Modelling of Stochastic Hybrid Systems with Applications to Accident Risk Assessment

The research described in this thesis was undertaken at the Department of Applied Mathematics, in the Faculty EWI, University of Twente, Enschede, The Netherlands. The funding of the research was provided by the European Commission project HYBRIDGE, IST-2001-32460.

Printed by Wöhrmann Printing Service, The Netherlands

ISBN: 90-365-2401-6

MODELLING OF STOCHASTIC HYBRID SYSTEMS WITH APPLICATIONS TO ACCIDENT RISK ASSESSMENT

DISSERTATION

to obtain
the doctor's degree at the University of Twente,
on the authority of the rector magnificus,
prof. dr. W.H.M. Zijm,
on account of the decision of the graduation committee,
to be publicly defended
on Wednesday 6 September 2006 at 13.15 hours

by

Jaroslav Krystul
born on 29 July 1979
in Vilnius, Lithuania

Dit proefschrift is goedgekeurd door de promotor
Prof. dr. A. Bagchi

Contents

Acronyms

ATM	Air Traffic Management
GB1	switching jump diffusion of Ghosh and Bagchi (2004)
GB2	switching diffusion with hybrid jumps at the boundary of Ghosh and Bagchi (2004)
HB1	switching hybrid-jump diffusion of Blom (2003)
HB2	switching hybrid-jump diffusion with hybrid jumps at the boundary of Blom *et al.* (2003*b*)
HIPS	Hybrid Interacting Particle System
ICAO	International Civil Aviation Organization
IPS	Interacting Particle Systems
KB1	switching hybrid-jump diffusion of Krystul and Blom (2005*a*)
KB2	switching hybrid-jump diffusion with hybrid jumps at the boundary of Krystul and Blom (2005*a*)
MC	Monte Carlo
SDE	Stochastic Differential Equation

Chapter 1

Introduction

1.1 Accident risk assessment

Stochastic dynamical modelling of accident risk is of high interest for the safe design of complex safety-critical systems and operations, such as in nuclear and chemical industries, and advanced air traffic management, e.g. see (Smidts *et al.*, 1998; Labeau *et al.*, 2000; Blom *et al.*, 2003*a*) and their references. In comparison with statistical analysis of collected data (Embrechts *et al.*, 1997), stochastic dynamical modelling approach has the advantage of enabling the use of stochastic analysis and advanced Monte Carlo (MC) simulation approaches (Doucet *et al.*, 2001). In some simple cases it is possible to use analytical calculation methods and these would normally be preferred. However, for complex situations, MC simulation often represents the only useful alternative. The advantage of standard MC simulation methods for accident risk assessment is that they do not require any specific assumptions on the system under consideration.

We define the *risk* as the probability that a particular adverse event occurs during a stated period of time. Usually this is the event when the system reaches a particular critical state. The events with very small probability of occurrence are called rare events . Obtaining accurate estimates of rare event probabilities, say about 10^{-9} to 10^{-12}, is not realistic just by using straightforward MC simulation because most of the realizations of the system never reach the rare event states. For example, in order to estimate the probabilities of order 10^{-9}, about 10^{11} simulation runs are needed, which is very time consuming. This makes MC simulation to be a practical alternative only when it is possible to realize a high speed up.

The techniques used in (Smidts *et al.*, 1998; Labeau *et al.*, 2000; Blom *et al.*, 2003*a*) for speeding up MC simulation are model specific risk decompositions. Hence there is a need for a more systematic and general approach.

A well known approach is importance sampling, which is based on a modifica-

tion of the underlying probability distribution in such a way that the rare events occur much more frequently. The effectiveness of such method depends critically on the ability to find the right change of measure. If it is done improperly, the importance sampling may produce worse results than straightforward simulation. Finding the right change of measure generally requires identifying at least the rough asymptotics of a rare event probability. This type of analysis can be formidable in complex dynamic models. See (Heidelberger, 1995; Shahabuddin, 1995; Liu, 2003) and (Juneja and Shahabuddin, 2006) for surveys.

An alternative approach for rare event estimation, that requires little analysis of system structure for its applicability, is to express the small probability of rare event to be estimated as the product of a certain number of larger probabilities, which can be efficiently estimated by the MC methods. This can be achieved by introducing sets of intermediate states that are visited one set after the other, in an ordered sequence, before reaching the final set of rare event states. The probability of rare event is then given by the product of the conditional probabilities of reaching a set of intermediate states given that the previous set of intermediate states have been reached. Each conditional probability is estimated by simulating in parallel several copies of the process, i.e. each copy is considered as a particle following the trajectory generated through the process dynamics. Each particle branches (i.e. the trajectory splits into a number of independent subpaths, which subsequently evolve independently of each other) as soon as it enters the intermediate states, which is usually characterized by crossing a threshold by a control parameter. Reaching intermediate states is more likely than reaching the rare event states, and by splitting at each threshold the chances to reach the rare event states are increasing.

A number of techniques based on the above mentioned idea of state space decomposition and splitting of trajectories are available in the literature: multilevel splitting techniques (for a complete review and detailed list of references see (Glasserman *et al.*, 1999; Lezaud *et al.*, 2004)); empirical method RESTART (Villén-Altamirano and Villén-Altamirano, 1991, 1994); and more recent Interacting Particle Systems (IPS) approaches (Cérou *et al.*, 2002; Del Moral, 2004; Cérou *et al.*, 2005). The IPS approach seems to be the most suitable for rare event estimation in stochastic dynamical systems. This will become clear in Chapter 2, where we give a brief overview of these splitting approaches.

In practice the IPS approach can be applied to any process. For the convergence proof to hold true, however a strong Markov process is required. The IPS approach works very well with diffusion processes (Krystul and Blom, 2004), but needs to be improved in order to provide computationally efficient and reliable estimates of rare event probabilities in stochastic dynamical systems that exhibit regime-switching behavior, i.e. the systems evolving in a hybrid state space. In this thesis we develop efficient model-independent IPS based MC simulation approaches for accident risk assessment in complex stochastic hybrid systems. This will be done in Chapter 5.

1.2 Stochastic dynamic modelling

Another essential part of the thesis deals with the modelling issues of dynamical hybrid systems. Herein we explain which modelling approach is taken in this thesis and the reasons of our choice.

While studying a wide variety of real-world phenomena we usually encounter processes the course of which cannot be predicted beforehand. For example: sudden deviation of the altitude of an aircraft from a prescribed flight-level; reproduction of bacteria in a favorable environment; movement of a stock price on a stock exchange. We cannot predict in advance whether at a particular time moment the altitude of the aircraft will be 19900 or 20100 feet, and what will be the coordinates. Such processes can be represented by stochastic movement of a point in a particular space specially selected for each problem. The proper choice of the phase space turns physical, mechanical or any other real-world system into dynamical system (it means that the current state of the system determines its future evolution). Similarly, by a proper choice of the phase space (or state space) an arbitrary stochastic process can be turned into a Markov process, i.e. a process the future evolution of which depends on the past only through its present state. This property is called the Markov property. From a whole set of stochastic processes this Markov property singles out a class of Markov processes which are a natural generalization of the dynamical system. In this thesis we deal only with Markov processes.

There exist two directions in the development of theory of Markov processes: an analytical and a stochastic direction. Transition densities or transition probabilities are the starting point of the analytical Markov process theory. It studies various classes of transition densities and transition probabilities, which are described by equations (for example, by partial differential equations), which in most cases can be solved only by numerical approximation methods. When proving the existence of corresponding Markov processes, any obtained conditions and properties on transition densities and probabilities are simply interpreted as certain properties of these processes. Broadly speaking, the approach taken by analytical Markov process theory could be compared with the analysis of the properties of random variables on the basis of their distribution functions or densities. In the stochastic Markov process theory a Markov process is constructed directly as a solution to stochastic differential equation (SDE). The main advantage of this approach is that it is easier to study a Markov process as a solution of a particular equation than a Markov process that is implicitly defined through its transition density or probability. In this thesis we consider only the stochastic Markov processes approach for modelling real-world dynamical systems.

Currently, the theory of stochastic differential equations, widely using martingale methods, has become a powerful tool for constructive description of various classes of stochastic processes including the processes which are semimartingales. Semimartingales form one of the most important and general class of stochastic

processes which includes diffusion-type processes, point processes, and diffusion-type processes with jumps that are widely used for stochastic modelling. Considering SDE with semimartingale solutions gives an advantage. It allows to use the powerful stochastic calculus available for the semimartingale processes when performing complex stochastic analysis.

Continuous time Markov processes such as diffusions, point processes and diffusions with jumps have been successfully used for years in stochastic modelling of various continuous time real-world dynamical systems with the Euclidean phase space. Recently, there is a great interest in more complex continuous time stochastic processes with the hybrid state space containing both Euclidean and discrete valued components. Such processes are called stochastic hybrid processes . Euclidean and discrete valued components may interact, i.e. Euclidean valued components may influence the dynamics of discrete valued component and vice versa. This makes the modelling and the analysis of stochastic hybrid processes quite involved and challenging. Several classes of stochastic hybrid processes have been studied in the literature, e.g. counting processes with diffusion intensity (Snyder, 1975; Marcus, 1978), diffusion processes with Markovian switching parameters (Wonham, 1970; Mariton, 1990), switching diffusions (Ghosh *et al.*, 1993, 1997), piecewise deterministic Markov processes (Davis, 1984, 1993; Jacod and Skorokhod, 1996), Markov decision drift processes (Duyn Schouten and Hordijk, 1983), stochastic hybrid systems of (Hu *et al.*, 2000; Pola *et al.*, 2003) and more recent SDE models on hybrid state spaces (Blom, 1990, 2003; Blom *et al.*, 2003*b*; Ghosh and Bagchi, 2004; Krystul *et al.*, 2006). All these stochastic hybrid processes arise in various kind of applications, have different degree of modelling power and have different properties inherent to the problems that they have been developed for. As we have already mentioned above, in this thesis we restrict our attention to the stochastic approaches of modelling of stochastic hybrid systems, i.e. modelling by Markov processes which are defined as solutions to SDE.

1.3 Numerical approximation

Most SDE do not admit closed form analytical solutions, the only alternative of studying them is then numerical simulation. The sample paths (realizations) or functionals of solutions of SDE are commonly simulated through discrete-time approximations which are implementable on digital computers.

The discrete-time numerical schemes for Itô diffusions are well explained in the literature (Kloeden and Platen, 1992). A discretization scheme for jump-diffusion processes with state-dependent intensities was considered in (Glasserman and Merener, 2004). Weak approximations of killed (or stopped) diffusions were studied in (Gobet, 1999*a*,*b*, 2000, 2001) and (Moon, 2003). They develop and prove the convergence of numerical schemes that approximate the expected value of a given function depending on the solution of an Itô stochastic differential equation and, in some cases, on the first exit time from a given domain. However,

despite the availability in the literature of a large number of numerical techniques for SDE, most of the focus has not been on SDE on hybrid state spaces. In this thesis we study the discrete-time approximations of a class of stochastic hybrid processes defined as solutions to SDE, namely the switching diffusions with state dependent switching rates (Ghosh *et al.*, 1993, 1997). We propose to approximate the switching diffusion by strong Euler-type discretization scheme and prove its convergence. Next, following the approach of Gobet (1999*a*) we prove that under weak conditions the first passage times of the discretized process weakly converge to the first passage times of the original process. These results are necessary in order to implement the IPS simulation approaches for a switching diffusions on a computer.

1.4 Layout and contributions

The thesis consists of six chapters. The contents of the remaining chapters are briefly summarized as follows.

Chapter 2 This chapter is an introductory discussion concerning accident risk assessment in stochastic dynamical systems. It starts with a simple motivating example from air traffic management domain. Two different risk measures are discussed: the in-crossing risk versus the hitting probability. Standard Monte Carlo, multilevel splitting and interacting particle systems approaches for rare event estimation in stochastic dynamical systems are explained. Interacting particle systems approach, as the most promising, is tested on a special test example for which the exact analytical solution is known. This chapter is mainly based on the following papers (Glasserman *et al.*, 1999; Cérou *et al.*, 2002; Krystul *et al.*, 2003; Krystul and Blom, 2004; Lezaud *et al.*, 2004), as well as other references in the chapter.

Chapter 3 This chapter deals with modelling of stochastic hybrid processes as strong solutions to stochastic differential equations on hybrid state space. A brief introduction to semimartingales is given. The existence and uniqueness results for $\mathbb{R}^n$-valued jump-diffusions are presented. Next, these results are extended to a class of hybrid state processes with Poisson and hybrid Poisson jumps. A general stochastic hybrid process which includes jumps at the boundaries is characterized. Several classes of stochastic hybrid models are discussed and compared. Finally, the Markov and the Strong Markov properties for the general stochastic hybrid process are shown. This chapter is based on (Blom, 2003; Blom *et al.*, 2003*b*; Ghosh and Bagchi, 2004; Krystul *et al.*, 2006).

Chapter 4 This chapter addresses some issues concerning time discrete approximations of stochastic hybrid processes. The stochastic Euler scheme for the switching diffusion with state dependent switching rates is presented. The

proof of strong convergence of Euler scheme is given. The approximation of first passage times is discussed and the convergence is proved. This chapter is based on (Krystul and Bagchi, 2004).

Chapter 5 In this chapter new Hybrid IPS algorithm is being developed for the efficient estimation of rare event probabilities in stochastic hybrid systems. First, the IPS approach is formulated for a switching diffusion case and subsequently extended to Hybrid IPS algorithm which is designed to cope with large differences in mode probabilities and rare switchings. Numerical evaluations and comparison of different versions of the IPS algorithms are given. This chapter is mainly based on (Cérou *et al.*, 2002; Krystul and Blom, 2004, 2005*b*).

Chapter 6 This chapter presents conclusions and recommendations on the possible directions for future research.

Chapter 2

Brief overview of Monte Carlo techniques for accident risk assessment

2.1 Introduction

The aim of this thesis is to develop a general framework for the modelling and assessment of accident risks in complex stochastic dynamical systems. In particular, we are interested in developing the framework which would facilitate the design of new advanced Air Traffic Management (ATM) concepts.

In the design of advanced ATM concepts, safety is recognized as a key factor. Traditional ATM design approaches tend first to design advanced ATM that provides sufficient capacity, and next to extend the design with safety features. The advantage of this approach is that ATM developments can be performed separately for a range of clusters of individual elements, i.e., the communication cluster, the navigation cluster, the human machine interfaces, the advanced procedures such as missed approaches, air traffic control sector transitions, overtake manoeuvres in unmanaged airspace, etc. The disadvantage of this traditional approach is that it fails to address the impact of interactions between ATM elements on the overall safety of the system and it is much more difficult to extend the already-existing design so that it meets safety requirements (Blom *et al.*, 2003*a*). Today, modelling seems to be the only feasible and cost effective way to develop new advanced ATM designs that are inherently safe at the capacity level required. In our intended unified modelling framework all individual ATM elements and interactions between them will be treated as one complex dynamical system, the evolution of which is described by a set of stochastic differential equations.

This chapter is organized as follows. Section 2.2 presents the motivating example from ATM domain and a general discussion. Section 2.3 explains standard Monte Carlo, multilevel splitting and interacting particle systems approaches for rare events estimation in stochastic dynamical systems.

2.2 Motivating example

Let us consider a simple scenario where two aircraft are flying towards each other on the parallel straight lanes on the same flight level. We assume that each aircraft is represented as a box having a fixed orientation, and we also assume that evolution of the pair of aircraft is represented by stochastic differential equations, one for each aircraft, i.e. for $i = 1, 2$,

$$dX_t^i = a^i(X_t)dt + b^i(X_t)dW_t^i,$$

with $X_t = \{X_t^1, X_t^2\}^T$, $W_t = \{W_t^1, W_t^2\}$, W_t^i an n-dimensional standard Brownian motion, $\{X_t^i\}$ assumes values in $\mathbb{R}^n$. Here the Brownian motion is used to model the effects of random wind disturbances on aircraft trajectories. Some elements of X_t^i form a 3D position s_t^i of aircraft i,

$$s_t^i = HX_t^i,$$

with H a $3 \times n$-matrix. To avoid the Brownian motion behavior in positions, the following assumption is adopted

$$Hb^i(X_t) = 0 \text{ for } i = 1, 2.$$

Hence

$$ds_t^i = v_t^i dt \text{ with } v_t^i \triangleq Ha^i(X_t).$$

Next, let s_t^1 and s_t^2 represent the positions of centers of aircraft pair. Then by

$$s_t = s_t^1 - s_t^2$$

we denote the relative position (separation process) . The relative velocity is represented in a similar way

$$v_t = v_t^1 - v_t^2.$$

Hence

$$ds_t = v_t dt.$$

These relative position/velocity equations were used in the previous studies by Bakker and Blom (1993); Blom and Bakker (2002); Blom *et al.* (2003*b*). Of course, in these references, the description of ATM scenarios involves much more than these relative position/velocity equations alone. In general, descriptions of real ATM scenarios are very complex. For example, an aircraft trajectory can

be subdivided into distinct segments corresponding to different phases of flight and operation modes. The aircraft can be in nominal mode, where it has small deviations from the lane; in a non-nominal mode caused by hardware failure or human error, where it exhibits large deviation from the predefined flight path; in a sharp turn mode and so forth. Thus for a realistic modelling of ATM scenarios it is more natural to use a hybrid state stochastic process $\{X_t, \theta_t\}$ where the Euclidean valued component X_t, for example, can represent the aircraft position, speed, acceleration, etc., and the discrete valued component θ_t can describe the flight phase, the operation modes of the hardware, the stress level and workload of the air traffic controller and pilots.

Let us continue with our relative position/velocity equations. We define a collision domain $D \equiv D_1 \times D_2 \times D_3$, $(D_k = [-m_k, m_k])$, as a box of a size of two aircraft with center in the origin of axis. If the relative position process $\{s_t\}$ enters set D then it means that at moment t the physical volumes of two aircraft are not separated any more, i.e. they have collided. Our aim is to assess the risk of collision between a pair of aircraft. Now, the natural question arises: how to measure the accident risk. Several safety metrics have been used in the literature for accident risk assessment in ATM. Let us discuss them in detail.

Incrossing risk. First, let us consider a safety metric used by the International Civil Aviation Organization (ICAO) known as *incrossing risk*. This metric arose from the studies of Rice (1945) and Reich (1964) and its development was continued by Bakker and Blom (1993); Blom *et al.* (1994) and Blom and Bakker (2002). The incrossing risk is defined as an expected number of incrossings of a process into a particular critical domain (in our example it is a collision domain D). Each time the separation process $\{s_t\}$ enters the set D, we say an incrossing occurs, and each time it leaves the set D, we say an outcrossing occurs. Following (Bakker and Blom, 1993) the expected number of incrossings $\mathscr{R}_{in}(0,T)$ between a pair of aircraft in the time interval $[0,T]$ satisfies:

$$\mathscr{R}_{in}(0,T) = \int_0^T \phi(t)dt \tag{2.2.1}$$

with $\phi(t)$ the in-crossing rate, which is defined, if the limit exists, as

$$\phi(t) \triangleq \lim_{\Delta \downarrow 0} \frac{P(s_{t-\Delta} \notin D, s_t \in D)}{\Delta}.$$

In (Bakker and Blom, 1993) and (Blom and Bakker, 2002) characterizations of the in-crossing rate $\phi(t)$ have been derived under very general conditions. In (Bakker and Blom, 1993) the following characterization for in-crossing rate has

been developed:

$$\phi(t) = \sum_{k=1}^{3} \int_{\underline{D}_k} \Big(\int_0^\infty v_k p_{s_t, v_{k,t}}(\underline{s}_k, -m_k, v_k) dv_k + \int_{-\infty}^0 -v_k p_{s_t, v_{k,t}}(\underline{s}_k, m_k, v_k) dv_k \Big) d\underline{s}_k$$

where

$$\begin{array}{lll} \underline{D}_1 \equiv D_2 \times D_3, & \underline{D}_2 \equiv D_1 \times D_3, & \underline{D}_3 \equiv D_1 \times D_2, \\ \underline{s}_1 \equiv (s_2, s_3), & \underline{s}_2 \equiv (s_1, s_3), & \underline{s}_3 \equiv (s_1, s_2). \end{array}$$

The above model, referred to as the generalized Reich model, assumes that the process $\{s_t, v_t\}$ is Markov and admits a joint density function $p_{s_t,v_t}(s, v)$ which satisfies the Chapman-Kolmogorov equation:

$$p_{s_t,v_t}(s, v) = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} p_{s_t,v_t|s_0,v_0}(s, v|s', v') p_{s_0,v_0}(s', v') ds' dv'.$$

The initial density function $p_{s_0,v_0}(s, v)$ is assumed to be known. Theoretically, the time evolution of the transition density $p_{s_t,v_t|s_0,v_0}(s, v|s', v')$ can be characterized through appropriate forward or backward Kolmogorov equations (see (Bakker and Blom, 1993; Krystul *et al.*, 2003) for overview and discussion). Unfortunately, solving these equations analytically or numerically is not realistic when the dimension of the process is as high as it is in ATM applications. For some of existing ATM operations there is an alternative way in using the generalized Reich equations. This consists of drawing appropriate samples from the true operation, and to use these samples as an empirical density in the evaluation of the generalized Reich model (Blom *et al.*, 2003*b*). This approach can only be applied to ATM designs that are sufficiently long in operation, but not for advanced designs.

Another important issue is that, in fact, in the above model the incrossing of collision domain D can happen more than once. The first incrossing is the actual collision of the aircraft, thus, the successive incrossings have no real physical meaning. Hence the incrossing risk is an upperbound of the collision probability. An exact characterization of the difference has been provided by Blom *et al.* (2003*b*). The upperbound approximation aspect has been identified for the first time by Hsu (1981). By appropriate modelling of the aircraft evolution one can achieve that the chances to have more than one incrossings, i.e. collisions, between one pair of aircraft are extremely small.

Hitting probability. In stochastic analysis the collision probability is commonly referred to as *hitting* (sometimes also called *first passage*, or *first exit*) probability. Define the hitting time of set D:

$$\tau = \inf\{t > 0 : s_t \in D\},$$

i.e. the first incrossing time of set D. If $\{s_t, v_t\}$ is an adapted càdlàg (i.e. all its paths are right-continuous with left limits) process and D is a closed set, then τ is a stopping time, the event $\{\omega : \tau(\omega) \leq T\}$ is well defined and its probability can be estimated. The probability to hit the target set D within the time interval $[0, T]$ is denoted as follows

$$P_{hit}(0, T) \triangleq P(\tau \leq T). \tag{2.2.2}$$

This risk metric agrees with the following definition of risk adopted by (RS, 1992).

Definition 2.2.1. We treat the *risk* as the probability that a particular adverse event occurs during a stated period of time. As a probability in the sense of statistical theory, risk obeys all the formal laws of combining probabilities. Explicitly or implicitly, it must always relate to the risk of a specific event or set of events and where appropriate must refer to an exposure to hazard specified in terms of its amount or intensity, time of starting or duration.

For particular types of Markov processes the risk metric (2.2.2) can be numerically evaluated through solving a backward Kolmogorov partial differential equation (PDE) with Dirichlet type boundary conditions (Friedman, 1975, Theorem 5.2, Chapter 6). However, the numerical evaluation of such PDEs is limited to two- or at most three- dimensional problems.

Since the Monte Carlo simulation techniques are almost insensitive to the dimensionality of the problem and do not require specific assumptions on the system under consideration, they seem to be the perfect alternative to numerical approximating techniques. In the following section we give a brief introduction to Monte Carlo techniques and discuss their suitability for rare event probability estimation in stochastic dynamical systems.

2.3 Monte Carlo and rare event estimation

Throughout this section, all stochastic processes are defined on a complete stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$ with $(\mathscr{F}_t)_{t\geq 0}$ a right continuous filtration.

2.3.1 Standard Monte Carlo

In this section we explain the main issues concerning the MC simulation of SDE.

Most SDE do not admit closed form analytical solutions, the only alternative then is numerical simulation. There are two types of problems connected with the simulation of solutions of SDE. The first type is where a good pathwise approximation is required, for instance in direct simulations or filtering problems. The second type of problems is where one is interested in approximating expectations of functionals of solution, such as its probability distribution and its moments.

For example, consider an $\mathbb{R}^n$-valued Ito process $\{X_t\}$, $t \in [0,T]$, satisfying the following SDE:

$$dX_t = a(X_t)dt + b(X_t)dW_t. \tag{2.3.1}$$

If one were able to solve (2.3.1) explicitly, the generation of sample paths and evaluation of functionals

$$\mathbb{E}[f(X)], \tag{2.3.2}$$

where a given function f depends on a solution $X = \{X_t, 0 \le t \le T\}$, would be an easy task.

Remark 2.3.1. For accident risk assessment function f would be of the following form:

$$f(X) = \mathbf{1}_{\left\{\inf\{t>0: X_t \in D\} \le T\right\}} = \mathbf{1}_{\{\tau \le T\}},$$

where a stopping time $\tau = \inf\{t > 0 : X_t \in D\}$ is a hitting time of a closed domain D. Then

$$\mathbb{E}[f(X)] = P(\tau \le T) = P_{hit}(0,T). \tag{2.3.3}$$

Unfortunately, very few specific SDE have explicitly known solutions. The well known exceptions are linear Gaussian SDE. A widely applicable approach to evaluating functionals like (2.3.2) and to generating sample paths of solutions of SDE is the simulation of sample paths of appropriate time discrete approximations. Usually this is done as follows.

For a given time interval $[0,T]$ we take a discretization $0 = t_0 < t_1 < \cdots < t_L = T$ with a constant step $h = T/L$. Then we should construct an appropriate approximating process $\left\{X_t^h\right\}$ for the process $\{X_t\}$. One of the simplest approximations is the explicit stochastic Euler approximation. This is a continuous time stochastic process $\{X_t^h\} = \left\{X_t^h,\ 0 \le t \le T\right\}$ satisfying the iterative scheme

$$X_{t_{n+1}}^h = X_{t_n}^h + a(X_{t_n}^h)(t_{n+1} - t_n) + b(X_{t_n}^h)(W_{t_{n+1}} - W_{t_n}), \tag{2.3.4}$$

$n = 0, 1, \ldots, L-1$, with initial value $X_0^h = X_0$. This scheme provides us with a method for pathwise approximation of the process $\{X_t\}$. Note that for a given time discretization the recursive Euler scheme (2.3.4) determines values of the approximating process $\{X_t^h\}$ at the discretization times only. If required, the values at the intermediate instants can be determined by an appropriate interpolation method.

Definition 2.3.2. We shall say that a time discrete approximation $\{X_t^h\}$ with maximum step size $h \in (0,T]$ *converges strongly* to $\{X_t\}$ if for all $t \in [0,T]$

$$\lim_{h \to 0} \mathbb{E}[|X_t^h - X_t|] = 0.$$

We shall also say that a time discrete approximation $\{X_t^h\}$ *converges strongly with order* $\gamma > 0$ to $\{X_t\}$ if there exist a positive constant C, which does not depend on h, such that for all $t \in [0,T]$

$$\mathbb{E}[|X_t^h - X_t|] \le Ch^\gamma,\ \ 0 < h \le T.$$

For computation of moments, probabilities or other functionals of the process we do not need so strong form of convergence as in the case of pathwise approximations. It is only required that the probability distribution of the processes $\{X_t\}$ and $\{X_t^h\}$ are sufficiently close to each other, but not necessarily the actual realizations of the random variables.

Definition 2.3.3. We shall say that a time discrete approximation $\{X_t^h\}$ with maximum step size $h \in (0, T]$ *converges weakly* to $\{X_t\}$ with respect to a wide class $\mathscr{C}$ of test functions f if for all $t \in [0, T]$

$$\lim_{h \to 0} \left| \mathbb{E}[f(X_t^h)] - \mathbb{E}[f(X_t)] \right| = 0, \text{ for all } f \in \mathscr{C}.$$

We shall also say that a time discrete approximation $\{X_t^h\}$ *converges weakly with order* $\gamma > 0$ to $\{X_t\}$ if there exist a positive constant C, which does not depend on h, such that for all $t \in [0, T]$

$$\left| \mathbb{E}[f(X_t^h)] - \mathbb{E}[f(X_t)] \right| \leq C h^\gamma, \;\; 0 < h \leq T.$$

Remark 2.3.4. In classical literature, class $\mathscr{C}$ is usually a class of sufficiently smooth functions. One can also consider a more general classes. For example, Gobet (1999*a*) has studied weak approximations for killed (or stopped) diffusion and he required functions f to be only bounded and measurable.

Recall that in our stochastic modelling framework the problem of accident risk assessment is reduced to the problem of estimating the functional (2.3.3), i.e. the probability that a process reaches a critical domain within a given time interval. It can be approximated using an appropriate weak time discrete approximation.

The general discrete-time numerical schemes for Itô diffusions are well explained in (Kloeden and Platen, 1992). A discretization scheme for jump-diffusion process with state-dependent intensities was considered in (Glasserman and Merener, 2004). Weak approximations of killed (or stopped) diffusions were studied in (Gobet, 1999*a*,*b*, 2000, 2001) and (Moon, 2003).

Suppose we have chosen a particular weak approximation $\{X_t^h\}$ (e.g. Euler) and want to use it in order to estimate a functional $\mathbb{E}[f(X)]$. We know that in theory $\mathbb{E}[f(X^h)] \approx \mathbb{E}[f(X)]$ with a certain accuracy when h is "small". Next, generate a large number of independent realizations $X^{h,i}$, $i = 1, 2, \ldots, N$, of random process $X^h = \{X_{t_n}^h, n = 1, 2, \ldots, L\}$ and calculate the sample average

$$\frac{1}{N} \sum_{i=1}^{N} f(X^{h,i}). \tag{2.3.5}$$

By Kolmogorov's strong law of large numbers[1]

$$\frac{1}{N} \sum_{i=1}^{N} f(X^{h,i}) \longrightarrow \mathbb{E}[f(X^h)] \text{ with probability 1, as } N \to \infty.$$

[1] provided that $\mathbb{E}[f(X^h)] < \infty$.

Hence, we can consider the sum (2.3.5) as an estimator of $\mathbb{E}[f(X)]$ with the following error:

$$\frac{1}{N}\sum_{i=1}^{N} f(X^{h,i}) - \mathbb{E}[f(X)]$$
$$= \Big(\frac{1}{N}\sum_{i=1}^{N} f(X^{h,i}) - \mathbb{E}[f(X^h)]\Big) + \Big(\mathbb{E}[f(X^h)] - \mathbb{E}[f(X)]\Big)$$
$$\triangleq \varepsilon_{stat} + \varepsilon_{sys}.$$

Here $\varepsilon_{sys} = \varepsilon_{sys}(h)$ is the *systematic error* and depends only on the approximation X^h. It tends to 0 as $h \to 0$. The random variable $\varepsilon_{stat} = \varepsilon_{stat}(\omega, h, N)$ is called the *statistical error*. If the number of independent simulations N is large then the Central Limit Theorem says that the statistical error is asymptotically Gaussian with mean zero and variance

$$Var(\varepsilon_{stat}) = \frac{1}{N} Var(f(X^h)).$$

Thus, if N tends to infinity the statistical error tends to 0.

For example, in order to obtain a reliable estimate of rare event probability of order 10^{-9}, we must run about 10^{11} independent simulations, which is very time consuming. To overcome this problem different efficient techniques for speeding up the simulations have been developed. In many applications the importance sampling techniques proved to be useful in reducing the variance of the simulated estimate and hence reducing the computational effort required to achieve a fixed degree of relative accuracy. The effectiveness of such method depends critically on the ability to find the right change of measure. If it is done improperly, the importance sampling may produce worse results than straightforward simulation. Finding the right change of measure generally requires identifying at least the rough asymptotics of the quantity to be estimated. This type of analysis can be formidable in complex dynamic models. See (Heidelberger, 1995; Shahabuddin, 1995; Liu, 2003) and (Juneja and Shahabuddin, 2006) for surveys.

2.3.2 Splitting techniques

In this subsection we discuss a class of splitting techniques for rare event estimation. The simulation approaches that are based on a splitting of trajectories require little analysis of the structure of the stochastic dynamic system under consideration. The idea is to express the small probability of rare event to be estimated as the product of a certain number of larger probabilities, which can be efficiently estimated by the MC methods. This can be achieved by introducing sets of intermediate states that are visited one after the other, in an ordered sequence, before reaching the final set of rare event states. The probability of rare

event is then given by the product of the conditional probabilities of reaching a set of intermediate states given that the previous set of intermediate states have been reached. Each conditional probability is estimated by simulating in parallel several copies of the system, i.e. each copy is considered as a particle following the trajectory generated through the system dynamics. Each particle branches (i.e. the trajectory splits into a number of independent subpaths, which subsequently evolve independently of each other) as soon as it enters the intermediate states, which is usually characterized by crossing a threshold by a control parameter. Reaching intermediate states is more likely than reaching the rare event states, and by splitting at each threshold the chances to reach the rare event states are increasing.

For example, let us consider a multidimensional diffusion process which is assumed to start in some Borel set $D_0 \subset \mathbb{R}^n$ with a given initial probability $P_{X_0}(\cdot)$:

$$dX_t = a(X_t)dt + b(X_t)dW_t, \tag{2.3.6}$$

where $a(x) : \mathbb{R}^n \to \mathbb{R}^n$, $b(x) : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ and $\{W_t\}$ is a Wiener process in $\mathbb{R}^n$ independent of X_0. For a given target Borel set $D \subset \mathbb{R}^n$ ($D \cap D_0 = \varnothing$), we define the first time the process $\{X_t\}$ hits D, namely

$$\tau_D = \inf\{t \geq 0 : X_t \in D\},\ \tau_D = \infty \text{ if this set is empty.}$$

We would like to estimate the quantity

$$P_{hit}(0, T) = P(\tau_D \leq T) \tag{2.3.7}$$

for some $T < \infty$, i.e. the probability that diffusion $\{X_t\}$ will hit the rare event set D before time T.

Let us introduce a sequence of nested Borel sets[2]

$$D = D_m \subset \cdots \subset D_1 \tag{2.3.8}$$

where D_k is a closed Borel set of $\mathbb{R}^n$, and D_1 such that $D_1 \cap D_0 = \varnothing$. The first moment that $\{X_t\}$ hits a set D_k is defined as the stopping time:

$$\tau_k \triangleq \inf\{t \geq 0 : X_t \in D_k\},$$

$\tau_k = \infty$ if this set is empty. Note that $P(\tau_D \leq T) = P(\tau_m \leq T)$. The process $\{X_t\}$, before hitting D, passes through a sequence of nested sets (2.3.8). An implicit assumption is that set D_{k+1} can not be reached from set D_{k-1} without reaching D_k, $k = 1, 2, \ldots, m-1$. Reaching the intermediate sets D_k's is more likely than reaching the rare event set D. Next, one can express the probability

[2]we will also use the term "sequence of nested level sets" or just "level sets"

of rare event $P_{hit}(0,T)$ as a product of conditional probabilities of intermediate "less rare" events leading to it (e.g. see Chapter 5):

$$P_{hit}(0,T) = \prod_{k=1}^{m} P(\tau_k \leq T | \tau_{k-1} \leq T). \tag{2.3.9}$$

All splitting techniques have in common that they all use the same idea of state space decomposition. The main differences lie in how the trajectories are being split and in the existence of convergence proofs. In what follows we give a brief description of the most interesting splitting approaches. For a complete review and detailed list of references see (Glasserman *et al.*, 1999; Lezaud *et al.*, 2004; Cérou *et al.*, 2005).

Multilevel splitting and RESTART

The multilevel splitting approach proceeds as follows. From the initial state $X_0 \in D_0$ we generate a trajectory until it reaches the set D_1. If the set D_1 is hit we split the trajectory into R_1 subtrajectories (trials) which independently continue evolution until the next set D_2 is hit. This procedure is repeated for all sets D_k, $k = 1, \ldots, m-1$, i.e. we split the trajectory into R_k subtrajectories (trials) each time the set D_k is hit. If a trajectory can not reach the rare event set D during the fixed time interval $[0,T]$ then we stop it. We independently repeat the whole procedure R_0 times, thus in total we consider $R_0 R_1 \ldots R_{m-1}$ trials. Note that if the process failed to hit the set D_k at the k-th step then in total $R_k \ldots R_{m-1}$ possible retrials have failed. Let I_D denote the total number of trajectories having reached the set D. Then, an unbiased estimator of $P_{hit}(0,T) = P(\tau_D \leq T)$ is given by

$$P_{hit}(0,T) \approx \frac{I_D}{R_0 \prod_{i=1}^{m-1} R_i}.$$

Villén-Altamirano and Villén-Altamirano (1991, 1994) describe a slightly different implementation of the multilevel splitting technique. In their empirical method called RESTART, a trajectory splits every time it crosses a nested set D_k, $k = 1, \ldots, m-1$, even the one it has reached before. The RESTART approach can also be used to estimate rare transient events not only the probability to reach the rare event set.

The central issue in implementing these methods is choosing the nested sets (2.3.8) and choosing the number of subtrajectories (i.e. $R_0, R_1, \ldots, R_{m-1}$) to generate when a trajectory splits. Too many splittings results in explosive computational requirements, and too few splittings eliminates any reduction in variance of the estimator, i.e. the resulting algorithm provides no speed up gains in comparison with the standard Monte Carlo simulation. Glasserman *et al.* (1999) analyzed the multilevel splitting method for a class of simple models. Under certain assumptions they identified the optimal degree of splitting of trajectories at each level set D_k, as the rarity of the event increases: it should be set so that the

expected number of subtrajectories reaching each level set D_k remains roughly constant. Among the restrictions they have imposed to obtain this result, the most significant is the requirement that there be either only finitely many ways of achieving each level set D_k or only one component of the process can take values in space $\mathbb{R}$ and the other components can assume only a finite number of values. Similar results were also obtain by Villén-Altamirano and Villén-Altamirano (1997) for their empirical method RESTART. Of course, these restricting requirements do not mean that the multilevel splitting and the RESTART techniques can not be used for rare event estimation in complex high dimensional dynamical systems. However, in this case we do not know anything about the choice of number of splittings and we also do not know how the rate of convergence of these approaches depends on the number of splittings. For RESTART we do not even know if there is convergence at all. This is much worse than not knowing the rate of convergence.

Interacting Particle Systems Approach

Another group of multilevel splitting techniques has been recently studied by Cérou *et al.* (2002); Del Moral (2004); Cérou *et al.* (2005); Le Gland and Oudjane (2005, 2006); Krystul and Blom (2005*b*, 2006); Del Moral and Lezaud (2006). These new splitting techniques are based on the well developed theory of branching and interacting particle systems (IPS) approximations of Feynman-Kac formulae (see (Del Moral, 2004)). The advantage of the IPS based approaches is that the existing precise and general results can be extended to the rare event analysis. In what follows we describe the idea behind the classical IPS approach of Cérou *et al.* (2002).

Let us denote $E' = \mathbb{R}^n$, and let $\mathscr{E}'$ be the Borel σ-algebra of E'. Recall that we consider the diffusion process described by SDE (2.3.6). It starts in a Borel set D_0 and, like in classical multilevel splitting, before hitting the target set D it has to pass through a sequence of nested Borel sets (2.3.8). To capture how $\{X_t\}$ enters each nested set $D = D_m \subset \cdots \subset D_1$ before finally hitting the target set D, we introduce the discrete time process $\{\xi_k, k = 0, 1, \ldots, m\}$ with values in space E' defined by $\xi_k \triangleq (X_{\tau_k \wedge T})$. By the strong Markov property of $\{X_t\}$, the process $\{\xi_k, k = 0, 1, \ldots, m\}$ is a Markov chain with transition kernel $\mathscr{Q}(\xi, d\xi') = P(\xi_k \in d\xi' | \xi_{k-1} = \xi)$. Now let us define the following conditional probabilities:

$$\pi_k(B) \triangleq P(\xi_k \in B | \tau_1 \leq T, \ldots, \tau_k \leq T),$$
$$p_k(B) \triangleq P(\xi_k \in B | \tau_1 \leq T, \ldots, \tau_{k-1} \leq T),$$

for any $B \in \mathscr{E}'$. It is easy to see that

$$p_k(B) = \int_{E'} \mathscr{Q}(\xi, B)\pi_{k-1}(d\xi) \text{ for all } B \in \mathscr{E}', \tag{2.3.10}$$

$$\pi_k(B) = \frac{\int_B \mathbf{1}_{\{\xi \in D_k\}} p_k(d\xi)}{\int_{E'} \mathbf{1}_{\{\xi' \in D_k\}} p_k(d\xi')} \text{ for all } B \in \mathscr{E}', \tag{2.3.11}$$

and

$$\gamma_k \triangleq P(\tau_k \leq T | \tau_{k-1} \leq T) = \int_{E'} \mathbf{1}_{\{\xi \in D_k\}} p_k(d\xi). \tag{2.3.12}$$

Recall that we want to estimate

$$P_{hit}(0,T) = \prod_{k=1}^{m} \gamma_k. \tag{2.3.13}$$

The evolution of the flow $\{\pi_k, p_k, \gamma_k; k = 0, 1, \ldots, m\}$ is described by the following diagram:

$$\pi_{k-1} \xrightarrow{\text{prediction}} p_k \xrightarrow{\text{conditioning}} \pi_k$$
$$\downarrow$$
$$\gamma_k$$

with initial condition $\pi_0(d\xi) = P_{\xi_0}(d\xi) = P(\xi_0 \in d\xi)$. In that way, each of the m terms γ_k in (2.3.13) is characterized as a solution of a sequence of Equations (2.3.10)-(2.3.12).

The IPS approach is based on the idea to approximate the flow $\{\pi_k, p_k, \gamma_k; k = 0, 1, \ldots, m\}$ by an approximating sequence $\{\pi_k^{N_p}, p_k^{N_p}, \gamma_k^{N_p}; k = 0, 1, \ldots, m\}$ which is described by the following diagram

$$\pi_{k-1}^{N_p} \xrightarrow{\text{prediction}} p_k^{N_p} \xrightarrow{\text{conditioning}} \pi_k^{N_p}$$
$$\downarrow$$
$$\gamma_k^{N_p}$$

with initial condition

$$\pi_0 \approx \pi_0^{N_p} = \frac{1}{N_p} \sum_{i=1}^{N_p} \delta_{\{\xi_0^i\}},$$

and approximations

$$p_k \approx p_k^{N_p} = \frac{1}{N_p} \sum_{i=1}^{N_p} \delta_{\{\xi_k^i\}},$$

$$\pi_k \approx \pi_k^{N_p} = \frac{1}{|I_k^{N_p}|} \sum_{i \in I_k^{N_p}} \delta_{\{\xi_k^i\}},$$

$$\gamma_k \approx \gamma_k^{N_p} = \frac{|I_k^{N_p}|}{N_p} = \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbf{1}_{\{\xi_k^i \in D_k\}},$$

in the form of the empirical measures associated with the particle system $\{t_k^i, \xi_k^i\}_{i=1}^{N_p}$, where N_p denotes the number of particles, $I_k^{N_p} = \{1 \leq i \leq N_p : \xi_k^i \in D_k\}$ is the set of the labels of the particles having succeeded to reach the k-th level set D_k and $|I_k^{N_p}|$ denotes the cardinal number of the set $I_k^{N_p}$. Each particle is a pair consisting of time index $t_k \triangleq \tau_k \wedge T$ and state ξ_k.

