

# Thinning and Multilevel Monte Carlo for Piecewise Deterministic (Markov) Processes. Application to a stochastic Morris-Lecar model.

Vincent Lemaire<sup>\*†</sup>

Michèle Thieullen<sup>\*‡</sup>

Nicolas Thomas<sup>\*§</sup>

## Abstract

In the first part of this paper we study approximations of trajectories of Piecewise Deterministic Processes (PDP) when the flow is not explicit 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 (PDMP). These estimates are the building blocks of the Multilevel Monte Carlo (MLMC) method which we study in the second part. The coupling required by MLMC is based on the thinning procedure. In the third part we apply these results to a 2-dimensional Morris-Lecar model with stochastic ion channels. In the range of our simulations the MLMC estimator outperforms the classical Monte Carlo one.

**Keywords:** Piecewise Deterministic (Markov) Processes, Multilevel Monte Carlo, Thinning, Strong error estimate, Weak error expansion, Morris-Lecar model.

**Mathematics Subject Classification:** 65C05, 65C20, 60G55, 60J25, 68U20

## 1 Introduction

In this paper we are interested in the approximation of the trajectories of PDPs. We establish strong error estimates for a PDP and a weak error expansion for a PDMP. Then we study the application of the Multilevel Monte Carlo (MLMC) method in order to approximate expectations of functional 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 coding, is considered as a 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 have been introduced by Davis in [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

---

<sup>\*</sup>Laboratoire de Probabilités, Statistique et Modélisation (LPSM), UMR CNRS 8001, Sorbonne Université-Campus Pierre et Marie Curie, Case 158, 4 place Jussieu, F-75252 Paris Cedex 5, France

<sup>†</sup>vincent.lemaire@upmc.fr

<sup>‡</sup>michele.thieullen@upmc.fr

<sup>§</sup>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 consider first a general PDP  $(x_t)$  which 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  $(x_t)$  that we denote  $(\bar{x}_t)$ . The process  $(\bar{x}_t)$  is a PDP constructed by thinning of a homogeneous Poisson process which enjoys explicitly solvable flows. Actually 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

$$\exists V_1 > 0, V_2 > 0, \quad \mathbb{E}[|F(\bar{x}_T) - F(x_T)|^2] \leq V_1 h + V_2 h^2. \quad (1)$$

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

$$\exists c_1 > 0, \quad \mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = c_1 h + o(h^2). \quad (2)$$

The estimate (1) is mainly based on the construction of the couple  $(x_t, \bar{x}_t)$  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. On the contrary, the estimate (2) relies on properties which are specific to PDMPs such as the Feynman-Kac formula.

The MLMC method relies simultaneously on estimates (1) and (2) that is why we study its application to the PDMP framework instead of the more general PDP one. MLMC extends the classical Monte Carlo (MC) method which is a very general approach to estimate expectations using stochastic simulations. The complexity (i.e the number of operations necessary in the simulation) associated to a 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 the classical MC method. MLMC can then greatly reduce the complexity of the classical MC by performing most simulations with low accuracy but with low complexity and only few simulations with high accuracy at high complexity. MLMC have been introduced by S. Heinrich in [18] and developed by M. Giles in [12]. The MLMC estimator has been efficiently used in various fields of numerical probability such as SDEs [12], Markov chains [1], [2], [14], Lévy processes [10], jump diffusions [28], [7], [8] or nested Monte Carlo [21], [13]. See [11] for more references. To the best of our knowledge, application of MLMC to PDMPs has not been considered.

For the sake of clarity, we describe here 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  $X^k, k \geq 1$  independent random variables identically distributed as  $X$ , provides an unbiased estimator. The associated  $L^2$  - error satisfies  $\|Y - \mathbb{E}[X]\|_2^2 = \text{Var}(Y) = \frac{1}{N} \text{Var}(X)$ . If we quantify the precision by the  $L^2$  - error, then a user-prescribed precision  $\epsilon^2 > 0$  is achieved for  $N = O(\epsilon^{-2})$  so that in this case the global complexity is of order  $O(\epsilon^{-2})$ .

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  $h \rightarrow 0$  in the following sense

$$\exists c_1 > 0, \alpha > 0, \quad \mathbb{E}[X_h] - \mathbb{E}[X] = c_1 h^\alpha + o(h^{2\alpha}), \quad (3)$$

and

$$\exists V_1 > 0, \beta > 0, \quad \mathbb{E}[|X_h - X|^2] \leq V_1 h^\beta. \quad (4)$$

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

$$Y = \frac{1}{N} \sum_{k=1}^N X_h^k, \quad (5)$$

where  $X_h^k, k \geq 1$  are independent random variables identically distributed as  $X_h$ . The bias and the variance of the estimator (5) are respectively given by  $\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[X_h] - \mathbb{E}[X] \simeq c_1 h^\alpha$  and  $\text{Var}(Y) = \frac{1}{N} \text{Var}(X_h)$ . From the strong estimate (4) we have that  $\text{Var}(X_h) \rightarrow \text{Var}(X)$  as  $h \rightarrow 0$  so that  $\text{Var}(X_h)$  is asymptotically a constant independent of  $h$ . If as above we quantify the precision by the  $L^2$  - error and use that  $\|Y - \mathbb{E}[X]\|_2^2 = (\mathbb{E}[Y] - \mathbb{E}[X])^2 + \text{Var}(Y)$ , we obtain that the estimator (5) achieves a user-prescribed precision  $\epsilon^2 > 0$  for  $h = O(\epsilon^{1/\alpha})$  and  $N = O(\epsilon^{-2})$  so that the global complexity of the estimator is now  $O(\epsilon^{-2-\frac{1}{\alpha}})$ .

The MLMC method takes advantage of the estimate (4) in order to reduce the global complexity. Let us fix  $L \geq 2$  and consider for  $l \in \{1, \dots, L\}$  a geometrically decreasing sequence  $(h_l, 1 \leq l \leq L)$  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 (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

$$\mathbb{E}[X_{h_L}] = \mathbb{E}[X_{h^*}] + \sum_{l=2}^L \mathbb{E}[X_{h_l} - X_{h_{l-1}}]. \quad (6)$$

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  $N_l \geq 1$  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 couples  $(X_{h_l}, X_{h_{l-1}})$  of correlated random variables, the MLMC estimator then reads

$$Y = \frac{1}{N_1} \sum_{k=1}^{N_1} X_{h^*}^k + \sum_{l=2}^L \frac{1}{N_l} \sum_{k=1}^{N_l} (X_{h_l}^k - X_{h_{l-1}}^k), \quad (7)$$

where  $(X_{h^*}^k, k \geq 1)$  is a sequence of independent and identically distributed random variables distributed as  $X_{h^*}$  and  $((X_{h_l}^k, X_{h_{l-1}}^k), k \geq 1)$  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 [12] or [21], that given a precision  $\epsilon > 0$  and provided that the family  $(X_h, h > 0)$  satisfies the strong and weak error estimates (4) and (3), the multilevel estimator (7) achieves a precision  $\|Y - \mathbb{E}[X]\|_2^2 = \epsilon^2$  with a global complexity of order  $O(\epsilon^{-2})$  if  $\beta > 1$ ,  $O(\epsilon^{-2}(\log(\epsilon))^2)$  if  $\beta = 1$  and  $O(\epsilon^{-2-(1-\beta)/\alpha})$  if  $\beta < 1$ . This complexity result shows the importance of the parameter  $\beta$ . Finally, let us mention that in the case  $\beta > 1$  it is possible to build an unbiased multilevel estimator, see [15].

Estimates (1) and (2) suggest to investigate 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 a MLMC estimator of  $\mathbb{E}[F(x_T)]$  just as in (7) (noted  $Y^{\text{MLMC}}$  in the paper) 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). In order to improve the convergence rate (to increase the parameter  $\beta$ ) in (1), we show that for a given PDMP  $(x_t)$  we have the following auxiliary representation

$$\mathbb{E}[F(x_T)] = \mathbb{E}[F(\tilde{x}_T)\tilde{R}_T]. \quad (8)$$

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 which depends on  $(\tilde{x}_t, t \in [0, T])$ . The representation (8) is inspired by the change of probability introduced in [28] and is actually valid for a general PDP (Proposition 2.2) so that  $\mathbb{E}[F(\tilde{x}_T)] = \mathbb{E}[F(\tilde{x}_T)\tilde{R}_T]$  where  $(\tilde{x}_t)$  is the Euler scheme corresponding to  $(\tilde{x}_t)$  and  $(\tilde{R}_t)$  is a process which depends on  $(\tilde{x}_t, t \in [0, T])$ . Letting  $X = F(\tilde{x}_T)\tilde{R}_T$  and  $X_h = F(\tilde{x}_T)\tilde{R}_T$  we define a second MLMC estimator (noted  $\tilde{Y}^{\text{MLMC}}$ ) where now the discrete components of the Euler schemes  $(\tilde{x}_t)$  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 ( $Y^{\text{MLMC}}$ ) derives from (6) where the corrective term at level  $l$  is  $\mathbb{E}[F(\tilde{x}_T^{h_l}) - F(\tilde{x}_T^{h_{l-1}})]$  whereas the corrective term of the second estimator ( $\tilde{Y}^{\text{MLMC}}$ ) is  $\mathbb{E}[F(\tilde{x}_T^{h_l})\tilde{R}_T^{h_l} - F(\tilde{x}_T^{h_{l-1}})\tilde{R}_T^{h_{l-1}}]$ . For readability, we no longer write the dependence of the approximations on the time step. For the processes  $(F(\tilde{x}_t)\tilde{R}_t)$  and  $(F(\tilde{x}_t)\tilde{R}_t)$  we show the following strong estimate

$$\exists \tilde{V}_1 > 0, \quad \mathbb{E}[|F(\tilde{x}_T)\tilde{R}_T - F(\tilde{x}_T)\tilde{R}_T|^2] \leq \tilde{V}_1 h^2,$$

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

As an application we consider the PDMP version of the 2-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 behavior of PDMPs to a stochastic 3-dimensional Morris-Lecar model. The original deterministic Morris-Lecar model has been introduced in [23] to account for various oscillating states in the barnacle giant muscle fiber. 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 cells 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 2-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 functional of PDMPs. This setting can then be considered as a natural extension of this work.

The paper is organised as follows. In section 2, we construct a general PDP by thinning and we give a representation of its distribution in term of the thinning data (Proposition 1). In section 3, we establish strong error estimates (Theorems 1-2). In section 4, we establish a weak error expansion (Theorem 3). In section 5, we compare the efficiency of the classical and the multilevel Monte Carlo estimators on the 2-dimensional stochastic Morris-Lecar model.

