



HAL
open science

Particle methods in finance

Shohruh Miryusupov

► **To cite this version:**

Shohruh Miryusupov. Particle methods in finance. Computational Finance [q-fin.CP]. Université Panthéon-Sorbonne - Paris I, 2017. English. NNT : 2017PA01E069 . tel-03035021

HAL Id: tel-03035021

<https://theses.hal.science/tel-03035021v1>

Submitted on 2 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Centre d'économie de la Sorbonne
Labex RéFi

Doctorat en Mathématiques Appliquées

THÈSE

pour obtenir le grade de docteur délivré par

Université Paris 1 Panthéon-Sorbonne

Spécialité doctorale "Sciences, technologies, santé"

présentée et soutenue publiquement par

Shohruh MIRYUSUPOV

le 20 décembre 2017

Particle Methods in Finance

Directeur de thèse : **Raphael DOUADY**

La présidente : **Dominique GUEGAN**

Jury

M. Rama Cont,	Professeur	Rapporteur
M. Andrew Mullhaupt,	Professeur	Rapporteur
M. Pierre Del Moral,	Directeur de Recherche	Examineur
M. Ricadro Perez-Marco,	Directeur de Recherche	Examineur
M. Jean-Paul Laurent,	Professeur	Examineur

Université Paris 1 Panthéon-Sorbonne
Centre d'économie de la Sorbonne (CES)
UMR 8174, Paris, France

Résumé

This thesis consists of two parts, namely rare event simulation and a homotopy transport for stochastic volatility model estimation.

Particle methods, that generalize hidden Markov models, are widely used in different fields such as signal processing, biology, rare events estimation, finance and etc. There are a number of approaches that are based on Monte Carlo methods that allow to approximate a target density such as Markov Chain Monte Carlo (MCMC), sequential Monte Carlo (SMC). We apply SMC algorithms to estimate default probabilities in a stable process based intensity process to compute a credit value adjustment (CVA) with a wrong way risk (WWR). We propose a novel approach to estimate rare events, which is based on the generation of Markov Chains by simulating the Hamiltonian system. We demonstrate the properties, that allows us to have ergodic Markov Chain and show the performance of our approach on the example that we encounter in option pricing.

In the second part, we aim at numerically estimating a stochastic volatility model, and consider it in the context of a transportation problem, when we would like to find "an optimal transport map" that pushes forward the measure. In a filtering context, we understand it as the transportation of particles from a prior to a posterior distribution in pseudotime. We also proposed to reweight transported particles, so as we can direct to the area, where particles with high weights are concentrated. We showed on the example of Stein-Stein stochastic volatility model the application of our method and illustrated the bias and variance.

Keywords : Hamiltonian flow Monte Carlo, Particle Monte Carlo, Sequential Monte Carlo, Monte Carlo, rare events, option pricing, stochastic volatility, optimal transport

Table des matières

Table des matières	iii
Liste des figures	v
Liste des tableaux	vii
1 Introduction	1
1.1 Motivation	2
1.2 Sampling Methods	5
1.3 Particle Methods and Sequential Monte Carlo	6
1.4 Hamiltonian Flow Monte Carlo	10
1.5 Filtering by Optimal Transport	13
1.6 Organization of the thesis and main contributions	14
2 Particle Levy tempered α-stable credit risk model simulation with application to CVA with WWR computation in multi-factor BGM setting	21
2.1 Introduction	24
2.2 Problem Formulation and Definitions	24
2.3 Market Risk and Exposure Profiles	25
2.4 Credit Risk	27
2.5 WWR Modelling	37
2.6 Particle Interpretations	37
2.7 Numerical Analysis and Discussion	44
2.8 Conclusion	52
3 Hamiltonian flow Simulation of Rare Events	55
3.1 Introduction	57
3.2 Monte Carlo and Interacting Particle System	57
3.3 Hamiltonian flow Monte Carlo	59
3.4 Convergence Analysis	65
3.5 Applications and Numerical Results	67
3.6 Conclusion and Further Research	69
4 Optimal Transport Filtering with Particle Reweighting in Finance	73
4.1 Introduction	75
4.2 Particle Filtering	75
4.3 Homotopy Transport	79
4.4 Homotopy Transport with Particle Reweighting	83
4.5 Numerical Applications and Results	85
4.6 Conclusions and Further Research	88