The approximation procedure goes as follows. At $t = 0$ we start with the empirical measure

$$\pi_0^{N_p} = \frac{1}{N_p} \sum_{i=1}^{N_p} \delta_{\{\xi_0^i\}},$$

where $\{\xi_0^i\}_{i=1}^{N_p}$ are independent samples from initial distribution $P_{\xi_0}(\cdot)$. From $t = 0$ to τ_1, and from τ_{k-1} to τ_k each particle evolves stochastically according to Equation (2.3.6) with initial condition (t_{k-1}^i, ξ_{k-1}^i) (prediction step) until it reaches the next level set D_k or the final time T. Let $\{\hat{t}_k^i, \hat{\xi}_k^i\}_{i=1}^{N_p}$ denote the values of the particles after the k-th prediction. Then the empirical distribution $p_k^{N_p}$ associated with the predicted cloud of particles is:

$$p_k^{N_p} = \frac{1}{N_p} \sum_{i=1}^{N_p} \delta_{\{\hat{\xi}_k^i\}}.$$

The particles which do not reach the set D_k before time T are deleted, and the labels of particles that have reached D_k are saved in the set $I_k^{N_p}$. Then

$$\gamma_k^{N_p} = \frac{|I_k^{N_p}|}{N_p}.$$

If all particles become deleted, i.e. $\gamma_k^{N_p} = 0$, then the algorithm is stopped and $P_{hit}(0, T) \approx 0$. The empirical distribution $\pi_k^{N_p}$ associated with the measurement updated cloud of particles is:

$$\pi_k^{N_p} = \frac{1}{|I_k^{N_p}|} \sum_{i \in I_k^{N_p}} \delta_{\{\tilde{\xi}_k^i\}}, \tag{2.3.14}$$

with

$$\tilde{\xi}_k^i = \hat{\xi}_k^i, \quad i = 1, \ldots, N_p.$$

Particles having reached the set D_k are used for a resampling step. We resample with replacement N_p independent particles according to empirical measure (2.3.14). After this step we again have N_p particles $\{t_k^i, \xi_k^i\}_{i=1}^{N_p}$ at level D_k. Next, set $k := k + 1$ and repeat the same simulation procedure iteratively until the target set D is reached. For $k = m$ we have

$$P_{hit}(0, T) \approx \prod_{k=1}^{m} \gamma_k^{N_p}.$$

In (Cérou *et al.*, 2002) it is proven that the particle estimates are unbiased, i.e.

$$\mathbb{E}\big[\prod_{k=1}^{m}\gamma_k^{N_p}\big] = P(\tau_m \le T) = P_{hit}(0,T)$$

and the rate of convergence is of order $\sqrt{N_p}$:

$$\Big(\mathbb{E}\Big[\big|\prod_{k=1}^{m}\gamma_k^{N_p} - \prod_{k=1}^{m}\gamma_k\big|^q\Big]\Big)^{\frac{1}{q}} \le \frac{a_q b_m}{\sqrt{N_p}},$$

finite constants a_q and b_m depend only on the parameters q and m respectively.

This means that by running the IPS algorithm many times (let us say N times) and by taking the sample average of random independent realizations of $Z \triangleq \prod_{k=1}^{m}\gamma_k^{N_p}$ we obtain an unbiased and consistent estimator of $P_{hit}(0,T)$:

$$P_{hit}(0,T) \approx \frac{1}{N}\sum_{i=1}^{N} Z^i,$$

it converges a.s. to $P_{hit}(0,T)$ as $N \to \infty$, by the strong law of large numbers.

2.3.3 Numerical example: IPS for diffusion

To illustrate the whole potential of the IPS simulation method we apply it to a special test example for which we know the exact analytical solution. We consider the Geometric Brownian motion process driven by the following SDE:

$$dX_t = \big(\mu + \frac{\sigma^2}{2}\big)X_t dt + \sigma X_t dW_t,\ X_0 = x \tag{2.3.15}$$

where μ and σ are strictly positive constants. We want to estimate the probability that process $\{X_t\}$ will hit barrier d before time T, i.e. $P(\tau_d \le T)$ where $\tau_d \triangleq \inf\{t > 0 : X_t \in [d, +\infty);\ X_0 = x\}$. Although the above model is rather simple, it allows us to check the IPS method accuracy. This is possible thanks to the following analytical formula (Tuckwell and Wan, 1984):

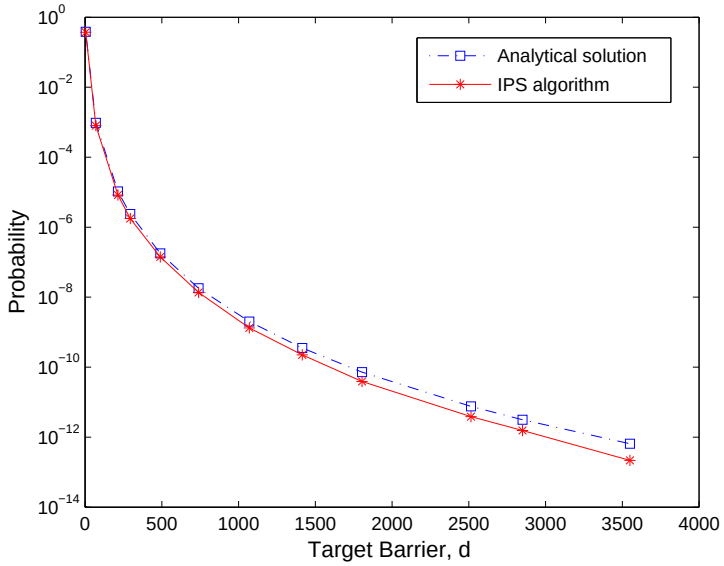
$$P(\tau_d \le t) = \int_0^t \frac{\ln(d/x)}{\sqrt{2\pi\sigma^3 s^3}} \exp\Big\{\frac{-(\ln(d/x) - \mu s)^2}{2\sigma^2 s}\Big\} ds. \tag{2.3.16}$$

Before starting the simulations we should define the nested sequence of sets: $D = D_m \subset \cdots \subset D_1$. For $\{D_j,\ j = 1,\ldots,m\}$ we choose an increasing sequence of real numbers $\{d_j,\ j = 1,\ldots,m\}$, with $0 < d_1 < \cdots < d_{m-1} < d_m = d$ and take $D_j = [d_j, +\infty)$. In our example we choose numbers d_j experimentally so that approximately $40\% - 50\%$ of particles started at level D_{j-1} manage to reach level D_j, $j = 1,\ldots,m$. In Figure 2.1(a) we see the probability $P(\tau_d \le T)$ as a function

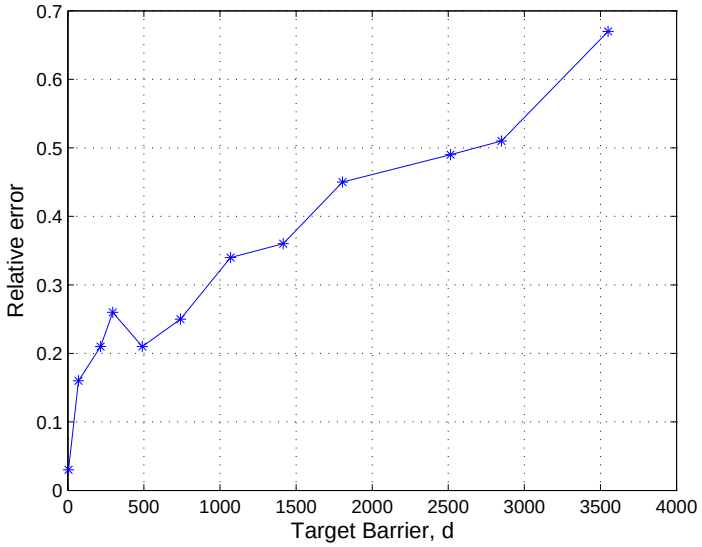
of d. The blue curve with squares is the numerically computed theoretical value, and the red curve with asterisks is the Monte-Carlo simulation result, with 1000 runs of 1000 particles each:

$$P_{hit}(0,T) \approx \frac{1}{1000} \sum_{i=1}^{1000} Z^i,$$

where Z^i denotes the i-th independent realization of random variable $Z = \prod_{k=1}^{m} \gamma_k^{N_p}$. We use the Euler scheme with 500 discretization time steps. The parameters of Geometric Brownian motion process are $\mu = 1$, $\sigma = 1$ and $X_0 = 1$. The largest value of d was 3550. This means that the probability for the process started at $X_0 = 1$ to reach the desired level is approximately 6.48×10^{-13}, so a standard Monte Carlo approach would not be practical. From Table 2.1 or Figure 2.1(b) we can see that the method works quite well; the relative error is less than 60%. We still get quite satisfactory results even with a fewer number of simulation runs. We repeat the above example with the same parameters, but instead of 1000 runs with 1000 particles now we run only 100 with 1000 particles. The results can be seen in Figure 2.2 and in Table 2.2.



(a) Probability to hit barrier d before time $T = 1$. 1000 runs with 1000 particles



(b) Relative error

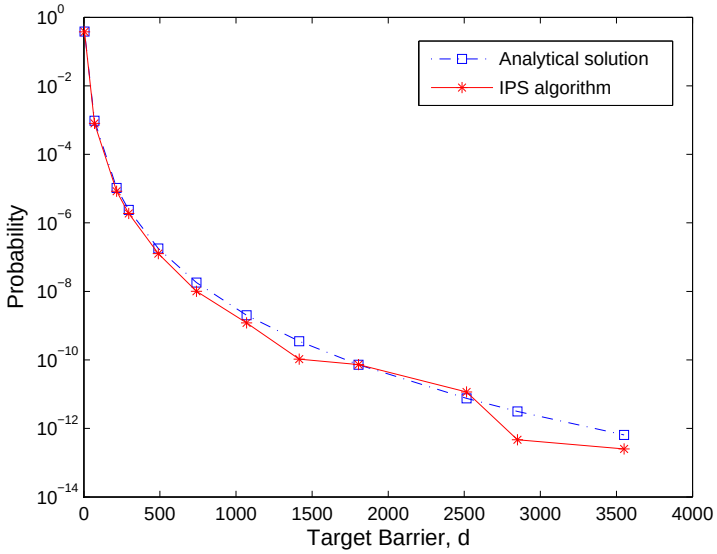
Figure 2.1: Simulation results: (a) Probability to hit barrier d before time $T = 1$. 1000 runs with 1000 particles; (b) Relative error

Table 2.1: Probability to hit barrier d before time $T = 1$. 1000 runs with 1000 particles

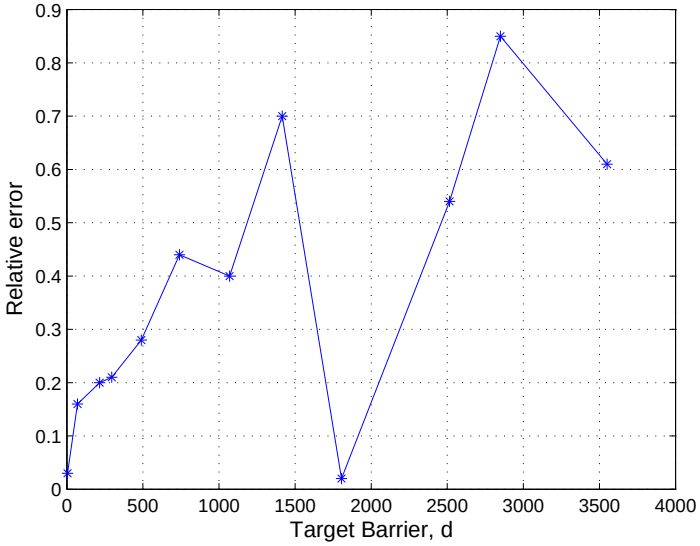
Target Level, d_j	Level's number, j	Probability to hit barrier d_j Analytical	Probability to hit barrier d_j IPS algorithm	Relative Error
5	2	0.38448	0.372872	0.03
70	12	0.00095	8.01E-04	0.16
215	19	1.05e-05	8.29E-06	0.21
295	21	2.38E-06	1.77E-06	0.26
490	26	1.78E-07	1.41E-07	0.21
740	31	1.80E-09	1.36E-08	0.25
1070	36	2.02E-09	1.32E-09	0.34
1415	40	3.51E-10	2.25E-10	0.36
1805	45	7.19E-11	3.92E-11	0.45
2515	51	7.55E-12	3.87E-12	0.49
2850	54	3.14E-12	1.55E-12	0.51
3550	58	6.48E-13	2.17E-13	0.67

Table 2.2: Probability to hit barrier d before time $T = 1$. 100 runs with 1000 particles

Target Level, d_j	Level's number, j	Probability to hit barrier d_j Analytical	Probability to hit barrier d_j IPS algorithm	Relative Error
5	2	0.38448	0.374263	0.03
70	12	0.00095	7.99E-04	0.16
215	19	1.05e-05	8.43E-06	0.20
295	21	2.38E-06	1.88E-06	0.21
490	26	1.78E-07	1.27E-07	0.28
740	31	1.80E-09	1.01E-08	0.44
1070	36	2.02E-09	1.21E-09	0.40
1415	40	3.51E-10	1.05E-10	0.70
1805	45	7.19E-11	7.30E-11	0.02
2515	51	7.55E-12	1.16E-11	0.54
2850	54	3.14E-12	4.67E-13	0.85
3550	58	6.48E-13	2.52E-13	0.61



(a) Probability to hit barrier d before time $T = 1$. 100 runs with 1000 particles



(b) Relative error

Figure 2.2: Simulation Results: (a) Probability to hit barrier d before time $T = 1$. 100 runs with 1000 particles; (b) Relative error

2.4 Concluding Remarks

In this chapter we have discussed the problem of accident risk assessment in stochastic dynamical systems, using relative position/velocity equations in ATM. We have explained that collision risk in ATM coincides with hitting probability in stochastic analysis. In case when the dimension of stochastic process describing the dynamical system is at most three, then the hitting probability can be numerically evaluated through solving a backward Kolmogorov PDE with Dirichlet type boundary conditions. For complex real-world applications Monte Carlo simulation remains the only feasible approach. Of course, it is unrealistic to estimate extremely small rare event probabilities just by using a naive Monte Carlo simulation. When the extremely small probabilities are considered, one has to use special techniques which allow to speed up the simulation. Importance sampling and multilevel splitting methods have been used to obtain dramatic improvements in efficiency in estimating small probabilities in queueing and reliability systems. However, these two methods have serious drawbacks. The importance sampling techniques, based on changing probability distributions to make rare events less rare, depend critically on the ability to find the right change of measure. Finding the right change of measure in complex dynamical models is practically impossible. The efficiency of multilevel splitting techniques depends on the choice of the level sets and especially on the optimal number of splittings. Too many splittings results in explosive computational requirements, and too few splittings provides no speed-up gains over straightforward Monte Carlo simulation. So far, the theoretical results regarding the optimal choice of splittings have been obtained only for very simple cases (Glasserman *et al.*, 1999; Villén-Altamirano and Villén-Altamirano, 1997). On the other hand, the IPS based approaches do not have such restrictive limitations as do multilevel splitting techniques have. Moreover, the IPS approaches have the advantage of being based on sound theory with many general results available such as proof of convergence and error estimates (Del Moral, 2004). Our numerical test of IPS approach showed that it has a great potential. In Chapter 5 we will extend it to a Hybrid IPS approach which is aimed for efficient estimation of rare event probabilities in stochastic hybrid systems with rare switches and large differences in mode probabilities.

Chapter 3

Generalized stochastic hybrid processes as strong solutions to stochastic differential equations

3.1 Introduction

In a series of recent studies by Blom (2003); Blom *et al.* (2003*b*); Ghosh and Bagchi (2004) and Krystul *et al.* (2006), several distinct classes of stochastic hybrid processes have been developed as solutions of SDE on hybrid state spaces. These classes have different modelling power and cover a wide range of interesting phenomena (see Table 3.1), though, all they contain, as a subclass, the switching diffusion processes of Ghosh *et al.* (1997). The entries in Table 3.1 have the following meaning:

- Switching diffusion: between the random switches of the discrete valued component, the Euclidean valued component evolves as diffusion.
- Random hybrid jumps: simultaneous and dependent jumps and switches of discrete and Euclidean valued components are driven by a Poisson random measure.
- Boundary hybrid jumps: simultaneous and dependent jumps and switches of discrete and Euclidean valued components are initiated by boundary hittings.
- Martingale inducing jumps: the Euclidean valued components driven by a compensated Poisson random measure may jump so frequently that it is no

longer a process of finite variation.

- Mode dependent dimension: the dimension of the Euclidean state space depends on the discrete valued component (i.e. the mode).

Table 3.1: Classes of stochastic hybrid processes

	Switching Diffusion	Random Hybrid Jumps	Boundary Hybrid Jumps	Martingale Inducing Jumps	Mode dependent dimension
Blom (2003) Ghosh and Bagchi (2004)	✓	✓	-	-	-
Blom *et al.* (2003*b*) Krystul *et al.* (2006)	✓	✓	✓	-	-
Ghosh and Bagchi (2004)	✓	-	✓	-	✓
Krystul *et al.* (2006)	✓	✓	-	✓	-

Blom (2003) studied the jump-diffusion SDE of Lepeltier and Marchal (1976) driven by Brownian motion and Poisson random measure. He placed the SDE of Lepeltier and Marchal (1976) on a hybrid state space and showed that strong uniqueness and existence, semimartingale and Markov properties identified by Lepeltier and Marchal (1976) carry over to the resulting stochastic hybrid process. This process includes diffusion, independent random jumps and switches of Euclidean valued and discrete valued components, and also the hybrid jumps. Hybrid jumps are special types of jumps where discontinuity in the Euclidean valued process components happens synchronous with a discontinuity in the discrete valued process component. The proof of the strong Markov property by Lepeltier and Marchal (1976) was sufficient to be carried over to the hybrid state space situation considered by Blom (2003), however it did not carry over to the stochastic hybrid processes with instantaneous jumps at the boundaries considered by Blom *et al.* (2003*b*).

Ghosh and Bagchi (2004) have developed two extensions to the switching diffusion model of Ghosh *et al.* (1997). Their first model, just as the model of Blom (2003), includes diffusion, hybrid jumps and independent jumps and switches of Euclidean valued and discrete valued components. The second model is the switching diffusion with instantaneous jumps at the boundaries and with possibility to reset dimensions of the hybrid state space after the jump. To prove the existence and uniqueness theorems for both models Ghosh and Bagchi (2004) used the method from (Ikeda and Watanabe, 1989). However, it has not been clarified if the strong Markov property holds and if the solutions are semimartingales or not.

The aim of this chapter is to significantly further the study of SDE on a hybrid state space, including characterizations of their solutions in terms of pathwise

uniqueness, semimartingale and strong Markov process properties. First, using (Gihman and Skorohod, 1982; Jacod and Shiryaev, 1987), we identify and characterize a very general class of jump-diffusions which are defined as semimartingale solutions of SDE. From this point on we follow a similar path as taken by (Blom, 1990, 2003) in transferring this pathwise uniqueness and semimartingale understanding to the class of stochastic hybrid processes. This subsequently allows to incorporate instantaneous jumps at a boundary within the same framework including pathwise uniqueness and semimartingale property. Finally we prove the strong Markov property for the stochastic hybrid process with instantaneous jumps at a boundary by considering it as a concatenation of killed strong Markov processes. To prove that this concatenation forms a strong Markov process we follow a novel and mathematically constructive approach.

This chapter is organized as follows. Section 3.2 provides a brief introduction to semimartingales. Section 3.3 presents the existence and uniqueness results for $\mathbb{R}^n$-valued jump-diffusions. Section 3.4 extends these results to hybrid state processes with Poisson and hybrid Poisson jumps. In Section 3.5 we characterize a general stochastic hybrid process which includes jumps at the boundaries. Section 3.6 presents a comparison of different stochastic hybrid models. Finally, the Markov and the Strong Markov properties for a general stochastic hybrid process are shown in Sections 3.7 and 3.8.

3.2 Semimartingales and characteristics

In this section following (Jacod and Shiryaev, 1987) we provide basic results concerning semimartingales, their canonical representation and their relation with the large class of SDE to be studied in this chapter.

Throughout this chapter we assume that a probability space $(\Omega, \mathscr{F}, P)$ is equipped with a right-continuous filtration $(\mathscr{F}_t)_{t\geq 0}$. The stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$ is called complete if the σ-algebra $\mathscr{F}$ is P-complete and if every $\mathscr{F}_t$ contains all P-null sets of $\mathscr{F}$. Note that it is always possible to "complete" a given stochastic basis, if it is not complete, by adding all subsets of P-null sets to $\mathscr{F}$ and $\mathscr{F}_t$. We will therefore assume throughout this chapter that the stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$ is complete.

The *predictable σ-algebra* is the σ-algebra $\mathscr{P}$ on $\Omega \times \mathbb{R}_+$ that is generated by all left-continuous adapted processes (considered as mappings from $\Omega \times \mathbb{R}_+$ into some set E). A process or random set that is $\mathscr{P}$-measurable is called *predictable*.

Definition 3.2.1. The canonical setting. Ω is the "canonical space" (also denoted by $\mathbb{D}(\mathbb{R}^n)$) of all càdlàg (right-continuous and admit left hand limits) functions $\omega : \mathbb{R}_+ \to \mathbb{R}^n$; X is the "canonical process" defined by $X_t(\omega) = \omega(t)$; $\mathscr{H} = \sigma(X_0)$; finally $(\mathscr{F}_t)_{t\geq 0}$ is generated by X and $\mathscr{H}$, by which we mean:

(i) $\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}^0_s$ and $\mathscr{F}^0_s = \mathscr{H} \vee \sigma(X_r : r \leq s)$ (in other words, $(\mathscr{F}_t)_{t\geq 0}$ is

the smallest filtration such that X is adapted and $\mathscr{H} \subset \mathscr{F}_0$);

(ii) $\mathscr{F} = \mathscr{F}_{\infty-} (= \bigvee_t \mathscr{F}_t)$.

Throughout this chapter we assume that canonical setting of Definition 3.2.1 is in force. The $\mathbb{R}^n$-valued càdlàg stochastic process $\{X_t\}$ defined on a probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$ is a *semimartingale* if X_t admits a decomposition of the form

$$X_t = X_0 + A_t + M_t,\ t \geq 0, \tag{3.2.1}$$

where X_0 is a finite-valued and $\mathscr{F}_0$-measurable, $\{A_t\} \in \mathscr{V}^n$ is a process of *bounded variation*, $\{M_t\} \in \mathscr{M}^n_{loc}$ is an n-dimensional *local martingale* starting at 0, and for each $t \geq 0$, A_t and M_t are $\mathscr{F}_t$-measurable. Recall that $\{M_t\} \in \mathscr{M}^n_{loc}$ if and only if there exists a sequence of $(\mathscr{F}_t)_{t\geq 0}$-stopping times $(\tau_k)_{k\geq 1}$ such that $\tau_k \uparrow \infty$ (P-a.s.) for $k \longrightarrow \infty$ and for each $k \geq 1$, the *stopped process*

$$\{M_t^{\tau_k}\} \text{ with } M_t^{\tau_k} = M_{t\wedge\tau_k},\ k \geq 1, \tag{3.2.2}$$

is a *martingale*:

$$\mathbb{E}|M_t^{\tau_k}| < \infty,\ \mathbb{E}[M_t^{\tau_k} \mid \mathscr{F}_s] = M_s^{\tau_k}\ (P - \text{a.s.}),\ s \leq t. \tag{3.2.3}$$

Denote by $\mu = \mu(\omega; ds, dx)$ the measure describing the jump structure of $\{X_t\}$:

$$\mu(\omega; (0,t] \times B) = \sum_{0<s\leq t} I_{\{\omega:\Delta X_s(\omega)\in B\}}(\omega),\ t > 0, \tag{3.2.4}$$

where $B \in \mathscr{B}(\mathbb{R}^n \setminus \{0\})$, i.e. the σ-algebra of Borel sets on $\mathbb{R}^n \setminus \{0\}$, $\Delta X_s = X_s - X_{s-}$ and $I_{\{\omega:\Delta X_s(\omega)\in B\}}(\omega)$ is the indicator function of set $\{\omega : \Delta X_s(\omega) \in B\}$. By $\nu = \nu(\omega; ds, dx)$ we denote a compensator of μ which is a predictable measure (unique up to a P-null set) with the property that $\mu - \nu$ is a local martingale measure. This means that for each $B \in \mathscr{B}(\mathbb{R}^n \setminus \{0\})$:

$$(\mu(\omega; (0,t] \times B) - \nu(\omega; (0,t] \times B))_{t>0} \tag{3.2.5}$$

is a local martingale with value 0 for $t = 0$.

A semimartingale $\{X_t\}$ is called *special* if there exists a decomposition (3.2.1) with a *predictable* process $\{A_t\}$. Every semimartingale with *bounded jumps* ($|\Delta X_t(\omega)| \leq b < \infty, \omega \in \Omega, t > 0$) is special (see Jacod and Shiryaev, 1987, Chapter I, 4.24).

Let h be a truncation function, i.e. $\Delta X_s - h(\Delta X_s) \neq 0$ if and only if $|\Delta X_s| > b$ for some $b > 0$. Hence

$$\widetilde{X}_t = \sum_{0<s\leq t} (\Delta X_s - h(\Delta X_s)) \tag{3.2.6}$$

denotes the jump part of $\{X_t\}$ corresponding to *large jumps*. The number of the large jumps is a.s. finite on $[0,t]$, for all $t>0$, because for all semimartingales (Jacod and Shiryaev, 1987, Chapter I, 4.47)

$$\sum_{0<s\leq t}(\Delta X_s)^2<\infty,\ P-a.s. \tag{3.2.7}$$

The process $\{X_t-\widetilde{X}_t\}$ is a semimartingale with *bounded jumps* and hence it is special:

$$X_t-\widetilde{X}_t=X_0+\widetilde{B}_t+\widetilde{M}_t \tag{3.2.8}$$

where $\{\widetilde{B}_t\}$ is a predictable process and $\{\widetilde{M}_t\}$ is a local martingale. The "tilde" above the process denotes the dependence on the truncation function h.

Every local martingale $\widetilde{M}_t$ can be decomposed as:

$$\widetilde{M}_t=M_t^c+\widetilde{M}_t^d \tag{3.2.9}$$

where M_t^c is a *continuous* (martingale) part and $\widetilde{M}_t^d$ is a *purely discontinuous* (martingale) part which satisfies:

$$\widetilde{M}_t^d=\int_0^t\int h(x)(\mu(ds,dx)-\nu(ds,dx)). \tag{3.2.10}$$

Note that the continuous martingale part M_t^c does not depend on h. By definition of μ and $\{\widetilde{X}_t\}$ we have

$$\widetilde{X}_t=\int_0^t\int(x-h(x))\mu(ds,dx). \tag{3.2.11}$$

Consequently, substitution of (3.2.9) - (3.2.11) into (3.2.8) yields the following canonical representation of semimartingale $\{X_t\}$:

$$X_t=X_0+\widetilde{B}_t+M_t^c+\int_0^t\int h(x)(\mu(ds,dx)-\nu(ds,dx))+\int_0^t\int(x-h(x))\mu(ds,dx). \tag{3.2.12}$$

Next we may assume $h(x)=x\cdot I_{\{x\,:\,|x|<1\}}(x)$ and replace $\widetilde{B}_t$ by B_t. Then (3.2.12) takes on the form:

$$X_t=X_0+B_t+M_t^c+\int_0^t\int_{|x|<1}x(\mu(ds,dx)-\nu(ds,dx))+\int_0^t\int_{|x|\geq 1}x\mu(ds,dx). \tag{3.2.13}$$

We denote by $\langle M_t^c\rangle$ the predictable quadratic variation of $\{M_t^c\}$, hence $(M_t^c)^2-\langle M_t^c\rangle$ is a local martingale.

We call the *characteristics* associated with h of the semimartingale $\{X_t\}$ (if there may be an ambiguity on h) the triplet (B_t,C_t,ν) consisting of:

(i) A predictable process $B_t = (B_t^i)_{i\leq n}$ in $\mathscr{V}^n$, namely the process $B_t = \widetilde{B}_t$ appearing in (3.2.8);

(ii) A continuous process $C_t = (C_t^{ij})_{i,j\leq n}$ in $\mathscr{V}^{n\times n}$, namely $C_t = \langle M_t^c\rangle$;

(iii) A predictable random measure ν on $\mathbb{R}_+ \times \mathbb{R}^n$, namely the compensator of random measure μ associated to the jumps of X by (3.2.4).

Definition 3.2.2. Jump diffusion. $\{X_t\}$ is called a *jump diffusion* on $(\Omega, \mathscr{F}, (\mathscr{F})_{t\geq 0}, P)$ if it is a semimartingale with the following characteristics:

$$\begin{cases} B_t^i(\omega) = \int_0^t \alpha^i(s, X_s(\omega))ds & (= +\infty \text{ if the integral diverges}) \\ C_t^{ij}(\omega) = \int_0^t \beta^{ij}(s, X_s(\omega))ds & (= +\infty \text{ if the integral diverges}) \\ \nu(\omega; dt \times dx) = dt \times K_t(\omega, X_t(\omega), dx) \end{cases} \tag{3.2.14}$$

where:

$$\begin{cases} \alpha : \mathbb{R}_+ \times \mathbb{R}^n \longrightarrow \mathbb{R}^n & \text{is Borel} \\ \beta : \mathbb{R}_+ \times \mathbb{R}^n \longrightarrow \mathbb{R}^n \times \mathbb{R}^n & \text{is Borel, } \beta(s,x) \text{ is symmetric nonnegative} \\ K_t(\omega, x, dy) & \text{is a Borel transition kernel from } \Omega \times \mathbb{R}^n \\ & \text{into } \mathbb{R}^n, \end{cases}$$

with $K_t(\omega, x, \{0\}) = 0$.

Next, we relate the above with stochastic differential equations, partially following (Jacod and Shiryaev, 1987).

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$ be a stochastic basis endowed with:

(i) $W = (W^i)_{i\leq m}$, an m-dimensional standard Wiener process (i.e., each W^i is a standard Wiener process, and the W^i's are independent);

(ii) p_i are Poisson random measures on $\mathbb{R}_+ \times U$ with intensity measure $dt \cdot m_i(du)$, $i = 1, 2$; here, $(U, \mathscr{U})$ is an arbitrary Blackwell space (one may take $U = \mathbb{R}^d$ for practical applications), and m_i, $i = 1, 2$, is a positive σ-finite measure on $U, \mathscr{U}$; We denote the compensated Poisson random measure by $q_i(dt, du) = p_i(dt, du) - dt \cdot m_i(du)$, $i = 1, 2$.

Let us also be given the coefficients:

$$\begin{cases} a = (a^i)_{i\leq n}, & \text{a Borel function: } \mathbb{R}_+ \times \mathbb{R}^n \longrightarrow \mathbb{R}^n \\ b = (b^{ij})_{i\leq n, j\leq m}, & \text{a Borel function: } \mathbb{R}_+ \times \mathbb{R}^n \longrightarrow \mathbb{R}^n \times \mathbb{R}^m \\ f_1 = (f_1^i)_{i\leq n} & \text{a Borel function: } \mathbb{R}_+ \times \mathbb{R}^n \times U \longrightarrow \mathbb{R}^n, \\ f_2 = (f_2^i)_{i\leq n} & \text{a Borel function: } \mathbb{R}_+ \times \mathbb{R}^n \times U \longrightarrow \mathbb{R}^n. \end{cases} \tag{3.2.15}$$

Let the initial variable be an $\mathscr{F}_0$-measurable $\mathbb{R}^n$-valued random variable X_0. The stochastic differential equation is as follows:

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t + \int_U f_1(t, X_{t-}, u)q_1(dt, du) + \int_U f_2(t, X_{t-}, u)p_2(dt, du), \quad (3.2.16)$$

Define two stochastic sets:

$$D_1 = \{(\omega, t) : p_1(\omega; \{t\} \times U) = 1\},$$
$$D_2 = \{(\omega, t) : p_2(\omega; \{t\} \times U) = 1\}.$$

If at least one of the Poisson random measures, p_1 or p_2, has a "jump" at point (t, u), then

$$\Delta X_t(\omega) = I_{D_1}(\omega, t) \cdot f_1(t, X_{t-}(\omega), u) + I_{D_2}(\omega, t) \cdot f_2(t, X_{t-}(\omega), u).$$

Next, let us assume that the following integrals make sense.

$$\int_0^t |a(s, X_s)|ds < \infty, \ P\text{-a.s.} \quad (3.2.17)$$

$$\int_0^t \int_U |f_1(s, X_{s-}, u)|^2 ds\, m_1(du) < \infty, \ P\text{-a.s.}, \quad (3.2.18)$$

$$\int_0^t \int_U |f_2(s, X_{s-}, u)|p_2(ds, du) < \infty, \ P\text{-a.s.}, \quad (3.2.19)$$

$$\int_0^t |b^{ij}(s, X_s)|^2 ds < \infty, \ P\text{-a.s. for any } i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (3.2.20)$$

for every $t \in \mathbb{R}_+$. By a solution to the SDE (3.2.16) we mean a càdlàg $\mathscr{F}_t$-adapted process $\{X_t\}$ such that the following equation is satisfied with probability one for every $t \in \mathbb{R}_+$

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s + \int_0^t \int_U f_1(s, X_{s-}, u)q_1(ds, du) + \int_0^t \int_U f_2(s, X_{s-}, u)p_2(ds, du). \quad (3.2.21)$$

If such process $\{X_t\}$ exists and conditions (3.2.17)-(3.2.20) are satisfied then it is a semimartingale with the characteristics, associated with truncation function

$h = x \cdot I_{\{x\,:\,|x|<1\}}(x)$, given by (3.2.14), where

$$\alpha(t, X_t(\omega)) = \Big[a(t, X_t(\omega)) - \int_{|f_1|\geq 1} f_1(t, X_{t-}(\omega), u) m_1(du) + \int_{|f_2|<1} f_2(t, X_{t-}(\omega), u) m_2(du)\Big],$$

$$\beta(t, X_t(\omega)) = b(t, X_t(\omega)) b^T(t, X_t(\omega)),$$

$$K_t(\omega, X_t(\omega), A) = I_{D_1}(\omega, t) \cdot \int_U I_{A\setminus\{0\}}\big(f_1(t, X_{t-}(\omega), u)\big) m_1(du) + I_{D_2}(\omega, t) \cdot \int_U I_{A\setminus\{0\}}\big(f_2(t, X_{t-}(\omega), u)\big) m_2(du).$$

3.3 Semimartingale strong solution of SDE

3.3.1 Existence and uniqueness concepts

There are two important notions of the sense in which a solution to stochastic differential equation can be said to *exist* and also two senses in which *uniqueness* is said to hold.

Definition 3.3.1. Strong Existence. We say that strong existence holds if given a probability space $(\Omega, \mathscr{F}, P)$, a filtration $\mathscr{F}_t$, an $\mathscr{F}_t$-Wiener process W, two $\mathscr{F}_t$-Poisson random measures p_1, p_2, and an $\mathscr{F}_0$-measurable initial condition X_0, then an $\mathscr{F}_t$-adapted process $\{X_t\}$ exists satisfying (3.2.21) for all $t \geq 0$.

Definition 3.3.2. Weak Existence. We say that weak existence holds if given any probability measure η on $\mathbb{R}^n$ there exists a probability space $(\Omega, \mathscr{F}, P)$, a filtration $\mathscr{F}_t$, an $\mathscr{F}_t$-Wiener process W, two $\mathscr{F}_t$-Poisson random measures p_1, p_2, and an $\mathscr{F}_t$-adapted process $\{X_t\}$ satisfying (3.2.21) for all $t \geq 0$ as well as $P(X_0 \in B) = \eta(B)$.

Strong existence of a solution requires that the probability space, filtration, and driving terms (W, p_1, p_2) be given first and that the solution $\{X_t\}$ then be found for the given data. Weak sense existence allows these objects to be constructed together with the process $\{X_t\}$. Clearly, strong existence implies weak existence.

Definition 3.3.3. Strong Uniqueness. Suppose that a fixed probability space $(\Omega, \mathscr{F}, P)$, a filtration $(\mathscr{F}_t)_{t\geq 0}$, an $\mathscr{F}_t$-Wiener process W, and two $\mathscr{F}_t$-Poisson random measures p_1 and p_2 are given. Let $\{X_t\}$ and $\{X'_t\}$ be two solutions of (3.2.16) for the given driving terms (W, p_1, p_2). We say that strong uniqueness holds if

$$P(X_0 = X'_0) = 1 \Longrightarrow P(X_t = X'_t \text{ for all } t \geq 0) = 1, \tag{3.3.1}$$

i.e. $\{X_t\}$ and $\{X'_t\}$ are indistinguishable.

Remark 3.3.4. Since solutions of (3.2.16) are càdlàg processes the requirement (3.3.1) can be relaxed to:

$$P(X_0 = X_0') = 1 \Longrightarrow P(X_t = X_t') = 1, \text{ for every } t \geq 0. \tag{3.3.2}$$

Definition 3.3.5. Weak Uniqueness. Suppose we are given weak sense solutions

$$\{(\Omega_i, \mathscr{F}_i, P_i), (\mathscr{F}_{i,t})_{t\geq 0}, \{X_{i,t}\}\},\ i = 1, 2,$$

to (3.2.16). We say that weak uniqueness holds if equality of the distributions induced on $\mathbb{R}^n$ by $X_{i,0}$ under P_i, $i = 1, 2$, implies the equality of the distributions induced on $\mathbb{D}(\mathbb{R}^n)$ by $\{X_{i,t}\}$ under P_i, $i = 1, 2$.

Strong uniqueness is also referred to as *pathwise uniqueness*, whereas weak uniqueness is often called *uniqueness in (the sense of probability) law*. Strong uniqueness implies weak uniqueness.

3.3.2 Strong Uniqueness

In what follows we will state and prove strong existence and strong uniqueness theorems for SDE (3.2.16), following (Gihman and Skorohod, 1982, pp.223-245).

We assume that Wiener process W and Poisson random measures p_1 and p_2 are mutually independent. Suppose $\{W_t\}$, p_1 and p_2 are adapted to the given filtration $(\mathscr{F}_t)_{t\geq 0}$. If τ is a stopping time relative to $\mathscr{F}_t$ and X_τ is an $\mathscr{F}_\tau$ measurable random variable, then we will be looking for an $\{\mathscr{F}_t\}$-adapted process $\{X_t\}$, defined for $t > \tau$, for which the following equation holds with probability 1

$$X_t = X_\tau + \int_\tau^t a(s, X_s)ds + \int_\tau^t b(s, X_s)dW_s + \int_\tau^t \int_U f_1(s, X_{s-}, u)q_1(ds, du) + \int_\tau^t \int_U f_2(s, X_{s-}, u)p_2(ds, du). \tag{3.3.3}$$

If equality (3.3.3) holds for all $t \in (\tau, \zeta)$, with ζ another stopping time, $\zeta > \tau$, then we will say that $\{X_t\}$ is the solution of SDE (3.2.16) on interval (τ, ζ), if started at X_τ.

Theorem 3.3.6. *A solution of Equation (3.2.16) for any given X_0 is strongly unique if the coefficients of Equation (3.2.16) satisfy the following conditions:*

(i) for each $r > 0$ there exist a constant l_r, for which

$$|a(s, x) - a(s, y)|^2 + |b(s, x) - b(s, y)|^2 + \int_U |f_1(s, x, u) - f_1(s, y, u)|^2 m_1(du) \leq l_r|x - y|^2,$$

for all $|x| \leq r$, $|y| \leq r$, $s \leq r$.

(ii) $\int_0^t \int_U |f_2(s, X_{s-}, u)| p_2(ds, du) < \infty$, *P-a.s.,*

(iii) $m_2(S_u) < \infty$, *where* S_u *is the projection on space* U *of the support of function* $f_2(\cdot, \cdot, \cdot)$.

Proof. We fix some admissible filtration $\{\mathscr{F}_t\}$ and consider only $\mathscr{F}_t$-measurable solutions. Suppose $\tau_1 < \tau_2 < \dots$ are all jump moments of the Poisson process $p_2(S_u, [0,t])$. Since it is a homogeneous process with parameter $m_2(S_u) < \infty$, there will be only finite number of jumps on every finite interval. Let $\tau_0 = 0$. Note, that it suffices to establish the uniqueness of a solution of Equation (3.2.16) on interval $[\tau_k, \tau_{k+1}]$, with assumption that X_{τ_k} is given. Then we establish by induction that a solution of (3.2.16) is unique on any interval $[0, \tau_k]$, and $\bigcup [0, \tau_k] = \mathbb{R}_+$. Suppose $\{X_t\}$ and $\{\overline{X}_t\}$ are two solutions of (3.2.16) on $[\tau_k, \tau_{k+1})$, for which $X_{\tau_k} = \overline{X}_{\tau_k}$. For $\tau_k \le t < \tau_{k+1}$

$$X_t = X_{\tau_k} + \int_{\tau_k}^t a(s, X_s) ds + \int_{\tau_k}^t b(s, X_s) dW_s + \int_{\tau_k}^t \int_U f_1(s, X_{s-}, u) q_1(ds, du), \tag{3.3.4}$$

since the last integral with respect to measure p_2 in (3.3.3) will be equal to zero. Similar equality holds for solution $\overline{X}_t$. Let $\zeta_r = \inf\{t > \tau_k, |X_t| + |\overline{X}_t| \ge r\} \wedge r$. Next, we provide the estimates of the right-hand side terms of the following expression:

$$\begin{aligned} X_{t\wedge\zeta_r} - \overline{X}_{t\wedge\zeta_r} = &\int_{\tau_k}^{t\wedge\zeta_r} [a(s, X_s) - a(s, \overline{X}_s)] ds \\ &+ \int_{\tau_k}^{t\wedge\zeta_r} [b(s, X_s) - b(s, \overline{X}_s)] dW_s \\ &+ \int_{\tau_k}^{t\wedge\zeta_r} [f_1(s, X_{s-}, u) - f_1(s, \overline{X}_{s-}, u)] q_1(ds, du). \end{aligned} \tag{3.3.5}$$

$$\mathbb{E}\Big(\Big| \int_{\tau_k}^{t\wedge\zeta_r} [a(s, X_s) - a(s, \overline{X}_s)] ds \Big|^2 | \mathscr{F}_{\tau_k} \Big) \le l_r (t - \tau_k) \mathbb{E}\Big(\int_{\tau_k}^{t\wedge\zeta_r} |X_s - \overline{X}_s|^2 ds | \mathscr{F}_{\tau_k} \Big),$$

$$\mathbb{E}\Big(\Big| \int_{\tau_k}^{t\wedge\zeta_r} [b(s, X_s) - b(s, \overline{X}_s)] dW_s \Big|^2 | \mathscr{F}_{\tau_k} \Big) \le l_r n m \mathbb{E}\Big(\int_{\tau_k}^{t\wedge\zeta_r} |X_s - \overline{X}_s|^2 ds | \mathscr{F}_{\tau_k} \Big),$$

where n and m are the dimensionalities of X_s and W_s correspondingly,

$$\mathbb{E}\Big(\Big|\int_{\tau_k}^{t\wedge\zeta_r}\int_U (f_1(s,X_{s-},u)-f_1(s,\overline{X}_{s-},u))q_1(ds,du)\Big|^2|\mathscr{F}_{\tau_k}\Big) \\ \leq l_r\mathbb{E}\Big(\int_{\tau_k}^{t\wedge\zeta_r}|X_s-\overline{X}_s|^2ds|\mathscr{F}_{\tau_k}\Big),$$

(we have made use of the properties of stochastic integrals and conditions of the theorem), then for some L (it is a $\mathscr{F}_{\tau_k}$-measurable quantity) we have the following estimate for expression (3.3.5):

$$\mathbb{E}(|X_{t\wedge\zeta_r}-\overline{X}_{t\wedge\zeta_r}|^2|\mathscr{F}_{\tau_k})\leq L\mathbb{E}\Big(\int_{\tau_k}^{t\wedge\zeta_r}|X_s-\overline{X}_s|^2ds|\mathscr{F}_{\tau_k}\Big).$$

But then the following holds

$$\mathbb{E}(|X_t-\overline{X}_t|^2I_{\{\zeta_r>t\}}|\mathscr{F}_{\tau_k})\leq L\int_{\tau_k}^{t}\mathbb{E}(|X_s-\overline{X}_s|^2I_{\{\zeta_r>s\}}|\mathscr{F}_{\tau_k})ds. \tag{3.3.6}$$

Hence, because of Gronwall's lemma :

$$\mathbb{E}|X_s-\overline{X}_s|^2I_{\{\zeta_r>t\}}=0.$$

Since $I_{\{\zeta_r>t\}}\longrightarrow 1$ as $r\longrightarrow\infty$, $X_t=\overline{X}_t$ for $\tau_k\leq t<\tau_{k+1}$. It remains to show that $X_{\tau_{k+1}}=\overline{X}_{\tau_{k+1}}$. Suppose X_t^* is a solution of (3.3.4). We have already shown that it is unique. Now note, that solution of Equation (3.3.3) at point τ_{k+1} can be expressed in terms of X_t^* on $[\tau_k,\tau_{k+1})$ in the following way:

$$X_{\tau_{k+1}}=X^*_{\tau_{k+1}-}+f_2(\tau_{k+1},X^*_{\tau_{k+1}-},\hat{u}_{k+1}),$$

where $\hat{u}_{k+1}$ - such a point from U, that $p_2(\{\hat{u}_{k+1}\}\times\{\tau_{k+1}\})=1$. From the coincidence of $X_{\tau_{k+1}-}$ and $\overline{X}_{\tau_{k+1}-}$ follows the coincidence $X_{\tau_{k+1}}=\overline{X}_{\tau_{k+1}}$. □

It is easy to see from the proof of Theorem 3.3.6 that not only two solutions of one equation coincide, but also solutions of two different equations with equal initial conditions coincide as long as their coefficients coincide. We formulate this statement precisely, known as the theorem of local uniqueness.

