Stochastic numerical methods for Piecewise Deterministic Markov Processes. Applications in Neuroscience

Nicolas Thomas

École doctorale de sciences mathématiques de Paris centre (ED 386)

Thèse de doctorat de Sorbonne Université

Discipline : Mathématiques Appliquées

présentée par

Nicolas Thomas

Stochastic numerical methods for Piecewise Deterministic Markov Processes. Applications in Neuroscience.

Soutenue le 20 juin 2019 devant le jury composé de :

M^{me} Evelyn Buckwar	Johannes Kepler University, Linz	Rapporteur
M. Fabien Campillo	INRIA, Montpellier	Rapporteur
M. Benjamin Jourdain	Ecole Nationale des Ponts et Chaussées	Examinateur
M. Vincent Lemaire	Sorbonne Université	Co-directeur
M. Gilles Pagès	Sorbonne Université	Examinateur
M^{me} Michèle Thieullen	Sorbonne Université	Co-directeur

Sorbonne Université - Laboratoire de Probabilités, Statistiques et Modélisation (LPSM). UMR 8001.
Case courrier 158
4 place Jussieu
75252 Paris Cedex 05

Sorbonne Université - Campus Pierre et Marie Curie.
École doctorale de sciences mathématiques de Paris centre.
Boîte courrier 290
4 place Jussieu
75252 Paris Cedex 05

Remerciements

Tout d'abord, je tiens à remercier mes directeurs de thèse Michèle Thieullen et Vincent Lemaire qui m'ont accompagné durant cette thèse avec beaucoup de patience, de sérénité, de disponibilité et qui ont su me soutenir dans les moments difficiles. Grâce à leur rigueur j'ai énormément appris, tant sur le plan théorique que numérique.

Je remercie aussi les thésards du LPSM, passés et présents, dont j'ai eu la chance de croiser la route. Merci à l'équipe administrative et informatique pour leur aide précieuse.

Merci à Evelyn Buckwar et à Fabien Campillo pour avoir raporté ma thèse d'une manière très constructive ainsi qu'à Gilles Pagès et à Benjamin Jourdain pour avoir accepté de faire parti du jury de soutenance.

Je souhaite également remercier mes parents, mes beaux-parents, mon frère, ma soeur, Sabine, Marcel, Arnaud et Louise pour leur présence et leur réconfort. Merci à mes poteaux de longue date, Clément, Thibault, Hugo, Adel, Benoit, Giovanna, Rémi, Fred, Nour, Cambyse pour toute la joie et la bonne humeur que vous m'apportez ainsi que pour tous les moments inoubliables passés ensemble.

Enfin, merci à toi Stacy pour ta patience admirable, ton amour inconditionnel et pour m'avoir encouragé et soutenu quotidiennement du début à la fin.

Abstract

In this thesis, motivated by applications in Neuroscience, we study efficient Monte Carlo (MC) and Multilevel Monte Carlo (MLMC) methods based on the thinning for piecewise deterministic (Markov) processes (PDMP or PDP) that we apply to stochastic conductance-based models. On the one hand, when the deterministic motion of the PDMP is explicitly known we end up with an exact simulation. On the other hand, when the deterministic motion is not explicit, we establish strong estimates and a weak error expansion for the numerical scheme that we introduce. The thinning method is fundamental in this thesis. Beside the fact that it is intuitive, we use it both numerically (to simulate trajectories of PDMP/PDP) and theoretically (to construct the jump times and establish error estimates for PDMP/PDP).

In a first part we consider the class of PDMP with explicit flow. In this case using the thinning we are able to simulate exactly the jump times and we obtain an exact simulation of the trajectories. We study both theoretically and numerically the efficiency of such thinning algorithms for different types of bounds, from the classical global bound to a path-adapted one.

In a second part, for PDMPs with no explicit flow, we efficiently approximate an expectation involving the process at a terminal time. We first discretise the flow with the classical deterministic Euler scheme and we apply the results of the first part to the discretisation of the PDMP (which turns out to be a PDP). We are able to prove strong error estimates for PDP and a weak error expansion for PDMP. We then use these error estimates to investigate the efficiency of the MLMC method

Stochastic conductance-based models are composed by two distinct variables, the first one is continuous and models the membrane potential whereas the second one is discrete and models the configuration of the ionic channels. These models naturally lead to the class of PDMPs. We are interested in applications to stochastic counterparts of Hodgkin-Huxley and Morris-Lecar models which are known as biologically realistic since they take into account the precise description of the ionic channels as opposed to the FitzHugh-Nagumo model for example. The flow is explicit for PDMP-type Hodgkin-Huxley and 3-dimensional Morris-Lecar models. However it is not explicit for the classically used 2-dimensional Morris-Lecar PDMP. We apply the theoretical part to simulate exactly and efficiently the first spike latency. Moreover, on these models our MLMC estimator does indeed outperform the classical MC one.

Keywords: Piecewise deterministic (Markov) process, Thinning, Exact simulation, Strong error estimate, Weak error expansion, Multilevel Monte Carlo, Hodgkin-Huxley, Morris-Lecar.

Contents

INTRODUCTION

The purpose of this thesis is to study probabilistic numerical methods for Piecewise Deterministic (Markov) Processes (PDMP or PDP) with a view towards applications in Neuroscience.

In the first section of the present introduction we position our work with respect to the literature. In section 2 we present the conductance-based models that will be used for applications purpose. In section 3 we introduce the general class of PDMP/PDP we consider. In sections 4 and 5 we review the general ideas behind thinning and Monte Carlo simulations (classical and multilevel). Section 6 describes, chapter by chapter, the main mathematical results obtained in this thesis. Some perspectives related to our work are discussed in Section 7.

1 Positioning of the thesis

In this thesis, motivated by applications in Neuroscience, we study efficient Monte Carlo (MC) and Multilevel Monte Carlo (MLMC) methods based on the thinning for piecewise deterministic (Markov) processes (PDMP or PDP) that we apply to stochastic conductance-based models. On the one hand, when the deterministic motion of the PDMP is explicitly known we end up with an exact simulation. On the other hand, when the deterministic motion is not explicit, we establish strong estimates and a weak error expansion for the numerical scheme that we introduce. The thinning method is fundamental in this thesis. Beside the fact that it is intuitive, we use it both numerically (to simulate trajectories of PDMP/PDP) and theoretically (to construct the jump times and establish error estimates for PDMP/PDP).

Stochastic conductance-based models are composed by two distinct variables, the first one is continuous and models the membrane potential whereas the second one is discrete and models the configuration of the ionic channels. Those variables are fully coupled in the sense that the evolution of the membrane potential is influenced by the proportion of ionic channels in a given configuration and the opening/closing dynamic of the channels depends continuously on the membrane potential. As a consequence, these models naturally lead to the class of PDMPs [67]. We are interested in applications to stochastic counterparts of Hodgkin-Huxley [47] and Morris-Lecar [63] models which are known as biologically realistic since they take into account the precise description of the ionic channels as opposed to the FitzHugh-Nagumo model [31] for example.

In a first part we consider the class of PDMP with explicit flow. In this case using the thinning we are able to simulate exactly the jump times and we obtain an exact simulation of the trajectories. We study both theoretically and numerically the efficiency of such thinning algorithms for different types of bounds from the classical global bound to a path-adapted one. We did not find such systematic study of these different bounds in the literature.

Suppose that we want to approximate an expectation $\mathbb{E}[g(x_T)]$ involving the process at a terminal time. Classical MC would be satisfactory when we have explicit flow.

However, when the flow is not explicit we have first to discretise it, for instance with a classical Euler scheme, in order to get back to an explicit flow. On the discretised flow we can use classical MC, but this would result in a high complexity. This motivated us to consider MLMC for PDMP which to the best of our knowledge have not been considered in the literature.

We apply the first part to the discretisation of the PDMP (which turns out to be a PDP) for which we obtain by thinning an exact simulation. Based on this construction, we are able to prove strong error estimates for PDP and a weak error expansion for PDMP. We then use these error estimates to investigate the efficiency of the MLMC method to approximate expectations of functions of the state of a PDMP at fixed time.

Regarding stochastic conductance-based models, the flow is explicit for PDMP-type Hodgkin-Huxley and 3-dimensional Morris-Lecar models. However it is not explicit for the classically used 2-dimensional Morris-Lecar PDMP. For example, our work can be applied to simulate exactly and efficiently quantities of interest such as first spike latency or inter-spike intervals in a PDMP-type Hodgkin-Huxley model and to efficiently approximate the moments of the proportion of open channels or the membrane potential at fixed time in a 2-dimensional Morris-Lecar PDMP in order to compute statistics on these biological variables.

In the literature, numerical schemes for PDMP/PDP have been the subject of different papers. In [7] and [8], the authors introduce numerical methods to compute expectations of functionals of a PDMP and optimal stopping times. Their approaches are based on the quantization of the underlying discrete-time Markov chain. In [71] and [2], the authors show that a PDMP with a specific jump distribution can be represented as the solution of a stochastic differential equation (SDE) where the noise comes from counting processes. Consequently, they build fixed time step numerical schemes where they simulate the number of jumps within each time step rather than the jump times explicitly. The numerical schemes introduced in [27] and [70] explicitly simulate the jump times and are both based on the numerical inversion of a survival function. In [27], the authors approximate the log-survival function (i.e the integrated jump rate) of the jump times using a numerical scheme together with a linear interpolation. By doing this, they approximate the distribution of the jump times with a piecewise exponential distribution. In [70], the author reformulates the problem of the inversion of the survival function of each jump time as a hitting time problem for a system of ordinary differential equations (ODE) with random threshold. The system of ODEs is non-linear, includes an equation on the jump rate along the flow of the PDMP and is different for each jump time. A numerical scheme which is related to [70] can be found in [77] where the author uses a change of time in the previous system of ODEs in order to obviate the hitting time problem. None of the numerical schemes discussed above uses the thinning and none of them produces an exact simulation even if the flow of the PDMP is explicit.

The author in [77] compares his ODE-based algorithm with a fictitious jump method proposed in [42]. This is a thinning algorithm which uses a constant global bound for the intensity of the PDMP and is exact when the flow is explicit. Another thinning algorithm using a constant global bound can be found in [12] in order to simulate a model

of chemostat.

Regarding the schemes specifically devoted to stochastic conductance-based models, several algorithms have been developed. The algorithms in [75], [15], [74], [14] consist in following the evolution of each channel separately (therefore the efficiency is low) and do not use thinning. Papers [14], [60], [72] simulate the number of channels in each possible states without considering each channel individually in the spirit of Gillespie algorithm [35]. Gillespie algorithm is used to simulate continuous time homogeneous Markov chains and has been popularized by Gillespie in order to simulate the stochastic time evolution of a system of chemical reactions. We emphasize that this algorithm is known under different names in the literature such as kinetic Monte Carlo, stochastic simulation algorithm or n-fold way algorithm. Generalisations of this algorithm have been considered for example by Gillespie himself to take into account the case of semi-Markov processes [36]. However Gillespie algorithm does not use thinning and is based on the inversion of a survival function. To be complete, let us mention that many papers aim to speed up the simulation using diffusion approximations of the Markov dynamic of the ionic channels [32], [58], [41], [65], [20], [48], [69]. Under this common theoretical approach, each implementation differs in how it handles various numerical difficulties such as bounding of channels proportion to $[0, 1]$. We precise that in this thesis we do not work in this direction.

To summarize, the thinning has been used in few papers dealing with numerical schemes for PDMP and only with a global constant bound for the intensity. No study of the efficiency of thinning algorithms has been conducted. Moreover, the MLMC method for PDMP has not been considered.

The MLMC method is a general method which allows to approximate efficiently the expectation of a random variable X. MLMC relies on the existence of a numerical scheme $(X_h, h > 0)$ which converges strongly (in the sense of squared L^2 norm) and weakly to X as h goes to 0. The main difficulty to efficiently use the MLMC method is to build a numerical scheme, well correlated to X, for which we have strong and weak estimates. Indeed, the orders of convergence play a crucial role in the complexity of a MLMC estimator.

The MLMC method has been popularized by Giles [34] in the case of SDEs with a view towards financial applications. Giles considers the fixed time step Euler-Maruyama scheme which is correlated to the underlying SDE by using the same brownian motion for both processes. Such scheme is known to verify a $O(h)$ strong and weak convergence (see [53], [76]).

For a jump-diffusion when the intensity is state-dependent the authors in [82], motivated by financial applications, use the jump-adapted Milstein scheme to build an approximation which they correlate to the original jump-diffusion by thinning. The weak convergence of such a Milstein scheme for jump-diffusions has been investigated first in [59] in which an $O(h^2)$ convergence was proved under strong regularity assumptions on the coefficients of the jump-diffusion and on the function g (remember that we want to approximate $\mathbb{E}[g(x_T)]$). When the jump coefficient is not regular (for instance not continuous) $O(h^2)$ is still valid under stronger assumptions on g as was proved in [38].

More recently, [81] obtains an $O(h)$ weak convergence for a Lipschitz jump coefficient and a class of g larger than those considered in [59] and [38]. Note that in [59] and [38] the numerical scheme is based on thinning. We emphasize that the discretisation grid in [59], [38] and [82] is constructed a priori and is composed by a fixed grid to which they add the jump times of the Poisson process introduced for the thinning. As a consequence, they simulate explicitly the jump times of the discretised process. Regarding the strong convergence, it is known that the Milstein scheme for diffusions provides an $O(h^2)$ convergence. However, because of the jumps and the fact that some proposed jump times may be accepted for one process but not for the other, or vice versa, the authors in [82] obtain only an $O(h)$ strong convergence. In order to improve this order, they introduce a change of probability under which the original process and its discretisation have the same probability of accepting a proposed jump time. Under this probability change they obtain an $O(h^2)$ strong convergence.

MLMC has been investigate for continuous time homogeneous Markov chains in [3] in the context of chemical reactions. The authors represent such processes as the solution of a random time change equation (see [29]) which is similar to a classical SDE without drift where the noise arises from a Poisson random measure. Consequently, they use the fixed time step Euler-Maruyama scheme where they simulate the number of jumps within each time step rather than the jump times explicitly. They correlate the original process and its approximation using the additivity property of independent Poisson variables and they prove $O(h)$ strong and weak convergence. In their paper they do not use thinning.

Our approach differs from the previous ones in the way we construct an approximation of our original PDMP. Since one of the characteristics of a PDMP is a family of vector fields indexed by its discrete component, we discretise a priori the flow corresponding to each vector field using a deterministic Euler scheme. Then, we build our scheme step by step inspired by the original iterative construction of PDMP/PDP (see [21]). In the first step, using the discretised initial flow and thinning, we simulate explicitly the first jump time and the new position. Then we start anew from this first jump (which is simulated exactly from the results in the first part of the thesis) and we iterate this procedure. This implies that our discretisation grid is constructed iteratively and starts anew from each jump times and so it differs from the ones in [34], [82] and [3]. Moreover we use the thinning also in the theoretical study of the error estimates. We first show that there is an $O(h)$ strong and weak convergence. Then, inspired by [82], we also introduce a new probability under which we obtain an $O(h^2)$ strong convergence. This is an important step in order to lower the complexity of the MLMC estimator as we detail in section 5. When implemented on a PDMP-type Morris-Lecar model, our MLMC estimator does indeed outperform the classical MC one.

2 Conductance-based models

Most neurons respond to incoming signals with action potentials (also called spikes) which are the building block of the neural coding. Action potentials are generated in the soma, propagate along the axon and produce inputs to the other neurons through dendrites.

Based on this biological phenomenon, two key Neuroscience fields have emerged in order to understand how neurons communicate: the neural encoding and the neural decoding. The neural encoding concerns the study of the response of a neuron to a given input (or stimulus) while the neural decoding is the opposite, it concerns the problem of recovering the stimulus observing the response (see [23] for more details). The response of a neuron usually refers to quantities of interest related to the spike train (sequence of spike times) it produces such as the first spike latency, the distribution of the inter-spike interval or the spike rate.

Neuron's membrane separates the intracellular environment from the extracellular one and allows exchanges of material and energy between these two environments. These exchanges are allowed by the opening and closing of ionic channels located on the membrane. A ionic channel is constituted by four gates which can be of different types (activation and inactivation) and is specific to one type of ions, for example, a sodium channel allows sodium ions only to pass the membrane. We say that a channel is active when all its gates are open. In most neurons, the intracellular environment contains a large proportion of potassium ions, whereas the extracellular environment contains a majority of sodium ones. A.L Hodgkin and A. Huxley discovered that the generation of action potentials principally relies on the movement of these two kind of ions across the membrane.

A stimulus (it can be an input from other neurons or external applied current) makes the sodium channels active, thus, sodium ions enter in the intracellular environment. It implies an increase of the membrane potential (voltage) above its resting value : the membrane is depolarized. The sodium channels open very fast leading to a fast increase of the membrane potential. When the membrane potential exceeds a certain threshold value, we say that the neuron emits an action potential or a spike (we also say the neuron discharges or fires). After being active, sodium channels become inactive, while potassium channels open at a much slower time scale. Potassium ions leave the intracellular environment to compensate the entry of sodium ions and the membrane potential goes back to its resting value : the membrane is re-polarized. Potassium channels stay active longer than sodium ones, thus, the membrane potential continues to decrease and goes below its resting value : the membrane is hyper-polarized. Finally, a protein makes the potassium ions go back into the intracellular environment and expels sodium ions in the extracellular one. The membrane potential then goes back to its resting value until the next action potential. These are the principal steps of the generation of an action potential.

The original deterministic Hodgkin-Huxley model and its stochastic counterpart

In 1952, Hodgkin and Huxley provided two models to describe the ionic mechanism underlying the generation and the propagation of action potentials in the squid giant axon [47]. We only consider models of generation of action potentials. Biologically, it

means that we clamp (isolate) a piece of the axon or the soma of the neuron and we study electrical properties in time in this clamped area (also called membrane-patch). The original four-dimensional Hodgkin-Huxley (HH) model is the following set of nonlinear differential equations.

$$\begin{cases} C\frac{dv}{dt} = I(t) - F(v,m,h,n), \\ \frac{dm}{dt} = (1-m)\alpha_m(v) - m\beta_m(v), \\ \frac{dh}{dt} = (1-h)\alpha_h(v) - h\beta_h(v), \\ \frac{dn}{dt} = (1-n)\alpha_n(v) - n\beta_n(v). \end{cases} \tag{1}$$

In this model, a channel is modelled by the gates that compose it. The $\mathbb{R}$ valued function v represents the membrane potential (voltage). The $[0,1]$ valued functions m, h, n correspond to the probability of a gate of type m, h (for the sodium) or n (for the potassium) to be open. The voltage-dependent functions α_z and β_z for $z = m, h, n$ are opening and closing rates of gates of type z respectively. I is a time-dependent function which represents the input current, C is the membrane capacity. The function F is given by $F(v,m,h,n) := I_{\mathrm{L}}(v) + I_{\mathrm{Na}}(v,m,h) + I_{\mathrm{K}}(v,n)$ where, for $z \in \{\mathrm{Na}, \mathrm{K}, \mathrm{L}\}$, $I_z = g_z(v - v_z)$ represents the ionic currents where $g_{\mathrm{Na}} = \overline{g}_{\mathrm{Na}} m^3 h$, $g_{\mathrm{K}} = \overline{g}_{\mathrm{K}} n^4$ and $g_{\mathrm{L}} = \overline{g}_{\mathrm{L}}$ are the conductances of the sodium, potassium and leak respectively. The constants $\overline{g}_{\mathrm{L}}$, $\overline{g}_{\mathrm{Na}}$, $\overline{g}_{\mathrm{K}}$ are the conductances when all the gates are open and v_{L}, v_{Na}, v_{K} are the resting potentials.

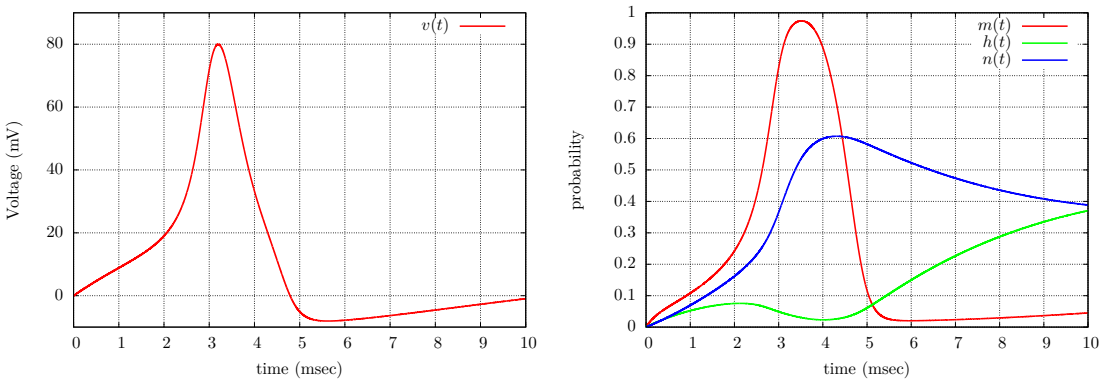


Figure 1 – Simulated trajectory of the membrane potential v (left) and of the corresponding gates m, h and n (right) in the deterministic Hodgkin-Huxley model (1) with a constant applied current $I(t) = 10$ and initial condition $(v,m,h,n) = (0,0,0,0)$.

This deterministic model successfully reproduces some of the main features of neural response such as the shape, amplitude and threshold of the action potential, the refractory period. However, the channels are considered to be in a very large number, it then fails to explain a fundamental experimental observation. When submitted to a repeated given stimulation (input current), the response of a single neuron is never exactly the same [78]. This observation suggests that there exists a stochastic component in the biological

mechanisms that generate action potentials. One explanation for that randomness is the fact that the opening and closing of ionic channels are subject to thermal noise, and are thus stochastic mechanisms [14], [80].

To obtain a stochastic version of the deterministic HH model (1), consider that the number of ionic channels in the neuron is small enough for the thermal noise to have an impact on the evolution of the membrane potential. Ionic channels are thus represented by finite-state pure jump processes with transitions depending on the membrane potential. Between jumps of these processes, the membrane potential evolves according to a deterministic dynamic which is influenced by the proportion of ionic channels in a given configuration. Such a model belongs to the class of Piecewise Deterministic Markov Processes (PDMP), see [21], [22], [51] (see also the section 3 of the present introduction). In this thesis, the stochastic version of the deterministic HH model (1) is called the subunit model. More precisely, the subunit model $(\nu, \theta^{(m)}, \theta^{(h)}, \theta^{(n)})$ can be described as follows:

- For each type of gates $z = m, h, n$, we consider that single gates $i \in \{1, ..., N_z\}$, where N_z is the number of gates of type z, are modelled by independent jump processes $u_i^{(z)}(t)$ with voltage-dependent transition rates α_z and β_z

$$0 \underset{\beta_z(.)}{\overset{\alpha_z(.)}{\rightleftarrows}} 1. \tag{2}$$

  The state 0 indicates that the gate is close and the state 1 that the gate is open.

- The proportion (empirical measure) of open gates of type $z = m, h, n$ is then defined as:

$$\theta_t^{(z)} = \frac{1}{N_z} \sum_{i=1}^{N_z} u_i^{(z)}(t).$$

- Between the jumps of $\theta = (\theta^{(m)}, \theta^{(h)}, \theta^{(n)})$, the dynamic of the membrane potential is given by the following ordinary differential equation (ODE):

$$C\frac{d\nu}{dt} = I(t) - F(\nu, \theta), \tag{3}$$

  where the function F is as in the model (1). We emphasize that for fixed θ, the above ODE is linear so that an explicit flow exits.

By noting $\theta = (\theta^{(m)}, \theta^{(h)}, \theta^{(n)})$, the jump rate of this model is then given by

$$\begin{aligned}\lambda(\theta, \nu) =& N_m\Big(\alpha_m(\nu)(1 - \theta^{(m)}) + \beta_m(\nu)\theta^{(m)}\Big) + N_h\Big(\alpha_h(\nu)(1 - \theta^{(h)}) + \beta_h(\nu)\theta^{(h)}\Big) + \\ & N_n\Big(\alpha_n(\nu)(1 - \theta^{(n)}) + \beta_n(\nu)\theta^{(n)}\Big)\end{aligned}$$

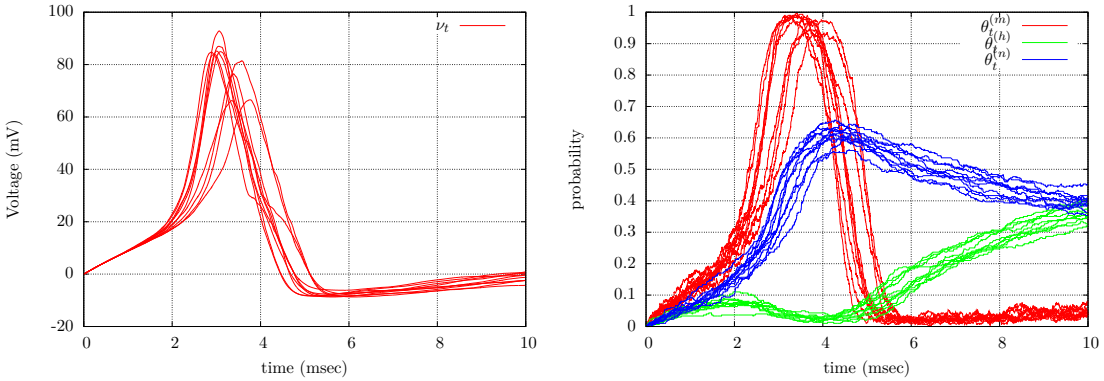


Figure 2 – 10 simulated trajectories of the membrane potential ν (left) and of the corresponding proportion of gates $\theta^{(m)}$,$\theta^{(h)}$ and $\theta^{(n)}$ (right) in the subunit model with a constant applied current $I(t) = 10$, initial condition $(\nu, \theta^{(m)}, \theta^{(h)}, \theta^{(n)}) = (0, 0, 0, 0)$ and $N_m = N_h = N_n = 300$.

so that if at time t_0 the model is in state (ν_0, θ_0) the survival function of the next transition time, say T, is given by $t \mapsto e^{-\int_{t_0}^{t} \lambda(\theta_0, \phi_{\theta_0}(s,\nu_0))ds}$ where $\phi_{\theta_0}(., \nu_0)$ denotes the solution of (3) with initial condition (ν_0, θ_0). Between times t_0 and T, the membrane potential evolves deterministically according to $\phi_{\theta_0}(., \nu_0)$. At time T a gate opens or closes according to a transition measure, say Q, which is proportional to the corresponding transition rate. For example, if we write $\lambda(\theta, \nu) = \sum_{i=1}^{6} \lambda_i(\theta, \nu)$ where $\lambda_1(\theta, \nu) = N_m \alpha_m(\nu)(1 - \theta^{(m)})$ denotes the rate for the opening of a gate of type m, λ_2 denotes the rate for the closing of a gate of type m and so on, then the probability that a gate of type m opens at time T is given by

$$Q\left((\theta_0, \phi_{\theta_0}(T - t_0, \nu_0)), \{\theta_0 + (1/N_m, 0, 0)\}\right) = \frac{\lambda_1(\theta_0, \phi_{\theta_0}(T - t_0, \nu_0))}{\lambda(\theta_0, \phi_{\theta_0}(T - t_0, \nu_0))}. \tag{4}$$

The electrical circuit introduced by Hodgkin and Huxley in [47] to model the membrane potential dynamic in the squid giant axon has become the basic formalism to represent most of the additional conductances encountered in neuron modelling. Among them, we can quote the transient potassium current [17] and the low-threshold calcium current [79], [43], [49]. This formalism is also used to model the membrane potential dynamic of different excitable cells such as cardiac cells [61] or muscle fibers [63]. Models that treat these aspects of ionic conductances, known as conductance-based models, can reproduce the rich and complex dynamics of real excitable cells quite accurately. For more details we refer to [46], [23], [50]. Moreover, the stochastic counterpart of these conductance-based models are obtained as for the HH model (1) by replacing the gain-loss equations describing the channels by jump processes. It is known that such stochastic models converge to their corresponding deterministic models when the number of channels goes to infinity [67].

Another stochastic Hodgkin-Huxley model: the channel model

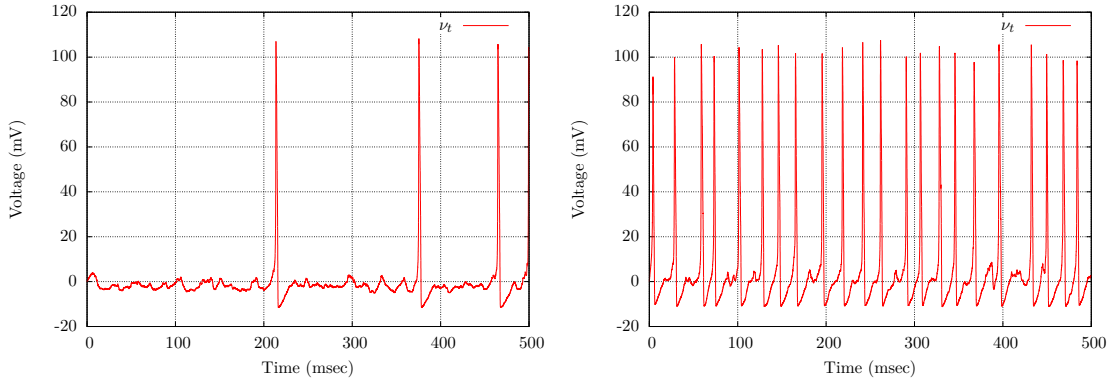


Figure 3 – Simulated spike train using the subunit model with a constant applied current $I(t) = 1$ (left) and $I(t) = 10$ (right), initial condition $(\nu, \theta^{(m)}, \theta^{(h)}, \theta^{(n)}) = (0, 0, 0, 0)$ and $N_m = N_h = N_n = 300$.

Actually, there is another stochastic Hodgkin-Huxley model which focuses on the channels themselves instead of the gates that compose it. Consequently, this model differs from the subunit model in the way the ionic conductances are modelled. We call it the channel model. It is also much used in computational Neuroscience since it describes the channel states more in detail. More precisely, let N_{Na} be the number of sodium channels and N_{K} be the number of potassium ones. We define independent jump processes $u_k^{(\mathrm{Na})}$ for $k = 1, \ldots, N_{\mathrm{Na}}$ and $u_k^{(\mathrm{K})}$ for $k = 1, \ldots, N_{\mathrm{K}}$ to model the sodium and potassium channels respectively. Unlike the dynamic of the gates (2), the dynamic of these jump processes can be represented by the diagrams in Figures 4 and 5 with voltage-dependent transition rates.

$$\begin{array}{ccccccc}
\mathbf{m_0h_0} & \underset{\beta_m}{\overset{3\alpha_m}{\rightleftarrows}} & \mathbf{m_1h_0} & \underset{2\beta_m}{\overset{2\alpha_m}{\rightleftarrows}} & \mathbf{m_2h_0} & \underset{3\beta_m}{\overset{\alpha_m}{\rightleftarrows}} & \mathbf{m_3h_0} \\
\beta_h \uparrow\downarrow \alpha_h & & \beta_h \uparrow\downarrow \alpha_h & & \beta_h \uparrow\downarrow \alpha_h & & \beta_h \uparrow\downarrow \alpha_h \\
\mathbf{m_0h_1} & \underset{\beta_m}{\overset{3\alpha_m}{\rightleftarrows}} & \mathbf{m_1h_1} & \underset{2\beta_m}{\overset{2\alpha_m}{\rightleftarrows}} & \mathbf{m_2h_1} & \underset{3\beta_m}{\overset{\alpha_m}{\rightleftarrows}} & \mathbf{m_3h_1}
\end{array}$$

Figure 4 – Sodium (Na) scheme.

$$\mathbf{n_0} \underset{\beta_n}{\overset{4\alpha_n}{\rightleftarrows}} \mathbf{n_1} \underset{2\beta_n}{\overset{3\alpha_n}{\rightleftarrows}} \mathbf{n_2} \underset{3\beta_n}{\overset{2\alpha_n}{\rightleftarrows}} \mathbf{n_3} \underset{4\beta_n}{\overset{\alpha_n}{\rightleftarrows}} \mathbf{n_4}$$

Figure 5 – Potassium (K) scheme.

The conducting state (the state that makes the channel active) of the sodium (respectively potassium) channels is $\{m_3h_1\}$ (respectively $\{n_4\}$) which corresponds to three open gates m and one open gate h (respectively four open gates n). The conductance of the membrane now depends on the empirical measures defined by the proportion of active channels of each types as follows

$$\theta_t^{(m_3h_1)} = \frac{1}{N_{\mathrm{Na}}} \sum_{k=1}^{N_{\mathrm{Na}}} \mathbf{1}_{\{m_3h_1\}}\Big(u_k^{(\mathrm{Na})}(t)\Big),$$
$$\theta_t^{(n_4)} = \frac{1}{N_{\mathrm{K}}} \sum_{k=1}^{N_{\mathrm{K}}} \mathbf{1}_{\{n_4\}}\Big(u_k^{(\mathrm{K})}(t)\Big).$$

Similarly, for $i = 0, 1, 2, 3$ and $j = 0, 1$, let $\theta^{(m_ih_j)}$ be the proportion of sodium channels in state $\{m_ih_j\}$ and for $k = 0, 1, 2, 3, 4$, let $\theta^{(n_k)}$ be the proportion of potassium channels in state $\{n_k\}$. Consequently, by noting $\theta = \Big((\theta^{(m_ih_j)})_{i,j}, (\theta^{(n_k)})_k\Big)$, the function F which defines the ODE for the evolution of the membrane potential in (3) now reads

$$F(\nu, \theta) = \overline{g}_{\mathrm{L}}(\nu - \nu_{\mathrm{L}}) - \overline{g}_{\mathrm{Na}}\theta^{(m_3h_1)}(\nu - \nu_{\mathrm{Na}}) - \overline{g}_{\mathrm{K}}\theta^{(n_4)}(\nu - \nu_{\mathrm{K}}),$$

and just as in (3), an explicit flow is available since F is linear in ν for fixed θ. Moreover the jump rate can be written in the following matrix form

$$\begin{aligned}\lambda(\theta, \nu) =& N_{\mathrm{Na}} \begin{pmatrix} \alpha_m(\nu) \\ \beta_m(\nu) \\ \alpha_h(\nu) \\ \beta_h(\nu) \end{pmatrix}^T \begin{pmatrix} 3 & 2 & 1 & 0 & 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \theta^{(m_0h_0)} \\ \vdots \\ \theta^{(m_3h_1)} \end{pmatrix} \\ &+ N_{\mathrm{K}} \begin{pmatrix} \alpha_n(\nu) \\ \beta_n(\nu) \end{pmatrix}^T \begin{pmatrix} 4 & 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} \theta^{(n_0)} \\ \vdots \\ \theta^{(n_4)} \end{pmatrix}.\end{aligned}$$

The channel model has 28 possible transitions (see Figures 4 and 5) compared to the subunit model which has only 6 (opening or closing of a gate of type m, h or n) and its jump distribution can be presented just as (4). For example, if we write $\lambda(\theta, \nu) = \sum_{i=1}^{28} \lambda_i(\theta, \nu)$ where $\lambda_1(\theta, \nu) = N_{\mathrm{Na}}\theta^{(m_0h_0)}3\alpha_m(\nu)$ denotes the rate of the transition from the state m_0h_0 to the state m_1h_0 and if we note (θ_0, ν_0) the state of the model just before a jump (or a transition), then the probability that the transition $m_0h_0 \to m_1h_0$ occurs is

$$Q\left((\theta_0, \nu_0), \{\theta_0 + (-1/N_{\mathrm{Na}}, 1/N_{\mathrm{Na}}, 0, \ldots, 0)\}\right) = \frac{\lambda_1(\theta_0, \nu_0)}{\lambda(\theta_0, \nu_0)}.$$

Deterministic and stochastic Morris-Lecar models

The original three-dimensional Morris-Lecar model has been introduced by C. Morris and H. Lecar in 1981 (see [63]) to account for various oscillating states in the barnacle

giant muscle fiber. This model also belongs to the family of conductance-based models and involves calcium and potassium conductances. It takes the following form

$$\begin{cases} C\frac{dv}{dt} = I(t) - F(v, M, N), \\ \frac{dM}{dt} = (1-M)\alpha_{\mathrm{Ca}}(v) - M\beta_{\mathrm{Ca}}(v), \\ \frac{dN}{dt} = (1-N)\alpha_{\mathrm{K}}(v) - N\beta_{\mathrm{K}}(v). \end{cases} \tag{5}$$

In this model, the $[0,1]$ valued functions M and N correspond to the probability of a calcium and potassium gate respectively to be open. The function F is given by

$$F(v, M, N) = \overline{g}_{\mathrm{L}}(v - v_{\mathrm{L}}) + \overline{g}_{\mathrm{Ca}}M(v - v_{\mathrm{Ca}}) + \overline{g}_{\mathrm{K}}N(v - v_{\mathrm{K}}).$$

The functions v, I, α_z and β_z for $z = \mathrm{Ca}, \mathrm{K}$ as well as the constants C, $\overline{g}_{\mathrm{L}}$, $\overline{g}_{\mathrm{Ca}}$, $\overline{g}_{\mathrm{K}}$, v_{L}, v_{Ca} and v_{K} have the same meaning as in the HH model (1). In their paper [63], Morris and Lecar reduce the dimension of the above model (5) by assuming that the variable M evolves much faster than N so that M is replaced by its steady state value M_∞. Consequently, they obtain the following set of equations

$$\begin{cases} C\frac{dv}{dt} = I(t) - F(v, M_\infty(v), N), \\ \frac{dN}{dt} = (1-N)\alpha_{\mathrm{K}}(v) - N\beta_{\mathrm{K}}(v). \end{cases} \tag{6}$$

The model (6) is called the reduced or the two-dimensional Morris-Lecar model. It is particularly interesting because of its low dimension. Indeed, many of its properties can be visualised on the (v, N) phase space. Stochastic counterparts of models (5) and (6) are obtained as explained above in the case of the HH model (1). We emphasize that because the function M_∞ is not linear, the ODE which gives the deterministic behaviour of the membrane potential in the stochastic two-dimensional Morris-Lecar model can not be solved explicitly.

3 Piecewise deterministic (Markov) processes

Piecewise Deterministic Processes (PDPs) have been introduced by M.H.A Davis as a general class of non-diffusive processes. These processes are based on an increasing sequence of random times in which the processes have a jump and on a deterministic evolution between two successive random times. The distribution of a PDP is thus determined by three parameters called the characteristics of the PDP: a vector field, a jump rate (intensity function) and a transition measure.

In this thesis we are interested in PDPs, that we denote by $(x_t, t \in \mathbb{R}_+)$, which has two distinct components: a discrete one which takes its values in a finite or countable set and a continuous one with values in a subset of $\mathbb{R}^d$, $d \geqslant 1$. Let Θ be a finite or countable set and let D be an open subset of $\mathbb{R}^d$. We denote by ∂D the boundary of D. Let

$$E = \{(\theta, \nu) : \theta \in \Theta, \nu \in D\}$$

be the state space of the process $(x_t) = (\theta_t, \nu_t)$. We denote by ∂E its boundary.

Let $(\Phi_\theta, \theta \in \Theta)$ be a family of functions such that $\Phi_\theta : \mathbb{R}_+ \times D \to D$ for all $\theta \in \Theta$. The functions (Φ_θ) will determine the deterministic motion of the continuous component of the PDP. Moreover, for $x = (\theta, \nu) \in E$, let us define

$$t_*(x) = \begin{cases} \inf\{t > 0 : \Phi_\theta(t, \nu) \in \partial D\}, \\ +\infty \text{ if no such time exists.} \end{cases}$$

For each $x \in E$, $t_*(x)$ is the time needed to reach the boundary ∂D following the curve $\Phi_\theta(., \nu)$ starting from the point x. Note that this time is deterministic given the starting point. For notational convenience in the sequel, we set $\Psi(t, x) := (\theta, \Phi_\theta(t, \nu))$ for all $t \geqslant 0$ and $x = (\theta, \nu) \in E$.

The jump mechanism of the PDP is described by a jump rate function $\lambda : E \to]0, +\infty[$ and a transition measure $Q : E \times \mathcal{B}(E) \to [0, 1]$ where $\mathcal{B}(E)$ denotes the σ-field generated by the Borel sets of E. We make the following hypotheses

1. The function λ is bounded,
2. $Q(x, \{x\}) = 0, \forall x \in E$.

If there were no jumps from the boundary, the assumption on λ would ensure that the resulting PDP does not blow up. In the presence of jumps from the boundary, we need an additional assumption on the transition measure to avoid blow up. Roughly speaking, this assumption ensures that the post-jump value from the boundary does not goes back to the boundary too fast (see [22] p.60).

We now present the classical construction of PDPs by M.H.A Davis [21]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space consisting in all sequences of independent uniformly distributed random variables on $[0, 1]$. We construct the PDP $(x_t, t \in \mathbb{R}_+) = ((\theta_t, \nu_t), t \in \mathbb{R}_+)$ from one such sequence. First, let

$$S_x(t) = \mathbb{1}_{t<t_*(x)} e^{-\int_0^t \lambda(\Psi(s,x))ds} \tag{7}$$

be the survival function of the inter jump times and let $G_x : [0, 1] \to \mathbb{R}_+$ be its generalised inverse defined by

$$G_x(u) = \begin{cases} \inf\{t > 0 : S_x(t) \leqslant u\}, \\ +\infty \text{ if the above set is empty.} \end{cases}$$

Moreover, there exits (see [22] p.56) a measurable function $H_x : [0, 1] \to E$ such that for all $x \in E$ and $A \in \mathcal{B}(E)$

$$\mathbb{P}(H_x(U) \in A) = Q(x, A),$$

where U is a random variable with uniform distribution on $[0, 1]$. The function H is the generalised inverse of Q.

Let $(U_k, k \geqslant 1)$ be a sequence of i.i.d random variables with uniform distribution on $[0,1]$. The sample path of the process (x_t), starting from a fixed initial point $x_0 = (\theta_0, \nu_0) \in E$ is defined as follows. Let $\omega \in \Omega$.

1. The initial condition is deterministic and is given by
$$(\theta_0(\omega), \nu_0(\omega)) = (\theta_0, \nu_0).$$

2. The component $\nu(\omega)$ follows the deterministic motion given by the function $\Phi_{\theta_0}(., \nu_0)$ as long as the discrete component $\theta(\omega)$ remains equal to θ_0. The first jump time of $\theta(\omega)$ is defined by
$$T_1(\omega) = G_{x_0}(U_1(\omega)).$$
Thus for $t \in [0, T_1(\omega)[$ we have
$$x_t(\omega) = (\theta_0, \Phi_{\theta_0}(t, \nu_0)).$$

3. At time $T_1(\omega)$ the process have a jump, its value is updated according to the distribution $Q\left((\theta_0, \Phi_{\theta_0}(T_1(\omega), \nu_0)), .\right)$, that is,
$$x_{T_1}(\omega) = H_{(\theta_0, \Phi_{\theta_0}(T_1(\omega), \nu_0))}(U_2(\omega)).$$

4. The algorithm is then repeated for $n \geqslant 2$
$$T_n(\omega) = T_{n-1}(\omega) + G_{x_{T_{n-1}}(\omega)}(U_{2n-1}(\omega)), \tag{8}$$
and
$$x_{T_n}(\omega) = H_{\left(\theta_{T_{n-1}}(\omega), \Phi_{\theta_{T_{n-1}}(\omega)}(T_n(\omega) - T_{n-1}(\omega), \nu_{T_{n-1}}(\omega))\right)}(U_{2n}(\omega)),$$
so that for $t \in [T_n(\omega), T_{n+1}(\omega)[$
$$x_t(\omega) = (\theta_{T_n}(\omega), \Phi_{\theta_{T_n}(\omega)}(t - T_n(\omega), \nu_{T_n}(\omega))).$$

As a particular case, we emphasize that the jump mechanism of stochastic conductance-based models of section 2 only concerns their discrete component and that there is no boundary. Indeed, the continuous component, which models the membrane potential, does not jump and $D = \mathbb{R}$. In this case, for $(\theta, j) \in \Theta^2$ and $\nu \in \mathbb{R}$, let $\lambda_j(\theta, \nu)$ be the rate for the transition from state θ to state j given that the membrane potential value is ν. Then, for all $x = (\theta, \nu) \in E$, the jump rate reads $\lambda(x) = \sum_{j \in \Theta} \lambda_j(x)$ and the transition measure (on the whole state space E) is given for $j \in \Theta$ and $B \in \mathcal{B}(\mathbb{R})$ by $Q(x, \{j\} \times B) = \tilde{Q}(x, \{j\})\delta_\nu(B)$ where $\tilde{Q} : E \times \mathcal{B}(\Theta) \to [0,1]$ is such that $\tilde{Q}(x, \{j\}) = \lambda_j(x)/\lambda(x)$.