Liste des figures

1.1	HMC sampling from Bivariate Gaussian distribution	13
2.1	Left : Positive exposure profiles. Right : EPE(Blue), ENE(Red)	27
2.2	The paths of CIR process (left) and survival function (right)	28
2.3	OU Jump-diffusion process with exponentially tempered α -stable jumps (left), Survival probability (right) with $\alpha = 0.8$	35
2.4	OU Jump-diffusion process with exponentially tempered α -stable jumps (left), Survival probability (right) with $\alpha = 1.5$	36
2.5	PMC simulation of OU Jump-diffusion process with exponentially tempered α -stable jumps (left), Survival probability (right) with $\alpha = 0.8$	43
2.6	PMC Euler, Euler jump-adapted, Milstein jump-adapted strong errors, $\alpha =$ 0.8 , #(particles)= 1000.	45
2.7	MC vs PMC Euler (left) and jump-adapted (right) strong errors. $\alpha = 0.8$, #(particles)= 1000, #(simulated paths)= 5000	46
2.8	MC vs PMC Euler/Milstein (left) and jump-adapted (right) paths. $\alpha = 1.5$, #(particles)= 1000, #(simulated paths)= 5000	47
2.9	EPE profile of the LIBOR swap (left), CVA process (right)	49
4.1	Volatility dynamics, PF (Blue), Homotopy (Red), RW-homotopy(Yellow)	86
4.2	Zoomed volatility dynamics. Homotopy (left), RW-homotopy (right)	87

Liste des tableaux

2.1	$\ \cdot\ _{L^2}$ error estimates of numerical schemes using MC and PMC density approximations, $\alpha = 0.8$	44
2.2	$\ \cdot\ _{L^1}$ error estimates of numerical schemes using MC and PMC density approximations . $\alpha = 0.8$	44
2.3	$\ \cdot\ _{L^2}$ error estimates of numerical schemes using MC and PMC density approximations, $\alpha = 1.5$	44
2.4	$\ \cdot\ _{L^1}$ error estimates of numerical schemes using MC and PMC density approximations . $\alpha = 1.5$	45
2.5	CVA WWR/free for LIBOR SWAP. PMC approach. $\alpha = 0.8$, $\#(scenarios) := \#(sc)$	51
2.6	LIBOR SWAP CVA WWR/free for different values of ρ and λ_0 , $\alpha = 0.8$. PMC approach.	51
2.7	LIBOR SWAP CVA WWR/free for different values of ρ and λ_0 , $\alpha = 0.8$. PMC approach.	51
3.1	DOC Barrier option estimates statistics. $B = 65, X_0 = 100, K = 100, r = 0.1, \sigma = 0.3, T = 1/2$, and $div = 0$; $\delta = 0.0001$, $\#(\text{Leap frog step}) : 35$. True price : 10.9064, $n_S = 50000, n_t = 750$	69
3.2	DOC Barrier option estimates statistics. $B = 65, X_0 = 100, K = 100, r = 0.1, \sigma = 0.3, T = 1/2$, and $div = 0$; $\delta = 0.0009$, $\#(\text{Leap frog step}) : 40$. True price : 10.9064, $n_S = 75000, n_t = 750$	69
4.1	Stein-Stein Stochastic volatility option price estimates statistics. $S_0 = 100, K = 90, r = 0.0953, \sigma = 0.2, \kappa = 4, \theta = 0.25, V_0 = 0.25, T = 1/2$, and dividends $d = 0$ True price : 16.05, $t = 20000, M = 64$	86
4.2	Stein-Stein Stochastic volatility option price estimates statistics. $S_0 = 100, K = 90, r = 0.0953, \sigma = 0.2, \kappa = 4, \theta = 0.25, V_0 = 0.25, T = 1/2$, and dividends $d = 0$ True price : 16.05, $t = 40000, M = 64$	87

Chapitre 1

Introduction

« " πάντες ἄνθρωποι τοῦ εἰδέναι ὀρέγονται φύσει. τὰ μετὰ τὰ φυσικά ' »

Αριστοτέλη

« "All men by nature desire to know" »

Aristotle

Sommaire

1.1 Motivation	2
1.1.1 Credit Risk Estimation	2
1.1.2 Pricing in Partial Observation Models	4
1.2 Sampling Methods	5
1.2.1 Monte Carlo	5
1.2.2 Importance Sampling	5
1.2.3 Markov Chain Monte Carlo	5
1.3 Particle Methods and Sequential Monte Carlo	6
1.3.1 Feynman-Kac Approximations	7
1.3.2 Hidden Markov Models	7
1.3.3 Sequential Monte Carlo	8
1.4 Hamiltonian Flow Monte Carlo	10
1.4.1 Sampling with HFMC	10
1.5 Filtering by Optimal Transport	13
1.6 Organization of the thesis and main contributions	14