Theorem 3.3.7. *Suppose $\{X_t\}$ is a solution of Equation (3.2.21), and $\{\tilde{X}_t\}$ is a solution of equation*

$$\tilde{X}_t=\tilde{X}_0+\int_0^t\tilde{a}(s,\tilde{X}_s)ds+\int_0^t\tilde{b}(s,\tilde{X}_s)dW_s \\ +\int_0^t\int_U\tilde{f}_1(s,\tilde{X}_{s-},u)q_1(ds,du)+\int_0^t\int_U\tilde{f}_2(s,\tilde{X}_{s-},u)p_2(ds,du).$$

If the conditions of Theorem 3.3.6 are satisfied and $X_0=\tilde{X}_0$, $a(s,x)=\tilde{a}(s,x)$, $b(s,x)=\tilde{b}(s,x)$, $f_k(s,x,u)=\tilde{f}_k(s,x,u)$ given $|x|\leq N$, then $X_s=\tilde{X}_s$ for $s\leq\tau$, where $\tau=\inf\{s:|X_s|\geq N\}$.

3.3.3 Strong Existence.

First, we state the classical existence results for the following equation (Gihman and Skorohod, 1982):

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s + \int_0^t \int_U f_1(s, X_{s-}, u)q_1(ds, du). \tag{3.3.7}$$

Theorem 3.3.8. *Assume that the coefficients of Equation (3.3.7) satisfy the following conditions:*

(i) $a(s,0)$, $b(s,0)$, $\int |f_1(s,0,u)|^2 m_1(du)$ *are locally bounded with respect to* s,

(ii) there exists increasing function $l(s)$ *such that*

$$|a(s,x) - a(s,y)|^2 + |b(s,x) - b(s,y)|^2 + \int_U |f_1(s,x,u) - f_1(s,y,u)|^2 m_1(du) \le l(s)|x-y|^2.$$

Let us denote by $\mathscr{F}_t$ *the* σ*-algebra generated by* X_0, $q_1(ds,du)$, W_s *with* $s \le t$. *If* X_0 *is independent of* W_s, $q_1(ds,du)$ *and* $\mathbb{E}|X_0|^2 < \infty$, *then equation (3.3.7) has* $\mathscr{F}_t$*-measurable solution, moreover* $\mathbb{E}|X_s|^2 < \infty$.

Theorem 3.3.9. *Assume that for the coefficients of Equation (3.3.7) the following conditions hold:*

$$|a(t,x)|^2 + |b(t,x)|^2 + \int_U |f_1(t,x,u)|^2 m_1(du) \le l(1+|x|^2),$$

and for any $r > 0$ *one can specify constant* l_r *such that*

$$|a(s,x) - a(s,y)|^2 + |b(s,x) - b(s,y)|^2 + \int_U |f_1(s,x,u) - f_1(s,y,u)|^2 m_1(du) \le l_r|x-y|^2$$

for $s \le r$, $|x| \le r$, $|y| \le r$. *If* X_0 *is independent of* $\{W_s, q_1(ds,du)\}$, *and* σ*-algebras* $\mathscr{F}_t$ *are constructed as in Theorem 3.3.8, then there exists an* $\mathscr{F}_t$*-measurable solution of (3.3.7) for every* $t \in \mathbb{R}_+$.

Remark 3.3.10. Suppose $\{\hat{\mathscr{F}}_t\}$ is some admissible filtration, τ is a stopping time relative to this filtration. Let us consider the SDE for $t > \tau$:

$$X_t = X_\tau + \int_\tau^t a(s, X_s)ds + \int_\tau^t b(s, X_s)dW_s + \int_\tau^t \int_U f_1(s, X_{s-}, u)q_1(ds, du). \tag{3.3.8}$$

Under conditions of Theorem 3.3.9, Equation (3.3.8) has $\hat{\mathscr{F}}_t$-measurable solution, no matter what the $\hat{\mathscr{F}}_\tau$-measurable variable X_τ is. To prove this, it suffices to consider the process $\hat{X}_t$ which is a solution of the following equation

$$\hat{X}_t = \hat{X}_0 + \int_0^t a(s+\tau, \hat{X}_s)ds + \int_0^t b(s+\tau, \hat{X}_s)d\hat{W}_s + \int_0^t \int_U f_1(s+\tau, \hat{X}_{s-}, u)\hat{q}_1(ds, du), \quad (3.3.9)$$

where

$$\hat{W}_s = W(s+\tau) - W_\tau; \quad \hat{q}_1([s_1, s_2] \times du) = q_1([s_1+\tau, s_2+\tau] \times du). \quad (3.3.10)$$

Obviously, $\hat{W}$ and $\hat{q}_1$ possess the same properties as W and q_1, and are independent of $\mathscr{F}_\tau$. Thus, for Equation (3.3.9), all derivations which were verified for Equation (3.3.7), hold as well, if expectations and conditional expectations with given X_0 are substituted by conditional expectation with respect to σ-algebra $\hat{\mathscr{F}}_\tau$. Obviously, then $X_t = \hat{X}_{t-\tau}$ will be the solution of Equation (3.3.8).

Now we prove the existence theorem for general SDE (3.2.16).

Theorem 3.3.11. *Assume that for Equation (3.2.16) the following conditions are satisfied:*

(i) The coefficients a, b, f_1 satisfy the conditions of Theorem 3.3.9.

(ii) X_0 is independent of $\{W_s, q_1(ds, du), p_2(ds, du)\}$.

(iii) Conditions (ii) and (iii) of Theorem 3.3.6 are satisfied.

Let $\mathscr{F}_t$ denote the σ-algebra generated by X_0 and $\{W_s,\ q_1([0,s], du),\ p_2([0,s], du),\ s \le t\}$. Then there exists an $\mathscr{F}_t$-measurable solution of equation (3.2.16).

Proof. Let $\tau_1 < \tau_2 < \cdots < \tau_n < \ldots$ denote all jump moments of the process $p_2(S_u, [0,t])$, (in other words these are the "growth" moments of the last integral term in (3.2.21)). Since $p_2(S_u, [0,t])$ is a homogeneous Poisson process with parameter $m_2(S_u) < \infty$ (condition *(iii)*), then there will be only finite number of jumps on every finite interval. It suffices to construct the solution of Equation (3.2.16) on each interval $[0, \tau_1)$, $[\tau_1,\ \tau_2)$, $\ldots$, $[\tau_n, \tau_{n+1})$, $\ldots$ Since $\int_{\tau_n}^t \int f_2(s, X_{s-}, u)p_2(ds, du) = 0$ when $t \in [\tau_n, \tau_{n+1})$, then on each of the specified intervals equation (3.2.16) turns into equation of type (3.3.8), where τ equals $0, \tau_1, \tau_2, \ldots$ and so on. As it was pointed out in remark 3.3.10, there exists a solution of this equation if X_τ is $\mathscr{F}_\tau$-measurable. Let us prove that this is indeed the case for the sequence of stopping times $\tau_1, \tau_2, \ldots$. If $\tau = 0$, then X_0 is $\mathscr{F}_0$-measurable by definition of σ-algebra $\mathscr{F}_t$. Suppose, that X_{τ_n} is $\mathscr{F}_{\tau_n}$-measurable.

We will show that then $X_{\tau_{n+1}}$ will be $\mathscr{F}_{\tau_{n+1}}$-measurable. Let X_t^n be the solution of the following equation

$$X_t^n = X_{\tau_n}^n + \int_{\tau_n}^t a(s, X_s^n)ds + \int_{\tau_n}^t b(s, X_s^n)dW_s + \int_{\tau_n}^t \int_U f_1(s, X_{s-}^n, u)q_1(ds, du)$$

for $t \geq \tau_n$. In consequence of Remark 3.3.10 such solution exists. Set $X_t = X_t^n$ for $t < \tau_{n+1}$. Let u_{n+1} be such a point in U that $p_2(\{\tau_{n+1}\} \times \{u_{n+1}\}) = 1$. Hence u_{n+1} is $\mathscr{F}_{\tau_{n+1}}$-measurable. Now let us define $X_{\tau_{n+1}}$ by the equality

$$X_{\tau_{n+1}} = X_{\tau_{n+1}-}^n + f_2(\tau_{n+1}, X_{\tau_{n+1}-}^n, u_{n+1}). \tag{3.3.11}$$

Both summands in the right hand side of (3.3.11) are $\mathscr{F}_{\tau_{n+1}}$-measurable, i.e. $X_{\tau_{n+1}}$ is $\mathscr{F}_{\tau_{n+1}}$-measurable. Thus, we can successively construct $\mathscr{F}_t$-measurable process X_t. In order to make certain that it is indeed a solution of (3.2.16), it suffices to see that

$$f_2(\tau_{n+1}, X_{\tau_{n+1}-}^n, u_{n+1}) = \int_{\tau_n}^{\tau_{n+1}} \int_U f_2(t, X_{t-}, u)p_2(dt, du).$$

□

Remark 3.3.12. The solution, whose existence was established in Theorem 3.3.11, is unique. Indeed, by Theorem 3.3.6 we have that for any enlargement of the initial probability space, any admissible filtration of σ-algebras $\tilde{\mathscr{F}}_t$, and any $\mathscr{F}_0$-measurable initial variable X_0, $\tilde{\mathscr{F}}_t$-measurable solution of Equation (3.2.16) is unique. Since $\mathscr{F}_t \subset \tilde{\mathscr{F}}_t$, the solution X_t constructed in Theorem 3.3.11 will be also $\tilde{\mathscr{F}}_t$-measurable, and therefore, there will be no other $\tilde{\mathscr{F}}_t$-measurable solutions of Equation (3.2.16).

Remark 3.3.13. The solution constructed in Theorem 3.3.11 is fully determined by the initial condition, Wiener process W and Poisson random measures p_1 and p_2, i.e. it is a "strong" solution (solution-process). Thus, Theorem 3.3.11 states that there exists a strong solution of Equation (3.2.16) (strong existence), and from the remark 3.3.12 it follows that under conditions of Theorem 3.3.11 any solution of (3.2.16) is unique (strong uniqueness).

Remark 3.3.14. Under the conditions of Theorem 3.3.11 the solution of SDE (3.2.16) admits the decomposition (3.2.1) with

$$A_t = \int_0^t a(s, X_s)ds + \int_0^t \int_U f_2(s, X_{s-}, u)p_2(ds, du) \in \mathscr{V}^n,$$
$$M_t = \int_0^t b(s, X_s)dW_s + \int_0^t \int_U f_1(s, X_{s-}, u)q_1(ds, du) \in \mathscr{M}_{loc}^n,$$

hence it is a semimartingale.

3.4 Stochastic Hybrid Processes as Solutions of SDE

3.4.1 SDE on hybrid state space

In this section we construct a switching jump diffusion $\{X_t, \theta_t\}$ taking values in $\mathbb{R}^n \times \mathbb{M}$, where $\mathbb{M} = \{e_1, e_2, \ldots, e_N\}$ is a finite set. We assume that for each $i = 1, \ldots, N$, e_i is the i-th unit vector, $e_i \in \mathbb{R}^N$. Let $\{X_t, \theta_t\}$ be an $\mathbb{R}^n \times \mathbb{M}$-valued process given by the following stochastic differential equation of Ito-Skorohod type.

$$dX_t = a(X_t, \theta_t)dt + b(X_t, \theta_t)dW_t + \int_{\mathbb{R}^d} g_1(X_{t-}, \theta_{t-}, u)q_1(dt, du) \tag{3.4.1}$$
$$+ \int_{\mathbb{R}^d} g_2(X_{t-}, \theta_{t-}, u)p_2(dt, du),$$
$$d\theta_t = \int_{\mathbb{R}^d} c(X_{t-}, \theta_{t-}, u)p_2(dt, du). \tag{3.4.2}$$

Here:

(i) for $t = 0$, X_0 is a prescribed $\mathbb{R}^n$-valued random variable.

(ii) for $t = 0$, θ_0 is a prescribed $\mathbb{M}$-valued random variable.

(iii) W is an m-dimensional standard Wiener process.

(iv) $q_1(dt, du)$ is a martingale random measure associated to a Poisson random measure p_1 with intensity $dt \times m_1(du)$.

(v) $p_2(dt, du)$ is a Poisson random measure with intensity $dt \times m_2(du) = dt \times du_1 \times \bar{\mu}(d\underline{u})$, where $\bar{\mu}$ is a probability measure on $\mathbb{R}^{d-1}$, $u_1 \in \mathbb{R}$, $\underline{u}$ refers to all components of $u \in \mathbb{R}^d$ except the first one.

The coefficients are assumed to be measurable.

$$\begin{aligned} a &: \mathbb{R}^n \times \mathbb{M} \to \mathbb{R}^n \\ b &: \mathbb{R}^n \times \mathbb{M} \to \mathbb{R}^{n \times m} \\ g_1 &: \mathbb{R}^n \times \mathbb{M} \times \mathbb{R}^d \to \mathbb{R}^n \\ g_2 &: \mathbb{R}^n \times \mathbb{M} \times \mathbb{R}^d \to \mathbb{R}^n \\ c &: \mathbb{R}^n \times \mathbb{M} \times \mathbb{R}^d \to \mathbb{R}^N. \end{aligned}$$

Function $c(\cdot, \cdot, \cdot)$ is defined by

$$c(x, e_i, u) = \begin{cases} e_j - e_i & \text{if } u_1 \in (\Sigma_{j-1}(x, e_i), \Sigma_j(x, e_i)], \\ 0 & \text{otherwise .} \end{cases} \tag{3.4.3}$$

Function $g_2(\cdot,\cdot,\cdot)$ is defined by

$$g_2(x,e_i,u)=\begin{cases}\phi(x,e_i,e_j,\underline{u}) & \text{if } u_1\in(\Sigma_{j-1}(x,e_i),\Sigma_j(x,e_i)],\\ 0 & \text{otherwise .}\end{cases} \tag{3.4.4}$$

The measurable mappings $\Sigma_k:\mathbb{R}^n\times\mathbb{M}\to\mathbb{R}_+$, $k=1,2,\dots,N$ determine the actual switching and jump rates of $\{\theta_t\}$ and $\{X_t\}$ components:

$$\Sigma_k(x,e_i)=\begin{cases}\sum_{j=1}^k\lambda(x,e_i,e_j) & k>0,\\ 0 & k=0,\end{cases} \tag{3.4.5}$$

here $\lambda:\mathbb{R}^n\times\mathbb{M}\times\mathbb{M}\to\mathbb{R}_+$ is a measurable mapping. Measurable function $\phi:\mathbb{R}^n\times\mathbb{M}\times\mathbb{M}\times\mathbb{R}^{d-1}\to\mathbb{R}^n$ determines the size of jumps of $\{X_t\}$. Let U_θ denote the projection of the support of function $\phi(\cdot,\cdot,\cdot,\cdot)$ on space $\underline{U}=\mathbb{R}^{d-1}$. The jump size of X_t and the new value of θ_t at the jump times generated by Poisson random measure p_2 are determined by the functions (3.4.3) and (3.4.4) correspondingly. There are three different situations possible:

(i) Simultaneous jump of X_t and θ_t

$$\begin{cases}c(\cdot,\cdot,u)\neq 0 & \text{if } u_1\in(\Sigma_{j-1}(x,e_i),\Sigma_j(x,e_i)],\ i,j=1,\dots,N \text{ and } j\neq i,\\ g_2(\cdot,\cdot,u)\neq 0 & \text{if } u_1\in(\Sigma_{j-1}(x,e_i),\Sigma_j(x,e_i)],\ i,j=1,\dots,N \text{ and } \underline{u}\in U_\theta.\end{cases}$$

(ii) Switch of θ_t only

$$\begin{cases}c(\cdot,\cdot,u)\neq 0 & \text{if } u_1\in(\Sigma_{j-1}(x,e_i),\Sigma_j(x,e_i)],\ i,j=1,\dots,N \text{ and } j\neq i,\\ g_2(\cdot,\cdot,u)=0 & \text{if } u_1\in(\Sigma_{j-1}(x,e_i),\Sigma_j(x,e_i)],\ i,j=1,\dots,N \text{ and } \underline{u}\notin U_\theta.\end{cases}$$

(iii) Jump of X_t only

$$\begin{cases}c(\cdot,\cdot,u)=0 & \text{if } u_1\in(\Sigma_{j-1}(x,e_j),\Sigma_j(x,e_j)],\ j=1,\dots,N,\\ g_2(\cdot,\cdot,u)\neq 0 & \text{if } u_1\in(\Sigma_{j-1}(x,e_j),\Sigma_j(x,e_j)],\ j=1,\dots,N, \text{ and } \underline{u}\in U_\theta.\end{cases}$$

We make the following assumptions on the coefficients of SDE (3.4.1)-(3.4.2).

(A1) There exists a constant l such that for all $i=1,2,\dots,N$

$$|a(x,e_i)|^2+|b(x,e_i)|^2+\int_{\mathbb{R}^d}|g_1(x,e_i,u)|^2m_1(du)\le l(1+|x|^2).$$

(A2) For any $r>0$ one can specify constant l_r such that for all $i=1,2,\dots,N$

$$|a(x,e_i)-a(y,e_i)|^2+|b(x,e_i)-b(y,e_i)|^2 + \int_{\mathbb{R}^d}|g_1(x,e_i,u)-g_1(y,e_i,u)|^2m_1(du)\le l_r|x-y|^2$$

for $|x|\le r$, $|y|\le r$.

(A3) Function c satisfies (3.4.5), (3.4.3) and for $i,j=1,2,\dots,N$, $\lambda(\cdot,e_i,e_j)$ are bounded and measurable, $\lambda(\cdot,e_i,e_j)\geq 0$.

(A4) Function g_2 satisfies (3.4.5), (3.4.4) and for all $t>0$, $i,j=1,\dots,N$

$$\int_0^t\int_{\mathbb{R}^d}|\phi(x,e_i,e_j,\underline{u})|p_2(ds,du)<\infty,\quad P\text{-a.s.}$$

3.4.2 Strong existence and uniqueness

Theorem 3.4.1. *Assume (A1)-(A4). Let p_1,p_2,W,X_0 and θ_0 be independent. Then SDE (3.4.1)-(3.4.2) has a unique strong solution which is a semimartingale.*

Proof. The switching jump diffusion $\{X_t,\theta_t\}$ governed by Equations (3.4.1)-(3.4.2) can be seen as the $\mathbb{R}^{n+N}$-valued jump diffusion $\{\xi_t\}\triangleq\{(X_t,\theta_t)^T\}$ governed by the stochastic differential equation

$$d\xi_t=\tilde{a}(\xi_t)dt+\tilde{b}(\xi_t)dW_t+\int_{\mathbb{R}^d}\tilde{f}_1(\xi_{t-},u)q_1(dt,du)+\int_{\mathbb{R}^d}\tilde{f}_2(\xi_{t-},u)p_2(dt,du) \tag{3.4.6}$$

with the following coefficients:

$$\begin{aligned}
&\tilde{a}:\mathbb{R}^{n+N}\to\mathbb{R}^{n+N} && \tilde{a}(\cdot)\triangleq[a(\cdot),O^N]^T\\
&\tilde{b}:\mathbb{R}^{n+N}\to\mathbb{R}^{(n+N)\times m} && \tilde{b}(\cdot)\triangleq[b(\cdot),O^{N\times m}]^T\\
&\tilde{f}_1:\mathbb{R}^{n+N}\times\mathbb{R}^d\to\mathbb{R}^{n+N} && \tilde{f}_1(\cdot,\cdot)\triangleq[g_1(\cdot,\cdot),O^N]^T\\
&\tilde{f}_2:\mathbb{R}^{n+N}\times\mathbb{R}^d\to\mathbb{R}^{n+N} && \tilde{f}_2(\cdot,\cdot)\triangleq[g_2(\cdot,\cdot),c(\cdot,\cdot)]^T
\end{aligned}$$

where by O^k and $O^{k\times s}$ we denote the k-dimensional zero vector and $k\times s$-dimensional zero matrix correspondingly.

Next we show that conditions (A1)-(A4) together with (3.4.5), (3.4.3), (3.4.4) imply the conditions of Theorems 3.3.6 and 3.3.11 thus the Equation (3.4.6) has an a.s. unique strong solution which implies that SDE (3.4.1)-(3.4.2) has an a.s. unique strong solution.

Let us verify all conditions.

Growth condition: by (A1) for every $\xi=(x,e_i)^T\in\mathbb{R}^{n+N}$ $i=1,\dots,N$ we have

$$\begin{aligned}
&|\tilde{a}(\xi)|^2+|\tilde{b}(\xi)|^2+\int_{\mathbb{R}^d}|\tilde{f}_1(\xi,u)|^2m_1(du)\\
&\quad=|\tilde{a}(x,e_i)|^2+|\tilde{b}(x,e_i)|^2+\int_{\mathbb{R}^d}|\tilde{f}_1(x,e_i,u)|^2m_1(du)\\
&\quad=|a(x,e_i)|^2+|b(x,e_i)|^2+\int_{\mathbb{R}^d}|g_1(x,e_i,u)|^2m_1(du)\\
&\quad\leq l(1+|x|^2)\leq l(1+|x|^2+|e_i|^2)=l(1+|\xi|^2).
\end{aligned}$$

Lipschitz condition: From (A1) and (A2) it follows that for any $r > 0$ one can specify a constant L_r such that for all $\xi = (x, e_i)^T \in \mathbb{R}^{n+N}$, $\zeta = (y, e_j)^T \in \mathbb{R}^{n+N}$ $i, j = 1, \dots, N$, and for $|x| < r$, $|y| < r$, i.e. $|\xi| \le \sqrt{r^2+1}$, $|\zeta| \le \sqrt{r^2+1}$, we have

$$
\begin{aligned}
&|\tilde{a}(\xi) - \tilde{a}(\zeta)|^2 + |\tilde{b}(\xi) - \tilde{b}(\zeta)|^2 + \int_{\mathbb{R}^d} |\tilde{f}_1(\xi, u) - \tilde{f}_1(\zeta, u)|^2 m_1(du) \\
&\qquad = |a(x, e_i) - a(y, e_j)|^2 + |b(x, e_i) - b(y, e_j)|^2 \\
&\qquad\quad + \int_{\mathbb{R}^d} |g_1(x, e_i, u) - g_1(y, e_j, u)|^2 m_1(du) \\
&\qquad \le 2\big(|a(x, e_i) - a(y, e_i)|^2 + |b(x, e_i) - b(y, e_i)|^2 \\
&\qquad\quad + \int_{\mathbb{R}^d} |g_1(x, e_i, u) - g_1(y, e_i, u)|^2 m_1(du) \\
&\qquad\quad + |a(y, e_i) - a(y, e_j)|^2 + |b(y, e_i) - b(y, e_j)|^2 \\
&\qquad\quad + \int_{\mathbb{R}^d} |g_1(y, e_i, u) - g_1(y, e_j, u)|^2 m_1(du)\big) \\
&\quad \le 2\Big(l_r|x-y|^2 + 4\big(|a(y, e_i)|^2 + |b(y, e_i)|^2 + \int_{\mathbb{R}^d} |g_1(y, e_i, u)|^2 m_1(du)\big)\Big) \\
&\qquad \le 2\big(l_r|x-y|^2 + 4l(1+|y|^2)\big) \le 2\big(l_r|x-y|^2 + 4l(1+r^2)\big) \\
&= 2\big(l_r|x-y|^2 + 2l(1+r^2)|e_i - e_j|^2\big) \le L_r\big(|x-y|^2 + |e_i - e_j|^2\big) = L_r|\xi - \zeta|^2,
\end{aligned}
$$

where $L_r = max(2l_r, 4l(1+r^2))$.

Let S be the support of $\tilde{f}_2$ and $S_u = S_{u_1} \times S_{\underline{u}}$ be the projection of S on $U = \mathbb{R}^d$. By (A3), (A4) and the fact that $\bar{\mu}$ is a probability measure, we have that $m_2(S_u) = m_L(S_{u_1}) \cdot \bar{\mu}(S_{\underline{u}}) < \infty$, where m_L is the Lebesgue measure.

By (A4) and definition of function c we have that for all $t > 0$, $i = 1, \dots N$

$$
\int_0^t \int_{\mathbb{R}^d} |\tilde{f}_2(x, e_i, u)| p_2(ds, du) < \infty, \quad P\text{-a.s.}
$$

We have shown that coefficients of Equation (3.4.6) satisfy the conditions of Theorems 3.3.6 and 3.3.11, thus Equation (3.4.6) (correspondingly (3.4.1)-(3.4.2)) has an a.s. unique strong solution.

It is clear that under conditions of the theorem the solution $\{\xi_t\} = \{(X_t, \theta_t)^T\}$ admits the decomposition (3.2.1) with

$$
\begin{aligned}
A_t &= \int_0^t \tilde{a}(\xi_s) ds + \int_0^t \int_U \tilde{f}_2(\xi_{s-}, u) p_2(ds, du) \in \mathscr{V}^n, \\
M_t &= \int_0^t \tilde{b}(\xi_s) dW_s + \int_0^t \int_U \tilde{f}_1(\xi_{s-}, u) q_1(ds, du) \in \mathscr{M}^n_{loc},
\end{aligned}
$$

hence it is a semimartingale. □

Following Blom (2003), one can show that solution of (3.4.1)-(3.4.2) is indistinguishable from the solution of the following set of equations:

$$d\theta_t = \sum_{i=1}^{N}(e_i - \theta_{t-})p_2\left(dt, (\Sigma_{i-1}(X_{t-}, \theta_{t-}), \Sigma_i(X_{t-}, \theta_{t-})] \times \mathbb{R}^{d-1}\right), \tag{3.4.7}$$

$$\begin{aligned} dX_t = a(X_t, \theta_t)dt + b(X_t, \theta_t)dW_t + \int_{\mathbb{R}^d} g_1(X_{t-}, \theta_{t-}, u)q_1(dt, du) \\ + \int_{\mathbb{R}^d} \phi(X_{t-}, \theta_{t-}, \theta_t, \underline{u})p_2\big(dt, \big(0, \Sigma_N(X_{t-}, \theta_{t-})\big] \times d\underline{u}\big). \end{aligned} \tag{3.4.8}$$

Corollary 3.4.2. *Assume (A1)-(A4). Let p_1, p_2, W, X_0 and θ_0 be independent. Then SDE (3.4.7)-(3.4.8) has a unique strong solution which is a semimartingale.*

Proof. The proof consists of showing that the solution of (3.4.7)-(3.4.8) is indistinguishable from the solution of (3.4.1)-(3.4.2). Subsequently Corollary 3.4.2 is the consequence of Theorem 3.4.1.

Indeed, rewriting of (3.4.7) yields (3.4.2):

$$\begin{aligned} d\theta_t &= \sum_{i=1}^{N}(e_i - \theta_{t-})p_2\left(dt, (\Sigma_{i-1}(X_{t-}, \theta_{t-}), \Sigma_i(X_{t-}, \theta_{t-})] \times \mathbb{R}^{d-1}\right) \\ &= \int_{\mathbb{R}^d} \sum_{i=1}^{N}(e_i - \theta_{t-})I_{(\Sigma_{i-1}(X_{t-}, \theta_{t-}), \Sigma_i(X_{t-}, \theta_{t-})]}(u_1)p_2(dt, du_1 \times d\underline{u}) \\ &= \int_{\mathbb{R}^d} c(X_{t-}, \theta_{t-}, u)p_2(dt, du). \end{aligned}$$

Next, since the first three right hand terms of (3.4.8) and (3.4.1) are equal, it remains to show that the fourth right hand term in (3.4.8) yields the fourth right hand term in (3.4.1) up to indistinguishability:

$$\begin{aligned} &\int_{\mathbb{R}^d} \phi(X_{t-}, \theta_{t-}, \theta_t, \underline{u})p_2\big(dt, \big(0, \Sigma_N(X_{t-}, \theta_{t-})\big] \times d\underline{u}\big) \\ &\quad = \int_{(0,\infty)} \int_{\mathbb{R}^{d-1}} \phi(X_{t-}, \theta_{t-}, \theta_t, \underline{u})I_{(0, \Sigma_N(X_{t-}, \theta_{t-})]}(u_1)p_2(dt, du_1 \times d\underline{u}) \\ &\quad = \int_{(0,\infty)} \int_{\mathbb{R}^{d-1}} \phi(X_{t-}, \theta_{t-}, \theta_t, \underline{u}) \times \\ &\qquad \times \sum_{i=1}^{N} I_{(\Sigma_{i-1}(X_{t-}, \theta_{t-}), \Sigma_i(X_{t-}, \theta_{t-})]}(u_1)p_2(dt, du_1 \times d\underline{u}) \end{aligned}$$

$$
\begin{aligned}
&= \int_{(0,\infty)} \int_{\mathbb{R}^{d-1}} \sum_{i=1}^{N} \big[\phi(X_{t-}, \theta_{t-}, \theta_t, \underline{u}) \times \\
&\qquad \times I_{(\Sigma_{i-1}(X_{t-},\theta_{t-}),\Sigma_i(X_{t-},\theta_{t-})]}(u_1)\big] p_2(dt, du_1 \times d\underline{u}) \\
&= \int_{(0,\infty)} \int_{\mathbb{R}^{d-1}} \sum_{i=1}^{N} \big[\phi(X_{t-}, \theta_{t-}, \theta_{t-} + \Delta\theta_t, \underline{u}) \times \\
&\qquad \times I_{(\Sigma_{i-1}(X_{t-},\theta_{t-}),\Sigma_i(X_{t-},\theta_{t-})]}(u_1)\big] p_2(dt, du_1 \times d\underline{u}) \\
&= \int_{(0,\infty)} \int_{\mathbb{R}^{d-1}} \sum_{i=1}^{N} \big[\phi(X_{t-}, \theta_{t-}, \theta_{t-} + (e_i - \theta_{t-}), \underline{u}) \times \\
&\qquad \times I_{(\Sigma_{i-1}(X_{t-},\theta_{t-}),\Sigma_i(X_{t-},\theta_{t-})]}(u_1)\big] p_2(dt, du_1 \times d\underline{u}) \\
&= \int_{(0,\infty)} \int_{\mathbb{R}^{d-1}} \sum_{i=1}^{N} \big[\phi(X_{t-}, \theta_{t-}, e_i, \underline{u}) \times \\
&\qquad \times I_{(\Sigma_{i-1}(X_{t-},\theta_{t-}),\Sigma_i(X_{t-},\theta_{t-})]}(u_1)\big] p_2(dt, du_1 \times d\underline{u}) \\
&= \int_{\mathbb{R}^d} g_2(X_{t-}, \theta_{t-}, u) p_2(dt, du).
\end{aligned}
$$

This completes the proof. □

Remark 3.4.3. We notice the interesting aspect that the presence of θ_t in ϕ (Equation (3.4.8)) explicitly shows that jump of $\{X_t\}$ depends on the switch from θ_{t-} to θ_t, i.e., it is a hybrid jump.

3.5 Instantaneous hybrid jumps at a boundary

Up to now we have considered $\mathbb{R}^n \times \mathbb{M}$-valued processes the jumps and switches of which are driven by Poisson random measure. In this section we will consider $\mathbb{R}^n \times \mathbb{M}$-valued processes which also have instantaneous jumps and switches when hitting boundaries of some given sets. In order to simplify the analysis we assume that the purely discontinuous martingale term is equal to zero (i.e. we take $g_1 \equiv 0$).

Suppose we have a collection of N_b open connected sets $E^i \subset \mathbb{R}^n$, with boundaries ∂E^i, $i = 1, 2, \ldots, N_b$. Let

$$
E = \{x \,|\, x \in E^i, \text{ for some } i = 1, \ldots, N_b\} = \bigcup_{i=1}^{N_b} E^i,
$$

$$
\partial E = \{x \,|\, x \in \partial E^i, \text{ for some } i = 1, \ldots, N_b\} = \bigcup_{i=1}^{N_b} \partial E^i.
$$

The interior of the set E is the jump "destination" set. Similarly as in (Blom *et al.*, 2003*b*, pp. 38-39) we consider an increasing sequence of stopping times

$$\tau_k^E \triangleq \inf\{t > \tau_{k-1}^E : X_t^k \in \partial E\}, \tag{3.5.1}$$

$$\tau_0^E \triangleq 0, \tag{3.5.2}$$

$k = 1, 2, \ldots,$ (in order that this sequence of stopping times is well defined we need additional conditions which we provide later) and a sequence of jump-diffusions $\{X_t^n \,;\, t \geq \tau_{n-1}^E\}$, $n = 1, 2, \ldots,$ governed by the following SDE (in integral form):

$$X_t^n = X^n_{\tau_{n-1}^E} + \int_{\tau_{n-1}^E}^t a(X_s^n, \theta_s^n)ds + \int_{\tau_{n-1}^E}^t b(X_s^n, \theta_s^n)dW_s \tag{3.5.3}$$

$$+ \int_{\tau_{n-1}^E}^t \int_{\mathbb{R}^d} g_2(X_{s-}^n, \theta_{s-}^n, u)p_2(ds, du),$$

$$\theta_t^n = \theta^n_{\tau_{n-1}^E} + \int_{\tau_{n-1}^E}^t \int_{\mathbb{R}^d} c(X_{s-}^n, \theta_{s-}^n, u)p_2(ds, du), \tag{3.5.4}$$

$$X^{n+1}_{\tau_n^E} = g^x(X^n_{\tau_n^E}, \theta^n_{\tau_n^E}, \beta_{\tau_n^E}), \tag{3.5.5}$$

$$\theta^{n+1}_{\tau_n^E} = g^\theta(X^n_{\tau_n^E}, \theta^n_{\tau_n^E}, \beta_{\tau_n^E}), \tag{3.5.6}$$

$$g^x : \partial E \times \mathbb{M} \times V \to \mathbb{R}^n, \tag{3.5.7}$$

$$g^\theta : \partial E \times \mathbb{M} \times V \to \mathbb{M}, \tag{3.5.8}$$

and $\{\beta_t, t \in [0, \infty)\}$ is the sequence of V-valued (one may take $V = \mathbb{R}^d$) i.i.d. random variables distributed according to some given distribution. The initial values X_0^1 and θ_0^1 are some prescribed random variables.

We define the process $\{X_t, \theta_t\}$ as follows

$$\left\{ \begin{array}{rcl} X_t(\omega) & = & \sum_{n=1}^\infty X_t^n(\omega) I_{\left[\tau_{n-1}^E(\omega), \tau_n^E(\omega)\right)}(t) \\ \theta_t(\omega) & = & \sum_{n=1}^\infty \theta_t^n(\omega) I_{\left[\tau_{n-1}^E(\omega), \tau_n^E(\omega)\right)}(t) \end{array} \right. \tag{3.5.9}$$

provided there exist solutions $\{X_t^n, \theta_t^n\}$ of SDE (3.5.3)-(3.5.6). On the open set E, process $\{X_t, \theta_t\}$ (provided it exists) evolves as a switching jump diffusion (3.4.1)-(3.4.2). At times τ_k^E jumps and switchings are determined by the mappings g^x and g^θ correspondingly, i.e. $X_{\tau_k^E} \neq X_{\tau_k^E -}$ and possibly $\theta_{\tau_k^E} \neq \theta_{\tau_k^E -}$.

In order that the sequence of stopping times (3.5.1) is well defined and $\tau_0^E < \tau_1^E < \cdots < \tau_k^E < \ldots$ and $\tau_k^E \uparrow \infty$ a.s. as $k \uparrow \infty$, we need the following assumptions.

(B1) Function g_2, defined by (3.4.4), in addition to requirement (A4) has the following property: $(x + g_2(x, \theta, u)) \in E^i$ for each $x \in E^i$, $\theta \in \mathbb{M}$, $u \in \mathbb{R}^d$, $i = 1, \ldots, N_b$.

(B2) $d(\partial E, g^x(\partial E, \mathbb{M}, V)) > 0$, i.e. when $\{X_t\}$ has reached the boundary ∂E it always jumps inside of open set E.

(B3) Process (3.5.9) hits the boundary ∂E a.s. finitely many times on any finite time interval.

Remark 3.5.1. Assumption (B1) means that if component $\{X_t\}$ evolves inside of set E^i, it can leave it only by hitting the boundary ∂E^i under the effect of the continuous dynamics between the jumps and/or switching times generated by Poisson random measure p_2.

Theorem 3.5.2. *Assume (A1)-(A4) and (B1)-(B3). Let W, p_2, $\{\beta_t, t \in [0,\infty)\}$, X_0 and θ_0 be independent. Then process (3.5.9) exists for every $t \in \mathbb{R}_+$, it is strongly unique and it is a semimartingale.*

Proof. Let $\mathscr{F}_t$ be the σ-algebra generated by X_0, W_s, $p_2(ds,du)$, and β_s with $s \le t$. Suppose $\tau_0^E < \tau_1^E < \dots$ is the sequence of all instantaneous jumps at the boundary ∂E. By assumptions (B1)-(B3) the number of these jumps is a.s. finite on every finite time interval and $\tau_k^E \uparrow \infty$ a.s. Similarly as in the proofs of Theorems 3.3.6 and 3.3.11 it suffices to establish uniqueness and existence of the process (3.5.9) on interval $[\tau_{k-1}^E, \tau_k^E]$ with assumption that $\mathscr{F}_{\tau_{k-1}^E}$-measurable random variable $(X_{\tau_{k-1}^E}, \theta_{\tau_{k-1}^E})$ is given. Then we establish by induction that (3.5.9) exists and is unique on $\bigcup_{k=1}^{\infty}[\tau_{k-1}^E, \tau_k^E] = \mathbb{R}_+$.

Suppose $(X_{\tau_{k-1}^E}, \theta_{\tau_{k-1}^E}) = (X^k_{\tau_{k-1}^E}, \theta^k_{\tau_{k-1}^E})$ is $\mathscr{F}_{\tau_{k-1}^E}$-measurable. Then under conditions (A1)-(A4), and (3.4.5), (3.4.3), (3.4.4) and using the same arguments as in Remark 3.3.10 it follows from Theorem 3.4.1 that for $\tau_{k-1}^E \le t < \tau_k^E$ there exists strongly unique process

$$\left\{ \begin{array}{rcl} X_t & = & X_t^k, \\ \theta_t & = & \theta_t^k. \end{array} \right. \tag{3.5.10}$$

It remains to show that $(X_{\tau_k^E}, \theta_{\tau_k^E})$ is $\mathscr{F}_{\tau_k^E}$-measurable and uniquely defined. By definition of the process (3.5.9) we have:

$$\left\{ \begin{array}{rcl} X_{\tau_k^E} & = & X^{k+1}_{\tau_k^E} = g^x(X^k_{\tau_k^E}, \theta^k_{\tau_k^E}, \beta_{\tau_k^E}), \\ \theta_{\tau_k^E} & = & \theta^{k+1}_{\tau_k^E} = g^\theta(X^k_{\tau_k^E}, \theta^k_{\tau_k^E}, \beta_{\tau_k^E}). \end{array} \right. \tag{3.5.11}$$

From adaptedness and right continuity of (X_t^k, θ_t^k) follows that it is progressively measurable, hence $(X^k_{\tau_k^E}, \theta^k_{\tau_k^E})$ is $\mathscr{F}_{\tau_k^E}$-measurable. $\beta_{\tau_k^E}$ is also $\mathscr{F}_{\tau_k^E}$-measurable. Thus the right hand side of (3.5.11) is $\mathscr{F}_{\tau_k^E}$-measurable, i.e. $(X_{\tau_k^E}, \theta_{\tau_k^E})$ is $\mathscr{F}_{\tau_k^E}$-measurable. From the strong uniqueness of $\{(X_t, \theta_t)\ ;\ t \in [\tau_{k-1}^E, \tau_k^E)\}$ follows strong uniqueness of $(X_{\tau_k^E}, \theta_{\tau_k^E})$:

$$X_{\tau_k^E} = g^x(X^k_{\tau_k^E}, \theta^k_{\tau_k^E}, \beta_{\tau_k^E}) = g^x(X^k_{\tau_k^E-}, \theta^k_{\tau_k^E-}, \beta_{\tau_k^E}) = g^x(X_{\tau_k^E-}, \theta_{\tau_k^E-}, \beta_{\tau_k^E}),$$

$$\theta_{\tau_k^E} = g^\theta(X^k_{\tau_k^E}, \theta^k_{\tau_k^E}, \beta_{\tau_k^E}) = g^\theta(X^k_{\tau_k^E-}, \theta^k_{\tau_k^E-}, \beta_{\tau_k^E}) = g^\theta(X_{\tau_k^E-}, \theta_{\tau_k^E-}, \beta_{\tau_k^E}).$$

By induction we obtain that process (3.5.9) exists and is strongly unique on $\bigcup_{k=1}^{\infty}[\tau_{k-1}^E, \tau_k^E] = \mathbb{R}_+$. Moreover, it is a semimartingale, since by the Theorem 3.4.1 each solution $\{X_t^k, \theta_t^k\}$, $k = 1, 2, \dots$ is a semimartingale. □

3.6 Related SDE models on hybrid state spaces

In this section we first present stochastic hybrid models developed by Blom (2003); Ghosh and Bagchi (2004) and then compare them with the models presented in Sections 3.4 and 3.5. We will use the same notations and definitions of coefficients as in Sections 3.4 and 3.5. Table 3.2 lists the models we are dealing within this section.

