Adv. Appl. Prob. **52,** 138–172 (2020)
*doi:*10.1017/apr.2019.55

THINNING AND MULTILEVEL MONTE CARLO METHODS FOR PIECEWISE DETERMINISTIC (MARKOV) PROCESSES WITH AN APPLICATION TO A STOCHASTIC MORRIS–LECAR MODEL

VINCENT LEMAIRE,*
MICHÈLE THIEULLEN,* AND
NICOLAS THOMAS,* *Sorbonne University*

Abstract

In the first part of this paper we study approximations of trajectories of piecewise deterministic processes (PDPs) when the flow is not given explicitly by the thinning method. We also establish a strong error estimate for PDPs as well as a weak error expansion for piecewise deterministic Markov processes (PDMPs). These estimates are the building blocks of the multilevel Monte Carlo (MLMC) method, which we study in the second part. The coupling required by the MLMC is based on the thinning procedure. In the third part we apply these results to a two-dimensional Morris–Lecar model with stochastic ion channels. In the range of our simulations the MLMC estimator outperforms classical Monte Carlo.

Keywords: Piecewise deterministic (Markov) processes; multilevel Monte Carlo; thinning; strong error estimate; weak error expansion; Morris–Lecar model

2010 Mathematics Subject Classification: Primary 65C05, 65C20
Secondary 60G55, 60J25, 68U20

1. Introduction

In this paper we are interested in the approximation of the trajectories of piecewise deterministic processes (PDPs). We establish strong error estimates for a PDP and a weak error expansion for a piecewise deterministic Markov process (PDMP). Then we study the application of the multilevel Monte Carlo (MLMC) method in order to approximate expectations of functionals of PDMPs. Our motivation comes from neuroscience, where the whole class of stochastic conductance-based neuron models can be interpreted as PDMPs. The response of a neuron to a stimulus, called neural encoding, is considered as relevant information to understand the functional properties of such excitable cells. Thus many quantities of interest, such as mean first spike latency, mean interspike intervals, and mean firing rate, can be modelled as expectations of functionals of PDMPs.

PDPs were introduced by Davis [5] as a general class of stochastic processes characterized by a deterministic evolution between two successive random times. In the case where the deterministic evolution part follows a family of ordinary differential equations (ODEs), the

Received 20 December 2018; revision received 25 October 2019.
* Postal address: Laboratory of Probability, Statistics and Modeling (LPSM), UMR CNRS 8001, Sorbonne University–Campus Pierre et Marie Curie, Case 158, 4 place Jussieu, F-75252 Paris Cedex 5, France. Email addresses: vincent.lemaire@sorbonne-universite.fr, michele.thieullen@upmc.fr, nicolas.thomas@upmc.fr

corresponding PDP enjoys the Markov property and is called a PDMP. The distribution of a PDMP is thus determined by three parameters called the characteristics of the PDMP: a family of vector fields, a jump rate (intensity function), and a transition measure.

We first consider a general PDP (x_t) that is not necessarily Markov on a finite time interval $[0, T]$ for which the flow is not explicitly solvable. Approximating its flows by the classical Euler scheme and using our previous work [22], we build a thinning algorithm which provides us with an exact simulation of an approximation of (x_t), which we denote by $(\bar{x}_t)$. The process $(\bar{x}_t)$ is a PDP constructed by thinning of a homogeneous Poisson process which enjoys explicitly solvable flows.

In fact this thinning construction provides a whole family of approximations indexed by the time step $h > 0$ of the Euler scheme. We prove that for any real-valued smooth function F the following strong estimate holds:

$$\text{there exist } V_1 > 0,\ V_2 > 0 \text{ such that } \mathbb{E}[|F(\bar{x}_T) - F(x_T)|^2] \le V_1 h + V_2 h^2. \tag{1.1}$$

Moreover, if (x_t) is a PDMP the following weak error expansion holds:

$$\text{there exists } c_1 > 0 \text{ such that } \mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = c_1 h + \mathrm{o}(h^2). \tag{1.2}$$

The estimate (1.1) is mainly based on the construction of the pair $(x_t, \bar{x}_t)$ and on the fact that the Euler scheme is of order 1; this is why it is valid for a general PDP and its Euler scheme. In contrast, the estimate (1.2) relies on properties which are specific to PDMPs, such as the Feynman–Kac formula.

The MLMC method relies simultaneously on estimates (1.1) and (1.2); this is why we study its application to the PDMP framework instead of the more general PDP framework. MLMC extends the classical Monte Carlo (MC) method, which is a very general approach to estimating expectations using stochastic simulations. The complexity (i.e. the number of operations necessary in the simulation) associated with an MC estimation can be prohibitive, especially when the complexity of an individual random sample is very high. MLMC relies on repeated independent random samplings taken on different levels of accuracy, which differs from classical MC. MLMC can then greatly reduce the complexity of classical MC by performing most simulations with low accuracy but with low complexity, and only a few simulations with high accuracy at high complexity. MLMC was introduced by Heinrich [18] and developed by Giles [11]. The MLMC estimator has been efficiently used in various fields of numerical probability such as SDEs [11], Markov chains [1, 2, 14], Lévy processes [10], jump diffusions [7, 8, 28], and nested Monte Carlo [13, 21]. See [12] for more references. To the best of our knowledge, application of MLMC to PDMPs has not been considered.

For the sake of clarity, we will describe the general improvement of MLMC. We are interested in the estimation of $\mathbb{E}[X]$, where X is a real-valued square-integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. When X can be simulated exactly, the classical MC estimator $(1/N)\sum_{k=1}^{N} X^k$ with X^k, $k \ge 1$ independent random variables identically distributed as X provides an unbiased estimator. The associated L^2-error satisfies $\|Y - \mathbb{E}[X]\|_2^2 = \mathrm{Var}(Y) = \frac{1}{N}\mathrm{Var}(X)$. If we quantify the precision by the L^2-error, then a user-prescribed precision $\epsilon^2 > 0$ is achieved for $N = \mathrm{O}(\epsilon^{-2})$, so that in this case the global complexity is of order $\mathrm{O}(\epsilon^{-2})$.

Assume now that X cannot be simulated exactly (or cannot be simulated at a reasonable cost) and that we can build a family of real-valued random variables $(X_h, h > 0)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ which converges weakly and strongly to X as $h \to 0$ in the following sense:

$$\text{there exist } c_1 > 0,\ \alpha > 0 \text{ such that } \mathbb{E}[X_h] - \mathbb{E}[X] = c_1 h^{\alpha} + \mathrm{o}(h^{2\alpha}), \tag{1.3}$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

and

$$\text{there exist } V_1 > 0,\ \beta > 0 \text{ such that } \mathbb{E}[|X_h - X|^2] \le V_1 h^\beta. \tag{1.4}$$

Assume, moreover, that for $h > 0$ the random variable X_h can be simulated at reasonable complexity (the complexity increases as $h \to 0$). The classical MC estimator now consists of a sequence of random variables,

$$Y = \frac{1}{N} \sum_{k=1}^{N} X_h^k, \tag{1.5}$$

where X_h^k, $k \ge 1$ are independent random variables identically distributed as X_h. The bias and the variance of the estimator (1.5) are given by $\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[X_h] - \mathbb{E}[X] \simeq c_1 h^\alpha$ and $\mathrm{Var}(Y) = \frac{1}{N}\mathrm{Var}(X_h)$ respectively. From the strong estimate (1.4) we have $\mathrm{Var}(X_h) \to \mathrm{Var}(X)$ as $h \to 0$, so that $\mathrm{Var}(X_h)$ is asymptotically a constant independent of h. If as above we quantify the precision by the L^2-error and use $\|Y - \mathbb{E}[X]\|_2^2 = (\mathbb{E}[Y] - \mathbb{E}[X])^2 + \mathrm{Var}(Y)$, we obtain that the estimator (1.5) achieves a user-prescribed precision $\epsilon^2 > 0$ for $h = \mathrm{O}(\epsilon^{1/\alpha})$ and $N = \mathrm{O}(\epsilon^{-2})$, so that the global complexity of the estimator is now $\mathrm{O}(\epsilon^{-2-1/\alpha})$.

The MLMC method takes advantage of the estimate (1.4) in order to reduce the global complexity. Let us fix $L \ge 2$ and consider for $l \in \{1, \dots, L\}$ a geometrically decreasing sequence $(h_l,\ 1 \le l \le L)$, where $h_l = h^* M^{-(l-1)}$ for fixed $h^* > 0$ and $M > 1$. The indexes l are called the levels of the MLMC and the complexity of X_{h_l} increases as the level increases. Thanks to the weak expansion (1.3), the quantity $\mathbb{E}[X_{h_L}]$ approximates $\mathbb{E}[X]$. Using the linearity of the expectation, the quantity $\mathbb{E}[X_{h_L}]$ can be decomposed over the levels $l \in \{1, \dots, L\}$ as follows:

$$\mathbb{E}[X_{h_L}] = \mathbb{E}[X_{h^*}] + \sum_{l=2}^{L} \mathbb{E}[X_{h_l} - X_{h_{l-1}}]. \tag{1.6}$$

For each level $l \in \{1, \dots, L\}$, a classical MC estimator is used to approximate $\mathbb{E}[X_{h_l} - X_{h_{l-1}}]$ and $\mathbb{E}[X_{h^*}]$. At each level, a number $N_l \ge 1$ of samples are required and the key point is that the random variables X_{h_l} and $X_{h_{l-1}}$ are assumed to be correlated in order to make the variance of $X_{h_l} - X_{h_{l-1}}$ small. Considering at each level $l = 2, \dots, L$ independent pairs $(X_{h_l}, X_{h_{l-1}})$ of correlated random variables, the MLMC estimator then reads

$$Y = \frac{1}{N_1} \sum_{k=1}^{N_1} X_{h^*}^k + \sum_{l=2}^{L} \frac{1}{N_l} \sum_{k=1}^{N_l} (X_{h_l}^k - X_{h_{l-1}}^k), \tag{1.7}$$

where $(X_{h^*}^k,\ k \ge 1)$ is a sequence of independent and identically distributed random variables distributed as X_{h^*} and $((X_{h_l}^k, X_{h_{l-1}}^k),\ k \ge 1)$ for $l = 2, \dots, L$ are independent sequences of independent copies of $(X_{h_l}, X_{h_{l-1}})$ and independent of $(X_{h^*}^k)$. It is known (see [11] or [21]) that given a precision $\epsilon > 0$ and provided that the family $(X_h,\ h > 0)$ satisfies the strong and weak error estimates (1.4) and (1.3), the multilevel estimator (1.7) achieves a precision $\|Y - \mathbb{E}[X]\|_2^2 = \epsilon^2$ with a global complexity of order $\mathrm{O}(\epsilon^{-2})$ if $\beta > 1$, $\mathrm{O}(\epsilon^{-2}(\log(\epsilon))^2)$ if $\beta = 1$, and $\mathrm{O}(\epsilon^{-2-(1-\beta)/\alpha})$ if $\beta < 1$. This complexity result shows the importance of the parameter β. Finally, let us mention that in the case $\beta > 1$ it is possible to build an unbiased multilevel estimator: see [15].

Estimates (1.1) and (1.2) suggest investigating the use of the MLMC method in the PDMP framework with $\beta = 1$ and $\alpha = 1$. Letting $X = F(x_T)$ and $X_h = F(\bar{x}_T)$ for $h > 0$ and F a smooth function, we define an MLMC estimator of $\mathbb{E}[F(x_T)]$ just as in (1.7) (denoted here by Y^{MLMC})

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

where the processes involved at the level l are correlated by thinning. Since these processes are constructed using two different time steps, the probability of accepting a proposed jump time differs from one process to the other. Moreover, the discrete components of the post-jump locations may also be different. This results in the presence of the term $V_1 h$ in the estimate (1.1). In order to improve the convergence rate (to increase the parameter β) in (1.1), we show that for a given PDMP (x_t) we have the following auxiliary representation:

$$\mathbb{E}[F(x_T)] = \mathbb{E}[F(\tilde{x}_T)\tilde{R}_T]. \tag{1.8}$$

The PDMP $(\tilde{x}_t)$ and its Euler scheme are such that their discrete components jump at the same times and in the same state. $(\tilde{R}_t)$ is a process that depends on $(\tilde{x}_t, t \in [0, T])$. The representation (1.8) is inspired by the change of probability introduced in [28] and is actually valid for a general PDP (Proposition 2.2) so that $\mathbb{E}[F(\bar{x}_T)] = \mathbb{E}[F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T]$, where $(\underline{\tilde{x}}_t)$ is the Euler scheme corresponding to $(\tilde{x}_t)$ and $(\underline{\tilde{R}}_t)$ is a process that depends on $(\underline{\tilde{x}}_t, t \in [0, T])$. Letting $X = F(\tilde{x}_T)\tilde{R}_T$ and $X_h = F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T$, we define a second MLMC estimator (denoted by $\tilde{Y}^{\mathrm{MLMC}}$) where now the discrete components of the Euler schemes $(\underline{\tilde{x}}_t)$ involved at the level l always jump in the same states and at the same times. To sum up, the first MLMC estimator we consider (Y^{MLMC}) derives from (1.6), where the corrective term at level l is $\mathbb{E}[F(\bar{x}_T^{h_l}) - F(\bar{x}_T^{h_{l-1}})]$, whereas the corrective term of the second estimator ($\tilde{Y}^{\mathrm{MLMC}}$) is $\mathbb{E}[F(\underline{\tilde{x}}_T^{h_l})\underline{\tilde{R}}_T^{h_l} - F(\underline{\tilde{x}}_T^{h_{l-1}})\underline{\tilde{R}}_T^{h_{l-1}}]$. For readability, we no longer write the dependence of the approximations on the time step. For the processes $(F(\underline{\tilde{x}}_t)\underline{\tilde{R}}_t)$ and $(F(\tilde{x}_t)\tilde{R}_t)$ we show the strong estimate

$$\text{there exists } \tilde{V}_1 > 0 \text{ such that } \mathbb{E}[|F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T - F(\tilde{x}_T)\tilde{R}_T|^2] \le \tilde{V}_1 h^2,$$

so that we end up with $\beta = 2$ and the complexity goes from a $\mathrm{O}(\epsilon^{-2}(\log(\epsilon))^2)$ to a $\mathrm{O}(\epsilon^{-2})$.

As an application we consider the PDMP version of the two-dimensional Morris–Lecar model (see [25]), which takes into account the precise description of the ionic channels and in which the flows are not explicit. Let us mention [3] for the application of quantitative bounds for the long-time behaviour of PDMPs to a stochastic three-dimensional Morris–Lecar model. The original deterministic Morris–Lecar model was introduced in [23] to account for various oscillating states in the barnacle giant muscle fibre. Because of its low dimension, this model is among the favourite conductance-based models in computational neuroscience. Furthermore, this model is particularly interesting because it reproduces some of the main features of excitable cell response, such as the shape, amplitude, and threshold of the action potential, the refractory period. We compare the classical MC and the MLMC estimators on the two-dimensional stochastic Morris–Lecar model to estimate the mean value of the membrane potential at fixed time. It turns out that in the range of our simulations the MLMC estimator outperforms the MC one. It suggests that MLMC estimators can be used successfully in the framework of PDMPs.

As mentioned above, the quantities of interest, such as mean first spike latency, mean interspike intervals, and mean firing rate, can be modelled as expectations of path-dependent functionals of PDMPs. This setting can then be considered as a natural extension of this work.

The paper is organized as follows. In Section 2 we construct a general PDP by thinning, and we give a representation of its distribution in terms of the thinning data (Proposition 2.1). In Section 3 we establish strong error estimates (Theorems 3.1–3.2). In Section 4 we establish a weak error expansion (Theorem 4.1). In Section 5 we compare the efficiency of the classical and multilevel Monte Carlo estimators on the two-dimensional stochastic Morris–Lecar model.

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

2. Piecewise deterministic process by thinning

2.1. Construction

In this subsection we introduce the setting and recall some results on the thinning method from our previous paper [22]. Let $E := \Theta \times \mathbb{R}^d$ where Θ is a finite or countable set and $d \geq 1$. A piecewise deterministic process (PDP) is defined via the following characteristics:

- a family of functions $(\Phi_\theta, \theta \in \Theta)$ such that $\Phi_\theta : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ for all $\theta \in \Theta$,
- a measurable function $\lambda : E \to]0, +\infty[$,
- a transition measure $Q : E \times \mathcal{B}(E) \to [0, 1]$.

We let $x = (\theta, \nu)$ denote a generic element of E. We only consider PDPs with continuous ν-component, so for $A \in \mathcal{B}(\Theta)$ and $B \in \mathcal{B}(\mathbb{R}^d)$ we write

$$Q(x, A \times B) = Q(x, A)\delta_\nu(B). \tag{2.1}$$

If we write $x = (\theta_x, \nu_x)$, then it holds that

$$Q((\theta_x, \Phi_{\theta_x}(t, \nu_x)),\ \mathrm{d}\theta\, \mathrm{d}\nu) = Q((\theta_x, \Phi_{\theta_x}(t, \nu_x)),\ \mathrm{d}\theta)\delta_{\Phi_{\theta_x}(t,\nu_x)}(\mathrm{d}\nu).$$

Our results do not depend on the dimension of the variable in $\mathbb{R}^d$ so we restrict ourselves to $\mathbb{R}$ $(d = 1)$ for readability. We work under the following assumption.

Assumption 2.1. *There exists $\lambda^* < +\infty$ such that, for all $x \in E$, $\lambda(x) \leq \lambda^*$.*

In [22] we considered a general upper bound λ^*. In the present paper λ^* is a constant (see Assumption 2.1). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which we define the following.

(1) A homogeneous Poisson process $(N_t^*, t \geq 0)$ with intensity λ^* (given in Assumption 2.1) whose successive jump times are denoted $(T_k^*, k \geq 1)$. We set $T_0^* = 0$.

(2) Two sequences of i.i.d. random variables with uniform distribution on $[0, 1]$, $(U_k, k \geq 1)$ and $(V_k, k \geq 1)$ independent of each other and independent of $(T_k^*, k \geq 1)$.