1.1 Motivation

Throughout last 20 years development of computation power allowed to use sophisticated Monte Carlo methods in signal-processing, rare events estimation, computational biology, queuing theory, computational statistics and etc. In finance we have to deal with large dimensionality of problems, where other techniques due to some constraints, such as curse dimensionality, computational burden makes us to look for alternative numerical techniques. Particle methods are a broad class of interacting type Monte Carlo algorithms for simulating from a sequence of probability distributions satisfying a nonlinear evolution equation. These flows of probability measures can always be interpreted as the distributions of the random states of a Markov process whose transition probabilities depends on the distributions of the current random states ([6], [5]).

Particle methods have a variety of applications in finance : rare events and stochastic volatility estimation. For instance, portfolio managers need to estimate a large portfolio loss for risk management. In banking industry, banks need to compute default probabilities, so as to compute different value adjustments and to comply with financial regulation. In insurance industry, companies are interested in estimating ruin probabilities in a given time horizon. All the above mentioned cases are the examples, where rare event simulation is applied.

Stochastic volatility models are widely used in the financial industry. In fact, while realized volatility could be hedged away by trading other options, stochastic volatility models are needed to model the dynamics of implied volatilities, which will provide their user with simple break-even accounting conditions for the Profit and Loss of a hedged position [1].

The current thesis consists of two main parts, namely computing rare events via simulation and general stochastic volatility estimation.

1.1.1 Credit Risk Estimation

Since the credit crisis 2007-2009, the importance of counterparty credit risk for regulators increased dramatically. According to Basel III regulation [33], banks are required to hold a regulatory capital based on CVA charges against each of their counterparties. There is already a number of articles on CVA valuation, the most common Credit Value Adjustment (CVA) formula ([5], [22], [22], [39] and [40]) is given by :

$$CVA = (1 - R) \int_0^T D_t \mathbb{E}[V_t^+ | \tau = t] dPD(t) \quad (1.1)$$

where R is a recovery rate, D_t - risk-free discount rate, $PD(t)$ probability of default up to time t and V_t - the value of an underlying asset.

Two problems related to CVA computation that usually arise : incorporation of wrong way risk (WWR) into the value of CVA and a high computational burden. WWR is a statistical dependence between the exposure and a counterparty's credit risk (CCR). Another challenge related to CVA is the computation of default probability. There are two main approaches in default probability computation : structural models (first-passage approach) [3] and reduced form credit models ([19], [26] and [28]).

Structural models

The first passage approach is modelled like in a barrier option pricing, and formulated in the following way. Assume that the default barrier b is constant valued. Then the default time τ is a continuous random variable (r.v.) values in $]0, +\infty[$ and given by

$$\tau = \inf\{t > 0 : V_t < b\} \quad (1.2)$$

then the first passage probability is given by :

$$P(T) = \mathbb{P}(M_T < b) = \mathbb{P}(\inf_{s \leq T} (\mu s + \sigma_D W_s) < \log(b/V_0)) \quad (1.3)$$

where $M_T = \inf_{s \leq T} V_s$.

Default intensity

The main difference between reduced form and structural models is in the fact that, in the latest one does not need any economic model of firm's default, i.e. defaults are exogenous.

$$D_t = \mathbb{1}_{\tau < t} = \begin{cases} 1 & \text{if } \tau \leq t \\ 0 & \text{otherwise} \end{cases} \quad (1.4)$$

Observe that default intensity is an increasing process, such that the conditional probability at time t that the firm defaults at time $s \geq t$ is at least as big as the process D_t . A process with such property is called submartingale. The Doob-Meyer decomposition theorem says that we can isolate upward trend from D . This very important result says that there exist an increasing process D^τ starting at zero such that $D - A^\tau$ becomes a martingale. The unique process A^τ counteracts the upward trend in D , and it is called a compensator. It describes the cumulative and conditional likelihood of default, which is parameterized by non-negative process λ .

$$A_t^\tau = \int_0^{t^\tau} \lambda_s ds = \int_0^t \lambda_s \mathbb{1}_{\tau > s} ds \quad (1.5)$$

where λ_t describes conditional default rate for small interval Δt and $\tau > t$. $\lambda \Delta t$ approximates the probability that default occurs in the interval $]t, t + \Delta t]$.