Table 3.2: List of models and their main features

	θ	$X1$	$X2$	$\theta\&X2$	B
HB1, (Blom, 2003)	✓	-	✓	✓	-
HB2, (Blom *et al.*, 2003*b*)	✓	-	✓	✓	✓
GB1, (Ghosh and Bagchi, 2004)	✓	-	✓	✓	-
GB2, (Ghosh and Bagchi, 2004)	✓	-	-	-	✓
KB1, (Krystul and Blom, 2005*a*)	✓	✓	✓	✓	-
KB2, (Krystul and Blom, 2005*a*)	✓	-	✓	✓	✓

The conventions used in Table 3.2 have the following meaning:

HB1 stands for switching hybrid-jump diffusion of Blom (2003);

HB2 stands for switching hybrid-jump diffusion with hybrid jumps at the boundary of Blom *et al.* (2003*b*);

GB1 stands for switching jump diffusion of Ghosh and Bagchi (2004);

GB2 stands for switching diffusion with hybrid jumps at the boundary of Ghosh and Bagchi (2004);

KB1 stands for switching hybrid-jump diffusion developed in Section 3.4;

KB2 stands for switching hybrid-jump diffusion with hybrid jumps at the boundary developed in Section 3.5.

θ stands for independent random switching of θ_t;

$X1$ stands for independent random jump of X_t generated by compensated Poisson random measure;

$X2$ stands for independent random jump of X_t generated by Poisson random measure;

$\theta\&X2$ stands for simultaneous jump of X_t and θ_t generated by Poisson random measure;

B stands for simultaneous jump of X_t and θ_t at the boundary.

3.6.1 Stochastic hybrid model HB1 of Blom

First, we present the model HB1 (see Table 3.2). Blom (2003) placed the SDE of Lepeltier and Marchal (1976) on a hybrid state space and showed that the strong uniqueness and existence identified by Lepeltier and Marchal (1976) carry over to the following resulting SDE of Itô-Skorohod type:

$$dX_t = a(X_t, \theta_t)dt + b(X_t, \theta_t)dW_t + \int_{\mathbb{R}^d} \phi(X_{t-}, \theta_{t-}, \theta_t, \underline{u}) p_2\big(dt, \big(0, \Sigma_N(X_{t-}, \theta_{t-})\big] \times d\underline{u}\big), \tag{3.6.1}$$

$$d\theta_t = \sum_{i=1}^{N} (e_i - \theta_{t-}) p_2\left(dt, (\Sigma_{i-1}(X_{t-}, \theta_{t-}), \Sigma_i(X_{t-}, \theta_{t-})] \times \mathbb{R}^{d-1}\right). \tag{3.6.2}$$

The solution $\{X_t, \theta_t\}$ governed by SDE (3.6.1)-(3.6.2) is an $\mathbb{R}^n \times \mathbb{M}$ valued stochastic process, the set $\mathbb{M} = \{e_1, \ldots, e_N\}$ is defined as in Section 3.4, and

(i) X_0 is an $\mathbb{R}^n$-valued random variable;

(ii) θ_0 is an $\mathbb{M}$-valued random variable;

(iii) W is an m-dimensional standard Wiener process;

(iv) $p_2(dt, du)$ is a Poisson random measure with intensity $dt \times m_2(du) = dt \times du_1 \times \bar{\mu}(d\underline{u})$, where $\bar{\mu}$ is a probability measure on $\mathbb{R}^{d-1}$, $u_1 \in \mathbb{R}$, $\underline{u}$ refers to all components of $u \in \mathbb{R}^d$ except the first one.

The coefficients are defined in the same way as in Section 3.4.

Let $S_{\underline{u}} \subset \mathbb{R}^{d-1}$ denote the projection of the support of function $\phi : \mathbb{R}^n \times \mathbb{M} \times \mathbb{M} \times \mathbb{R}^{d-1} \longrightarrow \mathbb{R}^n$ on $\mathbb{R}^{d-1}$. There are three different combinations of jumps and/or switches possible:

(i) If $u_1 \in (\Sigma_{j-1}(x, e_i), \Sigma_j(x, e_i)]$, $i, j = 1, \ldots, N$ and $j \neq i$, and $\underline{u} \in S_{\underline{u}}$ then simultaneous jumps of X_t and switches of θ_t are possible.

(ii) If $u_1 \in (\Sigma_{j-1}(x, e_i), \Sigma_j(x, e_i)]$, $i, j = 1, \ldots, N$ and $j \neq i$, and $\underline{u} \notin S_{\underline{u}}$ then only random switches of θ_t are possible.

(iii) If $u_1 \in (\Sigma_{j-1}(x, e_j), \Sigma_j(x, e_j)]$, $j = 1, \ldots, N$, and $\underline{u} \in S_{\underline{u}}$ then only random jumps of X_t are possible.

Blom (2003) showed that under the following assumptions there exists an a.s. unique strong solution of SDE (3.6.1)-(3.6.2).

(C1) There exists a constant l such that for each $i = 1, 2, \ldots, N$

$$|a(x, e_i)|^2 + |b(x, e_i)|^2 \leq l(1 + |x|^2).$$

(C2) for any $r > 0$ one can specify constant l_r such that for each $i = 1, 2, \ldots, N$

$$|a(x, e_i) - a(y, e_i)|^2 + |b(x, e_i) - b(y, e_i)|^2 \leq l_r|x - y|^2$$

for $|x| \leq r$, $|y| \leq r$.

(C3) Measurable mappings Σ_i, $i = 1, \ldots, N$, are defined by (3.4.5) and for $i, j = 1, 2, \ldots, N$, $\lambda(\cdot, e_i, e_j)$ are bounded and measurable, $\lambda(\cdot, e_i, e_j) \geq 0$.

(C4) For any $k > 0$ one can specify constant N_k such that for each $i, j = 1, 2, \ldots N$

$$\sup_{|x| \leq k} \int_{\mathbb{R}^{d-1}} |\phi(x, e_i, e_j, \underline{u})| \bar{\mu}(d\underline{u}) \leq N_k.$$

The stochastic model HB2 was first presented in (Blom *et al.*, 2003*b*). It is constructed in quite the same way as the model KB2 in Section 3.5. Actually HB2 and KB2 fall into one class of SDE. So we omit the description of model HB2 and refer the reader to (Blom *et al.*, 2003*b*) or Section 3.5. It is worth to mention that the difference between these two SDE is in the way the strong existence and uniqueness of the solution was proved.

3.6.2 Stochastic hybrid model GB1 of Ghosh and Bagchi

Now, let us consider the model GB1 of Ghosh and Bagchi (2004).

The evolution of $\mathbb{R}^n \times \mathbb{M}$-valued Markov process $\{X_t, \theta_t\}$ is governed by the following equations:

$$dX_t = a(X_t, \theta_t)dt + b(X_t, \theta_t)dW_t + \int_{\mathbb{R}} g(X_{t-}, \theta_{t-}, u)p(dt, du), \tag{3.6.3}$$

$$d\theta_t = \int_{\mathbb{R}} h(X_{t-}, \theta_{t-}, u)p(dt, du). \tag{3.6.4}$$

Here:

(i) For $t = 0$, X_0 is a prescribed $\mathbb{R}^n$-valued random variable.

(ii) For $t = 0$, θ_0 is a prescribed $\mathbb{M}$-valued random variable, $\mathbb{M} = \{e_1, \ldots, e_N\}$.

(iii) W is an n-dimensional standard Wiener process.

(iv) $p(dt, du)$ is a Poisson random measure with intensity $dt \times m(du)$, where m is the Lebesgue measure on $\mathbb{R}$. p is assumed to be independent of W.

The coefficients are defined as:

$$\begin{aligned} a &: \mathbb{R}^n \times \mathbb{M} \to \mathbb{R}^n \\ b &: \mathbb{R}^n \times \mathbb{M} \to \mathbb{R}^{n\times n} \\ g &: \mathbb{R}^n \times \mathbb{M} \times \mathbb{R} \to \mathbb{R}^n \\ h &: \mathbb{R}^n \times \mathbb{M} \times \mathbb{R} \to \mathbb{R}^N. \end{aligned}$$

Function h is defined as:

$$h(x, e_i, u) = \begin{cases} e_j - e_i & \text{if } u \in \Delta_{ij}(x) \\ 0 & \text{otherwise,} \end{cases} \tag{3.6.5}$$

where for $i, j \in \{1, \ldots, N\}$, $i \neq j$, $x \in \mathbb{R}^n$, $\Delta_{ij}(x)$ are the intervals of the real line defined as:

$$\begin{aligned} \Delta_{12}(x) &= [0, \lambda_{12}(x)) \\ \Delta_{13}(x) &= [\lambda_{12}(x), \lambda_{12}(x) + \lambda_{13}(x)) \\ &\vdots \\ \Delta_{1N}(x) &= \Big[\textstyle\sum_{j=2}^{N-1} \lambda_{1j}(x), \sum_{j=2}^{N} \lambda_{1j}(x)\Big) \\ \Delta_{21}(x) &= \Big[\textstyle\sum_{j=2}^{N} \lambda_{1j}(x), \sum_{j=2}^{N} \lambda_{1j}(x) + \lambda_{21}(x)\Big) \end{aligned}$$

and so on. In general,

$$\Delta_{ij}(x) = \Big[\sum_{i'=1}^{i-1} \sum_{\substack{j'=1 \\ j'\neq i'}}^{N} \lambda_{i'j'}(x) + \sum_{\substack{j'=1 \\ j'\neq i}}^{j-1} \lambda_{ij'}(x), \sum_{i'=1}^{i-1} \sum_{\substack{j'=1 \\ j'\neq i'}}^{N} \lambda_{i'j'}(x) + \sum_{\substack{j'=1 \\ j'\neq i}}^{j} \lambda_{ij'}(x)\Big).$$

For fixed x these are disjoint intervals, and the length of $\Delta_{ij}(x)$ is $\lambda_{ij}(x)$, $\lambda_{ij} : \mathbb{R}^n \to \mathbb{R}$, $i, j = 1, \ldots, N$, $i \neq j$.

Let K_1 be the support of $g(\cdot, \cdot, \cdot)$ and let U_1 be the projection of K_1 on $\mathbb{R}$. It is assumed that U_1 is bounded. Let K_2 denote the support of $h(\cdot, \cdot, \cdot)$ and U_2 the projection of K_2 on $\mathbb{R}$. By definition of c, U_2 is a bounded set. One can define function $g(\cdot, \cdot, \cdot)$ so that the sets U_1 and U_2 form three nonempty sets: $U_1 \setminus U_2$, $U_1 \cap U_2$ and $U_2 \setminus U_1$ (see Figure 3.1). Then, we have the following:

(i) For $u \in U_1 \cap U_2$

$$\begin{cases} g(\cdot, \cdot, u) \neq 0 \\ h(\cdot, \cdot, u) \neq 0 \end{cases}$$

i.e., simultaneous jumps of X_t and switches of θ_t are possible.

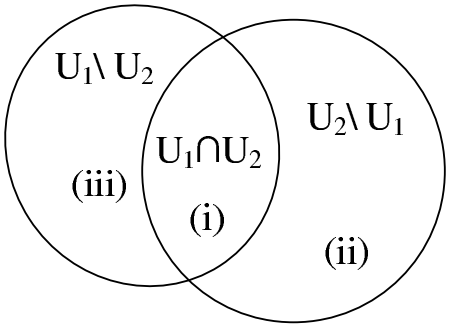


Figure 3.1: $U_1 \cup U_2$ is the projection of set $K_1 \cup K_2$ on $\mathbb{R}$

(ii) For $u \in U_2 \setminus U_1$

$$\begin{cases} g(\cdot,\cdot,u) = 0 \\ h(\cdot,\cdot,u) \neq 0 \end{cases}$$

i.e., only random switches of θ_t are possible.

(iii) For $u \in U_1 \setminus U_2$

$$\begin{cases} g(\cdot,\cdot,u) \neq 0 \\ h(\cdot,\cdot,u) = 0 \end{cases}$$

i.e., only random jumps of X_t are possible.

Ghosh and Bagchi (2004) proved that under the following conditions there exists an a.s. unique strong solution of SDE (3.6.3)-(3.6.4).

(D1) For each $e_i \in \mathbb{M}$, $i = 1, \ldots, N$, $a(\cdot, e_i)$ and $b(\cdot, e_i)$ are bounded and Lipschitz continuous.

(D2) For all $i, j \in \{1, \ldots, N\}$, $i \neq j$, functions $\lambda_{ij}(\cdot)$ are bounded and measurable, $\lambda_{ij}(\cdot) \geq 0$ for $i \neq j$ and $\sum_{j=1}^{N} \lambda_{ij}(\cdot) = 0$ for any $i \in \{1, \ldots, N\}$.

(D3) U_1, the projection of support of $g(\cdot,\cdot,\cdot)$ on $\mathbb{R}$, is bounded.

3.6.3 Stochastic hybrid model GB2 of Ghosh and Bagchi

Next, we present the GB2 model of Ghosh and Bagchi (2004). The state of the system at time t, denoted by (X_t, θ_t), takes values in $\bigcup_{n=1}^{\infty}(S_n \times \mathbb{M}_n)$, where $\mathbb{M}_n = \{e_1, e_2, \ldots, e_{N_n}\}$ and $S_n \subset \mathbb{R}^{d_n}$. Between the jumps of X_t the state equations are

of the form

$$dX_t = a^n(X_t, \theta_t)dt + b^n(X_t, \theta_t)dW_t^n, \tag{3.6.6}$$

$$d\theta_t = \int_{\mathbb{R}} h^n(X_{t-}, \theta_{t-}, u)p(dt, du), \tag{3.6.7}$$

where for each $n \in \mathbb{N}$

$$a^n : S_n \times \mathbb{M}_n \to \mathbb{R}^{d_n}$$
$$b^n : S_n \times \mathbb{M}_n \to \mathbb{R}^{d_n \times d_n}$$
$$h^n : S_n \times \mathbb{M}_n \times \mathbb{R} \to \mathbb{R}^{N_n}.$$

Function h^n is defined in a similar way as (3.6.5) with rates $\lambda_{ij}^n : S_n \to \mathbb{R}$, $\lambda_{ij}^n \geq 0$ for $i \neq j$, and $\sum_{j=1}^{N_n} \lambda_{ij}^n(\cdot) = 0$ for any $i \in \{1, \ldots, N\}$. W^n is a standard d_n-dimensional Wiener process, p is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with the intensity $dt \times m(du)$ as in the previous section.

For each $n \in \mathbb{N}$, let $A_n \subset S_n$, $D_n \subset S_n$. The set A_n is the set of instantaneous jumps, whereas D_n is the destination set. It is assumed that for each $n \in \mathbb{N}$, A_n and D_n are closed sets, $A_n \cap D_n = \varnothing$ and $\inf_n d(A_n, D_n) > 0$, where $d(\cdot, \cdot)$ denotes the distance between two sets. If at some random time X_t hits A_n, then it executes an instantaneous jump. The destination of (X_t, θ_t) at this juncture is determined by a map

$$g_n : A_n \times \mathbb{M}_n \to \cup_{m \in \mathbb{N}}(D_m \times \mathbb{M}_m).$$

After reaching the destination, the process $\{X_t, \theta_t\}$ follows the same evolutionary mechanism over and over again.

Let $\{\eta_t\}$ be an $\mathbb{N}$ valued process defined by

$$\eta_t = n \text{ if } (X_t, \theta_t) \in S_n \times \mathbb{M}_n. \tag{3.6.8}$$

The $\{\eta_t\}$ is a piecewise constant process that changes from n to m when (X_t, θ_t) jumps from the regime $S_n \times \mathbb{M}_n$ to the regime $S_m \times \mathbb{M}_m$. Thus η_t is an indicator of a regime and a change in η_t means a switching in the regimes in which $\{X_t, \theta_t\}$ evolves.

Let

$$\tilde{S} = \{(x, e_i, n) | x \in S_n, e_i \in \mathbb{M}_n\},$$
$$\tilde{A} = \{(x, e_i, n) | x \in A_n, e_i \in \mathbb{M}_n\},$$
$$\tilde{D} = \{(x, e_i, n) | x \in D_n, e_i \in \mathbb{M}_n\}.$$

Then $\{X_t, \theta_t, \eta_t\}$ is an $\tilde{S}$-valued process, the set $\tilde{A}$ is the set where jumps occur and $\tilde{D}$ is the destination set for this process. The sets $\cup_n(S_n \times \mathbb{M}_n)$, $\cup_n(A_n \times \mathbb{M}_n)$ and $\cup_n(D_n \times \mathbb{M}_n)$ can be embedded in $\tilde{S}$, $\tilde{A}$ and $\tilde{D}$ respectively.

Let d^0 denote the injection map of $\cup_n(D_n \times \mathbb{M}_n)$ into $\tilde{D}$. Define three maps

$$\tilde{g}_i : \tilde{A} \to \tilde{D},\ i = 1, 2,$$
$$\tilde{h} : \tilde{A} \to \mathbb{N}.$$

$$\begin{aligned}
\tilde{g}_1(x, e_i, n) &= \text{ the first component in } d^0(g_n(x, e_i)),\\
\tilde{g}_2(x, e_i, n) &= \text{ the second component in } d^0(g_n(x, e_i)),\\
\tilde{h}(x, e_i, n) &= \text{ the third component in } d^0(g_n(x, e_i)).
\end{aligned}$$

Let τ_{m+1} be the stopping time defined by

$$\tau_{m+1} = \inf\{t > \tau_m | X_{t-}, \theta_{t-}, \eta_{t-} \in \tilde{A}\}.$$

The equations for $\{X_t, \theta_t, \eta_t\}$ may thus be summarized as follows:

$$\begin{aligned}
dX_t = &\Big(a(X_t, \theta_t, \eta_t) + \sum_{m=0}^{\infty}[\tilde{g}_1(X_{\tau_m-}, \theta_{\tau_m-}, \eta_{\tau_m-}) - X_{\tau_m-})]\delta(t - \tau_m)\Big)dt \quad (3.6.9)\\
&+ b(X_t, \theta_t, \eta_t)dW_t^{\eta_t},\\
d\theta_t = &\int_{\mathbb{R}} h(X_{t-}, \theta_{t-}, \eta_{t-}, u)p(dt, du) \quad (3.6.10)\\
&+ \sum_{m=0}^{\infty}[\tilde{g}_2(X_{\tau_m-}, \theta_{\tau_m-}, \eta_{\tau_m-}) - \theta_{\tau_m-})]\delta(t - \tau_m)dt,\\
d\eta_t = &\sum_{m=0}^{\infty}[\tilde{h}(X_{\tau_m-}, \theta_{\tau_m-}, \eta_{\tau_m-}) - \eta_{\tau_m-})]I_{\{\tau_m \le t\}}, \quad (3.6.11)
\end{aligned}$$

where δ is the Dirac measure and $a(x, e_i, n) = a^n(x, e_i)$, $b(x, e_i, n) = b^n(x, e_i)$, $h(x, e_i, n, u) = h^n(x, e_i, u)$.

To ensure the existence of an a.s. unique strong solution of SDE (3.6.9)-(3.6.11), Ghosh and Bagchi (2004) adopted the following assumptions:

(E1) For each $n \in \mathbb{N}$ and $e_i \in \mathbb{M}_i$, $a^n(\cdot, e_i)$ and $b^n(\cdot, e_i)$ are bounded and Lipschitz continuous.

(E2) For each $n \in \mathbb{N}$, $i, j = 1, \ldots, M_n$, $i \neq j$, functions $\lambda_{ij}^n(\cdot)$ are bounded and measurable, $\lambda_{ij}^n(\cdot) \ge 0$ for $i \neq j$ and $\sum_{j=1}^N \lambda_{ij}^n(\cdot) = 0$ for any $i \in \{1, \ldots, N\}$.

(E3) The maps g_n, $n \in \mathbb{N}$, are bounded and uniformly continuous.

(E4) $\inf_n d(A_n, D_n) > 0$.

3.6.4 Hierarchy between stochastic hybrid models

In this subsection we discuss the differences between the models and determine the relative hierarchy of these models.

First, let us consider models HB1 and GB1 (see Sections 3.6.1 and 3.6.2). Both models allow either independent or simultaneous jumps and switches of X_t and θ_t. However, there are some differences in assumptions imposed on the coefficients and in construction of the jump and switching coefficients. The first two terms (i.e. the drift and the diffusion term) in (3.6.1) and in (3.6.3) are identical. However, when proving the existence of strong unique solution of SDE (3.6.3)-(3.6.4) Ghosh and Bagchi (2004) assume that the drift and the diffusion coefficients are bounded, i.e. condition (D1). To prove the similar result for SDE (3.6.1)-(3.6.2) more general growth condition (C1) is adopted. The construction of the "switching" terms (3.6.2) and (3.6.4) is almost identical with some minor differences in defining the "rate" intervals. The conditions on the "rate" functions $\lambda(\cdot, e_i, e_j)$ and $\lambda_{ij}(\cdot)$ are the same, i.e. these functions are assumed to be bounded and measurable for all $i, j = 1, \ldots, N$, i.e. conditions (C3) and (D2).

There is a substantial difference in the construction of the jump part of X_t in the HB1 and GB1 models. In GB1 the jumps of X_t are described by a stochastic integral of function g with respect to a Poisson random measure $p(dt, du)$ with intensity $dt \times m(du)$, where m is the Lebesgue measure on $U = \mathbb{R}$. In order to satisfy the existence and uniqueness of solution, U_1, the projection of support of function g on $U = \mathbb{R}$, must be bounded, i.e. condition (D3). In HB1 the jumps of X_t are also defined by a stochastic integral driven by Poisson random measure $p_2(dt, du)$ but with intensity $dt \times m(du_1) \times \bar{\mu}(\underline{u})$, where m is the Lebesgue measure on $U_1 = \mathbb{R}$ and $\bar{\mu}$ is a probability measure on $\underline{U} = \mathbb{R}^{d-1}$. The integrand function g_2, which determines the jump size of X_t, compared to function g, has an extra argument $\underline{u} \in \underline{U} = \mathbb{R}^{d-1}$, and, since the intensity of p_2 with respect to $\underline{u}$ is a probability measure $\bar{\mu}$ (which is always finite), the projection of support of g_2 on $\underline{U} = \mathbb{R}^{d-1}$ can be unbounded. This gives some extra freedom in modelling the jumps of X_t component. It is only required that function g_2 must satisfy condition (C4). From this follows that model HB1 includes model GB1 as a special case (GB1 $\subset$ HB1).

Now, let us look at models HB1 and KB1 (see Sections 3.6.1 and 3.4). At a first glance, one can see that KB1 has an extra integral term with respect to a compensated Poisson random measure q_1. This term represents a purely discontinuous martingale part of the X_t component. Conditions (A1)-(A3) adopted to SDE (3.4.1)-(3.4.2) are equivalent to conditions (C1)-(C3) in SDE (3.6.1)-(3.6.2). But condition (A4) is weaker than condition (C4). Using the same derivations as in Theorem 3.4.2 one can show that the solution of SDE (3.6.1)-(3.6.2) is indistinguishable from the solution of SDE (3.4.1)-(3.4.2) with $g_1 \equiv 0$. That means, model HB1 can be seen as a special case of KB1 (HB1 $\subset$ KB1).

Next, we compare HB2 and KB2 (see Sections 3.6.1 and 3.5). HB2 contains all models of HB1 extended with hybrid jumps at a boundary. The model KB2 is same as KB1 without integral term with respect to the compensated Poisson random measure q_1, but with hybrid jumps at the boundary. All together, KB2 and HB2 are the same class of SDE (HB2 = KB2).

Models KB2 and GB2 have some similarities (see Sections 3.5 and 3.6.3). Let us see what are the main differences between SDE (3.5.3)-(3.5.6) and SDE (3.6.9)-(3.6.11). Solutions of SDE (3.6.9)-(3.6.11) are the $\bigcup_{n=1}^{\infty}(S_n \times \mathbb{M}_n)$-valued switching diffusions with hybrid jumps at the boundary. Before hitting the boundary $\{X_t, \theta_t\}$ evolves as an $(S_n \times \mathbb{M}_n)$-valued switching diffusion in some regime $\eta_t = n \in \mathbb{N}$. The drift and the diffusion coefficients and the mapping determining a new starting point of the process after the hitting the boundary can be different for every different regime $n \in \mathbb{N}$. Solutions of SDE (3.5.3)-(3.5.6) are the $(\mathbb{R}^n \times \mathbb{M})$-valued switching-jump diffusions with hybrid jumps at the boundary. The dimension of the state space and the coefficients of SDE are fixed. Hence, on this specific point, model GB2 is more general. However the jump term in KB2, see Equation (3.5.3), is more general than the jump term in GB2, see Equation (3.6.9).

Now let us have a look at conditions (E1)-(E4). Condition (E1) implies that our local conditions (A1) and (A2) for SDE (3.4.1)-(3.4.2) are definitely satisfied. Conditions (E2) and (E3) imply that conditions (A3) and (A4) for SDE (3.4.1)-(3.4.2) are satisfied. Condition (E4) implies that (B1) and (B2) adopted to SDE (3.5.3)-(3.5.6) are satisfied. It ensures that after the jump the process starts inside of some open set, but not on a boundary. Condition (B3) of SDE (3.5.3)-(3.5.6) is missing for GB2 (Ghosh and Bagchi, 2004).

In general GB2 is not a subclass of KB2 (or HB2) since in GB2 the state of the system (X_t, θ_t) takes values in $\bigcup_{k=1}^{\infty}(S_k \times \mathbb{M}_k)$, where $\mathbb{M}_k = \{e_1, e_2, \ldots, e_{N_k}\}$ and $S_k \subset \mathbb{R}^{d_k}$ may be different for different k's. If $(S_k \times \mathbb{M}_k) = (\mathbb{R}^n \times \mathbb{M})$ for all $k \in \mathbb{N}$ then obviously GB2 $\subset$ KB2 (=HB2).

We summarize the "hierarchy" of models in Figure 3.2. The "hierarchy" is organized on the basis of the behaviors of the processes, e.g. different types of jumps, and not on the assumptions applied to the models.

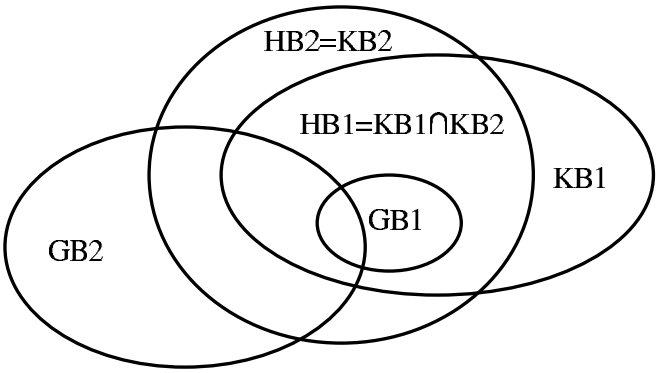


Figure 3.2: The hierarchy between stochastic hybrid models; the sets HB2=KB2 and GB2 fall within the set of Generalized Stochastic Hybrid Processes (Bujorianu and Lygeros, 2004). KB1 provides complementary modelling power in allowing processes that have infinite variation in jumps on a finite time interval.

3.7 Markov Property

In the following two sections we prove Markov and Strong Markov properties for model KB2 (Section 3.5). These results are crucial for developing efficient rare event simulation techniques as we will see in Chapter 5.

Assume we are given the following objects:

- a measurable space $(S, \mathscr{S})$;
- a measurable space $(\Omega, \mathscr{G})$ and a family of σ-algebras $\{\mathscr{G}_t^s, 0 \leq s \leq t \leq \infty\}$, such that $\mathscr{G}_t^s \subset \mathscr{G}_v^u \subset \mathscr{G}$ provided $0 \leq u \leq s \leq t \leq v$; $\mathscr{G}_t^s$ denotes a σ-algebra of events on time interval $[s, t]$; we write $\mathscr{G}_t$ in place of $\mathscr{G}_t^0$ and $\mathscr{G}^s$ in place of $\mathscr{G}_\infty^s$;
- a probability measure $P_{s,x}$ for each pair $(s, x) \in [0, \infty) \times S$ on $\mathscr{G}^s$;
- a function (stochastic process) $\xi_t(\omega) = \xi(t, \omega)$ defined on $[0, \infty) \times \Omega$ with values in S.

The system consisting of these four objects will be denoted by $\{\xi_t, \mathscr{G}_t^s, P_{s,x}\}$ (Gihman and Skorohod, 1975).

Definition 3.7.1. A system of objects $\{\xi_t, \mathscr{G}_t^s, P_{s,x}\}$ is called a Markov process provided:

(i) for each $t \in [0, \infty)$ $\xi_t(\omega)$ is measurable mapping of $(\Omega, \mathscr{G})$ into $(S, \mathscr{S})$;

(ii) for arbitrary fixed s, t and B ($0 \le s \le t, B \in \mathscr{S}$) the function $P(s,x,t,B) = P_{s,x}(\xi_t \in B)$ is $\mathscr{S}$-measurable with respect to x;

(iii) $P_{s,x}(\xi_s = x) = 1$ for all $s \ge 0$ and $x \in S$;

(iv) $P_{s,x}(\xi_u \in B \mid \mathscr{G}_t^s) = P_{t,\xi_t}(\xi_u \in B)$ for all $s, t, u, 0 \le s \le t \le u < \infty$, $x \in S$ and $B \in \mathscr{S}$.

The measure $P_{s,x}$ should be considered as a probability law which determines the probabilistic properties of the process $\xi_t(\omega)$ given that it starts at point x at the time s. Condition (iv) in Definition 3.7.1 expresses the Markov property of the processes. Let $\mathbb{E}_{s,x}$ denote the expectation with respect to measure $P_{s,x}$. For $\mathscr{G}^s$-measurable random variable $\xi(\omega)$

$$\mathbb{E}_{s,x}[\xi(\omega)] = \int \xi(\omega) P_{s,x}(d\omega).$$

It is not difficult to show that the Markov property (iv) in Definition 3.7.1 can be rewritten in terms of expectations as follows:

$$\mathbb{E}_{s,x}[f(\xi_u) \mid \mathscr{G}_t^s] = \mathbb{E}_{t,\xi_t}[f(\xi_u)],\ 0 \le s \le t \le u < \infty,$$

where f is an arbitrary $\mathscr{S}$-measurable bounded function.

Next, let us show that process

$$\left\{ \begin{array}{rcl} X_t(\omega) & = & \sum_{n=1}^{\infty} X_t^n(\omega) I_{\left[\tau_{n-1}^E(\omega), \tau_n^E(\omega)\right)}(t) \\ \theta_t(\omega) & = & \sum_{n=1}^{\infty} \theta_t^n(\omega) I_{\left[\tau_{n-1}^E(\omega), \tau_n^E(\omega)\right)}(t) \end{array} \right. \tag{3.7.1}$$

defined as a concatenation of solutions $\{X_t^n, \theta_t^n\}$ of the system of SDE (3.5.3)-(3.5.6) (see Sections 3.4 and 3.5), is Markov. We follow the approach used in (Gihman and Skorohod, 1982). Let $\xi_t^{s,\eta} = (X_t^{s,x}, \theta_t^{s,\theta})$ denote the process (3.7.1) on $[s, \infty)$ satisfying initial condition $\xi_s^{s,\eta} = \eta = (X_s^{s,x}, \theta_s^{s,\theta})$. Note that now $S = \mathbb{R}^n \times \mathbb{M}$ and $\mathscr{S} = \mathscr{B}_{\mathbb{R}^n \times \mathbb{M}}$ is the σ-algebra of Borel sets on $\mathbb{R}^n \times \mathbb{M}$. Assume that conditions of Theorem 3.5.2 are satisfied. Let $\mathscr{F}_t^s$, $s < t$ be the σ-algebras generated by $\{W_u - W_s, p_2([s,u], dz), \beta_u, u \in [s,t]\}$, $\mathscr{F}_t^0 = \mathscr{F}_t$, $\mathscr{F}_\infty^s = \mathscr{F}^s$. For $s \le t$ the σ-algebras $\mathscr{F}_s$ and $\mathscr{F}^s$ are independent. Process $\xi_t^{s,\eta}$ is $\mathscr{F}^s$-measurable, hence, it is independent of σ-algebra $\mathscr{F}_s$. Let η_s be an arbitrary $\mathbb{R}^n \times \mathbb{M}$-valued $\mathscr{F}_s$ measurable random variable. Then ξ_t^{s,η_s}, $t \ge s$, is unique $\mathscr{F}_t$-measurable process on $[s, \infty)$ satisfying the initial condition $\xi_s^{s,\eta_s} = \eta_s$. Since for $u < s$ process $\xi_t^{u,y}$ is $\mathscr{F}_t$-measurable on $[s, \infty)$ with initial condition $\xi_s^{u,y}$ then the following equality holds

$$\xi_t^{u,y} = \xi_t^{s,\xi_s^{u,y}},\ u < s < t. \tag{3.7.2}$$

Let φ be a bounded measurable function on $\mathbb{R}^n \times \mathbb{M}$, let ζ_s be an arbitrary bounded $\mathscr{F}_s$-measurable quantity. The independence of $\mathscr{F}_s$ and $\mathscr{F}^s$ and the Fubini theorem

imply that measure P on $\mathscr{F}_\infty$ is a product of measures P_s and P^s, where P_s is a restriction of P on $\mathscr{F}_s$, where P^s is a restriction of P on $\mathscr{F}^s$, and

$$\mathbb{E}[\varphi(\xi_t^{u,y})\zeta_s] = \mathbb{E}[\varphi(\xi_t^{s,\xi_s^{u,y}})\zeta_s] = \mathbb{E}\big[\zeta_s(\mathbb{E}[\varphi(\xi_t^{s,x})])_{x=\xi_s^{u,y}}\big].$$

Since $\xi_s^{u,y}$ is $\mathscr{F}_s$-measurable then $\mathbb{E}[\varphi(\xi_t^{u,y}) \mid \mathscr{F}_s] = \big[\mathbb{E}[\varphi(\xi_t^{s,x})]\big]_{x=\xi_s^{u,y}}$. Let

$$P(s,x,t,B) = P(\xi_t^{s,x} \in B),\ B \in \mathscr{B}_{\mathbb{R}^n\times\mathbb{M}}, \tag{3.7.3}$$

here $\mathscr{B}_{\mathbb{R}^n\times\mathbb{M}}$ is the σ-algebra of Borel sets on $\mathbb{R}^n \times \mathbb{M}$. Then, by taking $\varphi = I_B$, we obtain

$$P(\xi_t^{u,y} \in B \mid \mathscr{F}_s) = P(s,\xi_s^{u,y},t,B). \tag{3.7.4}$$

If ξ_t is an arbitrary process defined by (3.7.1), by the same reasoning with help of which equalities (3.7.2) and (3.7.4) have been obtained, one can show that $\xi_t = \xi_t^{s,\xi_s}$ for $s < t$ and that

$$P(\xi_t \in B \mid \mathscr{F}_s) = P(s,\xi_s,t,B).$$

Hence, the process defined by (3.7.1) is a Markov process with transition probability $P(s,x,t,B)$ defined by (3.7.4). To be precise, we have shown that the system of objects $\{(X_t,\theta_t),\mathscr{F}_t^s,P_{s,(x,\theta)}\}$, where $P_{s,(x,\theta)}\big((X_t,\theta_t)\in B\big) = P(s,(x,\theta),t,B) = P\big((X_t^{s,x},\theta_t^{s,\theta})\in B\big),\ B \in \mathscr{B}_{\mathbb{R}^n\times\mathbb{M}}$, is a Markov process.

3.8 Strong Markov property

In this section we prove the Markov property

$$P_{s,x}(\xi_u \in B|\mathscr{G}_t^s) = P_{t,\xi_t}(\xi_u \in B),\ \ s \le t \le u$$

remains valid also when a fixed time moment t is replaced by a stopping time.

Let $\{\xi_t(\omega),\mathscr{G}_t^s,P_{s,x}\}$ be a Markov process in the space $(S,\mathscr{S})$. Let $\mathscr{T}$ denote the σ-algebra of Borel sets on $[0,\infty)$.

Definition 3.8.1. A Markov process is called strong Markov if:

(i) the transition probability $P(s,x,t,B)$ for a fixed B is a $\mathscr{T}\times\mathscr{S}\times\mathscr{T}$-measurable function of (s,x,t) on the set $0 \le s \le t < \infty$, $x \in S$;

(ii) it is progressively measurable;

(iii) for any $s \ge 0$, $t \ge 0$, $\mathscr{S}$-measurable function $f(x)$ and arbitrary stopping time τ,

$$\mathbb{E}_{s,x}[f(\xi_{t+\tau}) \mid \mathscr{G}_\tau^s] = \mathbb{E}_{\tau,\xi_\tau}[f(\xi_{t+\tau})]. \tag{3.8.1}$$

Remark 3.8.2. For Equation (3.8.1) to be satisfied, it is necessary that the random variable $g(\xi_\tau, \tau, t+\tau) = \mathbb{E}_{\tau,\xi_\tau}[f(\xi_{t+\tau})]$ be $\mathscr{G}_\tau^s$-measurable. For this reason assumptions (i) and (ii) make part of the definition of the strong Markov property (Gihman and Skorohod, 1975).

Now we return to the process $\xi_t = (X_t, \theta_t)$ defined in Section 3.5. We have shown that it is a Markov process. The following proposition proves that it is a Strong Markov process also.

Proposition 3.8.3. *Assume (A1)-(A4) and (B1)-(B3). Let W, p_2, μ^E, X_0 and θ_0 be independent. Let $\mathscr{F}_t^s$, $s < t$ be the σ-algebras generated by $\{W_u - W_s, p_2(dz, [s,u]), \beta_u, u \in [s,t]\}$. For any bounded Borel function $f : \mathbb{R}^n \times \mathbb{M} \to \mathbb{R}$ and any $\mathscr{F}_t^s$-stopping time τ*

$$\mathbb{E}_{s,x}[f(\xi_{t+\tau}) \mid \mathscr{F}_\tau^s] = \mathbb{E}_{\tau,\xi_\tau}[f(\xi_{t+\tau})].$$

Proof. Let $\{\sigma_k, k = 0, 1, \dots\}$ denote the ordered set of the stopping times $\{\tau_k^E, k = 1, 2, \dots\}$ and $\{\tau_k, k = 0, 1, \dots\}$. The latter set is the set of the stopping times generated by Poisson random measure p_2. Then on each time interval $[\sigma_{k-1}, \sigma_k)$, $k = 1, 2, \dots$ process ξ_t evolves as a diffusion starting at point $\xi_{\sigma_{k-1}}$ at the time σ_{k-1}. It is known that for diffusions the strong Markov property holds (see (Øksendal, 2002)). This means that on each time interval $[\sigma_{k-1}, \sigma_k)$ the Strong Markov property for $\{X_t, \theta_t\}$ holds. Let $\mathscr{F}_\tau^s$ be the σ-algebra generated by the $\mathscr{F}_t^s$-stopping time τ. The sets $\{\omega : \tau(\omega) \in [\sigma_{k-1}(\omega), \sigma_k(\omega))\}$, $k = 1, 2, \dots$ are $\mathscr{F}_\tau^s$-measurable. Hence

$$\begin{aligned}
\mathbb{E}_{s,x}[f(\xi_{t+\tau}) \mid \mathscr{F}_\tau^s] &= \sum_{k=0}^{\infty} I_{[\sigma_{k-1},\sigma_k)}(\tau)\mathbb{E}_{s,x}\big[f(\xi_{t+\tau}) \mid \mathscr{F}_\tau^s\big] \\
&= \sum_{k=0}^{\infty} \mathbb{E}_{s,x}\big[I_{[\sigma_{k-1},\sigma_k)}(\tau) f(\xi_{t+\tau}) \mid \mathscr{F}_\tau^s\big] \\
&= \sum_{k=0}^{\infty} \mathbb{E}_{\tau,\xi_\tau}\big[I_{[\sigma_{k-1},\sigma_k)}(\tau) f(\xi_{t+\tau})\big] \\
&= \mathbb{E}_{\tau,\xi_\tau}\big[\sum_{k=0}^{\infty} I_{[\sigma_{k-1},\sigma_k)}(\tau) f(\xi_{t+\tau})\big] \\
&= \mathbb{E}_{\tau,\xi_\tau}\big[f(\xi_{t+\tau})\big].
\end{aligned}$$

This completes the proof. □

3.9 Concluding Remarks

The aim of this chapter was to significantly further the study of SDE on a hybrid space, including characterizations of its solutions in terms of pathwise uniqueness,

semimartingale and strong Markov process properties. We have used (Gihman and Skorohod, 1982) and (Jacod and Shiryaev, 1987) to identify and characterize the most general class of jump-diffusions which are defined as semimartingale solutions of SDE. This yielded a valuable improvement over the (Lepeltier and Marchal, 1976) regarding the understanding of semimartingale property and pathwise uniqueness of jump-diffusions. Next we have followed a similar path as taken by (Blom, 1990, 2003) in transferring this pathwise uniqueness and semimartingale understanding to the class of stochastic hybrid processes. This subsequently allowed to incorporate instantaneous jumps at a boundary within the same framework including pathwise uniqueness and semimartingale property. Finally we have introduced a novel approach in showing strong Markov property for general stochastic hybrid process, i.e. model KB2. The strong Markov property will prove to be very important in Chapter 5 for development of efficient rare event simulation techniques.

Chapter 4

Approximation of first passage times of switching diffusion

4.1 Introduction

In this chapter we study the problem of estimating the probability that a stochastic hybrid system reaches a given target set within some time horizon. This type of problems arise in numerous applications, e.g. reliability analysis of complex dynamical systems (Aldemir *et al.*, 1994), computer and communication systems, advanced air traffic management (Blom and Bakker, 2002), finance applications (Glasserman and Li, 2005) etc. An analytical solution to this problem is available only in some special simple cases. A widely applicable approach to estimation of expectations of functionals of a stochastic process, such as its probability distribution and its moments, is the simulation of sample paths of a corresponding discrete-time approximations. We refer the reader to Chapter 2 for an introduction to Monte Carlo simulation techniques and definitions of the concepts of weak and strong discrete-time approximations of stochastic processes.

Five distinct classes of stochastic hybrid processes, all containing the class of switching diffusions of Ghosh *et al.* (1997), have been studied in Chapter 3. In this chapter we select the stochastic hybrid model of Ghosh *et al.* (1997) in order to give a general insight on the specifics of discrete-time approximations and Monte Carlo simulations of the stochastic hybrid processes with non-trivial state dependent switching rates.

Let $\{X_t, \theta_t\}$ be the switching diffusion taking its values in $\mathbb{R}^n \times \mathbb{M}$ defined by

$$dX_t = a(\theta_t, X_t)dt + b(\theta_t, X_t)dW_t, \tag{4.1.1}$$

$$P_{\theta_{t+\delta}|\theta_t, X_t}(\theta|\eta, x) = \lambda_{\eta\theta}(x)\delta + o(\delta),\ \eta \neq \theta, \tag{4.1.2}$$

where $\mathbb{M}$ is a finite set of modes and $(W_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^n$ (Ghosh *et al.*, 1997). We set $\tau \triangleq \inf\{t > 0 : X_t \in D\}$ for the first passage time to the set $D \times \mathbb{M}$, where $D \subset \mathbb{R}^n$ is a closed connected set. We want to estimate

$$P(\tau \leq T) \tag{4.1.3}$$

where T is a fixed time. We propose to approximate the switching diffusion (4.1.1)-(4.1.2) by a discrete-time strong Euler approximation $(X^h_{t_i}, \theta^h_{t_i})_{t_i \in \mathscr{I}}$ (where $\mathscr{I}$ is a time discretization), and approximate probability (4.1.3) by

$$P(\tau^h \leq T), \tag{4.1.4}$$

where $\tau^h \triangleq \inf\{t > 0 : X^h_t \in D\}$. Using the Monte Carlo simulation method, the probability of first passage time (4.1.4) can be then approximated by a sample average of $\mathbf{1}_{\{\tau^h \leq T\}}$ where τ^h is the Euler approximation of the stopping time τ.