Given $T > 0$ we construct iteratively the sequence of jump times and post-jump locations $(T_n, (\theta_n, \nu_n), n \geq 0)$ of the E-valued PDP $(x_t, t \in [0, T])$ that we want to obtain in the end using its characteristics (Φ, λ, Q). Let $(\theta_0, \nu_0) \in E$ be fixed and let $T_0 = 0$. We construct T_1 by thinning of (T_k^*), that is,

$$T_1 := T_{\tau_1}^*, \tag{2.2}$$

where

$$\tau_1 := \inf\{k > 0 : U_k\lambda^* \leq \lambda(\theta_0, \Phi_{\theta_0}(T_k^*, \nu_0))\}. \tag{2.3}$$

We let $|\Theta|$ denote the cardinal of Θ (which may be infinite) and we set $\Theta = \{k_1, \ldots, k_{|\Theta|}\}$. For $j \in \{1, \ldots, |\Theta|\}$ we introduce the functions a_j defined on E by

$$a_j(x) := \sum_{i=1}^{j} Q(x, \{k_i\}) \quad \text{for all } x \in E. \tag{2.4}$$

By convention, we set $a_0 := 0$. We also introduce the function H defined by

$$H(x, u) := \sum_{i=1}^{|\Theta|} k_i \mathbb{1}_{a_{i-1}(x) < u \leq a_i(x)} \quad \text{for all } x \in E,\ u \in [0, 1].$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

For all $x \in E$, $H(x, .)$ is the inverse of the cumulative distribution function of $Q(x, .)$ (see e.g. [9]). Then we construct (θ_1, ν_1) from the uniform random variable V_1 and the function H as follows:

$$\begin{aligned}(\theta_1, \nu_1) &= (H((\theta_0, \Phi_{\theta_0}(T^*_{\tau_1}, \nu_0)), V_1), \phi_{\theta_0}(T^*_{\tau_1}, \nu_0)) \\ &= (H((\theta_0, \Phi_{\theta_0}(T_1, \nu_0)), V_1), \phi_{\theta_0}(T_1, \nu_0)).\end{aligned}$$

Thus, the distribution of (θ_1, ν_1) given $(\tau_1, (T^*_k)_{k\le\tau_1})$ is $Q((\theta_0, \Phi_{\theta_0}(T^*_{\tau_1}, \nu_0)), .)$ or, in view of (2.1),

$$\sum_{k\in\Theta} Q((\theta_0, \Phi_{\theta_0}(T^*_{\tau_1}, \nu_0)), \{k\})\delta_{(k,\phi_{\theta_0}(T^*_{\tau_1}, \nu_0))}.$$

For $n > 1$, assume that $(\tau_{n-1}, (T^*_k)_{k\le\tau_{n-1}}, (\theta_{n-1}, \nu_{n-1}))$ is constructed. Then we construct T_n by thinning of (T^*_k) conditionally to $(\tau_{n-1}, (T^*_k)_{k\le\tau_{n-1}}, (\theta_{n-1}, \nu_{n-1}))$, that is,

$$T_n := T^*_{\tau_n},$$

where

$$\tau_n := \inf\{k > \tau_{n-1} : U_k\lambda^* \le \lambda(\theta_{n-1}, \Phi_{\theta_{n-1}}(T^*_k - T^*_{\tau_{n-1}}, \nu_{n-1}))\}.$$

Then we construct (θ_n, ν_n) using the uniform random variable V_n and the function H as follows:

$$\begin{aligned}(\theta_n, \nu_n) &:= (H((\theta_{n-1}, \Phi_{\theta_{n-1}}(T^*_{\tau_n} - T^*_{\tau_{n-1}}, \nu_{n-1})), V_n), \Phi_{\theta_{n-1}}(T^*_{\tau_n} - T^*_{\tau_{n-1}}, \nu_{n-1})) \\ &= (H((\theta_{n-1}, \Phi_{\theta_{n-1}}(T_n - T_{n-1}, \nu_{n-1})), V_n), \Phi_{\theta_{n-1}}(T_n - T_{n-1}, \nu_{n-1})).\end{aligned}$$

We define the PDP x_t for all $t \in [0, T]$ from the process $(T_n, (\theta_n, \nu_n))$ by

$$x_t := (\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n)), \quad t \in [T_n, T_{n+1}[. \tag{2.5}$$

Thus $x_{T_n} = (\theta_n, \nu_n)$ and $x^-_{T_n} = (\theta_{n-1}, \nu_n)$. We also define the counting process associated with the jump times $N_t := \sum_{n\ge1} \mathbb{1}_{T_n\le t}$.

2.2. Approximation of a PDP

In applications we may not know the functions Φ_θ explicitly. In this case, we use a numerical scheme $\overline{\Phi}_\theta$ approximating Φ_θ. In this paper we consider schemes such that there exist positive constants C_1 and C_2 independent of h and θ such that

$$\sup_{t\in[0,T]} |\Phi_\theta(t, \nu_1) - \overline{\Phi}_\theta(t, \nu_2)| \le \mathrm{e}^{C_1T}|\nu_1 - \nu_2| + C_2h \quad \text{for all } \theta \in \Theta,\ (\nu_1, \nu_2) \in \mathbb{R}^2. \tag{2.6}$$

To each family $(\overline{\Phi}_\theta)$ we can associate a PDP constructed as above, which we denote by $(\overline{x}_t)$. We emphasize that there is a positive probability that (x_t) and $(\overline{x}_t)$ jump at different times and/or in different states, even if they are both constructed from the same data (N^*_t), (U_k), and (V_k). However, if the characteristics $(\Phi, \tilde{\lambda}, \tilde{Q})$ of a PDP $(\tilde{x}_t)$ are such that $\tilde{\lambda}$ and $\tilde{Q}$ depend only on θ, i.e. $\tilde{\lambda}(x) = \tilde{\lambda}(\theta)$ and $\tilde{Q}(x, .) = \tilde{Q}(\theta, .)$ for all $x = (\theta, \nu) \in E$, then its embedded Markov chain $(\tilde{T}_n, (\tilde{\theta}_n, \tilde{\nu}_n), n \ge 0)$ is such that $(\tilde{\theta}_n, n \ge 0)$ is an autonomous Markov chain with kernel $\tilde{Q}$ and $(\tilde{T}_n, n \ge 0)$ is a counting process with intensity $\tilde{\lambda}_t = \sum_{n\ge0} \tilde{\lambda}(\tilde{\theta}_n)\mathbb{1}_{\tilde{T}_n\le t<\tilde{T}_{n+1}}$. In particular, $(\tilde{\theta}_n)$ and $(\tilde{\tau}_n)$ do not depend on Φ. The particular form of the characteristics $\tilde{\lambda}$ and $\tilde{Q}$ implies that the PDP $(\tilde{x}_t)$ and its approximation $(\underline{\tilde{x}_t})$ are correlated via the same process $(\tilde{\tau}_n, \tilde{\theta}_n)$. In other words, these processes always jump at exactly the same times and their θ-components always

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

jump in the same states. Such processes $(\tilde{x}_t)$ are easier theoretically as well as numerically than the general case. They will be useful for us below.

The following lemma (which is important for several proofs below) gives a direct consequence of the estimate (2.6).

Lemma 2.1. *Let* (Φ_θ) *and* $(\overline{\Phi}_\theta)$ *satisfy* (2.6). *Let* $(t_n, n \geq 0)$ *be an increasing sequence of non-negative real numbers with* $t_0 = 0$ *and let* $(\alpha_n, n \geq 0)$ *be a sequence of* Θ*-valued components. For a given* $\nu \in \mathbb{R}$, *let us define iteratively the sequences* $(\beta_n, n \geq 0)$ *and* $(\overline{\beta}_n, n \geq 0)$ *as follows*:

$$\beta_n = \Phi_{\alpha_{n-1}}(t_n - t_{n-1}, \beta_{n-1}), \qquad \overline{\beta}_n = \overline{\Phi}_{\alpha_{n-1}}(t_n - t_{n-1}, \overline{\beta}_{n-1}),$$
$$\beta_0 = \nu, \qquad \overline{\beta}_0 = \nu.$$

Then for all $n \geq 1$ *we have*

$$|\overline{\beta}_n - \beta_n| \leq \mathrm{e}^{C_1 t_n} n C_2 h,$$

where C_1 *and* C_2 *are positive constants independent of* h.

Proof of Lemma 2.1. Let $n \geq 1$. From the estimate (2.6), we have for all $k \leq n$

$$|\overline{\beta}_k - \beta_k| \leq \mathrm{e}^{C_1(t_k - t_{k-1})}|\overline{\beta}_{k-1} - \beta_{k-1}| + C_2 h,$$

and therefore

$$\mathrm{e}^{-C_1 t_k}|\overline{\beta}_k - \beta_k| \leq \mathrm{e}^{-C_1 t_{k-1}}|\overline{\beta}_{k-1} - \beta_{k-1}| + C_2 h.$$

By summing up these inequalities for $1 \leq k \leq n$ and since $\beta_0 = \overline{\beta}_0$, we obtain

$$|\overline{\beta}_n - \beta_n| \leq \mathrm{e}^{C_1 t_n} n C_2 h. \qquad \square$$

2.3. Application to the construction of a PDMP and its associated Euler scheme

In this subsection we define a PDMP and its associated Euler scheme from the construction of the Section 2.1. Consider a family of vector fields $(f_\theta, \theta \in \Theta)$ satisfying the following.

Assumption 2.2. *For all* $\theta \in \Theta$, *the function* $f_\theta : \mathbb{R} \to \mathbb{R}$ *is bounded and Lipschitz with constant* L *independent of* θ.

If we choose $\Phi_\theta = \phi_\theta$ in the above construction, where for all $x = (\theta, \nu) \in E$ we let $(\phi_\theta(t, \nu), t \geq 0)$ denote the unique solution of the ODE

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = f_\theta(y(t)), \quad y(0) = \nu, \tag{2.7}$$

then the corresponding PDP is Markov since ϕ satisfies the semi-group property that reads $\phi_\theta(t + s, \nu) = \phi_\theta(t, \phi_\theta(s, \nu))$ for all $t, s \geq 0$ and for all $(\theta, \nu) \in E$. In this case, the process (x_t) is a piecewise deterministic Markov process (see [6] or [20]).

Let $h > 0$. We approximate the solution of (2.7) by the Euler scheme with time step h. First, we define the Euler subdivision of $[0, +\infty[$ with time step h, denoted $(\bar{t}_i, i \geq 0)$, by $\bar{t}_i := ih$.

Then, for all $x = (\theta, \nu) \in E$, we define the sequence $(\overline{y}_i(x), i \geq 0)$, the classical Euler scheme, iteratively by

$$\overline{y}_{i+1}(x) = \overline{y}_i(x) + h f_\theta(\overline{y}_i(x)), \quad \overline{y}_0(x) = \nu,$$

to emphasize its dependence on the initial condition. Finally, for all $x = (\theta, \nu) \in E$, we set

$$\overline{\phi}_\theta(t, \nu) := \overline{y}_i(x) + (t - \bar{t}_i) f_\theta(\overline{y}_i(x)) \quad \text{for all } t \in [\bar{t}_i, \bar{t}_{i+1}]. \tag{2.8}$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

We construct the approximating process $(\overline{x}_t)$ as follows. Its continuous component starts from ν_0 at time 0 and follows the flow $\overline{\phi}_{\theta_0}(t, \nu_0)$ until the first jump time $\overline{T}_1$, which we construct by (2.2) and (2.3) of Section 2.1, where we replace $\Phi_{\theta_0}(T_k^*, \nu_0)$ by $\overline{\phi}_{\theta_0}(T_k^*, \nu_0)$. At time $\overline{T}_1$ the continuous component of $\overline{x}_{\overline{T}_1}$ is equal to $\overline{\phi}_{\theta_0}(\overline{T}_1, \nu_0) := \overline{\nu}_1$ since there is no jump in the continuous component. The discrete component jumps to $\overline{\theta}_1$. We iterate this procedure with the new flow $\overline{\phi}_{\overline{\theta}_1}(t - \overline{T}_1, \overline{\nu}_1)$ until the next jump time $\overline{T}_2$ given by (2.2) and (2.3) with $\overline{\phi}_{\overline{\theta}_1}(T_k^* - \overline{T}_1, \overline{\nu}_1)$, and so on. We proceed by iteration to construct $(\overline{x}_t)$ on $[0, T]$.

Consequently, the discretization grid for $(\overline{x}_t)$ on the interval $[0, T]$ is random and is formed by the points $\overline{T}_n + kh$ for $n = 0, \ldots, \overline{N}_T$ and $k = 0, \ldots, \lfloor (\overline{T}_{n+1} \wedge T - \overline{T}_n)/h \rfloor$. This differs from the SDE case, where the classical grid is fixed.

By classical results of numerical analysis (see e.g. [17]), the continuous Euler scheme (2.8) (also called the Euler polygon) satisfies estimate (2.6). If we choose $\Phi_\theta = \overline{\phi}_\theta$ in the above construction then the corresponding PDP $(\overline{x}_t)$ is not Markov, since the functions $\overline{\phi}_\theta(., \nu)$ do not satisfy the semi-group property (see [20]).

2.4. Thinning representation for the marginal distribution of a PDP

The sequence $(T_n, (\theta_n, \nu_n), n \geq 0)$ is an $\mathbb{R}_+ \times E$-valued Markov chain with respect to its natural filtration $\mathcal{F}_n$ and with kernel K defined by

$$K((t, \theta, \nu), \mathrm{d}u\, \mathrm{d}j\, \mathrm{d}z) := \mathbb{1}_{u \geq t}\, \lambda(\theta, \Phi_\theta(u - t, \nu)) \exp\left(-\int_0^{u-t} \lambda(\theta, \Phi_\theta(s, \nu))\, \mathrm{d}s\right) Q((\theta, \Phi_\theta(u - t, \nu)), \mathrm{d}j\, \mathrm{d}z)\, \mathrm{d}u. \tag{2.9}$$

For $n \geq 0$, the law of the random variable $T_n - T_{n-1}$ given $\mathcal{F}_{n-1}$ admits the density given for $t \geq 0$ by

$$\lambda(\theta_{n-1}, \Phi_{\theta_{n-1}}(t, \nu_{n-1})) \exp\left(-\int_0^t \lambda(\theta_{n-1}, \Phi(s, \nu_{n-1}))\, \mathrm{d}s\right). \tag{2.10}$$

Classically, the marginal distribution of x_t is expressed using (2.5), the intensity λ via (2.10), and the kernel K (see (2.9)). Indeed, for fixed $x_0 = x \in E$ and for any bounded measurable function g, we can write

$$\begin{aligned}
\mathbb{E}[g(x_t)] &= \sum_{n \geq 0} \mathbb{E}[g(\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n)) \mathbb{1}_{N_t = n}] \\
&= \sum_{n \geq 0} \mathbb{E}[g(\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n)) \mathbb{1}_{T_n \leq t} \mathbb{E}[\mathbb{1}_{T_{n+1} > t} | \mathcal{F}_n]] \\
&= \sum_{n \geq 0} \mathbb{E}\left[g(\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n)) \mathbb{1}_{T_n \leq t} \exp\left(-\int_0^{t - T_n} \lambda(\theta_n, \Phi_{\theta_n}(u, \nu_n))\, \mathrm{d}u\right)\right] \qquad (2.11) \\
&= \sum_{n \geq 0} \int_0^t \int_E g(\theta, \Phi_\theta(t - s, \nu)) \exp\left(-\int_0^{t-s} \lambda(\theta, \Phi_\theta(u, \nu))\, \mathrm{d}u\right) K^n((0, x), \mathrm{d}s\, \mathrm{d}\theta\, \mathrm{d}\nu), \qquad (2.12)
\end{aligned}$$

where $K^0 := \delta$ and $K^n = K \circ \ldots \circ K$ n times, that is,

$$\int_0^t \int_E K^n((0, x), \mathrm{d}s\, \mathrm{d}y) = \int_0^t \int_E \int_{(\mathbb{R}_+ \times E)^{n-1}} K((0, x), \mathrm{d}t_1\, \mathrm{d}y_1) \ldots K((t_{n-1}, y_{n-1}), \mathrm{d}s\, \mathrm{d}y).$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

However, since we have constructed (x_t) by thinning, we would prefer to express the distribution of x_t using the upper bound λ^*, the Poisson process $(N_t^*, t \geq 0)$, and the sequences $(U_k, k \in \mathbb{N})$, $(V_k, k \in \mathbb{N})$. In Proposition 2.1 we give another representation of (2.11). The product term which appears in the expectation on the right-hand side of the equality in Proposition 2.1 should be interpreted as the conditional survival function,

$$t \to \exp\left(-\int_0^{t-T_n} \lambda(\theta_n, \Phi_{\theta_n}(u, \nu_n))\,\mathrm{d}u\right),$$

of T_{n+1} in (2.11).

Proposition 2.1. *Let $(x_t, t \in [0, T])$ be a PDP with characteristics (Φ, λ, Q) constructed in Section 2.1 and let $n \in \mathbb{N}$. Then*

$$\mathbb{E}[g(x_t)\mathbb{1}_{\{N_t=n\}}] = \sum_{1\leq p_1<p_2\cdots<p_n\leq m} \sum_{\theta\in\Theta} \mathbb{E}\Bigg[Q(x_{T^*_{p_n}}^-, \theta)\, g(\theta, \Phi_\theta(t - T^*_{p_n}, \nu_n)) \times \mathbb{1}_{\{\tau_i=p_i, 1\leq i\leq n, N_t^*=m\}} \prod_{q=p_n+1}^{m} \left(1 - \frac{\lambda(\theta, \Phi_\theta(T_q^* - T^*_{p_n}, \nu_n))}{\lambda^*}\right)\Bigg].$$

The following proposition and its corollaries will be useful in Section 3. In their statements $(x_t, t \in [0, T])$ and $(\tilde{x}_t, t \in [0, T])$ are PDPs constructed in Section 2.1 using the same data (N_t^*), (U_k), (V_k) and the same initial point $x \in E$ but with different sets of characteristics.

The following results are inspired by the change of probability introduced in [28] where the authors are interested in the application of the MLMC to jump-diffusion SDEs with state-dependent intensity. In our case we need a change of probability which guarantees not only that the processes jump at the same times but also in the same states.

Proposition 2.2. *Let (Φ, λ, Q) (resp. $(\Phi, \tilde{\lambda}, \tilde{Q})$) denote the characteristics of (x_t) (resp. $(\tilde{x}_t)$). Let us assume that $\tilde{\lambda}$ and $\tilde{Q}$ depend only on θ, that $\tilde{Q}$ is always positive, and $0 < \tilde{\lambda}(\theta) < \lambda^*$ for all $\theta \in \Theta$. For all integer n, let us define on the event $\{\tilde{N}_t = n\}$,*

$$\tilde{Z}_n = \frac{Q(\tilde{x}_{T^*_{\tilde{\tau}_n}}^-, \tilde{\theta}_n)}{\tilde{Q}(\tilde{\theta}_{n-1}, \tilde{\theta}_n)} \left(\left(1 - \frac{\tilde{\lambda}(\tilde{\theta}_n)}{\lambda^*}\right)^{N_t^*-\tilde{\tau}_n}\right)^{-1} \prod_{q=\tilde{\tau}_n+1}^{N_t^*} \left(1 - \frac{\lambda(\tilde{\theta}_n, \Phi_{\tilde{\theta}_n}(T_q^* - T^*_{\tilde{\tau}_n}, \tilde{\nu}_n))}{\lambda^*}\right),$$

the product being equal to 1 if $\tilde{\tau}_n = N_t^*$ and, for all $1 \leq \ell \leq n-1$,

$$\tilde{Z}_\ell = \frac{Q(\tilde{x}_{T^*_{\tilde{\tau}_\ell}}^-, \tilde{\theta}_\ell)}{\tilde{Q}(\tilde{\theta}_{\ell-1}, \tilde{\theta}_\ell)} \left(\frac{\tilde{\lambda}(\tilde{\theta}_\ell)}{\lambda^*}\left(1 - \frac{\tilde{\lambda}(\tilde{\theta}_\ell)}{\lambda^*}\right)^{\tilde{\tau}_{\ell+1}-\tilde{\tau}_\ell-1}\right)^{-1} \times \frac{\lambda(\tilde{\theta}_\ell, \Phi_{\tilde{\theta}_\ell}(T^*_{\tilde{\tau}_{\ell+1}} - T^*_{\tilde{\tau}_\ell}, \tilde{\nu}_\ell))}{\lambda^*} \prod_{q=\tilde{\tau}_\ell+1}^{\tilde{\tau}_{\ell+1}-1} \left(1 - \frac{\lambda(\tilde{\theta}_\ell, \Phi_{\tilde{\theta}_\ell}(T_q^* - T^*_{\tilde{\tau}_\ell}, \tilde{\nu}_\ell))}{\lambda^*}\right),$$

$$\tilde{Z}_0 = \left(\frac{\tilde{\lambda}(\tilde{\theta}_0)}{\lambda^*}\left(1 - \frac{\tilde{\lambda}(\tilde{\theta}_0)}{\lambda^*}\right)^{\tilde{\tau}_1-1}\right)^{-1} \frac{\lambda(\tilde{\theta}_0, \Phi_{\tilde{\theta}_0}(T^*_{\tilde{\tau}_1}, \tilde{\nu}_0))}{\lambda^*} \prod_{q=1}^{\tilde{\tau}_1-1} \left(1 - \frac{\lambda(\tilde{\theta}_0, \Phi_{\tilde{\theta}_0}(T_q^*, \tilde{\nu}_0))}{\lambda^*}\right),$$

$$\tilde{R}_n = \tilde{Z}_n \prod_{\ell=0}^{n-1} \tilde{Z}_\ell.$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Then, for all $n \geq 0$ we have

$$\mathbb{E}[g(\tilde{x}_t)\,\tilde{R}_n\,\mathbb{1}_{\{\tilde{N}_t=n\}}] = \mathbb{E}[g(x_t)\,\mathbb{1}_{\{N_t=n\}}].$$

Corollary 2.1. *Under the assumptions of Proposition 2.2, setting $\tilde{R}_t = \tilde{R}_{\tilde{N}_t}$, we have*

$$\mathbb{E}[g(\tilde{x}_t)\tilde{R}_t] = \mathbb{E}[g(x_t)].$$

Remark 2.1. Proposition 2.2 looks like Girsanov's theorem (see [26]), but we do not use martingale theory here.