Rare events

Rare events simulation is an important field of research in computational and numerical probability. It has a wide range of applications starting from catastrophe theory to finance, as an example one can consider pricing of a barrier option pricing or a credit risk estimation.

Let us consider the problem of rare event estimation, where the probability $\mathfrak{z} = \mathbb{P}(A)$ is very small. The crude Monte Carlo (CMC) estimates \mathfrak{z} through the proportion of times on the simulated points attain rare event area A over M independent trials :

$$\hat{\mathfrak{z}}_M = \frac{1}{M} \sum_{m=1}^M \mathbb{1}_{\xi^{(m)} \in A} \quad (1.6)$$

with variance $\sigma = \sqrt{\mathfrak{z}(1-\mathfrak{z})}$. In "easy problems" we can use the central limit theorem (CLT), which asserts that

$$\frac{1}{\sqrt{M}} (\hat{\mathfrak{z}}_M - \mathfrak{z}) \xrightarrow{M \rightarrow \infty} \sigma Z \quad (1.7)$$

where $Z \sim \mathcal{N}(0, 1)$ is a standard Gaussian distribution.

In a rare-events simulation we are not very concerned about the absolute error, but instead we would like to measure via relative error(RE) the precision of our estimation with respect to some "true" value of quantity we are computing. Relative error exposes the problems one encounters by using CMC in rare event setting :

$$\text{RE} = \frac{Z\sqrt{\mathfrak{z}(1-\mathfrak{z})}}{M^{1/2}\mathfrak{z}} = Z\sqrt{\frac{1-\mathfrak{z}}{M\mathfrak{z}}} \propto \frac{Z}{\sqrt{M\mathfrak{z}}} \xrightarrow{\mathfrak{z} \rightarrow 0} \infty \quad (1.8)$$

The approximation (1.8) shows that CMC requires $n \gg 1/\mathfrak{z}$. Lets show this on a simple example, assume that we would like to estimate rare event probability $\mathfrak{z} = 10^{-7}$, RE is targeted at 0.1 with a 95% confidence interval, then we have

$$\frac{1.96}{\sqrt{10^{-7}M}} \leq 0.1 \quad (1.9)$$

from the equation above we see that we need at least $M \geq 3.84 \times 10^{-9}$ samples to have a 10% RE.

Assume that we have a sequence of random variables $\{\xi(a)^{(m)}\}_{m=1}^M$. Define the rare event set as :

$$A^a = \{x \in \mathbb{R}^M, f(x) > a\} \quad (1.10)$$

then we can define a probability of interest and its estimator as

$$\mathfrak{z}(a) = \mathbb{E}[\mathbb{1}_{\xi(a) \in A^a}] \quad \text{and} \quad \hat{\mathfrak{z}}(a) = \frac{1}{M} \sum_{m=1}^M \mathbb{1}_{\xi^{(m)}(a) \in A^a} \quad (1.11)$$

We can show that if our estimator has a bounded RE, then the desired precision is independent of that the rarity of the set A^a .

$$\mathbb{P}\left(\frac{|\hat{\mathfrak{z}}(a) - \mathfrak{z}(a)|}{\mathfrak{z}(a)} > \epsilon\right) \leq \frac{\text{Var}(\xi(a))}{M\epsilon^2\mathfrak{z}(a)^2} \quad (1.12)$$

From the equation (1.12) we that the set will not depend on the "rarity" of the set A^a , if we impose the following condition :

$$\limsup_{a \rightarrow \infty} \frac{\text{Var}(\xi(a))}{\mathfrak{z}(a)^2} < \infty \quad (1.13)$$

1.1.2 Pricing in Partial Observation Models

Stochastic volatility models are the one of examples of the use of partial observation models (hidden Markov models) in finance, i.e. the situation, when one can observe the prices but not the dynamics of the stochastic volatility. In the last chapter we will show the application of different numerical approaches, such as particle filtering and homotopy transport on the example of a barrier option pricing in Stein-Stein stochastic volatility models. For example, in [1] author showed that the price of a barrier option is mostly dependent on the dynamics of the at-the-money skew conditional on price hitting the barrier. A stochastic volatility model for barrier options would need to provide a direct handle on this precise feature of the dynamics of the volatility surface so as to appropriately reflect its Profit and Loss impact in the option price. On of the first applications in option pricing could be found in [6]. Other applications and analysis of hidden Markov models could be found in [8] and [32].