Now let $(f_\theta, \theta \in \Theta)$ be a family of vector fields such that the functions $f_\theta : D \to D$ are bounded and Lipschitz continuous uniformly in θ. If we choose $\Phi_\theta = \phi_\theta$ in the

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

$$\begin{cases} \frac{dy(t)}{dt} = f_\theta\left(y(t)\right), \\ y(0) = \nu, \end{cases} \tag{9}$$

then the corresponding PDP is Markov since ϕ satisfies the semi-group property which reads $\phi_\theta(t+s, \nu) = \phi_\theta(t, \phi_\theta(s, \nu))$ for all $t, s \geqslant 0$ and for all $(\theta, \nu) \in E$. In this case, the process (x_t) is a piecewise deterministic Markov process (see [22] or [51]).

The following definition and theorems gather important properties of a stochastic process constructed as above. In their statement a PDP or a PDMP refers to a process constructed as above with characteristics (Φ, λ, Q) or (ϕ, λ, Q) respectively.

Definition 3.1. *Associated to a PDP $(x_t, t \in \mathbb{R}_+)$ we define the following processes for all $A \in \mathcal{B}(E)$.*

1. $p(t, A) = \sum_{i \geqslant 1} \mathbb{1}_{T_i \leqslant t} \mathbb{1}_{x_{T_i} \in A}$.
2. $p^*(t) = \sum_{i \geqslant 1} \mathbb{1}_{T_i \leqslant t} \mathbb{1}_{x_{T_i -} \in \partial E}$.
3. $\tilde{p}(t, A) = \int_0^t Q(x_s, A)\lambda(x_s)ds + \int_0^t Q(x_{s-}, A)p^*(ds)$.
4. $q(t, A) = p(t, A) - \tilde{p}(t, A)$.

In the above definition, p and p^* are counting processes where p^* counts the number of jumps from the boundary, $\tilde{p}$ is the compensator of p so that q is a local martingale.

Theorem 3.1 (Extended generator [22])**.** *A stochastic process $(x_t, t \in \mathbb{R}_+)$ constructed as above with characteristics (ϕ, λ, Q) is a homogeneous strong Markov càdlàg piecewise deterministic process. The domain $\mathcal{D}(\mathcal{A})$ of its extended generator $\mathcal{A}$ consists of the bounded measurable functions g defined on E such that*

1. *$t \to g(\theta, \phi_\theta(t, \nu))$ is absolutely continuous for all $x = (\theta, \nu) \in E$ and $t \in [0, t_*(x)[$.*
2. *For all $x \in \partial E$, the boundary condition $g(x) = \int_E g(y)Q(x, dy)$ is satisfied.*

For $g \in \mathcal{D}(\mathcal{A})$ the extended generator is given by

$$\mathcal{A}g(x) = (\nabla g.f)(x) + \lambda(x) \int_E (g(y) - g(x))\, Q(x, dy).$$

Theorem 3.2 (Itô formula [22])**.** *Let $(x_t, t \in \mathbb{R}_+)$ be a PDMP. Then, for $g \in \mathcal{D}(\mathcal{A})$ and for all $t \geqslant 0$*

$$g(x_t) = g(x_0) + \int_0^t \mathcal{A}g(x_s)ds + M_t^g,$$

where $M_t^g := \int_0^t \int_E (g(y) - g(x_{s-})\, q(ds, dy)$ is a true martingale with respect to the filtration generated by p.

Theorem 3.3 (Feynman-Kac formula [22]). *Let $(x_t, t \in \mathbb{R}_+)$ be a PDMP, $T > 0$ and $F : E \to \mathbb{R}$ be a bounded function. The following integro-differential equation*

$$\begin{cases} \frac{\partial}{\partial t}u(t,x) + \mathcal{A}u(t,x) = 0, & (t,x) \in [0,T[\times E, \\ u(t,x) = \int_E u(t,y)Q(x,dy), & (t,x) \in [0,T[\times \partial E, \\ u(T,x) = F(x), & x \in E, \end{cases}$$

admits a unique solution $u : \mathbb{R}_+ \times E \to \mathbb{R}$ which is given by

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

4 Simulation of non-homogeneous Poisson processes and thinning

Let λ be a positive real function defined on $[0, +\infty[$ and consider a non-homogeneous Poisson process $(N_t, t \geqslant 0)$ with intensity (or jump rate) λ and jump times $(T_n, n \geqslant 0)$. It is known that for all $n \geqslant 0$ and given that $T_n = t$ for some real $t \geqslant 0$, the survival function F of the inter-jump time $T_{n+1} - T_n$ is given by $F(s) = e^{-\int_t^{t+s} \lambda(u)du}$.
One way to simulate (N_t) (and probably the most natural) is to simulate iteratively the inter-jump times using the inverse of the survival function F^{-1} and a sequence of i.i.d random variables $(U_n, n \geqslant 1)$ with uniform distribution on $[0,1]$ according to $T_{n+1} - T_n = F^{-1}(U_{n+1})$. In this case we have $F^{-1}(u) = \Lambda^{-1}(-\ln(u) + \Lambda(t)) - t$ where $\Lambda(t) = \int_0^t \lambda(s)ds$ and Λ^{-1} denotes the inverse of Λ. Consequently, we have the following theorem.

Theorem 4.1. *Let $(T_n, n \geqslant 0)$ be a non-homogeneous Poisson process with jump rate $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$ such that $T_0 = 0$ and let $(E_n, n \geqslant 1)$ be an i.i.d sequence of exponential variables with parameter 1 independent of (T_n). Then, for $n \geqslant 0$, we have the following equality in distribution*

$$T_{n+1} = \Lambda^{-1}(E_{n+1} + \Lambda(T_n)).$$

Simulating the Poisson process (N_t) using Theorem 4.1 requires to compute Λ and Λ^{-1}. This task can be tedious especially if the jump rate λ is a complicated function which is not explicitly integrable. In this case, the computation can be done numerically using the Euler scheme for example. However, we emphasize that this method is numerically efficient when the jump rate is explicitly integrable with explicit inverse of its integral.

Another way to simulate the Poisson process (N_t) is through the thinning of a Poisson process $(N_t^*, t \geqslant 0)$ with jump times $(T_n^*, n \geqslant 0)$ and jump rate λ^* such that $\lambda(t) \leqslant \lambda^*(t)$ for all $t \geqslant 0$. The thinning method which has been introduced by Lewis and Shedler in [57] can be viewed as the analogue of the rejection method for point processes and in particular for Poisson processes. The idea of the thinning is the following. If we independently delete the points T_n^* with probability $1 - \lambda(T_n^*)/\lambda^*(T_n^*)$ then the remaining points form a non-homogeneous Poisson process with jump rate λ. The thinning method is formalised in the following theorem.

Theorem 4.2. *Let λ and λ^* be two real and positive functions defined on $[0,+\infty[$ such that $\lambda(t) \leqslant \lambda^*(t)$ for all $t \geqslant 0$. Let $(N_t, t \geqslant 0)$ and $(N_t^*, t \geqslant 0)$ be two Poisson processes with jump rate λ and λ^* respectively. Let us denote by $(T_n^*, n \geqslant 0)$ with $T_0^* = 0$ the jump times of (N_t^*). Let $(U_n, n \geqslant 0)$ be a sequence of i.i.d random variables with uniform distribution on $[0,1]$ independent of (T_n^*) and let $(\tau_n, n \geqslant 0)$ be a sequence of indexes defined iteratively by*

$$\begin{cases} \tau_0 = 0, \\ \tau_{n+1} = \inf\left\{k > \tau_n : U_k \leqslant \lambda(T_k^*)/\lambda^*(T_k^*)\right\}. \end{cases}$$

Then, the process $(T_{\tau_n}^*, n \geqslant 0)$ is a realisation of the non-homogeneous Poisson process (N_t) with jump rate λ.

The key point to efficiently simulate a non-homogeneous Poisson process by thinning is that the simulation of (N_t^*) must be simpler than the one of (N_t). The case where (N_t^*) is a homogeneous Poisson process has become classic and the resulting algorithm is easy to implement. However, it is intuitive that a constant upper bound λ^* could lead to many rejections especially if the jump rate presents significant variations over the time thus increasing the computation time too much. Consequently, if one can find a function λ^* such that $\lambda(t)/\lambda^*(t)$ is close to 1 for all $t \geqslant 0$ and such that the function $t \to \int_0^t \lambda^*(s)ds$ is explicit with explicit inverse, then, the combination of Theorems 4.2 and 4.1 will lead to an efficient simulation with few rejections and will obviates the need for numerical integration of the jump rate.

As a particular case, the thinning can be used to simulate random variables with survival function $t \mapsto e^{-\int_0^t h(s)ds}$ where h is a real and positive functions defined on $[0,+\infty[$ as follows.

Corollary 4.1. *Let h and λ^* be two real and positive functions defined on $[0,+\infty[$ such that $h(t) \leqslant \lambda^*(t)$ for all $t \geqslant 0$. Let $(T_n^*, n \geqslant 0)$ with $T_0^* = 0$ be a non-homogeneous Poisson process with jump rate λ^* independent of $(U_n, n \geqslant 1)$ a sequence of i.i.d random variables with uniform distribution on $[0,1]$. Let us define the random variable τ by*

$$\tau = \inf\left\{k > 0 : U_k \leqslant h(T_k^*)/\lambda^*(T_k^*)\right\}.$$

Then, the survival function of the random variable T_τ^* is $t \mapsto e^{-\int_0^t h(s)ds}$.

Finally, note that since the seminal paper [57] several variants and generalisations have been developed. As a variant, we can quote [26] (chapter 6) where a non homogeneous Poisson process is obtained from a two-dimensional one with unit rate. Moreover, we can find generalisations to multi-variate point process with stochastic intensity in [64], to spatial point processes in [62] and to random measures in [52].

5 Monte Carlo simulations

Let X be a random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the problem of the numerical approximation of $\mathbb{E}[X]$. To this purpose let us denote by Y

an estimator of such a quantity for which the computational complexity (or cost) is denoted by $\mathcal{C}(Y)$. The computational complexity of an estimator is usually defined as the number of operations necessary to its simulation. Generally speaking, we measure the error committed by approximating $\mathbb{E}[X]$ by Y using the Mean Squared Error (MSE) which is defined as the squared quadratic error (squared L^2-error), namely

$$\mathrm{MSE} := \mathbb{E}\left[(Y - \mathbb{E}[X])^2\right].$$

A direct computation shows that the MSE admits the following bias-variance decomposition

$$\begin{aligned}\mathrm{MSE} &= \mathbb{E}\left[(Y - \mathbb{E}[Y] + \mathbb{E}[Y] - \mathbb{E}[X])^2\right] \\ &= (\mathbb{E}[Y] - \mathbb{E}[X])^2 + \mathrm{Var}(Y). \end{aligned} \tag{10}$$

5.1 Non biased framework

Consider that the random variable X can be simulated (at a reasonable complexity) and let $(X_k, k \geqslant 1)$ denote a sequence of independent random variables distributed as X. The classical Monte Carlo (MC) estimator then reads

$$Y = \frac{1}{N}\sum_{k=1}^{N} X_k,$$

where $N \geqslant 1$ appears as a parameter. In this case $\mathbb{E}[Y] = \mathbb{E}[X]$ so that the estimator is not biased. A natural question is: how to choose N so that Y approximates $\mathbb{E}[X]$ quite accurately?

On the one hand, the strong law of large numbers which states that

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{k=1}^{N} X_k = \mathbb{E}[X],$$

suggests that if we choose N large enough then Y will be close to $\mathbb{E}[X]$. Moreover, if $\mathrm{Var(X)} < +\infty$, the central limit theorem states that the renormalized statistical error $\sqrt{N}(Y - \mathbb{E}[X])$ is approximatively distributed as a centred Gaussian with variance Var(X). This allows to build confidence intervals $\left[\mathbb{E}[X] - a\sqrt{\mathrm{Var}(X)/N}, \mathbb{E}[X] + a\sqrt{\mathrm{Var}(X)/N}\right]$ where a is a quantile of the centred Gaussian distribution with variance Var(X) and then to choose N in order to obtain the user desired precision (confidence level).

On the other hand, we can choose N in order to minimize the computational complexity of Y subject to the constraint that the MSE must be less or equal to ϵ^2 where $\epsilon > 0$ is a user prescribed error. Since Y is not biased, we have $\mathrm{MSE} = \mathrm{Var}(Y) = \frac{1}{N}\mathrm{Var}(X)$, moreover, $\mathcal{C}(Y) = N\kappa$ where κ is the complexity of a single simulation of X. The optimal sample size of the estimator is then obtained by saturation of the constraint $\mathrm{MSE} \leqslant \epsilon^2$, this leads to the following choice

$$N = \frac{\mathrm{Var}(X)}{\epsilon^2},$$

and the corresponding complexity is

$$\mathcal{C}(Y) = \frac{\kappa \mathrm{Var}(X)}{\epsilon^2} = O(\epsilon^{-2}). \tag{11}$$

We emphasize that it is not possible to build an estimator with an optimal complexity better than $O(\epsilon^{-2})$.

In the present non biased framework, both approaches (confidence interval and MSE) are similar since the precision ϵ is related to the quantity $\sqrt{\mathrm{Var}(X)/N}$. However, we will privilege the approach with the MSE in the biased framework because it explicitly takes into account the bias of the estimator through (10) whereas the approach with the confidence intervals does not.

5.2 Biased framework

Consider now that the random variable X can not be simulated (at a reasonable complexity). We then introduce a family of random variables $(X_h, h > 0)$ such that X_h can be simulated (at a reasonable complexity) for $h > 0$. Moreover, we assume that $(X_h, h > 0)$ converges to X when $h \to 0$ in the strong and weak following senses

$$\exists V_1 > 0, \beta > 0, \quad \mathbb{E}[|X_h - X|^2] \leqslant V_1 h^\beta, \tag{12}$$

and

$$\exists c_1 > 0, \alpha > 0, \quad \mathbb{E}[X_h] - \mathbb{E}[X] = c_1 h^\alpha + o(h^\alpha). \tag{13}$$

The family $(X_h, h > 0)$ will be used to construct an estimator and the strong and weak estimates (12) and (13) will be used to control respectively the variance and the bias of the estimator. We denote by $\kappa(h)$ the computational complexity induced by one simulation of X_h. It is natural to assume that $h \mapsto \kappa(h)$ is a decreasing function of h and that $\lim_{h\to 0} \kappa(h) = +\infty$ since X_h becomes closer to X when $h \to 0$, keeping in mind that X can not be simulated at a reasonable complexity. It is convenient to assume that $\kappa(h) = \overline{\kappa}/h$ where $\overline{\kappa}$ is a positive constant, see [54] or [66]. The complexity $\kappa(h)$ is usually interpreted as the number of time steps performed to simulate a realisation of X_h.

Classical Monte Carlo

Let $h > 0$ and let $(X_h^k, k \geqslant 1)$ be a sequence of independent random variables distributed as X_h. The classical MC estimator now reads

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

where h and $N \geqslant 1$ appear as parameters. In this case, the bias (using the weak estimate (13)), the variance and the complexity of the estimator read

$$\mathbb{E}[Y] - \mathbb{E}[X] = c_1 h^\alpha + o(h^\alpha), \quad \mathrm{Var}(Y) = \frac{1}{N}\mathrm{Var}(X_h), \quad \mathcal{C}(Y) = \frac{\overline{\kappa}}{h} N.$$

From the strong estimate (12) we have $\mathrm{Var}(X_h) \to \mathrm{Var}(X)$ as $h \to 0$ so that $\mathrm{Var}(X_h)$ is asymptotically a constant independent of h. Moreover, using (13) the bias of the estimator is independent of N and is asymptotically equals to $c_1 h^\alpha$. Thus, the optimal parameters $(h, N) = \mathrm{argmin}_{\mathrm{MSE} \leqslant \epsilon^2} \mathcal{C}(Y)$ are such that the bias parameter h must be of order $\epsilon^{\frac{1}{\alpha}}$ and the sample size N must be of order ϵ^{-2}. The resulting optimal complexity is then

$$\mathcal{C}(Y) = O(\epsilon^{-2-\frac{1}{\alpha}}). \tag{14}$$

Consequently, the unbiased MC complexity (11) is always better than the biased MC one (14).

Multilevel Monte Carlo

The Multilevel Monte Carlo (MLMC) method refers to the use of a MLMC estimator. This method has been introduced by S. Heinrich in [45] and developed by M. Giles in [34]. The main idea to obtain a MLMC estimator based on the family $(X_h, h > 0)$ is to consider the following telescopic sum with depth $L \geqslant 2$

$$\mathbb{E}[X_{h_L}] = \mathbb{E}[X_{h^*}] + \sum_{l=2}^{L} \mathbb{E}[X_{h_l} - X_{h_{l-1}}], \tag{15}$$

where $(h_l, 1 \leqslant l \leqslant L)$ is a geometrically decreasing sequence $h_l = h^* M^{-(l-1)}$ with $h^* > 0$ and $M > 1$ free parameters. In equality (15), the corrective term at level l, $(X_{h_l} - X_{h_{l-1}})$, is composed by two random variables, one with a fine time step, X_{h_l}, and the other with a coarse one, $X_{h_{l-1}}$. For each level $l \in \{1, \ldots, L\}$, a classical MC estimator is used to independently approximate $\mathbb{E}[X_{h_l} - X_{h_{l-1}}]$ and $\mathbb{E}[X_{h^*}]$. At each level, a number $N_l \geqslant 1$ 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 $l = 2, \ldots, L$ independent couples $(X_{h_l}, X_{h_{l-1}})$ of correlated random variables independent of X_{h^*}, the MLMC estimator then reads

$$Y = \frac{1}{N_1} \sum_{k=1}^{N_1} X_{h^*}^k + \sum_{l=2}^{L} \frac{1}{N_l} \sum_{k=1}^{N_l} (X_{h_l}^k - X_{h_{l-1}}^k), \tag{16}$$

where $(X_{h^*}^k, k \geqslant 1)$ is a sequence of independent and identically distributed random variables distributed as X_{h^*} and $\left((X_{h_l}^k, X_{h_{l-1}}^k), k \geqslant 1\right)$ for $l = 2, \ldots, L$ are independent sequences of independent copies of $(X_{h_l}, X_{h_{l-1}})$ and independent of $(X_{h^*}^k)$. From the weak estimate (13), the bias of the estimator (16) is given by

$$\mathbb{E}[Y] - \mathbb{E}[X] = c_1 h_L^\alpha + o\left(h_L^\alpha\right).$$

Note that an increase of L produces a decrease of the bias. Using the mutual independence of X_{h^*} and $(X_{h_l}, X_{h_{l-1}})$ for $l = 2, \ldots, L$, the variance of (16) is given by

$$\mathrm{Var}(Y) = \frac{1}{N_1} \mathrm{Var}(X_{h^*}) + \sum_{l=2}^{L} \frac{1}{N_l} \mathrm{Var}(X_{h_l} - X_{h_{l-1}}).$$

Note that from (13) and (12), $\mathrm{Var}(X_{h_l} - X_{h_{l-1}}) \to 0$ when $l \to +\infty$. Thus, we need only a small number N_l of samples on the finest levels. Moreover, the global computational complexity of this estimator is given by

$$\mathcal{C}(Y) = \overline{\kappa} \left(\frac{N_1}{h^*} + \sum_{l=2}^{L} N_l \left(h_l^{-1} + h_{l-1}^{-1} \right) \right). \tag{17}$$

It is known (see [34], [54], [66]) that provided that $(X_h, h > 0)$ satisfies (13) and (12) there exists values of the parameters L, $(N_l, 1 \leqslant l \leqslant L)$ such that the estimator (16) reaches a MSE less or equal to ϵ^2 with a computational complexity which satisfies

$$\mathcal{C}(Y) = \begin{cases} O(\epsilon^{-2}) & \text{if } \beta > 1, \\ O(\epsilon^{-2}(\log(\epsilon))^2) & \text{if } \beta = 1, \\ O(\epsilon^{-2-\frac{1-\beta}{\alpha}}) & \text{if } \beta < 1. \end{cases}$$

Consequently, in the case $\beta > 1$, the MLMC complexity is of the same order as the unbiased MC one (11). Moreover, in the worst case scenario $\beta < 1$, the MLMC complexity is still better than the biased MC one (14).
The computational complexity saving of a MLMC estimator depends on how fast the variance $\mathrm{Var}(X_{h_l} - X_{h_{l-1}})$ decreases as the level goes up. It is then an important matter to build MLMC estimators with strong order convergence $\beta > 1$.

Actually, a more general MLMC estimator can be obtained by considering at each level different approximations for fine and coarse simulations. Let $(X_h^f, h > 0)$ and $(X_h^c, h > 0)$ be two families of random variables which satisfy the following strong estimate

$$\forall l \in \{2, \dots, L\}, \exists V_1 > 0, \beta > 0, \quad \mathbb{E}[|X_{h_l}^f - X_{h_{l-1}}^c|^2] \leqslant V_1 h_l^{\beta}, \tag{18}$$

and which weakly converge to X in the same sense as (13), so that, for $h > 0$, $\mathbb{E}[X_h^f] = \mathbb{E}[X_h^c]$. In this case, equality (15) reads

$$\mathbb{E}[X_{h_L}^f] = \mathbb{E}[X_{h^*}^c] + \sum_{l=2}^{L} \mathbb{E}[X_{h_l}^f - X_{h_{l-1}}^c],$$

and the corresponding MLMC estimator reads

$$Y = \frac{1}{N_1} \sum_{k=1}^{N_1} X_{h^*}^{c,k} + \sum_{l=2}^{L} \frac{1}{N_l} \sum_{k=1}^{N_l} (X_{h_l}^{f,k} - X_{h_{l-1}}^{c,k}). \tag{19}$$

To finish this section, we emphasize that there is (at least) two ways to determine the parameters L, $(N_l, l = 1, \dots, L)$ of a multilevel estimator which respect the condition MSE $= \epsilon^2$. On the one hand, the initial way (see [34]) is through the use of an iterative algorithm in which the parameters are determined a posteriori. This algorithm is as

follows. Start with $L = 2$ and build a MLMC estimator with depth level L where the sample sizes N_l at levels $l = 1, \dots, L$ are given by

$$N_l = \left\lceil M\epsilon^{-2}\sqrt{V_l h_l}\left(\sum_{l=1}^{L}\sqrt{V_l/h_l}\right)\right\rceil,$$

where V_l is a rough estimation of the variance of the corrective term at level l. This makes the estimated variance of the estimator less than $\frac{1}{2}\epsilon^2$. Then, in order to ensure that the bias is less than $\frac{1}{\sqrt{2}}\epsilon$ use the following test

$$|Y_L| < \frac{1}{\sqrt{2}}(M-1)\epsilon, \tag{20}$$

where Y_l denotes the classical MC estimator used at level l. If (20) is verified then stop otherwise set $L = L + 1$ and continue until (20) is verified.
On the other hand, in [54] (see also [66]) the authors show that we can determine the parameters a priori through the following optimisation problem

$$(L, N_1, \dots, N_L) = \underset{\mathrm{MSE}\leqslant\epsilon^2}{\operatorname{argmin}}\,\mathcal{C}(Y). \tag{21}$$

In order to give some details about the resolution of (21), let us first introduce some notations. Consider that the sample size at level l is given by $N_l = \lceil Nq_l \rceil$ where N is the global sample size and $q = (q_l, 1 \leqslant l \leqslant L)$ is a stratification with $\sum_{l=1}^{L} q_l = 1$. The complexity of the estimator (see (17)) then reads $\mathcal{C}(Y) = NC(q, L)$ and the MSE is then asymptotically of the form $\mathrm{MSE} = \mu^2(L) + \frac{1}{N}\nu(q, L)$ where $\mu(L)$ denotes the bias of the estimator and $\nu(q, L)/N$ its variance. Let $\phi(q, L) = \nu(q, L)C(q, L)$ denotes the effort of the estimator (the product of the variance and complexity) which is independent of N. Since the problem (21) can not be solved directly, the authors decompose it in three steps. Firstly, they fix L and determine the stratification q^* which minimize the effort, that is

$$q^* = \operatorname{argmin}\phi(q, L).$$

Secondly, using the optimal stratification and always with fixed L they saturate the constraint $\mathrm{MSE} = \epsilon^2$ with respect to N in order to find the optimal sample size

$$N^* = \frac{\nu(q^*, L)}{\epsilon^2 - \mu^2(L)}.$$

Thirdly, they are able to choose the depth level L which minimize the complexity as follows

$$L^* = \underset{\mu(L)<\epsilon}{\operatorname{argmin}}\, N^* C(q^*, L).$$

6 Results of the thesis

In this section, we present the results of the present thesis chapter by chapter.

6.1 Chapter 1: Exact simulation of the jump times of a class of Piecewise Deterministic Markov Processes

This chapter has been published in the Journal of Scientific Computing (see [55]). The aim of this chapter is to introduce an exact simulation (perfect sampling) algorithm for the class of PDMPs whose flows are known. This means that we explicitly know the solution of each ordinary differential equation associated to each vector field. We focus on the (exact) simulation of the inter-jump times considering that the post-jump values can be simulated exactly. We emphasize that the post-jump values are discrete random variables (finite or countable) for which the exact simulation is, in general, not an issue. M.H.A Davis in [21] or [22] provides an iterative construction of PDMPs which suggests to simulate the inter-jump times by inversion of their survival function (see also the section 3 of the present introduction). The survival function is expressed using the integral of the jump rate along the flow. Thus, we are not guaranteed to be able to inverse the survival function even if the flows are known explicitly. To overcome this difficulty we use the thinning method [57] (see also the section 4 of this introduction). We propose different kinds of bounds for the intensity along the flow, from coarse to path-adapted:

- the *global bound*, the coarsest, which is constant (in particular it is independent of the state of the PDMP and of time),
- the *local bound*, which depends on the post-jump value of the PDMP and which is constant between two successive jump times,
- the *optimal bound*, the finest, which depends on the post-jump value of the PDMP and also on the time evolution of the process between two successive jump times.

The main interest of the optimal bound is that the thinning algorithm with such a bound applies with weaker hypotheses on the jump rate than with the classical global bound. More precisely, the optimal bound requires that the jump rate is locally bounded along a given flow whereas the global bound requires that it is globally bounded on the state space. Moreover, the optimal bound provides a powerful thinning algorithm. The drawback of this bound is that when the bound becomes very close to the actual jump rate, the computation time may be too long. It is thus necessary to look for a satisfactory balance.
Our main contribution is the theoretical study of their respective efficiency. We choose to define the efficiency as the mean value of the ratio between the number of selected jump times and the number of generated jump times. This indicator is between 0 and 1 and is easily understood, the closer it is to 1 the less we reject points, thus the more efficient the algorithm is.
As an application, we use the subunit and the channel model (see section 2) to numerically compare the efficiency of the different bounds (global, local, optimal). In this introduction we only present the results concerning the channel model since they are similar to those obtained with the subunit model. The comparison of the bounds enables us to show that the optimal bound speeds up simulation compared to the global and the local bound.

Let $E = \Theta \times D$ where Θ is a finite or countable set and D is an open subset of $\mathbb{R}^d$, $d \geqslant 1$. We consider a E-valued PDMP $(x_t, t \geqslant 0)$ with characteristics (ϕ, λ, Q) where ϕ is an explicit solution of the ODE (9) and λ, Q are as in section 3. Depending on which bound we use, we assume that one of the following assumptions is satisfied.

Additional assumptions on the jump rate:

$\mathbf{H^{glo}}$: $\sup_{x\in E} \lambda(x) < \infty$.

$\mathbf{H^{loc}}$: $\forall x \in E, \quad \sup_{s\geqslant 0} \lambda(\psi(s,x)) < \infty$.

$\mathbf{H^{opt}}$: $\forall x \in E, \forall I \subseteq \mathbb{R}_+, \quad \sup_{s\in I} \lambda(\psi(s,x)) < \infty$.

For all $x = (\theta, \nu) \in E$ and $t \geqslant 0$ we set $\psi(t,x) = (\theta, \phi_\theta(t,\nu))$. In this case, the survival function of the inter-jump times is given by (7) with $\Psi = \psi$ so that the jump-times occur either in a deterministic way (when the flow hits a boundary) or in a Poisson-like fashion. In order to construct (and simulate) the inter-jump times by thinning, we first prove the following lemma.

Lemma 6.1. *Let $T > 0$ and $g : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-negative, locally integrable function. Define*

$$S(t) \equiv \mathbf{1}_{t<T} e^{-\int_0^t g(s)ds}, \quad \overline{S}(t) \equiv e^{-\int_0^t g(s)ds}.$$

Let Y ($\overline{Y}$ respectively) be a random variable with survival function S ($\overline{S}$ respectively). Then, we have $Y \overset{law}{=} \overline{Y} \wedge T$.

Thinning algorithm

We describe the thinning algorithm with a generic function $\tilde{\lambda} : \mathbb{R}_+ \times E \to \mathbb{R}_+$ which is assumed to have the following properties:

- $\forall u \geqslant 0$, $\forall y \in E$,
$$\lambda(\psi(u,y)) \leqslant \tilde{\lambda}(u,y).$$
- $\forall u \geqslant 0$, $\forall y \in E$, the function $\tilde{\Lambda}_y(u) \equiv \int_0^u \tilde{\lambda}(v,y)dv$ is explicitly computable.
- $\forall y \in E$, the inverse of $\tilde{\Lambda}_y$, denoted by $\left(\tilde{\Lambda}_y\right)^{-1}$, is explicitly computable.

We simulate a sample path of the PDMP $(x_t, t \geqslant 0)$ with values in E, starting from a fixed initial point $x_0 \in E$ at time 0 as follows.

Let $(\tilde{T}_k^0, k \geqslant 0)$ be a non homogeneous Poisson process with jump rate $\tilde{\lambda}(t, x_0)$ for $t \geqslant 0$, and,

$$\tau_1 = \inf\{k > 0 : U_k^{(1)} \tilde{\lambda}(\tilde{T}_k^0, x_0) \leqslant \lambda(\psi(\tilde{T}_k^0, x_0))\},$$

where $(U_n^{(1)}, n \geqslant 1)$ is a sequence of independent random variables with uniform distribution on $[0,1]$, independent of $(\tilde{T}_k^0)$. Then, from Corollary 4.1 and Lemma 6.1 we have $T_1 = \tilde{T}_{\tau_1}^0 \wedge t_*(x_0)$. On $[0, T_1[$ the PDMP evolves deterministically according to the flow

$\psi(.,x_0)$. Then we simulate the post-jump value as a random variable whose conditional distribution is $Q\left(\psi(T_1,x_0),.\right)$.

Suppose we have simulated the PDMP up to time T_i. Then, conditionally on (T_i,x_{T_i}), the PDMP (x_t) restarts from x_{T_i} at time T_i independently from the past. Let $(\tilde{T}_k^i, k\geqslant 0)$ be a Poisson process with jump rate $\tilde{\lambda}(t-T_i,x_{T_i})$ for $t\geqslant T_i$, and,

$$\tau_{i+1}=\inf\{k>0 : U_k^{(i+1)}\tilde{\lambda}(\tilde{T}_k^i,x_{T_i}))\leqslant \lambda(\psi(\tilde{T}_k^i,x_{T_i}))\},$$

where $(U_n^{(i+1)}, n\geqslant 1)$ is a sequence of independent uniform random variables, independent of $(\tilde{T}_k^i)$ and x_{T_i}. Then, always from Corollary 4.1 and Lemma 6.1 we have $T_{i+1}=T_i+\tilde{T}^i_{\tau_{i+1}}\wedge t_*(x_{T_i})$. On $[T_i,T_{i+1}[$ the process evolves according to the flow $\psi(.-T_i,x_{T_i})$. The post-jump value has distribution $Q\left(\psi(T_{i+1}-T_i,x_{T_i}),.\right)$ and so on.

Note that given $(T_1,x_{T_1}),\dots,(T_i,x_{T_i})$ respectively, we efficiently simulate the Poisson processes $(\tilde{T}_k^0),(\tilde{T}_k^0),\dots,(\tilde{T}_k^i)$ respectively using Theorem 4.1.

We formally define the efficiency of a generic bound $\tilde{\lambda}$, that we call rate of acceptance, by $\mathbb{E}[N_t/\tilde{N}_t]$ where (N_t) is the counting process associated to the sequence of jump times of the PDMP and $(\tilde{N}_t)$ is the counting process whose (stochastic) intensity is $\tilde{\lambda}(t,x_t)$.

The bounds

The three different bounds we consider are formally defined by:

global bound: $\tilde{\lambda}^{\text{glo}}:=\sup_{x\in E}\lambda(x)$.

local bound: $\tilde{\lambda}^{\text{loc}}(y):=\sup_{s\geqslant 0}\lambda\left(\psi(s,y)\right),\ \ \forall y\in E$.

optimal bound: $\tilde{\lambda}^{\text{opt}}(u,y):=\sum_{k\in P}\sup_{s\in\mathcal{P}_k}\lambda(\psi(s,y))\mathbf{1}_{\mathcal{P}_k}(u),\ \ \forall u\geqslant 0,\forall y\in E$, where P is a finite or countable set and $(\mathcal{P}_k, k\in P)$ is a partition of $[0,+\infty[$.

The three additional hypotheses $\mathbf{H^{glo}}$, $\mathbf{H^{loc}}$, $\mathbf{H^{opt}}$ ensure that the functions $\tilde{\lambda}^z$, $\tilde{\Lambda}^z$ and $\left(\tilde{\Lambda}^z\right)^{-1}$ are well defined, $z\in\{\text{glo},\text{loc},\text{opt}\}$. Considering the optimal bound, note that we are free to choose the partition we want and for any partition the simulation remains exact. The simplest partition (also the one we use to expose the numerical results below) is obtained by letting $\epsilon>0$ and setting $P=\{0,1\}$ and $\mathcal{P}_0=[0,\epsilon[$, $\mathcal{P}_1=[\epsilon,+\infty[$. The optimal bound using this partition is called the optimal-$\mathcal{P}^\epsilon$ bound.

Numerical results

To numerically compare the efficiency of the three bounds (optimal-$\mathcal{P}^\epsilon$, local, global), we simulate 10^5 trajectories of the channel model on a finite time interval $[0,T]$ and we approximate the rate of acceptance and the simulation time using a classical Monte Carlo estimator. The numerical results are given in Figure 7 and Table 1. Moreover, Figure 6 shows the ratio $\lambda(x_t)/\tilde{\lambda}(t,x_t)$ for the three bounds as a function of time.

Figure 7 shows the computation time and the rate of acceptance using the optimal-$\mathcal{P}^\epsilon$ bound as a function of the parameter ϵ. In all cases (i.e $N_{\text{chan}}=30,300,3000$), there is a

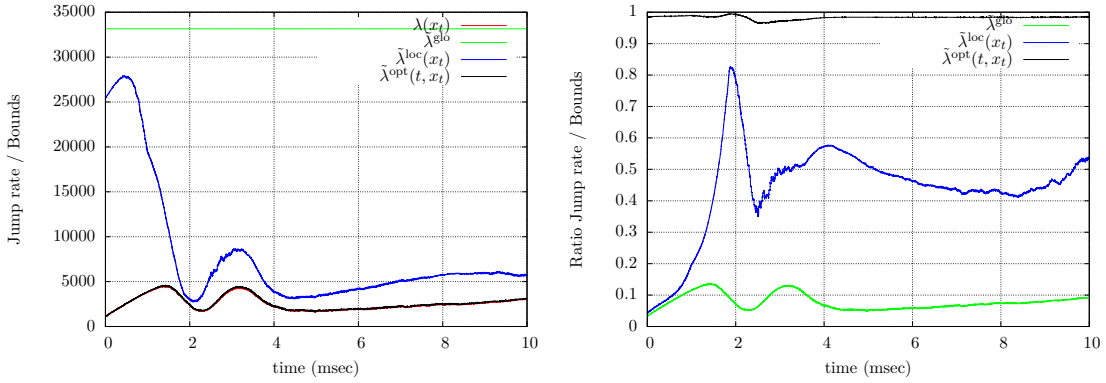


Figure 6 – Simulated trajectories of the jump rate and the different bounds (optimal-$\mathcal{P}^{0.005}$, local, global) (left) and of the corresponding ratio $\lambda/\tilde{\lambda}$ (right) in the channel model with $N_{\mathrm{Na}} = N_{\mathrm{K}} = 3000$.

value of ϵ which minimizes the computation time and maximizes the rate of acceptance. This optimal ϵ is inversely proportional to the jump rate. Thus, in order to efficiently use this optimal bound one has to take a small ϵ (respectively large) when the jumps frequency is high (respectively low). More precisely, the optimal computation time and rate of acceptance are obtained for ϵ of order $\max_n |T_{n+1} - T_n|$.

The results in Table 1 indicate that the simulation time is approximately reduced by 2 in going from the global bound to the local bound and it is again approximately reduced by 2 in going from the local to the optimal bound. However note that the rate of acceptance is refined by a factor of approximately 4 in going from the global to the local bound and is again refined by a factor 4 from the local to the optimal bound.

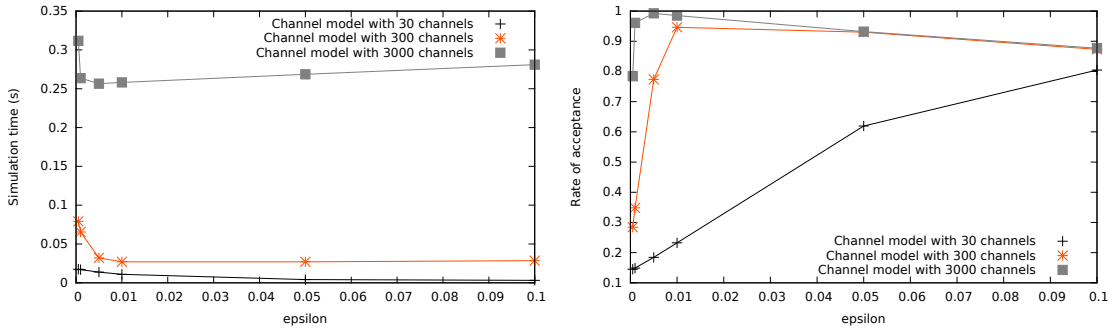


Figure 7 – Computation time and rate of acceptance with the optimal-$\mathcal{P}^{\epsilon}$ bound as a function of the parameter ϵ in the channel model with $N_{\mathrm{Na}} = N_{\mathrm{K}} = N_{\mathrm{chan}}$ where N_{chan} denotes the number of channels.

Table 1 – Computation time and rate of acceptance for the three bounds (optimal, local, global) in the channel model with $N_{\mathrm{Na}} = N_{\mathrm{K}} = N_{\mathrm{chan}}$ where N_{chan} denotes the number of channels.

N_{chan}	Bound	computation time (sec)	rate of acceptance
30	Optimal-$\mathcal{P}^{0.1}$	0,003 $(\pm 1.10^{-6})$	0,80 $(\pm 1.10^{-3})$
	Local	0,008 $(\pm 2.10^{-4})$	0,14 $(\pm 1.10^{-3})$
	Global	0,012 $(\pm 3.10^{-4})$	0,06 $(\pm 2.10^{-5})$
300	Optimal-$\mathcal{P}^{0.01}$	0,027 $(\pm 5.10^{-4})$	0,95 $(\pm 9.10^{-5})$
	Local	0,05 $(\pm 1.10^{-3})$	0,22 $(\pm 8.10^{-5})$
	Global	0,120 $(\pm 2.10^{-3})$	0,06 $(\pm 1.10^{-5})$
3000	Optimal-$\mathcal{P}^{0.005}$	0,26 $(\pm 3.10^{-2})$	0,99 $(\pm 9.10^{-6})$
	Local	0,474 $(\pm 4.10^{-2})$	0,24 $(\pm 7.10^{-6})$
	Global	1.184 $(\pm 3.10^{-1})$	0,06 $(\pm 9.10^{-8})$

6.2 Chapter 2: Thinning and Multilevel Monte Carlo for Piecewise Deterministic (Markov) Processes. Application to a stochastic Morris-Lecar model.

This chapter has been submitted to the Journal of Applied Probability, the preprint [56] is available on arXiv or HAL. The aim of this chapter is to extend the Multilevel Monte Carlo (MLMC) method to approximate expectations of a function of the state of a PDMP at fixed time. In the first part of this chapter we study approximations of trajectories of Piecewise Deterministic Processes (PDP) when the flow is not explicit by the thinning method. We also establish strong error estimates 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 the 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 does indeed outperform the classical Monte Carlo one.

Let $E = \Theta \times \mathbb{R}^d$ where Θ is a finite or countable set, $d \geqslant 1$ and let $T > 0$. We first consider a finite time horizon E-valued PDP $(x_t, t \in [0,T])$ with characteristics (Φ, λ, Q), without jumps from the boundary, that we construct by thinning of a homogeneous Poisson process as in chapter 1. We also consider that the functions $(\Phi_\theta, \theta \in \Theta)$ are not known explicitly and we use a numerical scheme $\overline{\Phi}_\theta$ (with implicit time step h) approximating Φ_θ for which there exits positive constants C_1 and C_2 independent of h and θ such that

$$\sup_{t\in[0,T]} |\Phi_\theta(t,\nu_1) - \overline{\Phi}_\theta(t,\nu_2)| \leqslant e^{C_1 T}|\nu_1 - \nu_2| + C_2 h, \qquad \forall \theta \in \Theta, \forall (\nu_1,\nu_2) \in \mathbb{R}^{2d}. \quad (22)$$

We associate to the family $(\overline{\Phi}_\theta, \theta \in \Theta)$ a PDP also constructed by thinning that we denote

$(\overline{x}_t, t \in [0,T])$. We emphasize that the processes (x_t) and $(\overline{x}_t)$ are correlated via the thinning of the same homogeneous Poisson process. We prove in the following theorem a strong error estimate for PDPs.

Theorem 6.1. *Let $(x_t, t \in [0,T])$ and $(\overline{x}_t, t \in [0,T])$ be two correlated PDPs with characteristics (Φ, λ, Q) and $(\overline{\Phi}, \lambda, Q)$ such that $x_0 = \overline{x}_0 = x$ for some $x \in E$. Assume that Θ is finite and that for all $\theta \in \Theta$ and for all $A \in \mathcal{B}(\Theta)$ the functions $\lambda(\theta, .)$ and $Q((\theta, .), A)$ are Lipschitz uniformly in θ. Then, for all bounded functions $F : E \to \mathbb{R}$ such that for all $\theta \in \Theta$ the function $\nu \mapsto F(\theta, \nu)$ is L_F-Lipschitz where L_F is positive and independent of θ, there exists constants $V_1 > 0$ and $V_2 > 0$ independent of the time step h such that*

$$\mathbb{E}\left[|F(\overline{x}_T) - F(x_T)|^2\right] \leqslant V_1 h + V_2 h^2. \tag{23}$$

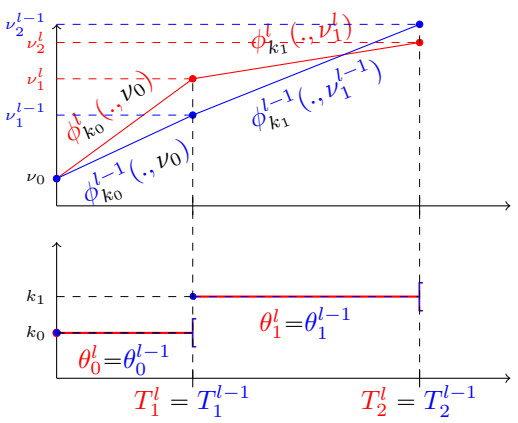
The result of Theorem 6.1 is mainly based on the construction of the couple $(x_t, \overline{x}_t)$ and on the fact that the Euler scheme is of order 1 this is why it is valid for a general PDP and its Euler scheme. Since the PDPs (x_t) and $(\overline{x}_t)$ are constructed using two different functions Φ and $\overline{\Phi}$ the probability of accepting a proposed jump time differs from one process to the other. Consequently, the sequence of jump times of both processes may be different. Moreover the discrete components of the post-jump locations may also be different. The presence of the term $V_1 h$ results from the trajectories of $((x_t, \overline{x}_t), t \in [0,T])$ where the jump times and/or the discrete components differ (see Figure 8b) whereas the term $V_2 h^2$ results from those where the jump times and the discrete components are equal (see Figure 8a).

