D_4 + 2D_2 L(t_{n+1}, z))\,\mathbb{E}_z\left[\|Z_{t_n} - \overline{Z}_{t_n}\|\right]\\
&\quad + \delta_{n+1}^2(\tilde{C} + C'(D_2\tilde{C} + mD_1 M_2(t_n, z) + 2mM_1 L(t_{n+1}, z))).
\end{aligned}
$$

Proof. Recall that E_{10} is the event in which there are no switches for Z_s for $s \in (t_n, t_{n+1}]$ and there is one event for the approximation. Taking advantage of the coupling of the two processes as described in Coupling 6.1 we find that E_{10} takes place as long as for some i

$$
\tilde{U}_{n+1}^i \in \left(1 - \exp\left(-\int_0^{\delta_{n+1}}\lambda_i(\varphi_s(Z_{t_n}))\mathrm{d}s\right),\, 1 - \exp\left(-\int_0^{\delta_{n+1}}\overline{\lambda}_i(\overline{Z}_{t_n}, s)\mathrm{d}s\right)\right].
$$

Then we can estimate the probability of this event as follows:

$$
\begin{aligned}
\mathbb{E}_z[\mathbb{1}_{E_{10}}|Z_{t_n}, \overline{Z}_{t_n}] &\leq \sum_{i=1}^m\left|\exp\left(-\int_0^{\delta_{n+1}}\lambda_i(\varphi_s(Z_{t_n}))\mathrm{d}s\right) - \exp\left(-\int_0^{\delta_{n+1}}\overline{\lambda}_i(\overline{Z}_{t_n}, s)\mathrm{d}s\right)\right|\\
&\leq \sum_{i=1}^m\int_0^{\delta_{n+1}}\left|\lambda_i(\varphi_s(Z_{t_n})) - \overline{\lambda}_i(\overline{Z}_{t_n}, s)\right|\mathrm{d}s
\end{aligned}
$$

where we used that $\exp(-z)$ is 1-Lipschitz for $z \geq 0$. Then we find bounds for $\mathbb{E}_z[\mathbb{1}_{E_{10}}]$ and $\mathbb{E}_z[\mathbb{1}_{E_{10}}|Z_{t_n}, \overline{Z}_{t_n}]$ respectively. For the first case we use the triangle inequality, followed by observing that λ and φ_s are Lipschitz by the inequality shown in (47) and then Assumption 4.6:

$$
\begin{aligned}
\mathbb{E}_z[\mathbb{1}_{E_{10}}] &\leq \sum_{i=1}^{m} \mathbb{E}_z \left[\int_0^{\delta_{n+1}} \left| \lambda_i(\varphi_s(Z_{t_n})) - \lambda_i(\varphi_s(\overline{Z}_{t_n})) \right| \mathrm{d}s \right. \\
&\quad \left. + \int_0^{\delta_{n+1}} \left| \lambda_i(\varphi_s(\overline{Z}_{t_n})) - \overline{\lambda}_i(\overline{Z}_{t_n}, s) \right| \mathrm{d}s \right] \\
&\leq \sum_{i=1}^{m} \mathbb{E}_z \left[\int_0^{\delta_{n+1}} D_4 C' \| Z_{t_n} - \overline{Z}_{t_n} \| \mathrm{d}s + \int_0^{\delta_{n+1}} \delta_{n+1} \overline{M}_2(\overline{Z}_{t_n}) \mathrm{d}s \right] \\
&\leq \delta_{n+1} m D_4 C' \mathbb{E}_z[\| Z_{t_n} - \overline{Z}_{t_n} \|] + m \delta_{n+1}^2 \mathbb{E}_z[\overline{M}_2(\overline{Z}_{t_n})] \\
&\leq \delta_{n+1} m D_4 C' \mathbb{E}_z[\| Z_{t_n} - \overline{Z}_{t_n} \|] + \delta_{n+1}^2 m M_2(t_n, z),
\end{aligned}
\tag{55}
$$

where in the last inequality we used again Assumption 4.6. Alternatively, we can bound the switching rates by Lemma A.3:

$$
\begin{aligned}
\mathbb{E}_z[\mathbb{1}_{E_{10}} | Z_{t_n}, \overline{Z}_{t_n}] &\leq \sum_{i=1}^{m} \int_0^{\delta_{n+1}} \left(|\lambda_i(\varphi_s(Z_{t_n}))| + |\overline{\lambda}_i(\overline{Z}_{t_n}, s)| \right) \mathrm{d}s \\
&\leq 2 m \delta_{n+1} L(t_{n+1}, z),
\end{aligned}
\tag{56}
$$

Let us now focus on bounding the distance between the two processes. On event E_{10} we have $Z_{t_{n+1}} = \varphi_{\delta_{n+1}}(Z_{t_n})$, while $\overline{Z}_{t_{n+1}} = \overline{\varphi}_{\delta_{n+1} - \overline{\tau}_{n+1}}(\overline{F}(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1}))$ where $\overline{\tau}_{n+1}$ is the time of the event for the approximation. By triangle inequality we can decompose the distance in the following terms

$$
\begin{aligned}
\mathbb{E}_z[\| Z_{t_{n+1}} - \overline{Z}_{t_{n+1}} \| \mathbb{1}_{E_{10}}] &= \\
&= \mathbb{E}_z \left[\| \varphi_{\delta_{n+1}}(Z_{t_n}) - \overline{\varphi}_{\delta_{n+1} - \overline{\tau}_{n+1}}(\overline{F}_{\overline{I}_{n+1}}(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})) \| \mathbb{1}_{E_{10}} \right] \\
&\leq \mathbb{E}_z \left[\| \varphi_{\delta_{n+1}}(Z_{t_n}) - \varphi_{\delta_{n+1} - \overline{\tau}_{n+1}}(F_{\overline{I}_{n+1}}(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1})) \| \mathbb{1}_{E_{10}} \right] && (*) \\
&\quad + \mathbb{E}_z \left[\| \varphi_{\delta_{n+1} - \overline{\tau}_{n+1}}(F_{\overline{I}_{n+1}}(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1})) \right. && (**) \\
&\quad \left. - \varphi_{\delta_{n+1} - \overline{\tau}_{n+1}}(F_{\overline{I}_{n+1}}(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})) \| \mathbb{1}_{E_{10}} \right] \\
&\quad + \mathbb{E}_z \left[\| \varphi_{\delta_{n+1} - \overline{\tau}_{n+1}}(F_{\overline{I}_{n+1}}(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})) \right. && (***) \\
&\quad \left. - \overline{\varphi}_{\delta_{n+1} - \overline{\tau}_{n+1}}(\overline{F}_{\overline{I}_{n+1}}(\overline{\varphi}_{\overline{\tau}_{n+1}}(\overline{Z}_{t_n}), U_{n+1})) \| \mathbb{1}_{E_{10}} \right] \\
&= (*) + (**) + (***).
\end{aligned}
$$

In order to estimate term $(*)$ we apply inequality (47), then Assumption 4.2(a) by conditioning on $Z_{t_n}, \overline{\tau}_{n+1}$, and then we apply (55):

$$
\begin{aligned}
(*) &\leq C' \mathbb{E}_z[\| \varphi_{\overline{\tau}_{n+1}}(Z_{t_n}) - F_{\overline{I}_{n+1}}(\varphi_{\overline{\tau}_{n+1}}(Z_{t_n}), U_{n+1}) \| \mathbb{1}_{E_{10}}] \\
&\leq C' D_1 \mathbb{E}_z[\mathbb{1}_{E_{10}}] \\
&\leq \delta_{n+1} m D_4 (C')^2 D_1 \mathbb{E}_z[\| Z_{t_n} - \overline{Z}_{t_n} \|] + \delta_{n+1}^2 m C' D_1 M_2(t_n, z).
\end{aligned}
\tag{57}
$$

For term $(**)$ we use the same reasoning of (49) together with the estimate (56):

$$
(**) \leq \delta_{n+1} m 2 L(t_{n+1}, z)(C')^2 D_2 \mathbb{E}_z[\| Z_{t_n} - \overline{Z}_{t_n} \|] + C' D_2 \tilde{C} \delta_{n+1}^2.
$$

Then for term $(***)$ we follow the reasoning in (50) and apply estimate (56) to obtain

$$
(***) \leq \delta_{n+1}^2 \left(2 \, m C' M_1 L(t_{n+1}, z) + \tilde{C} \right).
$$

The statement of the lemma follows then by combining estimates (*), (**), (***).　□

Lemma A.7. *Under Assumptions 4.1, 4.2(a), 4.3, 4.4, 4.6, 4.7, it holds that*

$$\mathbb{E}_z\left[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|\mathbb{1}_{E_{01}}\right] \leq \delta_{n+1}\,mC'\big(L(t_{n+1}, z) + C'D_1D_4\big)\,\mathbb{E}_z\left[\|Z_{t_n} - \overline{Z}_{t_n}\|\right]$$
$$+ \delta_{n+1}^2\big(2B(t_{n+1}, z)(L(t_{n+1}, z))^2 + mC'D_1M_2(t_n, z) + \tilde{C}\big).$$

Proof. Recall that E_{01} is the event in which for $s \in [t_n, t_{n+1})$ there are no switches for $\overline{Z}_s$, while there is at least one event for Z_s. Similarly to the proof of Lemma A.5, let us denote as $\overline{E}$ the event in which there is exactly one event for Z_s in the current time interval. On $\overline{E}^c$ we can use the bound (54). On $\overline{E}$

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\|\mathbb{1}_{E_{01}\cap\overline{E}}] =$$
$$= \mathbb{E}_z\left[\left\|\varphi_{\delta_{n+1}-\tau_{n+1}}(F_{I_{n+1}}(\varphi_{\tau_{n+1}}(Z_{t_n}), U_{n+1})) - \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n})\right\|\mathbb{1}_{E_{01}\cap\overline{E}}\right]$$
$$\leq \mathbb{E}_z\left[\left\|\varphi_{\delta_{n+1}-\tau_{n+1}}(F_{I_{n+1}}(\varphi_{\tau_{n+1}}(Z_{t_n}), U_{n+1})) - \varphi_{\delta_{n+1}}(Z_{t_n})\right\|\mathbb{1}_{E_{01}\cap\overline{E}}\right] \qquad (*)$$
$$+ \mathbb{E}_z\left[\|\varphi_{\delta_{n+1}}(Z_{t_n}) - \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n})\|\mathbb{1}_{E_{01}\cap\overline{E}}\right]. \qquad (**)$$