1.2 Sampling Methods

1.2.1 Monte Carlo

The application of CMC in a rare event setting was already demonstrated in previous sections. These methods are generally used to compute integral that can not be calculated analytically or very difficult to compute. For example, we would like to calculate the following expectation :

$$\mathbb{E}^{\mathbb{P}}[h(X)] = \int_{\mathcal{X}} h(x)\mathbb{P}(dx) \quad (1.14)$$

All we need is to generate M independent and identically distributed (i.i.d.) random variables $\{\xi^{(m)}\}_{m=1}^M$ according to the law \mathbb{P} . Using the law of large numbers (LLN) we have an unbiased estimator

$$\frac{1}{M} \sum_{m=1}^M \delta_{\xi^{(m)}}(x) h(x) \xrightarrow[M \rightarrow \infty]{a.s.} \mathbb{E}^{\mathbb{P}}[h(X)] \quad (1.15)$$

As we saw in the previous section, often we can not efficiently sample from the measure \mathbb{P} . Those problems arises in rare events setting and sampling from fat-tailed distributions, then we have to use advanced Monte Carlo techniques, such as an importance sampling (IS), control variates, a stratified sampling and etc.

1.2.2 Importance Sampling

One of the ways to deal with rare events probabilities is IS Monte Carlo method. The idea is to change a measure from \mathbb{P} , where samples rarely reach rare event sets, to the measure \mathbb{Q} , so that we can sample at low cost in the new measure .

$$\mathbb{E}^{\mathbb{P}}[h(X)] = \int_{\mathcal{X}} h(x)\mathbb{P}(dx) = \int_{\mathcal{X}} h(x)\omega(x)\mathbb{Q}(dx) = \mathbb{E}^{\mathbb{Q}}[h(X)\omega(X)] \quad (1.16)$$

$$\sum_{m=1}^M \delta_{\xi^{(m)}}(x)\omega(x)h(x) \xrightarrow[M \rightarrow \infty]{a.s.} \mathbb{E}^{\mathbb{Q}}[h(X)\omega(X)] \quad (1.17)$$

The disadvantage of this method is in the fact, that we do not know the explicit form of the Radon-Nykodim derivative $\frac{d\mathbb{P}}{d\mathbb{Q}} = \omega$. This method is unfeasible in many examples, if we exclude some simple cases.

1.2.3 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) method allows to approximate a target measure π , by constructing an ergodic Markov Chain, that admits the target measure as a stationary law. We do not need to know the law explicitly as in the case of the IS, but instead we construct a Markov kernel \mathcal{K} , that leaves the target measure π invariant. There are two very popular MCMC sampling techniques : Metropolis-Hastings(MH) and Gibbs algorithms. One can show that the Gibbs algorithm is a special case of MH. The idea behind Metropolis-Hastings algorithm is to propose new set of candidates, and accept them, so that the transition kernel \mathcal{K} is left invariant with respect to the target, where Markov transition kernel is given by :

$$\mathcal{K}(x, dy) = \alpha(x, y)Q(x, dy) + \left(1 - \int \alpha(x, z)Q(x, dz)\right) \delta_x(dy) \quad (1.18)$$

and MH transition kernel \mathcal{K} is reversible, i.e. $\pi\mathcal{K} = \pi$. By recursion we can show that the chain $\{X_n\}_{n \geq 0}$ follows the law $\pi_n = \pi_0\mathcal{K}^n$. Using reversibility property of the kernel \mathcal{K} , we can have convergence of the chain X_n to the target measure π , given that the kernel \mathcal{K} has contraction :

$$\pi_n - \pi = (\pi_0 - \pi)\mathcal{K}^n \quad \forall n > 0 \quad (1.19)$$

If the chain X_l is irreducible, or in other words it admits a unique invariant measure, then the paths of the chain satisfies ergodic theory :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{l=n_H+1}^{n_H+n} h(X_l) = \mathbb{E}^\pi[h(X)] \quad (1.20)$$

where n_H is the number of iterations needed for "burn-out period", so that the chain leaves its initial law, and converges to the invariant measure π . Usually one needs to make high number of iterations, so that MH algorithm will approach the target density.

The performance of MCMC algorithms, like in other Monte Carlo sampling algorithms depends on the experience and ability to tune and optimize it. In general, the quality of final samples depends on the kernel \mathcal{K} and how fast it is able to explore the state space. In chapter 3 we will show, that the transition kernel represented by a metropolized Hamiltonian dynamics allows to explore it fast.