Consider now that $\Phi_\theta = \phi_\theta$ for all $\theta \in \Theta$ where ϕ_θ is a non explicit solution of (9). In this case, the process (x_t) is a PDMP with characteristics (ϕ, λ, Q). Moreover, let us denote by $\overline{\phi}_\theta$ the continuous Euler scheme (also called Euler polygon) which approximate ϕ_θ with some time step $h > 0$ and let $(\overline{x}_t)$ be a PDP with characteristics $(\overline{\phi}, \lambda, Q)$. We emphasize that $(\overline{x}_t)$ is not Markov since the continuous Euler scheme $\overline{\phi}$ fails to satisfy the semi-group property and that ϕ and $\overline{\phi}$ satisfy estimate (22). In this case, we prove in the following theorem a weak error expansion for PDMPs.

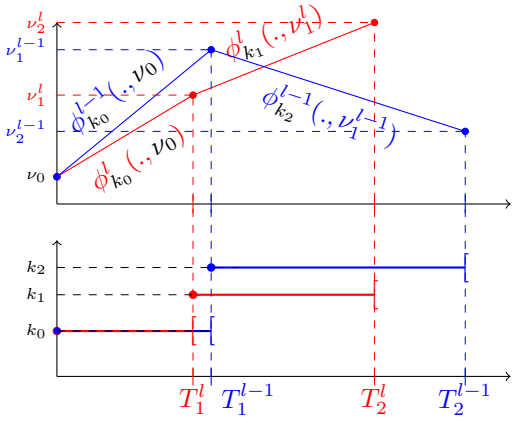
Theorem 6.2. *Let $(x_t, t \in [0,T])$ be a PDMP with characteristics (ϕ, λ, Q) and let $(\overline{x}_t, t \in [0,T])$ be a PDP with characteristics $(\overline{\phi}, \lambda, Q)$ such that $x_0 = \overline{x}_0 = x$ for some $x \in E$. Assume that for all $\theta \in \Theta$ and for all $A \in \mathcal{B}(\Theta)$, the functions $Q((\theta, .), A)$, $\lambda(\theta, .)$ and $f_\theta(.)$ are bounded and twice continuously differentiable with bounded derivatives. Assume moreover that the solution u of the integro differential equation*

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

with $F : E \to \mathbb{R}$ a bounded function and $\mathcal{A}$ the generator of the process (t, x_t) 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 and that the second derivatives of $(t,\nu) \mapsto u(t,\theta,\nu)$ are uniformly



(a) The jumps times and the discrete components are equals. Thus, the continuous components of both processes $(\overline{x}_t^{h_l})$ and $(\overline{x}_t^{h_{l-1}})$ are given, between two successive jump times, by the Euler discretisation of the same flows ϕ_{k_0}, ϕ_{k_1} with two time steps h_l and h_{l-1} and different but close initial points. These typical trajectories result in the presence of the term V_2h^2 in the estimate (23).



(b) The jumps times and the discrete components are not equal. First, we observe a temporal offset between the two processes. Moreover, when the discrete components are not equal, the continuous components of both processes $(\overline{x}_t^{h_l})$ and $(\overline{x}_t^{h_{l-1}})$ are given by the Euler discretisation of two different flows ϕ_{k_1} and ϕ_{k_2} with two time steps h_l and h_{l-1}. These typical trajectories result in the presence of the term V_1h in the estimate (23).

Figure 8 – Illustration of two typical behaviours of the couple of processes $(\overline{x}_t^{h_l}, \overline{x}_t^{h_{l-1}})$ involved at level l of the estimator (25).

Lipschitz in θ. Then, for any bounded function $F : E \to \mathbb{R}$ there exists a constant c_1 independent of h such that

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

The result of Theorem 6.2 mainly relies on the Feynman-Kac formula for PDMPs and so on the Markov property. Note that a similar weak error expansion has been proved for stochastic differential equations using the Feynman-Kac formula for those processes.

We now want to apply the MLMC method in order to approximate expectations of the form $\mathbb{E}[F(x_T)]$ where $(x_t, t \in [0,T])$ is a PDMP and $F : E \to \mathbb{R}$ is a smooth function. The MLMC method relies simultaneously on Theorems 6.1 and 6.2 that is why we study its application to the PDMP framework instead of the more general PDP one. The results of Theorems 6.1 and 6.2 indicate that the family (indexed by a time step h) of random variables $(F(\overline{x}_T), h > 0)$ converges strongly and weakly to $F(x_T)$ as in (12) and (13) which are the building blocks of biased Monte Carlo simulations (see section 5.2). Moreover, the same theorems suggest to investigate the use of the MLMC method in the PDMP framework with $\beta = 1$ and $\alpha = 1$.

Setting $X_h := F(\overline{x}_T)$ for $h > 0$ to emphasize the dependence of $\overline{x}_T$ on a time step h, we build a classical MC estimator (denoted by Y^{MC}) and a MLMC estimator (denoted by Y^{MLMC}) of $\mathbb{E}[F(x_T)]$ as follows

$$Y^{\mathrm{MC}} = \frac{1}{N}\sum_{k=1}^{N} X_h^k, \tag{24}$$

where $(X_h^k, k \geqslant 1)$ is an i.i.d sequence of random variables distributed like X_h and

$$Y^{\mathrm{MLMC}} = \frac{1}{N_1}\sum_{k=1}^{N_1} X_{h*}^k + \sum_{l=2}^{L}\frac{1}{N_l}\sum_{k=1}^{N_l}(X_{h_l}^k - X_{h_{l-1}}^k), \tag{25}$$

where $\left((X_{h_l}^k, X_{h_{l-1}}^k), k \geqslant 1\right)$ for $l = 2, \ldots, L$ are independent sequences of independent copies of the couple $(X_{h_l}, X_{h_{l-1}})$ and independent of the i.i.d sequence $(X_{h*}^k, k \geqslant 1)$.

In order to improve the convergence rate of the MLMC estimator (25) (to increase the parameter β in Theorem 6.1) we prove that the following representation holds.

Proposition 6.1. *Let $(x_t, t \in [0,T])$ and $(\tilde{x}_t, t \in [0,T])$ be two PDPs with characteristics (Φ, λ, Q) and $(\Phi, \tilde{\lambda}, \tilde{Q})$ respectively such that $x_0 = \tilde{x}_0 = x$ for a given $x = (\theta, \nu) \in E$. Assume that $\tilde{\lambda}$ and $\tilde{Q}$ depend only on θ, that $\tilde{Q}$ is always positive and $0 < \tilde{\lambda}(\theta) < \lambda^*$ for all $\theta \in \Theta$. Then, there exists a process $(\tilde{R}_t, t \in [0,T])$ which depends on $\Phi, \lambda, Q, \tilde{\lambda}, \tilde{Q}$ and $(\tilde{x}_t, t \in [0,T])$ such that for all $t \in [0,T]$ and for all bounded measurable functions $g : E \to \mathbb{R}$, we have*

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

The fact that $\tilde{\lambda}$ and $\tilde{Q}$ only depend on θ implies that the jump mechanism of the PDP $(\tilde{x}_t)$ is given by an autonomous Markov chain (independent of Φ). Consequently, the jump mechanism of the PDPs $(\underline{\tilde{x}}_t)$ with characteristics $(\overline{\Phi}, \tilde{\lambda}, \tilde{Q})$ for $h > 0$ is exactly the same as the one of $(\tilde{x}_t)$, that is, the jump times and the discrete components are the same. This situation is illustrated in Figure 9.

From Proposition 6.1, we can then decompose $\mathbb{E}[F(\overline{x}_T^{h_L})]$ over the levels using the scheme $(F(\underline{\tilde{x}}_T^h)\underline{\tilde{R}}_T^h, h > 0)$ as follows

$$\mathbb{E}[F(\overline{x}_T^{h_L})] = \mathbb{E}[F(\underline{\tilde{x}}_T^{h^*})\underline{\tilde{R}}_T^{h^*}] + \sum_{l=2}^{L}\mathbb{E}[F(\underline{\tilde{x}}_T^{h_l})\underline{\tilde{R}}_T^{h_l} - F(\underline{\tilde{x}}_T^{h_{l-1}})\underline{\tilde{R}}_T^{h_{l-1}}],$$

where, for $l = 1, \ldots, L$, the discrete components of the processes $(\underline{\tilde{x}}_t^{h_l})$ jump at the same times and in the same states and the processes $(\underline{\tilde{R}}_t^{h_l})$ are such that $\mathbb{E}[F(\overline{x}_T^{h_l})] = \mathbb{E}[F(\underline{\tilde{x}}_T^{h_l})\underline{\tilde{R}}_T^{h_l}]$. Then, letting $\tilde{X}_h = F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T$ for $h > 0$, we define a second MLMC estimator noted $\tilde{Y}^{\mathrm{MLMC}}$ as follows

$$\tilde{Y}^{\mathrm{MLMC}} = \frac{1}{N_1}\sum_{k=1}^{N_1} \tilde{X}_{h*}^k + \sum_{l=2}^{L}\frac{1}{N_l}\sum_{k=1}^{N_l}(\tilde{X}_{h_l}^k - \tilde{X}_{h_{l-1}}^k). \tag{26}$$

We prove in the following theorem a strong error estimate for the numerical scheme $(F(\underline{\tilde{x}}_T^h)\underline{\tilde{R}}_T^h, h > 0)$.

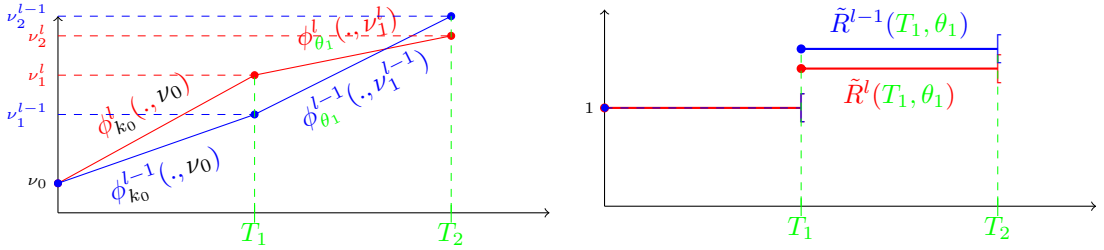


Figure 9 – Illustration of the typical behaviour of the couple of processes $\left((\underline{\tilde{x}}_t^{h_l}, \underline{\tilde{R}}_t^{h_l}), (\underline{\tilde{x}}_t^{h_{l-1}}, \underline{\tilde{R}}_t^{h_{l-1}})\right)$ involved at level l of the estimator (26). In this situation, the jump times and the discrete components of both processes $(\underline{\tilde{x}}_t^{h_l})$ and $(\underline{\tilde{x}}_t^{h_{l-1}})$ are the same and are equal to those of an autonomous Markov chain (independent of $\overline{\phi}^l$ and $\overline{\phi}^{l-1}$) represented in green in the above graphics. Thus, the continuous components of both $(\underline{\tilde{x}}_t^{h_l})$ and $(\underline{\tilde{x}}_t^{h_{l-1}})$ evolves as described on Figure 8a (see the left hand side graphic). Moreover, the corrective processes $(\underline{\tilde{R}}_t^{h_l})$ and $(\underline{\tilde{R}}_t^{h_{l-1}})$ are not necessarily close to 1 but are close to each other (see the right hand side graphic). Consequently, the corresponding L^2 error (see (27)) is of order $O(h_l^2)$.

Theorem 6.3. *Let (x_t) and $(\tilde{x}_t)$ be as in Proposition 6.1. Let $(\overline{x}_t, t \in [0,T])$ and $(\underline{\tilde{x}}_t, t \in [0,T])$ be two PDPs with characteristics $(\overline{\Phi}, \lambda, Q)$ and $(\overline{\Phi}, \tilde{\lambda}, \tilde{Q})$ respectively. Let $(\tilde{R}_t, t \in [0,T])$ and $(\underline{\tilde{R}}_t, t \in [0,T])$ be as in Proposition 6.1, that is such that, $\mathbb{E}[g(x_t)] = \mathbb{E}[g(\tilde{x}_t)\tilde{R}_t]$ and $\mathbb{E}[g(\overline{x}_t)] = \mathbb{E}[g(\underline{\tilde{x}}_t)\underline{\tilde{R}}_t]$. Assume that for all $\theta \in \Theta$ and for all $A \in \mathcal{B}(\Theta)$ the functions $\lambda(\theta,.)$ and $Q((\theta,.),A)$ are Lipschitz uniformly in θ. Then, for all bounded functions $F : E \to \mathbb{R}$ such that for all $\theta \in \Theta$ the function $\nu \mapsto F(\theta,\nu)$ is L_F-Lipschitz ($L_F > 0$), there exists a positive constant $\tilde{V}_1$ independent of the time step h such that*

$$\mathbb{E}\left[|F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T - F(\tilde{x}_T)\tilde{R}_T|^2\right] \leqslant \tilde{V}_1 h^2. \tag{27}$$

Thus, we end up with $\beta = 2$ in (12) with $X = F(\tilde{x}_T)\tilde{R}_T$ and $X_h = F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T$ so that the complexity goes from a $O(\epsilon^{-2}(\log(\epsilon))^2)$ to a $O(\epsilon^{-2})$. We also prove in the following proposition another representation which allows to build a MLMC estimator with two different numerical schemes (see (19)).

Proposition 6.2. *Let $(x_t, t \in [0,T])$ and $(\tilde{x}_t, t \in [0,T])$ be two PDPs with characteristics (Φ, λ, Q) and $(\tilde{\Phi}, \lambda, Q)$ respectively and let $x_0 = \tilde{x}_0 = x$ for a given $x = (\theta,\nu) \in E$. We denote by (θ_n) and (T_n) the discrete component and the jump times respectively of (x_t). Assume that Q is always positive and that $0 < \lambda(x) < \lambda^*$ for all $x \in E$. Let $(\mu_n, n \in \mathbb{N})$ 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 $n \geqslant 1$ and let us define for all $t \in [0,T]$, $y_t = (\theta_n, \tilde{\Phi}_{\theta_n}(t - T_n, \mu_n))$ if $t \in [T_n, T_{n+1}[$. Then, there exists a process $(\tilde{R}_t, t \in [0,T])$ which depends on $\Phi, \tilde{\Phi}, \lambda, Q$, (μ_n) and $(x_t, t \in [0,T])$ such that for all $t \in [0,T]$ and for all bounded measurable functions $g : E \to \mathbb{R}$, we have*

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

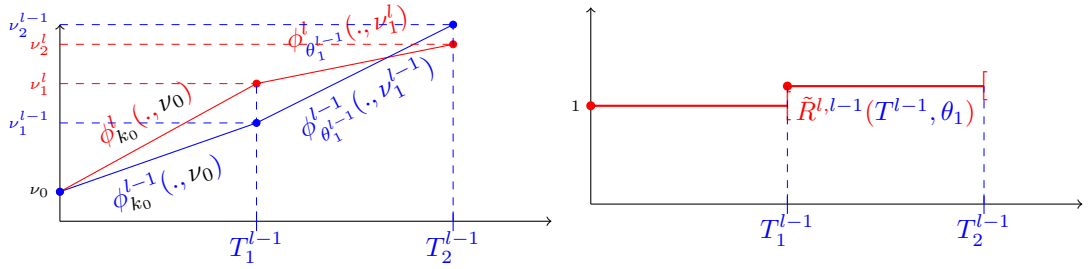


Figure 10 – Illustration of the typical behaviour of the couple of processes $\left((y_t^{(l,l-1)},\tilde{R}_t^{(l,l-1)}),(\overline{x}_t^{h_{l-1}})\right)$ involved at level l of the estimator (28). In this situation, the discrete component and the jump times of $(y_t^{(l,l-1)})$ are those of $(\overline{x}_t^{h_{l-1}})$. Thus, their corresponding continuous components evolves as described on Figure 8a (see the left hand side graphic). Moreover, the corrective process $(\tilde{R}_t^{(l,l-1)})$ is close to 1 (see the right hand side graphic). Consequently, the corresponding L^2 error (see (29)) is of order $O(h_l^2)$.

From Proposition 6.2 we can then decompose $\mathbb{E}[F(\overline{x}_T^{h_L})]$ over the levels as follows

$$\mathbb{E}[F(\overline{x}_T^{h_L})] = \mathbb{E}[F(\overline{x}_T^{h^*})] + \sum_{l=2}^{L} \mathbb{E}[F(y_T^{(l,l-1)})\tilde{R}_T^{(l,l-1)} - F(\overline{x}_T^{h_{l-1}})],$$

where for $l = 2,\dots,L$, the process $(y_t^{(l,l-1)}, t\in[0,T])$ is a PDP whose deterministic motions are given by the approximate flow $\overline{\phi}_\theta$ with time step h_l and whose discrete component jumps at the same times and in the same states as the Euler scheme $(\overline{x}_t)$ with time step h_{l-1} do. Moreover, the process $(\tilde{R}_t^{(l,l-1)}, t\in[0,T])$ is as in Proposition 6.2, that is such that $\mathbb{E}[F(y_T^{(l,l-1)})\tilde{R}_T^{(l,l-1)}] = \mathbb{E}[F(\overline{x}_T^{h_l})]$. We illustrate this situation in Figure 10. Letting $(X_{h_l}^f, X_{h_{l-1}}^c) = (F(y_T^{(l,l-1)})\tilde{R}_T^{(l,l-1)}, F(\overline{x}_T^{h_{l-1}}))$ for $l = 2,\dots,L$ we define a third MLMC estimator also noted $\tilde{Y}^{\mathrm{MLMC}}$ as follows

$$\tilde{Y}^{\mathrm{MLMC}} = \frac{1}{N_1}\sum_{k=1}^{N_1} X_{h^*}^{c,k} + \sum_{l=2}^{L}\frac{1}{N_l}\sum_{k=1}^{N_l}(X_{h_l}^{f,k} - X_{h_{l-1}}^{c,k}). \tag{28}$$

Following the same arguments as in the proof of Theorem 6.3 we are able to prove the following theorem.

Theorem 6.4. *For all $l\in\{2,\dots,L\}$, let $(y_t^{(l,l-1)})$, $(\tilde{R}_t^{(l,l-1)})$ and $(\overline{x}_t^{h_{l-1}})$ be as above. Then, there exists a constant $\tilde{\tilde{V}}_1$ independent of h_l such that*

$$\mathbb{E}[|F(y_T^{(l,l-1)})\tilde{R}_T^{(l,l-1)} - F(\overline{x}_T^{h_{l-1}})|^2] \leqslant \tilde{\tilde{V}}_1 h_l^2. \tag{29}$$

Thus, we end up with $\beta = 2$ in (18) with $(X_{h_l}^f, X_{h_{l-1}}^c) = \left(F(y_T^{(l,l-1)})\tilde{R}_T^{(l,l-1)}, F(\overline{x}_T^{h_{l-1}})\right)$.

Numerical Results

We first illustrate the strong convergence results of Theorems 6.1, 6.3 and 6.4 on a 2-dimensional Morris-Lecar PDMP, then, we numerically compare (on the same model) the classical MC estimator and the MLMC estimators. We chose the mean value of the membrane potential at fixed time as the quantity of interest. More precisely, we are interested in the approximation of $\mathbb{E}[F(x_T)]$ where $(x_t, t \in [0,T])$ denotes a 2-dimensional Morris-Lecar PDMP, $F(\theta,\nu) = \nu$ for $(\theta,\nu) \in E$ and $T > 0$ is fixed.

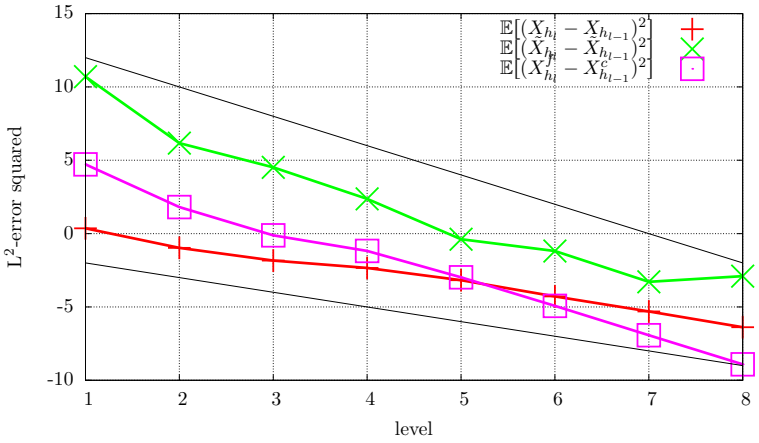


Figure 11 – The plot shows the decay of $\mathbb{E}[(X_{h_l} - X_{h_{l-1}})^2]$, $\mathbb{E}[(\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}})^2]$ and $\mathbb{E}[(X^f_{h_l} - X^c_{h_{l-1}})^2]$ (y-axis, $\log_M$ scale) as a function of l with $h_l = h \times M^{-(l-1)}$, $h = 1$, $M = 4$. For visual guide, we added black solid lines with slopes -1 and -2.

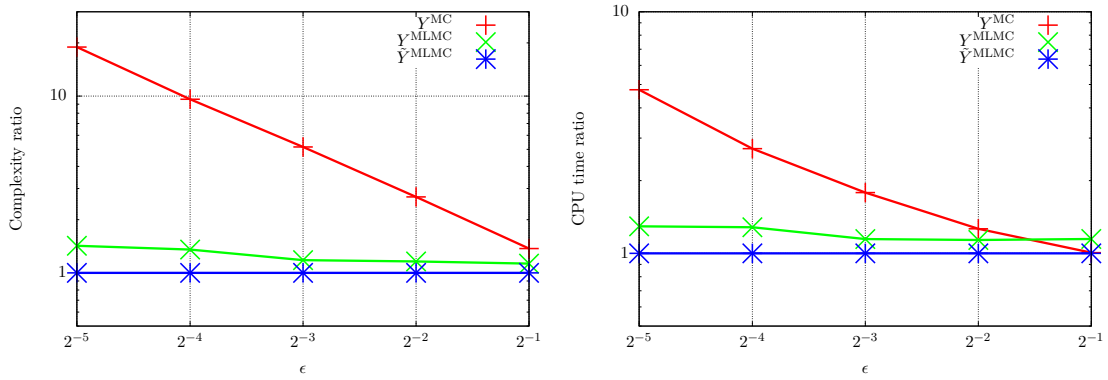
In Figure 11 we represent the L^2 errors $\mathbb{E}[(X_{h_l} - X_{h_{l-1}})^2]$, $\mathbb{E}[(\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}})^2]$ and $\mathbb{E}[(X^f_{h_l} - X^c_{h_{l-1}})^2]$ as a function of the level l where we set

$$
\begin{aligned}
(X_{h_l}, X_{h_{l-1}}) &= \left(F(\overline{x}_T^{h_l}), F(\overline{x}_T^{h_{l-1}})\right), \\
(\tilde{X}_{h_l}, \tilde{X}_{h_{l-1}}) &= \left(F(\underline{\tilde{x}}_T^{h_l})\underline{\tilde{R}}_T^{h_l}, F(\underline{\tilde{x}}_T^{h_{l-1}})\underline{\tilde{R}}_T^{h_{l-1}}\right), \\
(X^f_{h_l}, X^c_{h_{l-1}}) &= \left(F(y_T^{(l,l-1)})\tilde{R}_T^{(l,l-1)}, F(\overline{x}_T^{h_{l-1}})\right).
\end{aligned}
$$

The theoretical order of convergence are respected since the L^2 error $\mathbb{E}[(X_{h_l} - X_{h_{l-1}})^2]$ as a function of the level l behaves like a line with slope -1 and since $\mathbb{E}[(\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}})^2]$ and $\mathbb{E}[(X^f_{h_l} - X^c_{h_{l-1}})^2]$ behave like a line with slope -2. The green curve (representing $\mathbb{E}[(\tilde{X}_{h_l} - \tilde{X}_{h_{l-1}})^2]$) is above the purple one (representing $\mathbb{E}[(X^f_{h_l} - X^c_{h_{l-1}})^2]$) because the variance of $\underline{\tilde{R}}_T^{h_l}$ and $\underline{\tilde{R}}_T^{h_{l-1}}$ is bigger than the one of $\tilde{R}_T^{(l,l-1)}$. Consequently, the variance of the MLMC estimator (26) is bigger than the one of the estimator (28). For that reason we do not consider the estimator (26) in the comparison below.

In Figure 12 we compare the complexity and the CPU-time of the classical MC estimator (24) and the MLMC estimators (25) and (28) as a function of a prescribed

$\epsilon > 0$. We observe that the complexity of the classical MC estimator (24) and those of the MLMC estimators (25) and (28) do indeed behave as a $O(\epsilon^{-3})$, $O(\epsilon^{-2}(\log(\epsilon))^2)$ and $O(\epsilon^{-2})$ respectively as it is theoretically expected. The numerical results suggest that MLMC estimators can be successfully used in the framework of PDMPs.



(a) Ratio of the complexities.

(b) Ratio of the CPU-times.

Figure 12 – The plots (a) and (b) show the complexity and CPU-time ratios w.r.t the complexity and CPU-time of the estimator $\tilde{Y}^{\mathrm{MLMC}}$ (28) as a function of the prescribed ϵ ($\log_2$ scale for the x-axis, log scale for the y-axis).

7 Perspectives

In this section, we present perspectives which are linked to our work that we would like to develop in the future.

Exact simulation of action potentials

The thinning algorithm introduced in chapter 1 provides us with an exact simulation of the trajectories of conductance-based models with explicit flow such as a Hodgkin-Huxley PDMP. As mentioned above, several diffusion approximations have been developed to approximate such models. In [73], [69], the authors compare their respective accuracy and computational efficiency through a numerical analysis. They quantify the errors made in estimating quantities of interest related to action potentials (i.e distribution of inter-spike intervals, first spike latency or spike rate) using approximate algorithms such as the Euler-Maruyama scheme. They all use a model in which the channels are modelled by Markov chains with simplified or approximated transitions as a reference. This does not produce exact samples of action potentials or of spiking times. Since the thinning algorithm is exact, a possible perspective is to use it as a reference algorithm to conduct such a numerical analysis.

Multilevel Monte Carlo

In chapter 2 we consider a PDMP $(x_t, t \in [0,T])$ with no explicit flow, we propose a numerical scheme, $(\overline{x}_t)$, approximating the PDMP and we prove strong and weak convergences which are the building blocks of Monte Carlo simulations. Then, we address the problem of estimating $\mathbb{E}[F(x_T)]$ by MLMC where $F : E \to \mathbb{R}$ is a smooth function. This framework does not include the biologically relevant estimation of inter-spike intervals, first spike latency or spike rate which are modelled as path-dependent functionals of the PDMP. A perspective could be to investigate the MLMC in this setting. More precisely, if we denote by $\mathcal{D}([0,T],E)$ the space of E-valued cad-lag functions defined on $[0,T]$ one could investigate strong and weak convergences of $G\left(\overline{x}_t, t \in [0,T]\right)$ toward $G\left(x_t, t \in [0,T]\right)$ where $G : \mathcal{D}([0,T],E) \to \mathbb{R}$. The order of such convergences could then be used to estimate quantities of the form $\mathbb{E}\left[G\left(x_t, t \in [0,T]\right)\right]$.

Another related work could be to consider a different numerical scheme for the PDMP. More precisely, in this thesis we have chosen to approximate a PDMP with characteristics (ϕ, λ, Q) by a PDP with characteristics $(\overline{\phi}, \lambda, Q)$ where $\overline{\phi}$ denotes the classical Euler scheme associated to ϕ. This choice implies that both processes can be constructed using the same classical iterative construction. Consequently, the discretisation grid of the PDP on $[0,T]$ is random and is formed by the points $\overline{T}_n + kh$ for $n = 0, \dots, \overline{N}_T$ where $k = 0, \dots, \lfloor (\overline{T}_{n+1} \wedge T - \overline{T}_n)/h \rfloor$ and h denotes the time step of $\overline{\phi}$. This differs from the case where the numerical scheme is constructed from the regular fixed grid $(t_n, 0 \leqslant n \leqslant N)$ on $[0,T]$ defined by $t_n = nh$ for $n = 0, \dots N$ where $N > 0$ and $h = T/N$. It would be interesting to construct a scheme for PDMP on that fixed grid, that is, a sequence $(X_n, 0 \leqslant n \leqslant N)$ such that X_n approximates x_{t_n} in order to study strong and weak convergences and to compare the computational efficiency in both settings.

Chapter I

Exact simulation of the jump times of a class of Piecewise Deterministic Markov Processes

Abstract

In this paper, we are interested in the exact simulation of a class of Piecewise Deterministic Markov Processes (PDMP). We show how to perform an efficient thinning algorithm depending on the jump rate bound. For different types of bounds, we compare theoretically the efficiency of the algorithm (measured by the mean ratio between the total number of jump times generated by thinning and the number of selected ones) and we compare numerically the computation times. We use the thinning algorithm on Hodgkin-Huxley models with Markovian ion channels dynamics to illustrate our results.

1.1 Introduction

In many areas it is important to be able to simulate exactly and rapidly trajectories of a stochastic process. This is the case for Monte Carlo methods, statistical estimation, bootstrap. In this article, we are interested in the exact simulation (perfect sampling) of a class of Piecewise Deterministic Markov Processes (PDMP). These processes, introduced by M.H.A. Davis in [21], are based on an increasing sequence of random times in which the processes have a jump and on a deterministic evolution between two successive random times. The law 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.

In this study we consider the class of PDMPs whose flows are known, this means that we explicitly know the solution of each ordinary differential equation associated to each vector field. Explicit flows cover a wide-enough range of interesting applications. For example, we can quote the temporal evolution of the membrane potential and ionic channels in neuroscience (see [67]), the evolution of a food contaminant in the human body in pharmacokinetics (see [6]), the growth of bacteria in biology (see [28]), the data transmission in internet network (see [13]) or the evolution of a chemical network in chemistry (see [1]). For hybrid models in cell biology and gene networks see [9] and [18].

In this paper we focus on the exact simulation of the PDMP inter-jump times. Davis in [22] provides an iterative construction of PDMPs (cf. p. 59) which suggests to simulate the inter-jump times by inversion of their survival function. However he does not specify precisely how to do it numerically. The survival function is expressed using the integral of the jump rate along the flow. When the jump rate along the flow is explicitly integrable and when its integral is explicitly invertible, we can simulate the jump times exactly (by using the jump rate directly), see [26]. When the survival function is not explicitly invertible, several algorithms have been proposed in the literature (cf [70], [77], [27]) but none of them produced exact samples even if the flows are explicit.

We use the thinning method introduced by Lewis and Schedler in [57] to simulate Poisson processes and generalised by Ogata [64] to any point process. The thinning method obviates the need for numerical integration of the jump rate and produces an exact simulation. This method has become classic when the jump rate of the process admits a constant upper bound $\overline{\lambda}$. In this case, it consists in generating the jump times of a (homogeneous) Poisson process with intensity $\overline{\lambda}$ and then to select some of these times by a rejection argument. The times selected are realisations of the jump times. The resulting algorithm is easy to implement. However, it is intuitive that a constant upper bound $\overline{\lambda}$ could lead to many rejections especially if the jump rate presents significant variations thus increasing the computation time too much.

In the sequel we focus on path-adapted upper bounds. We propose different kinds of such bounds for the intensity along the flow, from coarse to path-adapted. Our main contribution is the theoretical study of their respective efficiency. We also provide a numerical study of the computation times for the different bounds. We will consider three types of jump rate bounds:

- the *global bound* (I.8), the coarsest, which is constant (in particular it is independent of the state of the PDMP and of time),
- the *local bound* (I.9), which depends on the post-jump value of the PDMP and which is constant between two successive jump times,
- the *optimal bound* (I.10), the finest, which depends on the post-jump value of the PDMP and also on the time evolution of the process between two successive jump times.

We see at least three interests in the *optimal bound*. The first is that the thinning algorithm with an *optimal bound* applies with weaker hypotheses on the jump rate than with the classical *global bound*. More precisely, the *optimal bound* requires that the jump rate is locally bounded along a given flow whereas the *global bound* requires that it is globally bounded on the state space. The second is that it provides a powerful thinning algorithm. The drawback of this bound is that when the bound becomes very close to the actual jump rate, the computation time may be too long. It is thus necessary to look for a satisfactory balance. We discuss this difficulty on a numerical example. Finally, the *optimal bound* is constructed by following each vector field of the family. This construction is thus natural in the context of switching processes such as PDMPs. For this reason we think that the algorithm studied in this article can be applied to a much larger family of processes such as Hawkes processes, switching stochastic differential equations or switching stochastic partial differential equations with state-dependent intensity.

As an indicator of the efficiency of our thinning algorithm, we choose the mean value of the ratio between the number of selected jump times and the number of generated jump times. We call it rate of acceptance. This indicator is between 0 and 1 and is easily understood, the closer it is to 1 the less we reject points, thus the more efficient the algorithm is. We explicitly express this rate of acceptance in terms of the transition measure of a discrete time Markov chain which carries information about the PDMP but also about all the rejected jump times. In particular this chain is different from the embedded Markov chain classically associated to a PDMP. We also express the rate of acceptance as a function of the ratio between the jump rate of the PDMP and the jump rate bound. This expression enables us to see that the closer the jump rate bound is to the PDMP jump rate the more efficient the algorithm is. Let us note that our rate of acceptance is different from the efficiency defined in [57] or [26] chap. 6 which is the ratio between the mean number of selected jump times and the mean number of generated jump times. However, both coincide in the case of Poisson processes.

As an application, we consider two stochastic versions of the deterministic Hodgkin-Huxley (HH) model (cf. [47]). The two biophysicist Hodgkin and Huxley proposed a four dimensional system of ordinary differential equations (ODE) based on their observations, in order to model the coupled evolution of the membrane potential of a neuron and of specific pores called channels located in the membrane. The circulation of ions through the channels create currents that modify the electric balance and the potential. (HH) model has become classic because it provides a way to express the conductance of the

membrane: the conductance is expressed using the potential dependent probability that specific subunits of channels (called gates) are open given that each gate can be in two states only, either open or closed. However since channels (and consequently gates) are in finite number it is natural to consider a stochastic version of (HH) that we call the *subunit model*. Actually a second stochastic version exists that focuses on the channels themselves. In this case the stochastic model is fourteen dimensional (see Section 1.6.2). We call it the *channel model*. It is also much used in computational Neuroscience since it describes the channel states more in detail. Both stochastic versions are PDMP. When the number of channels goes to infinity the *channel model* converges to a deterministic system of ODE (of dimension 14) such that the variable modelling the membrane potential coincides with the one in (HH) when the initial conditions satisfy a binomial relation [67].

The jump rates of the *subunit* and the *channel* models (which come from the modelling [47]) are complex functions with high variations especially when the membrane potential is in a depolarization phase. Thus, numerical inversion of the distribution function of the inter-jump times can be time consuming. We show in section 1.7.1 how to determine jump rate bounds in such stochastic (HH) models. We use these models to compare numerically the different bounds (I.8), (I.9), (I.10), and thus, to highlight the efficiency of the *optimal bound*. The comparison of the bounds enables us to show that the *optimal bound* speeds up simulation compared to the *global bound* and the *local bound*. We show that the computation time is reduced by 2 in going from the *global bound* to the *local bound* and that it is again reduced by 2 in going from the *local bound* to the *optimal bound*.

To be complete, let us mention some algorithms specific to (HH) models. When the number of channels or gates is high, some authors have used diffusion approximations to improve the computation time (cf [65], [41], [32]), which clearly does not produce exact samples. On the other hand, in many papers, the channels/gates are modelled by Markov chains with simplified or approximated transitions (cf [75], [14], [2], [15], [74]). This does not produce exact samples even if they are called exact in these papers in comparison to the diffusion approximation. For a review of these specific algorithms, see [60].

The paper is organized as follows. In section 2, we give the definition of PDMPs, the assumptions and set the notation used in other sections. In section 3, we present the construction of PDMPs by thinning. In section 4, we introduce the different jump rate bounds. In section 5, we give the theoretical results concerning the comparison of the jump rate bounds and the rate of acceptance without boundaries. In section 6, we introduce the Hodgkin-Huxley models. In section 7 we numerically illustrate the results. Section 1.8 is an appendix in which we compute the rate of acceptance for Poisson processes.

1.2 PDMPs and assumptions

A PDMP is a stochastic process in which the randomness comes from random jump times and post-jump locations [22],[21]. In this paper, we consider that such a process

takes the following general form

$$x_t = (\theta_t, V_t), \qquad \forall t \geqslant 0,$$

where

- $\theta : \mathbb{R}^+ \to K$ is a jump process that characterizes the mode of the system, K is a finite or countable space.
- $V : \mathbb{R}^+ \to D$ is a stochastic process which evolves deterministically between two successive jumps of θ, D is an open subset of $\mathbb{R}^d$.

Let us denote $E = K \times D$ so that $(x_t)_{t\geqslant 0}$ is an E-valued process. We note $(T_n)_{n\geqslant 0}$ the sequence of jump times of the PDMP and $(N_t)_{t\geqslant 0}$ the counting process, $N_t = \sum_{n\geqslant 1} \mathbf{1}_{T_n \leqslant t}$. We assume that for every starting point $x \in E$, $\mathbb{E}_x[N_t] < \infty$ for all $t \geqslant 0$. This assumption implies in particular that $T_n \to \infty$ almost surely.

Such a process is uniquely determined by three characteristics, namely, (ϕ, λ, Q). In the remainder of the paper, we consider that the characteristics verify the following.

Assumptions on the characteristics

- *The deterministic flow* $\phi : \mathbb{R}_+ \times E \to D$ is assumed continuous and induced by a conservative vector field $F : E \to D$, see [21].
- *The jump rate* $\lambda : E \to]0, +\infty[$ is assumed to be a measurable function such that for each $x = (\theta, \nu) \in E$ the function $s \to \lambda(\theta, \phi(s, x))$ is locally integrable. We also assume that λ has a uniformly bounded derivative along the flow and that $\inf_{(s,x)\in\mathbb{R}_+\times E} \lambda(\theta, \phi(s, x)) > 0$.
- *The transition measure* $Q : E \times \mathcal{B}(E) \to [0, 1]$ governs the post-jump location of the process. We assume that

$$Q\Big(x, \{x\}\Big) = 0, \qquad \forall x \in E.$$

For $t \in [T_n, T_{n+1}[$, V takes the following form $V_t = \phi(t - T_n, x_{T_n})$ and the trajectory of the process $(x_t)_{t\geqslant 0}$ is then given by

$$x_t = \sum_{n\geqslant 0} \Big(\theta_{T_n}, \phi(t - T_n, x_{T_n})\Big)\mathbf{1}_{T_n \leqslant t < T_{n+1}}.$$

For notational convenience, we define a vector field $G : E \to E$ such that, for $x \in E$, $G(x) = \begin{pmatrix} 0 \\ F(x) \end{pmatrix}$. We note ψ the flow induced by G. Then the PDMP can be written as follows

$$x_t = \sum_{n\geqslant 0} \psi(t - T_n, x_{T_n})\mathbf{1}_{T_n \leqslant t < T_{n+1}}, \tag{I.1}$$

Denote by ∂D the boundary of D. For all $x \in E$, let

$$t_*(x) = \begin{cases} \inf\{t > 0 : \phi(t,x) \in \partial D\}, \\ +\infty \text{ if no such time exists.} \end{cases}$$

For each $x \in E$, $t_*(x)$ is the time needed to reach the boundary from x. Note that this time is deterministic. In [22], M.H.A Davis shows that there exists a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}_x)$ such that the process $(x_t)_{t\geqslant 0}$ is a Markov process. He also shows that $(x_{T_k})_{k\geqslant 0}$ is a Markov chain with kernel Z such that for all $x \in E$

$$\begin{aligned} Z(x,A) = &\int_0^{t_*(x)} Q\Big(\psi(t,x)), A\Big)\lambda(\psi(t,x))e^{-\int_0^t \lambda(\psi(s,x))ds}dt \\ &+ e^{-\int_0^{t_*(x)} \lambda(\psi(s,x))ds} Q\Big(\psi(t_*(x),x), A\Big). \end{aligned}$$

Let $(S_n) := (T_n - T_{n-1})$ be the sequence of inter-jump times. The intensity of S_{n+1} conditionally on (T_n, x_{T_n}) is $\lambda(\psi(t - T_n, x_{T_n}))$ for $t \geqslant T_n$. We emphasise that the jump rate λ determines the law of the inter-jump times through the survival function S_x defined for all $t \geqslant 0$ and $x \in E$ by

$$S_x(t) = \mathbf{1}_{t<t_*(x)} e^{-\int_0^t \lambda(\psi(s,x))ds}. \tag{I.2}$$

Thus, the jump-times occur either in a deterministic way (when the flow hit a boundary) or in a Poisson-like fashion.

Identity (I.1) implies the representation

$$\lambda(x_t) = \sum_{n\geqslant 0} \lambda(\psi(t - T_n, x_{T_n}))\mathbf{1}_{T_n\leqslant t<T_{n+1}}, \tag{I.3}$$

for the jump rate along the trajectory of (x_t). In the sequel, and depending on the context, we assume that one of the following assumptions is satisfied.

Additional assumptions on the jump rate:

$\mathbf{H^{glo}}$: $\sup_{x\in E} \lambda(x) < \infty$.

$\mathbf{H^{loc}}$: $\forall x \in E, \quad \sup_{s\geqslant 0} \lambda(\psi(s,x)) < \infty$.

$\mathbf{H^{opt}}$: $\forall x \in E, \forall I \subseteq \mathbb{R}_+, \quad \sup_{s\in I} \lambda(\psi(s,x)) < \infty$.

Note that assumption $\mathbf{H^{glo}}$ is verified when λ is bounded. Assumption $\mathbf{H^{loc}}$ is verified when λ is continuous and ψ is bounded. In assumption $\mathbf{H^{opt}}$, when I is compact, λ continuous and ψ continuous is sufficient. Also, $\mathbf{H^{glo}}$ implies ($\mathbf{H^{loc}}$ and $\mathbf{H^{opt}}$) and $\mathbf{H^{loc}}$ implies $\mathbf{H^{opt}}$. Note also that $\mathbf{H^{opt}}$ implies $\mathbf{H^{loc}}$ so that these two assumptions are, in fact, equivalent.

1.3 Simulation of PDMPs and thinning

In [21] and [22] Davis provides an iterative construction of a PDMP and suggests to simulate its inter-jump times using the generalized inverse Ψ_x of (I.2) defined for all $u \in [0,1]$ and $x \in E$ by

$$\Psi_x(u) = \begin{cases} \inf\{t \geqslant 0 : S_x(t) \leqslant u\}, \\ +\infty \text{ if the above set is empty.} \end{cases} \tag{I.4}$$

Thus, the random variable $\Psi_x(U)$ where $U \sim \mathcal{U}([0,1])$, has survival function (I.2) (see, for example, [22] chap. 2 section 2.3 and 2.4). However, the problem of the exact computation of Ψ_x is not obvious. If the function (I.2) is explicitly invertible, the problem is solved. However, in most applications we cannot compute $\Lambda_x(t) \equiv \int_0^t \lambda(\psi(s,x))ds$ explicitly and even less invert it. Moreover the indicator function must be taken into account.

On the other hand, several papers have proposed methods to approximate the inverse (I.4). In [70] and [77], the authors use deterministic numerical methods which essentially consist in solving ODEs to compute (I.4). In [27], the authors use a piecewise linear approximation of (I.2) which can be explicitly invertible. In these three papers, the flows are not assumed explicit, the authors do not consider boundaries that is $t_*(x) = +\infty$ for all $x \in E$. We emphasise that the algorithms proposed in these papers would not produce exact samples even if the flows were explicit.