In order to find an estimate for term (*) observe that the probability of E_{01} can be estimated similarly to what done for the probability of E_{10}. Then following the reasoning in (57) we obtain

$$(*) \leq \delta_{n+1}m(C')^2D_1D_4\mathbb{E}_z[\|Z_{t_n} - \overline{Z}_{t_n}\|] + \delta_{n+1}^2mC'D_1M_2(t_n, z).$$

Similarly, for term (**) it is sufficient to apply Lemma A.2 and then to bound $\mathbb{E}[\mathbb{1}_{E_{01}}]$ by the probability that the continuous process has a random event:

$$(**) \leq \delta_{n+1}mC'L(t_{n+1}, z)\mathbb{E}_z\left[\|Z_{t_n} - \overline{Z}_{t_n}\|\right] + \tilde{C}\delta_{n+1}^2. \qquad □$$

A.2. Proof of Corollary 4.11

Proof. We only need to show that Assumptions 4.2 and 4.7 are verified in this setting under Assumption 4.10. Clearly the process moves with bounded velocity, and thus 4.7 holds. Then let us focus on verifying Assumption 4.2 and consider the ℓ^1-norm. For part (a) it is clear that for $z = (x, v)$

$$\mathbb{E}[\|z - F_i(z, \tilde{U})\|] = \mathbb{E}[\|v - F_i^v((x, v), \tilde{U})\|] \leq V_{max}.$$

Then consider part (b). For $z' = (y, w)$

$$\mathbb{E}[\|F_i(z, \tilde{U}) - F_i(z', \tilde{U})\|] \leq \|x - y\| + \mathbb{E}[\|F_i^v((x, v), \tilde{U}) - F_i^v((y, w), \tilde{U})\|]$$
$$\leq \|x - y\| + \frac{\|v - w\|}{V_{min}}V_{max} + \mathbb{E}[\|F_i^v((x, w), \tilde{U}) - F_i^v((y, w), \tilde{U})\|]$$
$$\leq \max\left\{\frac{V_{max}}{V_{min}}, 1 + D\right\}\|z - z'\|.$$

In the second inequality we used the triangle inequality and that $\|v - w\| \geq V_{min}$, while in the last inequality we bounded the rightmost term by Assumption 4.10. Let us focus on part (c). For the position part we have for $s \in [0, \delta]$ and $z = (x, v)$

$$\mathbb{E}[\|\varphi_{\delta-s}(\varphi_s(z), F_i^v((\varphi_s(z), v), \tilde{U})) - \varphi_\delta(z)\|] =$$

$$= \mathbb{E}\Big[\Big\| x + \int_0^s \Phi(v)\mathrm{d}s + \int_0^{\delta-s} \Phi(F_i^v((\varphi_s(z), v), U))\mathrm{d}r - x - \int_0^\delta \Phi(v)\mathrm{d}s \Big\|\Big]$$

$$\leq \mathbb{E}[\|s\,\Phi(v) + (\delta - s)\,\Phi(F_i^v((\varphi_s(z), v), U)) - \delta\,\Phi(v)\|]$$

$$\leq \mathbb{E}[\|(\delta - s)(\Phi(F_i^v((\varphi_s(z), v), U)) - \Phi(v))\|]$$

$$\leq \delta C \mathbb{E}[\| F_i^v((\varphi_s(z), v), U) - v\|]$$

$$\leq \delta C V_{max}.$$

On the other hand, for the velocity part we obtain using Assumptions 4.1 and 4.10

$$\mathbb{E}[\| F_i^v((\varphi_s(z), v), \tilde{U}) - F_i^v((\varphi_\delta(z), v), \tilde{U})\|] \leq D\|\varphi_s(z) - \varphi_\delta(z)\|$$

$$\leq DC'\|z - \varphi_{\delta-s}(z)\|$$

$$\leq \delta D(C')^2.$$

Therefore part (c) of Assumption 4.2 holds with $D_3 = D(C')^2 + C V_{max}$. □

A.3. Proof of Proposition 4.13

Proof. In the proof of Theorem 4.9 boundedness of the PDMP is used only in the case $p = 1$ to deal with the event in which the PDMP has two or more jumps in the same time step (see Lemmas A.5 and A.6, in particular Eq. (54)). Then it is sufficient to show that a similar bound holds also under Assumption 4.12 instead of Assumption 4.7. Let $p = 1$ and consider the case of Lemma A.6, i.e. restricting to event $E_{10} \cap \overline{E}$, which is the event in which the continuous time process has two or more jumps, while the approximation has zero jumps. Then we want to bound

$$\mathbb{E}_z[\| Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_{10} \cap \overline{E}}] =$$

$$= \sum_{\ell \geq 2} \int_{A_\ell} \mathbb{E}_z\Big[\|\varphi_{s_\ell} \circ F_{I_{n+1}^\ell}(\cdot, U_{\ell-1}) \circ \varphi_{s_{\ell-1}} \circ \ldots \circ F_{I_{n+1}^1}(\cdot, U_0) \circ \varphi_{s_0}(Z_{t_n}) - \overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n})\| p_{Z_{t_n}}(\mathrm{d}s)\Big],$$

where $s_0, \ldots, s_{\ell-1}$ are the interarrival times of the random jumps of Z_t, I_{n+1}^ℓ denote the index of the ℓ-th jump to occur between time t_n and t_{n+1}, $s_\ell = \delta_{n+1} - \sum_{i=1}^{\ell-1} s_i$, $U_0, \ldots, U_{\ell-1} \overset{iid}{\sim} \nu_{\mathcal{U}}$,

$$A_\ell = \Big\{ s = (s_0, \ldots, s_\ell) : \sum_{i=1}^\ell s_i = \delta_{n+1}, s_i > 0 \Big\},$$

and $p_{Z_{t_n}}(\mathrm{d}s)$ is the law of the interarrival times,. Then we have

$$\mathbb{E}_z[\| Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_{10} \cap \overline{E}}] \leq \mathbb{E}\Bigg[\sum_{\ell \geq 2} \int_{A_\ell} \mathbb{E}_z \Big[\Big(\|\varphi_{s_\ell} \circ F_{I_{n+1}^\ell}(\cdot, U_{\ell-1}) \circ \cdots \circ \varphi_{s_0}(Z_{t_n})\|$$

$$+ \|\overline{\varphi}_{\delta_{n+1}}(\overline{Z}_{t_n})\| \Big) \Big| Z_{t_n} \Big] p_{Z_{t_n}}(\mathrm{d}s) \Bigg]$$

Now we use (47) to conclude that φ_s has linear growth for some constant L, and so $\|\varphi_s(z)\| \leq L(\|z\| + 1)$. It follows that

$$\mathbb{E}_z\Big[\|\varphi_{s_\ell} \circ F_{I_{n+1}^\ell}(\cdot, U_{\ell-1}) \circ \varphi_{s_{\ell-1}} \circ F_{I_{n+1}^{\ell-1}}(\cdot, U_{\ell-2}) \circ \cdots \circ \varphi_{s_0}(Z_{t_n})\| | Z_{t_n} \Big] \leq$$

$$\leq L \mathbb{E}_z\Big[\Big(\|F_{I_{n+1}^\ell}(\cdot, U_{\ell-1}) \circ \varphi_{s_{m-1}} \circ F_{I_{n+1}^{\ell-1}}(\cdot, U_{\ell-2}) \circ \cdots \circ \varphi_{s_0}(Z_{t_n})\| + 1 \Big) | Z_{t_n} \Big]$$

$$\leq L(1 + D_1) + L\mathbb{E}_z\left[\left(\|\varphi_{s_{\ell-1}} \circ F_{I_{n+1}^{\ell-1}}(\cdot, U_{\ell-2}) \circ \cdots \circ \varphi_{s_0}(Z_{t_n})\|\right) | Z_{t_n}\right].$$

In the last inequality we used that $\mathbb{E}[\|F_i(z, U)\|] \leq \|z\| + D_1$ for any i, which is implied by Assumption 4.2(a). Therefore by recursion we have

$$\mathbb{E}_z\left[\|\varphi_{s_\ell} \circ F_{I_{n+1}^\ell}(\cdot, U_{\ell-1}) \circ \cdots \circ \varphi_{s_0}(Z_{t_n})\| | Z_{t_n}\right] \leq \sum_{i=1}^\ell (1 + D_1)L^i + L^{\ell+1}(1 + \|Z_{t_n}\|).$$

Moreover we also have that $\|\overline{\varphi}_s(z)\| \leq \delta_{n+1}^2 + L(\|z\| + 1)$ by Assumption 4.4. It follows that