Algorithm 1 : MH MCMC algorithm

```

1 Initialization : N - #(time steps),  $\pi_0$  - initial measure
2  $X_0 \sim \pi_0$ ;
3 for  $n = 1, \dots, N$  do
4     Generate  $x^*$  from  $Q(\cdot, X_{n-1})$ ;
5     Compute the weight :
           
$$a(X_n, x^*) = 1 \wedge \frac{\pi(x^*)q(x^*, X_n)}{\pi(X_n)q(X_n, x^*)}$$

           assuming that  $a(x, y) = 0$  if  $\pi(x)q(x, y) = 0$ ;
6     Draw  $u \sim \mathcal{U}nif(0, 1)$ ;
7     if  $u < a$  then
8         Set  $X_{n+1} = x^*$ ;
9     else
10        Reject, and set  $X_{n+1} = X_n$ 
11    end
12 end
13 end

```

1.3 Particle Methods and Sequential Monte Carlo

Interacting particle system(IPS) is very power method, that is applied in non-linear filtering problems, rare event estimation and many other applications. IPS allows to overcome the problem that we face in importance sampling Monte Carlo technique, where we needed to know explicitly the importance sampling measure in order to sample from it. On contrary, the importance measure is approximated by a collection of trajectories of a Markov process $\{X_n\}_{n \geq 0}$ weighted by a set of potential functions ω_n . IPS is related to the unnormalized Feynman-Kac models. In the next section we give a brief overview of these stochastic models, for more details we refer to monographs [6] and [5].

1.3.1 Feynman-Kac Approximations

We give some key notations that are used to describe Feynman-Kac models, that we use in the next chapters. First, we define normalised and unnormalized Feynman-Kac measures (η_n, γ_n) for any bounded function f by the formulae :

$$\gamma_N(f) = \mathbb{E}[f(X_N) \prod_{n=0}^{N-1} \omega_n(X_n)] \text{ and } \eta_N(f) = \frac{\gamma_N(f)}{\gamma_N(\mathbb{1}_X)} \quad (1.21)$$

Since non-negative measures $(\eta_n)_{n \geq 0}$ satisfy for any bounded function f the recursive linear equation : $\gamma_n(f) = \gamma_{n-1}(Q_n(f))$ we have

$$\gamma_N(f) = \int \dots \int f(X_N) \eta_0(dx_0) \prod Q_n(x_n, dx_{n+1}) \quad (1.22)$$

where $\{Q_n\}_{n \geq 0}$ are unnormalised transition kernels, for example, we can choose them of the following form $Q_n(x_n, dx_{n+1}) = \omega_n(x_n) k_n(x_n, dx_{n+1})$, where $\{k_n\}_{n \geq 0}$ is a sequence of elementary transitions.

The following definition was given in [6], for a linear semigroup $Q_{m,n}$, $0 \leq m \leq n$ associated with a measure γ_n and defined by : $Q_{m,n} = Q_{m+1}, \dots, Q_n$. For any bounded function f and $(m, n) \in \mathbb{N}$:

$$\gamma_n(f) = \gamma_m Q_m \dots Q_{n-1} f \text{ and } \eta_n(f) = \frac{\gamma_m Q_m \dots Q_{n-1}(f)}{\gamma_m Q_m \dots Q_{n-1}(\mathbb{1}_X)} = \frac{\eta_m Q_m \dots Q_{n-1}(f)}{\eta_m Q_m \dots Q_{n-1}(\mathbb{1}_X)} \quad (1.23)$$

with a convention $Q_m, \dots, Q_n = Id$, if $m > n$ by definition of a normalised measure η_n and a semigroup $Q_{m,n}$ we readily obtain

$$\eta_{n+1}(f) = \frac{\eta_n Q_n(f)}{\eta_n Q_n(\mathbb{1}_X)} \quad (1.24)$$

Let us consider the example of rare events estimation, assume that we have a sequence of rare event sets $\{A_p\}_{0 \leq p \leq n}$, X is a random variable on the probability space $(\Omega, \mathbb{P}, \mathcal{F})$. One of the application in barrier option pricing or credit risk estimation is computation of conditional expectation