We show how to simulate exactly the inter-jump times in the presence of a boundary when the flows are explicit. We proceed by thinning. Details on thinning may be found in [57] or [26]. First, let us introduce a modified survival function, $\overline{S}_x$, defined by

$$\overline{S}_x(t) = e^{-\int_0^t \lambda(\psi(s,x))ds}. \tag{I.5}$$

For $x \in E$, (I.5) is the survival function of a random variable with hazard rate $\lambda(\psi(.,x))$ (cf. [26] chap. 6). Let $\tilde{\lambda}$ be a function such that $\lambda(\psi(t,x)) \leqslant \tilde{\lambda}(t,x)$ for all $t \geqslant 0$. Let $(\tilde{T}_k)_{k\geqslant 1}$ be a Poisson process with jump rate $\tilde{\lambda}(.,x)$ independent of $(U_n)_{n\geqslant 1}$ a sequence of iid random variables with uniform distribution on $[0,1]$. We define the random variable τ by

$$\tau = \inf\{k > 0 : U_k \tilde{\lambda}(\tilde{T}_k, x) \leqslant \lambda(\psi(\tilde{T}_k, x))\}.$$

Then, we have the following lemma (see [26]).

Lemma 1.3.1. *The random variable $\tilde{T}_\tau$ has hazard rate $\lambda(\psi(.,x))$.*

Since the flows are assumed explicit, we can compute exactly the ratio $\lambda(\psi(t,x))/\tilde{\lambda}(t,x)$ for all $t \geqslant 0$. We use the following lemma to simulate the Poisson process $(\tilde{T}_k)_{k\geqslant 1}$.

Lemma 1.3.2 (Devroye [26], chap. 6)**.** *Let $(\mathcal{T}_n)_{n\geqslant 0}$ be a Poisson process with jump rate $f(t)$ and let $\mathcal{E}$ be an exponential variable with parameter 1 independent of the Poisson process, then, for $n \geqslant 0$, we have*

$$\mathcal{T}_{n+1} \overset{Law}{=} F^{-1}\left(\mathcal{E} + F(\mathcal{T}_n)\right),$$

where $F(t) = \int_0^t f(s)ds$.

Thus, we need an explicit expression of $\tilde{\Lambda}_x(t) \equiv \int_0^t \tilde{\lambda}(s,x)ds$ and of $\tilde{\Lambda}_x^{-1}(t)$ to simulate exactly the process $(\tilde{T}_k)$ (in section 1.4, we present bounds which verify these conditions). In the case where the bound $\tilde{\lambda}$ is constant, $\tilde{\Lambda}_x$ and its inverse are explicit and easy to compute. However, we can see from the expression of τ that such a bound will lead to many rejections especially when the function $t \to \lambda(\psi(t,x))$ present significant variations. Consequently many evaluations of the intensity along the flow and many generations of pseudo-random variables will be necessary to simulate $\tilde{T}_\tau$. This will potentially increase the computation time compared to a refined bound possibly complicated to integrate and invert but leading to fewer rejections. This balance between coarse and refined bounds will be illustrated both theoretically and numerically in the sequel.

So far, we have shown how to simulate exactly a random variable with survival function (I.5). To simulate exactly the inter-jump times with boundaries (i.e. a random variable with survival function (I.2)) by using the thinning described above, we need the following lemma.

Lemma 1.3.3. *Let $T > 0$ and $g : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-negative, locally integrable function. Define*

$$S(t) \equiv \mathbf{1}_{t<T} e^{-\int_0^t g(s)ds}, \qquad \overline{S}(t) \equiv e^{-\int_0^t g(s)ds}.$$

Let Y ($\overline{Y}$ respectively) be a random variable with survival function S ($\overline{S}$ respectively). Then, we have $Y \overset{law}{=} \overline{Y} \wedge T$.

Proof. A direct computation shows that, $\forall x \geqslant 0$, $\mathbb{P}(Y \geqslant x) = \mathbb{P}(\overline{Y} \wedge T \geqslant x)$. □

Since the flows are explicit, the deterministic time $t_*(x)$ can be computed exactly. Therefore we simulate the first jump of a PDMP starting from x by $\tilde{T}_\tau \wedge t_*(x)$ using lemmas 1.3.1 and 1.3.3 where the variables $(\tilde{T}_k)$ and τ are as above.

We now describe the construction of a PDMP (x_t) by thinning. In the remainder of this section and by analogy with the representation (I.3), we consider a generic bound of λ, namely β, defined by

$$\beta(t, x_t) := \sum_{n \geqslant 0} \tilde{\lambda}(t - T_n, x_{T_n}) \mathbf{1}_{T_n \leqslant t < T_{n+1}}. \tag{I.6}$$

We assume that the function $\tilde{\lambda} : \mathbb{R}_+ \times E \to \mathbb{R}_+$ has the following properties:

- $\forall u \geqslant 0$, $\forall y \in E$,
$$\lambda(\psi(u,y)) \leqslant \tilde{\lambda}(u,y).$$
- $\forall u \geqslant 0$, $\forall y \in E$, the function $\tilde{\Lambda}_y(u) \equiv \int_0^u \tilde{\lambda}(v,y)dv$ is explicitly computable.
- $\forall y \in E$, the inverse of $\tilde{\Lambda}_y$, denoted by $\left(\tilde{\Lambda}_y\right)^{-1}$, is explicitly computable.

The form of the generic bound (I.6) follows from the structure of the PDMP. In practice, one has to specify a bound $\tilde{\lambda}$ to implement the generic algorithm 1 below (three specifications are given in section 1.4). We construct a sample path of the PDMP $(x_t)_{t \geqslant 0}$ with values in E, starting from a fixed initial point $x_0 \in E$ at time 0 as follows.

Let $(\tilde{T}_k^0)_{k\geqslant 0}$ be a Poisson process defined on $[0,+\infty[$ with jump rate $\tilde{\lambda}(t,x_0)$ for $t\geqslant 0$, and,

$$\tau_1 = \inf\{k>0 : U_k^{(1)}\tilde{\lambda}(\tilde{T}_k^0,x_0)\leqslant \lambda(\psi(\tilde{T}_k^0,x_0))\},$$

where $(U_n^{(1)})_{n\geqslant 1}$ is a sequence of independent random variables with uniform distribution on $[0,1]$, independent of $(\tilde{T}_k^0)_{k\geqslant 0}$. By lemma 1.3.3, the first jump time $T_1 = S_1$ of the PDMP is the minimum between the first jump time of a non-homogeneous Poisson process defined on $[0,+\infty[$ with jump rate $\lambda(\psi(t,x_0))$ and $t_*(x_0)$. Thus, $T_1 = \tilde{T}^0_{\tau_1}\wedge t_*(x_0)$. On $[0,T_1[$ the PDMP evolves as follows

$$x_t = \psi(t,x_0).$$

The random variable x_{T_1} has distribution $Q\Big(\psi(T_1,x_0),.\Big)$. Note that conditionally on T_1 the process $(\tilde{T}_k^0)_{k\geqslant 1}$ is a Poisson process on $[0,T_1[$ with jump rate $\tilde{\lambda}(t,x_0)-\lambda(\psi(t,x_0))$, see [26] chap.6.

Suppose we have simulated T_i, then, conditionally on (T_i,x_{T_i}), the PDMP (x_t) restarts from x_{T_i} at time T_i independently from the past. Let $(\tilde{T}_k^i)_{k\geqslant 0}$ be a Poisson process on $[0,+\infty[$ with jump rate $\tilde{\lambda}(t-T_i,x_{T_i})$ for $t\geqslant T_i$, and,

$$\tau_{i+1} = \inf\{k>0 : U_k^{(i+1)}\tilde{\lambda}(\tilde{T}_k^i,x_{T_i}))\leqslant \lambda(\psi(\tilde{T}_k^i,x_{T_i}))\},$$

where $(U_n^{(i+1)})_{n\geqslant 1}$ is a sequence of independent uniform random variables, independent of $(\tilde{T}_k^i)_{k\geqslant 0}$ and x_{T_i}. By lemma 1.3.3 and the thinning procedure, we have $T_{i+1} = T_i + \tilde{T}^i_{\tau_{i+1}}\wedge t_*(x_{T_i})$. On $[T_i,T_{i+1}[$ the process evolves as follows

$$x_t = \psi(t-T_i,x_{T_i}).$$

The random variable $x_{T_{i+1}}$ has distribution $Q\Big(\psi(S_{i+1},x_{T_i}),.\Big)$. Note that, conditionally on (T_i,x_{T_i},T_{i+1}), the process $(T_i+\tilde{T}_k^i)_{k\geqslant 1}$ is a Poisson process on $[T_i,T_{i+1}[$ with jump rate $\tilde{\lambda}(t-T_i,x_{T_i})-\lambda(\psi(t-T_i,x_{T_i}))$.

Conditionally on $(T_1,x_{T_1},\ldots,T_i,x_{T_i},T_{i+1})$, the points in $[T_i,T_{i+1}[$ obtained from the Poisson process $(T_i+\tilde{T}_k^i)_{k\geqslant 1}$ are independent of the points in $[T_{j-1},T_j[$ obtained from the Poisson process $(T_j+\tilde{T}_k^j)_{k\geqslant 1}$ for $j=1,\ldots,i$. The construction above provides a generic algorithm associated to the bound (I.6) to simulate trajectories of PDMPs (see Algorithm 1 below).

To conclude this section, consider the case of PDMPs without boundaries. In this case, the construction above provides a point process, namely,

$$T_0<\tilde{T}_1^0<\ldots<\tilde{T}^0_{\tau_1-1}<T_1<T_1+\tilde{T}_1^1<\ldots<T_1+\tilde{T}^1_{\tau_2-1}<T_2<T_2+\tilde{T}_1^2<\ldots \quad \text{(I.7)}$$

Notation 1.3.1. *In the sequel, the process defined by* (I.7) *is noted* $(\tilde{T}_k)_{k\geqslant 0}$ *and the associated counting process is noted* $(\tilde{N}_t)_{t\geqslant 0}$. *We also denote by* $(\overline{T}_k)_{k\geqslant 0}$ *the process formed by all the rejected points (i.e the process* $(\tilde{T}_k)_{k\geqslant 0}$ *without the jump times* $(T_k)_{k\geqslant 0}$*) and* $(\overline{N}_t)_{t\geqslant 0}$ *the associated counting process.*

Algorithm 1 Simulation of a trajectory of (x_t) on $[0, T]$.

Require: Fix the initial condition $x_0 = (\theta_0, V_0)$, set a jump counter $n = 0$ and fix the initial time $T_n = 0$. Set also an auxiliary jump counter $k = 0$ and an auxiliary variable $\tilde{T}_k = T_n$.
repeat
  repeat
    $k \leftarrow k + 1$.
    Simulate $U_{2k-1} \sim \mathcal{U}(]0, 1[)$.
    Set $E_k = -\log(U_{2k-1})$.
    Set $\tilde{T}_k = \left(\tilde{\Lambda}_{x_{T_n}}\right)^{-1}\left(E_k + \tilde{\Lambda}_{x_{T_n}}(\tilde{T}_{k-1})\right)$.
    Simulate $U_{2k} \sim \mathcal{U}(]0, 1[)$.
  until $U_{2k}\tilde{\lambda}(\tilde{T}_k, x_{T_n}) \leqslant \lambda(\psi(\tilde{T}_k, x_{T_n}))$
  Set $T_{n+1} = T_n + \tilde{T}_k \wedge t_*(x_{T_n})$.
  Set $\tilde{T}_k = 0$.
  if $T_{n+1} \leqslant T$ **then**
    $V_t = \phi(t - T_n, x_{T_n})$ for $t \in [T_n, T_{n+1}[$.
    Simulate a post-jump value $x_{T_{n+1}}$ according to the Markovian kernel $Q\left(\psi(S_{n+1}, x_{T_n}), .\right)$.
  else
    $V_t = \phi(t - T_n, x_{T_n})$ for $t \in [T_n, T[$.
  end if
  $n \leftarrow n + 1$.
until $T_n \geqslant T$

The sequence $(\tilde{T}_k)_{k\geqslant 0}$ contains both generated and selected points and the subsequence noted $(T_k)_{k\geqslant 0}$ such that for $k \geqslant 1$, $T_k = \sum_{l=1}^{k} \tilde{T}_{\tau_l}^{l-1}$ defines the jump times of the PDMP. Thus, we have constructed the jump times of the PDMP by thinning the process $(\tilde{T}_k)_{k\geqslant 0}$ with non-constant (and random) probabilities (p_k) such that $p_k = \lambda(x_{\tilde{T}_k-})/\tilde{\lambda}(x_{\tilde{T}_k-}, \tilde{T}_k)$ is the probability to accept $\tilde{T}_k$. Note that the process $(\tilde{T}_k)_{k\geqslant 0}$ is composed by pieces of independent Poisson processes $(\tilde{T}_k^0), (\tilde{T}_k^1), \ldots, (\tilde{T}_k^i), \ldots$.

1.4 Jump rate bounds

In this section we introduce the different jump rate bounds considered in this paper, namely, the *optimal bound*, the *local bound* and the *global bound*. The *optimal bound* is particularly efficient in term of reject because it is as close as we want to the jump rate.

1.4.1 The global bound

We define the *global bound* by

$$\tilde{\lambda}^{\text{glo}}(u, y) := \sup_{x\in E} \lambda(x), \quad \forall u \geqslant 0, \forall y \in E. \tag{I.8}$$

By definition, this bound is constant and does not depend on the state of the PDMP nor on time, we will denote it by $\tilde{\lambda}^{\text{glo}}$. This bound is probably the most used and has the advantage to lead to an easy implementation. Indeed, to simulate the jump times of the PDMP we simulate a homogeneous Poisson process with jump rate $\tilde{\lambda}^{\text{glo}}$ disregarding the state of the PDMP. For $u \geqslant 0$ and $y \in E$, the integrated jump rate bound is given by $\tilde{\Lambda}_y^{\text{glo}}(u) = \tilde{\lambda}^{\text{glo}} u$ and the inverse is given by $(\tilde{\Lambda}_y^{\text{glo}})^{-1}(u) = u/\tilde{\lambda}^{\text{glo}}$.

1.4.2 The local bound

We define the *local bound* by

$$\tilde{\lambda}^{\text{loc}}(u, y) := \sup_{s \geqslant 0} \lambda\left(\psi(s, y)\right), \quad \forall u \geqslant 0, \forall y \in E. \tag{I.9}$$

By definition, this bound is constant between two successive jump times and has the advantage of being adapted to the state of the PDMP right-after a jump. We will denote it by $\tilde{\lambda}^{\text{loc}}(y)$. To each jump time of the PDMP corresponds a homogeneous Poisson process whose intensity depends on the state of the PDMP at the jump time. For $u \geqslant 0$ and $y \in E$, the integrated jump rate bound is $\tilde{\Lambda}_y^{\text{loc}}(u) = \Big(\sup_{s \geqslant 0} \lambda(\psi(s, y))\Big) u$ and the inverse is given by $(\tilde{\Lambda}_y^{\text{loc}})^{-1}(u) = \Big(u / \sup_{s \geqslant 0} \lambda(\psi(s, y))\Big)$.

1.4.3 The optimal bound

Let P be a finite or a countable space, for $i = 1, \ldots, \text{card}(P)$, we note p_i its elements. Let us denote by $(\mathcal{P}_k)_{k \in P}$ a partition of $[0, +\infty[$ formed by intervals. Thus, there exists $a_{p_{i-1}}$, $a_{p_i} \in \mathbb{R}$ such that $\mathcal{P}_{p_i} = [a_{p_{i-1}}, a_{p_i}[$ with $a_{p_0} := 0$. We assume that $a_{p_{i-1}} < a_{p_i}$ for $i = 1, \ldots, \text{card}(P)$. The partition $(\mathcal{P}_k)_{k \in P}$ can contain at most one element whose Lebesgue measure is infinite. Thus, if such an element exits, it is the last of the partition, i.e, $\left|\mathcal{P}_{p_{\text{Card}(P)}}\right| = \infty$, and, $\left|\mathcal{P}_{p_i}\right| < \infty$ for $i = 1, \ldots, \text{Card}(P) - 1$. We define the *optimal bound* by

$$\tilde{\lambda}^{\text{opt}}(u, y) := \sum_{k \in P} \sup_{s \in \mathcal{P}_k} \lambda(\psi(s, y)) \mathbf{1}_{\mathcal{P}_k}(u), \quad \forall u \geqslant 0, \forall y \in E. \tag{I.10}$$

By definition, this bound is piecewise constant between two successive jump times, thus it is adapted to the state of the PDMP right-after a jump but also to the evolution in time of the jump rate. To each jump time of the PDMP corresponds a non-homogeneous Poisson process whose intensity depends on the state of the PDMP at the jump time and on the flow starting from this state. For $u \geqslant 0$ and $y \in E$, the integrated jump rate bound is given by

$$\tilde{\Lambda}_y^{\text{opt}}(u) = \sum_{k \in P} \sup_{s \in \mathcal{P}_k} \lambda(\psi(s, y)) \Big|\mathcal{P}_k \cap [0, u]\Big|,$$

where $\left|\mathcal{P}_k \cap [0,u]\right|$ represents the length (Lebesgue measure) of $\mathcal{P}_k \cap [0,u]$. The inverse, $\left(\tilde{\Lambda}_y^{\text{opt}}\right)^{-1}$, is given by

$$\left(\tilde{\Lambda}_y^{\text{opt}}\right)^{-1}(u) = \sum_{i=1}^{\text{card}(P)} \Big(\frac{u - \sum_{k=1}^{i-1} \sup_{s\in\mathcal{P}_{p_k}} \lambda(\psi(s,y)) \left|\mathcal{P}_{p_k}\right|}{\sup_{s\in\mathcal{P}_{p_i}} \lambda(\psi(s,y))} + \sum_{l=1}^{i-1} \left|\mathcal{P}_{p_l}\right| \Big) \mathbf{1}_{[\kappa_{p_{i-1}},\kappa_{p_i}[}(u),$$

where $\kappa_{p_i} = \sum_{k=1}^{i} \sup_{s\in\mathcal{P}_k} \lambda(\psi(s,y)) \left|\mathcal{P}_{p_k}\right|$ for $i = 1,\dots,\text{Card}(P)$. By convention, we set $\sum_{l=1}^{0} \left|\mathcal{P}_{p_l}^{T_n}\right| = 0$ and $\kappa_{p_0} = 0$.

As an example of partition let $P = \mathbb{N}$, $\epsilon > 0$ and define $\mathcal{P}_k^\epsilon \equiv [k\epsilon, (k+1)\epsilon[$. Note that this partition is infinite. This is not a numerical problem since the time horizon is finite (λ is assumed positive, see section 1.2). Now, consider the *optimal bound* with this partition (see section 1.7.2 for a numerical study of this bound). We emphasise that the smaller the parameter ϵ is the fewer rejected points are. This point is theoretically illustrated by proposition 1.5.5. However, when ϵ is too small, possibly many iterations are required to compute $\tilde{\Lambda}(.)$ and $\tilde{\Lambda}^{-1}(.)$, this will increase the computation time. We will see in section 1.7.2 that taking ϵ of order $\max_n(T_{n+1} - T_n)$ leads to the optimal computation time.

Remark 1.4.1. *The three hypotheses* $\boldsymbol{H^{glo}}$, $\boldsymbol{H^{loc}}$, $\boldsymbol{H^{opt}}$ *(section 1.2) ensure that the functions* $\tilde{\lambda}^z$, $\tilde{\Lambda}^z$ *and* $\left(\tilde{\Lambda}^z\right)^{-1}$ *are well defined,* $z \in \{glo, loc, opt\}$. *Moreover, the numerical tractability of the suprema involved in the different bounds follows from the characteristics of the PDMP to simulate. We refer to section 1.7.1 for explicit formulas of these bounds for two stochastic HH models. Also, hypotheses* $\boldsymbol{H^{loc}}$ *and* $\boldsymbol{H^{opt}}$ *allow to implement the algorithm 1 when the jump rate is not globally bounded.*

Remark 1.4.2. *For the three jump rate bounds, the simulation is exact. In particular, for all finite or countable* P*, that is, for any partitions of* $[0,+\infty[$*, the simulation remains exact.*

Remark 1.4.3. *The choice of the bound depends on the PDMP we want to simulate. If the jump rate does not vary very much in time, the local bound or the global constant bound can be chosen but if the jump rate presents high variations in a small time interval, the optimal bound is preferable in term of computation time.*

Remark 1.4.4. *The local bound and the optimal bound along the trajectory of the PDMP* (x_t) *namely,* β^{loc} *and* β^{opt}*, are stochastic processes.*

1.5 Efficiency of the thinning algorithm

In this section, we do not consider boundaries. We compare the efficiency of the thinning algorithm in term of reject for the different bounds. The number of points needed to simulate one inter-jump time of a PDMP in state $x \in E$ is given by

$$\tau^z(x) = \inf\{k > 0 : U_k \tilde{\lambda}^z(\tilde{T}_k, x) \leqslant \lambda\left(\psi(\tilde{T}_k, x)\right)\},$$

for $z \in \{\text{glo}, \text{loc}, \text{opt}\}$, where $(U_n)_{n\geqslant 1}$ is a sequence of iid random variables with uniform distribution on $[0,1]$ independent of a Poisson process $(\tilde{T}_k)_{k\geqslant 1}$ with jump rate $\tilde{\lambda}^z(t,x)$.

The randomness of a PDMP (x_t) is contained in the associated jump process (η_t) defined by

$$\eta_t = x_{T_n} \qquad T_n \leqslant t < T_{n+1}. \tag{I.11}$$

Because $T_n = \inf\{t > T_{n-1} : \eta_{t-} \neq \eta_t\}$, the knowledge of $(\eta_t)_{t\geqslant 0}$ implies the knowledge of $(T_n)_{n\geqslant 0}$.

1.5.1 Comparison of the mean number of total jump times

In this section, the variables $\tau^z(x)$ for $z \in \{\text{glo}, \text{loc}, \text{opt}\}$ and $x \in E$ are called local reject. In proposition 1.5.1, we show that the best local reject is obtained with the *optimal bound*. The smaller the local reject the fewer pseudo-random variables have to be simulated. Thus, the computation time using the *optimal bound* is expected to be smaller than with the two other bounds.

Proposition 1.5.1. *For all $x \in E$, we have*

$$\mathbb{E}[\tau^{opt}(x)] \leqslant \mathbb{E}[\tau^{loc}(x)] \leqslant \mathbb{E}[\tau^{glo}(x)].$$

Proof. Let $x \in E$. From the definitions of the three bounds (I.8), (I.9) and (I.10), we have

$$\tilde{\lambda}^{\text{opt}}(t,x) \leqslant \tilde{\lambda}^{\text{loc}}(x) \leqslant \tilde{\lambda}^{\text{glo}}, \quad \forall t \geqslant 0.$$

Recall that, for fixed $x \in E$, $\tau^{\text{opt}}(x), \tau^{\text{loc}}(x)$ and $\tau^{\text{glo}}(x)$ denote the number of points (or iterations) needed to simulate one inter-jump time of a PDMP (which is in state x) by thinning using the upper bounds $\tilde{\lambda}^{\text{opt}}$, $\tilde{\lambda}^{\text{loc}}$ and $\tilde{\lambda}^{\text{glo}}$ respectively. Thus, we can use Theorem 2.2 (chapter 6) of [26] which gives a formula for the mean number of iterations in a thinning algorithm for random variables characterised by a hazard rate. We obtain the following equalities.

$$\begin{aligned}
\mathbb{E}[\tau^{\text{glo}}(x)] &= \tilde{\lambda}^{\text{glo}} \int_0^{+\infty} e^{-\int_0^t \lambda(\psi(s,x))ds} dt, \\
\mathbb{E}[\tau^{\text{loc}}(x)] &= \tilde{\lambda}^{\text{loc}}(x) \int_0^{+\infty} e^{-\int_0^t \lambda(\psi(s,x))ds} dt, \\
\mathbb{E}[\tau^{\text{opt}}(x)] &= \int_0^{+\infty} \tilde{\lambda}^{\text{opt}}(t,x) e^{-\int_0^t \lambda(\psi(s,x))ds} dt.
\end{aligned}$$

Since $\inf_{x\in E} \inf_{s\geqslant 0} \lambda(\psi(s,x)) > 0$, we have $0 < \int_0^{+\infty} e^{-\int_0^t \lambda(\psi(s,x))ds} dt < \infty$ and the conclusion follows. □

From Proposition 1.5.1, we deduce that $\mathbb{E}[\tilde{N}_t^{\text{opt}}] \leqslant \mathbb{E}[\tilde{N}_t^{\text{loc}}] \leqslant \mathbb{E}[\tilde{N}_t^{\text{glo}}]$ where $\tilde{N}_t^{\text{opt}}$, $\tilde{N}_t^{\text{loc}}$ and $\tilde{N}_t^{\text{glo}}$ are counting processes with stochastic intensity β^{opt}, β^{loc} and β^{glo} respectively.

1.5.2 Rate of acceptance

We are now interested in the rate of acceptance, that is, the mean proportion of selected points in an interval of the form $[0,t]$ for $t>0$. Let (N_t) be the counting process of the PDMP and $(\tilde{N}_t)$ the counting process with generic jump rate (I.6). In proposition 1.5.2 we give an explicit formula for the rate of acceptance defined as $\mathbb{E}[N_t/\tilde{N}_t|\tilde{N}_t \geqslant 1]$. This formula is valid for the three bounds $\tilde{\lambda}^{\text{opt}}$, λ^{loc} and λ^{glo} introduced in section 1.4. Note that for $k \geqslant 1$,

$$p_k = \lambda(\psi(\tilde{T}_k - T_{n_k}, x_{T_{n_k}}))/\tilde{\lambda}(\tilde{T}_k - T_{n_k}, x_{T_{n_k}}))$$

is the probability to accept the point $\tilde{T}_k$ where T_{n_k} denotes the last selected jump-time before $\tilde{T}_k$ and $(\tilde{T}_n)$ is defined by (I.7). Let $J:\mathbb{R}_+ \to \mathbb{R}_+$ be the process defined by $J_t = \sum_{k\geqslant 0}(t-T_k)\mathbf{1}_{T_k \leqslant t < T_{k+1}}$. Thus, for $t \geqslant 0$, J_t gives the age of the last selected jump-time before t. Then, for $k \geqslant 1$, we can write the probabilities p_k as follows

$$p_k = \lambda(\psi(J_{\tilde{T}_{k-1}} + \tilde{S}_k, \eta_{\tilde{T}_{k-1}}))/\tilde{\lambda}(J_{\tilde{T}_{k-1}} + \tilde{S}_k, \eta_{\tilde{T}_{k-1}}),$$

where $\tilde{S}_k = \tilde{T}_k - \tilde{T}_{k-1}$ and (η_t) is defined by (I.11). The process $(\tilde{S}_k, \tilde{X}_k)_{k\geqslant 0}$ where $\tilde{X}_k = (J_{\tilde{T}_k}, \eta_{\tilde{T}_k})$ defines a Markov chain on $\mathbb{R}_+ \times \tilde{E}$ where $\tilde{E} = \mathbb{R}_+ \times E$ with Markov kernel M defined by

$$M(j_0, x_0; ds, dj, dx) = \alpha(j_0, x_0; ds)\tilde{Q}(s, j_0, x_0; dj, dx),$$

where,

$$\alpha(j_0, x_0; ds) = \tilde{\lambda}(j_0+s, x_0)e^{-\int_0^s \tilde{\lambda}(j_0+z, x_0)dz}ds,$$

and,

$$\tilde{Q}(s, j_0, x_0; dj, dx) = \Big(1 - \frac{\lambda(\psi(j_0+s, x_0))}{\tilde{\lambda}(j_0+s, x_0), t)}\Big)\delta_{j_0+s}(dj)\delta_{x_0}(dx) + \frac{\lambda(\psi(j_0+s, x_0))}{\tilde{\lambda}(j_0+s, x_0)} Q\Big(\psi(j_0+s, x_0), dx\Big)\delta_0(dj).$$

The Markov kernel M should be understood as follows. Given that $\tilde{X}_{k_0} = (j_0, x_0)$ for some $k_0 \in \mathbb{N}$ (that is, at time $\tilde{T}_{k_0}$, the age of the last accepted jump time is $J_{\tilde{T}_{k_0}} = j_0$ and the state of the PDMP at the last accepted jump time is $\eta_{\tilde{T}_{k_0}} = x_0$), the next proposed inter-jump time $\tilde{S}_{k_0+1}$ has a density given by α. Then, conditionally on $\tilde{S}_{k_0+1} = s$, we accept or not this proposed inter-jump time according to the kernel $\tilde{Q}$. More precisely, we reject it with probability $1 - \lambda(.)/\tilde{\lambda}(.)$, in this case the age of the last accepted jump time is updated, $J_{\tilde{T}_{k_0+1}} = j_0 + s$, and the state of the PDMP is not updated, $\eta_{\tilde{T}_{k_0+1}} = x_0$. If we accept it (with probability $\lambda(.)/\tilde{\lambda}(.)$), then the age of the last accepted jump time is set to 0 and the state of the PDMP is updated according to the kernel Q.

Proposition 1.5.2. *Let $(N_t)_{t\geqslant 0}$ be the counting process of the PDMP $(x_t)_{t\geqslant 0}$, $(\tilde{N}_t)_{t\geqslant 0}$ be the counting process with jump times $(\tilde{T}_n)_{n\geqslant 0}$ and M be the kernel of the Markov chain $(\tilde{S}_k, J_{\tilde{T}_k}, \eta_{\tilde{T}_k})_{k\geqslant 0}$, we have*

$$\mathbb{E}\Big[\frac{N_t}{\tilde{N}_t}|\tilde{N}_t \geqslant 1\Big] = \frac{1}{\mathbb{P}(\tilde{N}_t \geqslant 1)} \sum_{n\geqslant 1} \frac{1}{n} \int_{(\mathbb{R}_+\times\tilde{E})^n} \left[\sum_{k=1}^{n} \frac{\lambda(\psi(j_{k-1}+s_k, x_{k-1}))}{\tilde{\lambda}(j_{k-1}+s_k, x_{k-1})}\right] e^{-\int_0^{t-t_n} \tilde{\lambda}(j_n+z, x_n)dz} \times \mathbf{1}_{t\geqslant t_n} \mu(dx_0) M(0, x_0; ds_1, dj_1, dx_1) \dots M(j_{n-1}, x_{n-1}; ds_n, dj_n, dx_n),$$

where $t_n := \sum_{i=1}^{n} s_i$, μ is the law of $\eta_{\tilde{T}_0}$ and the integration variables $s_\cdot$ and $(j_\cdot, x_\cdot)$ belong to $\mathbb{R}_+$ and $\tilde{E}$ respectively.

Proof. We provide a proof in two steps. First, we establish that, with an appropriate conditioning, the conditional law of N_t is the conditional law of a sum of independent Bernoulli random variables with different parameters. Then, we use this property as well as the kernel M to compute the rate of acceptance.

Let $n \geqslant 1$ and let us define n independent Bernoulli random variables X_i with parameters p_i such that

$$p_i = \frac{\lambda(\psi(J_{\tilde{T}_{i-1}} + \tilde{S}_i, \eta_{\tilde{T}_{i-1}}))}{\tilde{\lambda}(J_{\tilde{T}_{i-1}} + \tilde{S}_i, \eta_{\tilde{T}_{i-1}})}.$$

Let $X = \sum_{i=1}^{n} X_i$ and $A_{t,n} = \{\tilde{N}_t = n, p_1, \dots, p_n\}$. By noting that, for $0 \leqslant k \leqslant n$, we have

$$\{N_t = k | A_{t,n}\} = \bigcup_{1\leqslant i_1<\dots<i_k\leqslant n} \Big[\bigcap_{i\in I_k^n} \{U_i \leqslant p_i\} \bigcap_{i\in \overline{I}_k^n} \{U_i > p_i\}\Big] = \{X = k | p_1, \dots, p_n\},$$

where $I_k^n = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$, $\overline{I}_k^n$ is the complementary of I_k^n in $\{1, \dots, n\}$ and (U_i) are independent random variables uniformly distributed in $[0, 1]$ and independent of (p_i), we deduce that

$$\mathcal{L}(N_t | A_{t,n}) = \mathcal{L}(X | p_1, \dots, p_n).$$

In particular, $\mathbb{E}[N_t | A_{t,n}] = \mathbb{E}[X | p_1, \dots, p_n] = \sum_{i=1}^{n} p_i$. Thus, one can write

$$\begin{aligned}
\mathbb{E}\left[\frac{N_t}{\tilde{N}_t}|\tilde{N}_t \geqslant 1\right] &= \frac{1}{\mathbb{P}(\tilde{N}_t \geqslant 1)} \sum_{n\geqslant 1} \frac{1}{n} \mathbb{E}\Big[N_t \mathbf{1}_{\tilde{N}_t = n}\Big] \\
&= \frac{1}{\mathbb{P}(\tilde{N}_t \geqslant 1)} \sum_{n\geqslant 1} \frac{1}{n} \mathbb{E}\left[\mathbb{E}[N_t | A_{t,n}] | \tilde{N}_t = n\right] \mathbb{P}(\tilde{N}_t = n) \\
&= \frac{1}{\mathbb{P}(\tilde{N}_t \geqslant 1)} \sum_{n\geqslant 1} \frac{1}{n} \mathbb{E}\Big[\sum_{i=1}^{n} p_i \mathbf{1}_{\tilde{N}_t = n}\Big] \\
&= \frac{1}{\mathbb{P}(\tilde{N}_t \geqslant 1)} \sum_{n\geqslant 1} \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^{n} p_i \mathbf{1}_{t-\tilde{T}_n\geqslant 0} \mathbb{E}\left[\mathbf{1}_{\tilde{S}_{n+1}\geqslant t-\tilde{T}_n} | \eta_{\tilde{T}_0}, \tilde{S}_1, \dots, \tilde{S}_n, J_{\tilde{T}_n}, \eta_{\tilde{T}_n}\right]\right].
\end{aligned}$$

Conditionally to $(\eta_{\tilde{T}_0}, \tilde{S}_1, J_{\tilde{T}_1}, \eta_{\tilde{T}_1}, \ldots, \tilde{S}_n, J_{\tilde{T}_n}, \eta_{\tilde{T}_n})$, the random variable $\tilde{S}_{n+1}$ is a hazard law with rate $\tilde{\lambda}(J_{\tilde{T}_n}+t, \eta_{\tilde{T}_n})$ for $t \geqslant 0$. Thus,

$$\begin{aligned}\mathbb{E}\Big[\frac{N_t}{\tilde{N}_t}|\tilde{N}_t \geqslant 1\Big] &= \frac{1}{\mathbb{P}(\tilde{N}_t \geqslant 1)} \sum_{n\geqslant 1} \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} p_i e^{-\int_0^{t-\tilde{T}_n} \tilde{\lambda}(J_{\tilde{T}_n}+u,\eta_{\tilde{T}_n})du} \mathbf{1}_{t-\tilde{T}_n\geqslant 0}\Big] \\ &= \frac{1}{\mathbb{P}(\tilde{N}_t \geqslant 1)} \sum_{n\geqslant 1} \frac{1}{n}\mathbb{E}\Big[f(\eta_{\tilde{T}_0}, \tilde{S}_1, J_{\tilde{T}_1}, \eta_{\tilde{T}_1}, \ldots, \tilde{S}_n, J_{\tilde{T}_n}, \eta_{\tilde{T}_n})\Big],\end{aligned}$$

where,

$$\begin{aligned}&f(x_0, s_1, j_1, x_1, \ldots, s_n, j_n, x_n) \\ &= e^{-\int_0^{t-\sum_{i=1}^n s_n} \tilde{\lambda}(j_n+u,x_n)du} \mathbf{1}_{t-\sum_{i=1}^n s_n\geqslant 0} \sum_{i=1}^{n} \frac{\lambda(\psi(j_{i-1}+s_i, x_{i-1}))}{\tilde{\lambda}(j_{i-1}+s_i, x_{i-1})}.\end{aligned}$$

Since $(\tilde{S}_k, J_{\tilde{T}_k}, \eta_{\tilde{T}_k})_{k\geqslant 0}$ is a Markov chain with kernel M, we obtain

$$\begin{aligned}\mathbb{E}\left[f(\eta_{\tilde{T}_0}, \tilde{S}_1, \ldots, \tilde{S}_n, J_{\tilde{T}_n}, \eta_{\tilde{T}_n})\right] = \int_{\left(\mathbb{R}_+\times\tilde{E}\right)^n} & f(x_0, s_1, j_1, x_1 \ldots, s_n, j_n, x_n) \\ & \mu(dx_0)M(0, x_0; ds_1, dj_1, dx_1)\ldots M(j_{n-1}, x_{n-1}; ds_n, dj_n, dx_n),\end{aligned}$$

where μ is the law of $\eta_{\tilde{T}_0}$. Thus, we have the result. □

When $\tilde{\lambda}$ is close to λ, the rate of acceptance is expected to be close to 1. As an example, consider the case of two Poisson processes (N_t) and $(\tilde{N}_t)$ with intensity $\lambda(t)$ and $\tilde{\lambda}(t)$ respectively such that $\tilde{\lambda}(t) = \tilde{\lambda}$ for all $t \geqslant 0$. Thus, for $n \geqslant 1$, $\tilde{S}_1, \ldots, \tilde{S}_n$ are independent exponential variables with parameter $\tilde{\lambda}$. Let us also consider that $\lambda(t) \simeq \tilde{\lambda}$ for $t \geqslant 0$. In this case, the rate of acceptance is

$$\mathbb{E}\Big[\frac{N_t}{\tilde{N}_t}|\tilde{N}_t \geqslant 1\Big] \simeq \frac{1}{1-e^{-\tilde{\lambda}t}} \sum_{n\geqslant 1} \int_{(\mathbb{R}_+)^n} e^{-\tilde{\lambda}\Big(t-(s_1+\ldots+s_n)\Big)} \mathbf{1}_{t\geqslant s_1+\ldots+s_n} \alpha(ds_1), \ldots, \alpha(ds_n),$$

where $\alpha(ds) = \tilde{\lambda}e^{-\tilde{\lambda}s}ds$. Since, $\tilde{T}_n = \tilde{S}_1 + \ldots + \tilde{S}_n$ is gamma distributed with parameters n and $\tilde{\lambda}$, we have

$$\mathbb{E}\Big[\frac{N_t}{\tilde{N}_t}|\tilde{N}_t \geqslant 1\Big] \simeq \frac{1}{1-e^{-\tilde{\lambda}t}} \sum_{n\geqslant 1} \mathbb{E}[e^{-\tilde{\lambda}(t-\tilde{T}_n)}\mathbf{1}_{t\geqslant \tilde{T}_n}] \simeq \frac{1}{1-e^{-\tilde{\lambda}t}} \sum_{n\geqslant 1} \frac{(\tilde{\lambda}t)^n}{n!}e^{-\tilde{\lambda}t} \simeq 1.$$

1.5.3 Convergence of the counting process with a specific optimal bound as jump rate

We first show, in proposition 1.5.3, that $(\overline{T}_k)_{k\geqslant 0}$ (defined in section 1.3) is a Cox process with stochastic jump rate $\beta(t, x_t) - \lambda(x_t)$. Details on Cox processes can be found in [52].

Proposition 1.5.3. *The point process*

$$\xi([0,t]) = \sum_{n\geqslant 0} \mathbf{1}_{\overline{T}_n \leqslant t},$$

is a Cox process directed by the random measure μ such that $\mu([0,t]) = \int_0^t (\beta(s,x_s) - \lambda(x_s))ds$.

Proof. Let us first note that for $t \geqslant 0$, and, from (I.6) and (I.3), we have

$$\beta(t,x_t) - \lambda(x_t) = \sum_{n\geqslant 0} \Big[\tilde{\lambda}(t-T_n, x_{T_n}) - \lambda(\psi(t-T_n, x_{T_n}))\Big]\mathbf{1}_{T_n \leqslant t < T_{n+1}}.$$

The Laplace transform of a random measure completely characterises its distribution (see in [19], the discussion p.57 after Theorem 9.4.II). Moreover, the Laplace transform of a Cox process is given in [52], chap 12. Thus, we show that for any measurable and non-negative function f, the Laplace transform of ξ satisfies

$$\mathbb{E}[e^{-\xi f}] = \mathbb{E}[e^{-\mu(1-e^{-f})}], \tag{I.12}$$

where $\xi f = \int f d\xi$. Let f be a non-negative measurable function. Let us note $f_T(t) = f(t)\mathbf{1}_{t\leqslant T}$ for $T > 0$. Thus, $\lim_{T\to\infty} f_T(t) = f(t)$ with f_T increasing with T. Then, by Beppo-Levi Theorem, $\xi f_T \nearrow \xi f$ and $e^{-\xi f_T} \searrow e^{-\xi f}$ when T goes to infinity. Moreover, $e^{-\xi f_T} \leqslant 1$, thus by Lebesgue dominated convergence Theorem

$$\mathbb{E}[e^{-\xi f_T}] \to \mathbb{E}[e^{-\xi f}].$$

With the same type of arguments, we show that

$$\mathbb{E}[e^{-\mu(1-e^{-f_T})}] \to \mathbb{E}[e^{-\mu(1-e^{-f})}].$$

Thus, it is sufficient to show (I.12) for functions f_T. We have

$$\begin{aligned}
\mathbb{E}[e^{-\xi f_T}] &= \mathbb{E}[e^{-\sum_{n\geqslant 1} f_T(\overline{T}_n)}] \\
&= \sum_{k\geqslant 0} \mathbb{E}[e^{-\sum_{n\geqslant 1} f_T(\overline{T}_n)} | N_T = k]\mathbb{P}(N_T = k) \\
&= \sum_{k\geqslant 0} \mathbb{E}\Big[\mathbb{E}[\prod_{i=0}^{k} e^{-\sum_{n\geqslant 1} f_T(\overline{T}_n)\mathbf{1}_{T_i \leqslant \overline{T}_n < T_{i+1}}} | N_T = k, (\eta_t)_{0\leqslant t\leqslant T}] | N_T = k\Big]\mathbb{P}(N_T = k).
\end{aligned}$$

By the thinning procedure, the points $\overline{T}_n$ in $[T_i, T_{i+1}[$ may be written as $T_i + \tilde{T}_l^i$ for some $l \geqslant 1$ where $(\tilde{T}_l^i)_{l\geqslant 1}$ is, conditionally on (T_i, x_{T_i}), a Poisson process with jump rate $\tilde{\lambda}(t-T_i, x_{T_i}) - \lambda(\psi(t-T_i, x_{T_i}))$ for $t \geqslant T_i$. Since $(\tilde{T}_l^i)_{l\geqslant 0}$ is independent of $(\tilde{T}_l^j)$ for $i \neq j$, the random variables $X_i := e^{-\sum_{n\geqslant 1} f_T(\overline{T}_n)\mathbf{1}_{T_i \leqslant \overline{T}_n < T_{i+1}}}$ are independent conditionally

on $(\eta_t)_{0\leqslant t\leqslant T}$. Moreover, the Laplace functional of a Poisson process ξ with intensity μ verifies $\mathbb{E}[e^{-\xi f}] = e^{-\mu(1-e^{-f})}$. Thus, we obtain

$$\mathbb{E}[e^{-\xi f_T}] = \sum_{k\geqslant 0} \mathbb{E}\Big[\prod_{i=0}^{k} \mathbb{E}[e^{-\sum_{n\geqslant 1} f_T(\overline{T}_n)\mathbf{1}_{T_i\leqslant \overline{T}_n<T_{i+1}}} | N_T = k, (\eta_t)_{0\leqslant t\leqslant T}] | N_T = k\Big]\mathbb{P}(N_T = k)$$

$$= \sum_{k\geqslant 0} \mathbb{E}\Big[e^{-\sum_{i=0}^{N_T}\int \left(1-e^{-f_T(s)\mathbf{1}_{T_i\leqslant s<T_{i+1}}}\right)\left(\tilde{\lambda}(s-T_i,x_{T_i})-\lambda(\psi(s-T_i,x_{T_i}))\right)ds} | N_T = k\Big]\mathbb{P}(N_T = k)$$

$$= \mathbb{E}\Big[e^{-\sum_{i\geqslant 0}\int \left(1-e^{-f_T(s)}\right)\mathbf{1}_{T_i\leqslant s<T_{i+1}}\left(\tilde{\lambda}(s-T_i,x_{T_i})-\lambda(\psi(s-T_i,x_{T_i}))\right)ds}\Big]$$

$$= \mathbb{E}[e^{-\mu(1-e^{-f_T})}].$$

□

Now, let $P = \mathbb{N}$, $\epsilon > 0$ and let $(\mathcal{P}_k^\epsilon)_{k\in\mathbb{N}}$ be the partition such that $\mathcal{P}_k^\epsilon = [k\epsilon, (k+1)\epsilon[$ for $k \in \mathbb{N}$. Let us denote $\tilde{\lambda}^{\mathrm{opt},\epsilon}$ the *optimal bound* with the partition $(\mathcal{P}_k^\epsilon)_{k\in\mathbb{N}}$. In this case we have,

$$\tilde{\lambda}^{\mathrm{opt},\epsilon}(u, y) = \sum_{k\geqslant 0} \sup_{s\in[k\epsilon,(k+1)\epsilon[} \lambda(\psi(s, y))\mathbf{1}_{[k\epsilon,(k+1)\epsilon[}(u). \tag{I.13}$$

Moreover, we note $\beta^{\mathrm{opt},\epsilon}(t, x_t)$ the jump rate bound along the trajectory of (x_t) and we note $(\tilde{N}_t^{\mathrm{opt},\epsilon})$ the corresponding counting process. The number of points needed to simulate one inter-jump time of a PDMP in state $x \in E$ with this particular *optimal bound* is noted $\tau^{\mathrm{opt},\epsilon}(x)$.