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_{10} \cap \overline{E}}] =$$

$$\leq \sum_{\ell \geq 2} \mathbb{E}_z\left[\int_{A_\ell}\left(\sum_{i=1}^\ell (1 + D_1)L^i + L^{\ell+1}(1 + \|Z_{t_n}\|) + \delta_{n+1}^2 + L(\|\overline{Z}_{t_n}\| + 1)\right)p_{Z_{t_n}}(\mathrm{d}s)\right]$$

$$\leq \mathbb{E}_z\left[\sum_{\ell \geq 2}\left(\sum_{i=1}^\ell (1 + D_1)L^i + L^{\ell+1}(1 + \|Z_{t_n}\|) + \delta_{n+1}^2 + L(\|\overline{Z}_{t_n}\| + 1)\right)p_{Z_{t_n}}(A_\ell)\right]$$

$$\leq \tilde{L}\mathbb{E}_z\left[(1 + \|Z_{t_n}\| + \|\overline{Z}_{t_n}\|)\sum_{\ell \geq 2}\ell L^\ell p_{Z_{t_n}}(A_\ell)\right]$$

for some constant $\tilde{L}$ which depends only on D_1, L, δ_0. The function $f(\ell) = \ell L^\ell$ is increasing in the number of jumps and therefore because the switching rates have a global upper bound λ_{max} we obtain

$$\mathbb{E}_z[\|Z_{t_{n+1}} - \overline{Z}_{t_{n+1}}\| \mathbb{1}_{E_{10} \cap \overline{E}}] \leq$$

$$\leq \tilde{L}\mathbb{E}_z\left[(1 + \|Z_{t_n}\| + \|\overline{Z}_{t_n}\|)\sum_{\ell \geq 2}\ell L^\ell e^{-\delta_{n+1}\lambda_{max}}\frac{(\delta_{n+1}\lambda_{max})^\ell}{\ell!}\right]$$

$$\leq \tilde{L}(1 + 2B(t_n, z))\sum_{\ell \geq 2}\ell L^\ell e^{-\delta_{n+1}\lambda_{max}}\frac{(\delta_{n+1}\lambda_{max})^\ell}{\ell!},$$

where in the last inequality we used Assumption 4.12. It remains to show that the sum is of order δ_{n+1}^2. This can be proved as follows

$$\sum_{\ell \geq 2}\ell L^\ell e^{-\delta_{n+1}\lambda_{max}}\frac{(\delta_{n+1}\lambda_{max})^\ell}{\ell!} = e^{(L-1)\delta_{n+1}\lambda_{max}}\sum_{\ell \geq 2}e^{-L\delta_{n+1}\lambda_{max}}\frac{(L\delta_{n+1}\lambda_{max})^\ell}{(\ell-1)!}$$

$$= e^{(L-1)\delta_{n+1}\lambda_{max}}L\delta_{n+1}\lambda_{max}\sum_{\ell \geq 1}e^{-L\delta_{n+1}\lambda_{max}}\frac{(L\delta_{n+1}\lambda_{max})^\ell}{\ell!}$$

$$= e^{(L-1)\delta_{n+1}\lambda_{max}}L\delta_{n+1}\lambda_{max}(1 - e^{-L\delta_{n+1}\lambda_{max}})$$

$$\leq \delta_{n+1}^2 e^{(L-1)\delta_0\lambda_{max}}L\lambda_{max}.$$

In particular we used that $\delta_n \leq \delta_0$ for all $n \in \mathbb{N}$.

The same proof holds on the event $E_{10} \cap \overline{E}$, and thus we have proved the wanted result. □

Appendix B. Proofs of Section 4.2

B.1. Proof of Theorem 4.17: the case of $p = 1$

Proof of Lemma 7.2. Let us take advantage of the construction in Coupling 7.1. First consider the case in which $T_{i*}(Z_{t_{n-1}}) > \delta_n$. In this case there are no random events for either process in the time interval $(t_{n-1}, t_n]$ and therefore $Z_{t_n} = \overline{Z}_{t_n} = \varphi_{\delta_n}(Z_{t_{n-1}})$. Now, consider the case where $T_{i*} \leq \delta_n$. In this scenario, there are three disjoint events:

- The proposed switching time is accepted by both processes. Denote this event as E_1.
- The proposed switching time is accepted by one process, and rejected by the other. Denote this event as E_2.
- The proposed switching time is rejected for both processes. Denote this event as E_3.

Therefore we have

$$\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n} | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}) = \sum_{i=1}^{3} \mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}, E_i | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}).$$

We start with event E_1. In this case we have that $\overline{Z}_{t_n} \neq Z_{t_n}$ if the continuous time process has at least one more jump in time interval $(t_{n-1} + T_{i*}, t_n]$. Now let $\lambda(z) = \sum_{i=1}^{m} \lambda_i(z)$ and $\lambda_{tot}(z, t; \delta_n) = \sum_{i=1}^{m} \lambda_{tot}^i(z, t; \delta_n)$. Observe that, conditional on $Z_{t_{n-1}}$, the minimum of the m proposed random times is distributed as $\mathbb{P}(T_{i*} \leq t) = 1 - \exp(-\int_0^t \lambda_{tot}(Z_{t_{n-1}}, s; \delta_n) ds)$. Then bounding by 1 the probability that both proposals are accepted, and conditioning on $Z_{t_{n-1}}$ we obtain

$$\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}, E_1 | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}) \leq \mathbb{E}_z \Bigg[\int_0^{\delta_n} \lambda_{tot}(Z_{t_{n-1}}, t; \delta_n) e^{-\int_0^t \lambda_{tot}(Z_{t_{n-1}}, s; \delta_n) ds}$$
$$\left(1 - \exp\left(-\int_t^{\delta_n} \lambda(\varphi_s(F_{i*}(\varphi_t(Z_{t_{n-1}}), U_n))) ds \right) \right) dt \Bigg].$$

Then using that $1 - \exp(-z) \leq z$, that $\exp(-z) \leq 1$ for $z \geq 0$ and by Fubini's theorem we obtain the following bound:

$$\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}, E_1 | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}) \leq$$
$$\leq \mathbb{E}_z \Bigg[\int_0^{\delta_n} \int_t^{\delta_n} \lambda_{tot}(Z_{t_{n-1}}, t; \delta_n) \lambda(\varphi_s(F_{i*}(\varphi_t(Z_{t_{n-1}}), U_n))) ds dt \Bigg]$$
$$\leq \int_0^{\delta_n} \int_t^{\delta_n} \mathbb{E}_z \left[\lambda_{tot}(Z_{t_{n-1}}, t; \delta_n) \lambda(\varphi_s(F_{i*}(\varphi_t(Z_{t_{n-1}}), U_n))) \right] ds dt$$
$$\leq \delta_n^2 L_1(t_n, z)/2.$$

Note that in the last inequality the bound $L_1(t_n, z)$ follows from part (a) of Assumption 4.14.

Let us now consider event E_2. As the proposal $T_{i*}(Z_{t_{n-1}})$ is accepted for one process only, it must be that

$$\overline{U} \in \Bigg(\min \left\{ \frac{\lambda_{i*}(\varphi_{T_{i*}}(Z_{t_{n-1}}))}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}(Z_{t_{n-1}}); \delta_n)}, \frac{\overline{\lambda}_{i*}(Z_{t_{n-1}}, T_{i*}(Z_{t_{n-1}}))}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}(Z_{t_{n-1}}); \delta_n)} \right\},$$
$$\max \left\{ \frac{\lambda_{i*}(\varphi_{T_{i*}}(Z_{t_{n-1}}))}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}(Z_{t_{n-1}}); \delta_n)}, \frac{\overline{\lambda}_{i*}(Z_{t_{n-1}}, T_{i*}(Z_{t_{n-1}}); \delta_n)}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}(Z_{t_{n-1}}); \delta_n)} \right\} \Bigg].$$

Therefore using that $\overline{U}$ and T_{i*} are independent we obtain

$$\mathbb{P}(Z_{t_n} \neq \overline{Z}_{t_n}, E_2 | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}) =$$

$$= \mathbb{E}\left[\mathbb{1}_{\{T_{i*}(Z_{t_{n-1}})\leq\delta_n\}}\left|\frac{\overline{\lambda}_{i*}(Z_{t_{n-1}}, T_{i*}(Z_{t_{n-1}}); \delta_n) - \lambda_{i*}(\varphi_{T_{i*}}(Z_{t_{n-1}}))}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}(Z_{t_{n-1}}); \delta_n)}\right|\right].$$

By the definition given in [Coupling 7.1](#) we have $\lambda_{tot}^{i*}(z, t; \delta_n) \geq 1$. Using part (b) of [Assumption 4.14](#) and Fubini's theorem:

$$\mathbb{P}(Z_{t_n} \neq \overline{Z}_{t_n}, E_2 | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}) \leq \delta_n \mathbb{E}_z\left[\overline{M}_2(Z_{t_{n-1}})\mathbb{1}_{\{T_{i*}(Z_{t_{n-1}})\leq\delta_n\}}\right]$$

$$\leq \delta_n \mathbb{E}_z\left[\overline{M}_2(Z_{t_{n-1}})\int_0^{\delta_n} \lambda_{tot}(Z_{t_{n-1}}, t; \delta_n)e^{-\int_0^t \lambda_{tot}(Z_{t_{n-1}}, s; \delta_n)ds}dt\right]$$

$$\leq \delta_n \int_0^{\delta_n} \mathbb{E}_z\left[\lambda_{tot}(Z_{t_{n-1}}, t; \delta_n)\overline{M}_2(Z_{t_{n-1}})\right]dt$$

$$\leq \delta_n^2 L_2(t_n, z).$$

Finally, we focus on E_3. On this event, the processes remain equal unless there is (at least) a switch for either process for $t \in (t_{n-1} + T_{i*}(Z_{t_n}), t_n)$. Recall $\overline{\lambda}(z, s; \delta_n) = \sum_{i=1}^m \overline{\lambda}_i(z, s; \delta_n)$. Using this observation together with [Assumption 4.14](#) and the facts that on this event $T_{i*}(Z_{t_n}) \leq \delta_n$ and that $1 - \exp(-z) \leq z$ we obtain

$$\mathbb{P}(Z_{t_n} \neq \overline{Z}_{t_n}, E_3 | Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}) \leq \mathbb{E}_z\left[\int_0^{\delta_n}\left(\left(1 - \exp\left(-\int_t^{\delta_n} \lambda(\varphi_r(Z_{t_{n-1}}))dr\right)\right)\right.\right.$$

$$\left.+ \left(1 - \exp\left(-\int_t^{\delta_n} \overline{\lambda}(Z_{t_{n-1}}, r; \delta_n)dr\right)\right)\right)$$

$$\left.\lambda_{tot}(Z_{t_{n-1}}, t; \delta_n)\exp\left(-\int_0^t \lambda_{tot}(Z_{t_{n-1}}, s; \delta_n)ds\right)dt\right]$$

$$\leq \mathbb{E}_z\left[\int_0^{\delta_n}\int_t^{\delta_n} \lambda_{tot}(Z_{t_{n-1}}, t; \delta_n)\left(\lambda(\varphi_r(Z_{t_{n-1}})) + \overline{\lambda}(Z_{t_{n-1}}, r; \delta_n)\right)drdt\right]$$

$$= \int_0^{\delta_n}\int_t^{\delta_n} \mathbb{E}_z\left[\lambda_{tot}(Z_{t_{n-1}}, t; \delta_n)\left(\lambda(\varphi_r(Z_{t_{n-1}})) + \overline{\lambda}(Z_{t_{n-1}}, r; \delta_n)\right)\right]drdt$$

$$\leq \delta_n^2 L_3(t_n, z)/2.$$

Combining the three bounds on events E_1, E_2, E_3 we obtain the statement. $\square$

B.2. Proof of [Theorem 4.17](#): the case of $p > 1$

Proof of [Theorem 4.17](#). Observe that if $T_{i*} > \delta$ the two processes are equal at time δ and thus the probability that $Z_{t_n} \neq \overline{Z}_{t_n}$ is 0. We analyse in turn the three events E_1, E_2, E_3 which were defined in Section [7.1](#) in the proof of [Lemma 7.2](#). Define the event $E_= = \{Z_{t_{n-1}} = \overline{Z}_{t_{n-1}}\}$.