Remark 2.2. We have chosen to state Proposition 2.2 with a PDP $(\tilde{x}_t)$ whose intensity and transition measure only depend on θ for the sake of readability. In fact the arguments of the proof are valid for non-homogeneous intensity and transition measures of the form $\tilde{\lambda}(x,t)$ and $\tilde{Q}((x,t),\mathrm{d}y)$ for $x=(\theta,\nu)\in E$. A possible choice of such characteristics is $\tilde{\lambda}(x,t) = \lambda(\theta,\tilde{\Phi}_\theta(t,\nu))$ and $\tilde{Q}((x,t),\mathrm{d}y) = Q((\theta,\tilde{\Phi}_\theta(t,\nu)),\mathrm{d}y)$ for $\tilde{\Phi}$ a given function. This remark will be implemented in Section 5.4.

Corollary 2.2. *Let (Φ,λ,Q) (resp. $(\tilde{\Phi},\lambda,Q)$) be the set of characteristics of (x_t) (resp. $(\tilde{x}_t)$). We assume that Q is always positive and that $0<\lambda(x)<\lambda^*$ for all $x\in E$. Let (μ_n) be the sequence defined by $\mu_0=\nu$ and $\mu_n=\tilde{\Phi}_{\theta_{n-1}}(T_n-T_{n-1},\mu_{n-1})$ for $n\geq 1$. For all integer n, let us define on the event $\{N_t=n\}$,*

$$\tilde{Z}_n = \frac{Q((\theta_{n-1},\mu_n),\theta_n)}{Q((\theta_{n-1},\nu_n),\theta_n)}\left(\prod_{q=\tau_n+1}^{N_t^*} 1-\frac{\lambda(\theta_n,\Phi_{\theta_n}(T_q^*-T_{\tau_n}^*,\nu_n))}{\lambda^*}\right)^{-1}$$
$$\times\prod_{q=\tau_n+1}^{N_t^*}\left(1-\frac{\lambda(\theta_n,\tilde{\Phi}_{\theta_n}(T_q^*-T_{\tau_n}^*,\mu_n))}{\lambda^*}\right),$$

the products being equal to 1 if $\tau_n=N_t^*$ and for all $1\leq\ell\leq n-1$,

$$\tilde{Z}_\ell = \frac{Q((\theta_{\ell-1},\mu_\ell),\theta_\ell)}{Q((\theta_{\ell-1},\nu_\ell),\theta_\ell)}\left(\frac{\lambda(\theta_\ell,\Phi_{\theta_\ell}(T_{\tau_{\ell+1}}^*-T_{\tau_\ell}^*,\nu_\ell))}{\lambda^*}\prod_{q=\tau_\ell+1}^{\tau_{\ell+1}-1}\left(1-\frac{\lambda(\theta_\ell,\Phi_{\theta_\ell}(T_q^*-T_{\tau_\ell}^*,\nu_\ell))}{\lambda^*}\right)\right)^{-1}$$
$$\times\frac{\lambda(\theta_\ell,\tilde{\Phi}_{\theta_\ell}(T_{\tau_{\ell+1}}^*-T_{\tau_\ell}^*,\mu_\ell))}{\lambda^*}\prod_{q=\tau_\ell+1}^{\tau_{\ell+1}-1}\left(1-\frac{\lambda(\theta_\ell,\tilde{\Phi}_{\theta_\ell}(T_q^*-T_{\tau_\ell}^*,\mu_\ell))}{\lambda^*}\right),$$

$$\tilde{Z}_0 = \left(\frac{\lambda\big(\theta_0,\Phi_{\theta_0}(T_{\tau_1}^*,\nu_0)\big)}{\lambda^*}\prod_{q=1}^{\tau_1-1}\left(1-\frac{\lambda\big(\theta_0,\Phi_{\theta_0}(T_q^*,\nu_0)\big)}{\lambda^*}\right)\right)^{-1}$$
$$\times\frac{\lambda\big(\theta_0,\tilde{\Phi}_{\theta_0}(T_{\tau_1}^*,\mu_0)\big)}{\lambda^*}\prod_{q=1}^{\tau_1-1}\left(1-\frac{\lambda\big(\theta_0,\tilde{\Phi}_{\theta_0}(T_q^*,\mu_0)\big)}{\lambda^*}\right),$$

$$\tilde{R}_n = \tilde{Z}_n\prod_{\ell=0}^{n-1}\tilde{Z}_\ell.$$

Then, for all $n\geq 0$ we have

$$\mathbb{E}[g(\theta_n,\tilde{\Phi}_{\theta_n}(t-T_n,\mu_n))\,\tilde{R}_n\,\mathbb{1}_{\{N_t=n\}}] = \mathbb{E}[g(\tilde{x}_t)\,\mathbb{1}_{\{\tilde{N}_t=n\}}].$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Proof of Proposition 2.1. It holds that $\{N_t = n,\ \tau_i = p_i,\ 1 \le i \le n\} \subset \{N_t^* \ge p_n\}$. Then

$$\mathbb{E}[g(x_t)\mathbb{1}_{\{N_t=n\}}] = \sum_{1\le p_1<p_2<\cdots<p_n\le m} \mathbb{E}[g(x_t)\mathbb{1}_{\{N_t=n,\,\tau_i=p_i,\,1\le i\le n,\,N_t^*=m\}}]. \tag{2.13}$$

The set $\{N_t = n,\ \tau_i = p_i,\ 1 \le i \le n, N_t^* = m\}$ is equivalent to the following:

- $N_t^* = m$,
- among the times T_ℓ^*, $1 \le \ell \le m$, exactly n are accepted by the thinning method; these are the $T_{p_i}^*$, $1 \le i \le n$, and all the others are rejected.

We proceed by induction, starting from the fact that all the T_q^*, $p_n + 1 \le q \le m$ are rejected, which corresponds to the event

$$U_q > \frac{\lambda(\theta_n,\, \Phi_{\theta_n}(T_q^* - T_{p_n}^*,\, \nu_n))}{\lambda^*} \quad \text{for all } p_n + 1 \le q \le m.$$

The random variable $\mathbb{1}_{\{\tau_i=p_i,\,1\le i\le n\}}$ depends on

$$(\theta_\ell,\, \nu_\ell,\, 1 \le \ell \le n-1,\, T_i^*,\, 1 \le i \le p_n,\, U_j,\, 1 \le j \le p_n),$$

where by construction $\nu_\ell = \phi_{\theta_{\ell-1}}(T_{p_\ell}^* - T_{p_{\ell-1}}^*,\, \nu_{\ell-1})$, $\theta_\ell = H((\theta_{\ell-1},\, \nu_\ell),\, V_\ell)$, which implies that $(\theta_\ell,\, \nu_\ell,\, 1 \le \ell \le n-1)$ depend on $(T_i^*,\, 1 \le i \le p_{n-1},\, U_j,\, 1 \le j \le p_{n-1},\, V_k,\, 1 \le k \le n-1)$. Thus V_n is independent of all the other random variables of thinning that are present in $g(x_t)\mathbb{1}_{\{N_t=n,\,\tau_i=p_i,\,1\le i\le n,\,N_t^*=m\}}$. The conditional expectation of $g(x_t)\mathbb{1}_{\{N_t=n,\,\tau_i=p_i,\,1\le i\le n,N_t^*=m\}}$ with respect to the vector $(T_i^*,\, 1 \le i \le m+1,\, U_j,\, 1 \le j \le m,\, V_k,\, 1 \le k \le n-1)$ is therefore an expectation indexed by this vector as parameters. Since the law of $H(x, V_n)$ is $Q(x, \cdot)$ for all $x \in E$, we obtain for $p_1 < p_2 < \cdots < p_n \le m$,

$$\begin{aligned}
&\mathbb{E}[g(x_t)\mathbb{1}_{\{N_t=n,\,\tau_i=p_i,\,1\le i\le n,\,N_t^*=m\}}] \\
&\quad = \mathbb{E}\Bigg[\sum_{\theta\in\Theta} Q(x_{T_{p_n}^*}^-,\, \theta)\, g(\theta,\, \Phi_\theta(t - T_{p_n}^*,\, \nu_n)) \\
&\qquad \times F(\theta,\, U_j,\, 1 \le j \le m,\ T_\ell^*,\, 1 \le \ell \le m+1,\, V_k,\, 1 \le k \le n-1)\Bigg],
\end{aligned} \tag{2.14}$$

with

$$\begin{aligned}
&F(\theta,\, U_j,\, 1 \le j \le m,\ T_\ell^*,\, 1 \le \ell \le m+1,\, V_k,\, 1 \le k \le n-1) \\
&\quad = \mathbb{1}_{\{N_t^*=m,\,\tau_i=p_i,\,1\le i\le n\}} \prod_{q=p_n+1}^{m} \mathbb{1}_{U_q > \lambda(\theta,\,\Phi_\theta(T_q^* - T_{p_n}^*,\,\nu_n))/\lambda^*}.
\end{aligned}$$

In (2.14) the random variables $(U_q,\ p_n + 1 \le q \le m)$ are independent of the vector $(T_i^*,\, 1 \le i \le m+1,\, U_j,\, 1 \le j \le p_n,\, V_k,\, 1 \le k \le n-1)$. Conditioning by this vector, we obtain

$$\begin{aligned}
&\mathbb{E}[g(x_t)\mathbb{1}_{\{N_t=n,\,\tau_i=p_i,\,1\le i\le n,\,N_t^*=m\}}] \\
&\quad = \sum_{\theta\in\Theta} \mathbb{E}\Bigg[Q(x_{T_{p_n}^*}^-,\, \theta)\, g(\theta,\, \Phi_\theta(t - T_{p_n}^*,\, \nu_n))\mathbb{1}_{\{N_t^*=m,\,\tau_i=p_i,\,1\le i\le n\}} \\
&\qquad \times \prod_{q=p_n+1}^{m} \left(1 - \frac{\lambda(\theta,\, \Phi_\theta(T_q^* - T_{p_n}^*,\, \nu_n))}{\lambda^*}\right)\Bigg].
\end{aligned} \tag{2.15}$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Summing the above equality over $1 \le p_1 < p_2 < \cdots < p_n \le m$ and using equation (2.13) yields the result. □

Just as with the successive use of conditioning to obtain (2.12), we can iterate on the form (2.15) by first conditioning V_{n-1} by all the other random variables and then conditioning $(U_q,\ p_{n-1}+1 \le q \le p_n)$ by all the remaining ones, and so on. However, the terms that appear do not have the same structure, since the U_q correspond to rejection for $p_{n-1}+1 \le q \le p_n - 1$ whereas U_{p_n} corresponds to acceptance. Consequently the next step yields

$$\begin{aligned}
&\mathbb{E}[g(x_t)\mathbb{1}_{\{N_t=n,\tau_i=p_i,\ 1\le i\le n,\ N_t^*=m\}}] \\
&\quad= \sum_{\alpha\in\Theta}\sum_{\theta\in\Theta}\mathbb{E}\Bigg[Q(x^-_{T^*_{p_{n-1}}},\alpha)Q((\alpha,\nu_n),\theta) \\
&\qquad\times g(\theta,\Phi_\theta(t-T^*_{p_n},\nu_n))\mathbb{1}_{\{N_t^*=m,\tau_i=p_i,\ 1\le i\le n-1\}} \\
&\qquad\times \frac{\lambda(\alpha,\Phi_\alpha(T^*_{p_n}-T^*_{p_{n-1}},\nu_{n-1}))}{\lambda^*}\prod_{q=p_{n-1}+1}^{p_n-1}\left(1-\frac{\lambda(\alpha,\Phi_\alpha(T^*_q-T^*_{p_{n-1}},\nu_{n-1}))}{\lambda^*}\right) \\
&\qquad\times \prod_{q=p_n+1}^{m}\left(1-\frac{\lambda(\theta,\Phi_\theta(T^*_q-T^*_{p_n},\nu_n))}{\lambda^*}\right)\Bigg],
\end{aligned} \tag{2.16}$$

where we write ν_n for simplicity, keeping in mind that

$$\begin{aligned}
\nu_n &= \Phi_{\theta_{n-1}}(T^*_{p_n}-T^*_{p_{n-1}},\nu_{n-1}) \\
&= \Phi_{\theta_{n-1}}(T^*_{p_n}-T^*_{p_{n-1}},\Phi_{\theta_{n-2}}(T^*_{p_{n-1}}-T^*_{p_{n-2}},\nu_{n-2})) \\
&= \Phi_\alpha(T^*_{p_n}-T^*_{p_{n-1}},\Phi_{\theta_{n-2}}(T^*_{p_{n-1}}-T^*_{p_{n-2}},\nu_{n-2})).
\end{aligned}$$

In (2.16), the product term

$$\frac{\lambda(\alpha,\Phi_\alpha(T^*_{p_n}-T^*_{p_{n-1}},\nu_{n-1}))}{\lambda^*}\prod_{q=p_{n-1}+1}^{p_n-1}\left(1-\frac{\lambda(\alpha,\Phi_\alpha(T^*_q-T^*_{p_{n-1}},\nu_{n-1}))}{\lambda^*}\right)$$

should be interpreted as the density probability function of T_n which appears in (2.12) via the kernel K.

Moreover, the previous arguments apply to

$$\mathbb{E}(g(x_t)f(\theta_i,\nu_i,\ 1\le i\le n-1,\theta_n,\nu_n,T^*_k,\ 1\le k\le m)\ \mathbb{1}_{\{N_t=n,\tau_i=p_i,\ 1\le i\le n,\ N_t^*=m\}}),$$

where f is a measurable function, and provide

$$\begin{aligned}
&\mathbb{E}[g(x_t)f(\theta_i,\nu_i,\ 1\le i\le n-1,\theta_n,\nu_n,T^*_k,\ 1\le k\le m)\ \mathbb{1}_{\{N_t=n,\tau_i=p_i,\ 1\le i\le n,\ N_t^*=m\}}] \\
&\quad= \sum_{\theta\in\Theta}\mathbb{E}\Bigg[Q(x^-_{T^*_{p_n}},\theta)g(\theta,\Phi_\theta(t-T^*_{p_n},\nu_n)) \\
&\qquad\times f(\theta_i,\nu_i,\ 1\le i\le n-1,\theta,\nu_n,T^*_k,\ 1\le k\le m)\ \mathbb{1}_{\{N_t^*=m,\tau_i=p_i,\ 1\le i\le n\}} \\
&\qquad\times \prod_{q=p_n+1}^{m}\left(1-\frac{\lambda(\theta,\Phi_\theta(T^*_q-T^*_{p_n},\nu_n))}{\lambda^*}\right)\Bigg].
\end{aligned} \tag{2.17}$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Below we prove Proposition 2.2. The other statements can be proved analogously.

Proof of Proposition 2.2. By assumption the (jump) characteristics $(\tilde{\lambda}, \tilde{Q})$ of $(\tilde{x}_t)$ depend only on θ. Let $p_1 < p_2 < \cdots < p_n \le m$. Applying the same arguments as in (2.17) to $(\tilde{x}_t)$ and using the definitions of $\tilde{Z}_\ell$, $0 \le \ell \le n$ and $\tilde{R}_n$, we obtain

$$
\begin{aligned}
&\mathbb{E}[g(\tilde{x}_t)\,\tilde{R}_n\,\mathbb{1}_{\{\tilde{N}_t=n,\tilde{\tau}_i=p_i,1\le i\le n,N_t^*=m\}}]\\
&\quad=\sum_{\theta\in\Theta}\mathbb{E}\Bigg[\tilde{Q}(\tilde{\theta}_{n-1},\theta)\,g(\theta,\Phi_\theta(t-T^*_{p_n},\tilde{\nu}_n))\,\tilde{Z}_n\prod_{\ell=0}^{n-1}\tilde{Z}_\ell\,\mathbb{1}_{\{N_t^*=m,\tilde{\tau}_i=p_i,\,1\le i\le n\}}\Bigg]\left(1-\frac{\tilde{\lambda}(\theta)}{\lambda^*}\right)^{m-p_n}\\
&\quad=\sum_{\theta\in\Theta}\mathbb{E}\Bigg[\tilde{Q}(\tilde{\theta}_{n-1},\theta)\,g(\theta,\Phi_\theta(t-T^*_{p_n},\tilde{\nu}_n))\prod_{\ell=0}^{n-1}\tilde{Z}_\ell\;\mathbb{1}_{\{N_t^*=m,\tilde{\tau}_i=p_i,\,1\le i\le n\}}\left(1-\frac{\tilde{\lambda}(\theta)}{\lambda^*}\right)^{m-p_n}\\
&\qquad\times\left(\left(1-\frac{\tilde{\lambda}(\theta)}{\lambda^*}\right)^{m-p_n}\right)^{-1}\frac{Q(\tilde{x}^-_{T^*_{p_n}},\theta)}{\tilde{Q}(\tilde{\theta}_{n-1},\theta)}\prod_{q=p_n+1}^{m}\left(1-\frac{\lambda(\theta,\Phi_\theta(T^*_q-T^*_{p_n},\tilde{\nu}_n))}{\lambda^*}\right)\Bigg]\\
&\quad=\sum_{\theta\in\Theta}\mathbb{E}\Bigg[Q(\tilde{x}^-_{T^*_{p_n}},\theta)\,g(\theta,\Phi_\theta(t-T^*_{p_n},\tilde{\nu}_n))\,\tilde{Z}_{n-1}\prod_{\ell=0}^{n-2}\tilde{Z}_\ell\;\mathbb{1}_{\{N_t^*=m,\tilde{\tau}_i=p_i,\,1\le i\le n\}}\\
&\qquad\times\prod_{q=p_n+1}^{m}\left(1-\frac{\lambda(\theta,\Phi_\theta(T^*_q-T^*_{p_n},\tilde{\nu}_n))}{\lambda^*}\right)\Bigg].
\end{aligned}
$$

We iterate the above argument based on the use of (2.17) and we use the definition of $\tilde{Z}_{n-1}$ to obtain

$$
\begin{aligned}
&\mathbb{E}[g(\tilde{x}_t)\tilde{R}_n\mathbb{1}_{\{\tilde{N}_t=n,\tilde{\tau}_i=p_i,\,1\le i\le n,\,N_t^*=m\}}]\\
&\quad=\sum_{\alpha\in\Theta}\sum_{\theta\in\Theta}\mathbb{E}\Bigg[Q(\tilde{x}^-_{T^*_{p_{n-1}}},\alpha)Q((\alpha,\tilde{\nu}_n),\theta)\,g(\theta,\Phi_\theta(t-T^*_{p_n},\tilde{\nu}_n))\\
&\qquad\times\prod_{\ell=0}^{n-2}\tilde{Z}_\ell\;\mathbb{1}_{\{N_t^*=m,\tilde{\tau}_i=p_i,\,1\le i\le n-1\}}\prod_{q=p_n+1}^{m}\left(1-\frac{\lambda(\theta,\Phi_\theta(T^*_q-T^*_{p_n},\tilde{\nu}_n))}{\lambda^*}\right)\\
&\qquad\times\frac{\lambda(\alpha,\Phi_\alpha(T^*_{p_n}-T^*_{p_{n-1}},\tilde{\nu}_{n-1}))}{\lambda^*}\prod_{q=p_{n-1}+1}^{p_n-1}\left(1-\frac{\lambda(\alpha,\Phi_\alpha(T^*_q-T^*_{p_{n-1}},\tilde{\nu}_{n-1}))}{\lambda^*}\right)\Bigg],
\end{aligned}
$$

where for short $\tilde{\nu}_n=\phi_\alpha(T^*_{p_n}-T^*_{p_{n-1}},\tilde{\nu}_{n-1})$ and $\tilde{\nu}_{n-1}=\phi_{\tilde{\theta}_{n-2}}(T^*_{p_{n-1}}-T^*_{p_{n-2}},\tilde{\nu}_{n-2})$. Comparing the latter expression to (2.16) and using induction, we conclude that

$$
\mathbb{E}[g(\tilde{x}_t)\tilde{R}_n\mathbb{1}_{\{\tilde{N}_t=n,\tilde{\tau}_i=p_i,\,1\le i\le n,\,N_t^*=m\}}]=\mathbb{E}[g(x_t)\,\mathbb{1}_{\{N_t=n,\tau_i=p_i,1\le i\le n,N_t^*=m\}}].
$$

It remains to sum up on p_i, $1\le i\le n$ and m. □

3. Strong error estimates

In this section we are interested in strong error estimates. Below, we state the main assumptions and theorems of this section. The proofs are given in Sections 3.2 and 3.3 respectively.