We show, in proposition 1.5.4, that the counting process $(\tilde{N}_t^{\mathrm{opt},\epsilon})$ converges in distribution when ϵ goes to 0 to the counting process (N_t) of the PDMP. Finally, proposition 1.5.5 states that the smaller the parameter ϵ the fewer points have to be rejected. We begin by a lemma.

Lemma 1.5.1. *We have the following uniform convergence*

$$\sup_{x\in E}\sup_{s\geqslant 0} |\tilde{\lambda}^{opt,\epsilon}(s, x) - \lambda(\psi(s, x))| \underset{\epsilon\to 0}{\longrightarrow} 0,$$

where $\tilde{\lambda}^{opt,\epsilon}$ *is given by* (I.13).

Proof. For $n > 0$ we set $\epsilon = 1/n$, thus,

$$\tilde{\lambda}^{\mathrm{opt},1/n}(t, x) = \sum_{k\geqslant 0} \sup_{s\in[k/n,(k+1)/n[} \lambda(\psi(s, x))\mathbf{1}_{[k/n,(k+1)/n[}(t).$$

Let $M = \sup_{x\in E}\sup_{s\geqslant 0}\left|\frac{\partial\lambda}{\partial s}\Big(\psi(s, x)\Big)\right|$, $\nu > 0$, $N = \lceil M/\nu\rceil$ and $n \geqslant N$. Let $x \in E$ and $t \geqslant 0$, there exists $l \geqslant 0$ such that $t \in [l/n, (l+1)/n[$. Thus,

$$\tilde{\lambda}^{\mathrm{opt},1/n}(t, x) = \sup_{s\in[l/n,(l+1)/n[} \lambda(\psi(s, x)).$$

Let $t_0 \in [l/n, (l+1)/n]$ such that $\sup_{s\in[l/n,(l+1)/n[} \lambda(\psi(s,x)) = \lambda(\psi(t_0,x))$. The application of the mean value inequality to the function $t \to \lambda(\psi(t,x))$ gives

$$|\lambda(\psi(t_0,x)) - \lambda(\psi(t,x))| \leqslant M|t_0 - t| \leqslant M\frac{1}{n} \leqslant \nu.$$

The conclusion follows. □

Proposition 1.5.4. *Let (N_t) be the counting process of the PDMP (x_t). For $(\tilde{N}_t^{opt,\epsilon})$ defined above, we have the following convergence in distribution*

$$\tilde{N}^{opt,\epsilon} \underset{\epsilon\to 0}{\longrightarrow} N.$$

Proof. In order to show the convergence in distribution of $\tilde{N}^{\mathrm{opt},\epsilon}$ toward N when ϵ goes to 0, we show the convergence of the Laplace transform of $\tilde{N}^{\mathrm{opt},\epsilon}$ toward the one of N (see [16] Proposition 4.13, p.99 for example). More precisely, for all non-negative measurable function f, we show that

$$\mathbb{E}[e^{-\int f d\tilde{N}^{\mathrm{opt},\epsilon}}] \underset{\epsilon\to 0}{\longrightarrow} \mathbb{E}[e^{-\int f dN}].$$

Let f be a non-negative measurable function and let $T > 0$, following the same arguments as in the beginning of the proof of proposition 1.5.3, it is sufficient to show the convergence of the Laplace transform for functions $f_T(t) = f(t)\mathbf{1}_{t\leqslant T}$. Let $(\tilde{T}_n^\epsilon)$ be the points of the process $\tilde{N}^{\mathrm{opt},\epsilon}$. We have

$$\begin{aligned}
\mathbb{E}[e^{-\int f_T d\tilde{N}^{\mathrm{opt},\epsilon}}] &= \mathbb{E}[e^{-\sum_{n\geqslant 0} f_T(\tilde{T}_n^\epsilon)}] \\
&= \mathbb{E}\Big[\mathbb{E}[e^{-\sum_{n\geqslant 0} f_T(\tilde{T}_n^\epsilon)}|(\eta_t)_{0\leqslant t\leqslant T}]\Big] \\
&= \mathbb{E}\Big[e^{-\sum_{n\geqslant 0} f_T(T_n)}\mathbb{E}[e^{-\sum_{n\geqslant 0} f_T(\overline{T}_n^\epsilon)}|(\eta_t)_{0\leqslant t\leqslant T}]\Big],
\end{aligned}$$

where $(\overline{T}_n^\epsilon)$ denotes the rejected points. Since $(\overline{T}_n^\epsilon)$ is a Cox process with stochastic jump rate $\beta^{\mathrm{opt},\epsilon}(t,x_t) - \lambda(x_t)$, we obtain

$$\mathbb{E}[e^{-\int f_T d\tilde{N}^{\mathrm{opt},\epsilon}}] = \mathbb{E}\Big[e^{-\sum_{n\geqslant 0} f_T(T_n)}e^{-\int(1-e^{-f_T(s)})(\beta^{\mathrm{opt},\epsilon}(s,x_s)-\lambda(x_s))ds}\Big].$$

Since $e^{-\sum_{n\geqslant 0} f_T(T_n)}e^{-\int(1-e^{-f_T(s)})(\beta^{\mathrm{opt},\epsilon}(s,x_s)-\lambda(x_s))ds} \leqslant 1$, we obtain, by Lebesgue dominated convergence Theorem and by continuity of the exponential, that

$$\lim_{\epsilon\to 0}\mathbb{E}[e^{-\int f_T d\tilde{N}^{\mathrm{opt},\epsilon}}] = \mathbb{E}\Big[e^{-\sum_{n\geqslant 0} f_T(T_n)}e^{\lim_{\epsilon\to 0}\int -(1-e^{-f_T(s)})(\beta^{\mathrm{opt},\epsilon}(s,x_s)-\lambda(x_s))ds}\Big].$$

Moreover, we have

$$-T\sup_{y\in E}\sup_{u\geqslant 0}\Big(\tilde{\lambda}^{\mathrm{opt},\epsilon}(u,y) - \lambda(\psi(u,y))\Big) \leqslant \int -(1-e^{-f_T(s)})(\beta^{\mathrm{opt},\epsilon}(s,x_s) - \lambda(x_s))ds \leqslant 0,$$

Where $\tilde{\lambda}^{\mathrm{opt},\epsilon}$ is given by (I.13).

By lemma 1.5.1, we obtain that almost surely $e^{\lim_{\epsilon\to 0}\int -(1-e^{-f_T(s)})(\beta^{\mathrm{opt},\epsilon}(x_s,s)-\lambda(x_s))ds} = 1$. The conclusion follows since $\mathbb{E}\Big[e^{-\sum_{n\geqslant 0} f_T(T_n)}\Big] = \mathbb{E}\Big[e^{-\int f_T dN}\Big]$. □

Proposition 1.5.5. *For all $x \in E$, we have*

$$\mathbb{E}[\tau^{opt,\epsilon}(x)] \underset{\epsilon \to 0}{\longrightarrow} 1.$$

Proof. Let $x \in E$ and $\epsilon > 0$. From theorem 2.2 in chap.6 of [26], we have

$$\mathbb{E}[\tau^{\mathrm{opt},\epsilon}(x)] = \int_0^{+\infty} \tilde{\lambda}^{\mathrm{opt},\epsilon}(t,x) e^{-\int_0^t \lambda(\psi(s,x))ds} dt,$$

Where $\tilde{\lambda}^{\mathrm{opt},\epsilon}$ is given by (I.13).

From theorem 2.3 in [26] chap.6, we deduce that

$$\mathbb{E}[\tau^{\mathrm{opt},\epsilon}(x)] \leqslant \sup_{s \geqslant 0} \frac{\tilde{\lambda}^{\mathrm{opt},\epsilon}(s,x)}{\lambda(\psi(s,x))}.$$

By Lemma 1.5.1, we obtain

$$\lim_{\epsilon \to 0} \mathbb{E}[\tau^{\mathrm{opt},\epsilon}(x)] \leqslant 1.$$

Since $\mathbb{E}[\tau^{\mathrm{opt},\epsilon}(x)] \geqslant 1$ for all $\epsilon > 0$, the conclusion follows. □

1.6 Hodgkin-Huxley models

In this section, we introduce two deterministic Hodgkin-Huxley models and their stochastic versions, namely, the *subunit model* and the *channel model*.

1.6.1 Deterministic Hodgkin-Huxley models

In the celebrated paper [47] Alan Lloyd Hodgkin and Andrew Huxley proposed a deterministic model to explain the ionic mechanisms underlying the initiation of action potentials in the squid giant axon. They pointed out that the initiation of action potentials relies on three types of channels (sodium, potassium, and leak) which allow the transfer of ions across the membrane. A sodium (potassium respectively) channel is permeable to sodium ions (potassium ions respectively) only and is composed by three activation gates represented by the variable m and one inactivation gate h (four activation gates n and zero inactivation gates respectively). Leak channels are always open and allow all types of ions to pass the membrane. The gates are either open or closed and we say that a channel is open when all its gates are open. The classical four-dimensional Hodgkin-Huxley model is the following set of nonlinear differential equations.

$$\begin{cases} C\frac{dv}{dt} = I(t) - I_{\mathrm{L}}(v) - I_{\mathrm{Na}}(v,m,h) - I_{\mathrm{K}}(v,n), \\ \frac{dm}{dt} = (1-m)\alpha_m(v) - m\beta_m(v), \\ \frac{dh}{dt} = (1-h)\alpha_h(v) - h\beta_h(v), \\ \frac{dn}{dt} = (1-n)\alpha_n(v) - n\beta_n(v). \end{cases} \tag{I.14}$$

In this model, a channel is modelled by the gates that compose it. The $\mathbb{R}$ valued function v represents the membrane potential (voltage). The $[0,1]$ valued functions m, h, n correspond to the probability of a gate of type m, h (for the sodium) or n (for the potassium) to be open. The functions α_z and β_z for $z = m, h, n$ are opening and closing rates of gates z respectively. I is a time-dependent function which represents the input current, C is the membrane capacity. For $z \in \{\mathrm{Na}, \mathrm{K}, \mathrm{L}\}$, $I_z = g_z(v - v_z)$ represents the ionic currents where $g_{\mathrm{Na}} = \overline{g}_{\mathrm{Na}} m^3 h$, $g_{\mathrm{K}} = \overline{g}_{\mathrm{K}} n^4$ and $g_{\mathrm{L}} = \overline{g}_{\mathrm{L}}$ are the conductances of the sodium, potassium and leak respectively. Thus the constants $\overline{g}_{\mathrm{L}}$, $\overline{g}_{\mathrm{Na}}$, $\overline{g}_{\mathrm{K}}$ are the conductances when all the gates are open and v_{L}, v_{Na}, v_{K} are the resting potentials.

Now consider a channel in itself. Let $E_{\mathrm{Na}} = \{m_0h_0, m_1h_0, m_2h_0, m_3h_0, m_0h_1, m_1h_1, m_2h_1, m_3h_1\}$ be the set of the possible states of a sodium channel and $E_{\mathrm{K}} = \{n_0, n_1, n_2, n_3, n_4\}$ be the set of those of a potassium channel. The fourteen dimensional Hodgkin-Huxley model is given by the following set of nonlinear differential equations.

$$\begin{cases} C\frac{d\widehat{v}}{dt} = I(t) - I_{\mathrm{L}}(\widehat{v}) - I_{\mathrm{Na}}(\widehat{v}, \gamma^{\mathrm{Na}}_{m_3h_1}) - I_{\mathrm{K}}(\widehat{v}, \gamma^{\mathrm{K}}_{n_4}), \\ \frac{d\gamma^{\mathrm{Na}}_k}{dt} = \sum_{i \in E_{\mathrm{Na}}, i \neq k} \rho^{\mathrm{Na}}_{i,k}(\widehat{v})\gamma^{\mathrm{Na}}_i - \rho^{\mathrm{Na}}_{k,i}(\widehat{v})\gamma^{\mathrm{Na}}_k, \forall k \in E_{\mathrm{Na}}, \\ \frac{d\gamma^{\mathrm{K}}_l}{dt} = \sum_{j \in E_{\mathrm{K}}, j \neq l} \rho^{\mathrm{K}}_{j,l}(\widehat{v})\gamma^{\mathrm{K}}_j - \rho^{\mathrm{K}}_{l,j}(\widehat{v})\gamma^{\mathrm{K}}_l, \forall l \in E_{\mathrm{K}}. \end{cases} \tag{I.15}$$

The $[0,1]$ valued functions γ^{Na}_k for $k \in E_{\mathrm{Na}}$ (γ^{K}_l for $l \in E_{\mathrm{K}}$ respectively) represent the probability of a sodium (potassium respectively) channel to be in the state k (state l respectively). For all $(i,j) \in E_{\mathrm{Na}} \times E_{\mathrm{Na}}$ ($E_{\mathrm{K}} \times E_{\mathrm{K}}$ respectively), the function $\rho^{\mathrm{Na}}_{i,j}$ ($\rho^{\mathrm{K}}_{i,j}$ respectively) is the transition rate from state i to state j for a sodium (potassium respectively) channel. The possible sodium (potassium respectively) transitions are given in Figure (I.2) below ((I.3) respectively). For example, $\rho^{\mathrm{Na}}_{i,j} = 3\alpha_m$ if $i = m_0h_0$ and $j = m_1h_0$ and $\rho^{\mathrm{K}}_{i,j} = \alpha_n$ if $i = n_3$ and $j = n_4$. In this model, the functions $\widehat{v}$, I, I_{L}, I_{Na} and I_{K} and the constant C have the same meaning as in (I.14) but, the conductances of the sodium and potassium are now modelled by $g_{\mathrm{Na}} = \overline{g}_{\mathrm{Na}}\gamma^{\mathrm{Na}}_{m_3h_1}$ and $g_{\mathrm{K}} = \overline{g}_{\mathrm{K}}\gamma^{\mathrm{K}}_{n_4}$. Note that the conductance of the membrane depends on the probability of a channel to be open.

These models describe the electrical behaviour of a neuron with an infinite number of gates or channels. Thus, they do not reflect the variability observed experimentally. Note that, if a binomial relation is satisfied between the initial configuration of gates and channels and if $v(0) = \widehat{v}(0)$, then, the two models provide the same potential (i.e. $v(t) = \widehat{v}(t)$, for all $t \geqslant 0$), see [67]. Figure I.1 is obtained with the following set of parameters.

$$\alpha_n(x) = \tfrac{(0.1-0.01x)}{\exp(1-0.1x)-1}, \qquad \alpha_m(x) = \tfrac{(2.5-0.1x)}{\exp(2.5-0.1x)-1}, \quad \alpha_h(x) = 0.07\exp(-\tfrac{x}{20}),$$

$$\beta_n(x) = 0.125\exp(-\tfrac{x}{80}), \quad \beta_m(x) = 4\exp(-\tfrac{x}{18}), \qquad \beta_h(x) = \tfrac{1}{\exp(3-0.1x)+1},$$

$$V_{\mathrm{Na}} = 115, \quad g_{\mathrm{Na}} = 120, \quad V_{\mathrm{K}} = -12, \quad g_{\mathrm{K}} = 36, \quad V_{\mathrm{L}} = 0, \quad g_{\mathrm{L}} = 0.3, \quad C = 1.$$

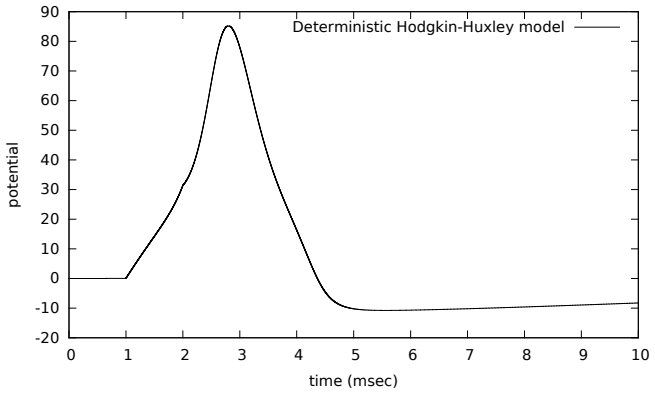


Figure I.1 – Simulated trajectory of the deterministic four-dimensional Hodgkin-Huxley model. The value of the parameters is given above and $I(t) = 30\mathbf{1}_{[1,2]}(t)$.

1.6.2 Stochastic Hodgkin-Huxley models

Neurons are subject to various sources of fluctuations, intrinsic (from the membrane) and extrinsic (from synapses). The intrinsic fluctuations are mainly caused by ion channels. To take into account these fluctuations in the models, we fix a finite number of gates or channels and replace their deterministic dynamic by stochastic processes. Here, we discuss two stochastic models, the *subunit model* and the *channel model*. These models belong to the class of Piecewise Deterministic Markov Processes. In the sequel we denote by V the stochastic membrane potentials of the *subunit model* and the *channel model* as opposed to the deterministic ones denoted by v and $\widehat{v}$ (see I.14 and I.15 respectively).

The subunit model

The *subunit model* is obtained by considering that the conductance of the membrane depends on the empirical measure defined by the proportion of open gates. We denote the number of gates of type m (respectively h, n) by N_m (respectively N_h, N_n). Let us consider that each gate is represented by a $\{0,1\}$-valued Markovian Jump Process (MJP) noted $u_k^{(z)}$ for $z = m, h, n$ and $k = 1, \ldots, N_z$. State 1 corresponds to the open configuration and 0 to the closed one. The opening and closing rates which depend on the voltage are noted $\alpha_z(.)$ and $\beta_z(.)$ respectively. The dynamics of a gate can be represented by the following diagram.

$$0 \underset{\beta_z(.)}{\overset{\alpha_z(.)}{\rightleftarrows}} 1. \tag{I.16}$$

We consider that all MJPs are independent conditionally on V_t, the value of the potential at time t, and we define the number of open gates z at time t by

$$\theta^{(z)}(t) = \sum_{k=1}^{N_z} u_k^{(z)}(t).$$

Furthermore, let $\Theta_{\text{sub}} = \{0, \dots, N_n\} \times \{0, \dots, N_m\} \times \{0, \dots, N_h\}$ be the state space of the process $\theta_t = \left(\theta^{(n)}(t), \theta^{(m)}(t), \theta^{(h)}(t)\right)$ which records the number of open gates at time t. Note that, $N_z - \theta^{(z)}(t)$ gives the number of closed gates z at time t. The *subunit model* takes the following form

$$(S) \quad \begin{cases} C\frac{dV_t}{dt} = f^{\text{sub}}(\theta_t, V_t, t), \\ (\theta_t), \end{cases}$$

where,

$$\begin{aligned} f^{\text{sub}}(\theta, V, t) = I(t) - \overline{g}_{\text{L}}(V - V_{\text{L}}) - \overline{g}_{\text{Na}} N_m^{-3} \left(\theta^{(m)}\right)^3 N_h^{-1} \theta^{(h)} (V - V_{\text{Na}}) \\ - \overline{g}_{\text{K}} N_n^{-4} \left(\theta^{(n)}\right)^4 (V - V_{\text{K}}). \end{aligned}$$

We also define the jump rate of the process by

$$\begin{aligned} \lambda^{\text{sub}}(\theta, V) = &\left(\alpha_m(V)(N_m - \theta^{(m)}) + \beta_m(V)\theta^{(m)}\right) + \left(\alpha_h(V)(N_h - \theta^{(h)}) + \beta_h(V)\theta^{(h)}\right) + \\ &\left(\alpha_n(V)(N_n - \theta^{(n)}) + \beta_n(V)\theta^{(n)}\right). \end{aligned}$$

The membrane potential is continuous thus the transition measure Q^{sub} is only concerned by the post-jump location of the jump process θ. For example, the probability of the event of exactly one gate n opens (conditionally on the last jump time being T_k) is given by

$$Q^{\text{sub}}\left((\theta_{T_{k-1}}, V_{T_k}), \{\theta_{T_{k-1}} + (1, 0, 0)\}\right) = \frac{\alpha_n(V_{T_k})(N_n - \theta^{(n)}(T_{k-1}))}{\lambda^{\text{sub}}(\theta_{T_{k-1}}, V_{T_k})}.$$

To summarize, the *subunit model* can be expressed as a PDMP $x_t^{\text{sub}} = (\theta_t, V_t, t) \in \Theta_{\text{sub}} \times \mathbb{R} \times \mathbb{R}_+$ with vector field $f^{\text{sub}} : \Theta_{\text{sub}} \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$, jump rate $\lambda^{\text{sub}} : \Theta_{\text{sub}} \times \mathbb{R} \to \mathbb{R}_+$, and transition measure $Q^{\text{sub}} : \Theta_{\text{sub}} \times \mathbb{R} \times \mathcal{B}(\Theta_{\text{sub}}) \to [0, 1]$. The *subunit model* converges to (I.14) when the number of gates goes to infinity [67].

The channel model

In the *channel model* we denote by N_{Na} the number of sodium channels and by N_{K} the number of potassium ones. We define MJPs $u_k^{(\text{Na})}$ for $k = 1, \dots, N_{\text{Na}}$ (respectively $u_k^{(\text{K})}$ for $k = 1, \dots, N_{\text{K}}$), conditionally independent on V_t, to model the sodium (respectively potassium) channels. The dynamics of these MJPs can be represented by the diagrams in Figures I.2 and I.3.

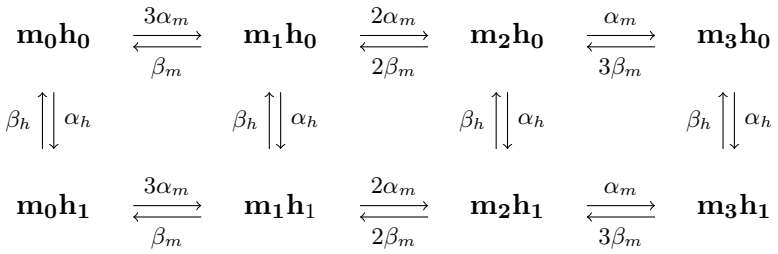


Figure I.2 – Sodium (Na) scheme

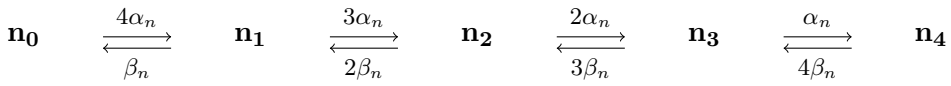


Figure I.3 – Potassium (K) scheme

The conducting state (the state that makes the channel active) of sodium (respectively potassium) channels is $\{m_3h_1\}$ (respectively $\{n_4\}$) which corresponds to three open gates m and one open gate h (respectively four open gates n). The conductance of the membrane depends on the empirical measure defined by the proportion of active channels. We define the number of active channels at time $t \geqslant 0$ by

$$\theta^{(m_3h_1)}(t) = \sum_{k=1}^{N_{\mathrm{Na}}} \mathbf{1}_{\{m_3h_1\}}\Big(u_k^{(\mathrm{Na})}(t)\Big), \quad \theta^{(n_4)}(t) = \sum_{k=1}^{N_{\mathrm{K}}} \mathbf{1}_{\{n_4\}}\Big(u_k^{(\mathrm{K})}(t)\Big).$$

For $i = 0, 1, 2, 3$ and $j = 0, 1$, let $\theta^{(m_ih_j)}$ be the number of channels in state $\{m_ih_j\}$ and for $k = 0, 1, 2, 3, 4$, let $\theta^{(n_k)}$ be the number of channels in state $\{n_k\}$. Let Θ_{chan} be the state space of the process $\theta_t = \Big((\theta^{(m_ih_j)}(t))_{i,j}, (\theta^{(n_k)}(t))_k\Big)$ which records the configuration of the channels at time t. The state space is defined by

$$\Theta_{\mathrm{chan}} = \{\theta \in \{0, \dots, N_{\mathrm{Na}}\}^8 \times \{0, \dots, N_{\mathrm{K}}\}^5 \ : \ \sum_{i=0}^{3}\sum_{j=0}^{1} \theta^{(m_ih_j)} = N_{\mathrm{Na}}, \ \sum_{k=0}^{5} \theta^{(n_k)} = N_{\mathrm{K}}\}.$$

The *channel model* takes the following form

$$(C) \quad \begin{cases} C\frac{dV_t}{dt} = f^{\mathrm{chan}}(\theta_t, V_t, t), \\ (\theta_t). \end{cases}$$

The vector field is given by

$$f^{\mathrm{chan}}(\theta, V, t) = I(t) - \overline{g}_{\mathrm{L}}(V - V_{\mathrm{L}}) - \overline{g}_{\mathrm{Na}}N_{\mathrm{Na}}^{-1}\theta^{(m_3h_1)}(V - V_{\mathrm{Na}}) - \overline{g}_{\mathrm{K}}N_{\mathrm{K}}^{-1}\theta^{(n_4)}(V - V_{\mathrm{K}}).$$

A change in the configuration of the channels happens when a gate opens or closes. We define the application $\eta : \Theta_{\mathrm{chan}} \to \Theta_{\mathrm{sub}}$ which, given a configuration of channels, returns

the configuration of the corresponding gates. We have

$$\eta(\theta) = \begin{bmatrix} \theta^{(n_1)} + 2\theta^{(n_2)} + 3\theta^{(n_3)} + 4\theta^{(n_4)} \\ \theta^{(m_1h_0)} + 2\theta^{(m_2h_0)} + 3\theta^{(m_3h_0)} + \theta^{(m_1h_1)} + 2\theta^{(m_2h_1)} + 3\theta^{(m_3h_1)} \\ \theta^{(m_0h_1)} + \theta^{(m_1h_1)} + \theta^{(m_2h_1)} + \theta^{(m_3h_1)} \end{bmatrix}.$$

The first component of the vector $\eta(\theta)$ contains θ^n_{open}, the number of open gates n, the second θ^m_{open}, the number of open gates m and the third θ^h_{open}, the number of open gates h. Thus, for $z = m, h, n$, $\theta^z_{\text{close}}(t) = N_z - \theta^z_{\text{open}}(t)$ gives the number of closed gates z at time t. We define the jump rate of the *channel model* by

$$\lambda^{\text{chan}}(\theta, V) = \lambda^{\text{sub}}(\eta(\theta), V),$$

where,

$$\begin{aligned} \lambda^{\text{sub}}(\eta(\theta), V) =& \Big(\alpha_m(V)(N_m - \theta^m_{\text{open}}) + \beta_m(V)\theta^m_{\text{open}}\Big) \\ &+ \Big(\alpha_h(V)(N_h - \theta^h_{\text{open}}) + \beta_h(V)\theta^h_{\text{open}}\Big) \\ &+ \Big(\alpha_n(V)(N_n - \theta^n_{\text{open}}) + \beta_n(V)\theta^n_{\text{open}}\Big). \end{aligned}$$

Since V is continuous, the kernel Q^{chan} is only concerned by the post-location of the process θ. Defining Q^{chan} classically done in the literature ([70] p.53 and [60] p.587) is computationally expensive because we have more transitions to deal with than in the *subunit model*. We propose to decompose the kernel Q^{chan} into a product of two kernels. The decomposition is based on the following observation: it is a change in the configuration of the gates that implies a change in the configuration of the channels. Thus, to determine which transition occurs at time t among the 28 transitions given above, we first determine which gate opens or closes by using the kernel Q^{sub} with $\lambda^{\text{sub}}(\eta(.), .)$ and then, depending on which gate changes state, we determine a channel transition by using another kernel. For example, suppose that at time t a gate m opens, thus, the possible channel transitions are: $\{m_0h_0 \to m_1h_0\}$, $\{m_1h_0 \to m_2h_0\}$, $\{m_2h_0 \to m_3h_0\}$, $\{m_0h_1 \to m_1h_1\}$, $\{m_1h_1 \to m_2h_1\}$, $\{m_2h_1 \to m_3h_1\}$ and the next transition is one of those. We define six kernels to take into account all the possibilities.

Let L^m_{open}, L^m_{close}, L^h_{open}, L^h_{close}, L^n_{open}, L^n_{close} be kernels defined on $\Theta_{\text{chan}} \times \mathbb{R} \times \mathcal{B}(\Theta_{\text{chan}})$ with values in $[0, 1]$ such that L^m_{open} is the kernel which chooses a transition as above, L^h_{open} is a kernel which choose a transition among the following ones $\{m_0h_0 \to m_0h_1\}$, $\{m_1h_0 \to m_1h_1\}$, $\{m_2h_0 \to m_2h_1\}$, $\{m_3h_0 \to m_3h_1\}$ and so on. For example, the probability of the event of having the transition $\{m_0h_0 \to m_1h_0\}$ (conditional on the last jump time being T_k) is given by

$$\begin{aligned} & Q^{\text{chan}}\Big((\theta_{T_{k-1}}, V_{T_k}), \{\theta_{T_{k-1}} + (-1, +1, 0, \ldots, 0)\}\Big) \\ =& Q^{\text{sub}}\Big((\eta(\theta_{T_{k-1}}), V_{T_k}), \{\eta(\theta_{T_{k-1}}) + (0, 1, 0)\}\Big) \\ & \times L^m_{\text{open}}\Big((\theta_{T_{k-1}}, V_{T_k}), \{\theta_{T_{k-1}} + (-1, +1, 0, \ldots, 0)\}\Big), \end{aligned}$$

where,

$$Q^{\text{sub}}\Big((\eta(\theta_{T_{k-1}}),V_{T_k}),\{\eta(\theta_{T_{k-1}})+(0,1,0)\}\Big)=\frac{\alpha_m(V_{T_k})\theta^m_{\text{close}}(T_{k-1})}{\lambda^{\text{sub}}(\eta(\theta_{T_{k-1}}),V_{T_k})},$$
$$L^m_{\text{open}}\Big((\theta_{T_{k-1}},V_{T_k}),\{\theta_{T_{k-1}}+(-1,+1,0,\dots,0)\}\Big)=\frac{3\theta^{(m_0h_0)}(T_{k-1})}{\theta^m_{\text{close}}(T_{k-1})}.$$

Finally, the probability of having the transition $\{m_0h_0\to m_1h_0\}$ is, as expected, given by the rate of this transition multiplied by the number of channels in the state $\{m_0h_0\}$ divided by the total rate.

For $x\in E$, the support K^{chan}_x of the discrete measure of probability $Q^{\text{chan}}(x,.)$ contains at most 28 elements (depending on the current state x), thus, in the worst case we have to do 28 "$if-then$" tests to determine the next transition. With the decomposition of Q^{chan}, we have, in the worst case 12 "$if-then$" tests to do. Indeed, for $x\in E$ the support K^{sub}_x of the discrete probability $Q^{\text{sub}}(\eta(x),.)$ contains at most six elements, and the support of the probabilities $L^m_{\text{open}}(x,.)$, $L^m_{\text{close}}(x,.)$, $L^h_{\text{open}}(x,.)$, $L^h_{\text{close}}(x,.)$, $L^n_{\text{open}}(x,.)$, $L^n_{\text{close}}(x,.)$ contains also at most six elements (when we deal with a transition of a gate m). Therefore, it is computationally cheaper to decompose the kernel.

Thus, the *channel model* can be expressed as a PDMP $x^{\text{chan}}_t=(\theta_t,V_t,t)\in\Theta_{\text{chan}}\times\mathbb{R}\times\mathbb{R}_+$ with vector field $f^{\text{chan}}:\Theta_{\text{chan}}\times\mathbb{R}\times\mathbb{R}_+\to\mathbb{R}$, jump rate $\lambda^{\text{chan}}:\Theta_{\text{chan}}\times\mathbb{R}\to\mathbb{R}_+$, and transition measure $Q^{\text{chan}}:\Theta_{\text{chan}}\times\mathbb{R}\times\mathcal{B}(\Theta_{\text{chan}})\to[0,1]$. The *channel model* converges to (I.15) when the number of channels goes to infinity [67].

Explicit flow between two successive jump times

In this section, we determine the explicit expression of the flow of both models. For $n\geqslant 0$, $t\geqslant T_n$ and $z\in\{\text{sub},\text{chan}\}$, the trajectory of the flow ϕ on $[T_n,+\infty[$ is given by the following ODE.

$$\begin{cases}\frac{d\phi(t-T_n,x_{T_n})}{dt}=f^z\Big(\theta_{T_n},\phi(t-T_n,x_{T_n}),t\Big)=-a^z_n\phi(t-T_n,x_{T_n})+b^z_n+\frac{1}{C}I(t),\\ \phi(0,x_{T_n})=V_{T_n},\end{cases}$$

where,

$$a^{\text{sub}}_n=\frac{1}{C}\left(g_{\text{L}}+g_{\text{Na}}N_m^{-3}\Big(\theta^{(m)}(T_n)\Big)^3N_h^{-1}\theta^{(h)}(T_n)+g_{\text{K}}N_n^{-4}\Big(\theta^{(n)}(T_n)\Big)^4\right),$$
$$b^{\text{sub}}_n=\frac{1}{C}\left(g_{\text{L}}V_{\text{L}}+g_{\text{Na}}V_{\text{Na}}N_m^{-3}\Big(\theta^{(m)}(T_n)\Big)^3N_h^{-1}\theta^{(h)}(T_n)+g_{\text{K}}V_{\text{K}}N_n^{-4}\Big(\theta^{(n)}(T_n)\Big)^4\right),$$
$$a^{\text{chan}}_n=\frac{1}{C}\Big(g_{\text{L}}+g_{\text{Na}}N_{\text{Na}}^{-1}\theta^{(m_3h_1)}(T_n)+g_{\text{K}}N_{\text{K}}^{-1}\theta^{(n_4)}(T_n)\Big),$$
$$b^{\text{chan}}_n=\frac{1}{C}\Big(g_{\text{L}}V_{\text{L}}+g_{\text{Na}}V_{\text{Na}}N_{\text{Na}}^{-1}\theta^{(m_3h_1)}(T_n)+g_{\text{K}}V_{\text{K}}N_{\text{K}}^{-1}\theta^{(n_4)}(T_n)\Big).$$

Then, the flow is given by

$$\phi(t-T_n,x_{T_n})=e^{-a^z_n(t-T_n)}\Big[V_{T_n}+\frac{b^z_n}{a^z_n}(e^{a^z_n(t-T_n)}-1)+\frac{1}{C}\int_{T_n}^t e^{a^z_n(s-T_n)}I(s)ds\Big]. \quad \text{(I.17)}$$

For both models we consider that the stimulation I takes the form $I(t) = K\mathbf{1}_{[t_1,t_2]}(t)$ with $K > 0$ and $t, t_1, t_2 \in \mathbb{R}_+$.

1.7 Simulations

We now proceed to the simulations of the *subunit model* and the *channel model* using the algorithm 1. Firstly, we explicit the three bounds for both models. Secondly, we numerically compare the efficiency of the bounds in term of reject and computation time. Finally, we use the algorithm 1 to compute a variable of biological interest for both models.

1.7.1 Determination of the jump rate bounds

For simplicity of presentation, we do not distinguish in the notation the flows of the *subunit model* and those of the *channel model*, one has to use a^{sub} and b^{sub} for the *subunit model* and a^{chan} and b^{chan} for the *channel model*. The determination of the bounds relies on the fact that α_n, α_m, β_h are increasing functions, β_n, β_m, α_h are decreasing, and that for $n \geqslant 0$, the flow (I.17) is bounded.

The global bound

To determine the *global bound* we use a result in [11] concerning the *channel model* which states that if $V_0 \in [V_-, V_+]$, then, $V_t \in [V_-, V_+]$ $\forall t \geqslant 0$, with, $V_- = \min\{V_{\mathrm{Na}}, V_{\mathrm{K}}, V_{\mathrm{L}}\}$ and $V_+ = \max\{V_{\mathrm{Na}}, V_{\mathrm{K}}, V_{\mathrm{L}}\}$. By using the monotony of the opening and closing rate functions, we find

$$\tilde{\lambda}^{\mathrm{glo}} = N_m\alpha_m(V_{\mathrm{Na}}) + N_h\beta_h(V_{\mathrm{Na}}) + N_n\alpha_n(V_{\mathrm{Na}}).$$

The result in [11] is also applicable to the *subunit model* and leads to the same expression of the *global bound* for this model.

The local bound

Let $n \geqslant 0$ and $t \geqslant T_n$. To determine the *local bound*, $\tilde{\lambda}^{\mathrm{loc}}(x_{T_n})$, we write the flow (I.17) as follows,

$$\phi(t - T_n, x_{T_n}) = f_n(t) + g_n(t),$$

where,

$$f_n(t) = e^{-a_n(t-T_n)}\Big(V_{T_n} + \frac{b_n}{a_n}(e^{a_n(t-T_n)} - 1)\Big),$$
$$g_n(t) = e^{-a_n(t-T_n)}\frac{1}{C}\int_{T_n}^{t} e^{a_n(s-T_n)}I(s)ds.$$

The purpose is to determine a lower and an upper bound of (I.17). We have $a_n > 0$, b_n may be negative or non-negative, and f_n is monotone. By using the fact that, $\forall t \geqslant 0$, $I(t) \leqslant K$, we find

$$\underline{V}_{T_n} \leqslant \phi(t - T_n, x_{T_n}) \leqslant \overline{V}_{T_n}, \tag{I.18}$$

where, $\overline{V}_{T_n} = V_{T_n} \vee \frac{b_n}{a_n} + \frac{K}{Ca_n}$, and $\underline{V}_{T_n} = V_{T_n} \wedge \frac{b_n}{a_n}$.

Then, by using the monotony of the opening and closing rate functions we obtain

$$\begin{aligned}\tilde{\lambda}^{\text{loc}}(x_{T_n}) =& \Big(\alpha_m(\overline{V}_{T_n})(N_m - \theta^m_{\text{open}}(T_n)) + \beta_m(\underline{V}_{T_n})\theta^m_{\text{open}}(T_n)\Big) \\ &+ \Big(\alpha_h(\underline{V}_{T_n})(N_h - \theta^h_{\text{open}}(T_n)) + \beta_h\overline{V}_{T_n})\theta^h_{\text{open}}(T_n)\Big) \\ &+ \Big(\alpha_n(\overline{V}_{T_n})(N_n - \theta^n_{\text{open}}(T_n)) + \beta_n(\underline{V}_{T_n})\theta^n_{\text{open}}(T_n)\Big).\end{aligned}$$

The expression of the *local bound* is the same for the *channel* and *subunit model* but the Markov chain θ is different.

The optimal bound

Let $n \geqslant 0$. We consider two partitions of $[0, +\infty[$. The first one is the same as in section 1.5.3 which is noted, for fixed $\epsilon > 0$, $(\mathcal{P}^\epsilon_k)_{k\in\mathbb{N}}$. We recall that, for $k \in \mathbb{N}$, $\mathcal{P}^\epsilon_k = [k\epsilon, (k+1)\epsilon[$ and that, in this case, the *optimal bound* is given by

$$\tilde{\lambda}^{\text{opt},\epsilon}(u, x_{T_n}) = \sum_{k\geqslant 0} \sup_{s\in[k\epsilon,(k+1)\epsilon[} \lambda(\psi(s, x_{T_n}))\mathbf{1}_{[k\epsilon,(k+1)\epsilon[}(u).$$

For $k \in \mathbb{N}$, we have

$$\begin{aligned}\sup_{s\in\mathcal{P}^\epsilon_k} \lambda(\psi(s, x_{T_n})) =& \Big(\alpha_m(\overline{V}^{k,\epsilon}_{T_n})(N_m - \theta^m_{open}(T_n)) + \beta_m(\underline{V}^{k,\epsilon}_{T_n})\theta^m_{open}(T_n)\Big) + \\ &\Big(\alpha_h(\underline{V}^{k,\epsilon}_{T_n})(N_h - \theta^h_{open}(T_n)) + \beta_h(\overline{V}^{k,\epsilon}_{T_n})\theta^h_{open}(T_n)\Big) + \\ &\Big(\alpha_n(\overline{V}^{k,\epsilon}_{T_n})(N_n - \theta^n_{open}(T_n)) + \beta_n(\underline{V}^{k,\epsilon}_{T_n})\theta^n_{open}(T_n)\Big),\end{aligned}$$

where,

$$\begin{aligned}\overline{V}^{k,\epsilon}_{T_n} &= f_n\left(T_n + k\epsilon\right) \vee f_n\left(T_n + (k+1)\epsilon\right) + e^{-a_n k\epsilon}\int_{T_n}^{T_n+(k+1)\epsilon} e^{a_n(s-T_n)}I(s)ds, \\ \underline{V}^{k,\epsilon}_{T_n} &= f_n\left(T_n + k\epsilon\right) \wedge f_n\left(T_n + (k+1)\epsilon\right) + e^{-a_n(k+1)\epsilon}\int_{T_n}^{T_n+k\epsilon} e^{a_n(s-T_n)}I(s)ds.\end{aligned}$$

The integrated *optimal bound* is given, for $u \geqslant 0$, by

$$\tilde{\Lambda}^{\text{opt},\epsilon}_{x_{T_n}}(u) = \sum_{k\geqslant 0} \sup_{s\in\mathcal{P}^\epsilon_k} \lambda(\psi(s, x_{T_n}))\Big[(k+1)\epsilon \wedge u - k\epsilon \wedge u\Big].$$

Its inverse is given by

$$\left(\tilde{\Lambda}^{\mathrm{opt},\epsilon}_{x_{T_n}}\right)^{-1}(u) = \sum_{p\geqslant 1}\Big(\frac{u-\epsilon\sum_{k=1}^{p-1}\sup_{s\in\mathcal{P}_k^\epsilon}\lambda(\psi(s,x_{T_n}))}{\sup_{s\in\mathcal{P}_p^\epsilon}\lambda(\psi(s,x_{T_n}))}+(p-1)\epsilon\Big)\mathbf{1}_{[\kappa_{p-1},\kappa_p[}(u),$$

where, $\kappa_p = \epsilon\sum_{k=1}^{p}\sup_{s\in\mathcal{P}_k^\epsilon}\lambda(\psi(s,x_{T_n}))$ and, by convention, $\kappa_0 = 0$.
The second partition is obtained for $P=\{0,1\}$ and is noted $(\mathcal{Q}_k^\epsilon)_{k\in P}$ where $\mathcal{Q}_0^\epsilon = [0,\epsilon[$ and $\mathcal{Q}_1^\epsilon = [\epsilon,+\infty[$. In this case, the *optimal bound* is given by

$$\tilde{\lambda}^{\mathrm{opt},\epsilon}(u,x_{T_n}) = \sup_{s\in[0,\epsilon[}\lambda(\psi(s,x_{T_n}))\mathbf{1}_{[0,\epsilon[}(u)+\tilde{\lambda}^{\mathrm{loc}}(x_{T_n})\mathbf{1}_{[\epsilon,+\infty[}(u).$$

The integrated *optimal bound* is

$$\tilde{\Lambda}^{\mathrm{opt},\epsilon}_{x_{T_n}}(u) = \sup_{s\in[0,\epsilon[}\lambda(\psi(s,x_{T_n}))(\epsilon\wedge u)+\tilde{\lambda}^{\mathrm{loc}}(x_{T_n})(u-\epsilon\wedge u).$$

The inverse is given by

$$\begin{aligned}\left(\tilde{\Lambda}^{\mathrm{opt},\epsilon}_{x_{T_n}}\right)^{-1}(u) =& \frac{u}{\sup_{s\in[0,\epsilon[}\lambda(\psi(s,x_{T_n}))}\mathbf{1}_{[0,\epsilon\sup_{s\in[0,\epsilon[}\lambda(\psi(s,x_{T_n}))[}(u)\\ &+\Big(\frac{u-\epsilon\sup_{s\in[0,\epsilon[}\lambda(\psi(s,x_{T_n}))}{\tilde{\lambda}^{\mathrm{loc}}(x_{T_n})}+\epsilon\Big)\mathbf{1}_{[\epsilon\sup_{s\in[0,\epsilon[}\lambda(\psi(s,x_{T_n})),+\infty[}(u).\end{aligned}$$

Once again, the expression of the *optimal bound* is the same for both models but the Markov chain is different. We precise that we used the *local bound* to define the *optimal bound* with the partition $(\mathcal{Q}_k^{T_n,\epsilon})_{k\in\{0,1\}}$.