On event E_1, the proposal T_{i*} is accepted by both processes. Then we reformulate $\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}, E_1 | E_=)$ in terms of the conditional probability

$$\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}, E_1 | E_=) = \mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n} | Z_{t_{n-1}+T_{i*}} = \overline{Z}_{t_{n-1}+T_{i*}}, T_{i*} < \delta)\mathbb{P}_z(E_1 | E_=).$$

The first term on the right hand side can be bounded by applying Inductive [Hypothesis 7.3](#). Moreover we can use the bound $\mathbb{P}_z(E_1 | E_=) \leq \mathbb{P}_z(T_{i*} < \delta | E_=)$ for the rightmost term to obtain

$$\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}, E_1 | E_=) \leq A\delta^{p+1}\mathbb{E}_z\left[\left(1 - \exp\left(-\int_0^\delta \lambda_{tot}(Z_{t_{n-1}}, t; \delta, p+1)dt\right)\right)\right]$$

$$\leq A\delta^{p+2} \sup_{s\in[0,\delta]} \mathbb{E}_z \left[\lambda_{tot}(Z_{t_n}, s; \delta, p+1) \right] \leq A\delta^{p+2} L_4(t_n, z).$$

In the last inequality we took advantage of the bound $1 - \exp(-z) \leq x$ which is true for $z > 0$.

On event E_2 the proposal T_{i*} is accepted for one process, and rejected for the other. This happens when

$$\overline{U} \in \left(\min\left\{ \frac{\lambda_{i*}(\varphi_{T_{i*}}(Z_{t_{n-1}}))}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}; \delta, p+1)}, \frac{\overline{\lambda}_{i*}(Z_{t_{n-1}}, T_{i*}; \delta, p+1)}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}; \delta, p+1)} \right\}, \right.$$

$$\left. \max\left\{ \frac{\lambda_{i*}(\varphi_{T_{i*}}(Z_{t_{n-1}}))}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}; \delta, p+1)}, \frac{\overline{\lambda}_{i*}(Z_{t_{n-1}}, T_{i*}; \delta, p+1)}{\lambda_{tot}^{i*}(Z_{t_{n-1}}, T_{i*}; \delta, p+1)} \right\} \right],$$

and therefore with probability

$$\left| \frac{\overline{\lambda}_{i*}(Z_{t_{n-1}}, T_{i*}; \delta, p+1) - \lambda_{i*}(\varphi_{T_{i*}}(Z_{t_{n-1}}))}{\lambda_{tot}^{i*}(z, T_{i*}(z); \delta, p+1)} \right| \leq \left| \overline{\lambda}_{i*}(Z_{t_{n-1}}, T_{i*}; \delta, p+1) - \lambda_{i*}(\varphi_{T_{i*}}(Z_{t_{n-1}})) \right|$$

$$\leq \delta^{p+1} \overline{M}_2(Z_{t_{n-1}})$$

where we used that by definition $\lambda_{tot}^{i*} \geq 1$ and then that $\overline{\lambda}_{i*}(\cdot, \cdot; \delta, p+1)$ is an approximation of $p+1$ order. Thus we have

$$\mathbb{P}_z(Z_{t_n} \neq \overline{Z}_{t_n}, E_2|E_=) \leq \delta^{p+1} \mathbb{E}_z \left[\overline{M}_2(Z_{t_{n-1}}) \mathbb{P}_z(T_{i*} < \delta|Z_{t_{n-1}}, E_=) \right]$$

$$\leq \delta^{p+2} \sup_{s\in[0,\delta]} \mathbb{E}_z \left[\overline{M}_2(Z_{t_{n-1}}) \lambda_{tot}(Z_{t_n}, s; \delta, p+1) \right]$$

$$\leq \delta^{p+2} L_4(t_n, z)$$

Finally consider event E_3. Similarly to the proof of Theorem 4.9 it is sufficient to bound the event that $p+2$ proposal times occur before the end of the time interval, which is bounded by Assumption 4.14. $\square$

Appendix C. Proofs of Section 4.3

C.1. Proofs of Theorem 4.24 and its corollaries

Lemma C.1. *Suppose λ and $\overline{\lambda}$ satisfy Assumption 4.6(a). We will consider the two algorithms separately. For Algorithm 2, let $p_{\overline{\tau}}^{z,\delta,2}$ be given by (38) then for any $h \in \mathcal{C}_b^1([0,\delta])$ we have*

$$\left| \int_0^\delta h_s \, p_{\overline{\tau}}^{z,\delta,2}(\mathrm{d}s) - h_s \lambda(\varphi_s(z))\mathrm{d}s \right| \leq \delta^2 \sup_{s,r\in[0,\delta]} \left(|\partial_r h_r| \overline{\lambda}(z, s; \delta) \right.$$

$$\left. + |h_s| (\overline{\lambda}(z, s; \delta)\overline{\lambda}(z, r; \delta) + \overline{M}_2(z)) \right).$$

For Algorithm 3, let $p_{\overline{\tau}}^{z,\delta,3}$ be given by (39) then for any $h \in \mathcal{C}_b([0,\delta])$

$$\left| \int_0^\delta h_s \, p_{\overline{\tau}}^{z,\delta,3}(\mathrm{d}s) - \int_0^\delta \lambda(\varphi_s(z))h_s \mathrm{d}s \right| \leq \delta^2 \sup_{s,r\in[0,\delta]} \left(|h_s| (\overline{\lambda}(z, s; \delta)\overline{\lambda}(z, r; \delta) + \overline{M}_2(z)) \right)$$

Proof of Lemma C.1. First consider the case where $p_{\overline{\tau}}^{z,\delta,2}$ is given by (38), and fix $h \in \mathcal{C}_b^1([0,\delta])$. Then

$$\left| \int_0^\delta h_s \, p_{\overline{\tau}}^{z,\delta,2}(\mathrm{d}s) - h_s \lambda(\varphi_s(z))\mathrm{d}s \right| = \left| h_\delta \left(1 - e^{-\int_0^\delta \overline{\lambda}(z,s;\delta)\mathrm{d}s} \right) - \int_0^\delta h_s \lambda(\varphi_s(z))\mathrm{d}s \right|.$$

We can rewrite

$$1 - e^{-\int_0^\delta \overline{\lambda}(z,s;\delta)\mathrm{d}s} = \int_0^\delta \overline{\lambda}(z,s;\delta)e^{-\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r}\mathrm{d}s.$$

Therefore we have

$$\left| \int_0^\delta h_s \, p_{\overline{\tau}}^{z,\delta,2}(\mathrm{d}s) - h_s \lambda(\varphi_s(z))\mathrm{d}s \right| = \left| h_\delta \int_0^\delta \overline{\lambda}(z,s;\delta)e^{-\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r}\mathrm{d}s - \int_0^\delta h_s \lambda(\varphi_s(z))\mathrm{d}s \right|$$

$$\leq \left| \int_0^\delta (h_\delta - h_s)\overline{\lambda}(z,s;\delta)e^{-\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r}\mathrm{d}s \right|$$

$$+ \left| \int_0^\delta h_s \left(\overline{\lambda}(z,s;\delta)e^{-\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r} - \lambda(\varphi_s(z)) \right)\mathrm{d}s \right|.$$

(58)

We can use Assumption 4.6(a) and that $1 - e^{-y} \leq y$ for $y > 0$ to bound the integrand of the second term on the right of (58),

$$\left| \overline{\lambda}(z,s;\delta)e^{-\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r} - \lambda(\varphi_s(z)) \right| \leq \left| \overline{\lambda}(z,s;\delta)(1 - e^{-\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r}) \right|$$

$$+ \left| \overline{\lambda}(z,s;\delta) - \lambda(\varphi_s(z)) \right|$$

$$\leq \overline{\lambda}(z,s;\delta)\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r + \delta \overline{M}_2(z).$$

(59)

For the first term on the right hand side of (58) we use that

$$|h_\delta - h_s| \leq (\delta - s) \sup_{r \in [0,\delta]} |\partial_r h_r|.$$

(60)

Applying (59) and (60) to (58) we have

$$\left| \int_0^\delta h_s \, p_{\overline{\tau}}^{z,\delta,2}(\mathrm{d}s) - h_s \lambda(\varphi_s(z))\mathrm{d}s \right| \leq$$

$$\leq \left| \sup_{r \in [0,\delta]} |\partial_r h_r| \int_0^\delta (\delta - s)\overline{\lambda}(z,s;\delta)e^{-\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r}\mathrm{d}s \right|$$

$$+ \left| \int_0^\delta |h_s| \left(\overline{\lambda}(z,s;\delta)\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r + \delta \overline{M}_2(z) \right)\mathrm{d}s \right|$$

$$\leq \delta^2 \sup_{s,r \in [0,\delta]} \left(|\partial_r h_r|\overline{\lambda}(z,s;\delta) + |h_s| \left(\overline{\lambda}(z,s;\delta)\overline{\lambda}(z,r) + \overline{M}_2(z) \right) \right).$$