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Assumption 3.1. *For all $\theta \in \Theta$ and for all $A \in \mathcal{B}(\Theta)$, the functions $\nu \mapsto \lambda(\theta, \nu)$ and $\nu \mapsto Q((\theta, \nu), A)$ are Lipschitz with constants $L_\lambda > 0$, $L_Q > 0$, respectively, independent of θ.*

Theorem 3.1. *Let Φ_θ and $\overline{\Phi}_\theta$ satisfy (2.6) and let $(x_t, t \in [0, T])$ and $(\bar{x}_t, t \in [0, T])$ be the corresponding PDPs constructed in Section 2.1* with *$x_0 = \bar{x}_0 = x$ for some $x \in E$. Assume that Θ is finite and that λ and Q satisfy Assumption 3.1. Then, for all bounded functions $F\colon E \to \mathbb{R}$ such that for all $\theta \in \Theta$ the function $\nu \mapsto F(\theta, \nu)$ is L_F-Lipschitz, where L_F is positive and independent of θ, there exist constants $V_1 > 0$ and $V_2 > 0$ independent of the time step h such that*

$$\mathbb{E}[|F(\bar{x}_T) - F(x_T)|^2] \le V_1 h + V_2 h^2.$$

Remark 3.1. When the numerical scheme $\overline{\Phi}_\theta$ is of order $p \ge 1$, which means

$$\sup_{t \in [0,T]} |\Phi_\theta(t, \nu_1) - \overline{\Phi}_\theta(t, \nu_2)| \le \mathrm{e}^{C_1 T} |\nu_1 - \nu_2| + C_2 h^p,$$

we have

$$\mathbb{E}[|F(\bar{x}_T) - F(x_T)|^2] \le V_1 h^p + V_2 h^{2p}.$$

Assumption 3.2. *There exist positive constants ρ, $\tilde{\lambda}_{\min}$, $\tilde{\lambda}_{\max}$ such that, for all $(i, j) \in \Theta^2$, $\rho \le \tilde{Q}(i, j)$ and $\tilde{\lambda}_{\min} \le \tilde{\lambda}(i) \le \tilde{\lambda}_{\max} < \lambda^*$.*

Theorem 3.2. *Let Φ_θ* and *$\overline{\Phi}_\theta$ satisfy (2.6) and let $(\tilde{x}_t, t \in [0, T])$ and $(\underline{\tilde{x}}_t, t \in [0, T])$ be the corresponding PDPs constructed in Section 2.1 with $\underline{\tilde{x}}_0 = \tilde{x}_0 = x$ for some $x \in E$. Let $(\tilde{R}_t, t \in [0, T])$ and $(\underline{\tilde{R}}_t, t \in [0, T])$ be defined as in Corollary 2.1. Under Assumptions 3.1 and 3.2 and for all bounded functions $F\colon E \to \mathbb{R}$ such that for all $\theta \in \Theta$ the function $\nu \mapsto F(\theta, \nu)$ is L_F-Lipschitz ($L_F > 0$), there exists a positive constant $\tilde{V}_1$ independent of the time step h* such that

$$\mathbb{E}[|F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T - F(\tilde{x}_T)\tilde{R}_T|^2] \le \tilde{V}_1 h^2.$$

We now introduce the random variable $\overline{\tau}^\dagger$ which will play an important role in the strong error estimate of Theorem 3.1 as well as in the identification of the coefficient c_1 in the weak error expansion in Section 4 (see the proof of Theorem 4.1 in Section 4.2).

Definition 3.1. Let us define $\overline{\tau}^\dagger := \inf\{k > 0\colon (\tau_k, \theta_k) \ne (\overline{\tau}_k, \overline{\theta}_k)\}$.

The random variable $\overline{\tau}^\dagger$ enables us to partition the trajectories of the pair $(x_t, \bar{x}_t)$ in a sense that we now make precise. Consider the event

$$\{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T\} = \{N_T = \overline{N}_T, (T_1, \theta_1) = (\overline{T}_1, \overline{\theta}_1), \ldots, (T_{N_T}, \theta_{N_T}) = (\overline{T}_{\overline{N}_T}, \overline{\theta}_{\overline{N}_T})\}, \quad (3.1)$$

where (T_n) and $(\overline{T}_n)$ denote the sequences of jump times of (x_t) and $(\bar{x}_t)$. On this event $\{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T\}$ the trajectories of the discrete time processes (T_n, θ_n) and $(\overline{T}_n, \overline{\theta}_n)$ are equal for all n such that $T_n \in [0, T]$ (or equivalently $\overline{T}_n \in [0, T]$). Moreover, the complement, i.e. $\{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \le T\}$, contains the trajectories for which (T_n, θ_n) and $(\overline{T}_n, \overline{\theta}_n)$ differ on $[0, T]$ (there exists $n \le N_T \vee \overline{N}_T$ such that $T_n \ne \overline{T}_n$ or $\theta_n \ne \overline{\theta}_n$).

3.1. Preliminary lemmas

In this subsection we start with two lemmas which will be useful for proving Theorems 3.2 and 3.3.

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Lemma 3.1. *Let K be a finite set. We let $|K|$ denote the cardinal of K, and for $i = 1, \dots, |K|$ we let k_i denote its elements. Let $(p_i, 1 \le i \le |K|)$ and $(\overline{p}_i, 1 \le i \le |K|)$ be two probabilities on K. Let $a_j := \sum_{i=1}^{j} p_i$ and $\overline{a}_j := \sum_{i=1}^{j} \overline{p}_i$ for all $j \in \{1, \dots, |K|\}$. By convention, we set $a_0 = \overline{a}_0 := 0$. Let X and $\overline{X}$ be two K-valued random variables defined by*

$$X := G(U), \quad \overline{X} := \overline{G}(U),$$

where

$$U \sim \mathcal{U}([0,1]), \quad G(u) = \sum_{j=1}^{|K|} k_j \mathbb{1}_{a_{j-1} < u \le a_j}, \quad \overline{G}(u) = \sum_{j=1}^{|K|} k_j \mathbb{1}_{\overline{a}_{j-1} < u \le \overline{a}_j} \quad \text{for all } u \in [0,1].$$

Then, we have

$$\mathbb{P}(X \neq \overline{X}) \le \sum_{j=1}^{|K|-1} |a_j - \overline{a}_j|.$$

Proof of Lemma 3.1. By definition of X and $\overline{X}$ and since the intervals $[a_{j-1}, a_j] \cap [\overline{a}_{j-1}, \overline{a}_j]$ are disjoint for $j = 1, \dots, K$, we have

$$\mathbb{P}(X = \overline{X}) = \sum_{j=1}^{|K|} \mathbb{P}(U \in [a_{j-1}, a_j] \cap [\overline{a}_{j-1}, \overline{a}_j]).$$

Moreover, for all $1 \le j \le |K|$, we have

$$\mathbb{P}(U \in [a_{j-1}, a_j] \cap [\overline{a}_{j-1}, \overline{a}_j]) = \begin{cases} 0 & \text{if } [a_{j-1}, a_j] \cap [\overline{a}_{j-1}, \overline{a}_j] = \emptyset, \\ a_j \wedge \overline{a}_j - a_{j-1} \vee \overline{a}_{j-1} & \text{if } [a_{j-1}, a_j] \cap [\overline{a}_{j-1}, \overline{a}_j] \neq \emptyset. \end{cases}$$

Thus, letting $x^+ := \max(x, 0)$ denote the positive part of $x \in \mathbb{R}$ and using $x^+ \ge x$, we obtain

$$\mathbb{P}(X = \overline{X}) \ge \sum_{j=1}^{|K|} (a_j \wedge \overline{a}_j - a_{j-1} \vee \overline{a}_{j-1}).$$

Adding and subtracting $a_j \vee \overline{a}_j$ in the the above sum yields

$$\mathbb{P}(X = \overline{X}) \ge \sum_{j=1}^{|K|} (a_j \vee \overline{a}_j - a_{j-1} \vee \overline{a}_{j-1}) + \sum_{j=1}^{|K|} (a_j \wedge \overline{a}_j - a_j \vee \overline{a}_j).$$

The first sum above is a telescopic sum. Since $a_{|K|} = \overline{a}_{|K|} = 1$ and $a_0 = \overline{a}_0 = 0$, we have

$$\mathbb{P}(X = \overline{X}) \ge 1 - \sum_{j=1}^{|K|-1} |a_j - \overline{a}_j|. \qquad \square$$

Lemma 3.2. *Let $(a_n, n \ge 1)$ and $(b_n, n \ge 1)$ be two real-valued sequences. For all $n \ge 1$, we have*

$$\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i = \sum_{i=1}^{n} (a_i - b_i) \prod_{j=i+1}^{n} a_j \prod_{j=1}^{i-1} b_j.$$

Proof of Lemma 3.2. We use induction. $\square$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

3.2. Proof of Theorem 3.1

First, we write

$$\begin{aligned}&\mathbb{E}[|F(\overline{x}_T) - F(x_T)|^2]\\&\quad= \mathbb{E}[\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \le T} |F(\overline{x}_T) - F(x_T)|^2] + \mathbb{E}[\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T} |F(\overline{x}_T) - F(x_T)|^2]\\&\quad=: \overline{P} + \overline{D},\end{aligned}$$

where $\overline{\tau}^\dagger$ is defined in Definition 3.1. The order of the term $\overline{P}$ is the order of the probability that the discrete processes (T_n, θ_n) and $(\overline{T}_n, \overline{\theta}_n)$ differ on $[0, T]$. The order of the term $\overline{D}$ is given by the order of the Euler scheme squared, because the discrete processes (T_n, θ_n) and $(\overline{T}_n, \overline{\theta}_n)$ are equal on $[0, T]$. In the following we prove that $\overline{P} = \mathrm{O}(h)$ and that $\overline{D} = \mathrm{O}(h^2)$.

Step 1: estimation of $\overline{P}$. The function F being bounded, we have

$$\overline{P} \le 4M_F^2 \mathbb{P}(\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \le T),$$

where $M_F > 0$. Moreover, for $k \ge 1$,

$$\{\overline{\tau}^\dagger = k\} = \{\overline{\tau}^\dagger > k-1\} \bigcap \{(\tau_k, \theta_k) \ne (\overline{\tau}_k, \overline{\theta}_k)\}.$$

Hence

$$\begin{aligned}\mathbb{P}(\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \le T) &= \sum_{k\ge1} \mathbb{E}[\mathbb{1}_{\min(T_k, \overline{T}_k) \le T} \mathbb{1}_{\overline{\tau}^\dagger = k}]\\&= \sum_{k\ge1} \mathbb{E}[\mathbb{1}_{\min(T_k, \overline{T}_k) \le T} \mathbb{1}_{\overline{\tau}^\dagger > k-1} \mathbb{1}_{(\tau_k, \theta_k) \ne (\overline{\tau}_k, \overline{\theta}_k)}]\\&\le \sum_{k\ge1} \overline{J}_k + 2\overline{I}_k,\end{aligned}$$

where

$$\begin{aligned}\overline{J}_k &:= \mathbb{E}[\mathbb{1}_{\min(T_k, \overline{T}_k) \le T} \mathbb{1}_{\overline{\tau}^\dagger > k-1} \mathbb{1}_{\tau_k = \overline{\tau}_k} \mathbb{1}_{\theta_k \ne \overline{\theta}_k}],\\\overline{I}_k &:= \mathbb{E}[\mathbb{1}_{\min(T_k, \overline{T}_k) \le T} \mathbb{1}_{\overline{\tau}^\dagger > k-1} \mathbb{1}_{\tau_k \ne \overline{\tau}_k}].\end{aligned} \tag{3.2}$$

We start with $\overline{J}_k$. First note that, for $k \ge 1$, $\{\tau_k = \overline{\tau}_k\} = \{T_k = \overline{T}_k\}$, and that on the event $\{T_k = \overline{T}_k\}$, we have $\min(T_k, \overline{T}_k) = T_k$, so

$$\overline{J}_k = \mathbb{E}[\mathbb{1}_{T_k \le T} \mathbb{1}_{\overline{\tau}^\dagger > k-1} \mathbb{1}_{\tau_k = \overline{\tau}_k} \mathbb{1}_{\theta_k \ne \overline{\theta}_k}].$$

We emphasize that it makes no difference in the rest of the proof if we choose $\min(T_k, \overline{T}_k) = \overline{T}_k$. Since

$$\left\{\overline{\tau}^\dagger > k-1\right\} = \bigcap_{i=0}^{k-1} \{(\tau_i, \theta_i) = (\overline{\tau}_i, \overline{\theta}_i)\},$$

we can rewrite $\overline{J}_k$ as follows:

$$\sum_{\substack{1 \le p_1 < \ldots < p_k\\ \alpha_1, \ldots, \alpha_{k-1} \in \Theta}} \mathbb{E}[\mathbb{1}_{\{\tau_i = \overline{\tau}_i = p_i, 1 \le i \le k\}} \mathbb{1}_{\{\theta_i = \overline{\theta}_i = \alpha_i, 1 \le i \le k-1\}} \mathbb{1}_{T^*_{p_k} \le T} \mathbb{1}_{\theta_k \ne \overline{\theta}_k}]. \tag{3.3}$$

By construction we have $\theta_k = H((\theta_{k-1}, \nu_k), V_k)$ and $\overline{\theta}_k = H((\overline{\theta}_{k-1}, \overline{\nu}_k), V_k)$. The random variable $\mathbb{1}_{\{\tau_i = \overline{\tau}_i = p_i, 1 \le i \le k\}} \mathbb{1}_{\{\theta_i = \overline{\theta}_i = \alpha_i, 1 \le i \le k-1\}} \mathbb{1}_{T^*_{p_k} \le T}$ depends on the vector $(U_i, 1 \le i \le$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

$p_k, T_j^*, 1 \le j \le p_k, V_q, 1 \le q \le k-1)$, which is independent of V_k. Conditioning by this vector in (3.3) and applying Lemma 3.1 yields

$$\mathbb{E}[\mathbb{1}_{\{\tau_i=\overline{\tau}_i=p_i, 1\le i\le k\}}\mathbb{1}_{\{\theta_i=\overline{\theta}_i=\alpha_i, 1\le i\le k-1\}}\mathbb{1}_{T_{p_k}^*\le T}\mathbb{1}_{\theta_k\neq\overline{\theta}_k}]$$
$$\le \mathbb{E}\left[\mathbb{1}_{\{\tau_i=\overline{\tau}_i=p_i, 1\le i\le k\}}\mathbb{1}_{\{\theta_i=\overline{\theta}_i=\alpha_i, 1\le i\le k-1\}}\mathbb{1}_{T_{p_k}^*\le T}\sum_{j=1}^{|\Theta|-1}|a_j(\alpha_{k-1}, \overline{\nu}_k) - a_j(\alpha_{k-1}, \nu_k)|\right].$$

From the definition of a_j (see (2.4)), the triangle inequality, and since Q is L_Q-Lipschitz, we have

$$\sum_{j=1}^{|\Theta|-1}|a_j(\alpha_{k-1}, \overline{\nu}_k) - a_j(\alpha_{k-1}, \nu_k)| \le \frac{(|\Theta|-1)|\Theta|}{2}L_Q|\overline{\nu}_k - \nu_k|.$$

Since we are on the event

$$\{\tau_i = \overline{\tau}_i = p_i, 1 \le i \le k\} \bigcap \{\theta_i = \overline{\theta}_i = \alpha_i, 1 \le i \le k-1\},$$

the application of Lemma 2.1 yields $|\overline{\nu}_k - \nu_k| \le \mathrm{e}^{LT_{p_k}^*}kCh$. Thus $\overline{J}_k \le C_1 h\mathbb{E}[\mathbb{1}_{T_k\le T}k]$, where C_1 is a constant independent of h. Moreover,

$$\sum_{k\ge1}\mathbb{1}_{T_k\le T}k = \sum_{k=1}^{N_T}k \le N_T^2 \quad \text{and} \quad \mathbb{E}[N_T^2] \le \mathbb{E}[(N_T^*)^2] < +\infty$$

so that

$$\sum_{k\ge1}\overline{J}_k = \mathrm{O}(h).$$

From the definition of $\overline{I}_k$ (see (3.2)), we can write

$$\begin{aligned}\overline{I}_k &= \mathbb{E}[\mathbb{1}_{\min(T_k,\overline{T}_k)\le T}\mathbb{1}_{\overline{\tau}^\dagger>k-1}(\mathbb{1}_{\tau_k<\overline{\tau}_k} + \mathbb{1}_{\tau_k>\overline{\tau}_k})]\\ &= \mathbb{E}[\mathbb{1}_{T_k\le T}\mathbb{1}_{\overline{\tau}^\dagger>k-1}\mathbb{1}_{\tau_k<\overline{\tau}_k}] + \mathbb{E}[\mathbb{1}_{\overline{T}_k\le T}\mathbb{1}_{\overline{\tau}^\dagger>k-1}\mathbb{1}_{\tau_k>\overline{\tau}_k}]\\ &=: \overline{I}_k^{(1)} + \overline{I}_k^{(2)}.\end{aligned}$$

The second equality above follows since $\{\tau_k < \overline{\tau}_k\} = \{T_k < \overline{T}_k\}$ and $\{\tau_k > \overline{\tau}_k\} = \{T_k > \overline{T}_k\}$. We only treat the term $\overline{I}_k^{(1)}$; the term $\overline{I}_k^{(2)}$ can be treated similarly by swapping (τ_k, T_k) and $(\overline{\tau}_k, \overline{T}_k)$. Just as in the previous case, we can rewrite $\overline{I}_k^{(1)}$ as follows:

$$\sum_{\substack{1\le p_1<\ldots<p_k\\ \alpha_1,\ldots,\alpha_{k-1}\in\Theta}}\mathbb{E}[\mathbb{1}_{\{\tau_i=\overline{\tau}_i=p_i, 1\le i\le k-1\}}\mathbb{1}_{\{\theta_i=\overline{\theta}_i=\alpha_i, 1\le i\le k-1\}}\mathbb{1}_{T_{p_k}^*\le T}\mathbb{1}_{\tau_k=p_k}\mathbb{1}_{p_k<\overline{\tau}_k}]. \tag{3.4}$$

In (3.4) we have

$$\{\tau_k = p_k\} \cap \{p_k < \overline{\tau}_k\}$$
$$\subseteq \{\lambda(\alpha_{k-1}, \overline{\Phi}_{\alpha_{k-1}}(T_{p_k}^* - T_{p_{k-1}}^*, \overline{\nu}_{k-1})) < U_{p_k}\lambda^* \le \lambda(\alpha_{k-1}, \Phi_{\alpha_{k-1}}(T_{p_k}^* - T_{p_{k-1}}^*, \nu_{k-1}))\}.$$

The random variable $\mathbb{1}_{\{\tau_i=\overline{\tau}_i=p_i, 1\le i\le k-1\}}\mathbb{1}_{\{\theta_i=\overline{\theta}_i=\alpha_i, 1\le i\le k-1\}}\mathbb{1}_{T_{p_k}^*\le T}$ depends on

$$(U_i, 1 \le i \le p_{k-1}, T_j^*, 1 \le j \le p_k, V_q, 1 \le q \le k-1),$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

which is independent of U_{p_k}. Conditioning by this vector in (3.4) yields

$$\begin{aligned}&\mathbb{E}[\mathbb{1}_{\{\tau_i=\overline{\tau}_i=p_i,1\le i\le k-1\}}\mathbb{1}_{\{\theta_i=\overline{\theta}_i=\alpha_i,1\le i\le k-1\}}\mathbb{1}_{T^*_{p_k}\le T}\mathbb{1}_{\tau_k=p_k}\mathbb{1}_{p_k<\overline{\tau}_k}]\\&\quad\le\mathbb{E}[\mathbb{1}_{\{\tau_i=\overline{\tau}_i=p_i,1\le i\le k-1\}}\mathbb{1}_{\{\theta_i=\overline{\theta}_i=\alpha_i,1\le i\le k-1\}}\mathbb{1}_{T^*_{p_k}\le T}\\&\qquad\times|\lambda(\alpha_{k-1},\overline{\Phi}_{\alpha_{k-1}}(T^*_{p_k}-T^*_{p_{k-1}},\overline{\nu}_{k-1}))-\lambda(\alpha_{k-1},\Phi_{\alpha_{k-1}}(T^*_{p_k}-T^*_{p_{k-1}},\nu_{k-1}))|].\end{aligned}$$

Using the Lipschitz continuity of λ and then Lemma 2.1, we obtain

$$\overline{I}^{(1)}_k\le C_2h\mathbb{E}[\mathbb{1}_{T_k\le T}k],$$

where C_2 is a constant independent of h. Concerning the term $\overline{I}^{(2)}_k$, we will end with the estimate $\overline{I}^{(2)}_k\le C_2h\mathbb{E}[\mathbb{1}_{\overline{T}_k\le T}k]$. We conclude in the same way as in the estimation of $\overline{J}_k$ above that $\sum_{k\ge1}\overline{I}_k=\mathrm{O}(h)$.