## 2 Piecewise Deterministic Process by thinning

### 2.1 Construction

In this section 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  $\Theta$  is a finite or countable set and  $d \geq 1$ . A piecewise deterministic process (PDP) is defined from the following characteristics

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

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

$$Q(x, A \times B) = Q(x, A)\delta_\nu(B). \quad (9)$$

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

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

Our results do not depend on the dimension of the variable in  $\mathbb{R}^d$  so we restrict ourself to  $\mathbb{R}$  ( $d = 1$ ) for the 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  $\lambda^*$ . In the present paper  $\lambda^*$  is constant (see Assumption 2.1). Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space on which we define

1. an homogeneous Poisson process  $(N_t^*, t \geq 0)$  with intensity  $\lambda^*$  (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 iid 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  $(\Phi, \lambda, Q)$ . 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

$$T_1 := T_{\tau_1}^*, \quad (10)$$

where

$$\tau_1 := \inf \{k > 0 : U_k \lambda^* \leq \lambda(\theta_0, \Phi_{\theta_0}(T_k^*, \nu_0))\}. \quad (11)$$

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

$$a_j(x) := \sum_{i=1}^j Q(x, \{k_i\}), \quad \forall x \in E. \quad (12)$$

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

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

For all  $x \in E$ ,  $H(x, \cdot)$  is the inverse of the cumulative distribution function of  $Q(x, \cdot)$  (see for example [9]). Then, we construct  $(\theta_1, \nu_1)$  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  $(\theta_1, \nu_1)$  given  $(\tau_1, (T_k^*)_{k \leq \tau_1})$  is  $Q((\theta_0, \Phi_{\theta_0}(T_{\tau_1}^*, \nu_0)), \cdot)$  or in view of (9),

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

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

$$T_n := T_{\tau_n}^*,$$

where

$$\tau_n := \inf \left\{ k > \tau_{n-1} : U_k \lambda^* \leq \lambda(\theta_{n-1}, \Phi_{\theta_{n-1}}(T_k^* - T_{\tau_{n-1}}^*, \nu_{n-1})) \right\}.$$

Then, we construct  $(\theta_n, \nu_n)$  using the uniform random variable  $V_n$  and the function  $H$  as follows

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

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

$$x_t := (\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n)), \quad t \in [T_n, T_{n+1}]. \quad (13)$$

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

## 2.2 Approximation of a PDP

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

$$\sup_{t \in [0, T]} |\Phi_\theta(t, \nu_1) - \bar{\Phi}_\theta(t, \nu_2)| \leq e^{C_1 T} |\nu_1 - \nu_2| + C_2 h, \quad \forall \theta \in \Theta, \forall (\nu_1, \nu_2) \in \mathbb{R}^2. \quad (14)$$

To the family  $(\bar{\Phi}_\theta)$  we can associate a PDP constructed as above that we denote  $(\bar{x}_t)$ . We emphasize that there is a positive probability that  $(x_t)$  and  $(\bar{x}_t)$  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  $\theta$ , that is  $\tilde{\lambda}(x) = \tilde{\lambda}(\theta)$  and  $\tilde{Q}(x, \cdot) = \tilde{Q}(\theta, \cdot)$  for all  $x = (\theta, \nu) \in E$ , then its embedded Markov chain  $(\tilde{T}_n, (\tilde{\theta}_n, \tilde{\nu}_n), n \geq 0)$  is such that  $(\tilde{\theta}_n, n \geq 0)$  is an autonomous Markov chain with kernel  $\tilde{Q}$  and  $(\tilde{T}_n, n \geq 0)$  is a counting process with intensity  $\tilde{\lambda}_t = \sum_{n \geq 0} \tilde{\lambda}(\tilde{\theta}_n) \mathbb{1}_{\tilde{T}_n \leq t < \tilde{T}_{n+1}}$ . In particular, both  $(\tilde{\theta}_n)$  and  $(\tilde{\tau}_n)$  do not depend on  $\Phi$ . 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 exactly at the same times and their  $\theta$ -component always

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 in the sequel.

The following lemma (which is important for several proofs below) gives a direct consequence of the estimate (14).

**Lemma 2.1.** *Let  $(\Phi_\theta)$  and  $(\bar{\Phi}_\theta)$  satisfying (14). 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  $\Theta$ -valued components. For a given  $\nu \in \mathbb{R}$  let us define iteratively the sequences  $(\beta_n, n \geq 0)$  and  $(\bar{\beta}_n, n \geq 0)$  as follows*

$$\begin{cases} \beta_n = \Phi_{\alpha_{n-1}}(t_n - t_{n-1}, \beta_{n-1}), \\ \beta_0 = \nu, \end{cases} \quad \text{and} \quad \begin{cases} \bar{\beta}_n = \bar{\Phi}_{\alpha_{n-1}}(t_n - t_{n-1}, \bar{\beta}_{n-1}), \\ \bar{\beta}_0 = \nu. \end{cases}$$

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

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

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

*Proof of Lemma 2.1.* Let  $n \geq 1$ . From the estimate (14), we have for all  $k \leq n$

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

and therefore

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

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

$$|\bar{\beta}_n - \beta_n| \leq e^{C_1 t_n} n C_2 h.$$

□

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

In this section we define a PDMP and its associated Euler scheme from the construction of the section 2.1. For all  $\theta \in \Theta$ , we consider a family of vector fields  $(f_\theta, \theta \in \Theta)$  satisfying

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

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

$$\begin{cases} \frac{dy(t)}{dt} = f_\theta(y(t)), \\ y(0) = \nu, \end{cases} \quad (15)$$

then the corresponding PDP is Markov since  $\phi$  satisfies the semi-group property which 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 (15) by the Euler scheme with time step  $h$ . First, we define the Euler subdivision of  $[0, +\infty[$  with time step  $h$ , noted  $(\bar{t}_i, i \geq 0)$ , by  $\bar{t}_i := ih$ .

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

$$\begin{cases} \bar{y}_{i+1}(x) = \bar{y}_i(x) + hf_\theta(\bar{y}_i(x)), \\ \bar{y}_0(x) = \nu, \end{cases}$$

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

$$\bar{\phi}_\theta(t, \nu) := \bar{y}_i(x) + (t - \bar{t}_i)f_\theta(\bar{y}_i(x)), \quad \forall t \in [\bar{t}_i, \bar{t}_{i+1}]. \quad (16)$$

We construct the approximating process  $(\bar{x}_t)$  as follows. Its continuous component starts from  $\nu_0$  at time 0 and follows the flow  $\bar{\phi}_{\theta_0}(t, \nu_0)$  until the first jump time  $\bar{T}_1$  that we construct by (10) and (11) of section 2.1 where we replace  $\Phi_{\theta_0}(T_k^*, \nu_0)$  by  $\bar{\phi}_{\theta_0}(T_k^*, \nu_0)$ . At time  $\bar{T}_1$  the continuous component of  $\bar{x}_{\bar{T}_1}$  is equal to  $\bar{\phi}_{\theta_0}(\bar{T}_1, \nu_0) := \bar{\nu}_1$  since there is no jump in the continuous component. The discrete component jumps to  $\bar{\theta}_1$ . We iterate this procedure with the new flow  $\bar{\phi}_{\bar{\theta}_1}(t - \bar{T}_1, \bar{\nu}_1)$  until the next jump time  $\bar{T}_2$  given by (10) and (11) with  $\bar{\phi}_{\bar{\theta}_1}(T_k^* - \bar{T}_1, \bar{\nu}_1)$  and so on. We proceed by iteration to construct  $(\bar{x}_t)$  on  $[0, T]$ .

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

By classical results of numerical analysis (see [17] for example), the continuous Euler scheme (16) (also called Euler polygon) satisfies estimate (14). If we choose  $\Phi_\theta = \bar{\phi}_\theta$  in the above construction then the corresponding PDP  $(\bar{x}_t)$  is not Markov since the functions  $\bar{\phi}_\theta(\cdot, \nu)$  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

$$K\left((t, \theta, \nu), dudjdz\right) := \mathbb{1}_{u \geq t} \lambda(\theta, \Phi_\theta(u - t, \nu)) e^{-\int_0^{u-t} \lambda(\theta, \Phi_\theta(s, \nu)) ds} Q((\theta, \Phi_\theta(u - t, \nu)), djdz) du. \quad (17)$$

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

$$\lambda(\theta_{n-1}, \Phi_{\theta_{n-1}}(t, \nu_{n-1})) e^{-\int_0^t \lambda(\theta_{n-1}, \Phi(s, \nu_{n-1})) ds}. \quad (18)$$

Classically the marginal distribution of  $x_t$  is expressed using (13), the intensity  $\lambda$  via (18) and the kernel  $K$  (see (17)). Indeed for fixed  $x_0 = x \in E$  and for any bounded measurable function  $g$  we can write,

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



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

$$\int_0^t \int_E K^n((0, x), ds dy) = \int_0^t \int_E \int_{(\mathbb{R}_+ \times E)^{n-1}} K((0, x), dt_1 dy_1) \dots K((t_{n-1}, y_{n-1}), ds dy).$$

However since we have constructed  $(x_t)$  by thinning, we would prefer to express the distribution of  $x_t$  using the upper bound  $\lambda^*$ , the Poisson process  $(N_t^*, t \geq 0)$  and the sequences  $(U_k, k \in \mathbb{N})$ ,  $(V_k, k \in \mathbb{N})$ .

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

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

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 us denote by  $(\Phi, \lambda, Q)$  ( resp.  $(\Phi, \tilde{\lambda}, \tilde{Q})$ ) the characteristics of  $(x_t)$  ( resp.  $(\tilde{x}_t)$ ). Let us assume that  $\tilde{\lambda}$  and  $\tilde{Q}$  depend only on  $\theta$ , 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  $1 \leq \ell \leq n-1$ ,

$$\begin{aligned} \tilde{Z}_\ell &= \frac{Q(\tilde{x}_{T_{\tilde{\tau}_\ell}^*}^-, \tilde{\theta}_\ell)}{\tilde{Q}(\tilde{\theta}_{\ell-1}, \tilde{\theta}_\ell)} \left( \frac{\tilde{\lambda}(\tilde{\theta}_\ell)}{\lambda^*} \left(1 - \frac{\tilde{\lambda}(\tilde{\theta}_\ell)}{\lambda^*}\right)^{\tilde{\tau}_{\ell+1} - \tilde{\tau}_\ell - 1} \right)^{-1} \\ &\quad \frac{\lambda(\tilde{\theta}_\ell, \Phi_{\tilde{\theta}_\ell}(T_{\tilde{\tau}_{\ell+1}}^* - T_{\tilde{\tau}_\ell}^*, \tilde{\nu}_\ell))}{\lambda^*} \prod_{q=\tilde{\tau}_\ell+1}^{\tilde{\tau}_{\ell+1}-1} \left(1 - \frac{\lambda(\tilde{\theta}_\ell, \Phi_{\tilde{\theta}_\ell}(T_q^* - T_{\tilde{\tau}_\ell}^*, \tilde{\nu}_\ell))}{\lambda^*}\right), \\ \tilde{Z}_0 &= \left( \frac{\tilde{\lambda}(\tilde{\theta}_0)}{\lambda^*} \left(1 - \frac{\tilde{\lambda}(\tilde{\theta}_0)}{\lambda^*}\right)^{\tilde{\tau}_1 - 1} \right)^{-1} \frac{\lambda(\tilde{\theta}_0, \Phi_{\tilde{\theta}_0}(T_{\tilde{\tau}_1}^*, \tilde{\nu}_0))}{\lambda^*} \prod_{q=1}^{\tilde{\tau}_1-1} \left(1 - \frac{\lambda(\tilde{\theta}_0, \Phi_{\tilde{\theta}_0}(T_q^*, \tilde{\nu}_0))}{\lambda^*}\right), \\ \tilde{R}_n &= \tilde{Z}_n \prod_{\ell=0}^{n-1} \tilde{Z}_\ell. \end{aligned}$$

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 a Girsanov theorem (see [26]) however we do not use the 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  $\theta$  for readability purposes. Actually the arguments of the proof are valid for non homogeneous intensity and transition measure of the form  $\tilde{\lambda}(x, t)$  and  $\tilde{Q}((x, t), dy)$  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  $\tilde{Q}((x, t), dy) = Q((\theta, \tilde{\Phi}_\theta(t, \nu)), dy)$  for  $\tilde{\Phi}$  a given function. This remark will be implemented in section 5.4.