Discrete-time approximations of an Itô diffusion are well explained in (Kloeden and Platen, 1992). A discretization scheme for jump-diffusion process with state-dependent intensities was considered in (Glasserman and Merener, 2004). Weak approximations of killed (or stopped) diffusions were studied in (Gobet, 1999*a*,*b*, 2000, 2001) and (Moon, 2003). They develop and prove the convergence of numerical schemes that approximate the expected value $E[g(x(\tau), \tau)]$ of a given function g depending on the solution x of an Itô stochastic differential equation and on the first exit time τ from a given domain. In this chapter we develop an Euler-type discretization scheme for hybrid model (4.1.1)-(4.1.2) and prove its convergence. Following the approach of Gobet (1999*a*), we then show that

$$P(\tau \leq T) - P(\tau^h \leq T) \longrightarrow 0$$

as the discretization step tends to zero.

The organization of this chapter is the following. Section 4.2 provides important details on switching diffusion model. The strong approximation scheme and its convergence proof are presented in Section 4.3. The approximation of first passage time is discussed in Section 4.4.

4.2 Switching diffusion

The switching diffusion SDE (4.1.1)-(4.1.2) can be rewritten in a form of Itô-Skorohod SDE which is just the same as the SDE of Ghosh and Bagchi (2004) only without the jump part of X_t, i.e. $g \equiv 0$ (see Ch.III, Sec.3.6.2 in this thesis).

In order to keep this chapter self-contained, we briefly recall the main features of the model.

The discrete valued component $\{\theta_t\}$ is assumed to take its values in a space $\mathbb{M} = \{e_1, e_2, \ldots, e_N\}$ defined as a finite set of unit vectors, i.e. $e_i \in \mathbb{M}$ is a i-th unit vector in $\mathbb{R}^N$. The coefficients of SDE (4.1.1)-(4.1.2) are defined as follows:

$$\begin{aligned} a &: \mathbb{R}^n \times \mathbb{M} \to \mathbb{R}^n \\ b &: \mathbb{R}^n \times \mathbb{M} \to \mathbb{R}^{n\times n} \\ \lambda_{ij} &: \mathbb{R}^n \to \mathbb{R},\ i,j = 1,2,\ldots,N \end{aligned}$$

For each $\theta \in \mathbb{M}$, $a(\cdot,\theta)$ and $b(\cdot,\theta)$ are assumed to be bounded, continuous and Lipschitz. For all $i,j \in \{1,\ldots,N\}$ $\lambda_{ij}(\cdot)$ is assumed to be bounded, continuous and Lipschitz, $\lambda_{ij}(\cdot) \geq 0$ for $i \neq j$ and $\sum_{j=1}^{N} \lambda_{ij}(\cdot) = 0$ for any $i \in \{1,\ldots,N\}$. These conditions ensure the existence of a unique strong solution to SDE (4.1.1)-(4.1.2) (Ghosh and Bagchi, 2004).

The transformation of SDE (4.1.1)-(4.1.2) into an Itô-Skorohod type SDE is done by replacing Equation (4.1.2) by an equivalent SDE driven by a Poisson random measure. For that, one first needs to define proper "rate" intervals, the meaning of which will soon become clear.

For $i,j \in \{1,\ldots,N\}$, $i \neq j$, $x \in \mathbb{R}^n$ we construct the intervals $\Delta_{ij}(x)$ of the real line in the following manner (Ghosh *et al.*, 1993, 1997):

$$\begin{aligned} \Delta_{12}(x) &= [0, \lambda_{12}(x)) \\ \Delta_{13}(x) &= [\lambda_{12}(x), \lambda_{12}(x) + \lambda_{13}(x)) \\ &\vdots \\ \Delta_{1N}(x) &= \Big[\sum_{j=2}^{N-1} \lambda_{1j}(x), \sum_{j=2}^{N} \lambda_{1j}(x)\Big) \\ \Delta_{21}(x) &= \Big[\sum_{j=2}^{N} \lambda_{1j}(x), \sum_{j=2}^{N} \lambda_{1j}(x) + \lambda_{21}(x)\Big) \end{aligned}$$

and so on. Thus, in general,

$$\Delta_{ij}(x) = \Big[\sum_{i'=1}^{i-1} \sum_{\substack{j'=1 \\ j'\neq i'}}^{N} \lambda_{i'j'}(x) + \sum_{\substack{j'=1 \\ j'\neq i}}^{j-1} \lambda_{ij'}(x), \sum_{i'=1}^{i-1} \sum_{\substack{j'=1 \\ j'\neq i'}}^{N} \lambda_{i'j'}(x) + \sum_{\substack{j'=1 \\ j'\neq i}}^{j} \lambda_{ij'}(x)\Big).$$

For fixed x these are disjoint intervals, and the length of $\Delta_{ij}(x)$ is $\lambda_{ij}(x)$. Now we define a function $c : \mathbb{R}^n \times \mathbb{M} \times \mathbb{R} \to \mathbb{R}^N$:

$$c(x, e_i, z) = \begin{cases} e_j - e_i & \text{if } z \in \Delta_{ij}(x) \\ 0 & \text{otherwise.} \end{cases} \tag{4.2.1}$$

Then the $(\mathbb{R}^n \times \mathbb{M})$-valued switching diffusion process (4.1.1)-(4.1.2) can be represented as a solution of the following SDE (Ghosh *et al.*, 1993, 1997):

$$dX_t = a(X_t, \theta_t)dt + b(X_t, \theta_t)dW_t, \tag{4.2.2}$$

$$d\theta_t = \int_{\mathbb{R}} c(X_t, \theta_{t-}, z)p(dt, dz), \tag{4.2.3}$$

for $t \geq 0$, with (X_0, θ_0) a prescribed $(\mathbb{R}^n \times \mathbb{M})$-valued random variable; $p(dt, dz)$ is a Poisson random measure with intensity $dt \cdot dz$; (W_t) is an n-dimensional Wiener process independent of (X_0, θ_0) and $p(dt, dz)$.

Define the following interval:

$$U(x) \triangleq \bigcup_{i=1}^{N} \Big(\bigcup_{\substack{j=1 \\ j \neq i}}^{N} \Delta_{ij}(x) \Big),$$

it includes all intervals $\Delta_{ij}(x), i, j = 1, \ldots, N,\ i \neq j$. Since the length of each interval $\Delta_{ij}(x)$ is $\lambda_{ij}(x)$, and this is continuous and bounded function for $i, j = 1, \ldots, N, i \neq j$, it follows that the length of interval $U(x)$ (we denote it by $l(U(x))$) is bounded and is a continuous function of x. Therefore, it has a maximum at some point $x^\star$:

$$l(U(y)) \leq l(U(x^\star)) \text{ for all } y \in \mathbb{R}^n.$$

Let $U_{\max} \triangleq U(x^\star)$ denote the interval of maximum length and let $\lambda_{\max} \triangleq l(U_{\max})$ denote its length. Then, Equation (4.2.3) can be written as follows:

$$d\theta_t = \int_{U_{\max}} c(X_t, \theta_{t-}, z)p(dt, dz).$$

One can think of a Poisson random measure $p(dt, dz)$ as assigning unit mass to $(\tau_n,\ z_n)$ if there is a jump at time τ_n of size z_n. Let $N(t)$ be a standard Poisson process with intensity $\lambda_{\max}$. We denote by $\tau_n,\ n = 1, 2, \ldots$ the jump times of $N(t)$. Let $U_{\max}$ be the "mark" space, and $(Z_n)_{n \geq 1}$ be a sequence of i.i.d. random variables with uniform distribution on $U_{\max}$, independent of $N(t)$. In this special case we can represent the random Poisson measure $p(dt, dz)$ with intensity $dt \cdot dz$ as a random counting measure associated to the marked point process $(\tau_n, Z_n)_{n \geq 0}$, i.e. for each Lebesgue measurable $A \subset U_{\max}$

$$p((0, t], A) = \sum_{n \geq 1} \mathbf{1}_{\{\tau_n \leq t\}} \cdot \mathbf{1}_{\{Z_n \in A\}}. \tag{4.2.4}$$

We check that

$$\mathbb{E}[p((0, t], A)] = \lambda_{\max} \cdot t \cdot \mathbb{P}(Z_n \in A) = \lambda_{\max} \cdot t \cdot \frac{l(A)}{\lambda_{\max}} = t \cdot l(A).$$

The representation (4.2.4) is very convenient for practical problems. We see that $p(dt, dz)$ can be generated just by sampling independent random variables τ_n and $Z_n,\ n = 1, 2, \ldots$.

4.3 Strong approximation of switching diffusion

4.3.1 Discretization Scheme

Now we turn our attention to numerical solution of SDE (4.2.2)-(4.2.3). We develop an Euler type discretization scheme which allows to obtain a strong approximation process to the solution of switching diffusion precess. To start with, we should define the appropriate discretization of time interval $[0,T]$. Let us denote by $\mathscr{I}_d = \left\{t_n^d : n = 0,1,\dots,L\right\}$ the usual equidistant time discretization of a bounded interval $[0,T]$ with discretization step $h = T/L$. Suppose $\tau_1, \tau_2, \dots$ are the jump times of the discrete valued component θ_t. Then we take a new time discretization $\mathscr{I} = \{t_n : n = 0,1,\dots\}$ which is the union of the random jump times τ_n of the component θ_t on interval $[0,T]$ and the deterministic grid $\mathscr{I}_d$.

For a given time discretization $\mathscr{I}$ an Euler type approximation is a continuous time stochastic process $\{X_t^h, \theta_t^h\}$ satisfying the following equation with "delayed" coefficients[1]:

$$X_t^h = X_0 + \int_0^t a^h(s, X^h, \theta^h)ds + \int_0^t b^h(s, X^h, \theta^h)dW_s, \tag{4.3.1}$$

$$\theta_t^h = \theta_0 + \int_0^t \int_{U_{\max}} c^h(s-, X^h, \theta^h, z)p(ds, dz), \tag{4.3.2}$$

here

$$\begin{aligned} a^h(s, X^h, \theta^h) &\triangleq a(X_{t_k}, \theta_{t_k}), \quad s \in [t_k, t_{k+1}), \\ b^h(s, X^h, \theta^h) &\triangleq b(X_{t_k}, \theta_{t_k}), \quad s \in [t_k, t_{k+1}), \\ c^h(s-, X^h, \theta^h, z) &\triangleq c(X_{t_k}, \theta_{t_k-}, z), \quad s \in [t_k, t_{k+1}). \end{aligned}$$

The corresponding recursive discretization scheme

$$X_{t_i}^h = X_{t_{i-1}}^h + a(X_{t_{i-1}}^h, \theta_{t_{i-1}}^h)(t_i - t_{i-1}) + b(X_{t_{i-1}}^h, \theta_{t_{i-1}}^h)(W_{t_i} - W_{t_{i-1}}), \tag{4.3.3}$$

$$\theta_{t_i}^h = \theta_{t_{i-1}}^h + \int_{U_{\max}} c(X_{t_i}^h, \theta_{t_{i-1}}^h, z)p(\{t_i\}, dz), \tag{4.3.4}$$

determines values of the approximating process (4.3.1)-(4.3.2) at discretization times only. Thus, approximation $(X_{t_i}^h, \theta_{t_i}^h)$ is iteratively computed starting from the initial condition (X_0, θ_0) using the scheme (4.3.3)-(4.3.4). At the grid point, (4.3.4) computes the jump of θ^h exactly, conditional on $(X_{t_i}^h, \theta_{t_i^-}^h) = (X_{t_i}^h, \theta_{t_{i-1}}^h)$, if t_i is indeed a point of the Poisson random measure. Otherwise, the jump term is zero. The integral in (4.3.4) entails at most a single evaluation of the function c because $p(\{t_i\}, dz)$ is a point mass at the mark z that arrives at time t_i if it is a jump time.

[1]Here h denotes the dependence on the time discretization step h.

4.3.2 Convergence

Theorem 4.3.1. *Suppose, functions a, b, c and λ_{ij} are defined as in Section 4.2 and the Euler type approximating process $\{X_t^h, \theta_t^h\}$ is defined as in Section 4.3.1. We assume, that*

(i) functions $\lambda_{ij}(\cdot)$ $(i, j = 1, \dots N)$ are bounded and Lipschitz

$$|\lambda_{ij}(x) - \lambda_{ij}(y)| \le C_\lambda |x - y|, \text{ for all } x, y \in \mathbb{R}^n; \tag{4.3.5}$$

(ii) for all $x, y \in \mathbb{R}^n$ and $\theta, \eta \in \mathbb{M}$

$$|a(x,\theta) - a(y,\eta)|^2 + |b(x,\theta) - b(y,\eta)|^2 \le C_{ab}(|x-y|^2 + 1) \tag{4.3.6}$$

for $\theta \neq \eta$ and

$$|a(x,\theta) - a(y,\eta)|^2 + |b(x,\theta) - b(y,\eta)|^2 \le C_{ab}(|x-y|^2) \tag{4.3.7}$$

for $\theta = \eta$.

Then

$$\sup_{s \le T} \mathbb{E}(|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2) \le e^{-\lambda_{\max} T} K^2 \sum_{k=0}^{\infty} h^{2^{-k}} \cdot \frac{(\lambda_{\max} T K)^k}{k!}, \tag{4.3.8}$$

$$\sup_{s \le T} \mathbb{E}(|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2) \longrightarrow 0, \text{ as } h \longrightarrow 0, \tag{4.3.9}$$

and

$$\mathbb{E}[\sup_{s \le T} |X_s^h - X_s|] \le \Big(2T(T+4) C_{ab} \cdot e^{-\lambda_{\max} T} K^2 \sum_{k=0}^{\infty} h^{2^{-k}} \cdot \frac{(\lambda_{\max} T K)^k}{k!}\Big)^{1/2}, \tag{4.3.10}$$

$$\mathbb{E}[\sup_{s \le T} |X_s^h - X_s|] \longrightarrow 0, \quad \text{as } h \longrightarrow 0. \tag{4.3.11}$$

Here the constant K does not depend on h.

To prove the Theorem 4.3.1 one needs the following lemmas.

Let $\mathscr{F}_t$ denote the σ-algebra generated by all random variables $W_s, p((0,s], U_{\max})$, $s \le t$ (see Section 4.2). Let $\mathscr{L}_T^2$ denote the space of all $\mathscr{F}_t$-adapted stochastic processes that are square integrable:

$$||f||_{\mathscr{L}_T^2} = \int_0^T \int_\Omega f^2(t,\omega) \mathbb{P}(d\omega) dt < \infty.$$

Remark 4.3.2. In order to shorten expressions we introduce the following notation:

$$\mathbb{E}^i[\cdot] \triangleq \mathbb{E}[\cdot \mid N(T) = i].$$

Recall that in Section 4.2 by $N(t)$ we denoted the standard Poisson process with intensity λ_{max}. Event $\{N(T) = i\}$ means that there were i switches of component $\{\theta_t\}$ in the time interval $[0, T]$.

Lemma 4.3.3. *Suppose, W_t is independent of $p(dt, dz)$. Then for every $f \in \mathscr{L}_T^2$,*

$$\mathbb{E}^i\Big[\Big(\int_0^T f(t,\omega)dW_t(\omega)\Big)^2\Big] = \mathbb{E}^i\Big[\int_0^T f^2(t,\omega)dt\Big]. \tag{4.3.12}$$

Proof. First we consider the step processes, and then extend the result to arbitrary processes.

Let ϕ be a bounded step process in $\mathscr{L}_T^2$:

$$\phi(t,\omega) = \sum_{j=0}^{n-1} c_j(\omega) 1_{[t_j, t_{j+1})}(t). \tag{4.3.13}$$

By adaptedness, c_i in (4.3.13) is independent of $\Delta W_j \triangleq W_{t_{j+1}} - W_{t_j}$ for $i \le j$. Therefore

$$\begin{aligned}
\mathbb{E}^i\Big[\Big(\int_0^T \phi(t)dW_t\Big)^2\Big] &= \mathbb{E}^i\Big[\Big(\sum_{j=0}^{n-1} c_j \Delta W_j\Big)^2\Big] \\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \mathbb{E}^i[c_i c_j \Delta W_i \Delta W_j] \\
&= \sum_{j=0}^{n-1} \mathbb{E}^i[c_j^2 \Delta W_j^2] + 2\sum_{i<j} \mathbb{E}^i[c_i c_j \Delta W_i]\mathbb{E}[\Delta W_j] \\
&= \sum_{j=0}^{n-1} \mathbb{E}^i[c_j^2]\mathbb{E}[\Delta W_j^2] = \sum_{j=0}^{n-1} \mathbb{E}^i[c_j^2]\Delta t_j \\
&= \mathbb{E}^i\Big[\sum_{j=0}^{n-1} c_j^2 \Delta t_j\Big] = \mathbb{E}^i\Big[\int_0^T \phi^2(t)dt\Big].
\end{aligned}$$

To go from step processes to arbitrary processes we use the known fact that every process $\phi \in \mathscr{L}_T^2$ can be approximated arbitrarily well by step processes in $\mathscr{L}_T^2$ (e.g. see (Øksendal, 2002)). Now, suppose $\phi \in \mathscr{L}_T^2$ is an arbitrary process. We can approximate it by step processes $\phi^n \in \mathscr{L}_T^2$, i.e. $\phi^n \longrightarrow \phi$ in $\mathscr{L}_T^2$. To get the claim (4.3.12) we pass to the limit in the following equality, $n \longrightarrow \infty$

$$\mathbb{E}^i\Big[\int_0^T \phi_t^n dW_t\Big]^2 = \mathbb{E}^i\Big[\int_0^T (\phi_t^n)^2 dt\Big],\ n \in \mathbb{N}.$$

Indeed, since $Y^n \triangleq \int_0^T \phi_t^n dW_t \xrightarrow{L_2} Y \triangleq \int_0^T \phi_t dW_t$, then

$$\begin{aligned}\left|(\mathbb{E}^i[(Y^n)^2])^{1/2} - (\mathbb{E}^i[Y^2])^{1/2}\right| &\leq (\mathbb{E}^i[(Y^n - Y)^2])^{1/2} \\ &\leq \left(\frac{\mathbb{E}[(Y^n - Y)^2]}{\mathbb{P}(N(T) = i)}\right)^{1/2} \longrightarrow 0,\ n \longrightarrow \infty,\end{aligned}$$

and

$$\begin{aligned}\Big|\big(\mathbb{E}^i[\int_0^T (\phi_t^n)^2 dt]\big)^{1/2} - \big(\mathbb{E}^i[\int_0^T (\phi_t)^2 dt]\big)^{1/2}\Big| &\leq \Big(\mathbb{E}^i\big[\int_0^T (\phi_t^n - \phi_t)^2 dt\big]\Big)^{1/2} \\ &\leq \left(\frac{\mathbb{E}[\int_0^T (\phi_t^n - \phi_t)^2 dt]}{\mathbb{P}(N(T) = i)}\right)^{1/2} \longrightarrow 0,\ n \longrightarrow \infty,\end{aligned}$$

Thus

$$\mathbb{E}^i\Big[\int_0^T \phi_t dW_t\Big]^2 = \mathbb{E}^i\Big[\int_0^T (\phi_t)^2 dt\Big].$$

□

Lemma 4.3.4. *Suppose functions $\lambda_{ij}(\cdot)$ $i, j = 1, \ldots, N$ satisfy the conditions of Theorem 4.3.1. Then there exist a constant C_c such that*

$$\int_{\mathbb{R}} |c(x, e_i, z) - c(y, e_k, z)|^2 dz \leq C_c(|x - y| + 1) \text{ for } i \neq k \tag{4.3.14}$$

and

$$\int_{\mathbb{R}} |c(x, e_i, z) - c(y, e_k, z)|^2 dz \leq C_c(|x - y|) \text{ for } i = k \tag{4.3.15}$$

for all $x, y \in \mathbb{R}^n$ and $e_i, e_k \in \mathbb{M}$.

Proof.

$$\begin{aligned}&\int_{\mathbb{R}} |c(x, e_i, z) - c(y, e_k, z)|^2 dz \\ &\quad = \int_{\mathbb{R}} \Big| \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{1}_{\Delta_{ij}(x)}(z) \cdot e_j - e_i - \big(\sum_{\substack{j=1 \\ j \neq k}}^N \mathbf{1}_{\Delta_{kj}(y)}(z) \cdot e_j - e_k\big)\Big|^2 dz \\ &\quad = \int_{U_{\max}} \Big|(e_k - e_i) + \big(\sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{1}_{\Delta_{ij}(x)}(z) \cdot e_j - \sum_{\substack{j=1 \\ j \neq k}}^N \mathbf{1}_{\Delta_{kj}(y)}(z) \cdot e_j\big)\Big|^2 dz \\ &\quad \leq 2\lambda_{\max}|e_k - e_i|^2 + 2\int_{U_{\max}} \Big| \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{1}_{\Delta_{ij}(x)}(z) \cdot e_j - \sum_{\substack{j=1 \\ j \neq k}}^N \mathbf{1}_{\Delta_{kj}(y)}(z) \cdot e_j\Big|^2 dz.\end{aligned}$$

Let us consider two cases:
1) suppose $i = k$, then

$$\begin{aligned}
&\int_{\mathbb{R}} |c(x, e_i, z) - c(y, e_k, z)|^2 dz \\
&\qquad \leq 2 \int_{U_{\max}} \Big| \sum_{\substack{j=1 \\ j\neq i}}^{N} \mathbf{1}_{\Delta_{ij}(x)}(z) \cdot e_j - \sum_{\substack{j=1 \\ j\neq k}}^{N} \mathbf{1}_{\Delta_{kj}(y)}(z) \cdot e_j \Big|^2 dz \\
&\qquad = 2 \int_{U_{\max}} \Big| \sum_{\substack{j=1 \\ j\neq i}}^{N} (\mathbf{1}_{\Delta_{ij}(x)}(z) - \mathbf{1}_{\Delta_{ij}(y)}(z)) \cdot e_j \Big|^2 dz \\
&\qquad \leq 2N \int_{U_{\max}} \sum_{\substack{j=1 \\ j\neq i}}^{N} |\mathbf{1}_{\Delta_{ij}(x)}(z) - \mathbf{1}_{\Delta_{ij}(y)}(z)|^2 dz \\
&\qquad = 2N \sum_{\substack{j=1 \\ j\neq i}}^{N} \int_{U_{\max}} |\mathbf{1}_{\Delta_{ij}(x)}(z) - \mathbf{1}_{\Delta_{ij}(y)}(z)|^2 dz \\
&\qquad = 2N \sum_{\substack{j=1 \\ j\neq i}}^{N} \Big(\int_{\Delta_{ij}(x)\setminus\Delta_{ij}(y)} 1 dz + \int_{\Delta_{ij}(y)\setminus\Delta_{ij}(x)} 1 dz \Big). \qquad (4.3.16)
\end{aligned}$$

(1a) suppose $\Delta_{ij}(x) \cap \Delta_{ij}(y) \neq \varnothing$. Then

$$\begin{aligned}
&\int_{\Delta_{ij}(x)\setminus\Delta_{ij}(y)} 1 dz + \int_{\Delta_{ij}(y)\setminus\Delta_{ij}(x)} 1 dz \\
&\quad = \Big| \sum_{i'=1}^{i-1} \sum_{\substack{j'=1 \\ j'\neq i'}}^{N} \lambda_{i'j'}(x) + \sum_{\substack{j'=1 \\ j'\neq i}}^{j-1} \lambda_{ij'}(x) - \sum_{i'=1}^{i-1} \sum_{\substack{j'=1 \\ j'\neq i'}}^{N} \lambda_{i'j'}(y) - \sum_{\substack{j'=1 \\ j'\neq i}}^{j-1} \lambda_{ij'}(y) \Big| \\
&\quad + \Big| \sum_{i'=1}^{i-1} \sum_{\substack{j'=1 \\ j'\neq i'}}^{N} \lambda_{i'j'}(x) + \sum_{\substack{j'=1 \\ j'\neq i}}^{j} \lambda_{ij'}(x) - \sum_{i'=1}^{i-1} \sum_{\substack{j'=1 \\ j'\neq i'}}^{N} \lambda_{i'j'}(y) - \sum_{\substack{j'=1 \\ j'\neq i}}^{j} \lambda_{ij'}(y) \Big| \\
&\qquad\qquad \leq 2N^2 C_\lambda |x - y|.
\end{aligned}$$

(1b) now suppose $\Delta_{ij}(x) \cap \Delta_{ij}(y) = \varnothing$. We denote by $\Delta_{ij}^{x,y}$ the interval that is

contiguous to intervals $\Delta_{ij}(x)$ and $\Delta_{ij}(y)$. Then

$$\begin{aligned}\int_{\Delta_{ij}(x)} 1dz + \int_{\Delta_{ij}(y)} 1dz &\le \int_{\Delta_{ij}(x)\cup\Delta_{ij}^{x,y}} 1dz + \int_{\Delta_{ij}(y)\cup\Delta_{ij}^{x,y}} 1dz \\ &= \Big| \sum_{i'=1}^{i-1}\sum_{\substack{j'=1\\ j'\neq i'}}^{N} \lambda_{i'j'}(x) + \sum_{\substack{j'=1\\ j'\neq i}}^{j-1} \lambda_{ij'}(x) - \sum_{i'=1}^{i-1}\sum_{\substack{j'=1\\ j'\neq i'}}^{N} \lambda_{i'j'}(y) - \sum_{\substack{j'=1\\ j'\neq i}}^{j-1} \lambda_{ij'}(y)\Big| \\ &+ \Big| \sum_{i'=1}^{i-1}\sum_{\substack{j'=1\\ j'\neq i'}}^{N} \lambda_{i'j'}(x) + \sum_{\substack{j'=1\\ j'\neq i}}^{j} \lambda_{ij'}(x) - \sum_{i'=1}^{i-1}\sum_{\substack{j'=1\\ j'\neq i'}}^{N} \lambda_{i'j'}(y) - \sum_{\substack{j'=1\\ j'\neq i}}^{j} \lambda_{ij'}(y)\Big| \\ &\le 2N^2 C_\lambda |x-y|.\end{aligned}$$

Now we can proceed with expression (4.3.16):

$$\begin{aligned}2N \sum_{\substack{j=1\\ j\neq i}}^{N} \Big(\int_{\Delta_{ij}(x)\setminus\Delta_{ij}(y)} 1dz + \int_{\Delta_{ij}(y)\setminus\Delta_{ij}(x)} 1dz \Big) \\ \le 2N \sum_{\substack{j=1\\ j\neq i}}^{N} (2N^2 C_\lambda |x-y|) \le 4N^4 C_\lambda |x-y|.\end{aligned}$$

2) suppose $i \neq k$, then

$$\begin{aligned}\int_{\mathbb{R}} |c(x,e_i,z) - c(y,e_k,z)|^2 dz \\ \le 4\lambda_{\max} + 2\int_{U_{\max}} \Big| \sum_{\substack{j=1\\ j\neq i}}^{N} \mathbf{1}_{\Delta_{ij}(x)}(z)\cdot e_j - \sum_{\substack{j=1\\ j\neq k}}^{N} \mathbf{1}_{\Delta_{kj}(y)}(z)\cdot e_j \Big|^2 dz \\ \le 4\lambda_{\max} + 2\cdot 4N^2\lambda_{\max} \le \lambda_{\max}(4+8N^2) + |x-y|.\end{aligned}$$

From the above estimations follows that there exists a constant C_c such that

$$\int_{\mathbb{R}} |c(x,e_i,z) - c(y,e_k,z)|^2 dz \le C_c(|x-y|+1) \text{ for } i \neq k,$$

and

$$\int_{\mathbb{R}} |c(x,e_i,z) - c(y,e_k,z)|^2 dz \le C_c(|x-y|) \text{ for } i = k.$$

□

Lemma 4.3.5. *Suppose all conditions of Theorem 4.3.1 are satisfied. Then*

$$\sup_{s\le T} \mathbb{E}^i[|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2] \le K^{i+2} h^{2^{-i}},$$

as $h \longrightarrow 0$, for some constant K depending only on T.

Proof. Let estimate the difference of the coefficients. For all $h > 0$ take $s \in [t_k, t_{k+1})$, then applying conditions (4.3.6)-(4.3.7) and Lemma 4.3.4, and since $\theta_s = \theta_{t_k}$ for any $t_k \in I$, and $\theta_t^h, \theta_t \in \mathbb{M}$ (unit vectors), we obtain

$$\begin{aligned}
&|a^h(s, X^h, \theta^h) - a(X_s, \theta_s)|^2 + |b^h(s, X^h, \theta^h) - b(X_s, \theta_s)|^2 \\
&\quad = |a(X^h_{t_k}, \theta^h_{t_k}) - a(X_s, \theta_s)|^2 + |b(X^h_{t_k}, \theta^h_{t_k}) - b(X_s, \theta_s)|^2 \\
&\quad \le C_{ab}(|X^h_{t_k} - X_s|^2 + |\theta^h_{t_k} - \theta_s|^2) \\
&\quad \le 2C_{ab}(|X^h_{t_k} - X_{t_k}|^2 + |X_{t_k} - X_s|^2 + |\theta^h_{t_k} - \theta_{t_k}|^2)
\end{aligned}$$

$$\begin{aligned}
&\int_{\mathbb{R}} \left|c^h(t_k-, X^h, \theta^h, z) - c(X_{t_k}, \theta_{t_k-}, z)\right|^2 dz \\
&\quad = \int_{\mathbb{R}} \left|c(X^h_{t_k}, \theta^h_{t_k-}, z) - c(X_{t_k}, \theta_{t_k-}, z)\right|^2 dz \\
&\quad \le C_c(|X^h_{t_k} - X_{t_k}| + |\theta^h_{t_k-} - \theta_{t_k-}|^2).
\end{aligned}$$

Denote $C \triangleq \max(2C_{ab}, C_c)$. Taking conditional expectation given $N(T) = i$, we obtain

$$\begin{aligned}
&\mathbb{E}^i[|a^h(s, X^h, \theta^h) - a(X_s, \theta_s)|^2 + |b^h(s, X^h, \theta^h) - b(X_s, \theta_s)|^2] \\
&\quad \le C(\mathbb{E}^i[|X^h_{t_k} - X_{t_k}|^2] + \mathbb{E}^i[|X_{t_k} - X_s|^2] + \mathbb{E}^i[|\theta^h_{t_k} - \theta_{t_k}|^2]), \quad (4.3.17)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}^i\Big[\int_{\mathbb{R}} |c^h(t_k-, X^h, \theta^h, z) - c(X_{t_k}, \theta_{t_k-}, z)|^2 dz\Big] \\
&\quad \le C(\mathbb{E}^i[|X^h_{t_k} - X_{t_k}|] + \mathbb{E}^i[|\theta^h_{t_k-} - \theta_{t_k-}|^2]). \quad (4.3.18)
\end{aligned}$$

Denote

$$\begin{cases}
\varphi_i^h(t) \triangleq \sup\limits_{u \le t} \mathbb{E}^i[|X^h_u - X_u|^2], & t \in [0, T], \\
\psi_i(h) \triangleq \sup\limits_{|s-u| \le h} \mathbb{E}^i[|X_s - X_u|^2], & h > 0, \\
\gamma_i^h(t) \triangleq \sup\limits_{u \le t} \mathbb{E}^i[|\theta^h_u - \theta_u|^2], & t \in [0, T].
\end{cases} \quad (4.3.19)$$

Then, using (4.3.19) and the inequality (4.3.17) we have

$$\begin{aligned}
&\mathbb{E}^i[|a^h(s, X^h, \theta^h) - a(X_s, \theta_s)|^2 + |b^h(s, X^h, \theta^h) - b(X_s, \theta_s)|^2] \\
&\quad \le 2C(\varphi_i^h(s) + \psi_i(h) + \gamma_i^h(s)),
\end{aligned}$$

$s \in [0, T]$, $h > 0$.

Thus, for $t < \tau_{k+1}$, $k < i$ (i.e. less then $(k+1)$-th jump time)

$$\begin{aligned}
\varphi_i^h(t) &= \sup_{s\le t} \mathbb{E}^i[|X_s^h - X_s|^2] \\
&= \sup_{s\le t} \mathbb{E}^i\Big[\Big(\int_0^s (a^h(u,X^h,\theta^h) - a(X_u,\theta_u))du \\
&\qquad + \int_0^s (b^h(u,X^h,\theta^h) - b(X_u,\theta_u))dW_u\Big)^2\Big] \\
&\le \sup_{s\le t} \mathbb{E}^i\Big[2T\int_0^s (a^h(u,X^h,\theta^h) - a(X_u,\theta_u))^2 du \\
&\qquad + \int_0^s (b^h(u,X^h,\theta^h) - b(X_u,\theta_u))^2 du\Big] \\
&\le K_1 \int_0^t (\varphi_i^h(u) + \psi_i(h) + \gamma_i^h(u))du \\
&\le K_1 \int_0^t \varphi_i^h(u)du + K_1 T\big(\psi_i(h) + \gamma_i^h(\tau_k)\big)
\end{aligned}$$

here $K_1 = \max(1, 2T, 2C)$.
By Gronwall's lemma:

$$\varphi_i^h(t) \le K_1 T(\psi_i(h) + \gamma_i^h(\tau_k))e^{K_1 t} \le K_1 T(\psi_i(h) + \gamma_i^h(\tau_k))e^{K_1 T}, \text{ for } t < \tau_{k+1}.$$

Note, that

$$\begin{aligned}
\mathbb{E}^i[|X_s - X_u|^2] &= \mathbb{E}^i\Big[\Big|\int_u^s a(X_v,\theta_v)dv + \int_u^s b(X_v,\theta_v)dW_v\Big|^2\Big] \\
&\le 2(s-u)\int_u^s \mathbb{E}^i[|a(X_v,\theta_v)|^2]dv + 2\int_u^s \mathbb{E}^i[|b(X_v,\theta_v)|^2]dv \\
&\le K_2(s-u), \qquad 0 \le u \le s,
\end{aligned}$$

and thus

$$\psi_i(h) = \sup_{|s-u|\le h} \mathbb{E}^i[|X_s - X_u|^2] \le K_2 h, \ h \to 0.$$

From here

$$\varphi_i^h(t) \le K_1 T(K_2 h + \gamma_i^h(\tau_k))e^{K_1 T}, \tag{4.3.20}$$

for $t < \tau_{k+1}$, $1 \le k < i$, and using the fact that $(X_s^h - X_s) = (X_{s-}^h - X_{s-})$, (i.e. continuity from the left) we get

$$\varphi_i^h(\tau_{k+1}) \le K_1 T(K_2 h + \gamma_i^h(\tau_k))e^{K_1 T}, \ 1 \le k < i. \tag{4.3.21}$$

Denote

$$q(dt, dz) = p(dt, dz) - dtdz.$$

Now we will derive the similar recurrent formula for $\gamma_i^h(\tau_{k+1})$:

$$\begin{aligned}
\gamma_i^h(\tau_{k+1}) &= \sup_{s\le\tau_{k+1}} \mathbb{E}^i[|\theta_s^h-\theta_s|^2] \\
&= \sup_{s\le\tau_{k+1}} \mathbb{E}^i\Big[\Big|\int_0^s\int_{U_{\max}} (c^h(u-,X^h,\theta^h,z)-c(X_u,\theta_{u-},z))p(du,dz)\Big|^2\Big] \\
&= \sup_{s\le\tau_{k+1}} \mathbb{E}^i\Big[\Big|\int_0^s\int_{U_{\max}} (c^h(u-,X^h,\theta^h,z)-c(X_u,\theta_{u-},z))q(du,dz) \\
&\qquad + \int_0^s\int_{U_{\max}} (c^h(u-,X^h,\theta^h,z)-c(X_u,\theta_{u-},z))dudz\Big|^2\Big] \\
&\le \sup_{s\le\tau_{k+1}} 2\mathbb{E}^i\Big[\int_0^s\int_{U_{\max}} |c^h(u-,X^h,\theta^h,z)-c(X_u,\theta_{u-},z)|^2 dudz \\
&\qquad + (\int_0^s\int_{U_{\max}} (c^h(u-,X^h,\theta^h,z)-c(X_u,\theta_{u-},z))dudz)^2\Big] \\
&\le \sup_{s\le\tau_{k+1}} 2\mathbb{E}^i\Big[\int_0^s\int_{U_{\max}} |c^h(u-,X^h,\theta^h,z)-c(X_u,\theta_{u-},z)|^2 dudz \\
&\qquad + T\lambda_{\max}\int_0^s\int_{U_{\max}} |c^h(u-,X^h,\theta^h,z)-c(X_u,\theta_{u-},z)|^2 dudz\Big] \\
&\le 2C(1+T\lambda_{\max})\int_0^{\tau_{k+1}} \mathbb{E}^i[|X_u^h-X_u|+|\theta_{u-}^h-\theta_{u-}|^2]du \\
&\le 2CT(1+T\lambda_{\max})\big(\sqrt{\varphi_i^h(\tau_{k+1})}+\gamma_i^h(\tau_k)\big), \quad \text{for } 1\le k<i. \qquad (4.3.22)
\end{aligned}$$

Assume $\tau_0 \triangleq 0$. Then $\gamma_i^h(\tau_0)=0$ and $\varphi_i^h(\tau_0)=0$. Define

$$K_3 \triangleq 2e^{K_1T}K_1K_2CT(1+T\lambda_{\max}).$$

Then, using the recurrent formulas (4.3.20), (4.3.21) and (4.3.22) for $\varphi_i^h(\tau_{k+1})$ and $\gamma_i^h(\tau_{k+1})$ we obtain:

$$\begin{aligned}
&\varphi_i^h(\tau_1) \le 2K_3^2h && \gamma_i^h(\tau_1) \le 2K_3^2\sqrt{h} \\
&\varphi_i^h(\tau_2) \le 4K_3^3\sqrt{h} && \gamma_i^h(\tau_2) \le 4K_3^3h^{\frac{1}{4}} \\
&\qquad\cdots && \qquad\cdots \\
&\varphi_i^h(\tau_k) \le K_3(2K_3)^k h^{2^{1-k}} && \gamma_i^h(\tau_k) \le K_3(2K_3)^k h^{2^{-k}} \\
&\qquad\cdots && \qquad\cdots \\
&\varphi_i^h(T) \le K_3(2K_3)^{i+1} h^{2^{-i}} && \gamma_i^h(T) \le K_3(2K_3)^i h^{2^{-i}}
\end{aligned}$$

Denote $K = 2K_3$. From the above estimates follows that

$$\begin{aligned}\sup_{s\le T}\mathbb{E}^i[|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2] &\le \varphi_i^h(T) + \gamma_i^h(T)\\ &\le K_3(2K_3)^{i+1}h^{2^{-i}} + K_3(2K_3)^i h^{2^{-i}}\\ &\le K^{i+2}h^{2^{-i}},\ h \longrightarrow 0.\end{aligned}$$

□

Proof of Theorem 4.3.1. Using the results of Lemma 4.3.5 we have

$$\begin{aligned}&\sup_{s\le T}\mathbb{E}[|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2]\\ &\quad= \sup_{s\le T}\sum_{k=0}^{\infty}\mathbb{E}^k[|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2]\cdot\mathbb{P}(N(T)=k)\\ &\quad\le \sum_{k=0}^{\infty}\big(\sup_{s\le T}\mathbb{E}^k[|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2]\big)\cdot\mathbb{P}(N(T)=k)\\ &\quad\le e^{-\lambda_{\max}T}K^2\sum_{k=0}^{\infty}h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!}.\end{aligned}$$

Denote

$$S_m(h) \triangleq \sum_{k=0}^{m}h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!},$$

and

$$S(h) \triangleq \lim_{m\to\infty}S_m(h) = \sum_{k=0}^{\infty}h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!}.$$

Since

$$\left|h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!}\right| \le \frac{(\lambda_{\max}TK)^k}{k!},\ h\in[0,1]$$

and

$$\sum_{k=0}^{\infty}\frac{(\lambda_{\max}TK)^k}{k!} = e^{\lambda_{\max}TK} < \infty,$$

then, by Weierstrass M-Test, $S(h) < \infty$ and the convergence is uniform on $[0,1]$, furthermore, function $S(h)$ is continuous on $[0,1]$. Thus

$$\lim_{h\to 0}S(h) = \lim_{h\to 0}\lim_{m\to\infty}S_m(h) = \lim_{m\to\infty}\lim_{h\to 0}S_m(h) = 0.$$

Hence, we have proven that

$$\sup_{s\le T}\mathbb{E}[|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2] \le e^{-\lambda_{\max}T}K^2\sum_{k=0}^{\infty}h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!}, \quad (4.3.23)$$

and

$$\sup_{s\leq T}\mathbb{E}[|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2] \longrightarrow 0, \text{ as } h \longrightarrow 0.$$

Using Jensen's inequality we obtain

$$\mathbb{E}[\sup_{s\leq T}|X_s^h - X_s|] \leq \Big(\mathbb{E}[\sup_{s\leq T}|X_s^h - X_s|^2]\Big)^{1/2}. \tag{4.3.24}$$

Next, using Doob's maximal martingale inequality, conditions (4.3.6)-(4.3.7) and (4.3.23) we show that

$$\begin{aligned}
\mathbb{E}[\sup_{s\leq T}|X_s^h - X_s|^2] &= \mathbb{E}\Big[\sup_{s\leq T}\Big|\int_0^s (a^h(u,X^h,\theta^h) - a(X_u,\theta_u))du \\
&\qquad + \int_0^s (b^h(u,X^h,\theta^h) - b(X_u,\theta_u))dW_u\Big|^2\Big] \\
&\leq 2\mathbb{E}\Big[\sup_{s\leq T}\Big(T\int_0^s |a^h(u,X^h,\theta^h) - a(X_u,\theta_u)|^2du \\
&\qquad + \Big|\int_0^s (b^h(u,X^h,\theta^h) - b(X_u,\theta_u))dW_u\Big|^2\Big)\Big] \\
&\leq 2\mathbb{E}\Big[T\int_0^T |a^h(u,X^h,\theta^h) - a(X_u,\theta_u)|^2du \\
&\qquad + 4\Big|\int_0^T (b^h(u,X^h,\theta^h) - b(X_u,\theta_u))dW_u\Big|^2\Big] \\
&\leq 2(T+4)C_{ab}\mathbb{E}\Big[\int_0^T (|X_u^h - X_u|^2 + |\theta_u^h - \theta_u|^2)du\Big] \\
&\leq 2T(T+4)C_{ab}\sup_{s\leq T}\mathbb{E}[|X_s^h - X_s|^2 + |\theta_s^h - \theta_s|^2] \\
&\leq 2T(T+4)C_{ab}\cdot e^{-\lambda_{\max}T}K^2\sum_{k=0}^{\infty}h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathbb{E}[\sup_{s\leq T}|X_s^h - X_s|] \\
&\qquad \leq \Big(2T(T+4)C_{ab}\cdot e^{-\lambda_{\max}T}K^2\sum_{k=0}^{\infty}h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!}\Big)^{1/2},
\end{aligned} \tag{4.3.25}$$

and

$$\mathbb{E}[\sup_{s\leq T}|X_s^h - X_s|] \longrightarrow 0, \text{ as } h \longrightarrow 0. \tag{4.3.26}$$

□

Remark 4.3.6. Expression (4.3.26) ensures that approximating process $\{X_t^h\}$ strongly converges to $\{X_t\}$ as $h \to 0$, i.e. we have that for all $t \in [0,T]$ $\lim_{h\to 0} \mathbb{E}[|X_t^h - X_t|] = 0$ (see Definition 2.3.2).