Step 2: estimation of $\overline{D}$. Note that for $n\ge0$ we have

$$\{N_T=n\}\cap\{\min(T_{\overline{\tau}^\dagger},\overline{T}_{\overline{\tau}^\dagger})>T\}=\{N_T=n\}\cap\{\overline{N}_T=n\}\cap\{\overline{\tau}^\dagger>n\},$$

where we can swap $\{N_T=n\}$ and $\{\overline{N}_T=n\}$. Thus, using the partition $\{N_T=n,n\ge0\}$, we have

$$\overline{D}=\sum_{n\ge0}\mathbb{E}[\mathbb{1}_{N_T=n}\mathbb{1}_{\overline{N}_T=n}\mathbb{1}_{\overline{\tau}^\dagger>n}|F(\theta_n,\overline{\Phi}_{\theta_n}(T-T_n,\overline{\nu}_n))-F(\theta_n,\Phi_{\theta_n}(T-T_n,\nu_n))|^2].$$

Application of the Lipschitz continuity of F and of Lemma 2.1 yields

$$|F(\theta_n,\overline{\Phi}_{\theta_n}(T-T_n,\overline{\nu}_n))-F(\theta_n,\Phi_{\theta_n}(T-T_n,\nu_n))|\le L_F\,\mathrm{e}^{LT}(n+1)Ch.$$

Then we have $\overline{D}\le C_3h^2\sum_{n\ge0}\mathbb{E}[\mathbb{1}_{N_T=n}(n+1)^2]$, where C_3 is a constant independent of h. Since

$$\sum_{n\ge0}\mathbb{E}[\mathbb{1}_{N_T=n}(n+1)^2]=\mathbb{E}[(N_T+1)^2]\le\mathbb{E}[(N^*_T+1)^2]<+\infty,$$

we conclude that $\overline{D}=\mathrm{O}(h^2)$.

3.3. Proof of Theorem 3.2

First we reorder the terms in $\tilde{R}_T$. We write $\tilde{R}_T=\tilde{\mathsf{Q}}_T\tilde{\mathsf{S}}_T\tilde{\mathsf{H}}_T$, where

$$\tilde{\mathsf{Q}}_T=\prod_{l=1}^{\tilde{N}_T}\frac{Q(\tilde{x}^-_{T^*_{\tilde{\tau}_l}},\tilde{\theta}_l)}{\tilde{Q}(\tilde{\theta}_{l-1},\tilde{\theta}_l)},\tag{3.5}$$

$$\tilde{\mathsf{S}}_T=\prod_{l=1}^{\tilde{N}_T}\frac{\lambda(\tilde{\theta}_{l-1},\Phi_{\tilde{\theta}_{l-1}}(T^*_{\tilde{\tau}_l}-T^*_{\tilde{\tau}_{l-1}},\tilde{\nu}_{l-1}))}{\lambda^*}\prod_{k=\tilde{\tau}_{l-1}+1}^{\tilde{\tau}_l}\left(1-\frac{\lambda(\tilde{\theta}_{l-1},\Phi_{\tilde{\theta}_{l-1}}(T^*_k-T^*_{\tilde{\tau}_{l-1}},\tilde{\nu}_{l-1}))}{\lambda^*}\right)\tag{3.6}$$

$$\times\prod_{l=\tilde{\tau}_{\tilde{N}_T}+1}^{N^*_T}\left(1-\frac{\lambda(\tilde{\theta}_{\tilde{N}_T},\Phi_{\tilde{\theta}_{\tilde{N}_T}}(T^*_l-T^*_{\tilde{\tau}_{\tilde{N}_T}},\tilde{\nu}_{\tilde{N}_T}))}{\lambda^*}\right),$$

$$\tilde{\mathsf{H}}_T=\prod_{l=1}^{\tilde{N}_T}\left(\frac{\tilde{\lambda}(\tilde{\theta}_{l-1})}{\lambda^*}\left(1-\frac{\tilde{\lambda}(\tilde{\theta}_{l-1})}{\lambda^*}\right)^{\tilde{\tau}_l-\tilde{\tau}_{l-1}-1}\right)^{-1}\left(\left(1-\frac{\tilde{\lambda}(\tilde{\theta}_{\tilde{N}_T})}{\lambda^*}\right)^{N^*_T-\tilde{\tau}_{\tilde{N}_T}}\right)^{-1}.\tag{3.7}$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Likewise we reorder the terms in $\tilde{\underline{R}}_T$, writing $\tilde{\underline{R}}_T = \tilde{\underline{\mathsf{Q}}}_T \tilde{\underline{\mathsf{S}}}_T \tilde{\mathsf{H}}_T$, where $\tilde{\underline{\mathsf{Q}}}_T$ and $\tilde{\underline{\mathsf{S}}}_T$ are defined as (3.5) and (3.6) replacing $\tilde{x}$ and Φ by $\underline{\tilde{x}}$ and $\overline{\Phi}$. Since the processes $(\tilde{\theta}_n)$ and $(\tilde{\tau}_n)$ do not depend on Φ or $\overline{\Phi}$, the term $\tilde{\mathsf{H}}$ is the same in $\tilde{R}$ and $\tilde{\underline{R}}$. To prove Theorem 3.2, let us decompose the problem and write

$$\begin{aligned}|F(\underline{\tilde{x}}_T)\tilde{\underline{R}}_T - F(\tilde{x}_T)\tilde{R}_T| &= |(F(\underline{\tilde{x}}_T) - F(\tilde{x}_T))\tilde{\underline{R}}_T + (\tilde{\underline{R}}_T - \tilde{R}_T)F(\tilde{x}_T)| \\ &\le |F(\underline{\tilde{x}}_T) - F(\tilde{x}_T)||\tilde{\underline{R}}_T| + |\tilde{\underline{R}}_T - \tilde{R}_T||F(\tilde{x}_T)|,\end{aligned}$$

so that

$$\begin{aligned}\mathbb{E}[|F(\underline{\tilde{x}}_T)\tilde{\underline{R}}_T - F(\tilde{x}_T)\tilde{R}_T|^2] &\le 2\mathbb{E}[|F(\underline{\tilde{x}}_T) - F(\tilde{x}_T)|^2|\tilde{\underline{R}}_T|^2] + 2\mathbb{E}[|\tilde{\underline{R}}_T - \tilde{R}_T|^2|F(\tilde{x}_T)|^2] \\ &=: 2\overline{D} + 2\overline{C}.\end{aligned}$$

In the following we show that $\overline{C} = \mathrm{O}(h^2)$ and $\overline{D} = \mathrm{O}(h^2)$.

Step 1: estimation of $\overline{C}$. The function F being bounded, we have

$$\overline{C} \le M_F^2 \mathbb{E}[|\tilde{\underline{R}}_T - \tilde{R}_T|^2],$$

where M_F is a positive constant. Moreover, for all $\theta \in \Theta$ we have

$$(1 - \tilde{\lambda}(\theta)/\lambda^*)^{-1} \le (1 - \tilde{\lambda}_{\max}/\lambda^*)^{-1} \quad \text{and} \quad (\tilde{\lambda}(\theta)/\lambda^*)^{-1} \le (\tilde{\lambda}_{\min}/\lambda^*)^{-1}.$$

Thus

$$\tilde{H}_T \le \left(\frac{\tilde{\lambda}_{\min}}{\lambda^*}\left(1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*}\right)\right)^{-N_T^*},$$

and using the definition of $\tilde{R}$ and $\tilde{\underline{R}}$ (see (3.5), (3.6), and (3.7)) we can write

$$|\tilde{\underline{R}}_T - \tilde{R}_T| \le \left(\frac{\tilde{\lambda}_{\min}}{\lambda^*}\left(1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*}\right)\right)^{-N_T^*} (|\tilde{\underline{\mathsf{Q}}}_T - \tilde{\mathsf{Q}}_T|\tilde{\mathsf{S}}_T + |\tilde{\underline{\mathsf{S}}}_T - \tilde{\mathsf{S}}_T|\tilde{\underline{\mathsf{Q}}}_T).$$

We set $\overline{J} = |\tilde{\underline{\mathsf{Q}}}_T - \tilde{\mathsf{Q}}_T|\tilde{\mathsf{S}}_T$ and $\overline{I} = |\tilde{\underline{\mathsf{S}}}_T - \tilde{\mathsf{S}}_T|\tilde{\underline{\mathsf{Q}}}_T$. To provide the desired estimate for $\overline{C}$, we proceed as follows. First, we work ω by ω to determine (random) bounds for $\overline{J}$ and $\overline{I}$, from which we deduce a (random) bound for $|\tilde{\underline{R}}_T - \tilde{R}_T|$. Finally, we take the expectation. We start with $\overline{I}$. For all $(\theta, \nu) \in E$ and for all $t \ge 0$, we have from Assumption 2.1 that $1 - \lambda(\theta, \Phi_\theta(t, \nu))/\lambda^* \le 1$ and $\lambda(\theta, \Phi_\theta(t, \nu))/\lambda^* \le 1$. Then, using Lemma 3.2 (twice), we have

$$\begin{aligned}|\tilde{\underline{\mathsf{S}}}_T - \tilde{\mathsf{S}}_T| \le \frac{1}{\lambda^*} \sum_{l=1}^{\tilde{N}_T+1} \sum_{k=\tilde{\tau}_{l-1}+1}^{\tilde{\tau}_l \wedge N_T^*} &|\lambda(\tilde{\theta}_{l-1}, \overline{\Phi}_{\tilde{\theta}_{l-1}}(T_k^* - T_{\tilde{\tau}_{l-1}}^*, \underline{\tilde{\nu}}_{l-1})) \\ &- \lambda(\tilde{\theta}_{l-1}, \Phi_{\tilde{\theta}_{l-1}}(T_k^* - T_{\tilde{\tau}_{l-1}}^*, \tilde{\nu}_{l-1}))|.\end{aligned}$$

Using the Lipschitz continuity of λ and Lemma 2.1, we find that, for all $l = 1, \ldots, \tilde{N}_T + 1$ and $k = \tilde{\tau}_{l-1} + 1, \ldots, \tilde{\tau}_l \wedge N_T^*$,

$$|\lambda(\tilde{\theta}_{l-1}, \overline{\Phi}_{\tilde{\theta}_{l-1}}(T_k^* - T_{\tilde{\tau}_{l-1}}^*, \underline{\tilde{\nu}}_{l-1})) - \lambda(\tilde{\theta}_{l-1}, \Phi_{\tilde{\theta}_{l-1}}(T_k^* - T_{\tilde{\tau}_{l-1}}^*, \tilde{\nu}_{l-1}))| \le \mathrm{e}^{LT} Chl.$$

Moreover, for all $l = 1, \ldots, \tilde{N}_T + 1$ we have $\tilde{\tau}_l \wedge N_T^* - \tilde{\tau}_{l-1} \le N_T^*$, so that

$$|\tilde{\underline{\mathsf{S}}}_T - \tilde{\mathsf{S}}_T| \le N_T^*(N_T^* + 1)^2 C_1 h,$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

where C_1 is a positive constant independent of h. Finally, since $\underline{\tilde{\mathsf{Q}}}_T \le \rho^{-N_T^*}$ we have

$$\bar{I} \le \rho^{-N_T^*} N_T^* (N_T^* + 1)^2 C_1 h. \tag{3.8}$$

Now, consider $\bar{J}$. Note that from Assumption 2.1 we have $\tilde{\mathsf{S}}_T \le 1$. We use the same type of arguments as for $\bar{I}$. That is, we successively use Lemma 3.2, the Lipschitz continuity of Q, and Lemma 2.1 to obtain

$$\bar{J} \le \rho^{-N_T^*} (N_T^*)^2 C_2 h, \tag{3.9}$$

where C_2 is a positive constant independent of h. Then we derive from the previous estimates (3.8) and (3.9) that

$$|\underline{\tilde{R}}_T - \tilde{R}_T| \le \Xi_1(N_T^*) C_3 h,$$

where

$$\Xi_1(n) = \left(\rho \frac{\tilde{\lambda}_{\min}}{\lambda^*}\left(1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*}\right)\right)^{-n} n(n+1)^2 \quad \text{and} \quad C_3 = \max(C_1, C_2).$$

Finally, we have $\mathbb{E}[|\underline{\tilde{R}}_T - \tilde{R}_T|^2] \le C_3 h^2 \mathbb{E}[\Xi_1(N_T^*)^2]$. Since

$$\mathbb{E}[\Xi_1(N_T^*)^2] < +\infty,$$

we conclude that $\overline{C} = \mathrm{O}(h^2)$.

Step 2: estimation of $\overline{D}$. Recall that $\tilde{x}_T = (\tilde{\theta}_{\tilde{N}_T}, \Phi_{\tilde{\theta}_{\tilde{N}_T}}(T - \tilde{T}_{\tilde{N}_T}, \tilde{\nu}_{\tilde{N}_T}))$ and $\underline{\tilde{x}}_T = (\tilde{\theta}_{\tilde{N}_T}, \overline{\Phi}_{\tilde{\theta}_{\tilde{N}_T}}(T - \tilde{T}_{\tilde{N}_T}, \underline{\tilde{\nu}}_{\tilde{N}_T}))$. Then, using the Lipschitz continuity of F, Lemma 2.1 and since $\tilde{N}_T \le N_T^*$, we get

$$|F(\underline{\tilde{x}}_T) - F(\tilde{x}_T)| \le L_F \, \mathrm{e}^{LT} (\tilde{N}_T + 1) C h \le L_F \, \mathrm{e}^{LT} (N_T^* + 1) C h.$$

Moreover,

$$|\underline{\tilde{R}}_T| \le \left(\rho \frac{\tilde{\lambda}_{\min}}{\lambda^*}\left(1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*}\right)\right)^{-N_T^*},$$

so that $\overline{D} \le C_4 h^2 \mathbb{E}[\Xi_2(N_T^*)^2]$, where C_4 is a positive constant independent of h and

$$\Xi_2(n) = (n+1)\left(\rho \frac{\tilde{\lambda}_{\min}}{\lambda^*}\left(1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*}\right)\right)^{-n}.$$

Since $\mathbb{E}[\Xi_2(N_T^*)^2] < +\infty$, we conclude that $\overline{D} = \mathrm{O}(h^2)$.

4. Weak error expansion

In this section we are interested in a weak error expansion for the PDMP (x_t) of Section 2.3 and its associated Euler scheme $(\bar{x}_t)$. First of all, we recall from [5] that the generator $\mathcal{A}$ of the process (t, x_t) which acts on functions g defined on $\mathbb{R}_+ \times E$ is given by

$$\mathcal{A}g(t,x) = \partial_t g(t,x) + f(x)\partial_v g(t,x) + \lambda(x) \int_E (g(t,y) - g(t,x)) Q(x, \mathrm{d}y), \tag{4.1}$$

where for notational convenience we have set

$$\partial_v g(t,x) := \frac{\partial g}{\partial v}(t,\theta,v), \quad \partial_t g(t,x) := \frac{\partial g}{\partial t}(t,x), \quad \text{and} \quad f(x) = f_\theta(v) \quad \text{for all } x = (\theta, v) \in E.$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Below, we state the assumptions and the main theorem of this section. Its proof, which is inspired by [27] (see also [24] or [16]), is deferred to Section 4.2.

Assumption 4.1. *For all $\theta \in \Theta$ and for all $A \in \mathcal{B}(\Theta)$, the functions $\nu \mapsto Q((\theta, \nu), A)$, $\nu \mapsto \lambda(\theta, \nu)$, and $\nu \mapsto f_\theta(\nu)$ are bounded and twice continuously differentiable with bounded derivatives.*

Assumption 4.2. *The solution u of the integro-differential equation*

$$\begin{cases} \mathcal{A}u(t, x) = 0 & (t, x) \in [0, T[\times E, \\ u(T, x) = F(x) & x \in E, \end{cases} \tag{4.2}$$

with $F\colon E \to \mathbb{R}$ a bounded function and $\mathcal{A}$ given by (4.1), is such that for all $\theta \in \Theta$, the function $(t, \nu) \mapsto u(t, \theta, \nu)$ is bounded and twice differentiable with bounded derivatives. Moreover, the second derivatives of $(t, \nu) \mapsto u(t, \theta, \nu)$ are uniformly Lipschitz in θ.

Theorem 4.1. *Let $(x_t, t \in [0, T])$ be a PDMP and $(\bar{x}_t, t \in [0, T])$ its approximation constructed in Section 2.3 with $x_0 = \bar{x}_0 = x$ for some $x \in E$. Under Assumptions 4.1 and 4.2, for any bounded function $F\colon E \to \mathbb{R}$ there exists a constant c_1 independent of h such that*

$$\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = hc_1 + \mathrm{O}(h^2). \tag{4.3}$$

Remark 4.1. If $(\tilde{x}_t)$ is a PDMP whose characteristics $\tilde{\lambda}, \tilde{Q}$ satisfy the assumptions of Proposition 2.2 and $(\underline{\tilde{x}}_t)$ is its approximation, we deduce from Theorem 4.1 that

$$\mathbb{E}[F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T] - \mathbb{E}[F(\tilde{x}_T)\tilde{R}_T] = hc_1 + \mathrm{O}(h^2). \tag{4.4}$$

4.1. Further results on PDMPs: Itô and Feynman–Kac formulas

Definition 4.1. Let us define the following operators, which act on functions g defined on $\mathbb{R}_+ \times E$:

$$\mathcal{T}g(t, x) := \partial_t g(t, x) + f(x)\partial_v g(t, x),$$
$$\mathcal{S}g(t, x) := \lambda(x) \int_E (g(t, y) - g(t, x))Q(x, \mathrm{d}y).$$

From Definition 4.1, the generator $\mathcal{A}$ defined by (4.1) reads $\mathcal{A}g(t, x) = \mathcal{T}g(t, x) + \mathcal{S}g(t, x)$. We introduce the random counting measure p associated with the PDMP (x_t) defined by $p([0, t] \times A) := \sum_{n \geq 1} \mathbb{1}_{T_n \leq t} \mathbb{1}_{Y_n \in A}$ for $t \in [0, T]$ and for $A \in \mathcal{B}(E)$. The compensator of p, denoted by p', is given from [5] by

$$p'([0, t] \times A) = \int_0^t \lambda(x_s)Q(x_s, A)\,\mathrm{d}s.$$

Hence, $q := p - p'$ is a martingale with respect to the filtration generated by p, denoted by $(\mathcal{F}_t^p)_{t \in [0,T]}$. Similarly, we introduce $\overline{p}, \overline{p}', \overline{q}$, and $(\mathcal{F}_t^{\overline{p}})_{t \in [0,T]}$ to be the same objects as above but corresponding to the approximation $(\bar{x}_t)$. The fact that $\overline{p}'$ is the compensator of $\overline{p}$ and that $\overline{q}$ is a martingale derives from arguments of the marked point processes theory: see [4].