**Corollary 2.2.** Let  $(\Phi, \lambda, Q)$  (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  $(\mu_n)$  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  $\{N_t = n\}$ ,

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

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

$$\begin{aligned} \tilde{Z}_\ell &= \frac{Q((\theta_{\ell-1}, \mu_\ell), \theta_\ell)}{Q((\theta_{\ell-1}, \nu_\ell), \theta_\ell)} \left( \frac{\lambda(\theta_\ell, \Phi_{\theta_\ell}(T_{\tau_\ell+1}^* - T_{\tau_\ell}^*, \nu_\ell))}{\lambda^*} \prod_{q=\tau_\ell+1}^{\tau_{\ell+1}-1} \left( 1 - \frac{\lambda(\theta_\ell, \Phi_{\theta_\ell}(T_q^* - T_{\tau_\ell}^*, \nu_\ell))}{\lambda^*} \right) \right)^{-1} \\ &\quad \frac{\lambda(\theta_\ell, \tilde{\Phi}_{\theta_\ell}(T_{\tau_\ell+1}^* - T_{\tau_\ell}^*, \mu_\ell))}{\lambda^*} \prod_{q=\tau_\ell+1}^{\tau_{\ell+1}-1} \left( 1 - \frac{\lambda(\theta_\ell, \tilde{\Phi}_{\theta_\ell}(T_q^* - T_{\tau_\ell}^*, \mu_\ell))}{\lambda^*} \right), \\ \tilde{Z}_0 &= \left( \frac{\lambda(\theta_0, \Phi_{\theta_0}(T_{\tau_1}^*, \nu_0))}{\lambda^*} \prod_{q=1}^{\tau_1-1} \left( 1 - \frac{\lambda(\theta_0, \Phi_{\theta_0}(T_q^*, \nu_0))}{\lambda^*} \right) \right)^{-1} \\ &\quad \frac{\lambda(\theta_0, \tilde{\Phi}_{\theta_0}(T_{\tau_1}^*, \mu_0))}{\lambda^*} \prod_{q=1}^{\tau_1-1} \left( 1 - \frac{\lambda(\theta_0, \tilde{\Phi}_{\theta_0}(T_q^*, \mu_0))}{\lambda^*} \right), \\ \tilde{R}_n &= \tilde{Z}_n \prod_{\ell=0}^{n-1} \tilde{Z}_\ell. \end{aligned}$$

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

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

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

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

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

-  $N_t^* = m$ ,

- among the times  $T_\ell^*, 1 \leq \ell \leq m$  exactly  $n$  are accepted by the thinning method they are the  $T_{p_i}^*, 1 \leq i \leq n$ , all the others are rejected.

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

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

The random variable  $\mathbb{1}_{\{\tau_i = p_i, 1 \leq i \leq n\}}$  depends on  $(\theta_\ell, \nu_\ell, 1 \leq \ell \leq n-1, T_i^*, 1 \leq i \leq p_n, U_j, 1 \leq j \leq p_n)$  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  $(\theta_\ell, \nu_\ell, 1 \leq \ell \leq n-1)$  depend on  $(T_i^*, 1 \leq i \leq p_{n-1}, U_j, 1 \leq j \leq p_{n-1}, V_k, 1 \leq k \leq n-1)$ . Thus  $V_n$  is independent of all the other random variables of thinning that are present in  $g(x_t)\mathbb{1}_{\{N_t = n, \tau_i = p_i, 1 \leq i \leq n, N_t^* = m\}}$ . The conditional expectation of  $g(x_t)\mathbb{1}_{\{N_t = n, \tau_i = p_i, 1 \leq i \leq n, N_t^* = m\}}$  w.r.t. the vector  $(T_i^*, 1 \leq i \leq m+1, U_j, 1 \leq j \leq m, V_k, 1 \leq k \leq n-1)$  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  $p_1 < p_2 < \dots < p_n \leq m$ ,

$$\begin{aligned} & \mathbb{E}[g(x_t)\mathbb{1}_{\{N_t = n, \tau_i = p_i, 1 \leq i \leq n, N_t^* = m\}}] \\ &= \mathbb{E}\left[\sum_{\theta \in \Theta} Q(x_{T_{p_n-1}^*}^-, \theta) g(\theta, \Phi_\theta(t - T_{p_n}^*, \nu_n)) \right. \\ & \quad \left. F(\theta, U_j, 1 \leq j \leq m, T_\ell^*, 1 \leq \ell \leq m+1, V_k, 1 \leq k \leq n-1)\right], \end{aligned} \quad (19)$$

with

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

In (19) the random variables  $(U_q, p_n + 1 \leq q \leq m)$  are independent of the vector  $(T_i^*, 1 \leq i \leq m+1, U_j, 1 \leq j \leq p_n, V_k, 1 \leq k \leq n-1)$ . Conditioning by this vector we obtain

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

We can iterate on the latter form by first conditioning  $V_{n-1}$  by all the other r.v. and then conditioning  $(U_q, p_{n-1} + 1 \leq q \leq p_n)$  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 a rejection for

$p_{n-1} + 1 \leq q \leq p_n - 1$  whereas  $U_{p_n}$  corresponds to an acceptance. So that the next step yields

$$\begin{aligned}
& \mathbb{E}[g(x_t) \mathbb{1}_{\{N_t=n, \tau_i=p_i, 1 \leq i \leq n, N_t^*=m\}}] \\
&= \sum_{\alpha \in \Theta} \sum_{\theta \in \Theta} \mathbb{E}[Q(x_{T_{p_n-2}^*}^-, \alpha) Q((\alpha, \nu_n), \theta) g(\theta, \Phi_\theta(t - T_{p_n}^*, \nu_n)) \mathbb{1}_{\{N_t^*=m, \tau_i=p_i, 1 \leq i \leq n-1\}}] \\
& \quad \frac{\lambda(\alpha, \Phi_\alpha(T_{p_n}^* - T_{p_{n-1}}^*, \nu_{n-1}))}{\lambda^*} \prod_{q=p_{n-1}+1}^{p_n-1} \left(1 - \frac{\lambda(\alpha, \Phi_\alpha(T_q^* - T_{p_{n-1}}^*, \nu_{n-1}))}{\lambda^*}\right) \\
& \quad \prod_{q=p_n+1}^m \left(1 - \frac{\lambda(\theta, \Phi_\theta(T_q^* - T_{p_n}^*, \nu_n))}{\lambda^*}\right)], \tag{20}
\end{aligned}$$

where we write  $\nu_n$  for simplicity keeping in mind that  $\nu_n = \Phi_{\theta_{n-1}}(T_{p_n}^* - T_{p_{n-1}}^*, \nu_{n-1}) = \Phi_{\theta_{n-1}}(T_{p_n}^* - T_{p_{n-1}}^*, \Phi_{\theta_{n-2}}(T_{p_{n-1}}^* - T_{p_{n-2}}^*, \nu_{n-2})) = \Phi_\alpha(T_{p_n}^* - T_{p_{n-1}}^*, \Phi_{\theta_{n-2}}(T_{p_{n-1}}^* - T_{p_{n-2}}^*, \nu_{n-2}))$ . Moreover the previous arguments apply to  $\mathbb{E}[g(x_t) f(\theta_i, \nu_i, 1 \leq i \leq n-1, \theta_n, \nu_n, T_k^*, 1 \leq k \leq m) \mathbb{1}_{\{N_t=n, \tau_i=p_i, 1 \leq i \leq n, N_t^*=m\}}]$  and provide

$$\begin{aligned}
& \mathbb{E}[g(x_t) f(\theta_i, \nu_i, 1 \leq i \leq n-1, \theta_n, \nu_n, T_k^*, 1 \leq k \leq m) \mathbb{1}_{\{N_t=n, \tau_i=p_i, 1 \leq i \leq n, N_t^*=m\}}] \\
&= \sum_{\theta \in \Theta} \mathbb{E}[Q(x_{T_{p_n-1}}^-, \theta) g(\theta, \Phi_\theta(t - T_{p_n}^*, \nu_n)) f(\theta_i, \nu_i, 1 \leq i \leq n-1, \theta, \nu_n, T_k^*, 1 \leq k \leq m) \\
& \quad \mathbb{1}_{\{N_t^*=m, \tau_i=p_i, 1 \leq i \leq n\}} \prod_{q=p_n+1}^m \left(1 - \frac{\lambda(\theta, \Phi_\theta(T_q^* - T_{p_n}^*, \nu_n))}{\lambda^*}\right)]. \tag{21}
\end{aligned}$$

□

We prove below 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  $\theta$ . Let  $p_1 < p_2 < \dots < p_n \leq m$ . Applying the same arguments as in (21) to  $(\tilde{x}_t)$  and using the definitions of  $\tilde{Z}_\ell$ ,  $0 \leq \ell \leq n$  and  $\tilde{R}_n$  we obtain,

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

We iterate the previous argument based on the use of (21) and we use the definition of  $\tilde{Z}_{n-1}$  to obtain

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

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 (20) and using an induction we conclude that

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

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

### 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, 3.3 respectively.

**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  $\theta$ .

**Theorem 3.1.** Let  $\Phi_\theta$  and  $\bar{\Phi}_\theta$  satisfying (14) 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  $\Theta$  is finite and that  $\lambda$  and  $Q$  satisfy Assumption 3.1. Then, for all bounded functions  $F : E \rightarrow \mathbb{R}$  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  $\theta$ , there exists constants  $V_1 > 0$  and  $V_2 > 0$  independent of the time step  $h$  such that

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

**Remark 3.1.** When the numerical scheme  $\bar{\Phi}_\theta$  is of order  $p \geq 1$ , which means  $\sup_{t \in [0, T]} |\Phi_\theta(t, \nu_1) - \bar{\Phi}_\theta(t, \nu_2)| \leq e^{C_1 T} |\nu_1 - \nu_2| + C_2 h^p$  we have  $\mathbb{E}[|F(\bar{x}_T) - F(x_T)|^2] \leq V_1 h^p + V_2 h^{2p}$ .

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

**Theorem 3.2.** Let  $\Phi_\theta$  and  $\bar{\Phi}_\theta$  satisfying (14) and let  $(\tilde{x}_t, t \in [0, T])$  and  $(\tilde{\underline{x}}_t, t \in [0, T])$  be the corresponding PDPs constructed in section 2.1 with  $\tilde{\underline{x}}_0 = \tilde{x}_0 = x$  for some  $x \in E$ . Let  $(\tilde{R}_t, t \in [0, T])$  and  $(\tilde{\underline{R}}_t, t \in [0, T])$  be defined as in Corollary 2.1. Under assumptions 3.1 and 3.2 and for all bounded functions  $F : E \rightarrow \mathbb{R}$  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

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

where  $\tilde{\underline{R}}_T$  has been defined in Corollary 2.1.

We now introduce the random variable  $\bar{\tau}^\dagger$  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  $\bar{\tau}^\dagger := \inf \{k > 0 : (\tau_k, \theta_k) \neq (\bar{\tau}_k, \bar{\theta}_k)\}$ .*

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

$$\{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) > T\} = \left\{ N_T = \bar{N}_T, (T_1, \theta_1) = (\bar{T}_1, \bar{\theta}_1), \dots, (T_{N_T}, \theta_{N_T}) = (\bar{T}_{N_T}, \bar{\theta}_{N_T}) \right\}, \quad (22)$$

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

### 3.1 Preliminary lemmas

In this section we start with two lemmas which will be useful to prove Theorems 3.2 and 3.3.