Note that, for $n\geqslant 0$, it is possible to define an ϵ_n which is "adapted" to the inter jump time $T_{n+1}-T_n$. To determine such an ϵ_n, we use the bounds of the flow in inequality (I.18) to define a lower *local bound*, $\underline{\lambda}(x_{T_n})$, as follows

$$\begin{aligned}\underline{\lambda}(x_{T_n}) =& \Big(\alpha_m(\underline{V}_{T_n})(N_m-\theta^m_{\mathrm{open}}(T_n))+\beta_m(\overline{V}_{T_n})\theta^m_{\mathrm{open}}(T_n)\Big)\\ &+\Big(\alpha_h(\overline{V}_{T_n})(N_h-\theta^h_{\mathrm{open}}(T_n))+\beta_h(\underline{V}_{T_n})\theta^h_{\mathrm{open}}(T_n)\Big)\\ &+\Big(\alpha_n(\underline{V}_{T_n})(N_n-\theta^n_{\mathrm{open}}(T_n))+\beta_n(\overline{V}_{T_n})\theta^n_{\mathrm{open}}(T_n)\Big).\end{aligned}$$

We note $(\underline{T}_n)_{n\geqslant 0}$ the corresponding point process. Then we have

$$0.05 = \mathbb{P}(T_{n+1}-T_n>\epsilon_n)\leqslant\mathbb{P}(\underline{T}_{n+1}-\underline{T}_n>\epsilon_n) = e^{-\epsilon_n\underline{\lambda}(x_{T_n})}.$$

We take $\epsilon_n = \frac{-\log(0.05)}{\underline{\lambda}(x_{T_n})}$. Note that ϵ_n is in fact adapted to the inter jump time $\underline{T}_{n+1}-\underline{T}_n$.

1.7.2 Numerical results

In this section, we numerically compare the three different jump rate bounds (I.8), (I.9), (I.10) and we use Algorithm 1 to simulate a variable of biological interest, the spiking time.

Numerical comparison of the jump rate bounds

In this part, we first show trajectories of the two stochastic Hodgkin-Huxley models obtained with Algorithm 1 using the *optimal bound* with the partition $(\mathcal{P}_k^\epsilon)_{k\in\mathbb{N}}$. Then, we collect in several tables and graphs the results concerning the computation time and the rate of acceptance of both models for the three types of bounds.

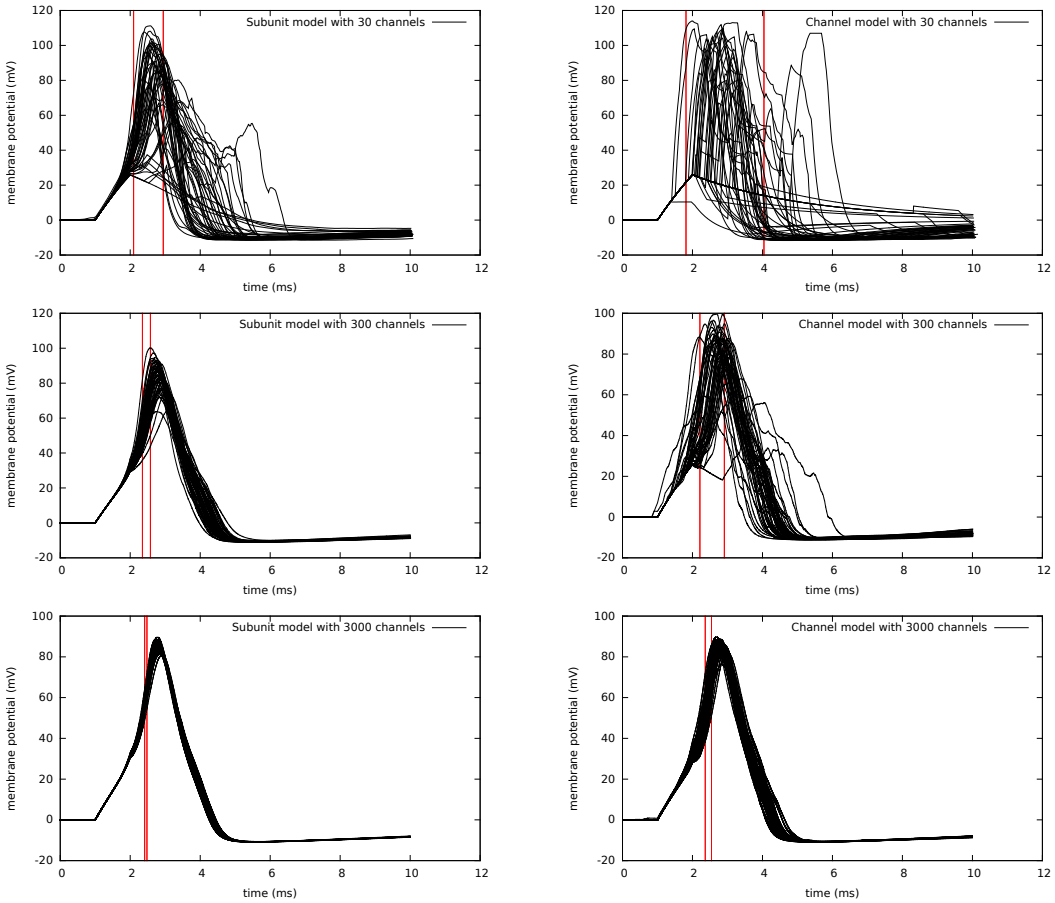


Figure I.4 – First column : *subunit model*. Second column : *channel model*. Vertical red rows are the standard deviation of the spiking times (see section 1.7.2).

In the sequel, for $\epsilon > 0$, the *optimal-$\mathcal{Q}^\epsilon$* (respectively *optimal-$\mathcal{P}^\epsilon$*) *bound* denotes the optimal bound using the partition $(\mathcal{Q}_k^\epsilon)_{k\in\{0,1\}}$ (respectively $(\mathcal{P}_k^\epsilon)_{k\in\mathbb{N}}$), see section 1.7.1. All numerical values are obtained from a classic Monte Carlo method with 100 000 trials. Parameters of the models are given in section 1.6.1 (the same set of parameters is used for both models). We denote by N_{chan} the common number of sodium and potassium channels, $N_{\text{chan}} = N_{\text{Na}} = N_{\text{K}}$. The input current is $I(t) = 30\mathbf{1}_{[1,2]}(t)$. The computation time represents the time needed to simulate one path of the PDMP on [0,10]. The simulations were carried out on a computer with a processor Intel Core i5-4300U CPU @

1.90GHz × 4. The code is written in C++ language.

Each rows of Figure 2 shows fifty trajectories of the *subunit* and the *channel model* with a different number of channels, $N_{\text{chan}} = 30, 300, 3000$. It allows to see the different behaviours of the two models. In each rows, we see that the behaviour of the *channel model* is more erratic than the *subunit model* one (except for the third row where the two models have approximately the same behaviour). Differences in trajectories are mainly explained by two distinct modelling approaches of the conductance of the membrane. In the *subunit model*, we consider that the conductance at time t depends on the fraction of open gates at time t, thus, the equation of the voltage changes rapidly at the same time as the state of the gates. In the *channel model*, the conductance at time t depends on the fraction of active channels at time t, therefore, a change in the state of the gates may not imply a change in the voltage's equation. Thus, the dynamic of the membrane potential changes less than in the first case and trajectories are more irregular. We also see that, the higher the number of channels the smaller the differences in trajectories. It illustrates a result in [67] where the authors showed that the deterministic limit (when the number of channels goes to infinity) of the variable V of both models are the same. However, it seems that the convergence speed is not the same.

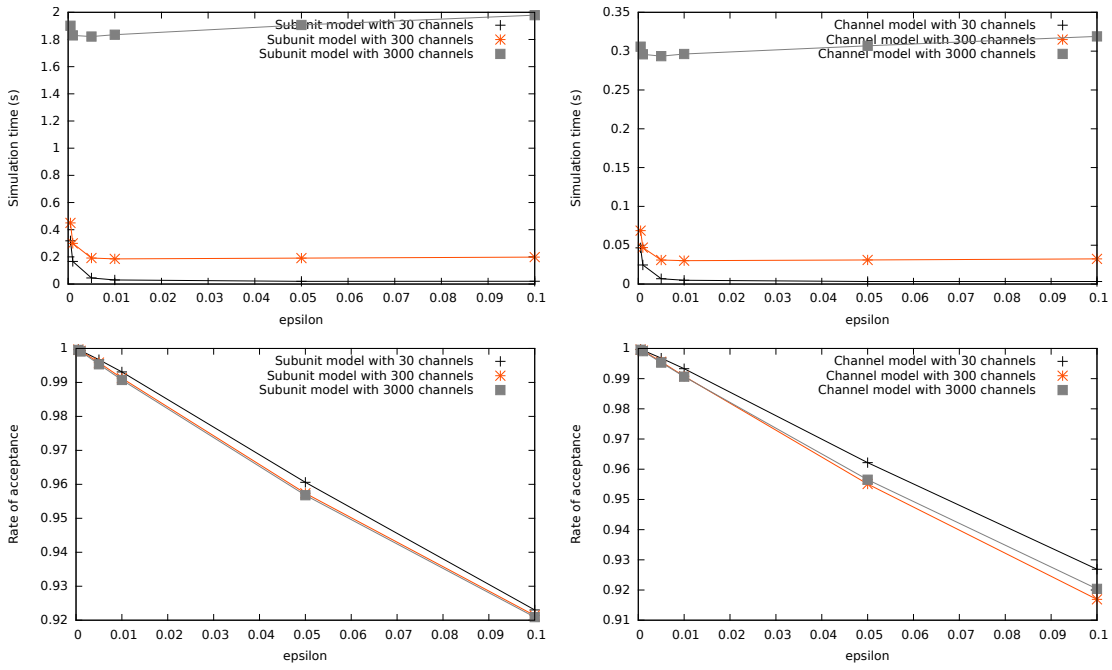


Figure I.5 – Computation time and rate of acceptance with the optimal-$\mathcal{P}^\epsilon$ bound as a function of the parameter ϵ.

Concerning the *optimal-*$\mathcal{P}^\epsilon$ bound, we see on Figure I.5 that in both models, the smaller $\epsilon$the fewer rejected points. It illustrates the fact that $\tilde{N}^\epsilon$ converge to N when ϵ goes to 0 (proposition 1.5.4). Figure 3 also shows that, for fixed N_{chan}, the computation

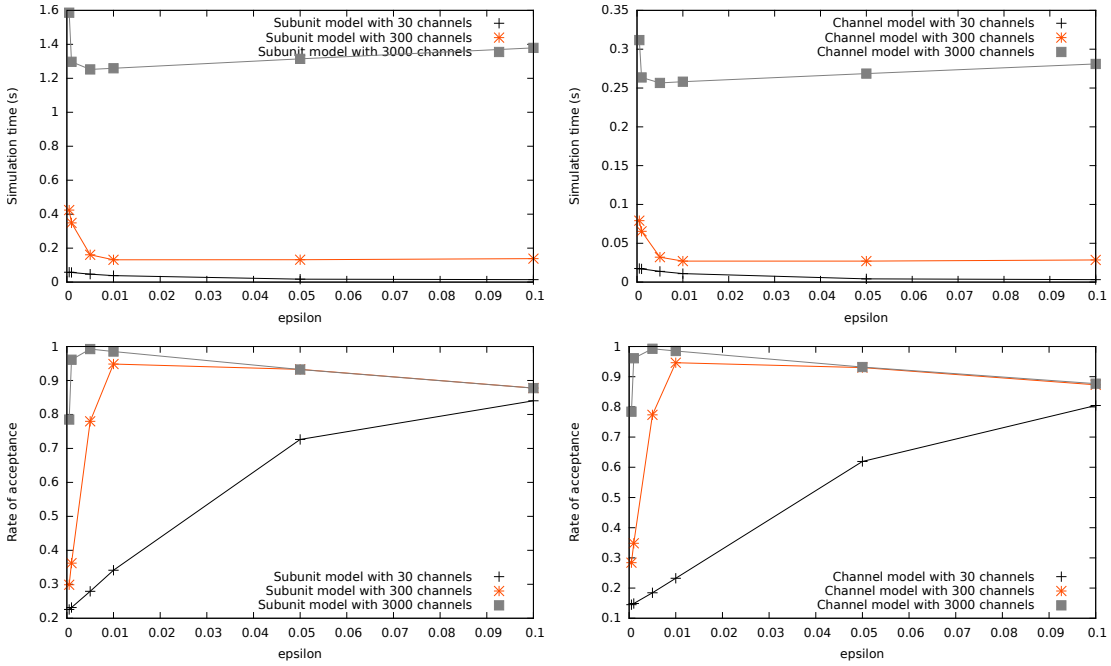


Figure I.6 – Computation time and rate of acceptance with the optimal-$\mathcal{Q}^\epsilon$ bound as a function of the parameter ϵ.

time varies with ϵ. For both models, the value of ϵ which minimizes the computation time is inversely proportional to the parameter N_{chan}. Let $\epsilon(N_{\text{chan}})$ be that optimal value of ϵ. For increasing $\epsilon > \epsilon(N_{\text{chan}})$, the rate of acceptance decreases, thus, we have to simulate more and more uniform pseudo-random variables and the computation time increases. For decreasing $\epsilon < \epsilon(N_{\text{chan}})$, the rate of acceptance increases but the computation time too because of the increasing number of iterations needed to compute the integrated jump rate bound and its inverse. Thus, one has to take a small (respectively large) ϵ when the jumps frequency is high (respectively low).

We see on Figure 4 that, the smaller ϵ the closer the rate of acceptance of the *optimal-*$\mathcal{Q}^\epsilon$ bound to the one of the *local bound*. Note that the value of ϵ which maximises the rate of acceptance is the same which minimizes the computation time. As in the case of the *optimal-*$\mathcal{P}^\epsilon$ bound, the optimal value of ϵ is inversely proportional to N_{chan}. For decreasing $\epsilon < \epsilon(N_{\text{chan}})$, the rate of acceptance decreases and the computation time increases because we mainly use the *local bound*, $\tilde{\lambda}^{\text{loc}}(x_{T_n})$, instead of the smaller bound, $\sup_{s\in[0,\epsilon[} \lambda(\psi(s, x_{T_n}))$, in the computation of $\tilde{\Lambda}^{\text{opt},\epsilon}$ and $(\tilde{\Lambda}^{\text{opt},\epsilon})^{-1}$ (see section 1.7.1). For increasing $\epsilon > \epsilon(N_{\text{chan}})$, the rate of acceptance decreases and the computation time increases because the bound $\sup_{s\in[0,\epsilon[} \lambda(\psi(s, x_{T_n}))$ becomes bigger and bigger.

By comparing the *optimal-*$\mathcal{Q}^{\epsilon(N_{\text{chan}})}$ and the *optimal-*$\mathcal{P}^{\epsilon(N_{\text{chan}})}$ bound we see that the first one is the most efficient in term of computation time, it is also the simplest to

implement. However, this bound does not exist when the jump rate or the flow is not bounded. In this case, one may use the *optimal-$\mathcal{P}^{\epsilon(N_{\text{chan}})}$ bound* which is efficient too but a little bit more complex to implement.

From Figures 3 and 4, we see that for both the *optimal-$\mathcal{Q}^{\epsilon}$* and the *optimal-$\mathcal{P}^{\epsilon}$ bounds* the best computation time is achieved for $\epsilon(30) = 0.1$, $\epsilon(300) = 0.01$ and $\epsilon(3000) = 0.005$. We saw in sections 1.6.2 and 1.6.2 that the *subunit model* and the *channel model* share the same jump rate. For both models, the maximum value of the inter-jump times is of order 10^{-1} for $N_{\text{chan}} = 30$, 10^{-2} for $N_{\text{chan}} = 300$ and 10^{-3} for $N_{\text{chan}} = 3000$. It coincides with the values $\epsilon(N_{\text{chan}})$ which, in this case, confirm that the optimal computation time is obtained for ϵ of order $\max_n |T_{n+1} - T_n|$.

Table I.1 – computation time and rate of acceptance for $N_{\text{chan}} = 30$. The lines ODE represent the algorithm in [70] with $h = 10^{-3}$ for both *subunit model* and *channel model*.

Model	Bound	computation time (sec)	rate of acceptance
Channel	Optimal-$\mathcal{Q}^{\epsilon_n}$	0,003 $(\pm 8.10^{-7})$	0,857 $(\pm 2.10^{-3})$
	Local	0,008 $(\pm 6.10^{-6})$	0,141 $(\pm 2.10^{-3})$
	Global	0,012 $(\pm 3.10^{-6})$	0,065 $(\pm 6.10^{-5})$
	ODE	0.009 $(\pm 1.10^{-7})$	
Subunit	Optimal-$\mathcal{Q}^{\epsilon_n}$	0,016 $(\pm 1.10^{-6})$	0,88 $(\pm 1.10^{-3})$
	Local	0,050 $(\pm 2.10^{-4})$	0,22 $(\pm 1.10^{-3})$
	Global	0,12 $(\pm 3.10^{-4})$	0,061 $(\pm 2.10^{-5})$
	ODE	0,016 $(\pm 2.10^{-7})$	

Table I.2 – computation time and rate of acceptance for $N_{\text{chan}} = 300$. The lines ODE represent the algorithm in [70] with $h = 10^{-4}$ for both *subunit model* and *channel model*.

Model	Bound	computation time (sec)	rate of acceptance
Channel	Optimal-$\mathcal{Q}^{\epsilon_n}$	0,030 $(\pm 3.10^{-5})$	0,962 $(\pm 9.10^{-5})$
	Local	0,050 $(\pm 1.10^{-4})$	0,223 $(\pm 3.10^{-4})$
	Global	0,120 $(\pm 3.10^{-4})$	0,062 $(\pm 7.10^{-5})$
	ODE	0.094 $(\pm 1.10^{-5})$	
Subunit	Optimal-$\mathcal{Q}^{\epsilon_n}$	0,148 $(\pm 5.10^{-4})$	0,957 $(\pm 9.10^{-5})$
	Local	0,244 $(\pm 1.10^{-3})$	0,237 $(\pm 8.10^{-5})$
	Global	0,322 $(\pm 2.10^{-3})$	0,061 $(\pm 1.10^{-5})$
	ODE	0,157 $(\pm 1.10^{-5})$	

Tables I.1-I.3 show results of the computation time and of the rate of acceptance of the thinning algorithm for the *global*, *local* and *optimal-$\mathcal{Q}^{\epsilon_n}$ bounds* using both the *channel* and the *subunit* models with different values of the parameter N_{chan}. For both

Table I.3 – computation time and rate of acceptance for $N_{\text{chan}} = 3000$. The lines ODE represent the algorithm in [70] with $h = 10^{-5}$ for both *subunit model* and *channel model*.

Model	Bound	computation time (sec)	rate of acceptance
Channel	optimal-$\mathcal{Q}^{\epsilon_n}$	0,296 $(\pm 3.10^{-3})$	0,965 $(\pm 2.10^{-5})$
	Local	0,474 $(\pm 6.10^{-3})$	0,236 $(\pm 3.10^{-5})$
	Global	1,184 $(\pm 2.10^{-2})$	0,060 $(\pm 3.10^{-7})$
	ODE	0.940 $(\pm 5.10^{-4})$	
Subunit	Optimal-$\mathcal{Q}^{\epsilon_n}$	1,471 $(\pm 3.10^{-2})$	0,964 $(\pm 9.10^{-6})$
	Local	2,478 $(\pm 4.10^{-2})$	0,238 $(\pm 7.10^{-6})$
	Global	3,315 $(\pm 3.10^{-1})$	0,060 $(\pm 9.10^{-8})$
	ODE	1,567 $(\pm 1.10^{-3})$	

models and for all the studied values of N_{chan}, the computation time using the *optimal bounds* ($\mathcal{Q}^{\epsilon(N_{\text{chan}})}$,$\mathcal{P}^{\epsilon(N_{\text{chan}})}$ and $\mathcal{Q}^{\epsilon_n}$) is better than the one obtained with both the *global* and *local bounds*. Note that the *optimal-*$\mathcal{Q}^{\epsilon_n}$ *bound* is more efficient than the *optimal-*$\mathcal{P}^{\epsilon(N_{\text{chan}})}$ *bound* to simulate the *subunit model*. Since the computation of ϵ_n requires the computation of the jump rate bound at each iterations, the *optimal-*$\mathcal{Q}^{\epsilon_n}$ *bound* will be more efficient when the jumps frequency is low. Thus, for all studied values of N_{chan}, the *optimal-*$\mathcal{Q}^{\epsilon(N_{\text{chan}})}$ *bound* is the most efficient.

The differences of computation time between the *subunit* and the *channel model* are explained by the fact that the numerical computation of the flow of the *channel model* is cheaper than the one of the *subunit model*. Note that the computation time using the three bounds (*global*, *local*, *optimal*) increases linearly as a function of N_{chan}.

In the ODE algorithm [70], we need to adapt the time step h when the parameter N_{chan} varies, otherwise, we do not simulate the expected trajectories of the models. Thinning algorithm in the *channel model* speeds up the simulation by a factor 3 compared to the ODE method whereas in the *subunit model* the factor is approximately 1. Such a difference is explained by the fact that the ratio of the computation times between the flows of the *subunit* and the *channel* (for thinning algorithm) is bigger than the ratio of the computation times between the vector fields of the *subunit* and the *channel* (for ODE algorithm).

Despite the complexity of the *optimal bound* compared to the two others, it is the most efficient one in terms of reject and computation time to simulate both the *channel model* and the *subunit model*.

Spiking times

Bio-scientists believe that the timing of action potentials is one of the characteristics of the nervous system which carries the most of information. It has been shown experimentally [78] that if a neuron is repeatedly stimulated by identical pulses, both the amplitude and

the timing of the action potentials is variable. In the sequel we numerically compare the mean value of the spiking time of the *subunit* and *channel model* to the one of the deterministic Hodgkin-Huxley model.

Let (x_t) be the *subunit model* or the *channel model* defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}_x)$. We consider that the stimulation is a monophasic current which produces only one action potential within a given time window $[0, T]$ as in Figure I.1. We suppose that a spike occurs when the membrane potential exceeds a certain value noted ν. Let $\mathcal{T}$ be the spiking time that we define by

$$\mathcal{T} = \inf\{t \in [0, T] : V_t \geqslant \nu\}.$$

We are interested in the numerical computation of the mean and the standard deviation of $\mathcal{T}$ as a function of the number of channels. For low values of the parameters N_{Na} and N_{K} a spike may never occur. In this case, $\mathcal{T} = T$ and we do not count these trajectories in the Monte Carlo procedure. Thus, we evaluate the mean value of the spiking time conditionally on having a spike, $\mathbb{E}[\mathcal{T}|\mathcal{T} < T]$, with the following estimator $I_M = (1/M)\sum_{k=1}^{M} \mathcal{T}_k$ where $(\mathcal{T}_k)$ are iid realizations of $\mathcal{T}$ conditionally on $\{\mathcal{T} < T\}$ and M denotes the sample size of the estimator. We define the proportion of spikes as follows. Consider that we simulate n independent trajectories of stochastic action potentials (with the *subunit model* or the *channel model*) on $[0, T]$. We define a sequence of independent random variables $X_1, \ldots, X_n$ as follows: for $i = 1, \ldots, n$,

$$X_i = \begin{cases} 1 \text{ if there exits } t \in [0, T] \text{ such that } V_t \geqslant \nu, \\ 0 \text{ if for all } t \in [0, T],\ V_t < \nu. \end{cases}$$

Then we define the proportion of spikes as $1/n \sum_{i=1}^{n} X_i$.

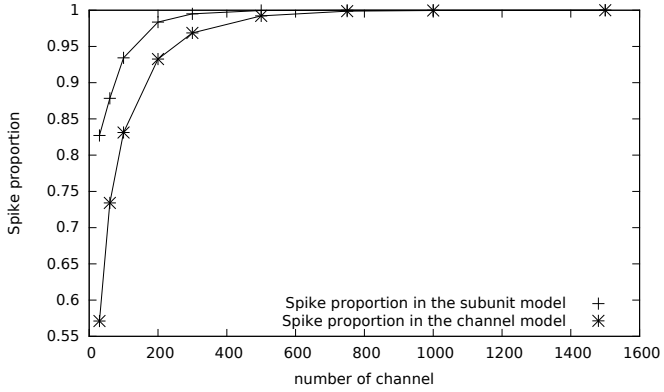


Figure I.7 – Proportion of spikes obtained with the *subunit model* and the *channel model* as a function of the number of channels N_{chan}.

It has been shown in [67] that the deterministic limits of both the *subunit* (Hodgkin-Huxley of dimension four [47]) and the *channel model* (Hodgkin-Huxley of dimension

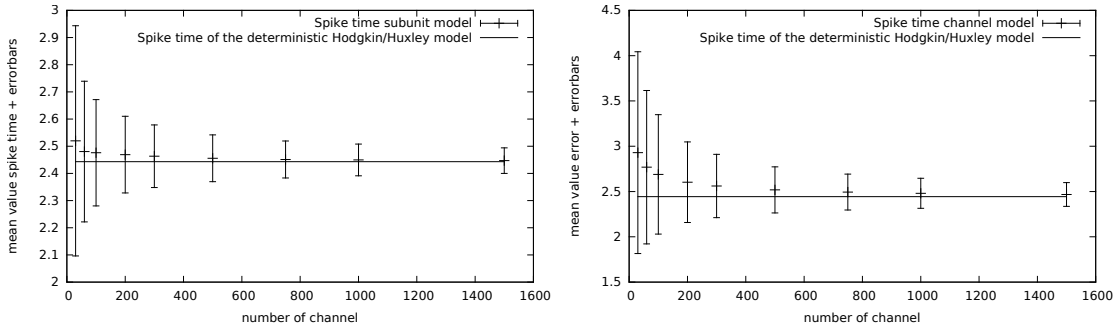


Figure I.8 – Mean value of the spiking time (ms) with standard deviation as a function of the number of channels N_{chan}. Left: *subunit model*. Right: *channel model*.

fourteen [67]) are equivalent when the initial conditions satisfy a combinatorial relationship. We consider that, at time $t = 0$, all the gates of the *subunit model* are closed and all the channels of the *channel model* are in the corresponding state, i.e state $\{m_0 h_0\}$ for the sodium and $\{n_0\}$ for the potassium. These initial conditions satisfy the combinatorial relationship in [67]. The initial conditions of both deterministic Hodgkin-Huxley models are also chosen so that they satisfy the binomial relation. Thus, the spiking time of these deterministic models is the same. In the simulations, we take $T = 10$, $\nu = 60$, we consider that the stimulation is given by $I(t) = 30\mathbf{1}_{[1,2]}(t)$ and that $N_{\text{Na}} = N_{\text{K}} = N_{\text{chan}}$. In this case, the spiking time of the deterministic model is $\mathcal{T}^{\text{deter}} = 2,443$.

Figure I.8 illustrates the convergence of the mean spiking time of both the *subunit* and the *channel model* when the number of channels goes to infinity. For $N_{\text{chan}} = 1500$ we see that the dispersion of the spiking time around its deterministic limit is approximately of order 10^{-1} ms for the *subunit model* and of order 10^{-2} ms for the *channel model*. Thus, a membrane patch with a number of channels superior to 1500 mimics the behaviour of the deterministic Hodgkin-Huxley model. For a number of channels inferior to 500, we see from Figure I.7 that the neuron may not respond to the stimuli. In this case, the dispersion of the spiking time ranges from approximately 10^{-1} and almost 1 ms which is consistent with the observations in [78]. Since the simulation is exact the estimator I_M is unbiased and errors due to the Monte Carlo procedure are of order of $M^{-1/2}$.

1.8 Appendix

In this section we compute the rate of acceptance for the thinning of Poisson processes. Let N and $\tilde{N}$ be two Poisson processes with jump rate λ and $\tilde{\lambda}$ respectively and jump times $(T_n)_{n\geqslant 1}$ and $(\tilde{T}_n)_{n\geqslant 1}$ respectively. Assume that N is the thinning of $\tilde{N}$. Since $\mathbb{P}(\tilde{N}_t = 0) = e^{-\int_0^t \tilde{\lambda}(s)ds}$, we define the rate of acceptance by $\mathbb{E}[N_t/\tilde{N}_t|\tilde{N}_t \geqslant 1]$. In the case of Poisson processes this indicator takes the following form

$$\mathbb{E}[\frac{N_t}{\tilde{N}_t}|\tilde{N}_t \geqslant 1] = \frac{\int_0^t \lambda(s)ds}{\int_0^t \tilde{\lambda}(s)ds}. \tag{I.19}$$

To get (I.19), we use the following result which is similar to the n-uplet of non-ordering uniform variables in the Poisson homogeneous case

$$f_{(\tilde{T}_1,\dots,\tilde{T}_n|\tilde{N}_t=n)}(t_1,\dots,t_n) = \frac{\tilde{\lambda}(t_1)\dots\tilde{\lambda}(t_n)}{\left(\int_0^t \tilde{\lambda}(s)ds\right)^n}\mathbf{1}_{(t_1,\dots,t_n)\in[0,t]^n}. \tag{I.20}$$

Equation (I.20) gives an explicit formula of the conditional density of the vector $(\tilde{T}_1,\dots,\tilde{T}_n|\tilde{N}_t = n)$. Note that we do not consider any ordering in points $(\tilde{T}_k)_{0\leqslant k\leqslant n}$ and that conditionally on $\{\tilde{N}_t = n\}$, the points $\tilde{T}_1,\dots,\tilde{T}_n$ are independent with density $\left(\tilde{\lambda}(s)/\int_0^t \tilde{\lambda}(u)du\right)\mathbf{1}_{s\in[0,t]}$. By noting that, for $k \leqslant n$,

$$\begin{aligned}&\{N_t = k|\tilde{N}_t = n\}\\ &= \bigcup_{1\leqslant i_1<\dots<i_k\leqslant n}\Big[\bigcap_{i\in\{i_1,\dots,i_k\}}\{U_i \leqslant \frac{\lambda}{\bar{\lambda}}(\tilde{T}_i)|\tilde{N}_t = n\}\bigcap_{i\in\{i_1,\dots,i_k\}^c}\{U_i > \frac{\lambda}{\bar{\lambda}}(\tilde{T}_i)|\tilde{N}_t = n\}\Big],\end{aligned}$$

where (U_i) are independent variables uniformly distributed in $[0,1]$, independent of $(\tilde{T}_i)$, we deduce that

$$\mathbb{P}(N_t = k|\tilde{N}_t = n) = \binom{n}{k}\mathbb{P}\left(U_i \leqslant \frac{\lambda}{\bar{\lambda}}(\tilde{T}_i)|\tilde{N}_t = n\right)^k \mathbb{P}\left(U_i > \frac{\lambda}{\bar{\lambda}}(\tilde{T}_i)|\tilde{N}_t = n\right)^{n-k}. \tag{I.21}$$

Thus, the law of the number of selected points is binomial conditionally on the number of generated points. With (I.20) and (I.21), one is able to determine that

$$\mathcal{L}(N_t|\tilde{N}_t = n) = \mathcal{B}(n,p),$$

with $p = \int_0^t \lambda(s)ds/\int_0^t \tilde{\lambda}(s)ds$. Then, we find (I.19) by using that

$$\mathbb{E}[\frac{N_t}{\tilde{N}_t}|\tilde{N}_t \geqslant 1] = \frac{1}{\mathbb{P}(\tilde{N}_t \geqslant 1)}\sum_{n\geqslant 1}\frac{1}{n}\mathbb{E}[N_t|\tilde{N}_t = n]\mathbb{P}(\tilde{N}_t = n).$$

Chapter II

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

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 the 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.

2.1 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 [21] 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 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 [55], we build a thinning algorithm which provides us with an exact simulation of an approximation of (x_t) that we denote $(\overline{x}_t)$. The process $(\overline{x}_t)$ 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

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

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

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

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

The MLMC method relies simultaneously on estimates (II.1) and (II.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 reduces 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 [45] and developed by M. Giles in [34]. The MLMC estimator has been efficiently used in various fields of numerical probability such as SDEs [34], Markov chains [3], [4], [40], Lévy processes [30], jump diffusions [82], [24], [25] or nested Monte Carlo [54], [37]. See [33] 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 $X^k, k \geqslant 1$ independent random variables identically distributed as X, provides an unbiased estimator. The associated L^2 - error satisfies $\| Y - \mathbb{E}[X] \|_2^2 = \mathrm{Var}(Y) = \frac{1}{N}\mathrm{Var}(X)$. If we quantify the precision by the L^2 - error, then a user-prescribed precision $\epsilon^2 > 0$ is achieved for $N = O(\epsilon^{-2})$ 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 $h \to 0$ in the following sense

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

and

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

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

$$Y = \frac{1}{N}\sum_{k=1}^{N} X_h^k, \tag{II.5}$$

where $X_h^k, k \geqslant 1$ are independent random variables identically distributed as X_h. The bias and the variance of the estimator (II.5) are respectively given by $\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[X_h] - \mathbb{E}[X] \simeq c_1 h^{\alpha}$ and $\mathrm{Var}(Y) = \frac{1}{N}\mathrm{Var}(X_h)$. From the strong estimate (II.4) we have that $\mathrm{Var}(X_h) \to \mathrm{Var}(X)$ as $h \to 0$ so that $\mathrm{Var}(X_h)$ is asymptotically a constant independent of h. If as above we quantify the precision by the L^2 - error and use that $\| Y - \mathbb{E}[X] \|_2^2 = (\mathbb{E}[Y] - \mathbb{E}[X])^2 + \mathrm{Var}(Y)$, we obtain that the estimator (II.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 (II.4) in order to reduce the global complexity. Let us fix $L \geqslant 2$ and consider for $l \in \{1, \dots, L\}$ a geometrically decreasing sequence $(h_l, 1 \leqslant l \leqslant L)$ 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 (II.3), the quantity $\mathbb{E}[X_{h_L}]$ approximates $\mathbb{E}[X]$. Using the linearity of the expectation the quantity $\mathbb{E}[X_{h_L}]$ can be decomposed over the levels $l \in \{1, \dots, L\}$ as follows

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

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 $N_l \geqslant 1$ 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

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

where $(X_{h^*}^k, k \geqslant 1)$ is a sequence of independent and identically distributed random variables distributed as X_{h^*} and $\left((X_{h_l}^k, X_{h_{l-1}}^k), k \geqslant 1\right)$ 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 [34] or [54], that given a precision $\epsilon > 0$ and provided that the family $(X_h, h > 0)$ satisfies the strong and weak error estimates (II.4) and (II.3), the multilevel estimator (II.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 β. Finally, let us mention that in the case $\beta > 1$ it possible to build an unbiased multilevel estimator, see [39].

Estimates (II.1) and (II.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 $X_h = F(\overline{x}_T)$ for $h > 0$ and F a smooth function, we define a MLMC estimator of $\mathbb{E}[F(x_T)]$ just as in (II.7) (noted Y^{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 (II.1). In order to improve the convergence rate (to increase the parameter β) in (II.1), we show that for a given PDMP (x_t) we have the following auxiliary representation

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

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 (II.8) is inspired by the change of probability introduced in [82] and is actually valid for a general PDP (Proposition 2.2.2) so that $\mathbb{E}[F(\overline{x}_T)] = \mathbb{E}[F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T]$ where $(\underline{\tilde{x}}_t)$ is the Euler scheme corresponding to $(\tilde{x}_t)$ and $(\underline{\tilde{R}}_t)$ is a process which depends on $(\underline{\tilde{x}}_t, t \in [0,T])$. Letting $X = F(\tilde{x}_T)\tilde{R}_T$ and $X_h = F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T$ we define a second MLMC estimator (noted $\tilde{Y}^{\mathrm{MLMC}}$) where now the discrete components of the Euler schemes $(\underline{\tilde{x}}_t)$ involved at the level l always jump in the same states and at the same times. To sum up, the first MLMC estimator we consider (Y^{MLMC}) derives from (II.6) where the corrective term at level l is $\mathbb{E}[F(\overline{x}_T^{h_l}) - F(\overline{x}_T^{h_{l-1}})]$ whereas the corrective term of the second estimator ($\tilde{Y}^{\mathrm{MLMC}}$) is $\mathbb{E}[F(\underline{\tilde{x}}_T^{h_l})\underline{\tilde{R}}_T^{h_l} - F(\underline{\tilde{x}}_T^{h_{l-1}})\underline{\tilde{R}}_T^{h_{l-1}}]$. For readability, we no longer write the dependence of the approximations on the time step. For the processes $(F(\underline{\tilde{x}}_t)\underline{\tilde{R}}_t)$ and $(F(\tilde{x}_t)\tilde{R}_t)$ we show the following strong estimate

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

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 [67], which takes into account the precise description of the ionic channels and in which the flows are not explicit. Let us mention [5] 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 [63] 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.

2.2 Piecewise Deterministic Process by thinning

2.2.1 Construction

In this section we introduce the setting and recall some results on the thinning method from our previous paper [55]. Let $E := \Theta \times \mathbb{R}^d$ where Θ is a finite or countable set and $d \geqslant 1$. A piecewise deterministic process (PDP) is defined from the following characteristics

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

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

$$Q(x, A \times B) = Q(x, A)\delta_\nu(B). \tag{II.9}$$

Then it holds that for all measurable function $f : E \to \mathbb{R}$, for all $x = (\theta, \nu) \in E$ and for all $t \geqslant 0$

$$\int_E f(i, z)Q((\theta, \Phi_\theta(t, \nu)), didz) = \sum_{i \in \Theta} f(i, \Phi_\theta(t, \nu))Q((\theta, \Phi_\theta(t, \nu)), i).$$

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

In [55] we considered a general upper bound λ^*. In the present paper λ^* is constant (see Assumption 2.2.1). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which we define

1. an homogeneous Poisson process $(N_t^*, t \geqslant 0)$ with intensity λ^* (given in Assumption 2.2.1) whose successive jump times are denoted $(T_k^*, k \geqslant 1)$. We set $T_0^* = 0$.
2. two sequences of iid random variables with uniform distribution on $[0, 1]$, $(U_k, k \geqslant 1)$ and $(V_k, k \geqslant 1)$ independent of each other and independent of $(T_k^*, k \geqslant 1)$.

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

$$T_1 := T_{\tau_1}^*, \tag{II.10}$$

where

$$\tau_1 := \inf \left\{ k > 0 : U_k \lambda^* \leqslant \lambda(\theta_0, \Phi_{\theta_0}(T_k^*, \nu_0)) \right\}. \tag{II.11}$$

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

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

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

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

For all $x \in E$, $H(x,.)$ is the inverse of the cumulative distribution function of $Q(x,.)$ (see for example [26]). Then, we construct (θ_1, ν_1) from the uniform random variable V_1 and the function H as follows

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

Thus, the distribution of (θ_1, ν_1) given $(\tau_1, (T^*_k)_{k\leqslant\tau_1})$ is $Q((\theta_0, \Phi_{\theta_0}(T^*_{\tau_1}, \nu_0)), .)$ or in view of (II.9),

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

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

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

where

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

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

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

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

$$x_t := \left(\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n)\right), \qquad t \in [T_n, T_{n+1}[. \tag{II.13}$$

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

2.2.2 Approximation of a PDP

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

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

To the family $(\overline{\Phi}_\theta)$ we can associate a PDP constructed as above that we denote $(\overline{x}_t)$. We emphasize that there is a positive probability that (x_t) and $(\overline{x}_t)$ jump at different times and/or in different states even if they are both constructed from the same data (N_t^*), (U_k) and (V_k). However if the characteristics $(\Phi,\tilde{\lambda},\tilde{Q})$ of a PDP $(\tilde{x}_t)$ are such that $\tilde{\lambda}$ and $\tilde{Q}$ depend only on θ, that is $\tilde{\lambda}(x)=\tilde{\lambda}(\theta)$ and $\tilde{Q}(x,.)=\tilde{Q}(\theta,.)$ for all $x=(\theta,\nu)\in E$, then its embedded Markov chain $(\tilde{T}_n,(\tilde{\theta}_n,\tilde{\nu}_n),n\geqslant 0)$ is such that $(\tilde{\theta}_n,n\geqslant 0)$ is an autonomous Markov chain with kernel $\tilde{Q}$ and $(\tilde{T}_n,n\geqslant 0)$ is a counting process with intensity $\tilde{\lambda}_t=\sum_{n\geqslant 0}\tilde{\lambda}(\tilde{\theta}_n)\mathbb{1}_{\tilde{T}_n\leqslant t<\tilde{T}_{n+1}}$. In particular, both $(\tilde{\theta}_n)$ and $(\tilde{\tau}_n)$ do not depend on Φ. The particular form of the characteristics $\tilde{\lambda}$ and $\tilde{Q}$ implies that the PDP $(\tilde{x}_t)$ and its approximation $(\underline{\tilde{x}_t})$ are correlated via the same process $(\tilde{\tau}_n,\tilde{\theta}_n)$. In other words, these processes always jump exactly at the same times and their θ-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 (II.14).

Lemma 2.2.1. *Let (Φ_θ) and $(\overline{\Phi}_\theta)$ satisfying* (II.14)*. Let $(t_n,n\geqslant 0)$ be an increasing sequence of non-negative real numbers with $t_0=0$ and let $(\alpha_n,n\geqslant 0)$ be a sequence of Θ-valued components. For a given $\nu\in\mathbb{R}$ let us define iteratively the sequences $(\beta_n,n\geqslant 0)$ and $(\overline{\beta}_n,n\geqslant 0)$ as follows*

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

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

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

where C_1 and C_2 are positive constants independent of h.

Proof of Lemma 2.2.1. Let $n\geqslant 1$. From the estimate (II.14), we have for all $k\leqslant n$

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

and therefore

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

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

$$\left|\overline{\beta}_n-\beta_n\right|\leqslant e^{C_1 t_n} n C_2 h.$$

□

2.2.3 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 2.2.1. Consider a family of vector fields $(f_\theta, \theta \in \Theta)$ satisfying

Assumption 2.2.2. *For all $\theta \in \Theta$, the function $f_\theta : \mathbb{R} \to \mathbb{R}$ is bounded and Lipschitz with constant L independent of θ.*

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

$$\begin{cases} \frac{dy(t)}{dt} = f_\theta\left(y(t)\right), \\ y(0) = \nu, \end{cases} \tag{II.15}$$

then the corresponding PDP is Markov since ϕ satisfies the semi-group property which reads $\phi_\theta(t + s, \nu) = \phi_\theta(t, \phi_\theta(s, \nu))$ for all $t, s \geqslant 0$ and for all $(\theta, \nu) \in E$. In this case, the process (x_t) is a piecewise deterministic Markov process (see [22] or [51]).