Let us consider the case where $p_{\overline{\tau}}^{z,\delta,3}$ is given by (39). We use (59) to bound

$$\left| \int_0^\delta h_s \, p_{\overline{\tau}}^{z,\delta,3}(\mathrm{d}s) - h_s \lambda(\varphi_s(z))\mathrm{d}s \right| \leq$$

$$\leq \int_0^\delta |h_s| \left| \overline{\lambda}(z,s;\delta)\exp\left(-\int_0^s \overline{\lambda}(z,r;\delta)\mathrm{d}r \right) - \lambda(\varphi_s(z)) \right| \mathrm{d}s$$

$$\leq \int_0^\delta |h_s| \left(\overline{\lambda}(z,s;\delta)\int_0^\delta \overline{\lambda}(z,r;\delta)\mathrm{d}r + \delta \overline{M}_2(z) \right)\mathrm{d}s$$

$$\leq \delta^2 \sup_{s,r \in [0,\delta]} \left(|h_s|(\overline{\lambda}(z,s;\delta)\overline{\lambda}(z,r;\delta) + \overline{M}_2(z)) \right). \quad \square$$

Proof of Corollary 4.27. First observe that (16) follows from (13). Then (17) is obtained by adding (11) and (13). To obtain (18) we use that

$$\left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_z[g(\overline{Z}_{t_n})] - \mu(g) \right| \leq \left| \frac{1}{N} \sum_{n=1}^{N} (\mathbb{E}_z[g(\overline{Z}_{t_n})] - \mathbb{E}_z[g(Z_{t_n})]) \right|$$
$$+ \left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_z[g(Z_{t_n})] - \mu(g) \right|.$$

We bound this using (11) and (13)

$$\left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}_z[g(\overline{Z}_{t_n})] - \mu(g) \right| \leq C\delta\overline{G}_2(z) + C\overline{G}_2(z) \frac{1}{N} \sum_{n=1}^{N} e^{-\omega t_n}$$
$$\leq C\overline{G}_2(z) \left(\delta + \frac{1}{t_N} \right). \quad \square$$

Proof of Corollary 4.28. It is sufficient to show for S_n given by (14) that $S_n \to 0$ as $n \to \infty$. Fix $\eta > 0$. Then we have

$$S_n = \sum_{k=0}^{\eta-1} \delta_{k+1}^2 e^{-\omega(t_n - t_{k+1})} + \sum_{k=\eta}^{n-1} \delta_{k+1}^2 e^{-\omega(t_n - t_{k+1})}.$$

Consider the first term:

$$\sum_{k=0}^{\eta-1} \delta_{k+1}^2 e^{-\omega(t_n - t_{k+1})} \leq \sup_k \delta_k \int_0^{t_n - \eta} e^{-\omega(t_n - s)} \mathrm{d}s = \sup_k \delta_k \frac{e^{-\omega\eta} - e^{-\omega t_n}}{\omega}.$$

Consider the second term:

$$\sum_{k=\eta}^{n-1} \delta_{k+1}^2 e^{-\omega(t_n - t_{k+1})} \leq \left(\sup_{k \in \{\eta,\dots,n\}} \delta_k \right) \int_{t_\eta}^{t_n} e^{-\omega(t_n - s)} \mathrm{d}s = \frac{1 - e^{-\omega(t_n - t_\eta)}}{\omega} \sup_{k \in \{\eta,\dots,n\}} \delta_k.$$

Therefore

$$\limsup_{n \to \infty} S_n \leq \left(\sup_{k \geq 0} \delta_k \right) \frac{e^{-\omega\eta}}{\omega} + \frac{1}{\omega} \sup_{k \geq \eta} \delta_k.$$

Since η is arbitrary we let η tend to ∞ which gives that $S_n \to 0$ as $n \to \infty$. $\quad \square$

C.2. Proofs of Example 5.8

Proof of Proposition 5.9. Fix $f \in C_b^1(\mathbb{R}^d \times \mathbb{R}^d)$. Then by the chain rule

$$\|\nabla_{q,p} \mathcal{P}_t f(q,p)\| = \|\mathbb{E}[\nabla_{q,p}(Q_t, P_t)(\nabla_{q,p} f)(Q_t, P_t)]\| \leq \|f\|_{C_b^1} \mathbb{E}[\|\nabla_{q,p}(Q_t, P_t)\|].$$

Notice that there is a version of (Q_t, P_t) which is differentiable with respect to the initial conditions since we can write (Q_t, P_t) as the composition of smooth operators. Let T_i denote the ith refreshment time and $\xi_i \sim N(0_d, I_d)$ the corresponding refreshed velocity. Set $T_0 = 0$. We shall track for which refreshment times we have that $\nu \leq T_i - T_{i-1} \leq K$. Let M_t denote the number of refreshment times before time t which have this property and let N_t denote the total number of refreshment times before time t. Note that conditional on N_t, M_t is distributed according to a Binomial distribution with N_t trials and success rate $e^{-\lambda\nu} - e^{-\lambda K_1}$.

To stress the dependence on the initial condition for the remainder of the proof we shall write $(Q_t^{q,p}, P_t^{q,p})$ to denote the process at time t with initial condition (q, p). Then by (23) we have

$$\|(Q_t^{q,p}, P_t^{q,p}) - (Q_t^{\bar{q},\bar{p}}, P_t^{\bar{q},\bar{p}})\| = \|\varphi_{t-T_{N_t}}(Q_{T_{N_t}}^{q,p}, \xi_{N_t}) - \varphi_{t-T_{N_t}}(Q_{T_{N_t}}^{\bar{q},\bar{p}}, \xi_{N_t})\|$$
$$\leq C\|Q_{T_{N_t}}^{q,p} - Q_{T_{N_t}}^{\bar{q},\bar{p}}\|.$$

There are now three possible events either $N_t = 0$, $\nu \leq T_{N_t} - T_{N_t-1} \leq K$ or $T_{N_t} - T_{N_t-1} \geq K$. If $\nu \leq T_{N_t} - T_{N_t-1} \leq K$ then we use (24), however if $T_{N_t} - T_{N_t-1} \geq K$ then we use (23). By doing this for each refreshment we have

$$\|(Q_t^{q,p}, P_t^{q,p}) - (Q_t^{\bar{q},\bar{p}}, P_t^{\bar{q},\bar{p}})\| \leq C^{1+N_t-M_t} \gamma^{M_t} \|Q_{T_1}^{q,p} - Q_{T_1}^{\bar{q},\bar{p}}\|.$$

Then by applying (23) once more we have

$$\|(Q_t^{q,p}, P_t^{q,p}) - (Q_t^{\bar{q},\bar{p}}, P_t^{\bar{q},\bar{p}})\| \leq C^{2+N_t-M_t} \gamma^{M_t} \|(q, p) - (\bar{q}, \bar{p})\|.$$

Dividing by $\|(q, p) - (\bar{q}, \bar{p})\|$ and taking the limit as $\|(q, p) - (\bar{q}, \bar{p})\| \to 0$ we have that

$$\|\nabla_{q,p}(Q_t^{q,p}, P_t^{q,p})\| \leq C^{2+N_t-M_t} \gamma^{M_t}.$$

It remains to bound $\mathbb{E}[C^{N_t-M_t} \gamma^{M_t}]$. By conditioning on N_t we can use the moment generating function of a Binomial distribution to find

$$\mathbb{E}[C^{N_t-M_t} \gamma^{M_t} | N_t] = C^{N_t}(1 - (e^{-\lambda\nu} - e^{-\lambda K})(1 - \gamma C^{-1}))^{N_t}.$$

Now N_t is a Poisson process with rate λ so we have

$$\mathbb{E}[C^{N_t-M_t} \gamma^{M_t}] = \exp\left(\lambda\left(C(1 - (e^{-\lambda\nu} - e^{-\lambda K})(1 - \gamma C^{-1})) - 1\right)\right).$$

This is decays exponentially provided

$$C(1 - (e^{-\lambda\nu} - e^{-\lambda K_1})(1 - \gamma C^{-1})) < 1. \quad \square$$

C.3. Proofs of Example 5.11

Proof of Lemma 5.12. Note that for the ZZS

$$[\Phi, Q]f(x, v) = \sum_{i=1}^{d}\left\langle v, \nabla_x\left(\frac{\lambda_i(x, v)}{\lambda(x, v)}f(x, F_i v)\right)\right\rangle - \sum_{i=1}^{d}\frac{\lambda_i(x, v)}{\lambda(x, v)}\langle F_i v, \nabla_x(f(x, F_i v))\rangle$$

$$= \sum_{i=1}^{d}\left\langle v, \nabla_x\left(\frac{\lambda_i(x, v)}{\lambda(x, v)}\right)\right\rangle f(x, F_i v) + 2\sum_{i=1}^{d}\left(\frac{\lambda_i(x, v)}{\lambda(x, v)}\right)v_i\partial_{x_i}f(x, F_i v).$$