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

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

where  $U \sim \mathcal{U}([0, 1])$ ,  $G(u) = \sum_{j=1}^{|K|} k_j \mathbb{1}_{a_{j-1} < u \leq a_j}$  and  $\bar{G}(u) = \sum_{j=1}^{|K|} k_j \mathbb{1}_{\bar{a}_{j-1} < u \leq \bar{a}_j}$  for all  $u \in [0, 1]$ . Then, we have

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

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

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

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

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

Thus, denoting by  $x^+ := \max(x, 0)$  the positive part of  $x \in \mathbb{R}$  and using that  $x^+ \geq x$ , we obtain

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

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

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

The first sum above is a telescopic sum. Since  $a_{|K|} = \bar{a}_{|K|} = 1$  and  $a_0 = \bar{a}_0 = 0$ , we have  $\mathbb{P}(X = \bar{X}) \geq 1 - \sum_{j=1}^{|K|-1} |a_j - \bar{a}_j|$ . □

**Lemma 3.2.** *Let  $(a_n, n \geq 1)$  and  $(b_n, n \geq 1)$  be two real-valued sequences. For all  $n \geq 1$ , 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.* By induction. □

### 3.2 Proof of Theorem 3.1

First, we write

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

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

*Step 1: estimation of  $\bar{P}$ .* The function  $F$  being bounded we have  $\bar{P} \leq 4M_F^2 \mathbb{P}(\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) \leq T)$  where  $M_F > 0$ . Moreover, for  $k \geq 1$ ,  $\{\bar{\tau}^\dagger = k\} = \{\bar{\tau}^\dagger > k-1\} \cap \{(\tau_k, \theta_k) \neq (\bar{\tau}_k, \bar{\theta}_k)\}$ . Hence

$$\begin{aligned} \mathbb{P}(\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) \leq T) &= \sum_{k \geq 1} \mathbb{E} \left[ \mathbb{1}_{\min(T_k, \bar{T}_k) \leq T} \mathbb{1}_{\bar{\tau}^\dagger = k} \right] \\ &= \sum_{k \geq 1} \mathbb{E} \left[ \mathbb{1}_{\min(T_k, \bar{T}_k) \leq T} \mathbb{1}_{\bar{\tau}^\dagger > k-1} \mathbb{1}_{(\tau_k, \theta_k) \neq (\bar{\tau}_k, \bar{\theta}_k)} \right] \\ &\leq \sum_{k \geq 1} \bar{J}_k + 2\bar{I}_k \end{aligned}$$

where

$$\bar{J}_k := \mathbb{E} \left[ \mathbb{1}_{\min(T_k, \bar{T}_k) \leq T} \mathbb{1}_{\bar{\tau}^\dagger > k-1} \mathbb{1}_{\tau_k = \bar{\tau}_k} \mathbb{1}_{\theta_k \neq \bar{\theta}_k} \right], \quad \bar{I}_k := \mathbb{E} \left[ \mathbb{1}_{\min(T_k, \bar{T}_k) \leq T} \mathbb{1}_{\bar{\tau}^\dagger > k-1} \mathbb{1}_{\tau_k \neq \bar{\tau}_k} \right]. \quad (23)$$

We start with  $\bar{J}_k$ . First note that, for  $k \geq 1$ ,  $\{\tau_k = \bar{\tau}_k\} = \{T_k = \bar{T}_k\}$  and that on the event  $\{T_k = \bar{T}_k\}$ , we have  $\min(T_k, \bar{T}_k) = T_k$ , so that  $\bar{J}_k = \mathbb{E} \left[ \mathbb{1}_{T_k \leq T} \mathbb{1}_{\bar{\tau}^\dagger > k-1} \mathbb{1}_{\tau_k = \bar{\tau}_k} \mathbb{1}_{\theta_k \neq \bar{\theta}_k} \right]$ . We

emphasize that it makes no difference in the rest of the proof if we choose  $\min(T_k, \bar{T}_k) = \bar{T}_k$ . Since  $\{\bar{\tau}^\dagger > k-1\} = \bigcap_{i=0}^{k-1} \{(\tau_i, \theta_i) = (\bar{\tau}_i, \bar{\theta}_i)\}$ , we can rewrite  $\bar{J}_k$  as follows

$$\sum_{\substack{1 \leq p_1 < \dots < p_k \\ \alpha_1, \dots, \alpha_{k-1} \in \Theta}} \mathbb{E}[\mathbb{1}_{\{\tau_i = \bar{\tau}_i = p_i, 1 \leq i \leq k\}} \mathbb{1}_{\{\theta_i = \bar{\theta}_i = \alpha_i, 1 \leq i \leq k-1\}} \mathbb{1}_{T_{p_k}^* \leq T} \mathbb{1}_{\theta_k \neq \bar{\theta}_k}]. \quad (24)$$

By construction we have  $\theta_k = H((\theta_{k-1}, \nu_k), V_k)$  and  $\bar{\theta}_k = H((\bar{\theta}_{k-1}, \bar{\nu}_k), V_k)$ . The random variable  $\mathbb{1}_{\{\tau_i = \bar{\tau}_i = p_i, 1 \leq i \leq k\}} \mathbb{1}_{\{\theta_i = \bar{\theta}_i = \alpha_i, 1 \leq i \leq k-1\}} \mathbb{1}_{T_{p_k}^* \leq T}$  depends on the vector  $(U_i, 1 \leq i \leq p_k, T_j^*, 1 \leq j \leq p_k, V_q, 1 \leq q \leq k-1)$  which is independent of  $V_k$ . Conditioning by this vector in (24) and applying Lemma 3.1 yields

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

From the definition of  $a_j$  (see (12)), the triangle inequality and since  $Q$  is  $L_Q$ -Lipschitz, we have  $\sum_{j=1}^{|\Theta|-1} |a_j(\alpha_{k-1}, \bar{\nu}_k) - a_j(\alpha_{k-1}, \nu_k)| \leq \frac{(|\Theta|-1)|\Theta|}{2} L_Q |\bar{\nu}_k - \nu_k|$ . Since we are on the event  $\{\tau_i = \bar{\tau}_i = p_i, 1 \leq i \leq k\} \cap \{\theta_i = \bar{\theta}_i = \alpha_i, 1 \leq i \leq k-1\}$ , the application of Lemma 2.1 yields  $|\bar{\nu}_k - \nu_k| \leq e^{LT_{p_k}^*} kCh$ . Thus  $\bar{J}_k \leq C_1 h \mathbb{E}[\mathbb{1}_{T_k \leq Tk}]$  where  $C_1$  is a constant independent of  $h$ . Moreover,  $\sum_{k \geq 1} \mathbb{1}_{T_k \leq Tk} = \sum_{k=1}^{N_T} k \leq N_T^2$  and  $\mathbb{E}[N_T^2] \leq \mathbb{E}[(N_T^*)^2] < +\infty$  so that  $\sum_{k \geq 1} \bar{J}_k = O(h)$ . From the definition of  $\bar{I}_k$  (see (23)), we can write

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

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

$$\sum_{\substack{1 \leq p_1 < \dots < p_k \\ \alpha_1, \dots, \alpha_{k-1} \in \Theta}} \mathbb{E}[\mathbb{1}_{\{\tau_i = \bar{\tau}_i = p_i, 1 \leq i \leq k-1\}} \mathbb{1}_{\{\theta_i = \bar{\theta}_i = \alpha_i, 1 \leq i \leq k-1\}} \mathbb{1}_{T_{p_k}^* \leq T} \mathbb{1}_{\tau_k = p_k} \mathbb{1}_{p_k < \bar{\tau}_k}]. \quad (25)$$

In (25) we have  $\{\tau_k = p_k\} \cap \{p_k < \bar{\tau}_k\} \subseteq \{\lambda(\alpha_{k-1}, \bar{\Phi}_{\alpha_{k-1}}(T_{p_k}^* - T_{p_{k-1}}^*, \bar{\nu}_{k-1})) < U_{p_k} \lambda^* \leq \lambda(\alpha_{k-1}, \Phi_{\alpha_{k-1}}(T_{p_k}^* - T_{p_{k-1}}^*, \nu_{k-1}))\}$ . The random variable  $\mathbb{1}_{\{\tau_i = \bar{\tau}_i = p_i, 1 \leq i \leq k-1\}} \mathbb{1}_{\{\theta_i = \bar{\theta}_i = \alpha_i, 1 \leq i \leq k-1\}} \mathbb{1}_{T_{p_k}^* \leq T}$  depends on  $(U_i, 1 \leq i \leq p_{k-1}, T_j^*, 1 \leq j \leq p_k, V_q, 1 \leq q \leq k-1)$  which is independent of  $U_{p_k}$ . Conditioning by this vector in (25) yields

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

Using the Lipschitz continuity of  $\lambda$  then Lemma 2.1 we get that  $\bar{I}_k^{(1)} \leq C_2 h \mathbb{E}[\mathbb{1}_{T_k \leq Tk}]$  where  $C_2$  is a constant independent of  $h$ . Concerning the term  $\bar{I}_k^{(2)}$ , we will end with the estimate



$\bar{T}_k^{(2)} \leq C_2 h \mathbb{E}[\mathbb{1}_{\bar{T}_k \leq T} k]$ . We conclude in the same way as in the estimation of  $\bar{J}_k$  above that  $\sum_{k \geq 1} \bar{T}_k = O(h)$ .

*Step 2: estimation of  $\bar{D}$ .* Note that for  $n \geq 0$  we have  $\{N_T = n\} \cap \{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) > T\} = \{N_T = n\} \cap \{\bar{N}_T = n\} \cap \{\bar{\tau}^\dagger > n\}$ , where we can interchange the role of  $\{N_T = n\}$  and  $\{\bar{N}_T = n\}$ . Thus, using the partition  $\{N_T = n, n \geq 0\}$ , we have

$$\bar{D} = \sum_{n \geq 0} \mathbb{E} \left[ \mathbb{1}_{N_T = n} \mathbb{1}_{\bar{N}_T = n} \mathbb{1}_{\bar{\tau}^\dagger > n} \left| F(\theta_n, \bar{\Phi}_{\theta_n}(T - T_n, \bar{\nu}_n)) - F(\theta_n, \Phi_{\theta_n}(T - T_n, \nu_n)) \right|^2 \right]$$

The application of the Lipschitz continuity of  $F$  and of Lemma 2.1 yields

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

Then, we have  $\bar{D} \leq C_3 h^2 \sum_{n \geq 0} \mathbb{E} [\mathbb{1}_{N_T = n} (n+1)^2]$  where  $C_3$  is a constant independent of  $h$ . Since  $\sum_{n \geq 0} \mathbb{E} [\mathbb{1}_{N_T = n} (n+1)^2] = \mathbb{E}[(N_T + 1)^2] \leq \mathbb{E}[(N_T^* + 1)^2] < +\infty$ , we conclude that  $\bar{D} = O(h^2)$ .  $\square$

### 3.3 Proof of Theorem 3.2

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

$$\tilde{Q}_T = \prod_{l=1}^{\tilde{N}_T} \frac{Q(\tilde{x}_{T_{\tilde{\tau}_l}^*}, \tilde{\theta}_l)}{\tilde{Q}(\tilde{\theta}_{l-1}, \tilde{\theta}_l)}, \quad (26)$$

$$\tilde{S}_T = \prod_{l=1}^{\tilde{N}_T} \frac{\lambda(\tilde{\theta}_{l-1}, \Phi_{\tilde{\theta}_{l-1}}(T_{\tilde{\tau}_l}^* - T_{\tilde{\tau}_{l-1}}^*, \tilde{\nu}_{l-1}))}{\lambda^*} \prod_{k=\tilde{\tau}_{l-1}+1}^{\tilde{\tau}_l} \left( 1 - \frac{\lambda(\tilde{\theta}_{l-1}, \Phi_{\tilde{\theta}_{l-1}}(T_k^* - T_{\tilde{\tau}_{l-1}}^*, \tilde{\nu}_{l-1}))}{\lambda^*} \right) \quad (27)$$

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

$$\tilde{H}_T = \prod_{l=1}^{\tilde{N}_T} \left( \frac{\tilde{\lambda}(\tilde{\theta}_{l-1})}{\lambda^*} \left( 1 - \frac{\tilde{\lambda}(\tilde{\theta}_{l-1})}{\lambda^*} \right)^{\tilde{\tau}_l - \tilde{\tau}_{l-1} - 1} \right)^{-1} \left( \left( 1 - \frac{\tilde{\lambda}(\tilde{\theta}_{\tilde{N}_T})}{\lambda^*} \right)^{N_T^* - \tilde{\tau}_{\tilde{N}_T}} \right)^{-1}. \quad (28)$$

Likewise we reorder the terms in  $\underline{\tilde{R}}_T$  writing  $\underline{\tilde{R}}_T = \underline{\tilde{Q}}_T \underline{\tilde{S}}_T \underline{\tilde{H}}_T$  where  $\underline{\tilde{Q}}_T$  and  $\underline{\tilde{S}}_T$  are defined as (26) and (27) replacing  $\tilde{x}$  and  $\Phi$  by  $\underline{\tilde{x}}$  and  $\underline{\Phi}$ . Since the processes  $(\tilde{\theta}_n)$  and  $(\tilde{\tau}_n)$  do not depend on  $\Phi$  or  $\bar{\Phi}$ , the term  $\tilde{H}$  is the same in  $\tilde{R}$  and  $\underline{\tilde{R}}$ . To prove Theorem 3.2, let us decompose the problem and write