4.4 Approximation of first passage times

Our aim is to show that first passage times of discretized switching diffusion converge in distribution to first passage times of original process, i.e.

$$P(\tau \le T) - P(\tau^h \le T) = \mathbb{E}[\mathbf{1}_{\{T<\tau^h\}}] - \mathbb{E}[\mathbf{1}_{\{T<\tau\}}] \longrightarrow 0, \quad h \to 0,$$

as the maximal discretization step h tends to zero. A similar problem was studied by Gobet (1999*a*,*b*), only in his case he was studying the approximations of expected value $E[g(x(\tau),\tau)]$ of a given function g depending on the Itô diffusion x and on its first exit time τ from a given domain. In our case $g(x(\tau),\tau) = \mathbf{1}_{\{T<\tau\}}$ and the evolution of x is more complex. Now it is a Euclidean valued component of the switching diffusion (4.1.1)-(4.1.2). Nevertheless, we still can follow the approach of Gobet (1999*a*,*b*), altering it where needed.

Theorem 4.4.1. *Suppose all conditions of Theorem 4.3.1 are satisfied. Assume that a and b are globally Lipschitz functions, and set $D = \mathbb{R}^n \backslash D^c$ is defined by $D^c = \{x \in \mathbb{R}^n : F(x) > 0\}$, $\partial D^c = \{x \in \mathbb{R}^n : F(x) = 0\}$ for some globally Lipschitz function F. Provided that the condition (C) below is satisfied*

$$\text{(C): } P(\exists t \in [0,T]\ X_t \notin D^c;\ \forall t \in [0,T]\ X_t \in D^c \cup \partial D^c) = 0,$$

then we have

$$\lim_{h\to 0} |P(\tau \le T) - P(\tau^h \le T)| = 0,$$

Remark 4.4.2. If D is of class C^2 with a compact boundary, the existence of such a function F holds.

Remark 4.4.3. Condition (C) rules out the pathological situation where with probability non-zero the paths may reach boundary ∂D^c without leaving $D^c \cup \partial D^c$. If this condition is not satisfied, then the approximation may not converge to the exact solution.

Remark 4.4.4. If a and b are bounded and D^c is of class C^3 with a compact boundary, then an uniform ellipticity condition on the diffusion implies condition (C).[2]

Proof. Obviously, we have

$$|P(\tau \le T) - P(\tau^h \le T)| = |P(T < \tau^h) - P(T < \tau)| \le \mathbb{E}[|\mathbf{1}_{\{T<\tau^h\}} - \mathbf{1}_{\{T<\tau\}}|].$$

[2]For more detail on this remarks see Gobet (1999*a*).

Fix $\delta > 0$. Then, elementary arguments lead to

$$\begin{aligned}|\mathbf{1}_{\{T<\tau^h\}} - \mathbf{1}_{\{T<\tau\}}| &= |\mathbf{1}_{\{\inf_{t\in[0,T]} F(X_t^h)>0\}} - \mathbf{1}_{\{\inf_{t\in[0,T]} F(X_t)>0\}}| \\ &\quad \times (\mathbf{1}_{\{|\inf_{t\in[0,T]} F(X_t)|<\delta\}} + \mathbf{1}_{\{|\inf_{t\in[0,T]} F(X_t)|\geq\delta\}}) \\ &\leq \mathbf{1}_{\{|\inf_{t\in[0,T]} F(X_t)|<\delta\}} + \mathbf{1}_{\{|\inf_{t\in[0,T]} F(X_t) - \inf_{t\in[0,T]} F(X_t^h)|\geq\delta\}}.\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbb{E}[|\mathbf{1}_{\{T<\tau^h\}} - \mathbf{1}_{\{T<\tau\}}|] \\ &\quad \leq P(|\inf_{t\in[0,T]} F(X_t)| < \delta) + P(|\inf_{t\in[0,T]} F(X_t) - \inf_{t\in[0,T]} F(X_t^h)| \geq \delta). \quad (4.4.1)\end{aligned}$$

Set

$$\delta \propto \Big(2T(T+4)C_{ab}\cdot e^{-\lambda_{\max}T}K^2\sum_{k=0}^{\infty} h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!}\Big)^{\frac{1}{4}}.$$

Then the first term in the r.h.s. of (4.4.1) converges to

$$P(\inf_{t\in[0,T]} F(X_t) = 0), \quad \text{as} \quad h \to 0,$$

which equals 0 using condition (C). The second one can be easily bounded by

$$\Big(2T(T+4)C_{ab}\cdot e^{-\lambda_{\max}T}K^2\sum_{k=0}^{\infty} h^{2^{-k}}\cdot\frac{(\lambda_{\max}TK)^k}{k!}\Big)^{\frac{1}{4}}.$$

using the Markov inequality and the estimate (4.3.10). This proves that $\lim_{h\longrightarrow+0}|P(\tau \leq T) - P(\tau^h \leq T)| = 0$. □

4.5 Concluding Remarks

The aim of this chapter was to study the problem of approximation of first passage times in stochastic hybrid processes with state dependent switching rates. We took a switching diffusion of Ghosh *et al.* (1997) as a non-trivial example of a stochastic hybrid process with state dependent switching rates; developed for it a strong Euler-type discretization scheme and proved its convergence. Then, following the approach of Gobet (1999*a*,*b*) we have shown that the first passage time of discretized switching diffusion converges in distribution to the first passage time of continuous time switching diffusion process, as the maximal discretization step h tends to zero. Using our Euler-type discretization scheme one can easily generate sample paths of switching diffusion and by taking a sample average of $\mathbf{1}_{\{\tau^h\leq T\}}$ one can obtain the estimate of the probability of reaching a given target set within the time interval $(0, T]$.

The results of this chapter will be important in the following chapter where we develop efficient Monte Carlo simulation techniques for estimation of probabilities of rare events in stochastic hybrid systems.

Chapter 5

Sequential Monte Carlo simulation of rare event probability in stochastic hybrid systems

5.1 Introduction

In Chapter 2 we have discussed the problem of rare event estimation in stochastic dynamical systems and came to a conclusion that the IPS based MC simulation approaches are very promising and suitable for this task. The numerical test in Chapter 2 has shown that the IPS approach of Cérou *et al.* (2002) works very well for a diffusion example. While in theory the IPS approach is applicable virtually to any strong Markov process, in practice the straightforward application of this approach to stochastic hybrid processes fails to produce reasonable estimates within a reasonable amount of simulation time. The aim of this chapter is to extend this IPS approach for estimation of the rare event probabilities in stochastic hybrid systems. We choose a switching diffusion with state dependent switching rates as a non-trivial example of stochastic hybrid system.

What are the reasons why the IPS approach may not work well for hybrid processes and how it can be improved? Suppose the stochastic hybrid system is modelled by a switching diffusion process and suppose that initial probabilities of starting in certain modes are very small. Then it is highly unlikely to draw particles in these "light" modes, and the bulk of the particles will be sampled in the "heavy" modes, i.e. the modes with higher probabilities. Subsequently, most, if not all, of the few particles in the "light" modes, most likely, will be lost after several resampling steps. We do not want this happening as it affects the

accuracy of estimate. To avoid this, at the initial sampling step we can sample a fixed number of particles for each mode separately and assign to each particle its "importance" weight in accordance with the initial probability. Note that introduction of weights will require instead of empirical measures in the form $\frac{1}{N_p}\sum_{i=1}^{N_p}\delta_{\{\xi_k^i\}}$ to use weighted empirical measures in the form $\sum_{i=1}^{N_p}\omega_k^i\delta_{\{\xi_k^i\}}$. In the classical IPS approach all particles actually have the same weights equal to $1/N_p$, where N_p is the number of particles. Now suppose at the initial sampling step, using the conditional sampling per mode, we sample a sufficient number of particles for "light" and for "heavy" modes and assign the weights. During the successive resampling steps the particles with larger weights have higher chances to be selected, and the particles with small weights in "light" modes tend to be discarded. Again, this can be avoided by resampling per mode and adjusting the weights appropriately.

The second problem is the problem of rare switches. If the probability of some mode transitions (switches) is very small, then it is highly unlikely to observe even one switch during a simulation run. In such cases, the possible switchings between the modes are not properly taken into account. Together with the problem of losing "light" particles this badly affects IPS estimator performance. By increasing the number of particles the IPS estimates should improve, but only at the cost of substantially increased simulation time. In order to avoid the need to increase the number of particles when the switching rates are decreasing we introduce a sequential importance switching technique.

The chapter is organized as follows. In Section 5.2 the IPS approach is formulated for a switching diffusion case and subsequently extended to Hybrid IPS algorithm which is designed to cope with large differences in mode probabilities and rare switchings. Numerical evaluations and comparison of different versions of the IPS algorithms are given in Section 5.3. Section 5.4 draws conclusions.

5.2 Rare event Monte Carlo simulation for switching diffusion

5.2.1 Factorization approach

Throughout this and the next sections, all stochastic processes are defined on a complete stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$ with $(\mathscr{F}_t)_{t\geq 0}$ a right continuous filtration.

Let $\{X_t, \theta_t\}$ be a switching diffusion taking its values in $\mathbb{R}^n \times \mathbb{M}$ according to

$$dX_t = a(X_t, \theta_t)dt + b(X_t, \theta_t)dW_t, \tag{5.2.1}$$

$$d\theta_t = \int_{\mathbb{R}} c(X_t, \theta_{t-}, z)p(dt, dz), \tag{5.2.2}$$

$\mathbb{M}$ is a finite set of modes and $(W_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^n$ independent of $\{\theta_t\}$ and of initial condition (X_0, θ_0), a prescribed $(\mathbb{R}^n \times \mathbb{M})-$valued random variable. The questions of existence and uniqueness and numerical approximations for the solutions of SDE (5.2.1)-(5.2.2) have been studied in Chapters 3 and 4 correspondingly.

Let us assume that switching diffusion (5.2.1)-(5.2.2) starts at $t = 0$ in a Borel set $\bar{D}_0 \subset \mathbb{R}^n \times \mathbb{M}$ with a known initial probability distribution $P_{X_0,\theta_0}(\cdot)$. We set $\tau_{\bar{D}} \stackrel{\triangle}{=} \inf\{t > 0 : (X_t, \theta_t) \in D \times \mathbb{M}\}$, i.e. the first passage time of $\{X_t\}$ to a closed connected Borel set D. The problem addressed in this chapter is to develop an efficient Monte Carlo simulation method for estimation of the probability $P_{hit}(0, T) \stackrel{\triangle}{=} P(\tau_{\bar{D}} \leq T)$, i.e. the probability that $\{X_t\}$ will hit the set D on the time interval $(0, T]$, $T < \infty$. Similarly as in (Cérou *et al.*, 2002) we will represent the probability of a rare event as a product of probabilities of intermediate "less rare" events and then use particle system approaches to estimate this product.

We assume a sequence of nested Borel sets, $\bar{D} = \bar{D}_m \subset \cdots \subset \bar{D}_1$ which are defined as follows:

$$\bar{D}_k \stackrel{\triangle}{=} D_k \times \mathbb{M}, \ k = 1, \ldots, m \tag{5.2.3}$$

where D_k is a closed Borel set of $\mathbb{R}^n$, and $\bar{D}_1$ such that $\bar{D}_1 \cap \bar{D}_0 = \varnothing$ (see Figure 5.2.1). The first moment that $\{X_t, \theta_t\}$ hits a set $\bar{D}_k$ is defined as the stopping

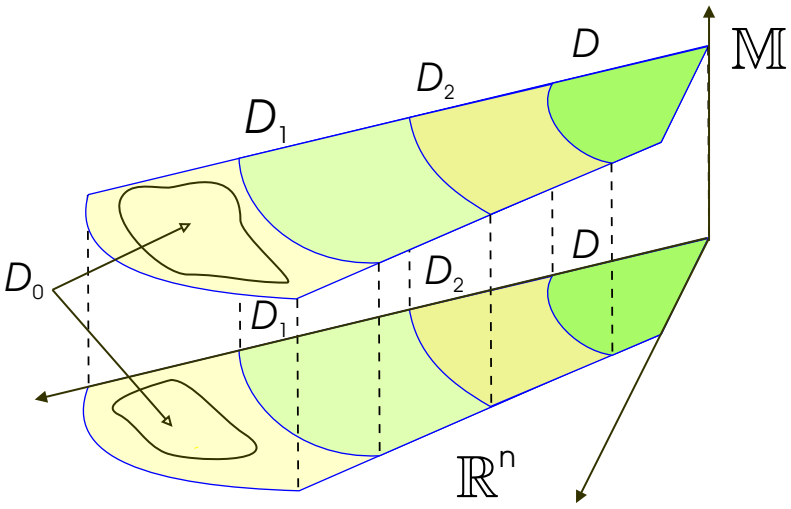


Figure 5.1: Schematic illustration of the hybrid space $\mathbb{R}^n \times \mathbb{M}$ and the nested sequence of level sets $\bar{D} = \bar{D}_m \subset \cdots \subset \bar{D}_1$ $(m = 3)$.

time:

$$\tau_k \stackrel{\Delta}{=} \inf\{t \geq 0 : (X_t, \theta_t) \in \bar{D}_k\} = \inf\{t \geq 0 : X_t \in D_k\},$$

$\tau_k = \infty$ if this set is empty. Note that $\tau_{\bar{D}} = \tau_m$. The process $\{X_t, \theta_t\}$, before hitting $\bar{D}$, passes through a sequence of nested sets (5.2.3). Following (Cérou *et al.*, 2002) we introduce the $\{0,1\}$-valued variables $\{y_k, k = 1, \ldots, m\}$ defined as follows:

$$y_k(\omega) \stackrel{\Delta}{=} \mathbf{1}_{\{\omega: X_{\tau_k \wedge T}(\omega) \in D_k\}} = \mathbf{1}_{\{\omega: \tau_k(\omega) \leq T\}}. \tag{5.2.4}$$

Hence, for each k we have

$$y_k(\omega) = \mathbf{1}_{\{\omega: \tau_k(\omega) \leq T\}} = \prod_{i=1}^{k} \mathbf{1}_{\{\omega:\ \tau_i(\omega) \leq T\}} = \prod_{i=1}^{k} y_i(\omega). \tag{5.2.5}$$

Next we characterize $P_{hit}(0,T)$ in terms of the sequence $\{y_k\}$. By its definition,

$$P_{hit}(0,T) = P(\tau_m \leq T) = \mathbb{E}[\mathbf{1}_{\{\tau_m \leq T\}}]$$

Subsequent substitution of (5.2.4) and (5.2.5) yields:

$$P_{hit}(0,T) = \mathbb{E}[y_m] = \mathbb{E}[\prod_{k=1}^{m} y_k]. \tag{5.2.6}$$

Since y_k assumes values from $\{0,1\}$,

$$\mathbb{E}[\prod_{k=1}^{m} y_k] = \prod_{k=1}^{m} \mathbb{E}[y_k | y_{k-1} = 1, \ldots, y_1 = 1]$$

Substituting this into (5.2.6) yields

$$\begin{aligned} P_{hit}(0,T) &= \prod_{k=1}^{m} \mathbb{E}[y_k | y_{k-1} = 1, \ldots, y_1 = 1] \\ &= \prod_{k=1}^{m} P(\tau_k \leq T | \tau_{k-1} \leq T, \ldots, \tau_1 \leq T) \\ &= \prod_{k=1}^{m} P(\tau_k \leq T | \tau_{k-1} \leq T) \end{aligned} \tag{5.2.7}$$

This means that (5.2.7) characterizes the probability $P_{hit}(0,T)$ of the rare event as a product of conditional probabilities of intermediate "less rare" events leading to it. Thus, if we define the conditional probabilities

$$\gamma_k \stackrel{\Delta}{=} P(\tau_k \leq T | \tau_{k-1} \leq T) \text{ for } k = 1, \ldots, m$$

and insert this in (5.2.7) then we get for $P_{hit}(0,T)$:

$$P_{hit}(0,T) = \prod_{k=1}^{m} \gamma_k \tag{5.2.8}$$

The estimation of the probabilities γ_k's is subsequently accomplished by the numerical approximation in terms of interacting particle system (IPS) (Cérou *et al.*, 2002).

5.2.2 IPS algorithm for switching diffusion

In this section we apply the IPS approach of Cérou *et al.* (2002) with some minor changes to the estimation of probability (5.2.8). We try to keep mathematical manipulations as simple as possible, thus we avoid complicated mathematical apparatus used in (Cérou *et al.*, 2002). However, the idea and the final result in principle remain the same.

Let us denote $E' = \mathbb{R}^n \times \mathbb{M}$, and let $\mathscr{E}'$ be the Borel σ-algebra of E'. Recall that the switching diffusion (5.2.1)-(5.2.2) starts in a Borel set $\bar{D}_0$ and before hitting the target set $\bar{D}$ it has to pass through a sequence of nested Borel sets (5.2.3). To capture how process $\{X_t, \theta_t\}$ enters each nested set $\bar{D} = \bar{D}_m \subset \cdots \subset \bar{D}_1$ before hitting finally the target set $\bar{D}$, we introduce the discrete time process $\{\xi_k, k = 0, 1, \ldots, m\}$ with values in space E' defined by $\xi_k \triangleq (X_{\tau_k \wedge T}, \theta_{\tau_k \wedge T})$. By the strong Markov property of $\{X_t, \theta_t\}$ (see Section 3.8), the process $\{\xi_k, k = 0, 1, \ldots, m\}$ is a Markov chain with transition kernel $\mathscr{Q}(\xi, d\xi') = P(\xi_k \in d\xi' | \xi_{k-1} = \xi)$. Observe also that

$$y_k(\omega) = \mathbf{1}_{\{\omega : X_{\tau_k \wedge T}(\omega) \in D_k\}} = \mathbf{1}_{\{\omega : \xi_k(\omega) \in \bar{D}_k\}} = \mathbf{1}_{\{\omega : \tau_k(\omega) \leq T\}}.$$

Now let us define the following conditional probabilities:

$$\pi_k(B) \triangleq P(\xi_k \in B | y_1 = 1, \ldots, y_k = 1), \tag{5.2.9}$$

$$p_k(B) \triangleq P(\xi_k \in B | y_1 = 1, \ldots, y_{k-1} = 1), \tag{5.2.10}$$

for any $B \in \mathscr{E}'$. It is easy to see that

$$p_k(B) = \int_{E'} \mathscr{Q}(\xi, B) \pi_{k-1}(d\xi) \text{ for all } B \in \mathscr{E}', \tag{5.2.11}$$

$$\pi_k(B) = \frac{\int_B \mathbf{1}_{\{\xi \in \bar{D}_k\}} p_k(d\xi)}{\int_{E'} \mathbf{1}_{\{\xi' \in \bar{D}_k\}} p_k(d\xi')} \text{ for all } B \in \mathscr{E}', \tag{5.2.12}$$

and

$$\gamma_k = \mathbb{E}[y_k | y_1 = 1, \ldots, y_{k-1} = 1] = \int_{E'} \mathbf{1}_{\{\xi \in \bar{D}_k\}} p_k(d\xi). \tag{5.2.13}$$

Thus, the evolution of the flow $\{\pi_k, p_k, \gamma_k; k = 0, 1, \ldots, m\}$ is described by the following diagram:

$$\pi_{k-1} \xrightarrow{\text{prediction}} p_k \xrightarrow{\text{conditioning}} \pi_k$$
$$\Big\downarrow$$
$$\gamma_k$$

with initial condition $\pi_0(d\xi) = P_{\xi_0}(d\xi) = P(\xi_0 \in d\xi)$. In that way, each of the m terms γ_k in (5.2.8) is characterized as a solution of a sequence of Equations (5.2.11)-(5.2.13).

The IPS approach is based on the idea to approximate the flow $\{\pi_k, p_k, \gamma_k; k = 0, 1, \ldots, m\}$ by an approximating sequence $\{\pi_k^{N_p}, p_k^{N_p}, \gamma_k^{N_p}; k = 0, 1, \ldots, m\}$ which is described by the following diagram

$$\pi_{k-1}^{N_p} \xrightarrow{\text{prediction}} p_k^{N_p} \xrightarrow{\text{conditioning}} \pi_k^{N_p}$$
$$\Big\downarrow$$
$$\gamma_k^{N_p}$$

with initial condition

$$\pi_0 \approx \pi_0^{N_p} = \sum_{i=1}^{N_p} \omega_0^i \delta_{\{\xi_0^i\}},$$

and approximations

$$p_k \approx p_k^{N_p} = \sum_{i=1}^{N_p} \omega_k^i \delta_{\{\xi_k^i\}},$$
$$\pi_k \approx \pi_k^{N_p} = \sum_{i=1}^{N_p} \frac{\mathbf{1}_{\{\xi_k^i \in \bar{D}_k\}} \omega_k^i}{\sum_{j=1}^{N_p} \omega_k^j \mathbf{1}_{\{\xi_k^j \in \bar{D}_k\}}} \delta_{\{\xi_k^i\}},$$
$$\gamma_k \approx \gamma_k^{N_p} = \sum_{i=1}^{N_p} \omega_k^i \mathbf{1}_{\{\xi_k^i \in \bar{D}_k\}},$$

in the form of the weighted empirical distributions associated with the particle system $\{t_k^i, \xi_k^i, \omega_k^i\}_{i=1}^{N_p}$, where N_p denotes the number of particles. Each particle is a triplet consisting of time index $t_k \triangleq \tau_k \wedge T$, state ξ_k and weight ω_k.

The approximation procedure goes as follows. At $t = 0$ we start with the empirical measure

$$\pi_0^{N_p} \triangleq \sum_{i=1}^{N_p} \omega_0^i \delta_{\{\xi_0^i\}},$$

where $\omega_0^i = 1/N_p$ are initial weights, and $\{\xi_0^i\}_{i=1}^{N_p}$ are independent samples from initial distribution $P_{\xi_0}(\cdot)$. See Figure 5.2(a). From $t = 0$ to τ_1, and from τ_{k-1}

to τ_k each particle evolves stochastically according to Equations (5.2.1)-(5.2.2) with initial condition (t^i_{k-1}, ξ^i_{k-1}) (prediction step) until it reaches the next level set $\bar{D}_k$ or the final time T. Let $\{\hat{t}^i_k, \hat{\xi}^i_k, \omega^i_k\}_{i=1}^{N_p}$ denote the values of the particles after the k-th prediction with $\omega^i_k = \omega^i_{k-1}$ (see Figure 5.2(b)). Then the empirical distribution $p_k^{N_p}$ associated with the predicted cloud of particles is:

$$p_k^{N_p} = \sum_{i=1}^{N_p} \omega^i_k \delta_{\{\hat{\xi}^i_k\}}.$$

The particles which do not reach the set $\bar{D}_k$ before time T are deleted, i.e. we set $\hat{\omega}^i_k = 0$ if $\hat{\xi}^i_k \notin \bar{D}_k$ and $\hat{t}^i_k = T$, else we set $\hat{\omega}^i_k = \omega^i_k$. See Figure 5.2(c). Then

$$\gamma_k \approx \gamma_k^{N_p} = \sum_{i=1}^{N_p} \hat{\omega}^i_k.$$

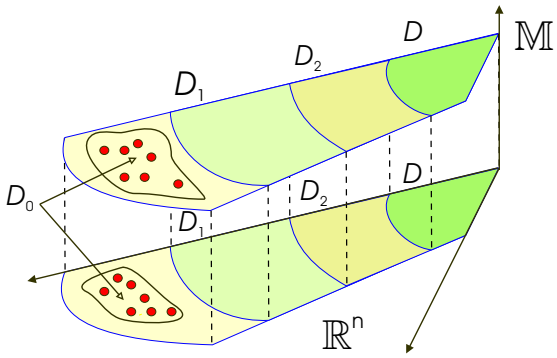
If all particles become deleted, i.e. $\gamma_k^{N_p} = 0$, then the algorithm is stopped and $P_{hit}(0,T) \approx 0$. The empirical distribution $\pi_k^{N_p}$ associated with the measurement updated cloud of particles is:

$$\pi_k^{N_p} = \sum_{i=1}^{N_p} \tilde{\omega}^i_k \delta_{\{\tilde{\xi}^i_k\}}. \tag{5.2.14}$$

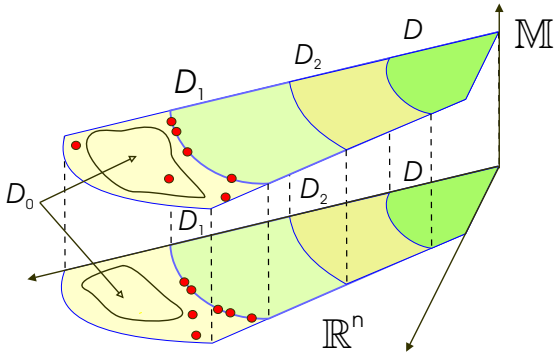
with

$$\tilde{\omega}^i_k = \frac{\hat{\omega}^i_k}{\sum_{j=1}^{N_p} \hat{\omega}^j_k}, \ \tilde{\xi}^i_k = \hat{\xi}^i_k, \quad i = 1, \ldots, N_p,$$

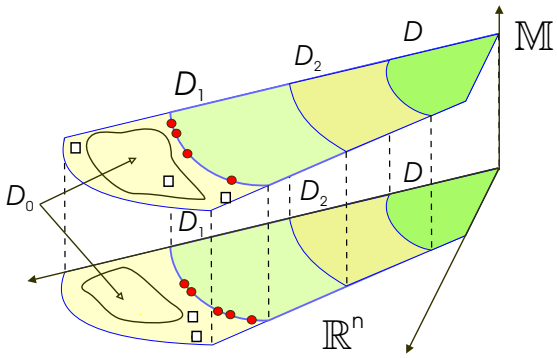
Particles having reached the set $\bar{D}_k$ are used for a resampling step. We resample with replacement N_p independent particles according to empirical measure (5.2.14). After this step we again have N_p particles $\{t^i_k, \xi^i_k, \omega^i_k\}_{i=1}^{N_p}$ at level $\bar{D}_k$. See Figure 5.3(a). Next, set $k := k+1$ and repeat the same simulation procedure iteratively until the target set $\bar{D}$ is reached (see Figures 5.3(b)-5.3(c)). For $k = m$ we have $P_{hit}(0,T) \approx \prod_{k=1}^{m} \gamma_k^{N_p}$.



(a) Step 1: Initial sampling

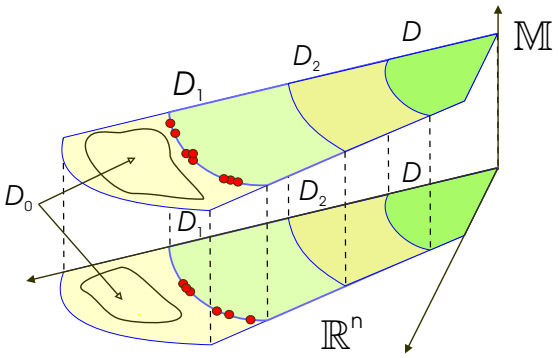


(b) Step 2: Prediction $\pi_{k-1}^{N_p} \rightarrow p_k^{N_p}$

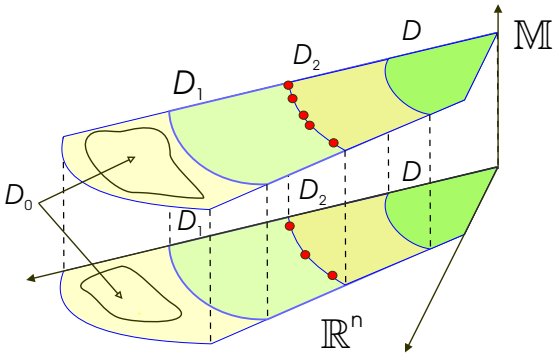


(c) Step 2: Evaluation of $\gamma_k^{N_p}$. The squares represent the particles with zero weight

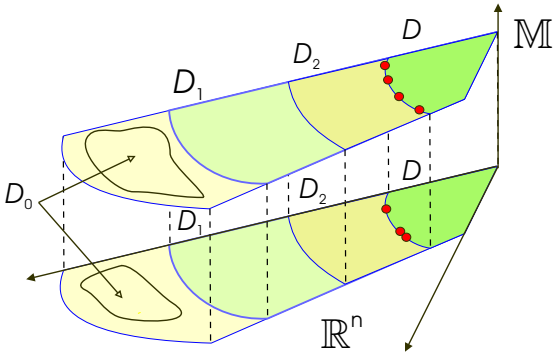
Figure 5.2: IPS algorithm. Steps 1 and 2



(a) Step 3: After resampling from $\pi_k^{N_p}$ we again obtain N_p particles at level $\bar{D}_k$



(b) Set $k := k+1$. Repeat steps 2 and 3. Particles move from one level set to the next



(c) Final stage: Particles reach the target set $\bar{D}$

Figure 5.3: IPS algorithm. Iterations of steps 2 and 3

In (Cérou *et al.*, 2002) it is proven that the particle estimates are unbiased, i.e.

$$\mathbb{E}\big[\prod_{k=1}^{m}\gamma_k^{N_p}\big] = P(\tau_m \le T) = P_{hit}(0,T)$$

and the rate of convergence is of order $\sqrt{N_p}$:

$$\Big(\mathbb{E}\Big[\big|\prod_{k=1}^{m}\gamma_k^{N_p} - \prod_{k=1}^{m}\gamma_k\big|^q\Big]\Big)^{\frac{1}{q}} \le \frac{a_q b_m}{\sqrt{N_p}},$$

finite constants a_q and b_m depend only on the parameters q and m respectively.

This means that by running the IPS algorithm many times (let us say N times) and by taking the sample average of random independent realizations of $Z \triangleq \prod_{k=1}^{m}\gamma_k^{N_p}$ we obtain an unbiased and consistent estimator of $P_{hit}(0,T)$:

$$P_{hit}(0,T) \approx \frac{1}{N}\sum_{i=1}^{N} Z^i,$$

it converges a.s. to $P_{hit}(0,T)$ as $N \to \infty$, by the strong law of large numbers.

Let us summarize the above and write down the algorithm step by step.

IPS algorithm for switching diffusion:

IPS Step 0. Level sets

- Choose appropriate nested sequence of closed subsets of $\mathbb{R}^n$: $D = D_m \subset D_{m-1} \subset \cdots \subset D_1$, and define $\bar{D}_k = D_k \times \mathbb{M}$, $k = 1, \ldots, m$.

IPS Step 1. Initial sampling; $k = 0$.

- For $i = 1, \ldots, N_p$ generate initial state value outside $\bar{D}_1$: $(X_0^i, \theta_0^i) \sim P_{X_0,\theta_0}(\cdot)$ and set $\xi_0^i = (X_0^i, \theta_0^i)$
- For $i = 1, \ldots, N_p$ set the initial weights and initial time indexes: $\omega_0^i = 1/N_p$, $t_0^i = 0$.
- Then

$$\pi_0^{N_p} = \sum_{i=1}^{N_p}\omega_0^i \delta_{\{\xi_0^i\}}.$$

Iteration k; $k = 1, \ldots, m$ over step 2 (prediction) and step 3 (resampling)

IPS Step 2. Prediction step: $\pi_{k-1} \longrightarrow p_k$;

- For $i = 1, \ldots, N_p$ simulate a new path (see Chapter 4) starting at ξ_{k-1}^i until the k-th set $\bar{D}_k$ is hit, or till time index $\hat{t}_k^i = T$.

- This yields new particles $\{\hat{t}_k^i, \hat{\xi}_k^i, \omega_k^i\}_{i=1}^{N_p}$. The weights are not changed: $\omega_k^i = \omega_{k-1}^i$.
- $p_k^{N_p}$ is the empirical distribution associated with the new cloud of particles:

$$p_k^{N_p} = \sum_{i=1}^{N_p} \omega_k^i \delta_{\{\hat{\xi}_k^i\}}.$$

- The particles which do not reach the set $\bar{D}_k$ are killed, i.e. we set $\hat{\omega}_k^i = 0$ if $\hat{t}_k^i = T$ and $\hat{\xi}_k^i \notin \bar{D}_k$, else $\hat{\omega}_k^i = \omega_k^i$.
- The new set of particles is $\{\hat{t}_k^i, \hat{\xi}_k^i, \hat{\omega}_k^i\}_{i=1}^{N_p}$.
- Approximation of γ_k:

$$\gamma_k \approx \gamma_k^{N_p} = \sum_{i=1}^{N_p} \hat{\omega}_k^i.$$

  If all particles are killed, i.e. $\gamma_k^{N_p} = 0$, then the algorithm stops and $P_{hit}(0,T) \approx 0$.

IPS Step 3. Resampling step: $p_k \longrightarrow \pi_k$

- For $i = 1, \ldots, N_p$ set $\tilde{\xi}_k^i = \hat{\xi}_k^i$ and

$$\tilde{\omega}_k^i = \frac{\hat{\omega}_k^i}{\sum_{j=1}^{N_p} \hat{\omega}_k^j}, \ i = 1, \ldots, N_p,$$

- Resample with replacement N_p particles ξ_k^i according to the empirical measure

$$\pi_k^{N_p} = \sum_{i=1}^{N_p} \tilde{\omega}_k^i \delta_{\{\tilde{\xi}_k^i\}}.$$

- The new set of particles is $\{t_k^i, \xi_k^i, \omega_k^i\}_{i=1}^{N_p}$, with weights $\omega_k^i = 1/N_p$.
- If $k < m$ then repeat steps 2, 3 for $k := k + 1$.
- Otherwise, stop with $P_{hit}(0,T) \approx \prod_{k=1}^{m} \gamma_k^{N_p}$.

5.2.3 Hybrid IPS

While in theory the IPS approach is applicable virtually to any strong Markov process, in practice the straightforward application of this approach to stochastic hybrid processes fails to produce reasonable estimates within a reasonable amount of simulation time. First, there will be few if any particles in modes with small

probabilities (i.e. "light" modes). This happens because after several resampling steps more and more "heavy" particles are being sampled from modes with higher probabilities, thus, "light" particles in the "light" modes become discarded. Second, if the switching rate is small then it is highly unlikely to observe even one switch during a simulation run. In such cases, the possible switching between modes is not properly taken into account. Together with the first problem this badly affects IPS estimator performance. By increasing the number of particles the IPS estimates should improve but only at the cost of substantially increased simulation time which makes the performance of IPS approach similar to one of the standard Monte Carlo.

In this section we propose two improvements over the IPS method: sampling per mode to cope with large differences in mode weights, and importance switching to cope with rare mode switching.

Sampling per Mode

We propose to replace the initial sampling step 1 and the resampling step 3 of the IPS algorithm by alternative steps which use conditional "sampling per mode".

If the initial probabilities of some particular modes are very small then it is highly unlikely to draw particles in these modes. To avoid this, at the initial sampling step we start with a fixed number of particles in each mode whatever small the initial probability is and adjust the weights appropriately. Let N_p^i denote number of particles in mode $e_i \in \mathbb{M} = \{e_1, \dots, e_N\}$. In total the system of particles will consist of $N_p = \sum_{i=1}^N N_p^i$ particles. Let J^i denote the set of indices of particles which are in mode e_i ($J^i \cap J^j = \varnothing$ for $i \neq j$). The whole set of indices is defined by

$$J \triangleq \bigcup_{i=1}^{N} J^i = \{1, 2, \dots, N_p\},\ |J| = N_p.$$

At the initial sampling step we take:

$$J^i = \Big\{ \sum_{k=0}^{i-1} N_p^k + 1, \dots, \sum_{k=0}^{i-1} N_p^k + N_p^i \Big\},\ N_p^0 = 0,\ |J^i| = N_p^i.$$

As particles evolve and switch from one mode to another one the numbers of particles in different modes will change, so will the index sets J^i's. But at each resampling step we will again sample N_p^i particles for each mode e_i from a conditional empirical distribution.

For $i = 1, \dots, N$ define the following probabilities:

$$\pi_k(B|e_i) \triangleq P(\xi_k \in B | y_1 = 1, \dots, y_k = 1, \theta_{\tau_k \wedge T} = e_i),$$
$$\pi_k(B, e_i) \triangleq \pi_k(B|e_i) \cdot P(\theta_{\tau_k \wedge T} = e_i | y_1 = 1, \dots, y_k = 1).$$

Then the following holds:

$$\pi_k(B) = \sum_{i=1}^{N} \pi_k(B, e_i).$$

Similarly for empirical approximation we have:

$$\pi_k^{N_p} = \sum_{j=1}^{N_p} \tilde{w}_k^j \delta_{\{\tilde{\xi}_k^j\}} = \sum_{i=1}^{N} \Big(\sum_{j \in J^i} \tilde{w}_k^j \delta_{\{\tilde{\xi}_k^j\}} \Big),$$

i.e.

$$\pi_k(B, e_i) \approx \pi_k^{i,N_p} \triangleq \sum_{j \in J^i} \tilde{w}_k^j \delta_{\{\tilde{\xi}_k^j\}}.$$

The idea behind the "sampling per mode" is to sample N_p^i particles for each mode e_i $(i = 1, \ldots, N)$ from the unnormalized conditional empirical measure $\pi_k^{i,N_p} = \sum_{j \in J^i} \tilde{w}_k^j \delta_{\{\tilde{\xi}_k^j\}}$ instead of sampling N_p particles from the empirical measure $\pi_k^{N_p} = \sum_{j=1}^{N_p} \tilde{w}_k^j \delta_{\{\tilde{\xi}_k^j\}}$. The weights of the particles in each mode e_i, i.e the particles with indices $j \in J^i$, are adjusted as follows:

$$\omega_k^j = \frac{\sum_{s \in J^i} \tilde{\omega}_k^s}{N_p^i}.$$

It may happen that $\sum_{j \in J^i} \tilde{\omega}_k^j = 0$, i.e. there are no particles left in mode e_i. But as long as there are particles left in other modes we still can continue the simulation. Below we present steps 1 and 3 of our new Hybrid IPS algorithm. In the following section we discuss how to improve step 2.

Step 1H. Initial sampling; $k = 0$.

- Choose for each mode $e_i \in \mathbb{M} = \{e_1, \ldots, e_N\}$, $i = 1, \ldots, N$ an integer N_p^i, so that $N_p = \sum_{i=1}^{N} N_p^i$.
- Set the initial index sets

$$J^i = \{ \sum_{k=0}^{i-1} N_p^k + 1, \ldots, \sum_{k=0}^{i-1} N_p^k + N_p^i \}, \; N_p^0 = 0.$$

- For each $i = 1, \ldots, N$ sample N_p^i initial state values outside $\bar{D}_1$:
  $X_0^j \sim P_{X_0|\theta_0}(\cdot|e_i)$. Set $\theta_0^j = e_i$, then $\xi_0^j = (X_0^j, \theta_0^j)$, $j \in J^i$.
  Assign initial weights and initial time indexes:
  $\omega_0^j = P_{\theta_0}(e_i)/N_p^i$, $t_0^j = 0$, $j \in J^i$.

- Then
$$\pi_0^{N_p} = \sum_{i=1}^{N_p} \omega_0^i \delta_{\{\xi_0^i\}}.$$

Step 3H. Resampling step: $p_k \longrightarrow \pi_k$

- For $i = 1, \ldots, N_p$ set $\tilde{\xi}_k^i = \hat{\xi}_k^i$ and
$$\tilde{\omega}_k^i = \frac{\hat{\omega}_k^i}{\sum_{s=1}^{N_p} \hat{\omega}_k^s}, \ i = 1, \ldots, N_p,$$

- For each mode $e_i \in \mathbb{M}$, $(i = 1, \ldots, N)$, resample with replacement N_p^i values of ξ_k from the unnormalized conditional empirical measure
$$\pi_k^{i,N_p} = \sum_{j \in J^i} \tilde{\omega}_k^j \delta_{\{\tilde{\xi}_k^j\}}$$
and adjust the weights as follows:
$$\omega_k^j = \frac{\sum_{s \in J^i} \tilde{\omega}_k^s}{N_p^i}.$$

- The new set of particles is $\{t_k^i, \xi_k^i, \omega_k^i\}_{i=1}^{N_p}$.
- If $k < m$ then repeat steps 2 and 3H for $k := k + 1$.
- Otherwise stop, with $P_{hit}(0, T) \approx \prod_{k=1}^{m} \gamma_k^{N_p}$.

Importance Switching

The possibility of small mode probabilities is covered well by resampling per mode. However there remains another problem to be solved; the problem of rare transitions (switches) of switching diffusion (5.2.1-5.2.2).

In the IPS approach during prediction step we generate random paths $(\xi_{k-1:k}^i)_{i=1}^{N_p} = (x^i_{\tau_{k-1}\wedge T:\tau_k \wedge T}, \theta^i_{\tau_{k-1}\wedge T:\tau_k\wedge T})_{i=1}^{N_p}$ to approximate the distribution p_k $(k = 1, \ldots, m)$. If the probability of some transitions (switches) is very small then, most probably, there will be few switches observed during the generation of these random paths. This may affect the accuracy of estimation of p_k and thus the accuracy of estimation of $P_{hit}(0, T)$ as well. To improve the quality of our estimates we will change the way the random trajectories are being generated. The idea is to use a kind of sequential importance sampling technique. We briefly explain the idea.

Let us consider a random trajectory $u_{t_0:t_N} = (u_{t_0}, \dots, u_{t_N})$ with the density

$$p_{u_{t_0},\dots,u_{t_N}}(u_0,\dots,u_N) = p_{u_{t_0}}(u_0)p_{u_{t_1}|u_{t_0}}(u_1|u_0)\dots p_{u_{t_N}|u_{t_0},\dots,u_{t_{N-1}}}(u_N|u_0,\dots,u_{N-1}).$$

The u_{t_k} is a state of some studied system at time t_k. The standard sequential sampling Monte Carlo works as follows: we pick $u_{t_0} = u_0$ according to the density $p_{u_{t_0}}$ and sequentially generate quantities $u_{t_k} = u_k$ according to the conditional density $p_{u_{t_k}|u_{t_0},\dots,u_{t_{k-1}}}$. In order to provoke more rare events during simulation we replace the original conditional densities $p_{u_{t_k}|u_{t_0},\dots,u_{t_{k-1}}}$ by some known conditional densities $q_{u_{t_k}|u_{t_0},\dots,u_{t_{k-1}}}$, $(k = 2,\dots N)$, with respect to which the rare events are more likely to occur. In such a way the new density of random trajectory $u_{t_0:t_N}$ is given by

$$q_{u_{t_0},\dots,u_{t_N}}(u_0,\dots,u_N) = q_{u_{t_0}}(u_0)q_{u_{t_1}|u_{t_0}}(u_1|u_0)\dots q_{u_{t_N}|u_{t_0},\dots,u_{t_{N-1}}}(u_N|u_0,\dots,u_{N-1}).$$

To compensate for this change (i.e. to keep the estimates unbiased) the importance weight for each random i-th realization of random path should be evaluated:

$$\omega_N^i = \frac{p_{u_{t_0},\dots,u_{t_N}}(u_0^i,\dots,u_N^i)}{q_{u_{t_0},\dots,u_{t_N}}(u_0^i,\dots,u_N^i)}.$$

This can be done recursively in time:

$$\omega_k^i = \omega_{k-1}^i \frac{p_{u_{t_k}|u_{t_0},\dots,u_{t_{k-1}}}(u_k^i|u_0^i,\dots,u_{k-1}^i)}{q_{u_{t_k}|u_{t_0},\dots,u_{t_{k-1}}}(u_k^i|u_0^i,\dots,u_{k-1}^i)}, \quad k = 1,\dots,N.$$

Now, we are going to apply the above sequential importance sampling technique to switching diffusion. But before this, we would like to remind that in order to conduct simulations the continuous time stochastic processes are commonly approximated by certain approximating continuous time processes which are easy to simulate (for example Euler or Milstein approximations) and then the values of the approximating process are determined recursively from the corresponding discretization schemes at discretization time points. In our case we use the Euler scheme (4.3.3-4.3.4). Therefore, all further derivations are done for the approximating stochastic process $\{X_t^h, \theta_t^h\}$ (see Chapter 4).