When we apply this with $f = \mathcal{P}_t g \circ \varphi_{\delta-s}$ and (x, v) replaced by $(x + vs, v)$ we have

$$[\Phi, Q]f(x, v) = \sum_{i=1}^{d}\left\langle v, \nabla_x\left(\frac{\lambda_i(x, v)}{\lambda(x, v)}\right)\right\rangle \mathcal{P}_t g(x + sv + (\delta - s)F_i v, F_i v)$$

$$+ 2\sum_{i=1}^{d}\left(\frac{\lambda_i(x, v)}{\lambda(x, v)}\right)v_i\partial_{x_i}(\mathcal{P}_t g)(x + sv + (\delta - s)F_i v, F_i v)$$

$$\leq \sum_{i=1}^{d}\left\langle v, \nabla_x\left(\frac{\lambda_i(x, v)}{\lambda(x, v)}\right)\right\rangle |\mathcal{P}_t g(x + sv + (\delta - s)F_i v, F_i v)|$$

$$+ 2\|\nabla_x(\mathcal{P}_t g)(x + sv + (\delta - s)F_i v, F_i v)\|.$$

Observe that

$$\partial_r \left(- \log \left(\phi(\exp(-r)) \right) \right) = \frac{1}{1 + e^{-r}}$$

then we have

$$\nabla_x \left(\frac{\lambda_i(x, v)}{\lambda(x, v)} \right) = \left(\frac{\nabla_x \lambda_i(x, v)}{\lambda(x, v)} \right) - (\nabla_x \lambda(x, v)) \left(\frac{\lambda_i(x, v)}{\lambda(x, v)^2} \right)$$

$$= \left(\frac{v_i \nabla_x \partial_{x_i} \psi(x, v)}{(1 + e^{-v_i \partial_{x_i} \psi(x)}) \lambda(x, v)} \right) - \sum_{j=1}^{d} \left(\frac{v_j \nabla_x \partial_{x_j} \psi(x, v) \lambda_i(x, v)}{(1 + e^{-v_i \partial_{x_i} \psi(x)}) \lambda(x, v)^2} \right)$$

Under our assumptions this is bounded. $\square$

Proof of Theorem 5.13. Fix $f \in \mathcal{G}_1$. We observe that $\mathcal{P}_t f$ satisfies

$$\partial_t \mathcal{P}_t f(x, v) = \mathcal{L} \mathcal{P}_t f(x, v).$$

We can differentiate this with respect to the ith component of x, denoted as x^i, to obtain an equation for $\partial_{x^i} \mathcal{P}_t$.

$$\partial_t \partial_{x^i} \mathcal{P}_t f(x, v) = \mathcal{L} \partial_{x^i} \mathcal{P}_t f + \sum_{j=1}^{d} \partial_{x^i} \lambda_j(x, v)[\mathcal{P}_t f(x, R_j v) - \mathcal{P}_t f(x, v)].$$

We can solve this equation using the variation of constants formula

$$\partial_{x^i} \mathcal{P}_t f(x, v) = \mathcal{P}_t \partial_{x^i} f(x, v) + \sum_{j=1}^{d} \int_0^t \mathcal{P}_{t-s} \left(g_s^{i,j} \right)(x, v) \mathrm{d}s$$

where

$$g_s^{i,j}(x, v) = \partial_{x^i} \lambda_j(x, v)[\mathcal{P}_s f(x, R_j v) - \mathcal{P}_s f(x, v)].$$

We can integrate this with respect to μ to obtain an expression for $\mu(\partial_{x^i} \mathcal{P}_t f)$

$$\mu(\partial_{x^i} \mathcal{P}_t f) = \mu(\partial_{x^i} f) + \sum_{j=1}^{d} \int_0^t \mu \left(g_s^{i,j} \right) \mathrm{d}s.$$

Here we have used that μ is an invariant measure for $\mathcal{P}_t$ to remove $\mathcal{P}_{t-s}$ terms. We shall bound this by comparing to $\mu(\partial_{x^i} \mathcal{P}_t f)$,

$$|\partial_{x^i} \mathcal{P}_t f(x, v) - \mu(\partial_{x^i} \mathcal{P}_t f)| \leq |\mathcal{P}_t \partial_{x^i} f(x, v) - \mu(\partial_{x^i} f)|$$

$$+ |\sum_{j=1}^{d} \int_0^t \mathcal{P}_{t-s} \left(g_s^{i,j} \right)(x, v) - \sum_{j=1}^{d} \int_0^t \mu \left(g_s^{i,j} \right) \mathrm{d}s|. \tag{61}$$

Observe that since λ_i is globally Lipschitz there exists a constant C_λ such that $|\partial_{x^i} \lambda| \leq C_\lambda$ for any $i \in \{1, \dots, d\}, x \in \mathbb{R}^d, v \in \{\pm 1\}^d$. Therefore we can bound $g_s^{i,j}$ by

$$|g_s^{i,j}(x, v)| \leq C_\lambda(|\mathcal{P}_s f(x, R_j v) - \mu(f)| + |\mathcal{P}_s f(x, v) - \mu(f)|).$$

Now we can bound $g_s^{i,j}$ using geometric ergodicity, (11), and that $f \in \mathcal{G}_1$

$$|g_s^{i,j}(x, v)| \leq C C_\lambda e^{-\kappa t} (\overline{G}_{\overline{\alpha}, \epsilon}(x, R_j v) + \overline{G}_{\overline{\alpha}, \epsilon}(x, v)).$$

147

By (25) there exists a constant K such that

$$|g_s^{i,j}(x,v)| \le K e^{-\kappa s} \overline{G}_{\alpha,\epsilon}(x,v).$$

By applying (11) we have

$$|\sum_{j=1}^{d} \int_0^t \mathcal{P}_{t-s}\left(g_s^{i,j}\right)(x,v) - \sum_{j=1}^{d} \int_0^t \mu\left(g_s^{i,j}\right) \mathrm{d}s| \le \sum_{j=1}^{d} \int_0^t |\mathcal{P}_{t-s}\left(g_s^{i,j}\right)(x,v) - \mu\left(g_s^{i,j}\right)|\mathrm{d}s$$

$$\le C K \overline{G}_{\alpha,\epsilon}(x,v) \int_0^t e^{-\kappa t} \mathrm{d}s$$

$$\le C K \overline{G}_{\alpha,\epsilon}(x,v) t e^{-\kappa t}. \tag{62}$$

Since $f \in \mathcal{G}_1$ we can also apply (11) for the function $\partial_i f$

$$|\mathcal{P}_t \partial_{x^i} f(x,v) - \mu(\partial_{x^i} f)| \le C \overline{G}_{\alpha,\epsilon}(x,v) e^{-\kappa t}. \tag{63}$$

Using (62), (63) to bound the right hand side of (61) we obtain

$$|\partial_{x^i} \mathcal{P}_t f(x,v) - \mu(\partial_{x^i} \mathcal{P}_t f)| \le C(1 + Kt) e^{-\kappa t} \overline{G}_{\alpha,\epsilon}(x,v). \tag{64}$$

It remains to consider $\mu(\partial_{x^i} \mathcal{P}_t f)$. Note that we have following integration by parts formula for μ

$$\int \partial_{x_k} g h \mathrm{d}\mu = -\int g \partial_{x_k} h \mathrm{d}\mu + \int \partial_{x_k} \psi g h \mathrm{d}\mu. \tag{65}$$

Setting $k = i$, $g = \mathcal{P}_t f$ and $h = 1$ we have

$$\mu(\partial_{x^i} \mathcal{P}_t f) = \mu(\mathcal{P}_t f \partial_{x^i} \psi)$$

Observe that $\mu(\partial_{x^i} \psi) = 0$ (this follows from (65) with $g = h = 1$ and $k = i$) then subtracting $\mu(\partial_{x^i} \psi)\mu(f)$ we have

$$|\mu(\partial_{x^i} \mathcal{P}_t f)| = |\mu(\partial_{x^i} \mathcal{P}_t f - \mu(f))|$$

$$\le \mu(|\partial_{x^i} \psi(x)||\mathcal{P}_t f - \mu(f)|) \tag{66}$$

$$\le C e^{-\kappa t} \mu(|\partial_{x^i} \psi(x)|\overline{G}_{\overline{\alpha},\epsilon}(x,v)).$$

By (25) we have $|\partial_{x^i} \psi(x)|\overline{G}_{\overline{\alpha},\epsilon}(x,v) \le C \overline{G}_{\alpha,\epsilon}$ and combining (64) and (66) we obtain a constant C' such that

$$|\partial_{x^i} \mathcal{P}_t f(x,v)| \le C'(1 + Kt) e^{-\kappa t} \overline{G}_{\alpha,\epsilon}(x,v). \quad \square$$

Lemma C.2. *Let $\{(\overline{X}_{t_n}, \overline{X}_{t_n})\}_{n \in \mathbb{N}}$ denote the Euler Zig Zag algorithm in 1-d using Algorithm 2 or 3. Let $\overline{\lambda}(x,v;\delta) = \lambda(x,v)$ and $\lambda(x,v) = (\psi'(x)v)_+ + \gamma(x)$ for $\gamma : \mathbb{R} \to [0,\overline{\gamma}]$ with $\overline{\gamma} < \infty$. Assume that $\psi \in \mathcal{C}^2$ is such that (25) is satisfied. Let $\alpha \in (0,1)$, $\beta > 0$ be such that $\alpha < 2\beta$ and define*

$$\overline{G}_{\alpha,\beta}(x,v;\delta) = \begin{cases} \exp\left(\alpha\psi(x) + \beta\delta\psi'(x)v\right), & \text{if } v\psi'(x) \ge 0, \\ \exp\left(\alpha\psi(x) - \beta\delta\psi'(x)v\right), & \text{if } v\psi'(x) < 0. \end{cases} \tag{67}$$

Then there exists a compact set C and $\kappa \in (0,1)$ such that

$$\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}_{t_n}, \overline{V}_{t_n}; \delta) \le \kappa^n \overline{G}_{\alpha,\beta}(x,v;\delta) \quad \text{for all } x \notin C.$$