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Definition 4.2. Let us define the following operators, which act on functions g defined on $\mathbb{R}_+ \times E$:

$$\overline{\mathcal{T}} g(t, x, y) := \partial_t g(t, x) + f(y)\partial_v g(t, x),$$
$$\overline{\mathcal{A}} g(t, x, y) := \overline{\mathcal{T}} g(t, x, y) + \mathcal{S} g(t, x).$$

Remark 4.2. For all functions g defined on $\mathbb{R}_+ \times E$, $\overline{\mathcal{T}} g(t, x, x) = \mathcal{T} g(t, x)$, so that $\overline{\mathcal{A}} g(t, x, x) = \mathcal{A} g(t, x)$.

The next theorem provides Itô formulas for the PDMP (x_t) and its approximation $(\bar{x}_t)$. For all $s \in [0, T]$, we set $\overline{\eta}(s) := \overline{T}_n + kh$ if $s \in [\overline{T}_n + kh, (\overline{T}_n + (k+1)h) \wedge \overline{T}_{n+1}[$ for some $n \geq 0$ and for some $k \in \{0, \ldots, \lfloor (\overline{T}_{n+1} - \overline{T}_n)/h \rfloor\}$.

Theorem 4.2. *Let $(x_t, t \in [0, T])$ and $(\bar{x}_t, t \in [0, T])$ be a PDMP and its approximation, respectively, constructed in Section 2.3 with $x_0 = \bar{x}_0 = x$ for some $x \in E$. For all bounded functions $g\colon \mathbb{R}_+ \times E \to \mathbb{R}$ continuously differentiable with bounded derivatives, we have*

$$g(t, x_t) = g(0, x) + \int_0^t \mathcal{A} g(s, x_s)\, \mathrm{d}s + M_t^g, \tag{4.5}$$

where

$$M_t^g := \int_0^t \int_E (g(s, y) - g(s, x_{s-}))q(\mathrm{d}s\, \mathrm{d}y)$$

is a true $\mathcal{F}_t^p$-martingale, and

$$g(t, \bar{x}_t) = g(0, x) + \int_0^t \overline{\mathcal{A}} g(s, \bar{x}_s, \bar{x}_{\overline{\eta}(s)})\, \mathrm{d}s + \overline{M}_t^g, \tag{4.6}$$

where

$$\overline{M}_t^g := \int_0^t \int_E (g(s, y) - g(s, \bar{x}_{s-}))\overline{q}(\mathrm{d}s\, \mathrm{d}y)$$

is a true $\mathcal{F}_t^{\overline{p}}$-martingale.

Proof of Theorem 4.2. The proof of (4.5) is given in [5]. We prove (4.6) following the same arguments. Since $\overline{q} = \overline{p} - \overline{p}'$, we have

$$\overline{M}_t^g = \sum_{k \geq 1} \mathbb{1}_{\overline{T}_k \leq t} (g(\overline{T}_k, \bar{x}_{\overline{T}_k}) - g(\overline{T}_k, \bar{x}_{\overline{T}_k}^-)) - \int_0^t \mathcal{S} g(s, \bar{x}_s)\, \mathrm{d}s.$$

Consider the above sum. As in [5], we write, on the event $\{\overline{N}_t = n\}$, that

$$\sum_{k \geq 1} \mathbb{1}_{\overline{T}_k \leq t} (g(\overline{T}_k, \bar{x}_{\overline{T}_k}) - g(\overline{T}_k, \bar{x}_{\overline{T}_k}^-))$$
$$= g(t, \bar{x}_t) - g(0, x) - \left[g(t, \bar{x}_t) - g(\overline{T}_n, \bar{x}_{\overline{T}_n}) + \sum_{k=0}^{n-1} g(\overline{T}_{k+1}, \bar{x}_{\overline{T}_{k+1}}^-) - g(\overline{T}_k, \bar{x}_{\overline{T}_k}) \right].$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

For all $k \le n-1$, we decompose the increment

$$g(\overline{T}_{k+1}, \overline{x}^{-}_{\overline{T}_{k+1}}) - g(\overline{T}_k, \overline{x}_{\overline{T}_k})$$

as a sum of increments on the intervals $[\overline{T}_k + ih, (\overline{T}_k + (i+1)h) \wedge \overline{T}_{k+1}] \subset [\overline{T}_k, \overline{T}_{k+1}]$. Without loss of generality we are led to consider increments of the form $g(t, \theta, \overline{\phi}_\theta(t, \nu)) - g(ih, \theta, \overline{y}_i(x))$ for some $i \ge 0$, $t \in [ih, (i+1)h]$ and for all $x = (\theta, \nu) \in E$, where we recall that $\overline{\phi}$ is defined by (2.8). The function g is smooth enough to write

$$g(t, \theta, \overline{\phi}_\theta(t, \nu)) - g(ih, \theta, \overline{y}_i(x)) = \int_{ih}^{t} (\partial_t g + f_\theta(\overline{y}_i(x))\partial_\nu g)(s, \theta, \overline{\phi}_\theta(s, \nu))\, \mathrm{d}s.$$

Then the above arguments together with Definition 4.2 yield

$$g(t, \overline{x}_t) - g(\overline{T}_n, \overline{x}_{\overline{T}_n}) + \sum_{k=0}^{n-1} g(\overline{T}_{k+1}, \overline{x}^{-}_{\overline{T}_{k+1}}) - g(\overline{T}_k, \overline{x}_{\overline{T}_k}) = \int_0^t \overline{\mathcal{T}} g(s, \overline{x}_s, \overline{x}_{\overline{\eta}(s)})\, \mathrm{d}s. \qquad \square$$

The following theorem gives us a way to represent the solution of the integro-differential equation (4.6) as the conditional expected value of a functional of the terminal value of the PDMP (x_t). It plays a key role in the proof of Theorem 4.1.

Theorem 4.3. (PDMP's Feynman–Kac formula [6].) *Let $F\colon E \to \mathbb{R}$ be a bounded function. Then the integro-differential equation (4.2) has a unique solution $u\colon \mathbb{R}_+ \times E \to \mathbb{R}$ given by*

$$u(t, x) = \mathbb{E}[F(x_T) | x_t = x], \quad (t, x) \in [0, T] \times E.$$

4.2. Proof of Theorem 4.1

We provide a proof in two steps. First, we give an appropriate representation of the weak error $\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)]$. Then we use this representation to identify the coefficient c_1 in (4.3).

Step 1: representing $\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)]$. Let u denote the solution of (4.2). From Theorem 4.3 we can write

$$\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)] = \mathbb{E}[u(T, \overline{x}_T)] - u(0, x).$$

Then, the application of the Itô formula (4.6) to u at time T yields

$$u(T, \overline{x}_T) = u(0, x) + \int_0^T \overline{\mathcal{A}} u(s, \overline{x}_s, \overline{x}_{\overline{\eta}(s)})\, \mathrm{d}s + \overline{M}^u_T.$$

Since $(\overline{M}^u_t)$ is a true martingale, we obtain

$$\mathbb{E}[u(T, \overline{x}_T) - u(0, x)] = \mathbb{E}\left[\int_0^T \overline{\mathcal{A}} u(s, \overline{x}_s, \overline{x}_{\overline{\eta}(s)})\, \mathrm{d}s \right].$$

For $s \in [0, T]$ we have $\overline{\mathcal{A}} u(s, \overline{x}_s, \overline{x}_{\overline{\eta}(s)}) = \partial_t u(s, \overline{x}_s) + f(\overline{x}_{\overline{\eta}(s)})\partial_\nu u(s, \overline{x}_s) + \mathcal{S} u(s, \overline{x}_s)$ (see Definition 4.2). From the regularity of λ, Q, and u (see Assumptions 4.1 and 4.2), the functions

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

$\partial_t u$, $\partial_v u$, and $\mathcal{S}u$ are smooth enough to apply the Itô formula (4.6) between $\overline{\eta}(s)$ and s respectively. This yields

$$\begin{aligned}
\partial_t u(s, \bar{x}_s) &= \partial_t u(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)}) + \int_{\overline{\eta}(s)}^{s} \overline{\mathcal{A}}(\partial_t u)(r, \bar{x}_r, \bar{x}_{\overline{\eta}(r)})\,\mathrm{d}r + \overline{M}_s^{\partial_t u} - \overline{M}_{\overline{\eta}(s)}^{\partial_t u},\\
\partial_v u(s, \bar{x}_s) &= \partial_v u(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)}) + \int_{\overline{\eta}(s)}^{s} \overline{\mathcal{A}}(\partial_v u)(r, \bar{x}_r, \bar{x}_{\overline{\eta}(r)})\,\mathrm{d}r + \overline{M}_s^{\partial_v u} - \overline{M}_{\overline{\eta}(s)}^{\partial_v u},\\
\mathcal{S}u(s, \bar{x}_s) &= \mathcal{S}u(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)}) + \int_{\overline{\eta}(s)}^{s} \overline{\mathcal{A}}(\mathcal{S}u)(r, \bar{x}_r, \bar{x}_{\overline{\eta}(r)})\,\mathrm{d}s + \overline{M}_s^{\mathcal{S}u} - \overline{M}_{\overline{\eta}(s)}^{\mathcal{S}u}.
\end{aligned}$$

Moreover, since $\overline{\eta}(r) = \overline{\eta}(s)$ for $r \in [\overline{\eta}(s), s]$, we have

$$\begin{aligned}
f(\bar{x}_{\overline{\eta}(s)})\partial_v u(s, \bar{x}_s) = {} & f(\bar{x}_{\overline{\eta}(s)})\partial_v u(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)})\\
& + \int_{\overline{\eta}(s)}^{s} f(\bar{x}_{\overline{\eta}(r)})\overline{\mathcal{A}}(\partial_v u)(r, \bar{x}_r, \bar{x}_{\overline{\eta}(r)})\,\mathrm{d}r + f(\bar{x}_{\overline{\eta}(s)})(\overline{M}_s^{\partial_v u} - \overline{M}_{\overline{\eta}(s)}^{\partial_v u}),
\end{aligned}$$

so that

$$\begin{aligned}
\overline{\mathcal{A}}u(s, \bar{x}_s, \bar{x}_{\overline{\eta}(s)}) = {} & \overline{\mathcal{A}}u(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)}, \bar{x}_{\overline{\eta}(s)}) + \int_{\overline{\eta}(s)}^{s} \Upsilon(r, \bar{x}_r, \bar{x}_{\overline{\eta}(r)})\,\mathrm{d}r\\
& + \overline{M}_s^{\partial_t u} - \overline{M}_{\overline{\eta}(s)}^{\partial_t u} + f(\bar{x}_{\overline{\eta}(s)})(\overline{M}_s^{\partial_v u} - \overline{M}_{\overline{\eta}(s)}^{\partial_v u}) + \overline{M}_s^{\mathcal{S}u} - \overline{M}_{\overline{\eta}(s)}^{\mathcal{S}u},
\end{aligned}$$

where

$$\Upsilon(t, x, y) := (\overline{\mathcal{A}}(\partial_t u) + f(y)\overline{\mathcal{A}}(\partial_v u) + \overline{\mathcal{A}}(\mathcal{S}u))(t, x, y). \tag{4.7}$$

Since $\overline{\mathcal{A}}u(t, x, x) = \mathcal{A}u(t, x)$, the first term in the above equality is 0 by Theorem 4.3. By using Fubini's theorem and the fact that $(\overline{M}_t^{\partial_t u})$ and $(\overline{M}_t^{\mathcal{S}u})$ are true martingales, we obtain

$$\mathbb{E}\left[\int_0^T \overline{M}_s^{\partial_t u} - \overline{M}_{\overline{\eta}(s)}^{\partial_t u}\,\mathrm{d}s\right] = \mathbb{E}\left[\int_0^T \overline{M}_s^{\mathcal{S}u} - \overline{M}_{\overline{\eta}(s)}^{\mathcal{S}u}\,\mathrm{d}s\right] = 0.$$

Moreover, since $(\overline{M}_t^{\partial_v u})$ is a $\mathcal{F}_t^{\overline{p}}$-martingale, we have

$$\mathbb{E}\left[\int_0^T f(\bar{x}_{\overline{\eta}(s)})(\overline{M}_s^{\partial_v u} - \overline{M}_{\overline{\eta}(s)}^{\partial_v u})\,\mathrm{d}s\right] = \int_0^T \mathbb{E}\left[f(\bar{x}_{\overline{\eta}(s)})\mathbb{E}[\overline{M}_s^{\partial_v u} - \overline{M}_{\overline{\eta}(s)}^{\partial_v u} | \mathcal{F}_{\overline{\eta}(s)}^{\overline{p}}]\right]\mathrm{d}s = 0.$$

Collecting the above results, we obtain

$$\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = \mathbb{E}\left[\int_0^T \int_{\overline{\eta}(s)}^{s} \Upsilon(r, \bar{x}_r, \bar{x}_{\overline{\eta}(r)})\,\mathrm{d}r\,\mathrm{d}s\right].$$

We can compute an explicit form of Υ in terms of u, f, λ, Q and their derivatives. Indeed, Υ is given by (4.7), and we have

$$\begin{aligned}
\overline{\mathcal{A}}(\partial_t u)(t, x, y) &= \partial_{tt}^2 u(t, x) + f(y)\partial_{tv}^2 u(t, x) + \mathcal{S}(\partial_t u)(t, x),\\
(f\overline{\mathcal{A}}(\partial_v u))(t, x, y) &= f(y)(\partial_{tv}^2 u(t, x) + f(y)\partial_{vv}^2 u(t, x) + \mathcal{S}(\partial_v u)(t, x)),\\
\overline{\mathcal{A}}(\mathcal{S}u)(t, x, y) &= \partial_t(\mathcal{S}u)(t, x) + f(y)\partial_v(\mathcal{S}u)(t, x) + \mathcal{S}(\mathcal{S}u)(t, x).
\end{aligned}$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Application of the Taylor formula to the functions $\partial^2_{tt}u$, $\partial^2_{tv}u$, $\partial^2_{vv}u$, $\mathcal{S}(\partial_t u)$, $\mathcal{S}(\partial_v u)$, $\partial_t(\mathcal{S}u)$, $\partial_v(\mathcal{S}u)$ and $\mathcal{S}(\mathcal{S}u)$ at the order 0 around $(\overline{\eta}(r), \bar{x}_{\overline{\eta}(r)})$ yields $\Upsilon(r, \bar{x}_r, \bar{x}_{\overline{\eta}(r)}) = \Upsilon(\overline{\eta}(r), \bar{x}_{\overline{\eta}(r)}, \bar{x}_{\overline{\eta}(r)}) + \mathrm{O}(h)$. Setting $\Psi(t, x) = \Upsilon(t, x, x)$ and recalling that for $r \in [\overline{\eta}(s), s]$, $\overline{\eta}(r) = \overline{\eta}(s)$ and that $|s - \overline{\eta}(s)| \leq h$, we obtain

$$\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = \mathbb{E}\left[\int_0^T (s - \overline{\eta}(s))\Psi(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)})\,\mathrm{d}s\right] + \mathrm{O}(h^2).$$

Consider the expectation on the right-hand side of the above equality. We decompose the integral into a (finite) sum of integrals on the intervals $[\overline{T}_n + kh, (\overline{T}_n + (k+1)h) \wedge \overline{T}_{n+1}]$, where Ψ is a constant. Without loss of generality, we are led to consider integrals of the form $\int_{kh}^t (s - kh)C\,\mathrm{d}s$ for some $k \geq 0$, $t \in [kh, (k+1)h]$ and C a bounded constant. We have

$$\int_{kh}^t (s - kh)C\,\mathrm{d}s = \frac{t - kh}{2}\int_{kh}^t C\,\mathrm{d}s,$$

and moreover adding and subtracting h in the numerator of $(t - kh)/2$ yields

$$\int_{kh}^t (s - kh)C\,\mathrm{d}s = \frac{h}{2}\int_{kh}^t C\,\mathrm{d}s + \frac{t - (k+1)h}{2}\int_{kh}^t C\,\mathrm{d}s.$$

Since C is bounded we deduce that

$$\int_{kh}^t (s - kh)C\,\mathrm{d}s = \frac{h}{2}\int_{kh}^t C\,\mathrm{d}s + \mathrm{O}(h^2).$$

Since Ψ is assumed bounded and $\mathbb{E}[\overline{N}_T] < +\infty$, the above arguments yield the following representation:

$$\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = \frac{h}{2}\mathbb{E}\left[\int_0^T \Psi(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)})\,\mathrm{d}s\right] + \mathrm{O}(h^2). \tag{4.8}$$

Step 2: from the representation (4.8) *to the expansion at the order one.* In this step we show that

$$\mathbb{E}\left[\int_0^T \Psi(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)})\,\mathrm{d}s\right] = \mathbb{E}\left[\int_0^T \Psi(s, x_s)\,\mathrm{d}s\right] + \mathrm{O}(h).$$

First, we introduce the random variables $\overline{\Gamma}$ and Γ defined by

$$\overline{\Gamma} := \int_0^T \Psi(\overline{\eta}(s), \bar{x}_{\overline{\eta}(s)})\,\mathrm{d}s \quad \text{and} \quad \Gamma := \int_0^T \Psi(\overline{\eta}(s), x_{\overline{\eta}(s)})\,\mathrm{d}s,$$

and write

$$\mathbb{E}[|\overline{\Gamma} - \Gamma|] = \mathbb{E}\big[\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \leq T}|\overline{\Gamma} - \Gamma|\big]x + \mathbb{E}\big[\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T}|\overline{\Gamma} - \Gamma|\big],$$

where $\overline{\tau}^\dagger$ is defined in Definition 3.1. Since Ψ is bounded and $\mathbb{P}(\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \leq T) = \mathrm{O}(h)$ (see the proof of Theorem 3.1), we have

$$\mathbb{E}\big[|\overline{\Gamma} - \Gamma|\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \leq T}\big] = \mathrm{O}(h).$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

Now, recall from (3.1) that, on the event $\{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T\}$, we have $T_k = \overline{T}_k$ and $\theta_k = \overline{\theta}_k$ for all $k \geq 1$ such that $T_k \in [0, T]$. Thus, for all $n \leq \overline{N}_T$ and for all $s \in [\overline{T}_n, \overline{T}_{n+1}$ [we have $\overline{x}_{\overline{\eta}(s)} = (\overline{\theta}_n, \overline{\phi}_{\overline{\theta}_n}(\overline{\eta}(s) - \overline{T}_n, \overline{\nu}_n))$ and $x_{\overline{\eta}(s)} = (\overline{\theta}_n, \phi_{\overline{\theta}_n}(\overline{\eta}(s) - \overline{T}_n, \nu_n))$. Consequently, on the event $\{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T\}$ we have

$$|\overline{\Gamma} - \Gamma| \leq \sum_{n=0}^{\overline{N}_T} \int_{\overline{T}_n}^{\overline{T}_{n+1} \wedge T} |\Psi(\overline{\eta}(s), \overline{\theta}_n, \overline{\phi}_{\overline{\theta}_n}(\overline{\eta}(s) - \overline{T}_n, \overline{\nu}_n)) - \Psi(\overline{\eta}(s), \overline{\theta}_n, \phi_{\overline{\theta}_n}(\overline{\eta}(s) - \overline{T}_n, \nu_n))| \, \mathrm{d}s.$$