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

so that

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

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

*Step 1: estimation of  $\overline{C}$ .* The function  $F$  being bounded we have  $\overline{C} \leq M_F^2 \mathbb{E} [|\tilde{R}_T - \tilde{R}_T|^2]$  where  $M_F$  is a positive constant. Moreover, for all  $\theta \in \Theta$ , we have  $(1 - \tilde{\lambda}(\theta)/\lambda^*)^{-1} \leq (1 - \tilde{\lambda}_{\max}/\lambda^*)^{-1}$  and  $(\tilde{\lambda}(\theta)/\lambda^*)^{-1} \leq (\tilde{\lambda}_{\min}/\lambda^*)^{-1}$ . Thus,  $\tilde{H}_T \leq \left(\frac{\tilde{\lambda}_{\min}}{\lambda^*}(1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*})\right)^{-N_T^*}$  and using the definition of  $\tilde{R}$  and  $\tilde{R}$  (see (26), (27) and (28)) we can write

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

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

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

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

$$|\lambda(\tilde{\theta}_{l-1}, \overline{\Phi}_{\tilde{\theta}_{l-1}}(T_k^* - T_{\tilde{\tau}_{l-1}}^*, \tilde{\nu}_{l-1})) - \lambda(\tilde{\theta}_{l-1}, \Phi_{\tilde{\theta}_{l-1}}(T_k^* - T_{\tilde{\tau}_{l-1}}^*, \tilde{\nu}_{l-1}))| \leq e^{LT} C h l.$$

Moreover, for all  $l = 1, \dots, \tilde{N}_T + 1$  we have  $\tilde{\tau}_l \wedge N_T^* - \tilde{\tau}_{l-1} \leq N_T^*$  so that  $|\tilde{S}_T - \tilde{S}_T| \leq N_T^* (N_T^* + 1)^2 C_1 h$  where  $C_1$  is a positive constant independent of  $h$ . Finally, since  $\tilde{Q}_T \leq \rho^{-N_T^*}$  we have

$$\overline{I} \leq \rho^{-N_T^*} N_T^* (N_T^* + 1)^2 C_1 h. \quad (29)$$

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

$$\overline{J} \leq \rho^{-N_T^*} (N_T^*)^2 C_2 h, \quad (30)$$

where  $C_2$  is a positive constant independent of  $h$ . Then, we derive from the previous estimates (29) and (30) that

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

where  $\Xi_1(n) = \left(\rho \frac{\tilde{\lambda}_{\min}}{\lambda^*} (1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*})\right)^{-n} n(n+1)^2$  and  $C_3 = \max(C_1, C_2)$ . Finally, we have  $\mathbb{E}[|\tilde{R}_T - \tilde{R}_T|^2] \leq C_3 h^2 \mathbb{E}[\Xi_1(N_T^*)^2]$ . Since  $\mathbb{E}[\Xi_1(N_T^*)^2] < +\infty$  we conclude that  $\overline{C} = O(h^2)$ .

*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  $\tilde{x}_T = (\tilde{\theta}_{\tilde{N}_T}, \overline{\Phi}_{\tilde{\theta}_{\tilde{N}_T}}(T - \tilde{T}_{\tilde{N}_T}, \tilde{\nu}_{\tilde{N}_T}))$ . Then, using the Lipschitz continuity of  $F$ , Lemma 2.1 and since  $\tilde{N}_T \leq N_T^*$  we get

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

Moreover,  $|\tilde{R}_T| \leq \left(\rho \frac{\tilde{\lambda}_{\min}}{\lambda^*} (1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*})\right)^{-N_T^*}$  so that  $\overline{D} \leq C_4 h^2 \mathbb{E}[\Xi_2(N_T^*)^2]$  where  $C_4$  is a positive constant independent of  $h$  and  $\Xi_2(n) = (n+1) \left(\rho \frac{\tilde{\lambda}_{\min}}{\lambda^*} (1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*})\right)^{-n}$ . Since  $\mathbb{E}[\Xi_2(N_T^*)^2] < +\infty$  we conclude that  $\overline{D} = O(h^2)$ .  $\square$

## 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

$$\mathcal{A}g(t, x) = \partial_t g(t, x) + f(x)\partial_\nu g(t, x) + \lambda(x) \int_E (g(t, y) - g(t, x))Q(x, dy), \quad (31)$$

where for notational convenience we have set  $\partial_\nu g(t, x) := \frac{\partial g}{\partial \nu}(t, \theta, \nu)$ ,  $\partial_t g(t, x) := \frac{\partial g}{\partial t}(t, x)$  and  $f(x) = f_\theta(\nu)$  for all  $x = (\theta, \nu) \in E$ . 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 delayed in 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*

$$\begin{cases} \mathcal{A}u(t, x) = 0, & (t, x) \in [0, T] \times E, \\ u(T, x) = F(x), & x \in E, \end{cases} \quad (32)$$

with  $F : E \rightarrow \mathbb{R}$  a bounded function and  $\mathcal{A}$  given by (31) is such that for all  $\theta \in \Theta$ , the function  $(t, \nu) \mapsto u(t, \theta, \nu)$  is bounded and two times differentiable with bounded derivatives. Moreover the second derivatives of  $(t, \nu) \mapsto u(t, \theta, \nu)$  are uniformly Lipschitz in  $\theta$ .

**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  $F : E \rightarrow \mathbb{R}$  there exists a constant  $c_1$  independent of  $h$  such that*

$$\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = hc_1 + O(h^2). \quad (33)$$

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

$$\mathbb{E}[F(\tilde{\underline{x}}_T)\tilde{\underline{R}}_T] - \mathbb{E}[F(\tilde{x}_T)\tilde{R}_T] = hc_1 + O(h^2). \quad (34)$$

### 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  $\mathbb{R}_+ \times E$ .*

$$\mathcal{T}g(t, x) := \partial_t g(t, x) + f(x)\partial_\nu g(t, x),$$

$$\mathcal{S}g(t, x) := \lambda(x) \int_E (g(t, y) - g(t, x))Q(x, dy).$$

From Definition 4.1, the generator  $\mathcal{A}$  defined by (31) reads  $\mathcal{A}g(t, x) = \mathcal{T}g(t, x) + \mathcal{S}g(t, x)$ . We introduce the random counting measure  $p$  associated to 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  $p$ , noted  $p'$ , is given from [5] by

$$p'([0, t] \times A) = \int_0^t \lambda(x_s)Q(x_s, A)ds.$$

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

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

$$\begin{aligned}\overline{\mathcal{T}}g(t, x, y) &:= \partial_t g(t, x) + f(y) \partial_\nu g(t, x), \\ \overline{\mathcal{A}}g(t, x, y) &:= \overline{\mathcal{T}}g(t, x, y) + \mathcal{S}g(t, x).\end{aligned}$$

**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  $(\overline{x}_t)$ . 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  $k \in \{0, \dots, \lfloor (\overline{T}_{n+1} - \overline{T}_n)/h \rfloor\}$ .

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

$$g(t, x_t) = g(0, x) + \int_0^t \mathcal{A}g(s, x_s) ds + M_t^g, \quad (35)$$

where  $M_t^g := \int_0^t \int_E (g(s, y) - g(s, x_{s-})) q(ds dy)$  is a true  $\mathcal{F}_t^P$ -martingale, and

$$g(t, \overline{x}_t) = g(0, x) + \int_0^t \overline{\mathcal{A}}g(s, \overline{x}_s, \overline{x}_{\overline{\eta}(s)}) ds + \overline{M}_t^g, \quad (36)$$

where,  $\overline{M}_t^g := \int_0^t \int_E (g(s, y) - g(s, \overline{x}_{s-})) \overline{q}(ds dy)$  is a true  $\mathcal{F}_t^{\overline{P}}$ -martingale.

*Proof of Theorem 4.2.* The proof of (35) is given in [5]. We prove (36) following the same arguments. Since  $\overline{q} = \overline{p} - \overline{p}'$ , we have

$$\overline{M}_t^g = \sum_{k \geq 1} \mathbf{1}_{\overline{T}_k \leq t} \left( g(\overline{T}_k, \overline{x}_{\overline{T}_k}) - g(\overline{T}_k, \overline{x}_{\overline{T}_k}^-) \right) - \int_0^t \mathcal{S}g(s, \overline{x}_s) ds.$$

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

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

For all  $k \leq n-1$ , 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  $i \geq 0$ ,  $t \in [ih, (i+1)h]$  and for all  $x = (\theta, \nu) \in E$  where we recall that  $\overline{\phi}$  is defined by (16). The function  $g$  is smooth enough to write

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

Then, the above arguments together with definition 4.2 yields

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

□

The following theorem gives us a way to represent the solution of the integro-differential equation (32) 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  $F : E \rightarrow \mathbb{R}$  be a bounded function. Then the integro-differential equation (32) has a unique solution  $u : \mathbb{R}_+ \times E \rightarrow \mathbb{R}$  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  $\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)]$ . Then, we use this representation to identify the coefficient  $c_1$  in (33).

*Step 1: Representing  $\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)]$ .* Let  $u$  denote the solution of (32). From Theorem 4.3 we can write  $\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = \mathbb{E}[u(T, \bar{x}_T)] - u(0, x)$ . Then, the application of the Itô formula (36) to  $u$  at time  $T$  yields

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

Since  $(\bar{M}_t^u)$  is a true martingale, we obtain

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

For  $s \in [0, T]$  we have  $\bar{\mathcal{A}}u(s, \bar{x}_s, \bar{x}_{\bar{\eta}(s)}) = \partial_t u(s, \bar{x}_s) + f(\bar{x}_{\bar{\eta}(s)})\partial_\nu u(s, \bar{x}_s) + \mathcal{S}u(s, \bar{x}_s)$  (see Definition 4.2). From the regularity of  $\lambda$ ,  $Q$  and  $u$  (see assumptions 4.1 and 4.2), the functions  $\partial_t u$ ,  $\partial_\nu u$  and  $\mathcal{S}u$  are smooth enough to apply the Itô formula (36) between  $\bar{\eta}(s)$  and  $s$  respectively. This yields

$$\begin{aligned} \partial_t u(s, \bar{x}_s) &= \partial_t u(\bar{\eta}(s), \bar{x}_{\bar{\eta}(s)}) + \int_{\bar{\eta}(s)}^s \bar{\mathcal{A}}(\partial_t u)(r, \bar{x}_r, \bar{x}_{\bar{\eta}(r)})dr + \bar{M}_s^{\partial_t u} - \bar{M}_{\bar{\eta}(s)}^{\partial_t u}, \\ \partial_\nu u(s, \bar{x}_s) &= \partial_\nu u(\bar{\eta}(s), \bar{x}_{\bar{\eta}(s)}) + \int_{\bar{\eta}(s)}^s \bar{\mathcal{A}}(\partial_\nu u)(r, \bar{x}_r, \bar{x}_{\bar{\eta}(r)})dr + \bar{M}_s^{\partial_\nu u} - \bar{M}_{\bar{\eta}(s)}^{\partial_\nu u}, \\ \mathcal{S}u(s, \bar{x}_s) &= \mathcal{S}u(\bar{\eta}(s), \bar{x}_{\bar{\eta}(s)}) + \int_{\bar{\eta}(s)}^s \bar{\mathcal{A}}(\mathcal{S}u)(r, \bar{x}_r, \bar{x}_{\bar{\eta}(r)})ds + \bar{M}_s^{\mathcal{S}u} - \bar{M}_{\bar{\eta}(s)}^{\mathcal{S}u}. \end{aligned}$$