Proof. Let $\{(\overline{X}^2_\delta, \overline{V}^2_\delta)\}$ ($\{(\overline{X}^3_\delta, \overline{V}^3_\delta)\}$ respectively) be given by Algorithm 2 (Algorithm 3 resp.). To simplify the notation in this proof suppress the δ dependence of $\overline{G}_{\alpha,\beta}$. Set $\beta_\pm = \beta$ if $v\psi'(x) \geq 0$ and $\beta_\pm = -\beta$ otherwise. Observe that

$$\mathbb{E}_{x,v}[\overline{G}_{\alpha,\beta}(\overline{X}^3_\delta, \overline{V}^3_\delta)] - \mathbb{E}_{x,v}[\overline{G}_{\alpha,\beta}(\overline{X}^2_\delta, \overline{V}^2_\delta)] =$$

$$= \int_0^\delta \lambda(x,v)e^{-\lambda(x,v)s}\left(e^{\alpha\psi(x+vs-(\delta-s)v)-\delta\beta_\pm v\psi'(x+sv-(\delta-s)v)} - e^{\alpha\psi(x+\delta v)-\delta\beta_\pm v\psi'(x+\delta v)}\right)ds$$

$$= \int_0^\delta \lambda(x,v)\exp\left(-\lambda(x,v)s + \alpha\psi(x+\delta v) - \delta\beta_\pm v\psi'(x+\delta v)\right)\left(e^{I(x,v,s;\delta)} - 1\right)ds,$$

where

$$I(x,v,s;\delta) := \alpha\psi(x+vs-(\delta-s)v) - \alpha\psi(x+\delta v)$$
$$- \delta\beta_\pm v\psi'(x+sv-(\delta-s)v) + \delta\beta_\pm v\psi'(x+\delta v).$$

By Taylor's theorem we can find ξ_1, ξ_2

$$I(x,v,s;\delta) = \alpha 2(s-\delta)v\psi'(x+\delta v) + 2\alpha(\delta-s)^2\psi''(\xi_1) + 2\beta_\pm\delta(s-\delta)v\psi''(\xi_2).$$

By taking x sufficiently large we can ensure that the sign of $I(x,v,s;\delta)$ is equal to the sign of $-v\psi'(x)$. Therefore,

$$\mathbb{E}_{x,v}[\overline{G}_{\alpha,\beta}(\overline{X}^3_\delta, \overline{V}^3_\delta)] \leq \mathbb{E}_{x,v}[\overline{G}_{\alpha,\beta}(\overline{X}^2_\delta, \overline{V}^2_\delta)] \quad \text{if } v\psi'(x) > 0,$$

$$\mathbb{E}_{x,v}[\overline{G}_{\alpha,\beta}(\overline{X}^3_\delta, \overline{V}^3_\delta)] \geq \mathbb{E}_{x,v}[\overline{G}_{\alpha,\beta}(\overline{X}^2_\delta, \overline{V}^2_\delta)] \quad \text{if } v\psi'(x) < 0.$$

In the first case it is sufficient to consider Algorithm 2, while in the latter it is sufficient to consider Algorithm 3. We shall consider these two cases separately.

Case $v\psi'(x) > 0$: Note that it is sufficient to show that outside of a sufficiently large compact set

$$\frac{\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}^2_\delta, \overline{V}^2_\delta)}{\overline{G}_{\alpha,\beta}(x,v)} < 1.$$

We can expand $\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}^2_\delta, \overline{V}^2_\delta)$ as

$$\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}^2_\delta, \overline{V}^2_\delta) = e^{-\delta\lambda(x,v)}\overline{G}_{\alpha,\beta}(x+v\delta, v) + (1 - e^{-\delta\lambda(x,v)})\overline{G}_{\alpha,\beta}(x+v\delta, -v).$$

Using the definition of $\overline{G}_{\alpha,\beta}$ we can write

$$\frac{\overline{G}_{\alpha,\beta}(x+v\delta, v)}{\overline{G}_{\alpha,\beta}(x,v)} = \exp\left(\alpha(\psi(x+v\delta) - \psi(x)) + \beta\delta v(\psi'(x+v\delta) - \psi'(x))\right),$$

$$\frac{\overline{G}_{\alpha,\beta}(x+v\delta, -v)}{\overline{G}_{\alpha,\beta}(x,v)} = \exp\left(\alpha(\psi(x+v\delta) - \psi(x)) - \beta\delta v(\psi'(x+v\delta) + \psi'(x))\right).$$

We can Taylor expand U to find some z_1, z_2, z_3 such that

$$\frac{\overline{G}_{\alpha,\beta}(x+v\delta, v)}{\overline{G}_{\alpha,\beta}(x,v)} = \exp\left(\alpha(\psi'(x)v\delta + \frac{1}{2}\psi''(z_1)\delta^2) + \beta\delta^2\psi''(z_2)\right),$$

$$\frac{\overline{G}_{\alpha,\beta}(x+v\delta, -v)}{\overline{G}_{\alpha,\beta}(x,v)} = \exp\left(\alpha(\psi'(x)v\delta + \frac{1}{2}\psi''(z_1)\delta^2) - \beta\delta(2\psi'(x)v + \psi''(z_3)\delta)\right).$$

Thus we have

$$
\frac{\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}_\delta^2, \overline{V}_\delta^2)}{\overline{G}_{\alpha,\beta}(x, v)} = e^{-\delta\lambda(x,v)}\exp\left(\alpha(\psi'(x)v\delta + \frac{1}{2}\psi''(z_1)\delta^2) + \beta\delta^2\psi''(z_2)\right)
$$
$$
+ (1 - e^{-\delta\lambda(x,v)})\exp\left(\alpha(\psi'(x)v\delta + \frac{1}{2}\psi''(z_1)\delta^2) - \beta\delta(2\psi'(x)v + \psi''(z_3)\delta)\right).
$$

Rearranging we can rewrite this as

$$
\frac{\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}_\delta^2, \overline{V}_\delta^2)}{\overline{G}_{\alpha,\beta}(x, v)} =
$$
$$
= \exp\left(-\delta\lambda(x, v) + \alpha\psi'(x)v\delta + \frac{\alpha}{2}\psi''(z_1)\delta^2\right)\left(e^{\left(\beta\delta^2\psi''(z_2)\right)} - e^{\left(-\beta\delta(2\psi'(x)v + \psi''(z_3)\delta)\right)}\right) \quad (68)
$$
$$
+ \exp\left((\alpha - 2\beta)\psi'(x)v\delta + \frac{1}{2}\alpha\psi''(z_1)\delta^2 - \beta\delta^2\psi''(z_3)\right). \quad (69)
$$

Recall that in this case $\lambda(x, v) \geq v\psi'(x) > 0$. Thus for the first term (68)

$$
\exp\left(-\delta\lambda(x, v) + \alpha\psi'(x)v\delta + \frac{\alpha}{2}\psi''(z_1)\delta^2\right)\left(e^{\left(\beta\delta^2\psi''(z_2)\right)} - e^{\left(-\beta\delta(2\psi'(x)v + \psi''(z_3)\delta)\right)}\right) \leq
$$
$$
\leq \exp\left(-\delta\lambda(x, v) + \alpha\psi'(x)v\delta + \frac{\alpha}{2}\psi''(z_1)\delta^2\right)e^{\left(\beta\delta^2\psi''(z_2)\right)}
$$
$$
\leq \exp\left(-(1 - \alpha)\delta v\psi'(x) + \frac{\alpha}{2}\psi''(z_1)\delta^2\right)e^{\left(\beta\delta^2\psi''(z_2)\right)}.
$$

Now choose $0 < \alpha < \min\{1, 2\beta\}$ and recall that by assumption ψ' diverges to infinity faster than ψ''. It follows that, outside of a large enough compact set, both (68) and (69) can be made arbitrarily small.

Case $v\psi'(x) < 0$: In this case $\lambda(x, v) = \gamma(x)$. We expand $\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}_\delta^2, \overline{V}_\delta^2)$ as

$$
\frac{\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}_\delta^2, \overline{V}_\delta^2)}{\overline{G}_{\alpha,\beta}(x, v)} = e^{-\delta\lambda(x,v)}\frac{\overline{G}_{\alpha,\beta}(x + v\delta, v)}{\overline{G}_{\alpha,\beta}(x, v)}
$$
$$
+ \int_0^\delta \lambda(x, v)e^{-s\lambda(x,v)}\frac{\overline{G}_{\alpha,\beta}(x + v(2s - \delta), -v)}{\overline{G}_{\alpha,\beta}(x, v)}\,ds.
$$

Similarly to above, we can use Taylor's theorem to find z_1, z_2, z_3, z_4 with

$$
\frac{\mathbb{E}_{x,v}\overline{G}_{\alpha,\beta}(\overline{X}_\delta^2, \overline{V}_\delta^2)}{\overline{G}_{\alpha,\beta}(x, v)} = e^{-\delta\lambda(x,v)}\exp\left(\alpha(\psi'(x)v\delta + \frac{1}{2}\psi''(z_1)\delta^2) - \beta\delta^2\psi''(z_2)\right)
$$
$$
+ \int_0^\delta \lambda(x, v)e^{-s\lambda(x,v)}\exp\left(\alpha(\psi'(x)v(2s - \delta) + \frac{1}{2}\psi''(z_3)(2s - \delta)^2) \right. \quad (70)
$$
$$
\left. + \beta\delta(2\psi'(x)v + \psi''(z_4)(2s - \delta))\right)\,ds.
$$

Taking advantage of $-\delta \leq 2s - \delta \leq \delta$ we obtain the bound

$$
\exp\left(\alpha(\psi'(x)v(2s - \delta)) + \beta\delta(2\psi'(x)v)\right) \leq \exp\left((-\psi'(x)v)(\alpha - 2\beta)\delta\right).
$$

Using this bound together with the assumption that $-v\psi'(x)$ diverges to $+\infty$ faster than ψ'', we obtain that for $0 < \alpha < 2\beta$ the right hand side of (70) can be made arbitrarily small for sufficiently large values of x.