Let $h > 0$. We approximate the solution of (II.15) by the Euler scheme with time step h. First, we define the Euler subdivision of $[0, +\infty[$ with time step h, noted $(\bar{t}_i, i \geqslant 0)$, by $\bar{t}_i := ih$.
Then, for all $x = (\theta, \nu) \in E$, we define the sequence $(\overline{y}_i(x), i \geqslant 0)$, the classical Euler scheme, iteratively by

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

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

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

We construct the approximating process $(\overline{x}_t)$ as follows. Its continuous component starts from ν_0 at time 0 and follows the flow $\overline{\phi}_{\theta_0}(t, \nu_0)$ until the first jump time $\overline{T}_1$ that we construct by (II.10) and (II.11) of section 2.2.1 where we replace $\Phi_{\theta_0}(T_k^*, \nu_0)$ by $\overline{\phi}_{\theta_0}(T_k^*, \nu_0)$. At time $\overline{T}_1$ the continuous component of $\overline{x}_{\overline{T}_1}$ is equal to $\overline{\phi}_{\theta_0}(\overline{T}_1, \nu_0) := \overline{\nu}_1$ since there is no jump in the continuous component. The discrete component jumps to $\overline{\theta}_1$. We iterate this procedure with the new flow $\overline{\phi}_{\overline{\theta}_1}(t - \overline{T}_1, \overline{\nu}_1)$ until the next jump time $\overline{T}_2$ given by (II.10) and (II.11) with $\overline{\phi}_{\overline{\theta}_1}(T_k^* - \overline{T}_1, \overline{\nu}_1)$ and so on. We proceed by iteration to construct $(\overline{x}_t)$ on $[0, T]$.
Consequently, the discretisation grid for $(\overline{x}_t)$ on the interval $[0, T]$ is random and is formed by the points $\overline{T}_n + kh$ for $n = 0, \ldots, \overline{N}_T$ and $k = 0, \ldots, \lfloor (\overline{T}_{n+1} \wedge T - \overline{T}_n)/h \rfloor$. This differs from the SDE case where the classical grid is fixed.

By classical results of numerical analysis (see [44] for example), the continuous Euler scheme (II.16) (also called Euler polygon) satisfies estimate (II.14). If we choose $\Phi_\theta = \overline{\phi}_\theta$ in the above construction then the corresponding PDP $(\overline{x}_t)$ is not Markov since the functions $\overline{\phi}_\theta(., \nu)$ do not satisfy the semi-group property (see [51]).

2.2.4 Thinning representation for the marginal distribution of a PDP

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

$$K\Big((t,\theta,\nu), dudjdz\Big) := \mathbb{1}_{u \geqslant t}\, \lambda(\theta, \Phi_\theta(u-t,\nu)) e^{-\int_0^{u-t} \lambda(\theta,\Phi_\theta(s,\nu))ds} Q((\theta, \Phi_\theta(u-t,\nu)), djdz)du\,. \tag{II.17}$$

That is to say, for $n \geqslant 0$, the law of the random variable T_n given $\mathcal{F}_{n-1}$ admits the density given for $u \geqslant 0$ by

$$\mathbb{1}_{u \geqslant T_n} \lambda(\theta_{n-1}, \Phi_{\theta_{n-1}}(u - T_n, \nu_{n-1})) e^{-\int_0^{u-T_n} \lambda(\theta_{n-1},\Phi(s,\nu_{n-1}))ds}, \tag{II.18}$$

and the law of (θ_n, ν_n) knowing $\mathcal{F}_{n-1}$ and T_n is given, in view of (II.9), by the following probability measure

$$Q\left((\theta_{n-1}, \Phi_{\theta_{n-1}}(T_n - T_{n-1}, \nu_{n-1})), .\right) \delta_{\Phi_{\theta_{n-1}}(T_n - T_{n-1}, \nu_{n-1})}(.). \tag{II.19}$$

The marginal distribution of x_t can then be expressed for $n \in \mathbb{N}$, for fixed $x_0 = x \in E$ and for any bounded measurable function g using (II.13), the intensity λ via (II.18) as follows

$$\begin{aligned}
\mathbb{E}\left[g(x_t)\mathbb{1}_{N_t=n}\right] &= \mathbb{E}\left[g(\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n))\mathbb{1}_{N_t=n}\right] \\
&= \mathbb{E}\left[g(\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n))\mathbb{1}_{T_n \leqslant t}\mathbb{E}[\mathbb{1}_{T_{n+1}>t}|\mathcal{F}_n]\right] \\
&= \mathbb{E}\left[g(\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n))\mathbb{1}_{T_n \leqslant t} e^{-\int_0^{t-T_n} \lambda(\theta_n,\Phi_{\theta_n}(s,\nu_n))ds}\right]
\end{aligned}$$

Moreover, using (II.19), we get

$$\begin{aligned}
\mathbb{E}\left[g(x_t)\mathbb{1}_{N_t=n}\right] &= \mathbb{E}\left[\mathbb{1}_{T_n \leqslant t}\mathbb{E}\left[g(\theta_n, \Phi_{\theta_n}(t - T_n, \nu_n)) e^{-\int_0^{t-T_n} \lambda(\theta_n,\Phi_{\theta_n}(s,\nu_n))ds} | \mathcal{F}_{n-1}, T_n\right]\right] \\
&= \sum_{i \in \Theta} \mathbb{E}[Q\left((\theta_{n-1}, \nu_n), i\right) g(i, \Phi_i(t - T_n, \nu_n))\mathbb{1}_{T_n \leqslant t} e^{-\int_0^{t-T_n} \lambda(i,\Phi_i(s,\nu_n))ds}]
\end{aligned} \tag{II.20}$$

where, for short, $\nu_n = \Phi_{\theta_{n-1}}(T_n - T_{n-1}, \nu_{n-1})$. We can iterate on (II.20) using successive conditioning and the kernel (II.17) to finally obtain

$$\begin{aligned}
\mathbb{E}\left[g(x_t)\mathbb{1}_{N_t=n}\right] = \int_{(\mathbb{R}_+ \times E)^n} & g(i_n, \Phi_{i_n}(t - t_n, z_n)) e^{-\int_0^{t-t_n} \lambda(i_n,\Phi_{i_n}(s,z_n))ds} \\
& K((0,x), dt_1 di_1 dz_1)...K((t_{n-1}, i_{n-1}, z_{n-1}), dt_n di_n dz_n).
\end{aligned}$$

However since we have constructed (x_t) by thinning, we would prefer to express the distribution of x_t using the upper bound λ^*, the Poisson process $(N_t^*, t \geqslant 0)$ and the sequences $(U_k, k \in \mathbb{N})$, $(V_k, k \in \mathbb{N})$. In Proposition 2.2.1, we give another representation of (II.20). Instead of using the conditional density of the jump times (II.18), we focus on the random indexes τ_n (recall that $T_n = T^*_{\tau_n}$) to make appears the acceptance and reject

probabilities ($\lambda(.)/\lambda^*$ and $1-\lambda(.)/\lambda^*$). The product term which appear in the expectation in the right hand side of the equality in Proposition 2.2.1 should be interpreted as the survival function of T_{n+1} in (II.20). Indeed, consider for example the first jump time T_1 and that $x_0 = (\theta, \nu) \in E$ is fixed. Using (II.18), we have

$$\mathbb{P}(T_1 > t) = e^{-\int_0^t \lambda(\theta, \Phi_\theta(s,\nu))ds}.$$

Moreover, we have

$$\mathbb{P}(T_1 > t) = \mathbb{E}\left[\prod_{j=1}^{N_t^*}(1 - \frac{\lambda(\theta, \Phi_\theta(T_j^*, \nu))}{\lambda^*})\right].$$

In order to derive the above equality, we use that

$$\begin{aligned}\{T_1 > t\} &= \{T_{\tau_1}^* > t\} \\ &= \{\tau_1 > N_t^*\} \\ &= \bigcup_{p\geqslant 0}\{N_t^* = p, U_1 > \frac{\lambda(\theta, \Phi_\theta(T_1^*, \nu))}{\lambda^*}, \ldots, U_p > \frac{\lambda(\theta, \Phi_\theta(T_p^*, \nu))}{\lambda^*}\}\end{aligned}$$

and the fact that the sequence $(U_k, k \geqslant 0)$ is independent of the Poisson process N^*. Then, using the partition $\cup_{p\geqslant 0}\{N_t^* = p\}$, the fact that $\{N_t^* = p\} = \{T_p^* \leqslant t < T_{p+1}^*\}$, the density of $(T_1^*, \ldots, T_p^*)$ given T_{p+1}^* and the fact that T_{p+1}^* is gamma distributed, we are able to show that

$$\mathbb{E}\left[\prod_{j=1}^{N_t^*}(1 - \frac{\lambda(\theta, \Phi_\theta(T_j^*, \nu))}{\lambda^*})\right] = e^{-\int_0^t \lambda(\theta, \Phi_\theta(s,\nu))ds}.$$

Proposition 2.2.1 allows us to identify corrective terms (which can be interpreted as Radon-Nikodym derivatives) in order to express the marginal distribution of a PDP (x_t) with characteristics (Φ, λ, Q) in term of a PDP $(\tilde{x}_t)$ with the same flow Φ and simplified jump characteristics $(\tilde{\lambda}, \tilde{Q})$. In Proposition 2.2.2 we state this change of representation and we define the corrective terms. Note that we modify the intensity of a PDP through the acceptance and reject probabilities. The new proposed representation (in Proposition 2.2.2) will be used to construct an efficient MLMC estimator in section 2.5.2.

Proposition 2.2.1. *Let $(x_t, t \in [0,T])$ be a PDP with characteristics (Φ, λ, Q) constructed in section 2.2.1. Let n and m be integers such that $n \leqslant m$ and let $p_1, \ldots, p_n$ be an ordered sequence of integers. Then, for all bounded measurable function g we have*

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

The following proposition and its corollaries will be useful in section 2.3. In their statements $(x_t, t \in [0,T])$ and $(\tilde{x}_t, t \in [0,T])$ are PDPs constructed in section 2.2.1 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 [82] 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 2.2.2. *Let us denote by (Φ, λ, 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 θ, that $\tilde{Q}$ is always positive and $0 < \tilde{\lambda}(\theta) < \lambda^*$ for all $\theta \in \Theta$. For all integer n, let us define on the event $\{\tilde{N}_t = n\}$,*

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

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

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

Then, for all $n \geqslant 0$ 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 2.2.1. *Under the assumptions of Proposition 2.2.2, 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 2.2.1. *Proposition 2.2.2 looks like a Girsanov theorem (see [68]) however we do not use the martingale theory here.*

Remark 2.2.2. *We have chosen to state Proposition 2.2.2 with a PDP $(\tilde{x}_t)$ whose intensity and transition measure only depend on θ 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 2.5.4.*

Corollary 2.2.2. *Let (Φ,λ,Q) (resp. $(\tilde\Phi,\lambda,Q)$) be the set of characteristics of (x_t) (resp. $(\tilde x_t)$). We assume that Q is always positive and that $0<\lambda(x)<\lambda^*$ for all $x\in E$. Let (μ_n) be the sequence defined by $\mu_0=\nu$ and $\mu_n=\tilde\Phi_{\theta_{n-1}}(T_n-T_{n-1},\mu_{n-1})$ for $n\geqslant 1$. For all integer n, let us define on the event $\{N_t=n\}$,*

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

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

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

$$\tilde Z_0=\left(\frac{\lambda\big(\theta_0,\Phi_{\theta_0}(T_{\tau_1}^*,\nu_0)\big)}{\lambda^*}\prod_{q=1}^{\tau_1-1}\left(1-\frac{\lambda\big(\theta_0,\Phi_{\theta_0}(T_q^*,\nu_0)\big)}{\lambda^*}\right)\right)^{-1}$$
$$\frac{\lambda\big(\theta_0,\tilde\Phi_{\theta_0}(T_{\tau_1}^*,\mu_0)\big)}{\lambda^*}\prod_{q=1}^{\tau_1-1}\left(1-\frac{\lambda\big(\theta_0,\tilde\Phi_{\theta_0}(T_q^*,\mu_0)\big)}{\lambda^*}\right),$$

$$\tilde R_n=\tilde Z_n\prod_{\ell=0}^{n-1}\tilde Z_\ell.$$

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

$$\mathbb{E}[g\left(\theta_n,\tilde\Phi_{\theta_n}(t-T_n,\mu_n)\right)\,\tilde R_n\,\mathbb{1}_{\{N_t=n\}}]=\mathbb{E}[g(\tilde x_t)\,\mathbb{1}_{\{\tilde N_t=n\}}].$$

Proof of Proposition 2.2.1. It holds that $\{N_t=n,\tau_i=p_i,\,1\leqslant i\leqslant n\}\subset\{N_t^*\geqslant p_n\}$. Then

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

The set $\{N_t=n,\tau_i=p_i,\,1\leqslant i\leqslant n,N_t^*=m\}$ is equivalent to the following
- $N_t^*=m$,
- among the times $T_\ell^*,1\leqslant\ell\leqslant m$ exactly n are accepted by the thinning method they are the $T_{p_i}^*,1\leqslant i\leqslant n$, all the others are rejected.
We proceed by induction starting from the fact that all the T_q^*, $p_n+1\leqslant q\leqslant m$ are rejected which corresponds to the event

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

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

$$\begin{aligned}&\mathbb{E}[g(x_t)\mathbb{1}_{\{N_t=n,\tau_i=p_i,\,1\leqslant i\leqslant n,\,N_t^*=m\}}]\\&=\mathbb{E}[\sum_{\theta\in\Theta}Q(x^-_{T^*_{p_{n-1}}},\theta)\,g(\theta,\Phi_\theta(t-T^*_{p_n},\nu_n))\\&F(\theta,U_j,1\leqslant j\leqslant m,\,T_\ell^*,1\leqslant\ell\leqslant m+1,V_k,1\leqslant k\leqslant n-1)],\end{aligned}\tag{II.21}$$

with

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

In (II.21) the random variables $(U_q,\,p_n+1\leqslant q\leqslant m)$ are independent of the vector $(T_i^*,1\leqslant i\leqslant m+1,U_j,1\leqslant j\leqslant p_n,V_k,1\leqslant k\leqslant n-1)$. Conditioning by this vector we obtain

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

We can iterate on the latter form by first conditioning V_{n-1} by all the other r.v. and then conditioning $(U_q,\,p_{n-1}+1\leqslant q\leqslant p_n)$ 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 $p_{n-1}+1\leqslant q\leqslant p_n-1$ whereas U_{p_n} corresponds to an acceptation. So that the next step yields

$$\begin{aligned}&\mathbb{E}[g(x_t)\mathbb{1}_{\{N_t=n,\tau_i=p_i,\,1\leqslant i\leqslant n,\,N_t^*=m\}}]\\=\;&\sum_{\alpha\in\Theta}\sum_{\theta\in\Theta}\mathbb{E}[Q(x^-_{T^*_{p_{n-2}}},\alpha)Q\left((\alpha,\nu_n),\theta\right)\,g(\theta,\Phi_\theta(t-T^*_{p_n},\nu_n))\mathbb{1}_{\{N_t^*=m,\tau_i=p_i,\,1\leqslant i\leqslant n-1\}}\\&\frac{\lambda(\alpha,\Phi_\alpha(T^*_{p_n}-T^*_{p_{n-1}},\nu_{n-1}))}{\lambda^*}\prod_{q=p_{n-1}+1}^{p_n-1}(1-\frac{\lambda(\alpha,\Phi_\alpha(T_q^*-T^*_{p_{n-1}},\nu_{n-1}))}{\lambda^*})\\&\prod_{q=p_n+1}^{m}(1-\frac{\lambda(\theta,\Phi_\theta(T_q^*-T^*_{p_n},\nu_n))}{\lambda^*})],\end{aligned}\tag{II.22}$$

where we write ν_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 $\mathbb{E}(g(x_t) f(\theta_i, \nu_i, 1 \leqslant i \leqslant n-1, \theta_n, \nu_n, T^*_k, 1 \leqslant k \leqslant m)\, \mathbb{1}_{\{N_t = n, \tau_i = p_i,\, 1 \leqslant i \leqslant n,\, N^*_t = m\}})$ and provide

$$
\begin{aligned}
&\mathbb{E}[g(x_t) f(\theta_i, \nu_i, 1 \leqslant i \leqslant n-1, \theta_n, \nu_n, T^*_k, 1 \leqslant k \leqslant m)\, \mathbb{1}_{\{N_t = n, \tau_i = p_i,\, 1 \leqslant i \leqslant n,\, N^*_t = m\}}] \\
&= \sum_{\theta \in \Theta} \mathbb{E}[Q(x^-_{T^*_{p_{n-1}}}, \theta) g(\theta, \Phi_\theta(t - T^*_{p_n}, \nu_n)) f(\theta_i, \nu_i, 1 \leqslant i \leqslant n-1, \theta, \nu_n, T^*_k, 1 \leqslant k \leqslant m) \\
&\mathbb{1}_{\{N^*_t = m, \tau_i = p_i,\, 1 \leqslant i \leqslant n\}} \prod_{q = p_n + 1}^{m} \Big(1 - \frac{\lambda(\theta, \Phi_\theta(T^*_q - T^*_{p_n}, \nu_n))}{\lambda^*}\Big)]. \qquad \text{(II.23)}
\end{aligned}
$$

□

We prove below Proposition 2.2.2. The other statements can be proved analogously.

Proof of Proposition 2.2.2. By assumption the (jump) characteristics $(\tilde\lambda, \tilde Q)$ of $(\tilde x_t)$ depend only on θ. Let $p_1 < p_2 < ... < p_n \leqslant m$. Applying the same arguments as in (II.23) to $(\tilde x_t)$ and using the definitions of $\tilde Z_\ell$, $0 \leqslant \ell \leqslant n$ and $\tilde R_n$ we obtain,

$$
\begin{aligned}
&\mathbb{E}[g(\tilde x_t)\, \tilde R_n\, \mathbb{1}_{\{\tilde N_t = n, \tilde\tau_i = p_i, 1 \leqslant i \leqslant n, N^*_t = m\}}] \\
&= \sum_{\theta \in \Theta} \mathbb{E}[\tilde Q(\tilde\theta_{n-1}, \theta)\, g(\theta, \Phi_\theta(t - T^*_{p_n}, \tilde\nu_n))\, \tilde Z_n \prod_{\ell=0}^{n-1} \tilde Z_\ell\, \mathbb{1}_{\{N^*_t = m, \tilde\tau_i = p_i,\, 1 \leqslant i \leqslant n\}}]\, \Big(1 - \frac{\tilde\lambda(\theta)}{\lambda^*}\Big)^{m - p_n} \\
&= \sum_{\theta \in \Theta} \mathbb{E}[\tilde Q(\tilde\theta_{n-1}, \theta)\, g(\theta, \Phi_\theta(t - T^*_{p_n}, \tilde\nu_n)) \prod_{\ell=0}^{n-1} \tilde Z_\ell\, \mathbb{1}_{\{N^*_t = m, \tilde\tau_i = p_i,\, 1 \leqslant i \leqslant n\}}\, \Big(1 - \frac{\tilde\lambda(\theta)}{\lambda^*}\Big)^{m - p_n} \\
&\Big(\Big(1 - \frac{\tilde\lambda(\theta)}{\lambda^*}\Big)^{m - p_n}\Big)^{-1} \frac{Q(\tilde x^-_{T^*_{p_{n-1}}}, \theta)}{\tilde Q(\tilde\theta_{n-1}, \theta)} \prod_{q = p_n + 1}^{m} \Big(1 - \frac{\lambda(\theta, \Phi_\theta(T^*_q - T^*_{p_n}, \tilde\nu_n))}{\lambda^*}\Big)] \\
&= \sum_{\theta \in \Theta} \mathbb{E}[Q(\tilde x^-_{T^*_{p_{n-1}}}, \theta)\, g(\theta, \Phi_\theta(t - T^*_{p_n}, \tilde\nu_n))\, \tilde Z_{n-1} \prod_{\ell=0}^{n-2} \tilde Z_\ell\, \mathbb{1}_{\{N^*_t = m, \tilde\tau_i = p_i,\, 1 \leqslant i \leqslant n\}} \\
&\prod_{q = p_n + 1}^{m} \Big(1 - \frac{\lambda(\theta, \Phi_\theta(T^*_q - T^*_{p_n}, \tilde\nu_n))}{\lambda^*}\Big)].
\end{aligned}
$$

We iterate the previous argument based on the use of (II.23) and we use the definition of $\tilde Z_{n-1}$ to obtain

$$
\begin{aligned}
&\mathbb{E}[g(\tilde x_t) \tilde R_n \mathbb{1}_{\{\tilde N_t = n, \tilde\tau_i = p_i,\, 1 \leqslant i \leqslant n,\, N^*_t = m\}}] \\
= \;& \sum_{\alpha \in \Theta} \sum_{\theta \in \Theta} \mathbb{E}[Q(\tilde x^-_{T^*_{p_{n-2}}}, \alpha) Q((\alpha, \tilde\nu_n), \theta)\, g(\theta, \Phi_\theta(t - T^*_{p_n}, \tilde\nu_n)) \\
& \prod_{\ell=0}^{n-2} \tilde Z_\ell\, \mathbb{1}_{\{N^*_t = m, \tilde\tau_i = p_i,\, 1 \leqslant i \leqslant n-1\}} \prod_{q = p_n + 1}^{m} \Big(1 - \frac{\lambda(\theta, \Phi_\theta(T^*_q - T^*_{p_n}, \tilde\nu_n))}{\lambda^*}\Big) \\
& \frac{\lambda(\alpha, \Phi_\alpha(T^*_{p_n} - T^*_{p_{n-1}}, \tilde\nu_{n-1}))}{\lambda^*} \prod_{q = p_{n-1} + 1}^{p_n - 1} \Big(1 - \frac{\lambda(\alpha, \Phi_\alpha(T^*_q - T^*_{p_{n-1}}, \tilde\nu_{n-1}))}{\lambda^*}\Big)],
\end{aligned}
$$

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 (II.22) and using an induction we conclude that

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

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

2.3 Strong error estimates

In this section we are interested in strong error estimates that we define as squared L^2 errors (mean squared errors) in order to respect the MLMC framework introduced in [54]. Below, we state the main assumptions and theorems of this section, the proofs are given in sections 2.3.2, 2.3.3 respectively.

Assumption 2.3.1. *For all $\theta \in \Theta$ and for all $A \in \mathcal{B}(\Theta)$, the functions $\nu \mapsto \lambda(\theta, \nu)$ and $\nu \mapsto Q((\theta, \nu), A)$ are Lipschitz with constants $L_\lambda > 0$, $L_Q > 0$ respectively independent of θ.*

Theorem 2.3.1. *Let Φ_θ and $\overline{\Phi}_\theta$ satisfying* (II.14) *and let $(x_t, t \in [0,T])$ and $(\overline{x}_t, t \in [0,T])$ be the corresponding PDPs constructed in section 2.2.1 with $x_0 = \overline{x}_0 = x$ for some $x \in E$. Assume that Θ is finite and that λ and Q satisfy Assumption 2.3.1. Then, for all bounded functions $F : E \to \mathbb{R}$ such that for all $\theta \in \Theta$ the function $\nu \mapsto F(\theta, \nu)$ is L_F-Lipschitz where L_F is positive and independent of θ, there exists constants $V_1 > 0$ and $V_2 > 0$ independent of the time step h such that*

$$\mathbb{E}\left[|F(\overline{x}_T) - F(x_T)|^2\right] \leqslant V_1 h + V_2 h^2.$$

Remark 2.3.1. *When the numerical scheme $\overline{\Phi}_\theta$ is of order $p \geqslant 1$, which means that $\sup_{t\in[0,T]} |\Phi_\theta(t, \nu_1) - \overline{\Phi}_\theta(t, \nu_2)| \leqslant e^{C_1 T}|\nu_1 - \nu_2| + C_2 h^p$ we have $\mathbb{E}\left[|F(\overline{x}_T) - F(x_T)|^2\right] \leqslant V_1 h^p + V_2 h^{2p}$.*

Assumption 2.3.2. *There exist positive constants ρ, $\tilde{\lambda}_{\min}$, $\tilde{\lambda}_{\max}$ such that for all $(i,j) \in \Theta^2$, $\rho \leqslant \tilde{Q}(i,j)$ and $\tilde{\lambda}_{\min} \leqslant \tilde{\lambda}(i) \leqslant \tilde{\lambda}_{\max} < \lambda^*$.*

Theorem 2.3.2. *Let Φ_θ and $\overline{\Phi}_\theta$ satisfying* (II.14) *and let $(\tilde{x}_t, t \in [0,T])$ and $(\underline{\tilde{x}}_t, t \in [0,T])$ be the corresponding PDPs constructed in section 2.2.1 with $\underline{\tilde{x}}_0 = \tilde{x}_0 = x$ for some $x \in E$. Let $(\tilde{R}_t, t \in [0,T])$ and $(\underline{\tilde{R}}_t, t \in [0,T])$ be defined as in Corollary 2.2.1. Under assumptions 2.3.1 and 2.3.2 and for all bounded functions $F : E \to \mathbb{R}$ such that for all $\theta \in \Theta$ the function $\nu \mapsto F(\theta, \nu)$ is L_F-Lipschitz ($L_F > 0$), there exists a positive constant $\tilde{V}_1$ independent of the time step h such that*

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

We now introduce the random variable $\overline{\tau}^\dagger$ which will play an important role in the strong error estimate of Theorem 2.3.1 as well as in the identification of the coefficient c_1 in the weak error expansion in section 2.4 (see the proof of Theorem 2.4.1 in section 2.4.2).

Definition 2.3.1. *Let us define* $\overline{\tau}^{\dagger} := \inf\left\{k > 0 : (\tau_k, \theta_k) \neq (\overline{\tau}_k, \overline{\theta}_k)\right\}$.

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

$$\{\min(T_{\overline{\tau}^{\dagger}}, \overline{T}_{\overline{\tau}^{\dagger}}) > T\} = \left\{N_T = \overline{N}_T, (T_1, \theta_1) = (\overline{T}_1, \overline{\theta}_1), \ldots, (T_{N_T}, \theta_{N_T}) = (\overline{T}_{\overline{N}_T}, \overline{\theta}_{\overline{N}_T})\right\}, \tag{II.24}$$

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

2.3.1 Preliminary lemmas

In this section we start with two lemmas which will be useful to prove Theorems 2.3.2 and 2.3.3.

Lemma 2.3.1. *Let K be a finite set. We denote by $|K|$ the cardinal of K and for $i = 1, \ldots, |K|$ we denote by k_i its elements. Let $(p_i, 1 \leqslant i \leqslant |K|)$ and $(\overline{p}_i, 1 \leqslant i \leqslant |K|)$ be two probabilities on K. Let $a_j := \sum_{i=1}^{j} p_i$ and $\overline{a}_j := \sum_{i=1}^{j} \overline{p}_i$ for all $j \in \{1, \ldots, |K|\}$. By convention, we set $a_0 = \overline{a}_0 := 0$. Let X and $\overline{X}$ be two K-valued random variables defined by*

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

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

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

Proof of Lemma 2.3.1. By definition of X and $\overline{X}$ and since the intervals $]a_{j-1}, a_j] \cap]\overline{a}_{j-1}, \overline{a}_j]$ are disjoints for $j = 1, \ldots, K$, we have

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

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

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

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

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

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

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

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

□

Lemma 2.3.2. *Let $(a_n, n \geqslant 1)$ and $(b_n, n \geqslant 1)$ be two real-valued sequences. For all $n \geqslant 1$, 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 2.3.2. By induction. □

2.3.2 Proof of Theorem 2.3.1

First, we write

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

where $\overline{\tau}^\dagger$ is defined in Definition 2.3.1. The term $\overline{P}$ has the same behaviour with respect to the time step h as the probability that the discrete processes (T_n, θ_n) and $(\overline{T}_n, \overline{\theta}_n)$ differ on $[0, T]$. Moreover, the term $\overline{D}$ behaves like h^2 because the discrete processes (T_n, θ_n) and $(\overline{T}_n, \overline{\theta}_n)$ are equal on $[0, T]$. In the following we prove that $\overline{P} = O(h)$ and that $\overline{D} = O(h^2)$.

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

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

where

$$\overline{J}_k := \mathbb{E}\left[\mathbb{1}_{\min(T_k, \overline{T}_k) \leqslant T} \mathbb{1}_{\overline{\tau}^\dagger > k-1} \mathbb{1}_{\tau_k = \overline{\tau}_k} \mathbb{1}_{\theta_k \neq \overline{\theta}_k}\right], \quad \overline{I}_k := \mathbb{E}\left[\mathbb{1}_{\min(T_k, \overline{T}_k) \leqslant T} \mathbb{1}_{\overline{\tau}^\dagger > k-1} \mathbb{1}_{\tau_k \neq \overline{\tau}_k}\right]. \tag{II.25}$$

We start with $\overline{J}_k$. First note that, for $k \geqslant 1$, $\{\tau_k = \overline{\tau}_k\} = \{T_k = \overline{T}_k\}$ and that on the event $\{T_k = \overline{T}_k\}$, we have $\min(T_k, \overline{T}_k) = T_k$, so that $\overline{J}_k = \mathbb{E}\left[\mathbb{1}_{T_k \leqslant T}\mathbb{1}_{\overline{\tau}^\dagger > k-1}\mathbb{1}_{\tau_k = \overline{\tau}_k}\mathbb{1}_{\theta_k \neq \overline{\theta}_k}\right]$. We emphasize that it makes no difference in the rest of the proof if we choose $\min(T_k, \overline{T}_k) = \overline{T}_k$. Since $\{\overline{\tau}^\dagger > k-1\} = \bigcap_{i=0}^{k-1}\{(\tau_i, \theta_i) = (\overline{\tau}_i, \overline{\theta}_i)\}$, we can rewrite $\overline{J}_k$ as follows

$$\sum_{\substack{1 \leqslant p_1 < \ldots < p_k \\ \alpha_1, \ldots, \alpha_{k-1} \in \Theta}} \mathbb{E}[\mathbb{1}_{\{\tau_i = \overline{\tau}_i = p_i, 1 \leqslant i \leqslant k\}}\mathbb{1}_{\{\theta_i = \overline{\theta}_i = \alpha_i, 1 \leqslant i \leqslant k-1\}}\mathbb{1}_{T^*_{p_k} \leqslant T}\mathbb{1}_{\theta_k \neq \overline{\theta}_k}]. \tag{II.26}$$

By construction we have $\theta_k = H((\theta_{k-1}, \nu_k), V_k)$ and $\overline{\theta}_k = H((\overline{\theta}_{k-1}, \overline{\nu}_k), V_k)$. The random variable $\mathbb{1}_{\{\tau_i = \overline{\tau}_i = p_i, 1 \leqslant i \leqslant k\}}\mathbb{1}_{\{\theta_i = \overline{\theta}_i = \alpha_i, 1 \leqslant i \leqslant k-1\}}\mathbb{1}_{T^*_{p_k} \leqslant T}$ depends on the vector $(U_i, 1 \leqslant i \leqslant p_k, T^*_j, 1 \leqslant j \leqslant p_k, V_q, 1 \leqslant q \leqslant k-1)$ which is independent of V_k. Conditioning by this vector in (II.26) and applying Lemma 2.3.1 yields

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

From the definition of a_j (see (II.12)), the triangle inequality and since Q is L_Q-Lipschitz, we have $\sum_{j=1}^{|\Theta|-1}|a_j(\alpha_{k-1}, \overline{\nu}_k) - a_j(\alpha_{k-1}, \nu_k)| \leqslant \frac{(|\Theta|-1)|\Theta|}{2}L_Q|\overline{\nu}_k - \nu_k|$. Since we are on the event $\{\tau_i = \overline{\tau}_i = p_i, 1 \leqslant i \leqslant k\}\bigcap\{\theta_i = \overline{\theta}_i = \alpha_i, 1 \leqslant i \leqslant k-1\}$, the application of Lemma 2.2.1 yields $|\overline{\nu}_k - \nu_k| \leqslant e^{LT^*_{p_k}}kCh$. Thus $\overline{J}_k \leqslant C_1 h\mathbb{E}[\mathbb{1}_{T_k \leqslant T}k]$ where C_1 is a constant independent of h. Moreover, $\sum_{k \geqslant 1}\mathbb{1}_{T_k \leqslant T}k = \sum_{k=1}^{N_T}k \leqslant N_T^2$ and $\mathbb{E}[N_T^2] \leqslant \mathbb{E}[(N_T^*)^2] < +\infty$ so that $\sum_{k \geqslant 1}\overline{J}_k = O(h)$. From the definition of $\overline{I}_k$ (see (II.25)), we can write

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

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

$$\sum_{\substack{1 \leqslant p_1 < \ldots < p_k \\ \alpha_1, \ldots, \alpha_{k-1} \in \Theta}} \mathbb{E}[\mathbb{1}_{\{\tau_i = \overline{\tau}_i = p_i, 1 \leqslant i \leqslant k-1\}}\mathbb{1}_{\{\theta_i = \overline{\theta}_i = \alpha_i, 1 \leqslant i \leqslant k-1\}}\mathbb{1}_{T^*_{p_k} \leqslant T}\mathbb{1}_{\tau_k = p_k}\mathbb{1}_{p_k < \overline{\tau}_k}]. \tag{II.27}$$

In (II.27) we have $\{\tau_k = p_k\} \cap \{p_k < \overline{\tau}_k\} \subseteq \{\lambda(\alpha_{k-1}, \overline{\Phi}_{\alpha_{k-1}}(T^*_{p_k} - T^*_{p_{k-1}}, \overline{\nu}_{k-1})) < U_{p_k}\lambda^* \leqslant \lambda(\alpha_{k-1}, \Phi_{\alpha_{k-1}}(T^*_{p_k} - T^*_{p_{k-1}}, \nu_{k-1}))\}$. The random variable $\mathbb{1}_{\{\tau_i = \overline{\tau}_i = p_i, 1 \leqslant i \leqslant k-1\}}$

$\mathbb{1}_{\{\theta_i=\overline{\theta}_i=\alpha_i,1\leqslant i\leqslant k-1\}}\,\mathbb{1}_{T^*_{p_k}\leqslant T}$ depends on $(U_i, 1\leqslant i\leqslant p_{k-1}, T^*_j, 1\leqslant j\leqslant p_k, V_q, 1\leqslant q\leqslant k-1)$ which is independent of U_{p_k}. Conditioning by this vector in (II.27) yields

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

Using the Lipschitz continuity of λ then Lemma 2.2.1 we get that $\overline{I}_k^{(1)}\leqslant C_2h\mathbb{E}[\mathbb{1}_{T_k\leqslant T}k]$ where C_2 is a constant independent of h. Concerning the term $\overline{I}_k^{(2)}$, we will end with the estimate $\overline{I}_k^{(2)}\leqslant C_2h\mathbb{E}[\mathbb{1}_{\overline{T}_k\leqslant T}k]$. We conclude in the same way as in the estimation of $\overline{J}_k$ above that $\sum_{k\geqslant1}\overline{I}_k=O(h)$.

Step 2: estimation of $\overline{D}$.
Note that for $n\geqslant0$ we have $\{N_T=n\}\cap\{\min(T_{\overline{\tau}^\dagger},\overline{T}_{\overline{\tau}^\dagger})>T\}=\{N_T=n\}\cap\{\overline{N}_T=n\}\cap\{\overline{\tau}^\dagger>n\}$, where we can interchange the role of $\{N_T=n\}$ and $\{\overline{N}_T=n\}$. Thus, using the partition $\{N_T=n, n\geqslant0\}$, we have

$$\overline{D}=\sum_{n\geqslant0}\mathbb{E}\left[\mathbb{1}_{N_T=n}\mathbb{1}_{\overline{N}_T=n}\mathbb{1}_{\overline{\tau}^\dagger>n}\left|F(\theta_n,\overline{\Phi}_{\theta_n}(T-T_n,\overline{\nu}_n))-F(\theta_n,\Phi_{\theta_n}(T-T_n,\nu_n))\right|^2\right]$$

The application of the Lipschitz continuity of F and of Lemma 2.2.1 yields

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

Then, we have $\overline{D}\leqslant C_3h^2\sum_{n\geqslant0}\mathbb{E}\left[\mathbb{1}_{N_T=n}(n+1)^2\right]$ where C_3 is a constant independent of h. Since $\sum_{n\geqslant0}\mathbb{E}\left[\mathbb{1}_{N_T=n}(n+1)^2\right]=\mathbb{E}[(N_T+1)^2]\leqslant\mathbb{E}[(N_T^*+1)^2]<+\infty$, we conclude that $\overline{D}=O(h^2)$.

□

2.3.3 Proof of Theorem 2.3.2

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

$$\tilde{\mathsf{Q}}_T=\prod_{l=1}^{\tilde{N}_T}\frac{Q(\tilde{x}^-_{T^*_{\tilde{\tau}_l}},\tilde{\theta}_l)}{\tilde{Q}(\tilde{\theta}_{l-1},\tilde{\theta}_l)},\tag{II.28}$$

$$\tilde{\mathsf{S}}_T=\prod_{l=1}^{\tilde{N}_T}\frac{\lambda(\tilde{\theta}_{l-1},\Phi_{\tilde{\theta}_{l-1}}(T^*_{\tilde{\tau}_l}-T^*_{\tilde{\tau}_{l-1}},\tilde{\nu}_{l-1}))}{\lambda^*}\prod_{k=\tilde{\tau}_{l-1}+1}^{\tilde{\tau}_l}(1-\frac{\lambda(\tilde{\theta}_{l-1},\Phi_{\tilde{\theta}_{l-1}}(T^*_k-T^*_{\tilde{\tau}_{l-1}},\tilde{\nu}_{l-1}))}{\lambda^*})\tag{II.29}$$

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

$$\tilde{\mathsf{H}}_T=\prod_{l=1}^{\tilde{N}_T}\left(\frac{\tilde{\lambda}(\tilde{\theta}_{l-1})}{\lambda^*}(1-\frac{\tilde{\lambda}(\tilde{\theta}_{l-1})}{\lambda^*})^{\tilde{\tau}_l-\tilde{\tau}_{l-1}-1}\right)^{-1}\left((1-\frac{\tilde{\lambda}(\tilde{\theta}_{\tilde{N}_T})}{\lambda^*})^{N^*_T-\tilde{\tau}_{\tilde{N}_T}}\right)^{-1}.\tag{II.30}$$

Likewise we reorder the terms in $\underline{\tilde{R}}_T$ writing $\underline{\tilde{R}}_T = \underline{\tilde{\mathsf{Q}}}_T \underline{\tilde{\mathsf{S}}}_T \tilde{\mathsf{H}}_T$ where $\underline{\tilde{\mathsf{Q}}}_T$ and $\underline{\tilde{\mathsf{S}}}_T$ are defined as (II.28) and (II.29) replacing $\tilde{x}$ and Φ by $\underline{\tilde{x}}$ and $\overline{\Phi}$. Since the processes $(\tilde{\theta}_n)$ and $(\tilde{\tau}_n)$ do not depend on Φ or $\overline{\Phi}$, the term $\tilde{\mathsf{H}}$ is the same in $\tilde{R}$ and $\underline{\tilde{R}}$. To prove Theorem 2.3.2, let us decompose the problem and write

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

so that

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

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

Step 1: estimation of $\overline{C}$.
The function F being bounded we have $\overline{C} \leqslant M_F^2 \mathbb{E}\left[|\underline{\tilde{R}}_T - \tilde{R}_T|^2\right]$ where M_F is a positive constant. Moreover, for all $\theta \in \Theta$, we have $(1 - \tilde{\lambda}(\theta)/\lambda^*)^{-1} \leqslant (1 - \tilde{\lambda}_{\max}/\lambda^*)^{-1}$ and $(\tilde{\lambda}(\theta)/\lambda^*)^{-1} \leqslant (\tilde{\lambda}_{\min}/\lambda^*)^{-1}$. Thus, $\tilde{H}_T \leqslant \left(\frac{\tilde{\lambda}_{\min}}{\lambda^*}(1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*})\right)^{-N_T^*}$ and using the definition of $\tilde{R}$ and $\underline{\tilde{R}}$ (see (II.28), (II.29) and (II.30)) we can write

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

We set $\overline{J} = |\underline{\tilde{\mathsf{Q}}}_T - \tilde{\mathsf{Q}}_T|\tilde{\mathsf{S}}_T$ and $\overline{I} = |\underline{\tilde{\mathsf{S}}}_T - \tilde{\mathsf{S}}_T|\underline{\tilde{\mathsf{Q}}}_T$. To provide the desired estimate for $\overline{C}$, we proceed as follows. First, we work ω by ω to determine (random) bounds for $\overline{J}$ and $\overline{I}$ from which we deduce a (random) bound for $|\underline{\tilde{R}}_T - \tilde{R}_T|$. Finally, we take the expectation. We start with $\overline{I}$. For all $(\theta, \nu) \in E$ and for all $t \geqslant 0$ we have, from Assumption 2.2.1, that $1 - \lambda(\theta, \Phi_\theta(t, \nu))/\lambda^* \leqslant 1$ and $\lambda(\theta, \Phi_\theta(t, \nu))/\lambda^* \leqslant 1$. Then, using Lemma 2.3.2 (twice) we have

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

Using the Lipschitz continuity of λ and Lemma 2.2.1, we find that, for all $l = 1, \ldots, \tilde{N}_T + 1$ and $k = \tilde{\tau}_{l-1} + 1, \ldots, \tilde{\tau}_l \wedge N_T^*$,

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

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

$$\overline{I} \leqslant \rho^{-N_T^*} N_T^*(N_T^* + 1)^2 C_1 h. \tag{II.31}$$

Now, consider $\overline{J}$. Note that from Assumption 2.2.1 we have $\tilde{\mathsf{S}}_T \leqslant 1$. We use the same type of arguments as for $\overline{I}$. That is, we successively use Lemma 2.3.2, the Lipschitz continuity of Q and Lemma 2.2.1 to obtain

$$\overline{J} \leqslant \rho^{-N_T^*}(N_T^*)^2 C_2 h, \tag{II.32}$$

where C_2 is a positive constant independent of h. Then, we derive from the previous estimates (II.31) and (II.32) that

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

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 $\mathbb{E}[|\underline{\tilde{R}}_T - \tilde{R}_T|^2] \leqslant C_3 h^2 \mathbb{E}[\Xi_1(N_T^*)^2]$. 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 $\underline{\tilde{x}}_T = (\tilde{\theta}_{\tilde{N}_T}, \overline{\Phi}_{\tilde{\theta}_{\tilde{N}_T}}(T - \tilde{T}_{\tilde{N}_T}, \underline{\tilde{\nu}}_{\tilde{N}_T}))$. Then, using the Lipschitz continuity of F, Lemma 2.2.1 and since $\tilde{N}_T \leqslant N_T^*$ we get

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

Moreover, $|\underline{\tilde{R}}_T| \leqslant \left(\rho \frac{\tilde{\lambda}_{\min}}{\lambda^*}(1 - \frac{\tilde{\lambda}_{\max}}{\lambda^*})\right)^{-N_T^*}$ so that $\overline{D} \leqslant C_4 h^2 \mathbb{E}[\Xi_2(N_T^*)^2]$ 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 $\overline{D} = O(h^2)$.

□

2.4 Weak error expansion

In this section we are interested in a weak error expansion for the PDMP (x_t) of section 2.2.3 and its associated Euler scheme $(\overline{x}_t)$. First of all, we recall from [21] 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

$$\mathcal{A}g(t,x) = \partial_t g(t,x) + f(x)\partial_\nu g(t,x) + \lambda(x)\int_E (g(t,y) - g(t,x))Q(x,dy), \tag{II.33}$$

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 [76] (see also [66] or [42]) is delayed in section 2.4.2.

Assumption 2.4.1. *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 2.4.2. *The solution u of the integro differential equation*

$$\begin{cases} \mathcal{A}u(t,x) = 0, & (t,x) \in [0,T[\times E, \\ u(T,x) = F(x), & x \in E, \end{cases} \tag{II.34}$$

with $F : E \to \mathbb{R}$ a bounded function and $\mathcal{A}$ given by (II.33) 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 θ.