Remark 5.2.1. It is important to point out that on the time interval $[0,T]$ the process $\{\theta_t^h\}$ is a continuous time Markov chain with piecewise constant switching rates, i.e. on each individual interval $[t_{k-1}, t_k)$ (where t_k, $k = 0,1,\dots$ are the discretization points) it is the homogeneous continuous time Markov chain with constant switching rate matrix $\Lambda(X_{t_{k-1}}^h) = (\lambda_{ij}(X_{t_{k-1}}^h))_{i,j=1}^N$, and on each discretization interval $[t_{k-1}, t_k)$ evolution of $\{\theta_t^h\}$ is independent of $\{X_t^h\}$.

Functions $\lambda_{ij}(\cdot)$ define the rates of switchings. To make the rare switches more frequent we replace $\lambda_{ij}(\cdot)$ with functions $\hat{\lambda}_{ij}(\cdot)$ which provide higher switching rates. The changed process is denoted by $\{\hat{X}_t^h, \hat{\theta}_t^h\}$ and its values at discretization points are evaluated from the following discretization scheme:

$$\hat{X}_{t_i}^h = \hat{X}_{t_{i-1}}^h + a(\hat{X}_{t_{i-1}}^h, \hat{\theta}_{t_{i-1}}^h)(t_i - t_{i-1}) + b(\hat{X}_{t_{i-1}}^h, \hat{\theta}_{t_{i-1}}^h)(W_{t_i} - W_{t_{i-1}}), \quad (5.2.15)$$

$$\hat{\theta}_{t_i}^h = \hat{\theta}_{t_{i-1}}^h + \int_{U_{\max}} \hat{c}(\hat{X}_{t_i}^h, \hat{\theta}_{t_{i-1}}^h, z)p(\{t_i\}, dz), \quad (5.2.16)$$

where $t_i \in I$, $i = 1, 2, \ldots,$ and coefficient $\hat{c}$ corresponds to the new switching rate matrix $\hat{\Lambda}(\cdot) = (\hat{\lambda}_{ij}(\cdot))_{i,j}^N$ (see the definition (4.2.1) of coefficient c in Chapter 4). In order to keep the estimates $\{\pi_k^{N_p}, p_k^{N_p}, \gamma_k^{N_p}, k = 0, 1, \ldots, m\}$ unbiased, we should calculate the appropriate importance weights. We find these weights from the following derivations. For $n > k$, $n, k = 1, 2, \ldots$:

$$\begin{aligned}
&P_{X_{t_n}^h, \theta_{t_n}^h | X_{t_k}^h, \theta_{t_k}^h}(A, B | x_k, \theta_k) = \\
&= \sum_{\theta_n \in B} \int_A \sum_{\theta_{n-1} \in \mathbb{M}} \int_{\mathbb{R}^n} \cdots \sum_{\theta_{k+1} \in \mathbb{M}} \int_{\mathbb{R}^n} \prod_{i=k+1}^n P_{X_{t_i}^h, \theta_{t_i}^h | X_{t_{i-1}}^h, \theta_{t_{i-1}}^h}(dx_i, \theta_i | x_{i-1}, \theta_{i-1}) \\
&= \sum_{\theta_n \in B} \int_A \sum_{\theta_{n-1} \in \mathbb{M}} \int_{\mathbb{R}^n} \cdots \sum_{\theta_{k+1} \in \mathbb{M}} \int_{\mathbb{R}^n} \prod_{i=k+1}^n \frac{P_{X_{t_i}^h, \theta_{t_i}^h | X_{t_{i-1}}^h, \theta_{t_{i-1}}^h}(dx_i, \theta_i | x_{i-1}, \theta_{i-1})}{P_{\hat{X}_{t_i}^h, \hat{\theta}_{t_i}^h | \hat{X}_{t_{i-1}}^h, \hat{\theta}_{t_{i-1}}^h}(dx_i, \theta_i | x_{i-1}, \theta_{i-1})} \times \\
&\qquad \times P_{\hat{X}_{t_i}^h, \hat{\theta}_{t_i}^h | \hat{X}_{t_{i-1}}^h, \hat{\theta}_{t_{i-1}}^h}(dx_i, \theta_i | x_{i-1}, \theta_{i-1}) \\
&= \sum_{\theta_n \in B} \int_A \sum_{\theta_{n-1} \in \mathbb{M}} \int_{\mathbb{R}^n} \cdots \sum_{\theta_{k+1} \in \mathbb{M}} \int_{\mathbb{R}^n} \prod_{i=k+1}^n \frac{P_{X_{t_i}^h | \theta_{t_i}^h, X_{t_{i-1}}^h, \theta_{t_{i-1}}^h}(dx_i | \theta_i, x_{i-1}, \theta_{i-1})}{P_{\hat{X}_{t_i}^h | \hat{\theta}_{t_i}^h, \hat{X}_{t_{i-1}}^h, \hat{\theta}_{t_{i-1}}^h}(dx_i | \theta_i, x_{i-1}, \theta_{i-1})} \times \\
&\qquad \times \frac{P_{\theta_{t_i}^h | X_{t_{i-1}}^h, \theta_{t_{i-1}}^h}(\theta_i | x_{i-1}, \theta_{i-1})}{P_{\hat{\theta}_{t_i}^h | \hat{X}_{t_{i-1}}^h, \hat{\theta}_{t_{i-1}}^h}(\theta_i | x_{i-1}, \theta_{i-1})} P_{\hat{X}_{t_i}^h, \hat{\theta}_{t_i}^h | \hat{X}_{t_{i-1}}^h, \hat{\theta}_{t_{i-1}}^h}(dx_i, \theta_i | x_{i-1}, \theta_{i-1}) = \\
&= \sum_{\theta_n \in B} \int_A \sum_{\theta_{n-1} \in \mathbb{M}} \int_{\mathbb{R}^n} \cdots \sum_{\theta_{k+1} \in \mathbb{M}} \int_{\mathbb{R}^n} \prod_{i=k+1}^n L_{t_i | t_{i-1}}(\theta_i | x_{i-1}, \theta_{i-1}) \times \\
&\qquad \times P_{\hat{X}_{t_i}^h, \hat{\theta}_{t_i}^h | \hat{X}_{t_{i-1}}^h, \hat{\theta}_{t_{i-1}}^h}(dx_i, \theta_i | x_{i-1}, \theta_{i-1}),
\end{aligned} \quad (5.2.17)$$

where

$$L_{t|s}(\theta|x',\theta') \triangleq \frac{P_{\theta^h_t|X^h_s,\theta^h_s}(\theta|x',\theta')}{P_{\hat{\theta}^h_t|\hat{X}^h_s,\hat{\theta}^h_s}(\theta|x',\theta')} \tag{5.2.18}$$

denotes the likelihood ratio (i.e. the importance weights). The last equality in derivation (5.2.17) follows from Remark 5.2.1, i.e.

$$P_{X^h_{t_i}|\theta^h_{t_i},X^h_{t_{i-1}},\theta^h_{t_{i-1}}}(dx_i|\theta_i,x_{i-1},\theta_{i-1}) = P_{\hat{X}^h_{t_i}|\hat{\theta}^h_{t_i},\hat{X}^h_{t_{i-1}},\hat{\theta}^h_{t_{i-1}}}(dx_i|\theta_i,x_{i-1},\theta_{i-1})$$

for $i = 1, 2, \ldots$. Hence, during the prediction step the unbiased estimates of distribution p_k $(k = 1, \ldots, m)$ can be obtained by generating random trajectories of the process (5.2.15)-(5.2.16) (i.e. sampling according to $P_{\hat{x}_t,\hat{\theta}_t|\hat{x}_s,\hat{\theta}_s}(\cdot|x',\theta')$, $t > s$) and adjusting the weight of each particle recursively in time:

$$\omega^i_{t_j} = \omega^i_{t_{j-1}} \cdot L_{t_j|t_{j-1}}(\theta^i_j|x^i_{j-1},\theta^i_{j-1}). \tag{5.2.19}$$

The key is picking a good importance switching distribution $P_{\hat{X}^h_t,\hat{\theta}^h_t|\hat{X}^h_s,\hat{\theta}^h_s}(\cdot|x',\theta')$, i.e. the switching rate matrix $\hat{\Lambda}(\cdot)$. Roughly speaking, we should pick it so as to make the rare switch more likely to occur.

Before we present the complete version of our new Hybrid IPS (HIPS) algorithm, we need to explain how the likelihood ratios are evaluated.

Evaluation of likelihood ratios

In HIPS algorithm, for each discretization interval $[t_{j-1}, t_j]$ we need to evaluate the likelihood ratios (5.2.18), i.e. the conditional transition probabilities of continuous time pure jump components $\{\theta^h_t\}$ and $\{\hat{\theta}^h_t\}$. Recall that the infinitesimal jump rate matrices of $\{\theta^h_t\}$ and $\{\hat{\theta}^h_t\}$, i.e. $\Lambda = (\lambda_{ij}(X^h_t))^N_{i,j=1}$ and $\hat{\Lambda} = (\hat{\lambda}_{ij}(\hat{X}^h_t))^N_{i,j=1}$ depend on fixed state values X^h_t and $\hat{X}^h_t$ respectively, and are constant on each discretization interval. This simply means that one should be able to estimate the transition probability $P_{y_t|y_0}(j|i)$, of some homogeneous continuous time Markov chain $\{y_t\}$ with constant rate matrix $Q = (q_{ij})^N_{i,j=1}$. In practice this is usually done using Jensen's method, also known as uniformization (Jensen, 1953). The idea behind this method is to convert the continuous time Markov chain (CTMC) $\{y_t\}$ to a probabilistically equivalent discrete time Markov chain (DTMC) $\{y^*_k$, $k = 0, 1, \ldots\}$, with transition probability matrix $P^* = \frac{1}{r}Q + I$, (I is the unit matrix), and jump times defined by a Poisson counting process $\{N_t\}$, with rate r, such that $\{y_t\}$ and $\{y^*_{N_t}\}$ are probabilistically identical. Indeed, for the transition probability matrix $P_t = (P_{y_t|y_0}(j|i))^N_{i,j=1}$ the following holds

$$P_t = \exp(tQ) = \exp\big(rt(I + \frac{1}{r}Q)\big)e^{-rt} = \sum_{n=0}^{\infty} \frac{(rt)^n}{n!} e^{-rt}(P^*)^n,$$

hence we obtain the following expression for the (i,j)-entry

$$\begin{aligned} P_{y_t|y_0}(j|i) = P(y_t = j|y_0 = i) &= \sum_{n=0}^{\infty} \frac{(rt)^n}{n!} e^{-rt} ((P^*)^n)_{ij} \\ &= \sum_{n=0}^{\infty} P(y_n^* = j|y_0^* = i) P(N_t = n) \\ &= \sum_{n=0}^{\infty} P(y_n^* = j,\ N_t = n|y_0^* = i) \\ &= \sum_{n=0}^{\infty} P(y_{N_t}^* = j, N_t = n|y_0^* = i) \\ &= P(y_{N_t}^* = j|y_0^* = i), \end{aligned} \tag{5.2.20}$$

i.e. $y_t = y_{N_t}^*$, where $\{N_t\}$ is a Poisson counting process with rate r, independent of the discrete time Markov chain $\{y_n^*\}$. If $r > \max_i |q_{ii}|$, then the convergence of the above power series (5.2.20) is guaranteed. In practice, only finite number of terms are used to evaluate the sum in (5.2.20). Truncation to the right takes place since the Poisson probabilities are negligible beyond a certain n. Truncation to the left takes place because the Poisson probabilities are negligible until a certain n for a large values of rt. Therefore, we approximate this sum as (Fox and Glynn, 1988):

$$\sum_{n=0}^{\infty} \frac{(rt)^n}{n!} e^{-rt} ((P^*)^n)_{ij} \approx \sum_{n=L}^{R} ((P^*)^n)_{ij} f_p(n, rt), \tag{5.2.21}$$

where

$$f_p(n,s) = f_p(n-1,s)\frac{s}{n}, \quad \text{and} \quad f_p(0,s) = e^{-s},$$

L and R are left and right truncation points respectively. The truncation error is bounded by 10^{-d} (d digits of precision) if

$$L = \max_{k \in \mathbb{N}} \Big\{ \sum_{n=0}^{k} f_p(n, rt) \le \frac{10^{-d}}{2} \Big\}, \tag{5.2.22}$$

$$R = \min_{k \in \mathbb{N}} \Big\{ 1 - \sum_{n=L}^{k} f_p(n, rt) \le \frac{10^{-d}}{2} \Big\}. \tag{5.2.23}$$

In this way, on each discretization interval $[t_{j-1}, t_j]$ the likelihood ratio (5.2.18)

can be approximated as follows:

$$L_{t_j|t_{j-1}}(\theta_j|x_{j-1},\theta_{j-1}) = \frac{P_{\theta^h_{t_j}|X^h_{t_{j-1}},\theta^h_{t_{j-1}}}(\theta|x',\theta')}{P_{\hat{\theta}^h_{t_j}|\hat{X}^h_{t_{j-1}},\hat{\theta}^h_{t_{j-1}}}(\theta|x',\theta')}$$

$$\approx \frac{\sum_{n=L}^{R}((P^*)^n)_{ij} f_p(n,rt)}{\sum_{n=\hat{L}}^{\hat{R}}((\hat{P}^*)^n)_{ij} f_p(n,\hat{r}t)}, \tag{5.2.24}$$

where

$$P^* = \frac{1}{r}\Lambda + I, \qquad \hat{P}^* = \frac{1}{\hat{r}}\hat{\Lambda} + I,$$

$$r = \max_{\substack{i=1,\dots,N \\ x\in\mathbb{R}^n}} \Big|\sum_{\substack{j=1 \\ j\neq i}}^{N} \lambda_{ij}(x)\Big| + \varepsilon, \qquad \hat{r} = \max_{\substack{i=1,\dots,N \\ x\in\mathbb{R}^n}} \Big|\sum_{\substack{j=1 \\ j\neq i}}^{N} \hat{\lambda}_{ij}(x)\Big| + \varepsilon,$$

$\varepsilon > 0$ is a small constant and $\hat{L}$, $\hat{R}$ are defined as L, R in (5.2.22)-(5.2.23) with $\hat{r}$ used instead of r.

In HIPS algorithm the likelihood ratios (5.2.18) must be evaluated repeatedly during simulation. In order to avoid repeating the necessary arithmetic operations we can perform some computations before simulation and store the obtained data in memory for further use. We precompute the quantities L, R, $\hat{L}$, $\hat{R}$ and all terms of approximating series, i.e. $f_p(n,rt))$ and $f_p(n,\hat{r}t)$, and store these data in memory. Then, performing the simulation, these precomputed values are used, thus reducing computations to simple summations of numbers. However, the matrix powers $(P^*)^n$, $(\hat{P}^*)^n$ have to be evaluated each time again, because the transition probability matrices $P^* = \frac{1}{r}\Lambda + I$ and $\hat{P}^* = \frac{1}{r}\hat{\Lambda} + I$ depend on matrices Λ and $\hat{\Lambda}$ which are different after each discretization step h if the switching rates $(\lambda_{ij}(\cdot))_{ij=1}^N$ and $(\hat{\lambda}_{ij}(\cdot))_{ij=1}^N$ are not constant. In case if these switching rates are constant then the likelihood ratio (5.2.18) can be evaluated only once on each interval $[\tau_{k-1},\tau_k]$, thus significantly increasing the speed of the algorithm.

HIPS algorithm

Next we present two versions of our new HIPS algorithm.

The case where the jump rates are constant

The following HIPS algorithm is applicable to the case where the jump rates are constants, i.e. the discrete valued component $\{\theta_t^h\}$ is independent of $\{X_t^h\}$. The particle is defined as a triplet $\{t, \xi, \omega\}$ - (current time, state and weight). In this special case the likelihood ratios are evaluated not on discretization intervals $[t_j, t_{j+1}]$ but on stochastic intervals $[\tau_k, \tau_{k+1}]$.

HIPS Step 0. Initial setup

- Choose appropriate nested sequence of closed subsets D_i, $(i = 1, \ldots, m)$, of $\mathbb{R}^n$ such that $D = D_m \subset D_{m-1} \subset \cdots \subset D_1$, and define $\bar{D}_k = D_k \times \mathbb{M}$, $k = 1, \ldots, m$.

- Choose small $\varepsilon > 0$ and compute

$$r = \max_{i=1,\ldots,N} \Big| \sum_{\substack{j=1 \\ j \neq i}}^{N} \lambda_{ij} \Big| + \varepsilon, \qquad \hat{r} = \max_{i=1,\ldots,N} \Big| \sum_{\substack{j=1 \\ j \neq i}}^{N} \hat{\lambda}_{ij} \Big| + \varepsilon,$$

  here $\hat{\lambda}_{ij}$ are the jump rates of $\{\hat{\theta}_t^h\}$.

HIPS Step 1H. Initial sampling; $k = 0$.

- Choose for each mode $e_i \in \mathbb{M} = \{e_1, \ldots, e_N\}$, $i = 1, \ldots, N$ an integer N_p^i, so that $N_p = \sum_{i=1}^{N} N_p^i$.

- Set the initial index sets

$$J^i = \Big\{ \sum_{k=0}^{i-1} N_p^k + 1, \ldots, \sum_{k=0}^{i-1} N_p^k + N_p^i \Big\}, \; N_p^0 = 0.$$

- For each $i = 1, \ldots, N$ sample N_p^i initial state values outside $\bar{D}_1$:
  $X_0^j \sim P_{X_0|\theta_0}(\cdot|e_i)$. Set $\theta_0^j = e_i$, then $\xi_0^j = (X_0^j, \theta_0^j)$, $j \in J^i$.
  Assign initial weights and initial time indexes:
  $\omega_0^j = P_{\theta_0}(e_i)/N_p^i$, $t_0^j = 0$, $j \in J^i$.

- Then

$$\pi_0^{N_p} = \sum_{i=1}^{N_p} \omega_0^i \delta_{\{\xi_0^i\}}.$$

Iteration k; $k = 1, \ldots, m$ over step 2 (prediction) and step 3 (resampling)
HIPS Step 2H. Prediction step: $\pi_{k-1} \longrightarrow p_k$

- For $i = 1, \ldots, N_p$ use (5.2.15)-(5.2.16) to simulate a new path (see Chapter 4) starting at ξ_{k-1}^i until the k-th set $\bar{D}_k$ is hit, or till time index $\hat{t}_k^i = T$.
- This yields new particles $\{\hat{t}_k^i, \hat{\xi}_k^i, \omega_{k-1}^i\}_{i=1}^{N_p}$. The weights are not changed yet.
- Evaluate the likelihood ratios and adjust the weights:

  Find the truncation points

$$L = \max_{k\in\mathbb{N}} \Big\{ \sum_{n=0}^{k} f_p(n, r(\hat{t}_k^i - t_{k-1}^i)) \leq \frac{10^{-d}}{2} \Big\},$$

$$R = \min_{k\in\mathbb{N}} \Big\{ 1 - \sum_{n=L}^{k} f_p(n, r(\hat{t}_k^i - t_{k-1}^i)) \leq \frac{10^{-d}}{2} \Big\}$$

$$\hat{L} = \max_{k\in\mathbb{N}} \Big\{ \sum_{n=0}^{k} f_p(n, \hat{r}(\hat{t}_k^i - t_{k-1}^i)) \leq \frac{10^{-d}}{2} \Big\},$$

$$\hat{R} = \min_{k\in\mathbb{N}} \Big\{ 1 - \sum_{n=\hat{L}}^{k} f_p(n, \hat{r}(\hat{t}_k^i - t_{k-1}^i)) \leq \frac{10^{-d}}{2} \Big\}.$$

  Evaluate the likelihood ratio

$$L_{\hat{t}_k|t_{k-1}}(\theta_k^i|\theta_{k-1}^i) \approx \frac{\sum_{n=L}^{R}((P^*)^n)_{\theta_{k-1}^i\theta_k^i} f_p(n, r(\hat{t}_k^i - t_{k-1}^i))}{\sum_{n=\hat{L}}^{\hat{R}}((\hat{P}^*)^n)_{\theta_{k-1}^i\theta_k^i} f_p(n, \hat{r}(\hat{t}_k^i - t_{k-1}^i))}$$

  here $P^* = \frac{1}{r}\Lambda + I$, $\hat{P}^* = \frac{1}{\hat{r}}\hat{\Lambda} + I$, and $\Lambda = (\lambda_{ij})_{i,j=1}^N$, $\hat{\Lambda} = (\hat{\lambda}_{ij})_{i,j=1}^N$ are the corresponding infinitesimal jump rate matrices of $\{\theta_t^h\}$ and $\{\hat{\theta}_t^h\}$.

  Update the weights:

$$\omega_k^i = \omega_{k-1}^i \cdot L_{\hat{t}_k|t_{k-1}}(\theta_k^i|\theta_{k-1}^i).$$

- $p_k^{N_p}$ is the empirical distribution associated with the new cloud of particles:

$$p_k^{N_p} = \sum_{i=1}^{N_p} \omega_k^i \delta_{\{\hat{\xi}_k^i\}}.$$

- The particles which do not reach the set $\bar{D}_k$ are killed, i.e. we set $\hat{\omega}_k^i = 0$ if $\hat{t}_k^i = T$ and $\hat{\xi}_k^i \notin \bar{D}_k$, else $\hat{\omega}_k^i = \omega_k^i$.

- The new set of particles is $\{\hat{t}_k^i, \hat{\xi}_k^i, \hat{\omega}_k^i\}_{i=1}^{N_p}$.
- Approximation of γ_k:

$$\gamma_k \approx \gamma_k^{N_p} = \sum_{i=1}^{N_p} \hat{\omega}_k^i.$$

If all particles are killed, i.e. $\gamma_k^{N_p} = 0$, then the algorithm stops and $P_{hit}(0,T) \approx 0$.

HIPS Step 3H. Resampling step: $p_k \longrightarrow \pi_k$

- For $i = 1, \ldots, N_p$ set $\tilde{\xi}_k^i = \hat{\xi}_k^i$ and

$$\tilde{\omega}_k^i = \frac{\hat{\omega}_k^i}{\sum_{s=1}^{N_p} \hat{\omega}_k^s}, \ i = 1, \ldots, N_p,$$

- For each mode $e_i \in \mathbb{M}$, $(i = 1, \ldots, N)$, resample with replacement N_p^i values of ξ_k from the unnormalized conditional empirical measure

$$\pi_k^{i,N_p} = \sum_{j \in J^i} \tilde{\omega}_k^j \delta_{\{\tilde{\xi}_k^j\}}$$

and adjust the weights as follows:

$$\omega_k^j = \frac{\sum_{s \in J^i} \tilde{\omega}_k^s}{N_p^i}.$$

- The new set of particles is $\{t_k^i, \xi_k^i, \omega_k^i\}_{i=1}^{N_p}$.
- If $k < m$ then repeat steps 2 and 3 for $k := k + 1$.
- Otherwise stop, with $P_{hit}(0,T) \approx \prod_{k=1}^{m} \gamma_k^{N_p}$.

The case where the jump rates are functions of state

The following HIPS algorithm is applicable to the case where the switching rates of discrete valued component may depend on state.

HIPS Step 0. Initial setup

- Choose appropriate nested sequence of closed subsets D_i, $(i = 1, \ldots, m)$, of $\mathbb{R}^n$ such that $D = D_m \subset D_{m-1} \subset \cdots \subset D_1$, and define $\bar{D}_k = D_k \times \mathbb{M}$, $k = 1, \ldots, m$.

- Choose small $\varepsilon > 0$ and compute

$$r = \max_{\substack{i=1,\dots,N \\ x\in\mathbb{R}^n}} \big|\sum_{\substack{j=1 \\ j\neq i}}^{N} \lambda_{ij}\big| + \varepsilon, \qquad \hat{r} = \max_{\substack{i=1,\dots,N \\ x\in\mathbb{R}^n}} \big|\sum_{\substack{j=1 \\ j\neq i}}^{N} \hat{\lambda}_{ij}\big| + \varepsilon,$$

  here $\hat{\lambda}_{ij}$ are the jump rates of $\{\hat{\theta}^h_t\}$.

- Precompute the truncation points

$$L = \max_{k\in\mathbb{N}} \Big\{ \sum_{n=0}^{k} f_p(n, rh) \leq \frac{10^{-d}}{2} \Big\},$$
$$R = \min_{k\in\mathbb{N}} \Big\{ 1 - \sum_{n=L}^{k} f_p(n, rh) \leq \frac{10^{-d}}{2} \Big\}.$$

$$\hat{L} = \max_{k\in\mathbb{N}} \Big\{ \sum_{n=0}^{k} f_p(n, \hat{r}h) \leq \frac{10^{-d}}{2} \Big\},$$
$$\hat{R} = \min_{k\in\mathbb{N}} \Big\{ 1 - \sum_{n=\hat{L}}^{k} f_p(n, \hat{r}h) \leq \frac{10^{-d}}{2} \Big\}.$$

  and the following quantities

$$f_p(n, rh) = f_p(n-1, rh)\frac{rh}{n}, \quad \text{and} \quad f_p(0, rh) = e^{-rh}.$$

$$f_p(n, \hat{r}h) = f_p(n-1, \hat{r}h)\frac{\hat{r}h}{n}, \quad \text{and} \quad f_p(0, \hat{r}h) = e^{-\hat{r}h}.$$

- Then store the following data in memory for further use in the subsequent steps: $r, \hat{r}, L, R, \hat{L}, \hat{R}$, $f_p(n, rh)$ and $f_p(n, \hat{r}h)$.

HIPS Step 1H. Initial sampling: $k = 0$

- Choose for each mode $e_i \in \mathbb{M} = \{e_1, \dots, e_N\}$, $i = 1, \dots, N$ an integer N_p^i, so that $N_p = \sum_{i=1}^{N} N_p^i$.

- Set the initial index sets

$$J^i = \{ \sum_{k=0}^{i-1} N_p^k + 1, \dots, \sum_{k=0}^{i-1} N_p^k + N_p^i \}, \; N_p^0 = 0.$$

- For each $i = 1, \dots, N$ sample N_p^i initial state values outside $\bar{D}_1$: $X_0^j \sim P_{X_0|\theta_0}(\cdot|e_i)$. Set $\theta_0^j = e_i$, then $\xi_0^j = (X_0^j, \theta_0^j)$, $j \in J^i$. Assign initial weights and initial time indexes: $\omega_0^j = P_{\theta_0}(e_i)/N_p^i$, $t_0^j = 0$, $j \in J^i$.

- Then

$$\pi_0^{N_p} = \sum_{i=1}^{N_p} \omega_0^i \delta_{\{\xi_0^i\}}.$$

Iteration k; $k = 1, \dots, m$ over step 2 (prediction) and step 3 (resampling)

HIPS Step 2H. Prediction step:[1] $\pi_{k-1} \longrightarrow p_k$

- For $i = 1, \dots, N_p$ use (5.2.15)-(5.2.16) to simulate a new path (see Chapter 4) starting at ξ_{k-1}^i until the k-th set $\bar{D}_k$ is hit, or till time index $\hat{t}_k^i = T$.

  Path simulation is done iteratively per time discretization interval $[\tilde{t}_j, \tilde{t}_{j-1}]$ by performing the following substeps[2]:

  Compute the matrices $P^* = \frac{1}{r}\Lambda + I$ and $\hat{P}^* = \frac{1}{\hat{r}}\hat{\Lambda} + I$, where $\Lambda = (\lambda_{i'j'}(X_{j-1}^i))_{i',j'=1}^N$ and $\hat{\Lambda} = (\hat{r}_{i'j'}(X_{j-1}^i))_{i',j'=1}^N$ are the corresponding infinitesimal jump rate matrices of processes $\{\theta_t^h\}$ and$\{\hat{\theta}_t^h\}$.

  If $\tilde{t}_j - \tilde{t}_{j-1} = h$, i.e. $\tilde{t}_j$ and $\tilde{t}_{j-1}$ are fixed equidistant deterministic discretization points with distance h, then using the precomputed data evaluate the likelihood ratio:

$$L_{\tilde{t}_j|\tilde{t}_{j-1}}(\theta_j^i|X_{j-1}^i, \theta_{j-1}^i) \approx \frac{\sum_{n=L}^{R}((P^*)^n)_{\theta_{j-1}^i\theta_j^i} f_p(n, rh)}{\sum_{n=\hat{L}}^{\hat{R}}((\hat{P}^*)^n)_{\theta_{j-1}^i\theta_j^i} f_p(n, \hat{r}h)}$$

  If $\tilde{t}_j - \tilde{t}_{j-1} < h$, i.e. one of these discretization points is a random switch time, then we can not use the precomputed data, thus we perform

[1] each prediction step consists of a sequence of local subsets; one local subset for each discretization interval $[\tilde{t}_j, \tilde{t}_{j-1}]$, here $j = 1, 2, \dots$

[2] to simplify notations we use: $X_j^i \triangleq X_{\tilde{t}_j}^i$ and $\theta_j^i \triangleq \theta_{\tilde{t}_j}^i$

the computation of likelihood ratio in full:

$$L = \max_{k\in\mathbb{N}} \Big\{ \sum_{n=0}^{k} f_p(n, r(\tilde{t}_j - \tilde{t}_{j-1})) \leq \frac{10^{-d}}{2} \Big\},$$

$$R = \min_{k\in\mathbb{N}} \Big\{ 1 - \sum_{n=L}^{k} f_p(n, r(\tilde{t}_j - \tilde{t}_{j-1}) \leq \frac{10^{-d}}{2} \Big\},$$

$$\hat{L} = \max_{k\in\mathbb{N}} \Big\{ \sum_{n=0}^{k} f_p(n, \hat{r}(\tilde{t}_j - \tilde{t}_{j-1}) \leq \frac{10^{-d}}{2} \Big\},$$

$$\hat{R} = \min_{k\in\mathbb{N}} \Big\{ 1 - \sum_{n=\hat{L}}^{k} f_p(n, \hat{r}(\tilde{t}_j - \tilde{t}_{j-1}) \leq \frac{10^{-d}}{2} \Big\},$$

$$L_{\tilde{t}_j|\tilde{t}_{j-1}}(\theta_j^i | X_{j-1}^i, \theta_{j-1}^i) \approx \frac{\sum_{n=L}^{R} ((P^*)^n)_{\theta_{j-1}^i \theta_j^i} f_p(n, r(\tilde{t}_j - \tilde{t}_{j-1}))}{\sum_{n=\hat{L}}^{\hat{R}} ((\hat{P}^*)^n)_{\theta_{j-1}^i \theta_j^i} f_p(n, \hat{r}(\tilde{t}_j - \tilde{t}_{j-1}))}.$$

Update the weights:

$$\omega_{\tilde{t}_j}^i = \omega_{\tilde{t}_{j-1}}^i \cdot L_{\tilde{t}_j|\tilde{t}_{j-1}}(\theta_j^i | X_{j-1}^i, \theta_{j-1}^i).$$

- This yields new particles $\{\hat{t}_k^i, \hat{\xi}_k^i, \omega_k^i\}_{i=1}^{N_p}$.
- $p_k^{N_p}$ is the empirical distribution associated with the new cloud of particles:

$$p_k^{N_p} = \sum_{i=1}^{N_p} \omega_k^i \delta_{\{\hat{\xi}_k^i\}}.$$

- The particles which do not reach the set $\bar{D}_k$ are killed, i.e. we set $\hat{\omega}_k^i = 0$ if $\hat{t}_k^i = T$ and $\hat{\xi}_k^i \notin \bar{D}_k$, else $\hat{\omega}_k^i = \omega_k^i$.
- The new set of particles is $\{\hat{t}_k^i, \hat{\xi}_k^i, \hat{\omega}_k^i\}_{i=1}^{N_p}$.
- Approximation of γ_k:

$$\gamma_k \approx \gamma_k^{N_p} = \sum_{i=1}^{N_p} \hat{\omega}_k^i.$$

If all particles are killed, i.e. $\gamma_k^{N_p} = 0$, then the algorithm stops and $P_{hit}(0,T) \approx 0$.

HIPS Step 3H. Resampling step: $p_k \longrightarrow \pi_k$

- For $i = 1, \ldots, N_p$ set $\tilde{\xi}_k^i = \hat{\xi}_k^i$ and

$$\tilde{\omega}_k^i = \frac{\hat{\omega}_k^i}{\sum_{s=1}^{N_p} \hat{\omega}_k^s},\ i = 1, \ldots, N_p,$$

- For each mode $e_i \in \mathbb{M}$, $(i = 1, \ldots, N)$, resample with replacement N_p^i values of ξ_k from the unnormalized conditional empirical measure

$$\pi_k^{i,N_p} = \sum_{j \in J^i} \tilde{\omega}_k^j \delta_{\{\tilde{\xi}_k^j\}}$$

  and adjust the weights as follows:

$$\omega_k^j = \frac{\sum_{s \in J^i} \tilde{\omega}_k^s}{N_p^i}.$$

- The new set of particles is $\{t_k^i, \xi_k^i, \omega_k^i\}_{i=1}^{N_p}$.
- If $k < m$ then repeat steps 2 and 3 for $k := k + 1$.
- Otherwise stop, with $P_{hit}(0, T) \approx \prod_{k=1}^{m} \gamma_k^{N_p}$.

5.3 Numerical evaluations

5.3.1 Switching diffusion example

This section illustrates the performance of the Monte Carlo (MC) approach, the IPS algorithm of Cérou *et al.* (2002) and the effect of the alternative steps 1H, 2H and 3H of HIPS algorithm for a switching diffusion. Table 5.1 presents the list of tested algorithms. There IPS stands for the algorithm of Cérou *et al.* (2002) in case of a switching diffusion (Section 5.2.2); HIPS 1 is IPS with improved initial sampling step 1H and with resampling per mode step 3H (Section 5.2.3); and HIPS 2 is HIPS 1 plus importance switching (Section 5.2.3).

Table 5.1: Tested Algorithms

Algorithm	Particle system	Initial sampling per mode	Resampling per mode	Importance switching
MC	–	Yes (1H)	–	–
IPS	Yes	–	–	–
HIPS1	Yes	Yes (1H)	Yes (3H)	–
HIPS2	Yes	Yes (1H)	Yes (3H)	Yes (2H)

We use the following stochastic hybrid process as a test example. We extend the diffusion (2.3.15) to an $\mathbb{R} \times \{e_1, e_2, e_3\}$-valued Markovian switching diffusion $\{X_t, \theta_t\}$ which is described by the following SDE

$$dX_t = \Big(\mu(\theta_t) + \frac{\sigma(\theta_t)^2}{2}\Big)X_t dt + \sigma(\theta_t)X_t dW_t,$$
$$d\theta_t = \int_{\mathbb{R}} c(\theta_{t-}, z)p(dt, dz).$$

Initial conditions:

$$P(X_0 = 1) = 1; \qquad P(\theta_0 = e_1) = 1 - 10^{-7} - 10^{-9};$$
$$P(\theta_0 = e_2) = 10^{-7}; \qquad P(\theta_0 = e_3) = 10^{-9};$$

Parameters:

$$\begin{array}{rclrclrcl} \mu(e_1) & = & 1, & \mu(e_2) & = & 4, & \mu(e_3) & = & 3, \\ \sigma(e_1) & = & 1, & \sigma(e_2) & = & 0.9, & \sigma(e_3) & = & 1.7. \end{array}$$

$$c(e_i, z) = \begin{cases} e_j - e_i & \text{if } z \in \Delta_{ij} \\ 0 & \text{otherwise.} \end{cases}$$

$|\Delta_{ij}| = \lambda_{ij}$ (see Chapter 4) and the switching rates are:

$$\begin{array}{rclrcl} \lambda_{12} & = & 1 \cdot 10^{-4}, & \lambda_{13} & = & 1 \cdot 10^{-6}, \\ \lambda_{21} & = & 5 \cdot 10^{-5}, & \lambda_{23} & = & 1 \cdot 10^{-5}, \\ \lambda_{31} & = & 5 \cdot 10^{-5}, & \lambda_{32} & = & 1 \cdot 10^{-4}. \end{array}$$

Let σ_1 be the first jump time of continuous time Markov chain $\{\theta_t\}$. The probability that $\{\theta_t\}$ will stay in initial state $\theta_0 = e_i$ for more than t time units is given by

$$P(\sigma_1 > t | \theta_0 = e_i) = \exp\Big\{ -\sum_{\substack{j=1 \\ j \neq i}}^{3} \lambda_{ij} t \Big\} \approx 1 \text{ for } t \in [0, 1],$$

$i = 1, 2, 3$. Actually

$$P(\sigma_1 > 1 | \theta_0 = e_1) = 0.99990, \tag{5.3.1}$$
$$P(\sigma_1 > 1 | \theta_0 = e_2) = 0.99994, \tag{5.3.2}$$
$$P(\sigma_1 > 1 | \theta_0 = e_3) = 0.99985, \tag{5.3.3}$$

i.e. probabilities that $\{\theta_t\}$ will switch within time interval $[0, T]$, $T = 1$ are very small.

As in the preceding example, we want to estimate the probability that continuous valued component $\{X_t\}$ will hit level d before time $T = 1$, i.e. $P(\tau_d \leq 1)$

where $\tau_d \triangleq \inf\{t > 0 : X_t \in [d, +\infty)\}$. We are also interested in estimating the probabilities of hitting the barrier in particular modes:

$$P(\tau_d \leq 1, \theta_{\tau_d} = e_i),\ e_i \in \mathbb{M}. \tag{5.3.4}$$

Unfortunately, there is no analytical solution available for our switching diffusion example to compare with the results of simulation algorithms. The evolution of continuous component X_t is more complicated now. It switches from one diffusion path to another as the discrete component θ_t jumps from one state to another. Luckily, we can use the analytical formula (2.3.16) to evaluate the conditional probabilities of hitting the barrier d within time interval $(0, T]$ given that the switching diffusion started in a particular mode and never switched (i.e. X_t evolves just as diffusion (2.3.15) only with different coefficients):

$$P(\tau_d \leq 1, \theta_{\tau_d} = e_i | \theta_s = e_i, s \in [0, \tau_d \wedge 1)),\ e_i \in \mathbb{M}. \tag{5.3.5}$$

Figure 5.4 illustrates the conditional probabilities (5.3.5) as functions of d. We see that reaching the target level d is more likely in mode 2 or 3 and less likely in mode 1. It is also obvious that higher number of particles will survive in modes 2 and 3 than in mode 1, but their weights can be relatively small because of initial conditions. Using (5.3.1)-(5.3.3), (5.3.5) and the initial conditions we can analytically evaluate close approximations of the following probabilities (see Table 5.2):

$$\begin{aligned} &P(\tau_d \leq 1, \theta_s = e_i, s \in [0, \tau_d]) \\ &\qquad = P(\tau_d \leq 1, \theta_{\tau_d} = e_i | \theta_s = e_i, s \in [0, \tau_d \wedge 1)) \times \\ &\qquad\qquad \times P(\theta_s = e_i, s \in (0, \tau_d \wedge 1) | \theta_0 = e_i) P(\theta_0 = e_i) \\ &\qquad \approx P(\tau_d \leq 1, \theta_{\tau_d} = e_i | \theta_s = e_i, s \in [0, \tau_d \wedge 1)) \times \\ &\qquad\qquad \times P(\sigma_1 > 1 | \theta_0 = e_i) P(\theta_0 = e_i), \end{aligned} \tag{5.3.6}$$

$i = 1, 2, 3$.

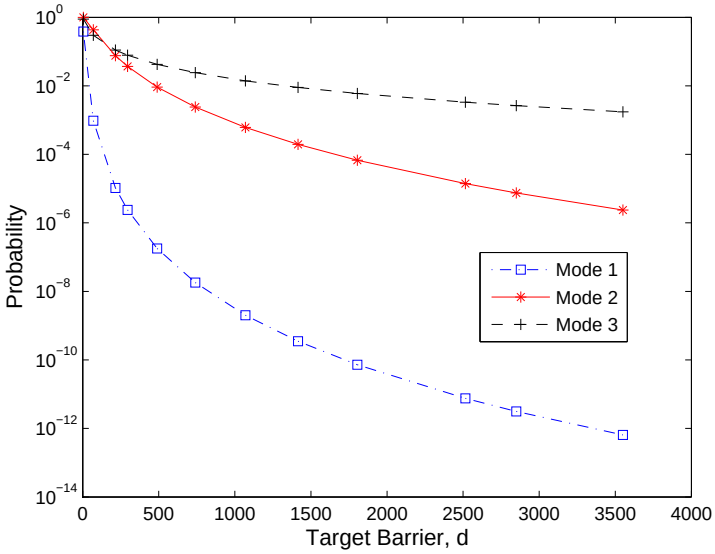


Figure 5.4: Probabilities $P(\tau_d \leq 1, \theta_{\tau_d} = e_i | \theta_s = e_i, s \in [0, \tau_d \wedge 1))$, $i = 1, 2, 3$.

Table 5.2: Analytical estimates of probabilities $P(\tau_d \leq 1, \theta_s = e_i, s \in [0, \tau_d])$

Target Level, d	Sum of three Modes	Probability $P(\tau_d \leq 1, \theta_s = e_i, s \in [0, \tau_d])$		
		Mode 1	Mode 2	Mode 3
5	3.84E-01	3.84E-01	9.98E-08	8.88E-10
70	9.50E-04	9.50E-04	4.33E-08	2.99E-10
215	1.05E-05	1.05e-05	7.58E-09	1.11E-10
295	2.38E-06	2.38E-06	3.68E-09	7.86E-11
490	1.79E-07	1.78E-07	9.17E-10	4.23E-11
740	1.83E-08	1.80E-08	2.40E-10	2.42E-11
1070	2.09E-09	2.02E-09	6.11E-11	1.40E-11
1415	3.79E-10	3.51E-10	1.95E-11	8.97E-12
1805	8.46E-11	7.19E-11	6.70E-12	5.97E-12
2515	1.23E-11	7.55E-12	1.39E-12	3.33E-12
2850	6.53E-12	3.14E-12	7.46E-13	2.64E-12
3550	2.64E-12	6.48E-13	2.38E-13	1.75E-12

These values provide insight into evolution of the process in different modes. We can use these analytical estimates as a reference for interpretation of our simulation results.

In the following sections the performance of several algorithms is demonstrated and their relative merits are discussed.

For the IPS, HIPS1 and HIPS2 we use same sequence of nested sets $D = D_m \subset \cdots \subset D_1$, $D_j = [d_j, +\infty)$, $d = d_m > \cdots > d_1 > 0$, as in example for diffusion (Section 2.3.3). We run 1000 simulations with 1000 particles ($N_p^1 = 500$ for Mode 1, $N_p^2 = 300$ for Mode 2 and $N_p^3 = 200$ for Mode 3). For algorithm HIPS2 the importance switching rates are $\hat{\lambda}_{ij} = \frac{1}{30}$ for $i \neq j$.