$$\eta_N(f) = \mathbb{E}[f(X_N) | X_p \in A_p] \quad (1.25)$$

Feynman-Kac interpretation for any bounded function f and potential function $\omega_n(x) = \mathbb{1}_{A_n}(x)$, using the fact that $\mathbb{1}_{A_n} \mathbb{1}_{A_{n+1}} = \mathbb{1}_{A_n}(x)$, since $A_{n+1} \subset A_n$

$$\eta_N(f) = \frac{1}{\mathbb{P}(A_N)} \int \mathbb{1}_{A_N}(x) h(x) \mathbb{P}(dx) = \frac{\mathbb{E}[f(X_N) \prod_{l=1}^{N-1} \omega_l(X_l)]}{\mathbb{E}[\prod_{l=1}^{N-1} \omega_l(X_l)]} = \frac{\mathbb{E}[f(X_N) \mathbb{1}_{A_N}(X)]}{\mathbb{E}[\mathbb{1}_{A_N}(X)]} \quad (1.26)$$

1.3.2 Hidden Markov Models

As we mentioned in the motivation section, hidden Markov Models (HMM) can be formulated in the context of a filtering problem. Under the filtering problem one assumes that we have a couple of processes in discrete time $(X_n, Y_n)_{n \geq 0}$, where $(X_n)_{n \geq 0}$ is a sequence of hidden variables and $(Y_n)_{n \geq 0}$ is partially observed data.

$$X_{n+1} | X_n = x_n \sim k_n(x_n, dx_{n+1}) \quad (1.27)$$

$$Y_n | X_n = x_n \sim \rho_n(x_n, dy_n) \quad (1.28)$$

where k_n is an elementary transition kernel, and ρ_n is a likelihood function. Using Bayes formula for a bounded function f we can derive

$$\mathbb{E}[f(X_0, \dots, X_n) | Y_0, \dots, Y_{n-1}] = \frac{\int \dots \int f(X_0, \dots, X_n) \prod \rho_p(X_p, Y_p) \nu(dx_0) k_0(X_0, dx_1) \dots k_{n-1}(X_{n-1}, dx_n)}{\int \dots \int \prod \rho_p(X_p, Y_p) \nu(dx_0) k_0(X_0, dx_1) \dots k_{n-1}(X_{n-1}, dx_n)} \quad (1.29)$$

Motivating example for studying and using HMM in finance is stochastic volatility model. It is defined by the following system :

$$\begin{aligned} dY_t &= \mu(X_t) dt + \sigma_Y(X_t) dW_t \\ dX_t &= \beta(X_t) dt + \sigma_X(X_t) dB_t \end{aligned} \quad (1.30)$$

Using Euler discretization scheme we have

$$\begin{aligned} Y_n &= Y_{n-1} + \mu(X_n) \Delta t + \sigma_Y(X_n) \sqrt{\Delta t} Z_n^{(1)} \\ X_{n+1} &= X_n + \beta(X_n) \Delta t + \sigma_X(X_n) \sqrt{\Delta t} Z_n^{(2)} \end{aligned} \quad (1.31)$$

where $Z^{(1)}$ and $Z^{(2)}$ are two independent standard Gaussian r.v., Δt is a discretization step.

In this case, the unnormalised transition kernel has the following form $Q_n(x_n, dx_{n+1}) = \rho_n(x_n, dy_n) k_n(x_n, x_{n+1})$. For the sake of illustration, assume that we have constant volatilities $\sigma_X(X_n) = \sigma_X$, $\sigma_Y(X_n) = \sigma_Y$, and other parameters are given by $\Delta t = 1$, $\beta(X_n) = 0$, $\mu(X_n) = X_n$ and $Y_{n-1} = 0$, then the likelihood is given by

$$\rho_n(x_n, dy_n) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(-\frac{(y_n - x_n)^2}{2\sigma_Y^2}\right) \quad (1.32)$$

and an elementary Markov transition kernel

$$k_n(x_n, dx_{n+1}) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x_{n+1} - x_n)^2}{2\sigma_X^2}\right) \quad (1.33)$$

If we exclude toy examples, there does not exist the exact simulation method for such type of models. Sequential Monte Carlo based particle algorithms allow to approximate normalised measures η_n .