Theorem 2.4.1. *Let $(x_t, t \in [0,T])$ be a PDMP and $(\overline{x}_t, t \in [0,T])$ its approximation constructed in section 2.2.3 with $x_0 = \overline{x}_0 = x$ for some $x \in E$. Under assumptions 2.4.1. and 2.4.2. for any bounded function $F : E \to \mathbb{R}$ there exists a constant c_1 independent of h such that*

$$\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)] = hc_1 + O(h^2). \tag{II.35}$$

Remark 2.4.1. *If $(\tilde{x}_t)$ is a PDMP whose characteristics $\tilde{\lambda}, \tilde{Q}$ satisfy the assumptions of Proposition 2.2.2 and $(\underline{\tilde{x}}_t)$ is its approximation we deduce from Theorem 2.4.1 that*

$$\mathbb{E}[F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T] - \mathbb{E}[F(\tilde{x}_T)\tilde{R}_T] = hc_1 + O(h^2). \tag{II.36}$$

2.4.1 Further results on PDMPs: Itô and Feynman-Kac formulas

Definition 2.4.1. *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 2.4.1, the generator $\mathcal{A}$ defined by (II.33) 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 $p([0,t] \times A) := \sum_{n \geqslant 1} \mathbb{1}_{T_n \leqslant t}\mathbb{1}_{Y_n \in A}$ for $t \in [0,T]$ and for $A \in \mathcal{B}(E)$. The compensator of p, noted p', is given from [21] 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 $\overline{p}$, $\overline{p}'$, $\overline{q}$ and $(\mathcal{F}_t^{\overline{p}})_{t\in[0,T]}$ to be the same objects as above but corresponding to the approximation $(\overline{x}_t)$. The fact that $\overline{p}'$ is the compensator of $\overline{p}$ and that $\overline{q}$ is a martingale derives from arguments of the marked point processes theory, see [10].

Definition 2.4.2. *Let us define the following operators which act on functions g defined on $\mathbb{R}_+ \times E$.*

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

Remark 2.4.2. *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 $n \geqslant 0$ and for some $k \in \{0, \dots, \lfloor (\overline{T}_{n+1} - \overline{T}_n)/h \rfloor\}$.

Theorem 2.4.2. *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 2.2.3 with $x_0 = \overline{x}_0 = x$ for some $x \in E$. For all bounded functions $g : \mathbb{R}_+ \times E \to \mathbb{R}$ continuously differentiable with bounded derivatives, we have*

$$g(t, x_t) = g(0,x) + \int_0^t \mathcal{A}g(s, x_s)ds + M_t^g, \tag{II.37}$$

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

$$g(t, \overline{x}_t) = g(0,x) + \int_0^t \overline{\mathcal{A}}g(s, \overline{x}_s, \overline{x}_{\overline{\eta}(s)})ds + \overline{M}_t^g, \tag{II.38}$$

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

Proof of Theorem 2.4.2. The proof of (II.37) is given in [21]. We prove (II.38) following the same arguments. Since $\overline{q} = \overline{p} - \overline{p}'$, we have

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

Consider the above sum. As in [21], we write, on the event $\{\overline{N}_t = n\}$, that

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

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

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

Then, the above arguments together with definition 2.4.2 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 (II.34) 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 2.4.1.

Theorem 2.4.3 (PDMP's Feynman-Kac formula [22]). *Let $F : E \to \mathbb{R}$ be a bounded function. Then the integro-differential equation* (II.34) *has a unique solution $u : \mathbb{R}_+ \times E \to \mathbb{R}$ given by*

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

2.4.2 Proof of Theorem 2.4.1

We provide a proof in two steps. First, we give an appropriate representation of the weak error $\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)]$. Then, we use this representation to identify the coefficient c_1 in (II.35).

Step 1: Representing $\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)]$.
Let u denote the solution of (II.34). From Theorem 2.4.3 we can write $\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)] = \mathbb{E}[u(T,\overline{x}_T)] - u(0,x)$. Then, the application of the Itô formula (II.38) to u at time T yields

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

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

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

For $s \in [0,T]$ we have $\overline{\mathcal{A}}u(s,\overline{x}_s,\overline{x}_{\overline{\eta}(s)}) = \partial_t u(s,\overline{x}_s) + f(\overline{x}_{\overline{\eta}(s)})\partial_\nu u(s,\overline{x}_s) + \mathcal{S}u(s,\overline{x}_s)$ (see Definition 2.4.2). From the regularity of λ, Q and u (see assumptions 2.4.1 and 2.4.2), the functions $\partial_t u$, $\partial_\nu u$ and $\mathcal{S}u$ are smooth enough to apply the Itô formula (II.38) between $\overline{\eta}(s)$ and s respectively. This yields

$$\partial_t u(s,\overline{x}_s) = \partial_t u(\overline{\eta}(s),\overline{x}_{\overline{\eta}(s)}) + \int_{\overline{\eta}(s)}^s \overline{\mathcal{A}}(\partial_t u)(r,\overline{x}_r,\overline{x}_{\overline{\eta}(r)})dr + \overline{M}_s^{\partial_t u} - \overline{M}_{\overline{\eta}(s)}^{\partial_t u},$$

$$\partial_\nu u(s,\overline{x}_s) = \partial_\nu u(\overline{\eta}(s),\overline{x}_{\overline{\eta}(s)}) + \int_{\overline{\eta}(s)}^s \overline{\mathcal{A}}(\partial_\nu u)(r,\overline{x}_r,\overline{x}_{\overline{\eta}(r)})dr + \overline{M}_s^{\partial_\nu u} - \overline{M}_{\overline{\eta}(s)}^{\partial_\nu u},$$

$$\mathcal{S}u(s,\overline{x}_s) = \mathcal{S}u(\overline{\eta}(s),\overline{x}_{\overline{\eta}(s)}) + \int_{\overline{\eta}(s)}^s \overline{\mathcal{A}}(\mathcal{S}u)(r,\overline{x}_r,\overline{x}_{\overline{\eta}(r)})ds + \overline{M}_s^{\mathcal{S}u} - \overline{M}_{\overline{\eta}(s)}^{\mathcal{S}u}.$$

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

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

so that

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

where,

$$\Upsilon(t,x,y) := \left(\overline{\mathcal{A}}(\partial_t u) + f(y)\overline{\mathcal{A}}(\partial_\nu u) + \overline{\mathcal{A}}(\mathcal{S}u)\right)(t,x,y). \tag{II.39}$$

Since $\overline{\mathcal{A}}u(t,x,x) = \mathcal{A}u(t,x)$, the first term in the above equality is 0 by Theorem 2.4.3. By using Fubini's theorem and the fact that $(\overline{M}_t^{\partial_t u})$ and $(\overline{M}_t^{\mathcal{S}u})$ are true martingales, we obtain

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

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

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

Collecting the previous results, we obtain

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

We can compute an explicit form of Υ in term of u, f, λ, Q and their derivatives. Indeed, Υ is given by (II.39), and we have

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

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 $(\overline{\eta}(r),\overline{x}_{\overline{\eta}(r)})$ yields $\Upsilon(r,\overline{x}_r,\overline{x}_{\overline{\eta}(r)}) = \Upsilon(\overline{\eta}(r),\overline{x}_{\overline{\eta}(r)},\overline{x}_{\overline{\eta}(r)}) + O(h)$. Setting $\Psi(t,x) = \Upsilon(t,x,x)$ and recalling that for $r \in [\overline{\eta}(s), s]$, $\overline{\eta}(r) = \overline{\eta}(s)$ and that $|s - \overline{\eta}(s)| \leqslant h$, we obtain

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

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 $[\overline{T}_n + kh, (\overline{T}_n + (k+1)h) \wedge \overline{T}_{n+1}]$ where Ψ is constant. Without loss of generality, we are led to consider integrals of the form $\int_{kh}^t (s - kh)Cds$ for some $k \geqslant 0$, $t \in [kh, (k+1)h]$ and C a bounded constant. We

have $\int_{kh}^{t}(s-kh)Cds = \frac{t-kh}{2}\int_{kh}^{t}Cds$ moreover adding and subtracting h in the numerator of $(t-kh)/2$ yields

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

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

$$\mathbb{E}[F(\overline{x}_T)] - \mathbb{E}[F(x_T)] = \frac{h}{2}\mathbb{E}\left[\int_0^T \Psi(\overline{\eta}(s), \overline{x}_{\overline{\eta}(s)})ds\right] + O(h^2). \tag{II.40}$$

Step 2: From the representation (II.40) *to the expansion at the order one.*
In this step, we show that $\mathbb{E}\left[\int_0^T \Psi(\overline{\eta}(s), \overline{x}_{\overline{\eta}(s)})ds\right] = \mathbb{E}\left[\int_0^T \Psi(s, x_s)ds\right] + O(h)$. First, we introduce the random variables $\overline{\Gamma}$ and Γ defined by $\overline{\Gamma} := \int_0^T \Psi(\overline{\eta}(s), \overline{x}_{\overline{\eta}(s)})ds$ and $\Gamma := \int_0^T \Psi(\overline{\eta}(s), x_{\overline{\eta}(s)})ds$ and write

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

where $\overline{\tau}^\dagger$ is defined in Definition 2.3.1. Since Ψ is bounded and $\mathbb{P}(\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \leqslant T) = O(h)$ (see the proof of Theorem 2.3.1), we have $\mathbb{E}\left[|\overline{\Gamma} - \Gamma|\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) \leqslant T}\right] = O(h)$. Now, recall from (II.24) that, on the event $\{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T\}$, we have $T_k = \overline{T}_k$ and $\theta_k = \overline{\theta}_k$ for all $k \geqslant 1$ such that $T_k \in [0, T]$. Thus, for all $n \leqslant \overline{N}_T$ and for all $s \in [\overline{T}_n, \overline{T}_{n+1}[$ we have $\overline{x}_{\overline{\eta}(s)} = (\overline{\theta}_n, \overline{\phi}_{\overline{\theta}_n}(\overline{\eta}(s) - \overline{T}_n, \overline{\nu}_n))$ and $x_{\overline{\eta}(s)} = (\overline{\theta}_n, \phi_{\overline{\theta}_n}(\overline{\eta}(s) - \overline{T}_n, \nu_n))$. Consequently, on the event $\{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T\}$ we have

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

From the regularity assumptions 2.4.1 and 2.4.2, the function $\nu \mapsto \Psi(t, \theta, \nu)$ is uniformly Lipschitz in (t, θ) with constant L_Ψ as sum and product of bounded Lipschitz functions. Thus, from this Lipschitz property and the application of Lemma 2.2.1, we get

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

From the above inequality, we find that

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

Since $\overline{N}_T \leqslant N_T^*$ and $\mathbb{E}[N_T^*(N_T^*+1)] < +\infty$ we conclude that $\mathbb{E}\left[\mathbb{1}_{\min(T_{\overline{\tau}^\dagger}, \overline{T}_{\overline{\tau}^\dagger}) > T}|\overline{\Gamma} - \Gamma|\right] = O(h)$. We have shown that $\mathbb{E}\left[\int_0^T \Psi(\overline{\eta}(s), \overline{x}_{\overline{\eta}(s)})ds\right] = \mathbb{E}\left[\int_0^T \Psi(\overline{\eta}(s), x_{\overline{\eta}(s)})ds\right] + O(h)$. Secondly, from the regularity assumptions 2.4.1 and 2.4.2, the function $(t, \nu) \mapsto \Psi(t, \theta, \nu)$

is uniformly Lipschitz in θ. Moreover, for all $s \in [0,T]$ there exits $k \geqslant 0$ such that both s and $\overline{\eta}(s)$ belong to the same interval $[\overline{T}_k, \overline{T}_{k+1}[$ so that $x_s = (\theta_k, \phi_{\theta_k}(s - \overline{T}_k, \nu_k))$ and $x_{\overline{\eta}(s)} = (\theta_k, \phi_{\theta_k}(\overline{\eta}(s) - \overline{T}_k, \nu_k))$. Thus, from the Lipschitz continuity of Ψ, from the fact that $|s - \overline{\eta}(s)| \leqslant h$ and since f_θ is uniformly bounded in θ we have $|\Psi(s, x_s) - \Psi(\overline{\eta}(s), x_{\overline{\eta}(s)})| \leqslant Ch$ where C is a constant independent of h. Then, we obtain $\sup_{s\in[0,T]} |\mathbb{E}[\Psi(s, x_s)] - \mathbb{E}[\Psi(\overline{\eta}(s), x_{\overline{\eta}(s)})]| \leqslant Ch$ from which we deduce that $\left|\mathbb{E}\left[\int_0^T \Psi(\overline{\eta}(s), x_{\overline{\eta}(s)})ds\right] - \mathbb{E}\left[\int_0^T \Psi(s, x_s)ds\right]\right| \leqslant CTh$. Finally, the weak error expansion reads

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

□

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).

2.5.1 The PDMP 2-dimensional Morris-Lecar

The deterministic Morris-Lecar model has been introduced in 1981 by Catherine Morris and Harold Lecar in [63] 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 [47]) and takes the following form

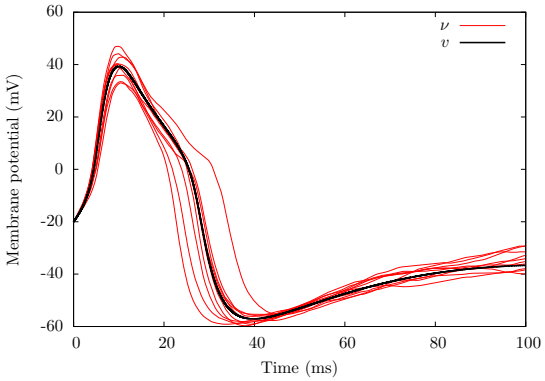
$$\begin{cases} \frac{dv}{dt} = \frac{1}{C}\Big(I - g_{\text{Leak}}(v - V_{\text{Leak}}) - g_{\text{Ca}}M_\infty(v)(v - V_{\text{Ca}}) - g_{\text{K}}n(v - V_{\text{K}})\Big), \\ \frac{dn}{dt} = (1 - n)\alpha_{\text{K}}(v) - n\beta_{\text{K}}(v), \end{cases} \tag{II.41}$$

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

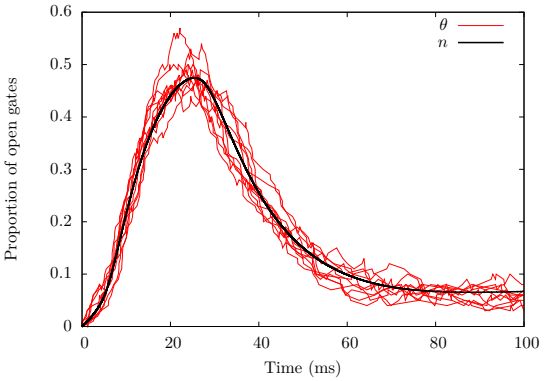
In this section we consider the PDMP version of (II.41) that we denote by $(x_t, t \in [0,T])$, $T > 0$, whose characteristics (f, λ, Q) are given by

- $f(\theta, \nu) = \frac{1}{C}\Big(I - g_{\text{Leak}}(\nu - V_{\text{Leak}}) - g_{\text{Ca}}M_\infty(\nu)(\nu - V_{\text{Ca}}) - g_{\text{K}}\frac{\theta}{N_{\text{K}}}(\nu - V_{\text{K}})\Big)$,
- $\lambda(\theta, \nu) = (N_{\text{K}} - \theta)\alpha_{\text{K}}(\nu) + \theta\beta_{\text{K}}(\nu)$,
- $Q\Big((\theta, \nu), \{\theta + 1\}\Big) = \frac{(N_{\text{K}} - \theta)\alpha_{\text{K}}(\nu)}{\lambda(\theta,\nu)}$, $\quad Q\Big((\theta, \nu), \{\theta - 1\}\Big) = \frac{\theta\beta_{\text{K}}(\nu)}{\lambda(\theta,\nu)}$.

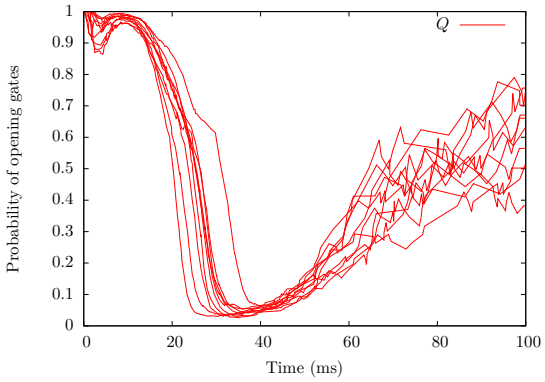
The state space of the model is $E = \{0, \ldots, N_{\text{K}}\} \times \mathbb{R}$ where $N_{\text{K}} \geqslant 1$ stands for the number of potassium gates. The values of the parameters used in the simulations are $V_1 = -1.2$, $V_2 = 18$, $V_3 = 2$, $V_4 = 30$, $\overline{\lambda}_{\text{K}} = 0.04$,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



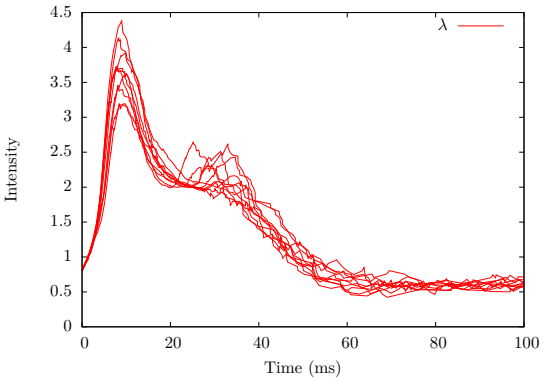
(a) Membrane potential as a function of time. Red curves: stochastic potential. Black curve: deterministic potential.



(b) Proportion of opened gates as a function of time. Red curves: stochastic gates (θ/N_{K}). Black curve: deterministic gates (n).



(c) Probability of opening a gate ($Q(x_t, \{\theta_t + 1\})$) as a function of time.



(d) Jump rate ($\lambda(x_t)$) as a function of t.

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

2.5.2 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 $n \geqslant 2$. 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 $(\overline{x}_t)$ on a time step h. We introduce a family of random variables $(X_h, h > 0)$ defined by $X_h := F(\overline{x}_T)$ where for a given $h > 0$ the corresponding PDP $(\overline{x}_t)$ is constructed as in section 2.2.3 with time step h. In particular, the processes $(\overline{x}_t)$ for $h > 0$ are correlated through the same randomness (U_k), (V_k) and (N_t^*). We build a

classical Monte Carlo estimator of $\mathbb{E}[X]$ based on the family $(X_h, h > 0)$ as follows

$$Y^{\mathrm{MC}} = \frac{1}{N} \sum_{k=1}^{N} X_h^k, \tag{II.42}$$

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

$$Y^{\mathrm{MLMC}} = \frac{1}{N_1} \sum_{k=1}^{N_1} X_{h^*}^k + \sum_{l=2}^{L} \frac{1}{N_l} \sum_{k=1}^{N_l} (X_{h_l}^k - X_{h_{l-1}}^k), \tag{II.43}$$

where $\left((X_{h_l}^k, X_{h_{l-1}}^k), k \geqslant 1\right)$ for $l = 2, \ldots, L$ are independent sequences of independent copies of the couple $(X_{h_l}, X_{h_{l-1}})$ and independent of the i.i.d sequence $(X_{h^*}^k, k \geqslant 1)$. The parameter h^* is a free parameter that we fix in section 2.5.4. The parameters $L \geqslant 2$, $M \geqslant 2$, $N \geqslant 1$ and $q = (q_1, \ldots, q_L) \in]0,1[^L$ with $\sum_{l=1}^{L} q_l = 1$ have to be determined, then we set $N_l := \lceil N q_l \rceil$, $h_l := h^* M^{-(l-1)}$.

We also set $\tilde{X} := F(\tilde{x}_T)\tilde{R}_T$ where $\tilde{R}_T$ is defined as in Proposition 2.2.2 with an intensity $\tilde{\lambda}$ and a kernel $\tilde{Q}$ that will be specified in section 2.5.4 and let $(\tilde{X}_h, h > 0)$ be such that $\tilde{X}_h := F(\underline{\tilde{x}}_T)\underline{\tilde{R}}_T$ for all $h > 0$. By Proposition 2.2.2, we have $\mathbb{E}[X] = \mathbb{E}[\tilde{X}]$ and $\mathbb{E}[X_h] = \mathbb{E}[\tilde{X}_h]$ for $h > 0$. Consequently, we build likewise a multilevel estimator $\tilde{Y}^{\mathrm{MLMC}}$ based on the family $(\tilde{X}_h, h > 0)$.

The complexity of the classical Monte Carlo estimator Y^{MC} depends on the parameters (h, N) and the one of the multilevel estimators Y^{MLMC} and $\tilde{Y}^{\mathrm{MLMC}}$ depends on (L, q, N). In order to compare those estimators we proceed as in [54] (see also [66]), 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 [54], we call V_1, c_1, α, β and $\mathrm{Var}(X)$ the structural parameters associated to the family $(X_h, h > 0)$ and X. We know theoretically from Theorem 2.3.1 (strong estimate) and Theorem 2.4.1 (weak expansion) that $(\alpha, \beta) = (1, 1)$ whereas V_1, c_1 and $\mathrm{Var}(X)$ are not explicit (we explain how we estimate them in section 2.5.3). Moreover, the structural parameters $\tilde{V}_1$, $\tilde{c}_1$, $\tilde{\alpha}$, $\tilde{\beta}$ and $\mathrm{Var}(\tilde{X})$ associated to $(\tilde{X}_h, h > 0)$ and $\tilde{X}$ are such that $\tilde{\alpha} = \alpha$, $\tilde{c}_1 = c_1$ (see (II.36)), $\tilde{\beta} = 2$ (see Theorem 2.3.2) and $\tilde{V}_1$, $\mathrm{Var}(\tilde{X})$ are not explicit.

The classical and the multilevel estimators defined above are linear and of Monte Carlo type in the sense described in [54]. The optimal parameters of those estimators are then expressed in term of the corresponding structural parameters as follows (see [54] or [66]). For a user prescribed $\epsilon > 0$, the classical Monte Carlo parameters h and N are

$$h(\epsilon) = (1 + 2\alpha)^{\frac{-1}{2\alpha}} \left(\frac{\epsilon}{|c_1|}\right)^{\frac{1}{\alpha}}, \quad N(\epsilon) = \left(1 + \frac{1}{2\alpha}\right) \frac{\mathrm{Var}(X)\left(1 + \rho h^{\beta/2}(\epsilon)\right)^2}{\epsilon^2}, \tag{II.44}$$

where $\rho = \sqrt{V_1/\mathrm{Var}(X)}$. The parameters of the estimator Y^{MLMC} are given in Table II.1 where $n_l := M^{l-1}$ for $l = 1, \ldots, L$ with the convention $n_0 = n_0^{-1} = 0$. The parameters of

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

Table II.1 – Optimal parameters for the MLMC estimator (II.43).

$\tilde{Y}^{\mathrm{MLMC}}$ are given in a similar way using $\tilde{V}_1$, $\tilde{\beta}$ and $\mathrm{Var}(\tilde{X})$. Finally, the parameter $M(\epsilon)$ is determined as in [54] section 5.1.

2.5.3 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 $\| Y - \mathbb{E}[X] \|_2 = \sqrt{(\mathbb{E}[Y] - \mathbb{E}[X])^2 + \mathrm{Var}(Y)}$ 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 (II.42) or (II.43). We estimate the bias of Y by

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

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

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

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

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

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 $l \geqslant 2$, $m_{N_l}^{(l)} = \frac{1}{N_l}\sum_{k=1}^{N_l} X_{h_l}^k - X_{h_{l-1}}^k$. Then, we define the empirical RMSE $\widehat{\epsilon}_R$ by

$$\widehat{\epsilon}_R = \sqrt{\widehat{b}_R^2 + \widehat{v}_R}. \tag{II.45}$$

The numerical computation of (II.45) for both estimators (II.42) and (II.43) requires the computation of the optimal parameters given by (II.44) and in table II.1 of section 2.5.2 which are expressed in term of the structural parameters c_1, V_1 and $\mathrm{Var}(X)$. 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 [54] section 5.1, that is

$$\widehat{V_1} = (1 + M^{-\beta/2})^{-2} h^{-\beta} \mathbb{E}\left[|X_h - X_{h/M}|^2\right], \tag{II.46}$$

and we use the following estimator of c_1

$$\widehat{c_1} = \left(1 - M^{-\alpha}\right)^{-1} h^{-\alpha} \mathbb{E}\left[X_{h/M} - X_h\right]. \tag{II.47}$$

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 (II.46) we use $(h, M) = (0.1, 4)$ and in (II.47), 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 interested in the order of c_1 and V_1 so that we do not need a precise estimation here.

2.5.4 Numerical results

In this section we first illustrate the results of Theorems 2.3.1 and 2.3.2 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 × 4. The code is written in C++ language. We implement the estimator $\tilde{Y}^{\mathrm{MLMC}}$ (see section 2.5.2) for the following choices of the parameters $(\tilde{\lambda}, \tilde{Q})$.

Case 1: $\tilde{\lambda}(\theta) = 1$ and $\tilde{Q}\Big(\theta, \{\theta + 1\}\Big) = \frac{N_{\mathrm{K}} - \theta}{N_{\mathrm{K}}}$, $\tilde{Q}\Big(\theta, \{\theta - 1\}\Big) = \frac{\theta}{N_{\mathrm{K}}}$.

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 (II.41).

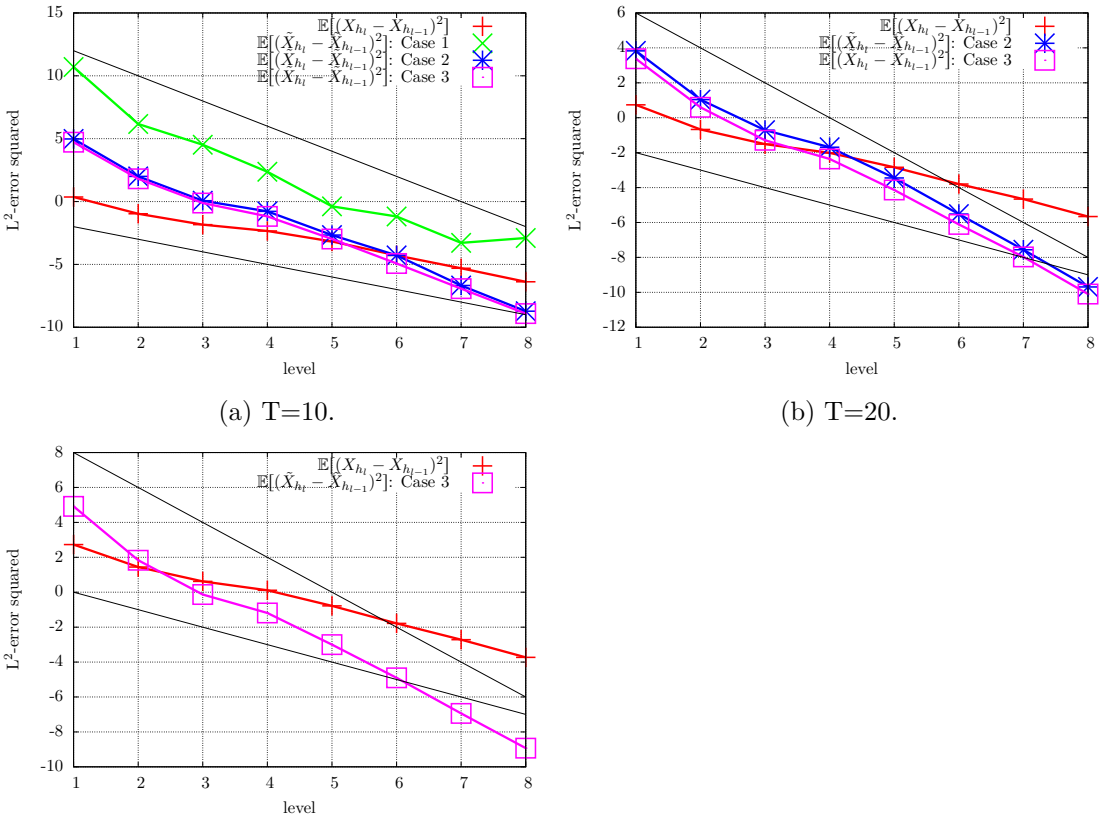
Cases 1 and 2 correspond to the application of Proposition 2.2.2. Based on Corollary 2.2.2 we also consider the following case.

Case 3: Consider the quantity $\mathbb{E}[F(x_T) - F(\tilde{x}_T)]$ where (x_t) and $(\tilde{x}_t)$ are PDPs with characteristics (Φ, λ, Q) and $(\tilde{\Phi}, \lambda, Q)$ respectively. By Corollary 2.2.2, 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 (II.43).

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

where $\left((X_{h_l}^k, \tilde{X}_{h_{l-1}}^k), k \geqslant 1\right)$ for $l = 2, \ldots, L$ are independent sequences of independent copies of the couple $(X_{h_l}, \tilde{X}_{h_{l-1}}) = (F(\overline{x}_T), F(\overline{y}_T)\tilde{R}_T)$ where $(\overline{y}_t)$ is a PDP whose discrete component jumps in the same states and at the same times as the Euler scheme $(\overline{x}_t)$ 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 2.2.2).



(a) T=10.

(b) T=20.

(c) T=30.

Figure II.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 figure II.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 2.3.1 and 2.3.2 respectively).

Indeed, for $T = 10$ (see figure II.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 do not consider this case in the sequel. As T increases (see figures II.2b and II.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 $\mathrm{Var}(X) = 335$. Note that $v(30) = -35.3083$ where v is the deterministic membrane potential solution of (II.41) 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 (II.45).

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

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

k	$\epsilon = 2^{-k}$	$\widehat{\epsilon}_{100}$	$\widehat{b}_{100}$	$\widehat{v}_{100}$	time (sec)	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

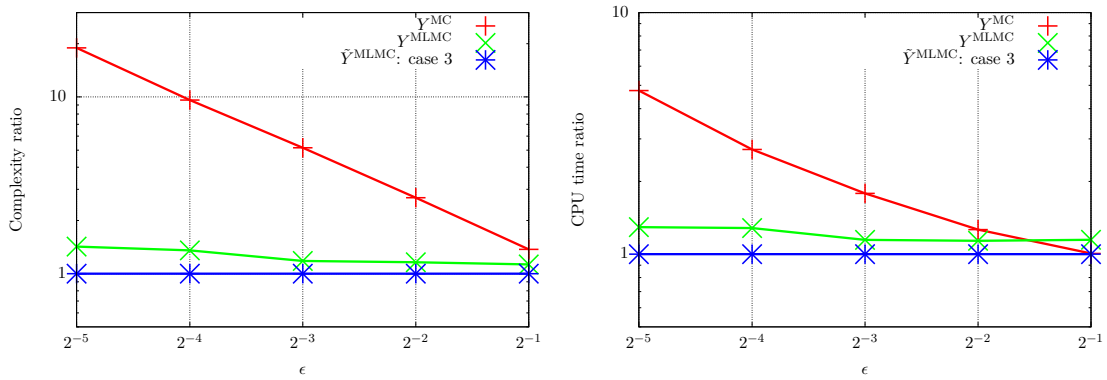
Table II.3 – Results and parameters of the Multilevel Monte Carlo estimator Y^{MLMC}. Estimated values of the structural parameters: $c_1 = 4.58$, $V_1 = 7.25$.

The results of the Monte Carlo simulations are shown in tables II.2 for the classical Monte Carlo estimator Y^{MC} and in tables II.3 and II.4 for the multilevel estimators

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

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

Y^{MLMC} and $\tilde{Y}^{\mathrm{MLMC}}$ (case 3). As an example, the first line of table II.3 reads as follows: for a user prescribed $\epsilon = 2^{-1} = 0.5$, the MLMC estimator Y^{MLMC} is implemented with $L = 2$ levels, the time step at the first level is $h^* = 0.1$, this time step is refined by a factor $n_l = M^{l-1}$ with $M = 2$ at each levels and the sample size is $N = 2600$. For such parameters, the numerical complexity of the estimator is $\mathrm{Cost}(Y^{\mathrm{MLMC}}) = 28200$, the empirical RMSE $\widehat{\epsilon}_{100} = 0.389$ and the computational time of one realisation of Y^{MLMC} is 0.362 seconds. We also reported the empirical bias $\widehat{b}_{100}$ and the empirical variance $\widehat{v}_{100}$ in view of (II.45).



(a) Ratio of the complexities.

(b) Ratio of the CPU-times.

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

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

complexity of these two MLMC estimators increases more slowly than the one between a MC and a MLMC estimator. Recall that the computational benefit of the MLMC over the MC grows as the prescribed ϵ decreases.
Both classical and multilevel estimators provide an empirical RMSE which is close to the prescribed precision (see tables II.2, II.3 and II.4). We can conclude that the choice of the parameters is well adapted. For the readability, figures II.3a, II.3b show the ratios of the complexities and the CPU-times of the three estimators Y^{MC}, Y^{MLMC} and $\tilde{Y}^{\mathrm{MLMC}}$ as a function of ϵ.

Bibliography

[1] A. Alfonsi, E. Cancés, G. Turinci, B. Di Ventura, and W. Huisinga. Adaptive hybrid simulation of hybrid stochastic and deterministic models for biochemical reactions. *ESAIM: proceedings*, 14:1–13, 2005.

[2] D.F. Anderson, B. Ermentrout, and P.J. Thomas. Stochastic Representations of Ion Channel Kinetics and Exact Stochastic Simulation of Neuronal Dynamics. *J. Comput. Neurosci.*, 38(1):67–82, February 2015.

[3] 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.

[4] 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.

[5] 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.

[6] F. Bouguet. Quantitative speeds of convergence for exposure to food contaminants. *ESAIM: Probability and Statistics*, 19:482–501, 2015.

[7] A. Brandejsky, B. De Saporta, and F. Dufour. Numerical method for expectations of piecewise deterministic Markov processes. *Commun. Appl. Math. Comput. Sci.*, 7(1):63–104, 2012.

[8] A. Brandejsky, B. De Saporta, and F. Dufour. Optimal stopping for partially observed piecewise-deterministic Markov processes. *Stochastic Processes and their Applications*, 123(8):3201 – 3238, 2013.

[9] P.C. Bressloff. *Stochastic Processes in Cell Biology.* Springer, Interdisciplinary Applied Mathematics, 2014.

[10] P. Brémaud. *Point Processes and Queues, Martingale Dynamics.* Springer-Verlag New York Inc, 1981.

[11] E. Buckwar and M.G. Riedler. An exact stochastic hybrid model of excitable membranes including spatio-temporal evolution. *Journal of Mathematical Biology*, 63:1051–1093, 2011.

[12] F. Campillo and C. Fritsch. A mass-structured individual-based model of the chemostat: convergence and simulation. *arXiv e-prints*, page arXiv:1308.2411, August 2013.

[13] D. Chafaï, F. Malrieu, and K. Paroux. On the long time behavior of the TCP window size process. *Stochastic Processes and their Applications*, 120:1518–1534, 2010.

[14] C.C. Chow and J.A. White. Spontaneous action potentials due to channel fluctuations. *Biophysical Journal*, 71(6):3013–3021, 1996.

[15] J.R. Clay and L.J. DeFelice. Relationship between membrane excitability and single channel open-close kinetics. *Biophysical Journal*, 42:151–157, 1983.

[16] C. Cocozza-Thivent. *Processus stochastiques et fiabilité des systèmes*. Springer, 1997.

[17] J.A. Connor and C.F. Stevens. Prediction of repetitive firing behaviour from voltage clamp data on an isolated neurone soma. *The Journal of Physiology*, 213(1):31–53, 1971.

[18] A. Crudu, A. Debussche, A. Muller, and O. Radulescu. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *The Annals of Applied Probability*, 22(5):1822 – 1859, 2012.

[19] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Vol. II.* Probability and its Applications (New York). Springer, second edition, 2008. General Theory and Structure.

[20] C.E. Dangerfield, D. Kay, and K. Burrage. Modeling ion channel dynamics through reflected stochastic differential equations. *Phys. Rev. E*, 85:051907, May 2012.

[21] 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.

[22] M.H.A. Davis. *Markov Models and Optimization*. Chapman and Hall, London, 1993.

[23] P. Dayan and L.F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[24] S. Dereich. Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *The Journal of Applied Probability*, 21(1):283–311, 2011.

[25] 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.

[26] L. Devroye. *Non-uniform random variate generation*. Springer-Verlag, New York Inc., 1986.

[27] S. Ding, M. Qian, H. Qian, and X. Zhang. Numerical simulations of piecewise deterministic Markov processes with an application to the stochastic Hodgkin-Huxley model. *The Journal of Chemical Physics*, 154:244107, 2016.

[28] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli Society for Mathematical Statistics and Probability*, 21:1760–1799, 2015.

[29] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., 1986.

[30] 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.

[31] R. FitzHugh. Mathematical models of threshold phenomena in the nerve membrane. *The bulletin of mathematical biophysics*, 17(4):257–278, Dec 1955.

[32] R.F. Fox. Stochastic Versions of the Hodgkin-Huxley Equations. *Biophysical Journal*, 72:2068–2074, 1997.

[33] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.

[34] M.B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[35] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical equations. *Journal of computational physics*, 22:403–434, 1976.

[36] D.T Gillespie. Monte Carlo Simulation of Random Walks with Residence Time Dependent Transition Probability Rates. *Journal of Computational Physics*, 28(3):395 – 407, 1978.

[37] D. Giorgi. *Théorémes limites pour estimateurs Multilevel avec et sans poids. Comparaisons et applications.* PhD thesis, Université Pierre et Marie Curie - Paris 6, 2017.

[38] P. Glasserman and N. Merener. Convergence of a discretization scheme for jump-diffusion processes with state-dependent intensities. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2041):111–127, 2004.

[39] P.W. Glynn and C-H. Rhee. Unbiased Estimation with Square Root Convergence for SDE models. *Operations Research*, 63(5):1026–1043, 2015.

[40] P.W. Glynn, C-H. Rhee, et al. Exact estimation for markov chain equilibrium expectations. *Journal of Applied Probability*, 51:377–389, 2014.

[41] J.H. Goldwyn, N.S. Imennov, M. Famulare, and E. Shea-Brown. Stochastic differential equation models for ion channel noise in Hodgkin-Huxley neurons. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 83 4 Pt 1:041908, 2011.

[42] C. Graham and D. Talay. *Stochastic simulation and Monte Carlo*. Springer, 2013.

[43] C. Gutierrez, C.L. Cox, J. Rinzel, and S.M Sherman. Dynamics of Low-Threshold Spike Activation in Relay Neurons of the Cat Lateral Geniculate Nucleus. *Journal of Neuroscience*, 21(3):1022–1032, 2001.

[44] E. Hairer, S.P. Norsett, and G. Wanner. *Solving Ordinary Differential Equations I*. Second Revised Edition, Springer, 2008.

[45] S. Heinrich. Multilevel Monte Carlo methods. *Large-scale scientific computing*, pages 58–67, 2001.

[46] B. Hille. *Ionic channels of excitable membranes*. Sinauer Associates Inc, 1992.

[47] 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.

[48] Y. Huang, S. Rüdiger, and J. Shuai. Channel-based Langevin approach for the stochastic Hodgkin-Huxley neuron. *Phys. Rev. E*, 87:012716, Jan 2013.

[49] J. R. Huguenard and D. A. McCormick. Simulation of the currents involved in rhythmic oscillations in thalamic relay neurons. *Journal of Neurophysiology*, 68(4):1373–1383, 1992. PMID: 1279135.

[50] E.M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, 2008.

[51] M. Jacobsen. *Point Process Theory and Applications, Marked Point and Piecewise Deterministic Processes*. Birkhäuser Boston, 2006.

[52] O. Kallenberg. *Foundations of Modern Probability*. Springer,New York, 2002.

[53] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, New York, corrected edition, June 2011.

[54] V. Lemaire and G. Pagès. Multilevel Richardson-Romberg extrapolation. *Bernoulli*, 23(4A):2643–2692, 2017.

[55] 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.

[56] V. Lemaire, M. Thieullen, and N. Thomas. Thinning and Multilevel Monte Carlo for Piecewise Deterministic (Markov) Processes. Application to a stochastic Morris-Lecar model. *arXiv e-prints*, page arXiv:1812.08431, December 2018.

[57] P.A.W. Lewis and G.S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26:403–413, 1979.

[58] D. Linaro, M. Storace, and M. Giugliano. Accurate and Fast Simulation of Channel Noise in Conductance-Based Model Neurons by Diffusion Approximation. *PLOS Computational Biology*, 7(3):1–17, 03 2011.

[59] R. Mikulevicius and E. Platen. Time Discrete Taylor Approximations for Itô Processes with Jump Component. *Math. Nachr.*, 138:93–104, 1988.

[60] H. Mino, J.T. Rubinstein, and J.A White. Comparison of algorithms for the simulation of action potentials with stochastic sodium channels. *Annals of Biomedical Engineering*, 30:578–587, 2002.

[61] C.C. Mitchell and D.G. Schaeffer. A two-current model for the dynamics of cardiac membrane. *Bulletin of Mathematical Biology*, 65(5):767–793, Sep 2003.

[62] J. Moller and R.P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2003.

[63] C. Morris and H. Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*, 35:193–213, 1981.

[64] Y. Ogata. On Lewis' Simulation Method for Point Processes. *IEEE Transactions on Information Theory*, 27:23 – 31, 1981.

[65] P. Orio and D. Soudry. Simple, Fast and Accurate Implementation of the Diffusion Approximation Algorithm for Stochastic Ion Channels with Multiple States. *PLOS ONE*, 7(5):1–16, 05 2012.

[66] G. Pagès. *Numerical Probability: An Introduction with Applications to Finance*. Universitext, Springer, Cham, 2018.

[67] 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.

[68] Z. Palmowski and T. Rolski. A technique for exponential change of measure for Markov processes. *Bernoulli*, 8(6):767–785, 12 2002.

[69] D. Pezo, D. Soudry, and P. Orio. Diffusion approximation-based simulation of stochastic ion channels: which method to use? *Frontiers in Computational Neuroscience*, 8:139, 2014.

[70] M.G. Riedler. Almost sure convergence of numerical approximations for Piecewise Deterministic Markov Processes. *Journal of Computational and Applied Mathematics*, pages 50–71, 2012.

[71] M.G. Riedler and G. Notarangelo. Strong error Analysis of the Theta-Method for Stochastic Hybrid Systems. *arXiv preprint arXiv:1310.0392*, 2013.

[72] P.F. Rowat. Interspike Interval Statistics in the Stochastic Hodgkin-Huxley Model: Coexistence of Gamma Frequency Bursts and Highly Irregular Firing. *Neural Computation*, 19:1215–1250, 2007.

[73] P.F. Rowat and P.E. Greenwood. The ISI distribution of the stochastic Hodgkin-Huxley neuron. In *Front. Comput. Neurosci.*, 2014.

[74] J.T. Rubinstein. Threshold fluctuations in an N sodium channel model of the node of ranvier. *Biophysical Journal*, 68:779–785, 1995.

[75] E. Skaugen and L. Walloe. Firing behaviour in a stochastic nerve membrane model based upon the Hodgkin-Huxley equations. *Acta Physiologica*, 107:343–363, 1979.

[76] 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.

[77] R. Veltz. A new twist for the simulation of hybrid systems using the true jump method. *arXiv[math]*, 2015.

[78] A.A. Verveen and H.E. Derksen. Fluctuation Phenomena in Nerve Mebrane. *proceedings of the IEEE*, 56:906–916, 1968.

[79] X. J. Wang, J. Rinzel, and M. A. Rogawski. A model of the T-type calcium current and the low-threshold spike in thalamic neurons. *Journal of Neurophysiology*, 66(3):839–850, 1991.

[80] J. White, R. Klink, A. Alonso, and A. Kay. Noise from voltage-gated ion channels may influence neuronal dynamics in the entorhinal cortex. *Journal of Neurophysiology*, 80(1):262–269, 1998.

[81] Y. Xia. *Multilevel Monte Carlo for jump processes*. PhD thesis, University of Oxford, 2013.

[82] 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.