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

$$\begin{aligned} f(\bar{x}_{\bar{\eta}(s)})\partial_\nu u(s, \bar{x}_s) &= f(\bar{x}_{\bar{\eta}(s)})\partial_\nu u(\bar{\eta}(s), \bar{x}_{\bar{\eta}(s)}) \\ &\quad + \int_{\bar{\eta}(s)}^s f(\bar{x}_{\bar{\eta}(r)})\bar{\mathcal{A}}(\partial_\nu u)(r, \bar{x}_r, \bar{x}_{\bar{\eta}(r)})dr + f(\bar{x}_{\bar{\eta}(s)})(\bar{M}_s^{\partial_\nu u} - \bar{M}_{\bar{\eta}(s)}^{\partial_\nu u}), \end{aligned}$$

so that

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

where,

$$\Upsilon(t, x, y) := (\bar{\mathcal{A}}(\partial_t u) + f(y)\bar{\mathcal{A}}(\partial_\nu u) + \bar{\mathcal{A}}(\mathcal{S}u))(t, x, y). \quad (37)$$

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

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

Moreover, since  $(\bar{M}_t^{\partial_\nu u})$  is a  $\mathcal{F}_t^{\bar{\mathcal{P}}}$ -martingale, we have

$$\mathbb{E} \left[ \int_0^T f(\bar{x}_{\bar{\eta}(s)}) (\bar{M}_s^{\partial_\nu u} - \bar{M}_{\bar{\eta}(s)}^{\partial_\nu u}) ds \right] = \int_0^T \mathbb{E} \left[ f(\bar{x}_{\bar{\eta}(s)}) \mathbb{E}[\bar{M}_s^{\partial_\nu u} - \bar{M}_{\bar{\eta}(s)}^{\partial_\nu u} | \mathcal{F}_{\bar{\eta}(s)}^{\bar{\mathcal{P}}}] \right] ds = 0.$$

Collecting the previous results, we obtain  $\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = \mathbb{E} \left[ \int_0^T \int_{\bar{\eta}(s)}^s \Upsilon(r, \bar{x}_r, \bar{x}_{\bar{\eta}(r)}) dr ds \right]$ . We can compute an explicit form of  $\Upsilon$  in term of  $u, f, \lambda, Q$  and their derivatives. Indeed,  $\Upsilon$  is given by (37), and we have

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

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

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

Consider the expectation in the right-hand side of the above equality. We decompose the integral into a (finite) sum of integrals on the intervals  $[\bar{T}_n + kh, (\bar{T}_n + (k+1)h) \wedge \bar{T}_{n+1}]$  where  $\Psi$  is constant. Without loss of generality, we are led to consider integrals of the form  $\int_{kh}^t (s - kh)C ds$  for some  $k \geq 0, t \in [kh, (k+1)h]$  and  $C$  a bounded constant. We have  $\int_{kh}^t (s - kh)C ds = \frac{t-kh}{2} \int_{kh}^t C ds$  moreover adding and subtracting  $h$  in the numerator of  $(t - kh)/2$  yields

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

Since  $C$  is bounded we deduce that  $\int_{kh}^t (s - kh)C ds = \frac{h}{2} \int_{kh}^t C ds + O(h^2)$ . Since  $\Psi$  is assumed bounded and  $\mathbb{E}[\bar{N}_T] < +\infty$ , the above arguments yields the following representation

$$\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = \frac{h}{2} \mathbb{E} \left[ \int_0^T \Psi(\bar{\eta}(s), \bar{x}_{\bar{\eta}(s)}) ds \right] + O(h^2). \quad (38)$$

*Step 2: From the representation (38) to the expansion at the order one.* In this step, we show that  $\mathbb{E} \left[ \int_0^T \Psi(\bar{\eta}(s), \bar{x}_{\bar{\eta}(s)}) ds \right] = \mathbb{E} \left[ \int_0^T \Psi(s, x_s) ds \right] + O(h)$ . First, we introduce the random variables  $\bar{\Gamma}$  and  $\Gamma$  defined by  $\bar{\Gamma} := \int_0^T \Psi(\bar{\eta}(s), \bar{x}_{\bar{\eta}(s)}) ds$  and  $\Gamma := \int_0^T \Psi(\bar{\eta}(s), x_{\bar{\eta}(s)}) ds$  and write

$$\mathbb{E}[|\bar{\Gamma} - \Gamma|] = \mathbb{E} \left[ \mathbf{1}_{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) \leq T} |\bar{\Gamma} - \Gamma| \right] + \mathbb{E} \left[ \mathbf{1}_{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) > T} |\bar{\Gamma} - \Gamma| \right],$$

where  $\bar{\tau}^\dagger$  is defined in Definition 3.1. Since  $\Psi$  is bounded and  $\mathbb{P}(\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) \leq T) = O(h)$  (see the proof of Theorem 3.1), we have  $\mathbb{E} \left[ |\bar{\Gamma} - \Gamma| \mathbb{1}_{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) \leq T} \right] = O(h)$ . Now, recall from (22) that, on the event  $\{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) > T\}$ , we have  $T_k = \bar{T}_k$  and  $\theta_k = \bar{\theta}_k$  for all  $k \geq 1$  such that  $T_k \in [0, T]$ . Thus, for all  $n \leq \bar{N}_T$  and for all  $s \in [\bar{T}_n, \bar{T}_{n+1}[$  we have  $\bar{x}_{\bar{\eta}(s)} = (\bar{\theta}_n, \bar{\phi}_{\bar{\theta}_n}(\bar{\eta}(s) - \bar{T}_n, \bar{\nu}_n))$  and  $x_{\bar{\eta}(s)} = (\bar{\theta}_n, \phi_{\bar{\theta}_n}(\bar{\eta}(s) - \bar{T}_n, \nu_n))$ . Consequently, on the event  $\{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) > T\}$  we have

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

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

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

From the above inequality, we find that  $\mathbb{E} \left[ \mathbb{1}_{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) > T} |\bar{\Gamma} - \Gamma| \right] \leq L_\Psi C e^{LT} T h \mathbb{E}[\bar{N}_T(\bar{N}_T + 1)]$ .

Since  $\bar{N}_T \leq N_T^*$  and  $\mathbb{E}[N_T^*(N_T^* + 1)] < +\infty$  we conclude that  $\mathbb{E} \left[ \mathbb{1}_{\min(T_{\bar{\tau}^\dagger}, \bar{T}_{\bar{\tau}^\dagger}) > T} |\bar{\Gamma} - \Gamma| \right] = O(h)$ .

We have shown that  $\mathbb{E} \left[ \int_0^T \Psi(\bar{\eta}(s), \bar{x}_{\bar{\eta}(s)}) ds \right] = \mathbb{E} \left[ \int_0^T \Psi(\bar{\eta}(s), x_{\bar{\eta}(s)}) ds \right] + O(h)$ . Secondly, from the regularity assumptions 4.1 and 4.2, the function  $(t, \nu) \mapsto \Psi(t, \theta, \nu)$  is uniformly Lipschitz in  $\theta$ . Moreover, for all  $s \in [0, T]$  there exists  $k \geq 0$  such that both  $s$  and  $\bar{\eta}(s)$  belong to the same interval  $[\bar{T}_k, \bar{T}_{k+1}[$  so that  $x_s = (\theta_k, \phi_{\theta_k}(s - \bar{T}_k, \nu_k))$  and  $x_{\bar{\eta}(s)} = (\theta_k, \phi_{\theta_k}(\bar{\eta}(s) - \bar{T}_k, \nu_k))$ . Thus, from the Lipschitz continuity of  $\Psi$ , from the fact that  $|s - \bar{\eta}(s)| \leq h$  and since  $f_\theta$  is uniformly bounded in  $\theta$  we have  $|\Psi(s, x_s) - \Psi(\bar{\eta}(s), x_{\bar{\eta}(s)})| \leq Ch$  where  $C$  is a constant independent of  $h$ . Then, we obtain  $\sup_{s \in [0, T]} |\mathbb{E}[\Psi(s, x_s)] - \mathbb{E}[\Psi(\bar{\eta}(s), x_{\bar{\eta}(s)})]| \leq Ch$  from which we deduce that  $\left| \mathbb{E} \left[ \int_0^T \Psi(\bar{\eta}(s), x_{\bar{\eta}(s)}) ds \right] - \mathbb{E} \left[ \int_0^T \Psi(s, x_s) ds \right] \right| \leq CTh$ . Finally, the weak error expansion reads

$$\mathbb{E}[F(\bar{x}_T)] - \mathbb{E}[F(x_T)] = \frac{h}{2} \mathbb{E} \left[ \int_0^T \Psi(s, x_s) ds \right] + O(h^2).$$

□

## 5 Numerical experiment

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

### 5.1 The PDMP 2-dimensional Morris-Lecar

The deterministic Morris-Lecar model has been introduced in 1981 by Catherine Morris and Harold Lecar in [23] to explain the dynamics of the barnacle muscle fiber. This model belongs to the family of conductance-based models (just as the Hodgkin-Huxley model [19]) and takes the following form

$$\begin{cases} \frac{dv}{dt} = \frac{1}{C} \left( I - g_{\text{Leak}}(v - V_{\text{Leak}}) - g_{\text{Ca}} M_\infty(v)(v - V_{\text{Ca}}) - g_{\text{K}} n(v - V_{\text{K}}) \right), \\ \frac{dn}{dt} = (1 - n)\alpha_{\text{K}}(v) - n\beta_{\text{K}}(v), \end{cases} \quad (39)$$

where  $M_\infty(v) = (1 + \tanh[(v - V_1)/V_2])/2$ ,  $\alpha_K(v) = \lambda_K(v)N_\infty(v)$ ,  $\beta_K(v) = \lambda_K(v)(1 - N_\infty(v))$ ,  $N_\infty(v) = (1 + \tanh[(v - V_3)/V_4])/2$ ,  $\lambda_K(v) = \bar{\lambda}_K \cosh((v - V_3)/2V_4)$ .

In this section we consider the PDMP version of (39) that we denote by  $(x_t, t \in [0, T])$ ,  $T > 0$ , whose characteristics  $(f, \lambda, Q)$  are given by

- $f(\theta, \nu) = \frac{1}{C} \left( I - g_{\text{Leak}}(\nu - V_{\text{Leak}}) - g_{\text{Ca}} M_\infty(\nu)(\nu - V_{\text{Ca}}) - g_K \frac{\theta}{N_K} (\nu - V_K) \right),$
- $\lambda(\theta, \nu) = (N_K - \theta)\alpha_K(\nu) + \theta\beta_K(\nu),$
- $Q((\theta, \nu), \{\theta + 1\}) = \frac{(N_K - \theta)\alpha_K(\nu)}{\lambda(\theta, \nu)}, \quad Q((\theta, \nu), \{\theta - 1\}) = \frac{\theta\beta_K(\nu)}{\lambda(\theta, \nu)}.$

The state space of the model is  $E = \{0, \dots, N_K\} \times \mathbb{R}$  where  $N_K \geq 1$  stands for the number of potassium gates. The values of the parameters used in the simulations are  $V_1 = -1.2$ ,  $V_2 = 18$ ,  $V_3 = 2$ ,  $V_4 = 30$ ,  $\bar{\lambda}_K = 0.04$ ,  $C = 20$ ,  $g_{\text{Leak}} = 2$ ,  $V_{\text{Leak}} = -60$ ,  $g_{\text{Ca}} = 4.4$ ,  $V_{\text{Ca}} = 120$ ,  $g_K = 8$ ,  $V_K = -84$ ,  $I = 60$ ,  $N_K = 100$ .