From the regularity Assumptions 4.1 and 4.2, the function $\nu \mapsto \Psi(t, \theta, \nu)$ is uniformly Lipschitz in (t, θ) with constant L_Ψ as sum and product of bounded Lipschitz functions. Thus, from this Lipschitz property and the application of Lemma 2.1, we get

$$|\Psi(\overline{\eta}(s), \overline{\theta}_n, \overline{\phi}_{\overline{\theta}_n}(\overline{\eta}(s) - \overline{T}_n, \overline{\nu}_n)) - \Psi(\overline{\eta}(s), \overline{\theta}_n, \phi_{\overline{\theta}_n}(\overline{\eta}(s) - \overline{T}_n, \nu_n))| \leq L_\Psi C \, \mathrm{e}^{LT}(n+1)h.$$

From the above inequality, we find that

$$\mathbb{E}\big[\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T} |\overline{\Gamma} - \Gamma|\big] \leq L_\Psi C \, \mathrm{e}^{LT} T h \mathbb{E}[\overline{N}_T(\overline{N}_T + 1)].$$

Since $\overline{N}_T \leq N_T^*$ and $\mathbb{E}[N_T^*(N_T^* + 1)] < +\infty$, we conclude that

$$\mathbb{E}\big[\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T} |\overline{\Gamma} - \Gamma|\big] = \mathrm{O}(h).$$

We have shown that

$$\mathbb{E}\bigg[\int_0^T \Psi(\overline{\eta}(s), \overline{x}_{\overline{\eta}(s)}) \, \mathrm{d}s\bigg] = \mathbb{E}\bigg[\int_0^T \Psi(\overline{\eta}(s), x_{\overline{\eta}(s)}) \, \mathrm{d}s\bigg] + \mathrm{O}(h).$$

Secondly, from the regularity Assumptions 4.1 and 4.2, the function $(t, \nu) \mapsto \Psi(t, \theta, \nu)$ is uniformly Lipschitz in θ. Moreover, for all $s \in [0, T]$ there exists $k \geq 0$ such that both s and $\overline{\eta}(s)$ belong to the same interval $[\overline{T}_k, \overline{T}_{k+1}$ [so that $x_s = (\theta_k, \phi_{\theta_k}(s - \overline{T}_k, \nu_k))$ and $x_{\overline{\eta}(s)} = (\theta_k, \phi_{\theta_k}(\overline{\eta}(s) - \overline{T}_k, \nu_k))$. Thus, from the Lipschitz continuity of Ψ, from the fact that $|s - \overline{\eta}(s)| \leq h$ and since f_θ is uniformly bounded in θ, we have $|\Psi(s, x_s) - \Psi(\overline{\eta}(s), x_{\overline{\eta}(s)})| \leq Ch$, where C is a constant independent of h. Then we obtain

$$\sup_{s \in [0,T]} |\mathbb{E}[\Psi(s, x_s)] - \mathbb{E}[\Psi(\overline{\eta}(s), x_{\overline{\eta}(s)})]| \leq Ch,$$

from which we deduce that

$$\bigg|\mathbb{E}\bigg[\int_0^T \Psi(\overline{\eta}(s), x_{\overline{\eta}(s)}) \, \mathrm{d}s\bigg] - \mathbb{E}\bigg[\int_0^T \Psi(s, x_s) \, \mathrm{d}s\bigg]\bigg| \leq CTh.$$

Finally, the weak error expansion reads

$$\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)] = \frac{h}{2}\mathbb{E}\bigg[\int_0^T \Psi(s, x_s) \, \mathrm{d}s\bigg] + \mathrm{O}(h^2).$$

5. Numerical experiment

In this section we use the theoretical results above to apply the MLMC method to the PDMP two-dimensional Morris–Lecar (shortened to PDMP 2d-ML).

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

5.1. The PDMP two-dimensional Morris–Lecar

The deterministic Morris–Lecar model was introduced in 1981 by Catherine Morris and Harold Lecar in [23] to explain the dynamics of the barnacle muscle fibre. This model belongs to the family of conductance-based models (like the Hodgkin–Huxley model [19]) and takes the following form:

$$\begin{aligned}\frac{\mathrm{d}v}{\mathrm{d}t} &= \frac{1}{C}(I - g_{\mathrm{Leak}}(v - V_{\mathrm{Leak}}) - g_{\mathrm{Ca}}M_\infty(v)(v - V_{\mathrm{Ca}}) - g_{\mathrm{K}}n(v - V_{\mathrm{K}})),\\ \frac{\mathrm{d}n}{\mathrm{d}t} &= (1 - n)\alpha_{\mathrm{K}}(v) - n\beta_{\mathrm{K}}(v),\end{aligned} \tag{5.1}$$

where

$$\begin{aligned}M_\infty(v) &= (1 + \tanh[(v - V_1)/V_2])/2,\\ \alpha_{\mathrm{K}}(v) &= \lambda_{\mathrm{K}}(v)N_\infty(v),\\ \beta_{\mathrm{K}}(v) &= \lambda_{\mathrm{K}}(v)(1 - N_\infty(v)),\\ N_\infty(v) &= (1 + \tanh[(v - V_3)/V_4])/2,\\ \lambda_{\mathrm{K}}(v) &= \overline{\lambda}_{\mathrm{K}}\cosh((v - V_3)/2V_4).\end{aligned}$$

In this section we consider the PDMP version of (5.1), which we denote by $(x_t, t \in [0, T])$, $T > 0$, whose characteristics (f, λ, Q) are given by

$$f(\theta, v) = \frac{1}{C}\left(I - g_{\mathrm{Leak}}(v - V_{\mathrm{Leak}}) - g_{\mathrm{Ca}}M_\infty(v)(v - V_{\mathrm{Ca}}) - g_{\mathrm{K}}\frac{\theta}{N_{\mathrm{K}}}(v - V_{\mathrm{K}})\right),$$

$$\lambda(\theta, v) = (N_{\mathrm{K}} - \theta)\alpha_{\mathrm{K}}(v) + \theta\beta_{\mathrm{K}}(v),$$

$$Q((\theta, v), \{\theta + 1\}) = \frac{(N_{\mathrm{K}} - \theta)\alpha_{\mathrm{K}}(v)}{\lambda(\theta, v)}, \quad Q((\theta, v), \{\theta - 1\}) = \frac{\theta\beta_{\mathrm{K}}(v)}{\lambda(\theta, v)}.$$

The state space of the model is $E = \{0, \ldots, N_{\mathrm{K}}\} \times \mathbb{R}$ where $N_{\mathrm{K}} \geq 1$ stands for the number of potassium gates. The values of the parameters used in the simulations are $V_1 = -1.2$, $V_2 = 18$, $V_3 = 2$, $V_4 = 30$, $\overline{\lambda}_{\mathrm{K}} = 0.04$, $C = 20$, $g_{\mathrm{Leak}} = 2$, $V_{\mathrm{Leak}} = -60$, $g_{\mathrm{Ca}} = 4.4$, $V_{\mathrm{Ca}} = 120$, $g_{\mathrm{K}} = 8$, $V_{\mathrm{K}} = -84$, $I = 60$, $N_{\mathrm{K}} = 100$. See Figure 1.

5.2. Classical and multilevel Monte Carlo estimators

In this subsection we introduce the classical and multilevel Monte Carlo estimators in order to estimate the quantity $\mathbb{E}[F(x_T)]$, where $(x_t, t \in [0, T])$ is the PDMP 2d-ML and $F(\theta, v) = v$ for $(\theta, v) \in E$ so that $F(x_T)$ gives the value of the membrane potential at time T. Note that other possible choices are $F(\theta, v) = v^n$ or $F(\theta, v) = \theta^n$ for some $n \geq 2$. In those cases, the quantity $\mathbb{E}[F(x_T)]$ gives the moments of the membrane potential or the number of open gates at time T so that we can compute statistics on these biological variables.

Let $X := F(x_T)$. It will be convenient below to emphasize the dependence of the Euler scheme $(\overline{x}_t)$ on a time step h. We introduce a family of random variables $(X_h, h > 0)$ defined by $X_h := F(\overline{x}_T)$, where for a given $h > 0$ the corresponding PDP $(\overline{x}_t)$ is constructed as in Section 2.3 with time step h. In particular, the processes $(\overline{x}_t)$ for $h > 0$ are correlated through the same randomness (U_k), (V_k), and (N_t^*). We build a classical Monte Carlo estimator of $\mathbb{E}[X]$ based on the family $(X_h, h > 0)$ as follows:

$$Y^{\mathrm{MC}} = \frac{1}{N}\sum_{k=1}^{N} X_h^k, \tag{5.2}$$

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

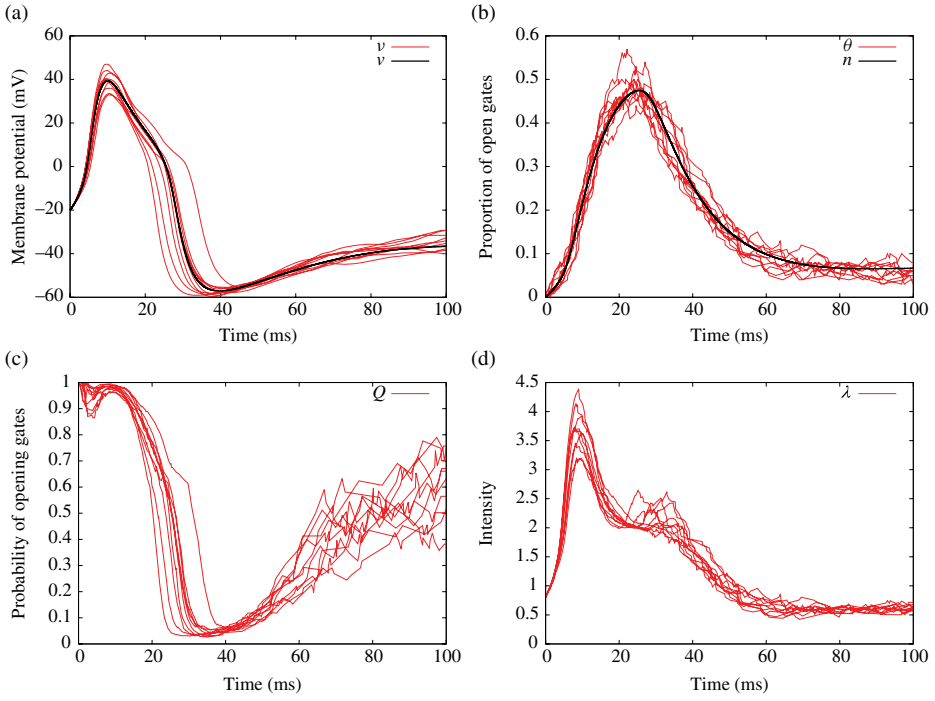

FIGURE 1. Ten trajectories of the characteristics of the PDMP 2d-ML on [0,100]. (a) Membrane potential as a function of time: red curves, stochastic potential; black curve, deterministic potential. (b) Proportion of opened gates as a function of time: red curves, stochastic gates (θ/N_{K}); black curve, deterministic gates (n). (c) Probability of opening a gate ($Q(x_t, \{\theta_t + 1\})$) as a function of time. (d) Jump rate ($\lambda(x_t)$) as a function of t.

where $(X_h^k, k \geq 1)$ is an i.i.d. sequence of random variables distributed like X_h. The parameters $h > 0$ and $N \in \mathbb{N}$ have to be determined. We build a multilevel Monte Carlo estimator based on the family $(X_h, h > 0)$ as follows:

$$Y^{\mathrm{MLMC}} = \frac{1}{N_1} \sum_{k=1}^{N_1} X_{h^*}^k + \sum_{l=2}^{L} \frac{1}{N_l} \sum_{k=1}^{N_l} (X_{h_l}^k - X_{h_{l-1}}^k), \tag{5.3}$$

where $((X_{h_l}^k, X_{h_{l-1}}^k), k \geq 1)$ for $l = 2, \ldots, L$ are independent sequences of independent copies of the pair $(X_{h_l}, X_{h_{l-1}})$ and independent of the i.i.d. sequence $(X_{h^*}^k, k \geq 1)$. The parameter h^* is a free parameter that we fix in Section 5.4. The parameters $L \geq 2$, $M \geq 2$, $N \geq 1$ and $q = (q_1, \ldots, q_L) \in]0, 1[^L$ with $\sum_{l=1}^{L} q_l = 1$ have to be determined; then we set $N_l := \lceil N q_l \rceil$, $h_l := h^* M^{-(l-1)}$.

We also set $\tilde{X} := F(\tilde{x}_T)\tilde{R}_T$, where $\tilde{R}_T$ is defined as in Proposition 2.2 with an intensity $\tilde{\lambda}$ and a kernel $\tilde{Q}$ that will be specified in Section 5.4, and let $(\tilde{X}_h, h > 0)$ be such that $\tilde{X}_h := F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T$ for all $h > 0$. By Proposition 2.2 we have $\mathbb{E}[X] = \mathbb{E}[\tilde{X}]$ and $\mathbb{E}[X_h] = \mathbb{E}[\tilde{X}_h]$ for $h > 0$. Consequently, we build likewise a multilevel estimator $\tilde{Y}^{\mathrm{MLMC}}$ based on the family $(\tilde{X}_h, h > 0)$.

The complexity of the classical Monte Carlo estimator Y^{MC} depends on the parameters (h, N) and that of the multilevel estimators Y^{MLMC} and $\tilde{Y}^{\mathrm{MLMC}}$ depends on (L, q, N). In order

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

TABLE 1: Optimal parameters for the MLMC estimator (5.3).

L	$\left\lceil 1 + \dfrac{\log(\lvert c_1\rvert^{1/\alpha} h^*)}{\log(M)} + \dfrac{\log(A/\epsilon)}{\alpha \log(M)} \right\rceil, \quad A = \sqrt{1+2\alpha}$
Q	$q_1 = \mu^*(1+\rho(h^*)^{\beta/2})$
	$q_j = \mu^* \rho (h^*)^{\beta/2} \left(\dfrac{n_{j-1}^{-\beta/2} + n_j^{-\beta/2}}{\sqrt{n_{j-1}+n_j}} \right), \quad j = 2, \ldots, L, \quad \mu^* = 1/\sum_{1\le j\le L} q_j$
N	$\left(1+\dfrac{1}{2\alpha}\right) \dfrac{\mathrm{Var}(X)\left(1+\rho(h^*)^{\beta/2}\sum_{j=1}^{L}(n_{j-1}^{-\beta/2}+n_j^{-\beta/2})\sqrt{n_{j-1}+n_j}\right)^2}{\epsilon^2 \sum_{j=1}^{L} q_j (n_{j-1}+n_j)}$

to compare those estimators we proceed as in [21] (see also [24]), that is to say, for each estimator we determine the parameters which minimize the global complexity (or cost) subject to the constraint that the resulting L^2-error must be lower than a prescribed $\epsilon > 0$.

As in [21], we let V_1, c_1, α, β, and $\mathrm{Var}(X)$ be the structural parameters associated with the family $(X_h, h > 0)$ and X. We know theoretically from Theorem 3.1 (strong estimate) and Theorem 4.1 (weak expansion) that $(\alpha, \beta) = (1, 1)$, whereas V_1, c_1, and $\mathrm{Var}(X)$ are not explicit (we explain how we estimate them in Section 5.3). Moreover, the structural parameters $\tilde{V}_1$, $\tilde{c}_1$, $\tilde{\alpha}$, $\tilde{\beta}$, and $\mathrm{Var}(\tilde{X})$ associated with $(\tilde{X}_h, h > 0)$ and $\tilde{X}$ are such that $\tilde{\alpha} = \alpha$, $\tilde{c}_1 = c_1$ (see (4.4)), $\tilde{\beta} = 2$ (see Theorem 3.2), and $\tilde{V}_1$, $\mathrm{Var}(\tilde{X})$ are not explicit.

The classical and multilevel estimators defined above are linear and of Monte Carlo type in the sense described in [21]. The optimal parameters of those estimators are then expressed in terms of the corresponding structural parameters as follows (see [21] or [24]). For a user-prescribed $\epsilon > 0$, the classical Monte Carlo parameters h and N are

$$h(\epsilon) = (1+2\alpha)^{-1/(2\alpha)} \left(\frac{\epsilon}{|c_1|} \right)^{1/\alpha}, \quad N(\epsilon) = \left(1+\frac{1}{2\alpha}\right) \frac{\mathrm{Var}(X)(1+\rho h^{\beta/2}(\epsilon))^2}{\epsilon^2}, \tag{5.4}$$

where $\rho = \sqrt{V_1/\mathrm{Var}(X)}$. The parameters of the estimator Y^{MLMC} are given in Table 1, where $n_l := M^{l-1}$ for $l = 1, \ldots, L$ with the convention $n_0 = n_0^{-1} = 0$. The parameters of $\tilde{Y}^{\mathrm{MLMC}}$ are given in a similar way using $\tilde{V}_1$, $\tilde{\beta}$, and $\mathrm{Var}(\tilde{X})$. Finally, the parameter $M(\epsilon)$ is determined as in [21, Section 5.1].

5.3. Methodology

We compare the classical and multilevel Monte Carlo estimators in terms of *precision*, *CPU-time*, and *complexity*. The *precision* of an estimator Y is defined by the L^2-error

$$\|Y - \mathbb{E}[X]\|_2 = \sqrt{(\mathbb{E}[Y] - \mathbb{E}[X])^2 + \mathrm{Var}(Y)},$$

also known as the root mean square error (RMSE). The *CPU-time* represents the time needed to compute one realization of an estimator. The *complexity* is defined as the number of time

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

steps involved in the simulation of an estimator. Let Y denote the estimator (5.2) or (5.3). We estimate the bias of Y by

$$\widehat{b}_R = \frac{1}{R}\sum_{k=1}^{R} Y^k - \mathbb{E}[X],$$

where $Y^1, \ldots, Y^R$ are R independent replications of the estimator. We estimate the variance of Y by

$$\widehat{v}_R = \frac{1}{R}\sum_{k=1}^{R} v^k,$$

where $v^1, \ldots, v^R$ are R independent replications of v, the empirical variance of Y. In the case where Y is the crude Monte Carlo estimator, we set

$$v = \frac{1}{N(N-1)}\sum_{k=1}^{N}(X_h^k - m_N)^2, \quad m_N = \frac{1}{N}\sum_{k=1}^{N} X_h^k.$$

If Y is the MLMC estimator, we set

$$v = \frac{1}{N_1(N_1-1)}\sum_{k=1}^{N_1}(X_h^k - m_{N_1}^{(1)})^2 + \sum_{l=2}^{L}\frac{1}{N_l(N_l-1)}\sum_{k=1}^{N_l}(X_{h_l}^k - X_{h_{l-1}}^k - m_{N_l}^{(l)})^2,$$

where

$$m_{N_1}^{(1)} = \frac{1}{N_1}\sum_{k=1}^{N_1} X_h^k,$$

and for $l \geq 2$,

$$m_{N_l}^{(l)} = \frac{1}{N_l}\sum_{k=1}^{N_l} X_{h_l}^k - X_{h_{l-1}}^k.$$

Then we define the empirical RMSE $\widehat{\epsilon}_R$ by

$$\widehat{\epsilon}_R = \sqrt{\widehat{b}_R^2 + \widehat{v}_R}. \tag{5.5}$$

The numerical computation of (5.5) for both estimators (5.2) and (5.3) requires computation of the optimal parameters given by (5.4) and in Table 1 of Section 5.2, which are expressed in terms of the structural parameters c_1, V_1, and $\mathrm{Var}(X)$. Moreover, computation of the bias requires the value $\mathbb{E}[X]$. Since there is no closed formula for the mean and variance of X, we estimate them using a crude Monte Carlo estimator with $h = 10^{-5}$ and $N = 10^6$. The constants c_1 and V_1 are not explicit; we use the same estimator of V_1 as in [21, Section 5.1], that is,

$$\widehat{V}_1 = (1 + M^{-\beta/2})^{-2} h^{-\beta}\mathbb{E}[|X_h - X_{h/M}|^2], \tag{5.6}$$

and we use the following estimator of c_1:

$$\widehat{c_1} = (1 - M^{-\alpha})^{-1} h^{-\alpha}\mathbb{E}[X_{h/M} - X_h]. \tag{5.7}$$

The estimator of c_1 is obtained by writing the weak error expansion for the two time steps h and h/M, summing and neglecting the $\mathrm{O}(h^2)$ term. In (5.6) we use $(h, M) = (0.1, 4)$ and in

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

(5.7) we use $(h, M) = (1, 4)$, and the expectations are estimated using a classical Monte Carlo of size $N = 10^4$ on $(X_{h/M}, X_h)$. We emphasize that we are interested in the order of c_1 and V_1, so we do not need a precise estimate here.