Standard Monte Carlo

To begin with, we want to present the results obtained by running weighted standard Monte Carlo simulation. The simulation starts with 5000000 particles in Mode 1, 3000000 in Mode 2 and 2000000 in Mode 3 (in total 10^7 particles), and each particle is assigned its weight according to initial conditions (see step 1H in Section 5.2.3). Then, each trial is independently simulated until it reaches barrier d or the end time $T = 1$. Table 5.3 presents the results of 10^7 runs.

Table 5.3: Probability to hit barrier d. Standard Monte Carlo, 10^7 runs.

Target Level, d	Total probability $P(\tau_d \leq 1)$	Probability $P(\tau_d \leq 1, \theta_{\tau_d} = e_i)$		
		Mode 1	Mode 2	Mode 3
5	3.74E-01	3.74E-01	6.35E-05	1.40E-06
70	8.39E-04	8.29E-04	9.64E-06	2.89E-10
215	1.02E-05	9.00E-06	1.21E-06	1.05E-10
295	2.60E-06	2.20E-06	4.03E-07	7.38E-11
490	4.01E-07	4.00E-07	7.82E-10	3.91E-11
740	2.18E-10	0.00E+00	1.96E-10	2.20E-11
1070	5.87E-11	0.00E+00	4.62E-11	1.25E-11
1415	2.17E-11	0.00E+00	1.38E-11	7.94E-12
1805	1.01E-11	0.00E+00	4.80E-12	5.27E-12
2515	3.38E-12	0.00E+00	5.00E-13	2.88E-12
2850	2.60E-12	0.00E+00	3.00E-13	2.30E-12
3550	1.58E-12	0.00E+00	6.67E-14	1.51E-12

The results show that standard Monte Carlo (MC) method does not allow to obtain reasonable estimates of rare events by running reasonable number of simulations (compare Tables 5.2 and 5.3). 10^7 runs of MC enable reasonable estimates to be made for events with probabilities of order 10^{-6}. The probability to reach the highest barrier $d = 3550$ in modes 2 and 3, given that the process

starts in one of these modes and never switches, is higher than 10^{-6} (see Figure 5.4). That is why MC shows rather better performance in these modes than in mode 1. However, when switching rates are low, an MC simulation also needs numerous trials to achieve reliable estimate. We conclude that MC method is very time consuming and not appropriate for estimating event probabilities of orders less than 10^{-6}.

IPS

Table 5.4 presents the results obtained by running IPS algorithm for switching diffusion. As we have expected, the IPS algorithm fails to keep enough particles in "light" modes 2 and 3. After several resampling steps more and more "heavy" particles are being sampled from mode 1 with the highest probability, and the "light" particles from modes 2 and 3 become discarded. The sampling per mode can help to solve this problem.

Table 5.4: Probability to hit barrier d. IPS approach, 1000 simulations of 1000 particles.

Target Level, d	Total probability $P(\tau_d \leq 1)$	Probability $P(\tau_d \leq 1, \theta_{\tau_d} = e_i)$		
		Mode 1	Mode 2	Mode 3
5	3.74E-01	3.74E-01	5.24E-05	0.00E+00
70	8.18E-04	8.15E-04	3.45E-06	1.57E-08
215	8.78E-06	8.57E-06	2.08E-07	6.23E-10
295	1.99E-06	1.92E-06	6.98E-08	1.90E-10
490	1.42E-07	1.34E-07	8.47E-09	6.28E-11
740	1.34E-08	1.27E-08	6.74E-10	1.26E-11
1070	1.34E-09	1.31E-09	2.66E-11	1.75E-12
1415	2.17E-10	2.15E-10	1.28E-12	2.34E-13
1805	3.33E-11	3.31E-11	9.11E-14	1.89E-14
2515	3.15E-12	3.15E-12	3.10E-15	8.74E-18
2850	1.21E-12	1.21E-12	7.06E-16	0.00E+00
3550	2.85E-13	2.85E-13	3.01E-17	0.00E+00

HIPS1

Our next results show how the situation can be improved by using "sampling per mode" for the initial sampling step and for the resampling step. Resampling per mode allows us to avoid loss of "light" particles in "light" modes, e.g. in mode 3, and helps to maintain fixed number of particles in each mode. Compare Tables 5.2 and 5.5. The results for mode 1 and 3 are almost the same as analytical

estimates in Table 5.2. However, the probability to hit barrier d, especially for low values of d, in mode 2 differs from the analytical estimates of probability $P(\tau_d \leq 1, \theta_s = e_2, s \in [0, \tau_d]])$ presented in Table 5.2. This means there were some switches from mode 1 to mode 2, i.e. heavy particles from mode 1 "arrived" to mode 2 before they have hit the level. It is important to understand that even rare switches may have considerable effect on final result. In our example switching rates are very low, so we should not expect many switches during simulation. We therefore believe that the probabilities of hitting the target barrier in some modes are underestimated.

Table 5.5: Probability to hit barrier d. HIPS1 approach, 1000 simulations of 1000 particles.

Target Level, d	Total probability $P(\tau_d \leq 1)$	Probability $P(\tau_d \leq 1, \theta_{\tau_d} = e_i)$		
		Mode 1	Mode 2	Mode 3
5	3.73E-01	3.73E-01	7.75E-05	8.86E-10
70	8.13E-04	8.03E-04	9.49E-06	2.27E-08
215	9.26E-06	8.60E-06	6.55E-07	3.21E-10
295	2.16E-06	1.94E-06	2.16E-07	1.15E-10
490	1.76E-07	1.38E-07	3.81E-08	4.06E-11
740	1.83E-08	1.27E-08	5.57E-09	2.18E-11
1070	1.88E-09	1.31E-09	5.61E-10	1.23E-11
1415	3.16E-10	2.17E-10	9.11E-11	7.78E-12
1805	5.68E-11	3.68E-11	1.52E-11	4.77E-12
2515	7.99E-12	3.85E-12	1.54E-12	2.60E-12
2850	3.97E-12	1.51E-12	5.13E-13	1.95E-12
3550	1.78E-12	3.54E-13	1.15E-13	1.31E-12

HIPS2

Next, we will see how the results change if we use algorithm HIPS2, i.e. HIPS1 with the combination of Importance Switching technique. Table 5.6 presents the results of HIPS2 algorithm. The increase of the frequencies of switches has a considerable effect on probabilities of all modes and thus on total probability. This is what one should expect. First of all, at the beginning heavy particles from mode 1 can switch to mode 2 or mode 3, and thus, to increase the weights of these modes, which in turn increases the hitting probabilities in these modes. Second, particles in modes 2 and 3 approach the rare set faster than in mode 1, and if the switching to mode 1 occurs, then there is high chance that the particle is close to the target level and there is high chance that it will hit that level. Thus, there is no big surprise that we observe higher hitting probabilities in modes 1,2

and 3 after increasing frequencies of switchings.

This example shows how it is important to take care of rare switches, since it may have a substantial effect on total result. But it is not obvious how much we can change the frequencies (rates of θ's) without introducing perceptible distortion of the results. The tests of the HIPS2 algorithm had shown that it is stable with different choices of switching rates $\hat{\lambda}_{ij}$.

Table 5.6: Probability to hit barrier d. HIPS2 approach, 1000 simulations of 1000 particles.

Target Level, d	Total probability $P(\tau_d \leq 1)$	Probability $P(\tau_d \leq 1, \theta_{\tau_d} = e_i)$		
		Mode 1	Mode 2	Mode 3
5	4.04E-01	4.04E-01	4.82E-05	4.32E-07
70	1.47E-03	1.45E-03	1.23E-05	1.27E-07
215	2.34E-05	2.14E-05	1.93E-06	4.78E-08
295	6.52E-06	5.51E-06	9.74E-07	3.59E-08
490	8.14E-07	5.22E-07	2.70E-07	2.18E-08
740	1.63E-07	6.58E-08	8.25E-08	1.46E-08
1070	4.35E-08	8.65E-09	2.46E-08	1.02E-08
1415	1.87E-08	1.72E-09	9.29E-09	7.70E-09
1805	1.00E-08	3.96E-10	3.67E-09	5.94E-09
2515	5.55E-09	7.53E-11	1.18E-09	4.29E-09
2850	4.75E-09	4.04E-11	8.57E-10	3.85E-09
3550	3.67E-09	1.41E-11	4.75E-10	3.18E-09

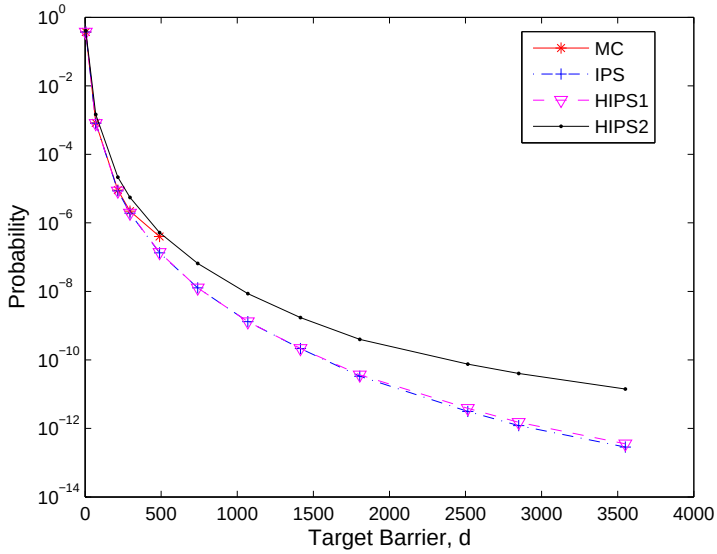
5.3.2 Discussion of numerical results

Figures 5.5(a), 5.5(b) and 5.6(a) present the estimated values of probabilities $P(\tau_d \leq 1, \theta_{\tau_d} = e_i)$, $e_i \in \mathbb{M}$, obtained by running algorithms listed in Table 5.1. The estimated values of total hitting probability $P(\tau_d \leq 1)$ are given in Figure 5.6(b). The results in Figure 5.5(a) show that MC stops at $d = 490$ in mode 1. This implies the same for the total probability in Figure 5.6(b).

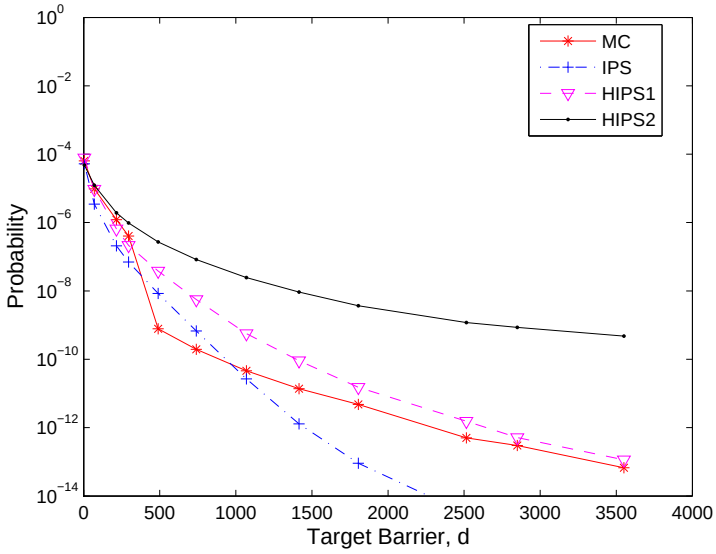
IPS algorithm looses most of the "light" particles in modes 2 and 3 (see Figures 5.5(b) and 5.6(a)).

The HIPS1 algorithm allows to avoid loss of light particles in "light" modes, e.g. in mode 3, and helps to maintain fixed number of particles in each mode. However, it is not able to cope with rare switches. It may happen that the switch itself is a rare event but it may have a significant influence on the total result. If this is the case (as in our example), then it is unlikely that there will be any switches during the simulation.

Algorithm HIPS2 copes well with both the problem of rare switches and problem of "light" modes. It forces interaction between the modes by making rare switches more frequent and properly adjusting the weights of particles. The results in Figures 5.5(a), 5.5(b) and 5.6(a) show that the increase of the frequencies of switches has a considerable effect on probabilities of mode 2 and 3 and thus on total probability (Figure 5.6(b)). This is what one should expect because the heavy particles can leave mode 1, although very rarely, and have a great influence on modes 2 and 3. The interaction between modes 2 and 3 is not really noticeable but these modes have an influence on mode 1.

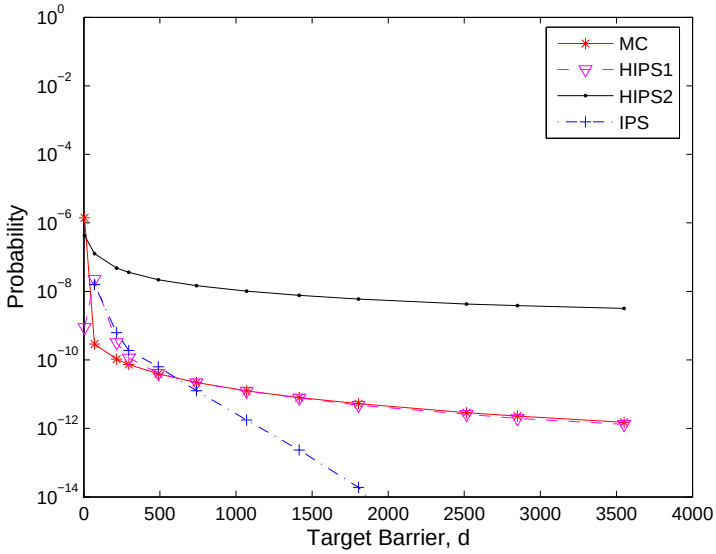


(a) Probability to hit barrier d in mode 1. MC stops at $d = 490$.

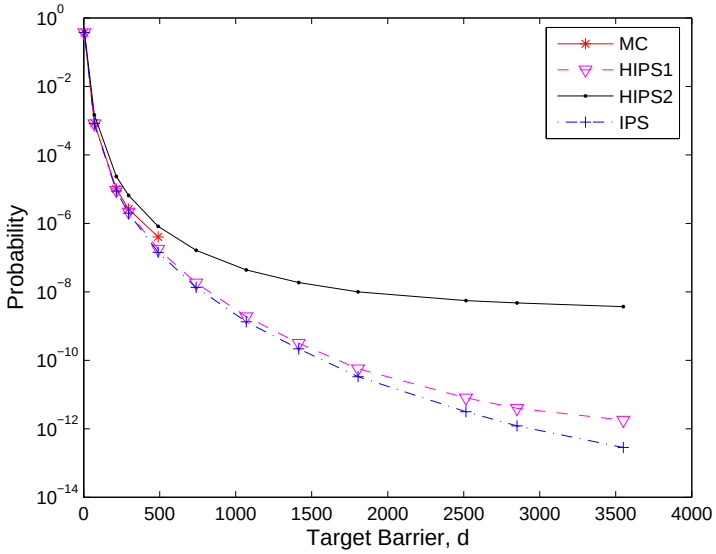


(b) Probability to hit barrier d in mode 2.

Figure 5.5: Comparison of algorithms



(a) Probability to hit barrier d in mode 3.



(b) Total hitting probability of barrier d. MC stops at $d = 490$.

Figure 5.6: Comparison of algorithms

5.4 Concluding Remarks

This chapter aimed to extend the sequential Monte Carlo approach of Cérou *et al.* (2002) to estimating rare events of rarely switching diffusions. We have formulated the approach of Cérou *et al.* (2002) to explicitly include the switching diffusion situation. Then we have developed two extensions: sampling per mode to cope with large differences in mode weights, and importance switching to cope with rare mode switching. Next, we evaluated the algorithms for a switching diffusion example. Table 5.1 lists the tested algorithms. The obtained results show that all the proposed extensions are in fact needed for estimating rare events for a rarely switching diffusion. The best performing HIPS2 algorithm is able to cope with differences in weights (sampling per mode), rare switches of discrete valued component (importance switching) and rare visits of Euclidean valued component to the target set (decomposition of rare event probability).

Chapter 6

Conclusions and Future Research

This chapter summarizes the main contributions of this thesis. Afterwards, we present a few directions regarding future research.

6.1 Conclusions

In this thesis we develop a general framework for the modelling and assessment of accident risks in complex stochastic dynamical systems. In particular, our modelling framework is aimed to facilitate the design of new advanced Air Traffic Management (ATM) concepts.

In Chapter 2 we have discussed the problem of accident risk assessment in stochastic dynamical systems, using the simple example from ATM. We have chosen to express the risk as a hitting probability, i.e. the probability that the system under consideration will reach a certain critical state within a given time frame. In case when the dimension of a stochastic process describing the dynamical system is at most three, the hitting probability can be numerically evaluated through solving a backward Kolmogorov PDE with Dirichlet type boundary conditions. For complex real-world applications Monte Carlo simulation remains the only feasible approach. Of course, it is unrealistic to estimate extremely small rare event probabilities just by using a naive Monte Carlo simulation. When the extremely small probabilities are considered one has to use special techniques which allow the speeding of the simulation. Importance sampling and multilevel splitting methods have been used to obtain dramatic improvements in efficiency in estimating small probabilities in queueing and reliability systems. However, these two methods have serious drawbacks. The importance sampling techniques, based on changing probability distributions to make rare events less rare, depend critically on the ability to find the right change of measure. Finding the right

change of measure in complex dynamical models is practically impossible. The efficiency of multilevel splitting techniques depends on the choice of the level sets and especially on the optimal number of splittings. Too many splittings results in explosive computational requirements, and too few splittings provides no speed-up gains over straightforward Monte Carlo simulation. So far, the theoretical results regarding the optimal choice of splittings have been obtained only for very simple cases. On the other hand, the IPS based approaches do not have such restrictive limitations as the multilevel splitting techniques. Moreover, the IPS approaches have the advantage of being based on sound theory with many general results available. Our numerical tests in Chapter 2 showed that IPS approach has a great potential for accident risk assessment in stochastic dynamical systems.

Chapter 3 deals with the modelling issues of stochastic hybrid systems. We restrict our attention to the stochastic approach of modelling, i.e. modelling by Markov processes which are defined as solutions to SDE. In addition, we require the solutions to be semimartingale processes. Considering SDE with semimartingale solutions gives an advantage. It allows to use the powerful stochastic calculus available for the semimartingale processes when performing complex stochastic analysis. We have used (Gihman and Skorohod, 1982) and (Jacod and Shiryaev, 1987) to identify and characterize a very general class of jump-diffusions which are defined as semimartingale solutions of SDE. Next we have followed a similar path as taken by (Blom, 1990, 2003) in transferring these results to the class of stochastic hybrid processes. As a result, we have obtained two stochastic models: KB1 and KB2. These two models cover a range of interesting phenomena:

(i) Switching diffusion: between the random switches of the discrete valued component, the Euclidean valued component evolves as diffusion.

(ii) Random hybrid jumps: simultaneous and dependent jumps and switches of discrete and Euclidean valued components are driven by a Poisson random measure.

(iii) Boundary hybrid jumps: simultaneous and dependent jumps and switches of discrete and Euclidean valued components are initiated by boundary hittings.

(iv) Martingale inducing jumps: the Euclidean valued components driven by a compensated Poisson random measure may jump so frequently that it is no longer a process of finite variation. These jumps represent the pure discontinuous martingale part of the process.

To be precise, model KB1 combines the features (i), (ii) and (iv), and model KB2 combines the features (i), (ii) and (iii). We have shown that the Markov and the strong Markov properties hold true for the stochastic hybrid model KB2. The strong Markov property is required for the development of efficient IPS based rare event simulation techniques.

Most SDE do not admit closed form analytical solutions, the only alternative of studying them is then numerical simulation. The sample paths (realizations) or functionals of solutions of SDE are commonly simulated through discrete-time approximations which are implementable on digital computers. Despite the availability in the literature of a large number of numerical techniques for SDE, there is hardly any focus on SDE on hybrid state spaces. In Chapter 4 we study the problem of approximation of first passage times in stochastic hybrid processes with state dependent switching rates. We took a switching diffusion of Ghosh *et al.* (1997) as a non-trivial example of a stochastic hybrid process with state dependent switching rates; developed for it a strong Euler-type discretization scheme and proved its convergence. Then, following the approach of Gobet (1999*a*,*b*) we have shown that the first passage time of discretized switching diffusion converges in distribution to the first passage time of continuous time switching diffusion process as the maximal discretization step h tends to zero. Using our Euler-type discretization scheme one can easily generate sample paths of switching diffusion and by taking a sample average of $\mathbf{1}_{\{\tau^h \leq T\}}$, where τ^h is the Euler approximation of the stopping time τ (i.e. the first passage time), one can obtain the estimate of the probability of reaching a given target set within the time interval $(0, T]$. These results are necessary in order to implement the IPS simulation approaches for a switching diffusions on a computer.

In Chapter 5 we extend the IPS approach of Cérou *et al.* (2002) to estimating rare event probabilities in stochastic hybrid systems. While in theory the IPS approach is applicable virtually to any strong Markov process, in practice the straightforward application of this approach to stochastic hybrid processes fails to produce good estimates within a reasonable amount of time. First of all, if the initial probabilities of starting in certain modes are very small, it is highly unlikely that one would draw particles in these "light" modes, and the bulk of the particles is being sampled in the "heavy" modes, i.e. the modes with higher probabilities. Subsequently, most, if not all, of the few particles in the "light" modes, most likely, will be lost after several resampling steps. To avoid this, at the initial sampling step we sample a fixed number of particles for each mode separately and assign to each particle its "importance" weight in accordance with the initial probability. During the successive resampling steps the particles with larger weights have higher chances to be selected, and the particles with small weights in "light" modes tend to be discarded. We solve this problem by sampling fixed number of particles for each mode separately and adjusting the weights appropriately. Introduction of weights requires the use of weighted empirical measures in the form $\sum_{i=1}^{N_p} \omega_k^i \delta_{\{\xi_k^i\}}$, instead of empirical measures in the form $\frac{1}{N_p} \sum_{i=1}^{N_p} \delta_{\{\xi_k^i\}}$. In the classical IPS approach all particles actually have the same weights equal to $1/N_p$, where N_p is the number of particles. The second important problem is the problem of rare switches. If the probability of some mode transitions (switches) is very small, it is highly unlikely to observe even one switch during a simulation run. In such cases, the possible switchings between the modes are not properly

taken into account. Together with the problem of losing "light" particles, this severely affects IPS estimator performance. By increasing the number of particles the IPS estimates should improve, but only at the cost of substantially increased simulation time. In order to avoid the need to increase the number of particles when the switching rates are decreasing, we introduce a sequential importance switching technique. Our numerical studies have shown that all the proposed extensions are significantly contributing to efficient estimation of rare event probabilities in stochastic hybrid systems. The best performing HIPS2 algorithm is able to cope with differences in weights (sampling per mode), rare switches of discrete valued component (importance switching) and rare visits of Euclidean valued component to the target set (decomposition of rare event probability).

6.2 Future Research

In this section we review several issues which need attention in the near future.

6.2.1 Extension of SDE models on hybrid state space

In Chapter 3 of this thesis we have studied the question of existence and uniqueness of semimartingale strong solutions to SDE on a hybrid state space (models KB1 and KB2). We have also reviewed several related SDE models of Blom (2003) and Ghosh and Bagchi (2004) (models HB1, HB2, GB1 and GB2). The hierarchy between these stochastic hybrid models is depicted in Figure 3.2. The next step is to develop even more general stochastic hybrid model which will combine the most interesting features of models KB1, KB2=HB2 and GB2:

- switching hybrid-jump diffusion as in KB1;
- instantaneous jump reflection at the boundary as in KB2=HB2 and GB2;
- Mode dependent dimension as in GB2.

It will be important to enquire whether the extended model would retain the strong Markov property.

6.2.2 Estimating rare event probabilities in large scale stochastic hybrid systems

In Chapter 5 hybrid versions of the IPS algorithm have been developed. Our new HIPS algorithms are able to cope with large differences in discrete state (mode) probabilities or with rare mode switchings. However, these hybrid versions would need impractically many particles when the state space of the discrete valued component is very large. Such situation typically occurs when the stochastic hybrid process considered is highly distributed and incorporates many local discrete

valued switching processes. In such cases, the large number of interacting stochastic hybrid processes makes the size of the state space of the total discrete valued component exponentially large. Future research should aim at extending the HIPS algorithm for estimation of extremely small rare event probabilities in large scale stochastic hybrid systems. Currently, we are already working on the development of the Hierarchical version of HIPS algorithm. The effectiveness of the approach is initially tested on an advanced ATM scenario (Blom *et al.*, 2006). This showed that the speed up was very good for such a complex Monte Carlo simulation model. However it also appeared necessary to develop another factor 10 to 100 of additional speed up factor. Several directions to accomplish this already have been identified:

- Develop an effective combination of Interacting Particle System based rare event simulation with Markov Chain Monte Carlo speed up technique;
- Develop a method to assess the sensitivity of multiple aircraft encounter geometries to collision risk, and develop importance sampling approaches which take advantage of these sensitivities;
- Develop novel ways how Interacting Particle System speed up techniques that apply to a pair of aircraft can effectively be extended to situations of multiple aircraft;
- Find a way for optimal selection of importance sampling (switching) distributions in Hybrid Interacting Particle System based rare event simulation approach;
- Develop a technique for optimal choice of nested level sets $D = D_m \subset \cdots \subset D_2 \subset D_1$ in high-dimensional hybrid state spaces.

Bibliography

Aldemir, T., N. Siu, P. Cacciabue, B. Gül Göktepe and A. Mosleh (1994). *Reliability and safety assessment for dynamic process systems*. Vol. 120 of *NATO ASI Series*. Springer. Berlin.

Bakker, G.J. and H.A.P Blom (1993). Air traffic collision risk modelling. In: *32nd IEEE Conference on Decision and Control*. Vol. 2. Institute of Electrical and Electronics Engineers, New York. pp. 1464–1469.

Blom, H.A.P. (1990). Bayesian estimation for decision-directed stochastic control. PhD thesis. Delft University of Technology.

Blom, H.A.P. (2003). Stochastic hybrid processes with hybrid jumps. In: *Proc. IFAC Conf. Analysis and Design of Hybrid Systems*. Saint-Malo, Brittany, France.

Blom, H.A.P. and G.J. Bakker (2002). Conflict probability and incrossing probability in air traffic management. In: *Proceedings IEEE Conference on Decision and Control*. Las Vegas, NV. pp. 2421–2426.

Blom, H.A.P., G.J. Bakker, M.H.C. Everdij and M.N.J. van der Park (2003*a*). Collision risk modeling of air traffic. In: *Proc. European Control Conf.*. Cambridge, UK.

Blom, H.A.P., G.J. Bakker, M.H.C. Everdij and M.N.J. van der Park (2003*b*). *Stochastic analysis background of accident risk assessment for Air Traffic Management*. National Aerospace Laboratory NLR. Hybridge Report D2.2, `http://www.nlr.nl/public/hosted-sites/hybridge`.

Blom, H.A.P., J. Krystul, G.J. Bakker, M.B. Klompstra and B. Klein Obbink (2006). Estimation of collision risk in free flight air traffic by sequential monte carlo simulation. In: *Stochastic Hybrid Systems: Recent Developments and Research Trends* (C.G. Cassandras and J. Lygeros, Eds.). CRC Press.

Blom, H.A.P., M.B. Klompstra and G.J. Bakker (1994). Air traffic management as a multiagent stochastic dynamic game under partial state observation. In:

Proc. IFAC 7th Symposium on Transportation systems: Theory and application of advanced technology. Tianjijn, China. pp. 249–254.

Bujorianu, M. and J. Lygeros (2004). General stochastic hybrid systems: Modelling and optimal control. In: *Proc. IEEE Conference on Decision and Control*. Bahamas.

Cérou, F., P. Del Moral, F. Le Gland and P. Lezaud (2005). Limit theorems for the multilevel splitting algorithm in the simulation of rare events. In: *Proceedings of the 2005 Winter Simulation Conference* (M.E. Kuhl, N.M. Steiger, F.B. Armstrong and J.A. Joines, Eds.). Orlando.

Cérou, F.P., P. Del Moral, F. Le Gland and P. Lezaud (2002). Genetic genealogical models in rare event analysis. Research Report. `http://www.nlr.nl/public/hosted-sites/hybridge/`.

Davis, M.H.A. (1984). Piecewise-Deterministic Markov Processes: A general class of non-diffusion stochastic models. *J.R. Statist. Soc. B* **46**(3), 353–388.

Davis, M.H.A. (1993). *Markov models and optimization*. Chapman & Hall.

Del Moral, P. (2004). *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Probability and its Applications. Springer-Verlag. New York.

Del Moral, P. and P. Lezaud (2006). Branching and interacting particle interpretation of rare event probabilities. In: *Stochastic Hybrid Systems: Theory and Safety Critical Applications* (H. Blom and J. Lygeros, Eds.). pp. 351–389. Number 337 In: *Lecture Notes in Control and Information Sciences*. Springer-Verlag. Berlin.

Doucet, A., N. de Freitas and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag.

Duyn Schouten, F.A.van der and A. Hordijk (1983). Average optimal policies in Markov decision drift processes with applications to a queueing and a replacement model. *Adv. Appl. Prob.* **15**, 274303.

Embrechts, P., C. Kluppelberg and T. Mikosh (1997). *Modeling extremal events for insurance and finance*. Springer.

Fox, B.L. and P.W. Glynn (1988). Computing Poisson probabilities. *Communications of the ACM*.

Friedman, A.V. (1975). *Stochastic Differential Equations and Applications*. Volume 28-1 of Probability and Mathematical Statistics. Academic Press. New York.

Ghosh, M.K., A. Arapostathis and S.I. Marcus (1993). Optimal control of switching diffusions with application to flexible manufacturing systems. *SIAM J. Control Optimization* **31**, 1183–1204.

Ghosh, M.K., A. Arapostathis and S.I. Marcus (1997). Ergodic control of switching diffusions. *SIAM J. Control Optimization* **35**, 1952–1988.

Ghosh, M.K. and A. Bagchi (2004). Modeling stochastic hybrid systems. In: *System Modeling and Optimization. Proceedings of the* 21^{st} *IFIP TC7 Conference* (J. Cagnol and J.P. Zolèsio, Eds.). Sophia Antipolis, France. pp. 269–279.

Gihman, I.I. and A.V. Skorohod (1975). *The theory of stochastic processes II.* Springer-Verlag. Berlin.

Gihman, I.I. and A.V. Skorohod (1982). *Stochastic differential equations and their applications*. Naukova Dumka. Kiev. (in Russian).

Glasserman, P. and J. Li (2005). Importance sampling for portfolio credit risk. To appear in Management Science.

Glasserman, P. and N. Merener (2004). *Stochastic Analysis with Applications to Mathematical Finance*. Chap. Convergence of a Discretization Scheme for Jump-Diffusion Processes with State-Dependent Intensities. London, UK.

Glasserman, P., P. Heidelberger, P. Shahabuddin and T. Zajic (1999). Multilevel splitting for estimating rare event probabilities. *Operations research* **47**(4), 585–600.

Gobet, E. (1999*a*). Weak approximation of killed diffusion. Part I: Continuous Euler scheme. preprint.

Gobet, E. (1999*b*). Weak approximation of killed diffusion. Part II: Discrete Euler scheme. preprint.

Gobet, E. (2000). Weak approximation of killed diffusion using Euler schemes. *Stochastic Processes and their Applications* **87**, 167–197.

Gobet, E. (2001). Euler schemes and half-space approximation for the simulation of diffusion in a domain. *ESAIM Probability and Statistics* **5**, 261–297.

Heidelberger, P. (1995). Fast simulation of rare events in queueing and reliability models. *ACM Trans. Modeling and Computer Simulation* **5**, 43–85.

Hsu, D.A. (1981). The evaluation of aircraft collision probabilities at intersecting air routes. *J. of Navigation* **34**, 78–102.

Hu, J., J. Lygeros and S. Sastry (2000). Towards a theory of stochastic hybrid systems. In: *Lecture notes in Computer Science*. Vol. 1790. pp. 160–173.

Ikeda, N. and S. Watanabe (1989). *Stochastic Differential Equations and Diffusion Processes*. Second ed.. North-Holland. Kodansha.

Jacod, J. and A.N. Shiryaev (1987). *Limit Theorems for Stochastic Processes*. Springer. Berlin.

Jacod, J. and A.V. Skorokhod (1996). Jumping Markov Processes. *Annales de l'I. H. P.* **32**(1), 11–67.

Jensen, A. (1953). Markov chains as an aid in the study of Markov processes. *Skand. Aktuarietidskr* **36**, 87–91.

Juneja, S. and P. Shahabuddin (2006). *Rare Event Simulation Techniques: An Introduction and Recent Advances*. To appear in Handbook on Simulation.

Kloeden, P.E. and E. Platen (1992). *Numerical solution of stochastic differential equations*. Springer-Verlag.

Krystul, J., A. Bagchi and H.A.P. Blom (2003). Risk decomposition and assessment methods. Hybridge Report D8.1. `http://www.nlr.nl/public/hosted-sites/hybridge/`.

Krystul, J. and A. Bagchi (2004). Approximation of first passage times of switching diffusion. In: *Proc. Sixteenth International Symposium on Mathematical Theory of Networks and Systems*. Leuven, Belgium.

Krystul, J. and H.A.P. Blom (2004). Monte Carlo simulation of rare events in hybrid system. Hybridge Report D8.3. `http://www.nlr.nl/public/hosted-sites/hybridge/`.

Krystul, J. and H.A.P. Blom (2005*a*). Generalized stochastic hybrid processes as strong solutions of stochastic differential equations. Hybridge Report D2.3. `http://www.nlr.nl/public/hosted-sites/hybridge/`.

Krystul, J. and H.A.P. Blom (2005*b*). Sequential Monte Carlo simulation of rare event probability in stochastic hybrid systems. In: *16th IFAC World congress*. Prague, Czech Republic.

Krystul, J. and H.A.P. Blom (2006). Sequential Monte Carlo simulation for the estimation of small reachability probabilities for stochastic hybrid systems. In: *IEEE-EURASIP Int. Symp. on Control, Communications and Signal Processing*. Marrakech, Morocco.

Krystul, J., H.A.P. Blom and A. Bagchi (2006). Stochastic hybrid processes as solutions to stochastic differential equations. In: *Stochastic Hybrid Systems: Recent Developments and Research Trends* (C.G. Cassandras and J. Lygeros, Eds.). CRC Press.

Labeau, P.E., C. Smidts and S. Swaminathan (2000). Dynamic reliability: Towards an integrated platform for probabilistic risk assessment. *Reliability Engineering and System Safety* **68**, 219–254.

Le Gland, F. and N. Oudjane (2005). A sequential algorithm that keeps the particle system alive. In: *Proceedings of the 13th European Signal Processing Conference*. EURASIP. Antalya.

Le Gland, F. and N. Oudjane (2006). A sequential algorithm that keeps the particle system alive. In: *Stochastic Hybrid Systems : Theory and Safety Critical Applications* (H. Blom and J. Lygeros, Eds.). pp. 351–389. Number 337 In: *Lecture Notes in Control and Information Sciences*. SpringerVerlag. Berlin.

Lepeltier, J.P. and B. Marchal (1976). Probleme des martingales et equations differentielles stochastiques associees a un operateur integro-differentiel. *Ann. Inst. Henri Poincare* **12**(Section B), 43–103.

Lezaud, P., J. Krystul and H.A.P. Blom (2004). Accident risk assessment and Monte Carlo simulation methods. Hybridge Report D8.2. `http://www.nlr.nl/public/hosted-sites/hybridge/`.

Liu, J.S. (2003). *Monte Carlo Strategies in Scientific Computing*. Springer.

Marcus, S.I. (1978). Modeling and analysis of stochastic differential equations driven by point processes. *IEEE Trans. Information Theory* **24**, 164–172.

Mariton, M. (1990). *Jump Linear Systems in Automatic Control*. Marcel Dekker.

Moon, K.S. (2003). Adaptive Algorithms for Deterministic and Stochastic Differential Equations. PhD thesis.

Øksendal, B. (2002). *Stochastic Differential Equations. An Introduction with Applications*. Springer, 5th edition.

Pola, G., M. Bujorianu, J. Lygeros and M. D. Di Benedetto (2003). Stochastic hybrid models: An overview. In: *Proc. IFAC Conference on Analysis and Design of Hybrid Systems ADHS03*. Saint Malo, France. p. 4550.

Reich, P.G. (1964). *A theory of safe separation standards for Air Traffic Control*. Royal Aircraft Establishment. U.K. Technical Report 64041.

Rice, S.O. (1945). Mathematical analysis of random noise, Part II. *Bell System Technical Journal* **24**, 46–156.

RS (1992). *Risk: Analysis, Perception and Management*. Report of a Royal Society Study Group. London.

Shahabuddin, P. (1995). Rare event simulation in stochastic models. In: *Proc. 1995 Winter Simulation Conf.*. IEEE. Piscataway, NJ.

Smidts, C., J. Devooght and P.E. Labeau (1998). Dynamic reliability: Future directions. In: *International Workshop Series on Advanced Topics in Reliability and Risk Analysis*. Greenbelt. MD, USA.

Snyder, D.L. (1975). *Random Point Processes*. Wiley.

Tuckwell, H.C. and F.Y.M. Wan (1984). First-passage time of Markov processes to moving barriers. *J. Appl. Prob.* **21**, 695–709.

Villén-Altamirano, M. and J. Villén-Altamirano (1991). RESTART: A method for accelerating rare event simulation. In: *Queueing, Performance and Control in ATM: Proc. of the 13rd International Telegraffic Congress* (J.W. Cohen and C.D. Pack, Eds.). number 15 In: *North-Holland Studies in Telecommunications*. Amsterdam. pp. 71–76.

Villén-Altamirano, M. and J. Villén-Altamirano (1994). RESTART: A straightforward method for fast simulation of rare events. In: *Proc. of the 1994 Winter Simulation Conference* (J.D. Tew, M.S. Manivannan, D.A. Sadowski and A.F. Seila, Eds.). pp. 282–289.

Villén-Altamirano, M. and J. Villén-Altamirano (1997). RESTART: An efficient and general method for fast simulation of rare event. Technical report. Technical Report 7, Departamento de Mætmatica Aplicada, E.U. Inform'atica, Universidad Politécnica de Madrid.

Wonham, W.M. (1970). *Probability Analysis in Applied Mathematics*. Chap. Random differential equations in control theory. Vol. 2. Academic Press.

Index

Summary

Stochastic dynamical modelling of accident risk is of high interest for the safe design of complex safety-critical systems and operations, such as in nuclear and chemical industries, and advanced air traffic management. In comparison with statistical analysis of collected data, stochastic dynamical modelling approach has the advantage of enabling the use of stochastic analysis and advanced Monte Carlo simulation approaches.

Currently, the theory of stochastic differential equations has become a powerful tool for modelling real-world dynamical systems. In this thesis the stochastic approach to modelling of stochastic hybrid systems by Markov processes is presented. A very general class of jump-diffusions, which are defined as strongly unique semimartingale solutions of stochastic differential equations, is identified and characterized. These results are extended to two classes of stochastic hybrid processes which cover a range of interesting phenomena: switching diffusion, random hybrid jumps, boundary hybrid jumps and martingale inducing jumps. The Markov and the strong Markov properties are shown.

The numerical issues concerning time discrete approximations of stochastic hybrid processes are discussed. The stochastic Euler scheme for the switching diffusion with state dependent switching rates is presented. The proof of strong convergence of Euler scheme is given. The approximation of first passage times is discussed and the convergence is proved.

New version of Interacting Particle System (IPS) algorithm, called Hybrid IPS, is developed for the efficient estimation of rare event probabilities in stochastic hybrid systems. The HIPS approach is designed to cope with large differences in mode probabilities and rare switchings. Numerical evaluations on a test problem confirm the advantage of new HIPS algorithm over the standard Monte Carlo and the IPS simulation approaches.

Samenvatting

Stochastische dynamische modellering van risico van ongelukken is van groot belang voor het ontwerp van complexe veiligheids-kritische systemen zoals in de nucleaire en chemische industrie en geavanceerde luchtverkeersmanagement. In vergelijking met statistische analyse van vergaarde data heeft stochastische dynamische modellering als voordeel dat het stochastische analyse en geavanceerde Monte-Carlosimulatie mogelijk maakt.

De theorie van stochastische differentiaalvergelijkingen heeft zich ontwikkeld tot een krachtig gereedschap voor het modelleren van dynamische systemen uit de praktijk. Dit proefschrift presenteert een probabilistische benadering van modellering van stochastische hybride systemen d.m.v. Markov processen. Een zeer algemene klasse van sprong-diffusies, gedefinieerd als sterk-unieke semimartingaaloplossingen van stochastische differentiaalvergelijkingen, wordt geïdentificeerd en gekarakteriseerd. De resultaten zijn uitgebreid met twee klassen van stochastische hybride systemen die verscheidene interessante fenomenen bevatten: schakelende diffusies, random hybride sprongen, rand-hybride sprongen en martingaal-geïnduceerde sprongen. De Markoveigenschappen en sterke Markoveigenschappen zijn aangetoond.

Numerieke kwesties van tijddiscrete benaderingen van stochastische hybride processen worden beschouwd. Het stochastische Eulerschema voor schakelende diffusies met toestandsafhankelijke schakelingfrequenties wordt gepresenteerd. Een bewijs van sterke convergentie van het Eulerschema wordt gegeven. De benadering van de eerste passagetijd wordt bediscussieerd en de convergentie ervan is aangetoond.

Een nieuwe versie van het interacterende-deeltjessysteemalgoritme, hybride IPS-algoritme genoemd, is ontwikkeld voor efficiënte schatting van zeldzame gebeurtenissen in stochastische hybride systemen. Het hybride IPS-algoritme is ontworpen om overweg te kunnen met grote verschillen in geschatte conditionele kansen en zeldzame schakelingen. Numerieke evaluatie van een testprobleem bevestigt het voordeel het nieuwe hybride IPS-algoritme vergeleken met de standaard IPS-simulatie en Monte-Carlosimulatie.

Acknowledgements

I would like to express my gratitude to those who helped and assisted me in accomplishing this thesis.

First of all, I would like to express my sincere appreciation to my research supervisors, Prof. Arun Bagchi and Dr. Henk Blom, for their kind help, guidance, support and encouragement throughout the past four years.

I am grateful to all the members of my promotion committee: Prof. Arun Bagchi, Dr. Henk Blom, Prof. François Le Gland, Prof. Martin van Maarseveen, Dr. Peter Spreij, Prof. René Boel, Prof. Richard Boucherie and Dr. Michel Vellekoop for their attentive reading of a draft of this thesis and subsequently supplying me with numerous helpful comments and suggestions.

I would like to thank my former colleagues at National Aerospace Laboratory NLR in Amsterdam for their assistance with this work, particularly Bert Bakker, Marko van der Park and all other members of NLR's ATM group (Edwin, Mariken, Margriet, Sybert and Bart).

Dr. Pranab Mandal is greatly acknowledged for his careful reading of some of my papers and his insightful comments which helped to improve the papers greatly.

I also thank all faculty members and employees of the Systems, Signals and Control group for their pleasant company. In particular, I would like to thank Marja Langkamp and Marjo Mulder for their help in administrative matters, and Gjerrit Meinsma for writing the Dutch summary.

Also many thanks go to all my friends that I have met in The Netherladns. Special thanks to Viktor and Olga, Emad and Kitam, Arianto, Stefan, Zaher, Beata, Piotr, Vera, Alex, Maria and Maxim for the nice moments that we have spent together.

Finally, I reserve infinite thanks to my family and especially my wife Olga. It would be impossible to accomplish this work without their endless love, encouragement and support. I would like to dedicate this thesis to them.

Enschede, August 2006

Jaroslav Krystul