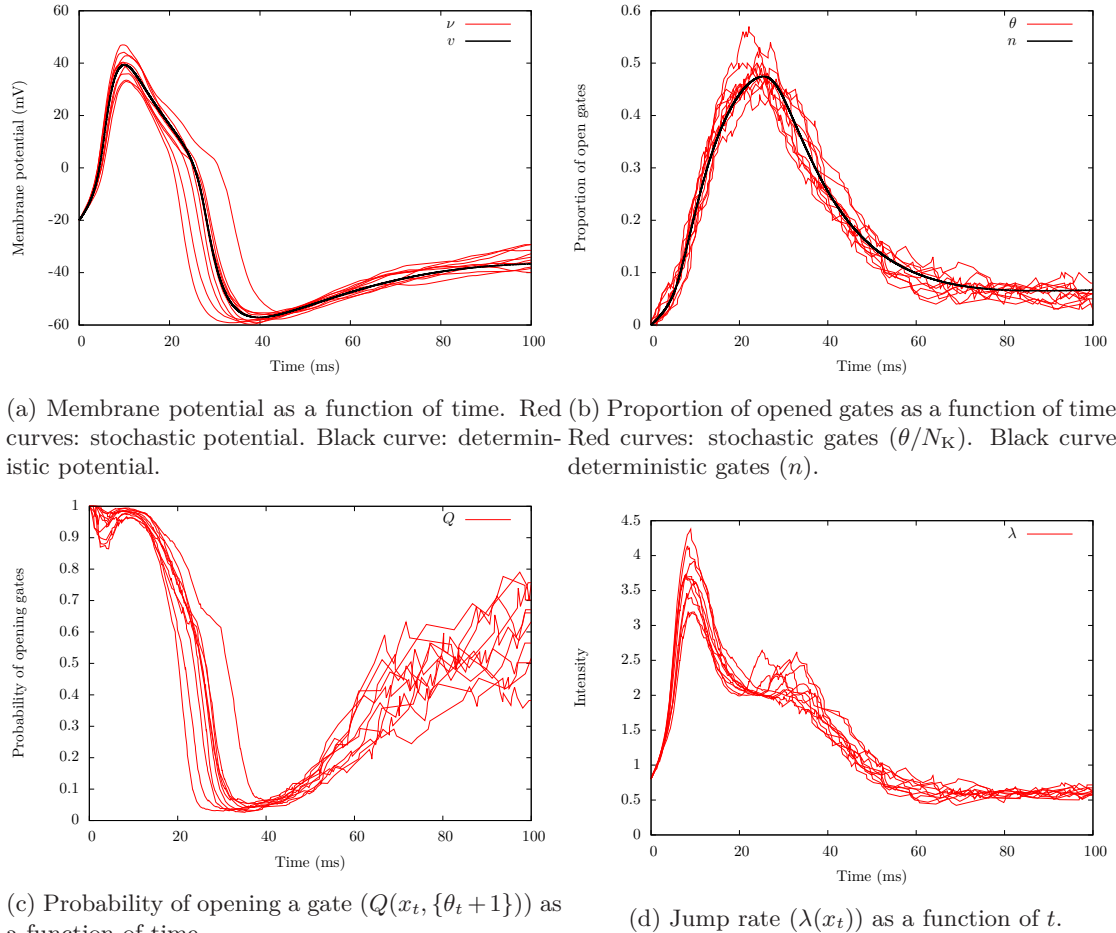


Figure 1: 10 trajectories of the characteristics of the PDMP 2d-ML on  $[0, 100]$ .



## 5.2 Classical and Multilevel Monte Carlo estimators

In this section 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  $F(\theta, \nu) = \nu$  for  $(\theta, \nu) \in E$  so that  $F(x_T)$  gives the value of the membrane potential at time  $T$ . Note that other possible choices are  $F(\theta, \nu) = \nu^n$  or  $F(\theta, \nu) = \theta^n$  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  $X := F(x_T)$ . In the sequel it will be convenient to emphasize the dependence of the Euler scheme  $(\bar{x}_t)$  on a time step  $h$ . We introduce a family of random variables  $(X_h, h > 0)$  defined by  $X_h := F(\bar{x}_T)$  where for a given  $h > 0$  the corresponding PDP  $(\bar{x}_t)$  is constructed as in section 2.3 with time step  $h$ . In particular, the processes  $(\bar{x}_t)$  for  $h > 0$  are correlated through the same randomness  $(U_k)$ ,  $(V_k)$  and  $(N_k^*)$ . We build a classical Monte Carlo estimator of  $\mathbb{E}[X]$  based on the family  $(X_h, h > 0)$  as follows

$$Y^{\text{MC}} = \frac{1}{N} \sum_{k=1}^N X_h^k, \quad (40)$$

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

$$Y^{\text{MLMC}} = \frac{1}{N_1} \sum_{k=1}^{N_1} X_{h^*}^k + \sum_{l=2}^L \frac{1}{N_l} \sum_{k=1}^{N_l} (X_{h_l}^k - X_{h_{l-1}}^k), \quad (41)$$

where  $((X_{h_l}^k, X_{h_{l-1}}^k), k \geq 1)$  for  $l = 2, \dots, L$  are independent sequences of independent copies of the couple  $(X_{h_l}, X_{h_{l-1}})$  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  $q = (q_1, \dots, q_L) \in ]0, 1[^L$  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  $\tilde{X}_h := F(\tilde{x}_T) \tilde{R}_T$  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  $\tilde{Y}^{\text{MLMC}}$  based on the family  $(\tilde{X}_h, h > 0)$ .

The complexity of the classical Monte Carlo estimator  $Y^{\text{MC}}$  depends on the parameters  $(h, N)$  and the one of the multilevel estimators  $Y^{\text{MLMC}}$  and  $\tilde{Y}^{\text{MLMC}}$  depends on  $(L, q, N)$ . In order 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 call  $V_1$ ,  $c_1$ ,  $\alpha$ ,  $\beta$  and  $\text{Var}(X)$  the structural parameters associated to 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  $\text{Var}(X)$  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  $\text{Var}(\tilde{X})$  associated to  $(\tilde{X}_h, h > 0)$  and  $\tilde{X}$  are such that  $\tilde{\alpha} = \alpha$ ,  $\tilde{c}_1 = c_1$  (see (34)),  $\tilde{\beta} = 2$  (see Theorem 3.2) and  $\tilde{V}_1$ ,  $\text{Var}(\tilde{X})$  are not explicit.

The classical and the 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 term of the corresponding structural parameters as follows (see [21] or [24]). For a user prescribed

$L$	$\left\lceil 1 + \frac{\log( c_1 ^{\frac{1}{\alpha}} h^*)}{\log(M)} + \frac{\log(A/\epsilon)}{\alpha \log(M)} \right\rceil, \quad A = \sqrt{1 + 2\alpha}$
$q$	$q_1 = \mu^*(1 + \rho(h^*)^{\frac{\beta}{2}})$ $q_j = \mu^* \rho(h^*)^{\frac{\beta}{2}} \left( \frac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1} + n_j}} \right), j = 2, \dots, L; \mu^* = 1 / \sum_{1 \leq j \leq L} q_j$
$N$	$(1 + \frac{1}{2\alpha}) \frac{\text{Var}(X) \left( 1 + \rho(h^*)^{\frac{\beta}{2}} \sum_{j=1}^L \left( \frac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1} + n_j}} \right) \sqrt{n_{j-1} + n_j} \right)^2}{\epsilon^2 \sum_{j=1}^L q_j (n_{j-1} + n_j)}$

Table 1: Optimal parameters for the MLMC estimator (41).

$\epsilon > 0$ , the classical Monte Carlo parameters  $h$  and  $N$  are

$$h(\epsilon) = (1 + 2\alpha)^{\frac{-1}{2\alpha}} \left( \frac{\epsilon}{|c_1|} \right)^{\frac{1}{\alpha}}, \quad N(\epsilon) = \left( 1 + \frac{1}{2\alpha} \right) \frac{\text{Var}(X) (1 + \rho h^{\beta/2}(\epsilon))^2}{\epsilon^2}, \quad (42)$$

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

### 5.3 Methodology

We compare the classical and the multilevel Monte Carlo estimators in term of *precision*, *CPU-time* and *complexity*. The *precision* of an estimator  $Y$  is defined by the  $L^2$ -error  $\|Y - \mathbb{E}[X]\|_2 = \sqrt{(\mathbb{E}[Y] - \mathbb{E}[X])^2 + \text{Var}(Y)}$  also known as the Root Mean Square Error (RMSE). The *CPU-time* represents the time needed to compute one realisation of an estimator. The *complexity* is defined as the number of time steps involved in the simulation of an estimator. Let  $Y$  denote the estimator (40) or (41). We estimate the bias of  $Y$  by

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

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

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

where  $v^1, \dots, v^R$  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  $\hat{\epsilon}_R$  by

$$\hat{\epsilon}_R = \sqrt{\hat{b}_R^2 + \hat{v}_R}. \quad (43)$$

The numerical computation of (43) for both estimators (40) and (41) requires the computation of the optimal parameters given by (42) and in table 1 of section 5.2 which are expressed in term of the structural parameters  $c_1$ ,  $V_1$  and  $\text{Var}(X)$ . Moreover the 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

$$\widehat{V}_1 = (1 + M^{-\beta/2})^{-2} h^{-\beta} \mathbb{E}[|X_h - X_{h/M}|^2], \quad (44)$$

and we use the following estimator of  $c_1$

$$\hat{c}_1 = (1 - M^{-\alpha})^{-1} h^{-\alpha} \mathbb{E}[X_{h/M} - X_h]. \quad (45)$$

The estimator of  $c_1$  is obtained writing the weak error expansion for the two time steps  $h$  and  $h/M$ , summing and neglecting the  $O(h^2)$  term. In (44) we use  $(h, M) = (0.1, 4)$  and in (45), 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 that we do not need a precise estimation here.

## 5.4 Numerical results

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

**Case 1:**  $\tilde{\lambda}(\theta) = 1$  and  $\tilde{Q}(\theta, \{\theta + 1\}) = \frac{N_K - \theta}{N_K}$ ,  $\tilde{Q}(\theta, \{\theta - 1\}) = \frac{\theta}{N_K}$ .

**Case 2:**  $\tilde{\lambda}(x, t) = \lambda(\theta, v(t))$  and  $\tilde{Q}((x, t), dy) = Q((\theta, v(t)), dy)$  where  $v$  denotes the first component of the solution of (39).

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  $(\Phi, \lambda, Q)$  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 times as the discrete component of  $(x_t)$  do 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  $\mathbb{E}[F(x_T) - F(\tilde{x}_T)]$ .

The case 3 implies to use the following MLMC estimator which is slightly different from (41).

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

where  $\left((X_{h_l}^k, \tilde{X}_{h_{l-1}}^k), k \geq 1\right)$  for  $l = 2, \dots, L$  are independent sequences of independent copies of the couple  $(X_{h_l}, \tilde{X}_{h_{l-1}}) = (F(\bar{x}_T), F(\bar{y}_T)\tilde{R}_T)$  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  $h_l$  do, 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).

The figure 2 confirms numerically that  $\mathbb{E}[|X_{h_l} - X_{h_{l-1}}|^2] = O(h_l)$  and that  $\mathbb{E}[|\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}}|^2] = O(h_l^2)$  for the cases 1,2 and 3 (see Theorems 3.1 and 3.2 respectively). Indeed, for  $T = 10$  (see figure 2a), 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 the cases 1,2 and 3 are parallel to a line of slope -2. We also see that the curves corresponding to the 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 the case 1 is always above all the other ones, this indicates that the  $L^2$ -error (or the variance) in the case 1 is too big (w.r.t the others) and that is why we do not consider this case in the sequel. As  $T$  increases (see figures 2b and 2c), the theoretical order of the numerical schemes is still observed. However, for  $T = 20$ , a slight difference begin to emerge between the cases 2 and 3 (the case 3 being better) and this difference is accentuated for  $T = 30$  so that we do not represent the case 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  $\mathbb{E}[X] = -31.4723$  and  $\text{Var}(X) = 335$ . Note that  $v(30) = -35.3083$  where  $v$  is the deterministic membrane potential solution of (39) so that 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 (43).

$k$	$\epsilon = 2^{-k}$	$\hat{\epsilon}_{100}$	$\hat{b}_{100}$	$\hat{v}_{100}$	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 2: Results and parameters of the Monte Carlo estimator  $Y^{\text{MC}}$ . Estimated values of the structural parameters:  $c_1 = 4.58$ ,  $V_1 = 7.25$ .