5.4. Numerical results

In this subsection we first illustrate the results of Theorems 3.1 and 3.2 on the Morris–Lecar PDMP and then compare the MC and MLMC estimators. The simulations were carried out on a computer with an Intel Core i5-4300U CPU @ 1.90GHz × 4 processor. The code is written in C++. We implement the estimator $\tilde{Y}^{\mathrm{MLMC}}$ (see Section 5.2) for the following choices of the parameters $(\tilde{\lambda}, \tilde{Q})$.

Case 1: $\tilde{\lambda}(\theta) = 1$ and

$$\tilde{Q}(\theta, \{\theta + 1\}) = \frac{N_{\mathrm{K}} - \theta}{N_{\mathrm{K}}}, \quad \tilde{Q}(\theta, \{\theta - 1\}) = \frac{\theta}{N_{\mathrm{K}}}.$$

Case 2: $\tilde{\lambda}(x, t) = \lambda(\theta, v(t))$ and $\tilde{Q}((x, t), \mathrm{d}y) = Q((\theta, v(t)), \mathrm{d}y)$, where v denotes the first component of the solution of (5.1).

Cases 1 and 2 correspond to the application of Proposition 2.2. Based on Corollary 2.2, we also consider the following case.

Case 3: Consider the quantity $\mathbb{E}[F(x_T) - F(\tilde{x}_T)]$, where (x_t) and $(\tilde{x}_t)$ are PDPs with characteristics (Φ, λ, Q) and $(\tilde{\Phi}, \lambda, Q)$ respectively. By Corollary 2.2, we have $\mathbb{E}[F(\tilde{x}_T)] = \mathbb{E}[F(y_T)\tilde{R}_T]$, where (y_t) is a PDP whose discrete component jumps in the same states and at the same times as the discrete component of (x_t), and $(\tilde{R}_t)$ is the corresponding corrective process. Thus, we consider the quantity $\mathbb{E}[F(x_T) - F(y_T)\tilde{R}_T]$ instead of $\mathbb{E}[F(x_T) - F(\tilde{x}_T)]$.

Case 3 implies using the following MLMC estimator, which is slightly different from (5.3):

$$\tilde{Y}^{\mathrm{MLMC}} = \frac{1}{N_1}\sum_{k=1}^{N_1} X_{h^*}^k + \sum_{l=2}^{L} \frac{1}{N_l}\sum_{k=1}^{N_l} X_{h_l}^k - \tilde{X}_{h_{l-1}}^k,$$

where $((X_{h_l}^k, \tilde{X}_{h_{l-1}}^k), k \geq 1)$ for $l = 2, \ldots, L$ are independent sequences of independent copies of the pair $(X_{h_l}, \tilde{X}_{h_{l-1}}) = (F(\bar{x}_T), F(\bar{y}_T)\tilde{R}_T)$, where $(\bar{y}_t)$ is a PDP whose discrete component jumps in the same states and at the same times as the Euler scheme $(\bar{x}_t)$ with time step h_l, whose deterministic motions are given by the approximate flows with time step h_{l-1} and $(\tilde{R}_t)$ is the corresponding corrective process (see Corollary 2.2).

Figure 2 confirms numerically that $\mathbb{E}[|X_{h_l} - X_{h_{l-1}}|^2] = \mathrm{O}(h_l)$ and that $\mathbb{E}[|\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}}|^2] = \mathrm{O}(h_l^2)$ for Cases 1, 2, and 3 (see Theorems 3.1 and 3.2 respectively). Indeed, for $T = 10$ (see Figure 2(a)), we observe that the curve corresponding to the decay of $\mathbb{E}[|X_{h_l} - X_{h_{l-1}}|^2]$ as l increases is approximately parallel to a line of slope -1 and that the curves corresponding to the decay of $\mathbb{E}[|\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}}|^2]$ in Cases 1, 2, and 3 are parallel to a line of slope -2. We also see that the curves corresponding to Cases 2 and 3 are approximately similar, and that for some value of l those curves go below the one corresponding to $\mathbb{E}[|X_{h_l} - X_{h_{l-1}}|^2]$. The curve corresponding to Case 1 is always above all the others; this indicates that the L^2-error (or the variance) in Case 1 is too big (with respect to the others) and that is why we do not consider this case below. As T increases (see Figures 2(b) and 2(c)), the theoretical order of the numerical schemes is still observed. However, for $T = 20$, a slight difference begins to emerge between Cases 2 and 3 (Case 3 being better) and this difference is accentuated for $T = 30$, so we do not represent Case 2.

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

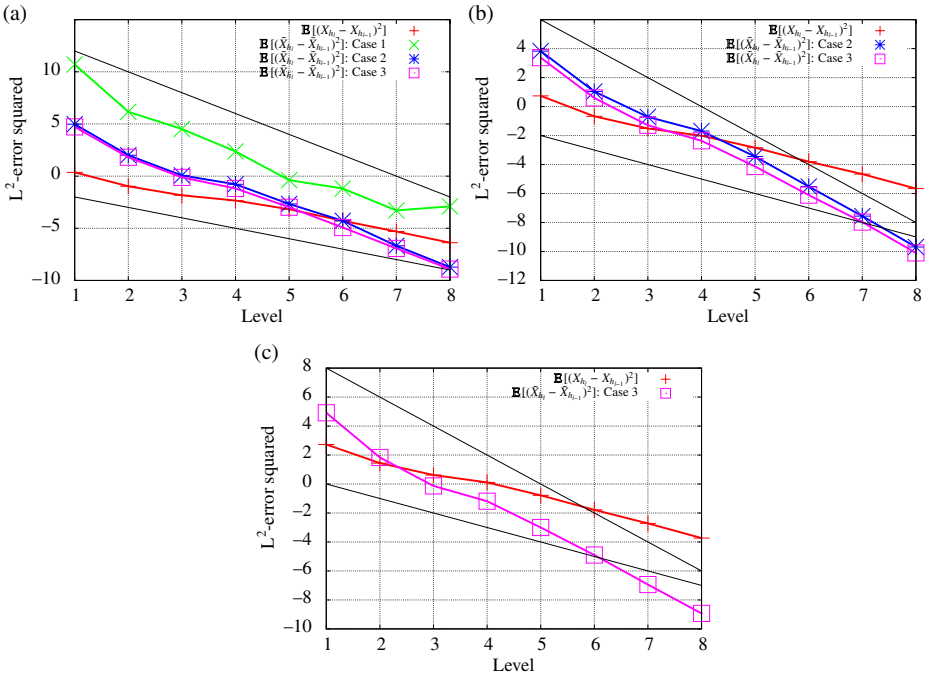

FIGURE 2. Plots (a), (b), and (c) show the decay of $\mathbb{E}[(X_{h_l} - X_{h_{l-1}})^2]$ and $\mathbb{E}[\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}})^2]$ (y-axis, $\log_M$ scale) as a function of l with $h_l = h \times M^{-(l-1)}$, $h = 1$, $M = 4$, for different values of the final time T: (a) $T = 10$, (b) $T = 20$, (c) $T = 30$. For a visual guide, we added black solid lines with slopes -1 and -2.

For the Monte Carlo simulations we set $T = 30$, $\lambda^* = 10$, and the time step involved in the first level of the MLMC is set to $h^* = 0.1$. We choose this value for h^* because it represents (on average) the size of an interval $[T_n^*, T_{n+1}^*]$ of two successive jump times of the auxiliary Poisson process (N_t^*). The estimation of the true value and variance leads to $\mathbb{E}[X] = -31.4723$ and $\mathrm{Var}(X) = 335$. Note that $v(30) = -35.3083$, where v is the deterministic membrane potential solution of (5.1), so there is an offset between the deterministic potential and the mean of the stochastic potential. We replicate 100 times the simulation of the classical and multilevel estimators to compute the empirical RMSE so that $R = 100$ in (5.5).

The results of the Monte Carlo simulations are shown in Table 2 for the classical Monte Carlo estimator Y^{MC} and in Tables 3 and 4 for the multilevel estimators Y^{MLMC} and $\tilde{Y}^{\mathrm{MLMC}}$ (Case 3). As an example, the first line of Table 3 reads as follows: for a user-prescribed $\epsilon = 2^{-1} = 0.5$, the MLMC estimator Y^{MLMC} is implemented with $L = 2$ levels, the time step at the first level is $h^* = 0.1$, this time step is refined by a factor $n_l = M^{l-1}$ with $M = 2$ at each level, and the sample size is $N = 2600$. For such parameters, the numerical complexity of the estimator is $\mathrm{Cost}(Y^{\mathrm{MLMC}}) = 28\,200$, the empirical RMSE $\widehat{\epsilon}_{100} = 0.389$ and the computational time of one realization of Y^{MLMC} is 0.362 seconds. We also reported the empirical bias $\widehat{b}_{100}$ and the empirical variance $\widehat{v}_{100}$ in view of (5.5).

The results indicate that the MLMC outperforms classical MC. More precisely, for small values of ϵ (i.e. $k = 1, 2, 3$) the complexity and the CPU-time of the classical and multilevel MC estimators are of the same order. As ϵ decreases (i.e. as k increases) the difference in complexity and CPU-time between classical and multilevel MC increases. Indeed, for $k = 5$ the

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

TABLE 2: Results and parameters of the Monte Carlo estimator Y^{MC}. Estimated values of the structural parameters: $c_1 = 4.58$, $V_1 = 7.25$.

k	$\epsilon = 2^{-k}$	$\widehat{\epsilon}_{100}$	$\widehat{b}_{100}$	$\widehat{v}_{100}$	Time (sec)	N	h	Cost
1	5.00e-01	4.32e-01	2.34e-01	1.52e-01	3.10e-01	2.16e+03	6.30e-02	3.43e+04
2	2.50e-01	2.59e-01	1.69e-01	3.87e-02	1.55e+00	8.47e+03	3.15e-02	2.69e+05
3	1.25e-01	1.17e-01	6.25e-02	9.78e-03	8.80e+00	3.34e+04	1.58e-02	2.12e+06
4	6.25e-02	5.67e-02	2.73e-02	2.47e-03	5.62e+01	1.32e+05	7.88e-03	1.68e+07
5	3.12e-02	2.50e-02	−1.78e-03	6.21e-04	3.93e+02	5.24e+05	3.94e-03	1.33e+08

TABLE 3: Results and parameters of the multilevel Monte Carlo estimator Y^{MLMC}. Estimated values of the structural parameters: $c_1 = 4.58$, $V_1 = 7.25$.

k	$\epsilon = 2^{-k}$	$\widehat{\epsilon}_{100}$	$\widehat{b}_{100}$	$\widehat{v}_{100}$	Time (sec)	L	M	h	N	Cost
1	5.00e-01	3.89e-01	1.14e-01	1.38e-01	3.62e-01	2	2	0.1	2.60e+03	2.82e+04
2	2.50e-01	2.29e-01	1.19e-01	3.83e-02	1.44e+00	2	4	0.1	1.04e+04	1.16e+05
3	1.25e-01	1.21e-01	6.24e-02	1.07e-02	5.76e+00	2	7	0.1	4.22e+04	4.85e+05
4	6.25e-02	5.91e-02	1.38e-02	3.30e-03	2.69e+01	3	4	0.1	1.90e+05	2.37e+06
5	3.12e-02	3.47e-02	−1.39e-02	1.01e-03	1.08e+02	3	6	0.1	7.71e+05	9.99e+06

TABLE 4: Results and parameters of the multilevel Monte Carlo estimator $\tilde{Y}^{\mathrm{MLMC}}$ (Case 3). Estimated values of the structural parameters: $\tilde{c}_1 = 3.91$, $\tilde{V}_1 = 34.1$.

k	$\epsilon = 2^{-k}$	$\widehat{\epsilon}_{100}$	$\widehat{b}_{100}$	$\widehat{v}_{100}$	Time (sec)	L	M	h	N	Cost
1	5.00e-01	4.28e-01	1.98e-01	1.44e-01	3.13e-01	2	2	0.1	2.38e+03	2.50e+04
2	2.50e-01	2.47e-01	1.55e-01	3.72e-02	1.26e+00	2	3	0.1	9.46e+03	1.00e+05
3	1.25e-01	1.36e-01	8.90e-02	1.05e-02	5.00e+00	2	6	0.1	3.80e+04	4.11e+05
4	6.25e-02	6.22e-02	2.15e-02	3.41e-03	2.09e+01	3	4	0.1	1.58e+05	1.75e+06
5	3.12e-02	3.17e-02	6.07e-03	9.71e-04	8.35e+01	3	5	0.1	6.30e+05	7.02e+06

complexity of the estimator Y^{MC} is approximately 13 times superior to that of Y^{MLMC} and 19 times superior to that of $\tilde{Y}^{\mathrm{MLMC}}$. The same fact appears when we look at the complexity ratio of the estimators Y^{MLMC} and $\tilde{Y}^{\mathrm{MLMC}}$ (i.e. $\mathrm{Cost}(Y^{\mathrm{MLMC}})/\mathrm{Cost}(\tilde{Y}^{\mathrm{MLMC}})$) as ϵ decreases. However, the difference between the complexity of these two MLMC estimators increases more slowly than the one between an MC and an MLMC estimator. Recall that the computational benefit of the MLMC over the MC grows as the prescribed ϵ decreases.

Both classical and multilevel estimators provide an empirical RMSE which is close to the prescribed precision (see Tables 2, 3, and 4). We can conclude that the choice of the parameters is well adapted. For readability, Figures 3(a) and 3(b) show the ratios of the complexities and the CPU-times of the three estimators Y^{MC}, Y^{MLMC}, and $\tilde{Y}^{\mathrm{MLMC}}$ as a function of ϵ.

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

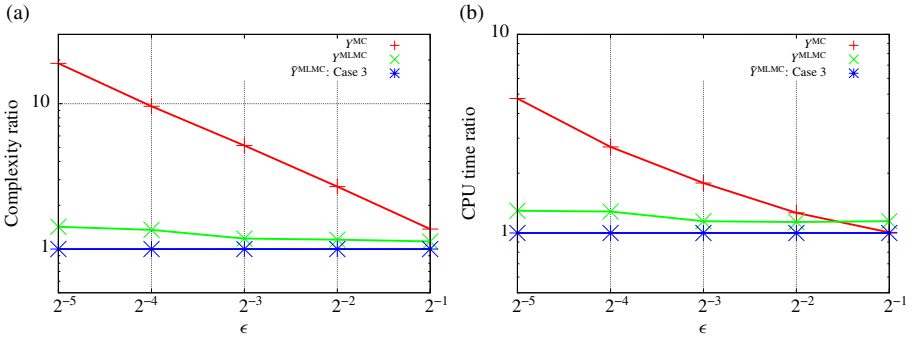

FIGURE 3. (a) Ratio of the complexities and (b) ratio of the CPU-times with respect to the complexity and CPU-time of the estimator $\tilde{Y}^{\mathrm{MLMC}}$ as a function of the prescribed ϵ ($\log_2$ scale for the x-axis, log scale for the y-axis).

References

[1] ANDERSON, D. F. AND HIGHAM, D. J. (2012). Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics. *Multiscale Model. Simul.* **10**, 146–179.

[2] ANDERSON, D. F., HIGHAM, D. J. AND SUN, Y. (2014). Complexity of multilevel Monte Carlo tau-leaping. *SIAM J. Numer. Anal.* **52**, 3106–3127.

[3] BENAM, M., LE BORGNE, S., MALRIEU, F. AND ZITT, P.-A. (2012). Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.* **17**, 1–14, 56.

[4] BRÉMAUD, P. (1981). *Point Processes and Queues, Martingale Dynamics*. Springer, New York.

[5] DAVIS, M. H. A. (1984). Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. *J. R. Statist. Soc.* **46**, 353–388.

[6] DAVIS, M. H. A. (1993). *Markov Models and Optimization*. Chapman & Hall, London.

[7] DEREICH, S. (2011). Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *J. Appl. Prob.* **21**, 283–311.

[8] DEREICH, S. AND HEIDENREICH, F. (2011). A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stoch. Process. Appl.* **121**, 1565–1587.

[9] DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.

[10] FERREIRO-CASTILLA, A., KYPRIANOU, A. E., SCHEICHL, R. AND SURYANARAYANA, G. (2014). Multilevel Monte Carlo simulation for Lévy processes based on the Wiener–Hopf factorisation. *Stoch. Process. Appl.* **124**, 985–1010.

[11] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Operat. Res.* **56**, 607–617.

[12] GILES, M. B. (2015). Multilevel Monte Carlo methods. In *Acta Numerica 24*, pp. 259–328. Cambridge University Press.

[13] GIORGI, D. (2017). Théorèmes limites pour estimateurs Multilevel avec et sans poids: comparaisons et applications. Doctoral thesis, Université Pierre et Marie Curie – Paris 6.

[14] GLYNN, P. W. AND RHEE, C.-H. (2014). Exact estimation for Markov chain equilibrium expectations. *J. Appl. Prob.* **51**, 377–389.

[15] GLYNN, P. W. AND RHEE, C.-H. (2015). Unbiased estimation with square root convergence for SDE models. *Operat. Res.* **63**, 1026–1043.

[16] GRAHAM, C. AND TALAY, D. (2013). *Stochastic Simulation and Monte Carlo*. Springer.

[17] HAIRER, E., NØRSETT, S. P. AND WANNER, G. (2008). *Solving Ordinary Differential Equations I*, 2nd revised edn. Springer.

[18] HEINRICH, S. (2001). Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing* (Lecture Notes Comput. Sci. 2179), eds S. Margenov *et al.*, pp. 58–67. Springer, Berlin and Heidelberg.

[19] HODGKIN, A. L. AND HUXLEY, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544.

[20] JACOBSEN, M. (2006). *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhäuser, Boston.

[21] LEMAIRE, V. AND PAGÉS, G. (2017). Multilevel Richardson–Romberg extrapolation. *Bernoulli* **23** (4A), 2643–2692.

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press

[22] LEMAIRE, V., THIEULLEN, M. AND THOMAS, N. (2018). Exact simulation of the jump times of a class of piecewise deterministic Markov processes. *J. Sci. Comput.* **75**, 1776–1807.
[23] MORRIS, C. AND LECAR, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.* **35**, 193–213.
[24] PAGÉS, G. (2018). *Numerical Probability: An Introduction with Applications to Finance* (Universitext). Springer, Cham.
[25] PAKDAMAN, K., THIEULLEN, M. AND WAINRIB, G. (2010). Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Adv. Appl. Prob.* **42** (3), 761–794.
[26] PALMOWSKI, Z. AND ROLSKI, T. (2002). A technique for exponential change of measure for Markov processes. *Bernoulli* **8** (6), 767–785.
[27] TALAY, D. AND TUBARO, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8** (4), 483–509.
[28] XIA, Y. AND GILES, M. B. (2012). Multilevel path simulation for jump-diffusion SDEs. In *Monte Carlo and Quasi-Monte Carlo Methods 2010* (Springer Proc. Math. Statist. 23), eds L. Plaskota and H. Woźniakowski, pp. 695–708. Springer, Berlin and Heidelberg.

https://doi.org/10.1017/apr.2019.55 Published online by Cambridge University Press