Combining the two cases above we obtain the statement of the lemma. $\quad\square$

Lemma C.3. *Assume that $\psi \in \mathcal{C}^2$ satisfies (25). Let $\overline{G}_{\alpha,\beta}(x, v; \delta)$ be given by (67) and $\overline{G}_{\alpha,\epsilon}(x, v)$ be given by (26). Then for any $0 < \alpha_1 < \overline{\alpha} < \alpha_2 < 1$ there exist positive constants $C, C' > 0$ with*

$$\overline{G}_{\alpha_1,\epsilon}(x, v) \leq C' e^{\overline{\alpha}\psi(x)} \leq C\overline{G}_{\alpha_2,\beta}(x, v; \delta). \tag{71}$$

Proof of Lemma C.3. Let us first consider $\overline{G}_{\alpha_1,\epsilon}(x, v)$, since $|\phi_\epsilon(s)| \leq \epsilon|s|/2$ we have

$$\overline{G}_{\alpha_1,\epsilon}(x, v) \leq \exp\left(\alpha_1\psi(x) + \frac{\epsilon}{2}|\psi'(x)|\right).$$

By (25) there exists $R > 0$ such that for any $|x| > R$ we have $|\psi'(x)| \leq 2\epsilon^{-1}(\overline{\alpha} - \alpha_1)\psi(x)$. Therefore for $|x| > R$ we have

$$\overline{G}_{\alpha_1,\epsilon}(x, v) \leq \exp\left(\overline{\alpha}\psi(x)\right).$$

Setting $C' = \exp(\sup_{|x|\leq R}|\psi'(x)|)$ we have the left hand side of (71).

Similarly, we have

$$\overline{G}_{\alpha_2,\beta}(x, v; \delta) \geq \exp\left(\alpha_2\psi(x) - \beta\delta_0|\psi'(x)|\right).$$

Using (25) for x sufficiently large we have that $\beta\delta_0|\psi'(x)| \leq (\alpha_2 - \overline{\alpha})\psi(x)$ and hence the right hand side of (71) follows. $\quad\square$

References

[1] A.B. Abdessalem, R. Azaïs, M. Touzet-Cortina, A. Gégout-Petit, M. Puiggali, Stochastic modelling and prediction of fatigue crack propagation using piecewise-deterministic Markov processes, Proc. Inst. Mech. Eng. O 230 (4) (2016) 405–416.

[2] C. Andrieu, A. Durmus, N. Nüsken, J. Roussel, Hypocoercivity of piecewise deterministic Markov process-Monte Carlo, 2019, arXiv:1808.08592.

[3] C. Andrieu, S. Livingstone, Peskun-tierney ordering for Markov chain and process Monte Carlo: beyond the reversible scenario, 2019, arXiv:1906.06197.

[4] M. Benaïm, S. Le Borgne, F. Malrieu, P.-A. Zitt, Qualitative properties of certain piecewise deterministic Markov processes, Annales de L'I.H.P. Probabilités Et Statistiques 51 (3) (2015) 1040–1075.

[5] H.C. Berg, D.A. Brown, Chemotaxis in Escherichia coli analysed by three-dimensional tracking, Nature 239 (5374) (1972) 500–504.

[6] A. Bertazzi, J. Bierkens, Adaptive schemes for piecewise deterministic Monte Carlo algorithms, Bernoulli 28 (4) (2022) 2404 – 2430.

[7] J. Bierkens, P. Fearnhead, G. Roberts, The zig-zag process and super-efficient sampling for Bayesian analysis of big data, Ann. Statist. 47 (2019a).

[8] J. Bierkens, K. Kamatani, G.O. Roberts, High-dimensional scaling limits of piecewise deterministic sampling algorithms, 2018, arXiv:1807.11358.

[9] J. Bierkens, G. Roberts, A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model, Ann. Appl. Probab. 27 (2) (2017) 846–882.

[10] J. Bierkens, G.O. Roberts, P.-A. Zitt, Ergodicity of the zigzag process, Ann. Appl. Probab. 29 (4) (2019b) 2266–2301.

[11] N. Bou-Rabee, A. Eberle, R. Zimmer, Coupling and convergence for Hamiltonian Monte Carlo, Ann. Appl. Probab. 30 (3) (2020).

[12] N. Bou-Rabee, J.M. Sanz-Serna, Randomized hamiltonian monte carlo, Ann. Appl. Probab. 27 (4) (2017) 2159–2194.

[13] A. Bouchard-Côté, S.J. Vollmer, A. Doucet, The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method, J. Amer. Statist. Assoc. 113 (522) (2018) 855–867.

[14] B. Cloez, R. Dessalles, A. Genadot, F. Malrieu, A. Marguet, R. Yvinec, Probabilistic and piecewise deterministic models in biology, ESAIM: Procs 60 (2017) 225–245.

[15] C. Cocozza-Thivent, Markov Renewal and Piecewise Deterministic Processes, Springer, 2021.

[16] O.L.V. Costa, F. Dufour, Stability and ergodicity of piecewise deterministic Markov processes, SIAM J. Control Optim. 47 (2) (2008) 1053–1077.

[17] S. Cotter, T. House, F. Pagani, The nuzz: Numerical ZigZag sampling for general models, 2020, arXiv:2003 .03636.

[18] D. Crisan, P. Dobson, M. Ottobre, Uniform in time estimates for the weak error of the Euler method for SDEs and a pathwise approach to derivative estimates for diffusion semigroups, Trans. Amer. Math. Soc. 374 (5) (2021) 3289–3330.

[19] D. Crisan, M. Ottobre, Pointwise gradient bounds for degenerate semigroups (of UFG type), in: Proc. R. Soc. A, Vol. 472, The Royal Society, 2016, 20160442.

[20] A. Dassios, P. Embrechts, Martingales and insurance risk, Commun. Stat. Stoch. Model. 5 (2) (1989) 181–217.

[21] M.H.A. Davis, Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models, J. R. Stat. Soc. Ser. B Stat. Methodol. 46 (3) (1984) 353–388.

[22] M. Davis, Markov Models & Optimization, in: Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1993.

[23] G. Deligiannidis, A. Bouchard-Côté, A. Doucet, Exponential ergodicity of the bouncy particle sampler, Ann. Statist. 47 (3) (2019) 1268–1287.

[24] L. Devroye, Non-Uniform Random Variate Generation, Springer-Verlag, 1986.

[25] A. Durmus, A. Guillin, P. Monmarché, Geometric ergodicity of the bouncy particle sampler, Ann. Appl. Probab. 30 (5) (2020) 2069–2098.

[26] A. Durmus, A. Guillin, P. Monmarché, Piecewise deterministic Markov processes and their invariant measures, Annales de L'Institut Henri Poincaré, Probabilités Et Statistiques 57 (3) (2021) 1442–1475.

[27] A. Durmus, É. Moulines, Nonasymptotic convergence analysis for the unadjusted Langevin algorithm, Ann. Appl. Probab. 27 (3) (2017) 1551–1587.

[28] A. Durmus, É. Moulines, High-dimensional Bayesian inference via the unadjusted Langevin algorithm, Bernoulli 25 (4A) (2019) 2854–2882.

[29] P. Embrechts, H. Schmidli, Ruin estimation for a general insurance risk model, Adv. Appl. Probab. 26 (2) (1994) 404–422.

[30] P. Fearnhead, J. Bierkens, M. Pollock, G.O. Roberts, Piecewise deterministic Markov processes for continuous-time Monte Carlo, Statist. Sci. 33 (3) (2018) 386–412.

[31] A. Guillin, B. Nectoux, Low lying eigenvalues and convergence to the equilibrium of some piecewise deterministic Markov processes generators in the small temperature regime, 2020, Hal-02436593, working paper or preprint.

[32] E. Horton, A.E. Kyprianou, D. Villemonais, Stochastic methods for the neutron transport equation I: Linear semigroup asymptotics, Ann. Appl. Probab. 30 (6) (2020) 2573–2612.

[33] J. Huggins, J. Zou, Quantifying the accuracy of approximate diffusions and Markov chains, in: Artificial Intelligence and Statistics, PMLR, 2017, pp. 382–391.

[34] P.E. Kloeden, E. Platen, Stochastic differential equations, in: Numerical Solution of Stochastic Differential Equations, Springer, 1992, pp. 103–160.

[35] V. Lemaire, M. Thieullen, N. Thomas, Thinning and multilevel Monte Carlo methods for piecewise deterministic (Markov) processes with an application to a stochastic Morris–Lecar model, Adv. Appl. Probab. 52 (1) (2020) 138–172.

[36] P. Lewis, G. Shedler, Simulation of nonhomogeneous Poisson processes by thinning, 1978.

[37] J. Lu, L. Wang, On explicit L^2-convergence rate estimate for piecewise deterministic Markov processes, 2020, arXiv:2007.14927.

[38] M. Michel, S.C. Kapfer, W. Krauth, Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps, J. Chem. Phys. 140 (5) (2014).

[39] C. Morris, H. Lecar, Voltage oscillations in the barnacle giant muscle fiber, Biophys. J. 35 (1) (1981) 193–213.

[40] K. Pakdaman, M. Thieullen, G. Wainrib, Fluid limit theorems for stochastic hybrid systems with application to neuron models, Adv. Appl. Probab. 42 (2010).

[41] E.A.J.F. Peters, G. De With, Rejection-free Monte Carlo sampling for general potentials, Phys. Rev. E 85 (2) (2012) 1–5.

[42] R. Rudnicki, M. Tyran-Kamińska, Piecewise Deterministic Processes in Biological Models, Vol. 1, Springer, 2017.

[43] J.M. Sanz-Serna, K.C. Zygalakis, Wasserstein distance estimates for the distributions of numerical approximations to ergodic stochastic differential equations, J. Mach. Learn. Res. (21) (2021) 1–37.

[44] C. Sherlock, A.H. Thiery, A discrete bouncy particle sampler, 2021, arXiv:1707.05200.

[45] A. Singh, J.P. Hespanha, Stochastic hybrid systems for studying biochemical processes, Phil. Trans. R. Soc. A 368 (1930) (2010) 4995–5011.

[46] D.W. Stroock, Some stochastic processes which arise from a model of the motion of a bacterium, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 28 (4) (1974) 305–315.

[47] K.S. Turitsyn, M. Chertkov, M. Vucelja, Irreversible Monte Carlo algorithms for efficient sampling, Physica D 240 (4–5) (2011) 410–414.

[48] P. Vanetti, A. Bouchard-Côté, G. Deligiannidis, A. Doucet, Piecewise-deterministic Markov chain Monte Carlo, 2017, arXiv:1707.05296.