$k$	$\epsilon = 2^{-k}$	$\hat{\epsilon}_{100}$	$\hat{b}_{100}$	$\hat{v}_{100}$	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 3: Results and parameters of the Multilevel Monte Carlo estimator  $Y^{\text{MLMC}}$ . Estimated values of the structural parameters:  $c_1 = 4.58$ ,  $V_1 = 7.25$ .

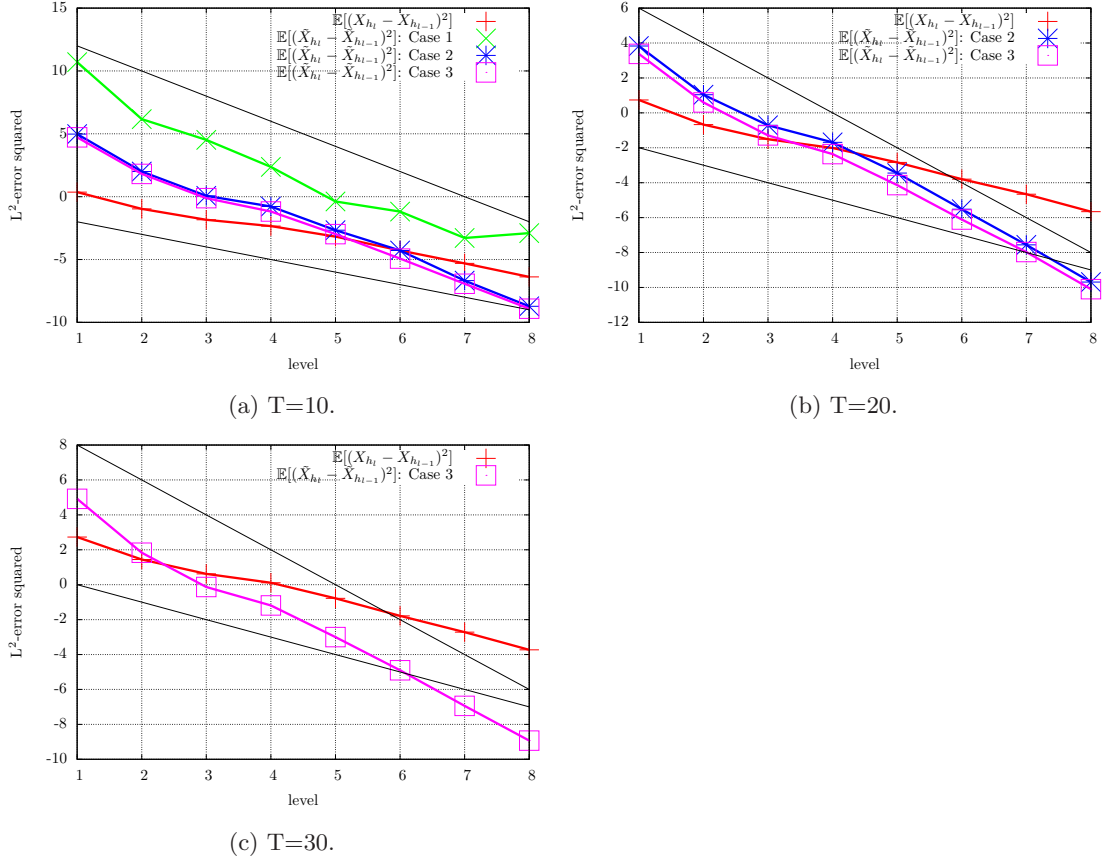


Figure 2: The plots (a),(b) and (c) show the decay of  $\mathbb{E}[(X_{h_l} - X_{h_{l-1}})^2]$  and  $\mathbb{E}[\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}}]^2$  (y-axis,  $\log_M$  scale) as a function of  $l$  with  $h_l = h \times M^{-(l-1)}$ ,  $h = 1$ ,  $M = 4$ , for different values of the final time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

The results of the Monte Carlo simulations are shown in tables 2 for the classical Monte Carlo estimator  $Y^{\text{MC}}$  and in tables 3 and 4 for the multilevel estimators  $Y^{\text{MLMC}}$  and  $\tilde{Y}^{\text{MLMC}}$  (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  $Y^{\text{MLMC}}$  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 levels and the sample size is  $N = 2600$ . For such parameters, the numerical complexity of the estimator is  $\text{Cost}(Y^{\text{MLMC}}) = 28200$ , the empirical RMSE  $\hat{\epsilon}_{100} = 0.389$  and the computational time of one realisation of  $Y^{\text{MLMC}}$  is 0.362 seconds. We also reported the empirical bias  $\hat{b}_{100}$  and the empirical variance  $\hat{v}_{100}$  in view of (43).

The results indicate that the MLMC outperforms the classical MC. More precisely, for small values of  $\epsilon$  (i.e  $k = 1, 2, 3$ ) the complexity and the CPU-time of the classical and the multilevel MC estimators are of the same order. As  $\epsilon$  decreases (i.e as  $k$  increases) the difference in complexity and CPU-time between classical and multilevel MC increases. Indeed, for  $k = 5$  the complexity of the estimator  $Y^{\text{MC}}$  is approximately 13 times superior to the one of  $Y^{\text{MLMC}}$  and 19 times superior to the one of  $\tilde{Y}^{\text{MLMC}}$ . The same fact appears when we look at the complexity ratio of the estimators  $Y^{\text{MLMC}}$  and  $\tilde{Y}^{\text{MLMC}}$  (i.e  $\text{Cost}(Y^{\text{MLMC}})/\text{Cost}(\tilde{Y}^{\text{MLMC}})$ ) as  $\epsilon$  decreases.

$k$	$\epsilon = 2^{-k}$	$\hat{\epsilon}_{100}$	$\hat{b}_{100}$	$\hat{v}_{100}$	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

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

However, the difference between the complexity of these two MLMC estimators increases more slowly than the one between a MC and a MLMC estimator. Recall that the computational benefit of the MLMC over the MC grows as the prescribed  $\epsilon$  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 the readability, figures 3a, 3b show the ratios of the complexities and the CPU-times of the three estimators  $Y^{\text{MC}}$ ,  $Y^{\text{MLMC}}$  and  $\tilde{Y}^{\text{MLMC}}$  as a function of  $\epsilon$ .

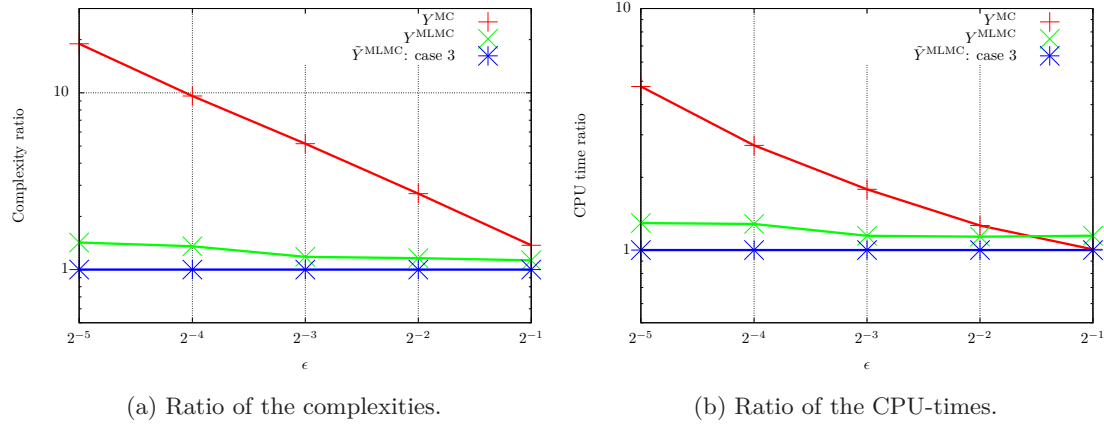


Figure 3: The plots (a) and (b) show the complexity and CPU-time ratios w.r.t the complexity and CPU-time of the estimator  $\tilde{Y}^{\text{MLMC}}$  as a function of the prescribed  $\epsilon$  ( $\log_2$  scale for the  $x$ -axis, log scale for the  $y$ -axis).

## References

- [1] D.F. Anderson and D.J. Higham. Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics. *Multiscale Model. Simul.*, 10(1):146–179, 2012.
- [2] D.F. Anderson, D.J. Higham, and Y. Sun. Complexity of Multilevel Monte Carlo tau-leaping. *SIAM Journal on Numerical Analysis*, 52(6):3106–3127, 2014.
- [3] M. Benaïm, S. Le Borgne, F. Malrieu, and P-A. Zitt. Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.*, 17:14 pp., 2012.

- [4] P. Brémaud. *Point Processes and Queues, Martingale Dynamics*. Springer-Verlag New York Inc, 1981.
- [5] M.H.A. Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal statistical Society*, 46:353–388, 1984.
- [6] M.H.A. Davis. *Markov Models and Optimization*. Chapman and Hall, London, 1993.
- [7] S. Dereich. Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *The Journal of Applied Probability*, 21(1):283–311, 2011.
- [8] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo algorithm for Lévy-driven Stochastic Differential Equations. *Stochastic Processes and their Applications*, 121(7):1565–1587, 2011.
- [9] L. Devroye. *Non-uniform random variate generation*. Springer-Verlag, New York Inc., 1986.
- [10] A. Ferreiro-Castilla, A.E. Kyprianou, R. Scheichl, and G. Suryanarayana. Multilevel Monte Carlo simulation for Lévy processes based on the Wiener–Hopf factorisation. *Stochastic Processes and their Applications*, 124(2):985 – 1010, 2014.
- [11] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.
- [12] M.B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.
- [13] D. Giorgi. *Théorèmes limites pour estimateurs Multilevel avec et sans poids. Comparaisons et applications*. PhD thesis, Université Pierre et Marie Curie - Paris 6, 2017.
- [14] P.W. Glynn and C-H. Rhee. Exact estimation for markov chain equilibrium expectations. *Journal of Applied Probability*, 51:377–389, 2014.
- [15] P.W. Glynn and C-H. Rhee. Unbiased Estimation with Square Root Convergence for SDE models. *Operations Research*, 63(5):1026–1043, 2015.
- [16] C. Graham and D. Talay. *Stochastic simulation and Monte Carlo*. Springer, 2013.
- [17] E. Hairer, S.P. Norsett, and G. Wanner. *Solving Ordinary Differential Equations I*. Second Revised Edition, Springer, 2008.
- [18] S. Heinrich. Multilevel Monte Carlo methods. *Large-scale scientific computing*, pages 58–67, 2001.
- [19] A.L. Hodgkin and A.F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544, 1952.
- [20] M. Jacobsen. *Point Process Theory and Applications, Marked Point and Piecewise Deterministic Processes*. Birkhäuser Boston, 2006.
- [21] V. Lemaire and G. Pagès. Multilevel Richardson-Romberg extrapolation. *Bernoulli*, 23(4A):2643–2692, 2017.
- [22] V. Lemaire, M. Thieullen, and N. Thomas. Exact simulation of the jump times of a class of Piecewise Deterministic Markov Processes. *Journal of Scientific Computing*, 75:1776 – 1807, 2018.

- [23] C. Morris and H. Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*, 35:193–213, 1981.
- [24] G. Pagès. *Numerical Probability: An Introduction with Applications to Finance*. Universitext, Springer, Cham, 2018.
- [25] K. Pakdaman, M. Thieullen, and G. Wainrib. Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Advances in Applied Probability*, 42(3):761–794, 2010.
- [26] Z. Palmowski and T. Rolski. A technique for exponential change of measure for Markov processes. *Bernoulli*, 8(6):767–785, 12 2002.
- [27] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509, 1990.
- [28] Y. Xia and M.B. Giles. Multilevel Path Simulation for Jump-Diffusion SDEs. In Leszek Plaskota and Henryk Woźniakowski, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2010*, pages 695–708, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.