1.3.3 Sequential Monte Carlo

Assume that on the measurable space (X, \mathcal{X}) , there exists an unnormalized transition $Q_n : X_n \times \mathcal{X}_n \rightarrow \mathbb{R}_+$, which is absolutely continuous with respect to a kernel $K_n : X_n \times \mathcal{X}_{n+1} \rightarrow \mathbb{R}_+$, i.e. $Q_n(x_n, \cdot) \ll K_n(x_n, \cdot)$ for all $x_n \in X_n$, then we can define an importance weight function ω_n :

$$\omega_n(x_n, x_{n+1}) = \frac{dQ_n(x_n, \cdot)}{dK_n(x_n, \cdot)}(x_{n+1}) \quad (1.34)$$

Sequential Monte Carlo (SMC) produces a fixed number of weighted samples (particles). At each time instant we generate a couple of particles and their corresponding weights $\{\xi_n^m, \omega_n^{(m)}\}_{m=1}^M$, where M is a fixed number of particles. If at time $n \in \mathbb{N}$, the set of weighted particles $\{\xi_n^{(m)}, \omega_n^{(m)}\}_{m=1}^M$ approximates measure η_n , then using IS method, we can also approximate η_{n+1} by a couple of the set of particles and their corresponding weights. $\{\xi_{n+1}^{(m)}, \omega_{n+1}^{(m)}\}_{m=1}^M$. This procedure is called a sequential importance sampling (SIS), which

allows us recursively construct the measure η_N from the initial measure $\eta_0 \in \mathcal{P}(\mathcal{X})$ and the set of importance weights $\{\omega_n^{(m)}\}_{m=1}^M$.

Algorithm 2 : Sequential Importance Sampling algorithm

```

1 Initialization : M - #(simulations), N - #(time steps),  $\eta_0$  - initial measure
2 for  $m = 1, \dots, M$  do
3    $\xi_0^{(m)} \sim \eta_0$ ;
4    $\xi_1^{(m)} \sim K_0(\xi_0^{(m)}, \cdot)$ ;
5    $\omega_1^{(m)} = \omega_0^{(m)}(\xi_0^{(m)}, \xi_1^{(m)})$ 
6 end
7 for  $n = 1, \dots, N$  do
8   for  $m = 1, \dots, M$  do
9     Generate  $\xi_n^{(m)}$  from  $K_n(\cdot, \xi_{n-1}^{(m)})$  and set  $\widehat{\xi}_n^{(m)} = (\widehat{\xi}_n^{(m)}, \xi_{n-1}^{(m)})$ ;
10    Compute the weight :  $\omega_n^{(m)}(\widehat{\xi}_n^{(m)})$ ;
11   end
12 end

```

In some cases, we can observe the weight degeneracy, when the variance of importance weights increase over time, as the fact that most of weights are negligible. To overcome this problem, in [24] authors proposed to resample particles at each iteration, which was called sequential importance resampling (SIRS) or bootstrap algorithm. There was an extensive research to optimize the bootstrap using, for example, effective sample size (ESS), we refer to [27], [16] and [14] for details.

In [6] and in [5] it was shown, that SMC estimators are unbiased, i.e. we can prove that

$$\sup_{\|f\| \leq 1} \left\| \mathbb{E} \left[\frac{1}{M} \sum_{m=1}^M f(\xi_N^{(m)}) - \eta_N(f) \right] \right\| \leq \frac{c(N)}{M} \quad (1.35)$$

where $c(N)$ is some positive constant, whose values depend on time horizon N .

Algorithm 3 : Sequential Importance Resampling algorithm

```

1 Initialization : M - #(simulations), N - #(time steps),  $\eta_0$  - initial measure
2 for  $m = 1, \dots, M$  do
3    $\xi_0^{(m)} \sim \eta_0$ ;
4    $\xi_1^{(m)} \sim K_0(\xi_0^{(m)}, \cdot)$ ;
5    $\omega_1^{(m)} = \omega_0^{(m)}(\xi_0^{(m)}, \xi_1^{(m)})$ 
6 end
7 for  $n = 1, \dots, N$  do
8   for  $m = 1, \dots, M$  do
9     if  $n < N$  then
10      Resample using probability weight :  $W_n(\widehat{\xi}_n^{(m)}) = \frac{\omega_n(\widehat{\xi}_n^{(m)})}{\frac{1}{M} \sum_{j=1}^M \omega_n(\widehat{\xi}_n^{(j)})}$ ;
11     end
12      Generate  $\xi_n^{(m)}$  from  $K_n(\cdot, \xi_{n-1}^{(m)})$  and set  $\widehat{\xi}_n^{(m)} = (\widehat{\xi}_n^{(m)}, \xi_{n-1}^{(m)})$ ;
13      Compute the weight :  $\omega_n^{(m)}(\widehat{\xi}_n^{(m)})$ ;
14   end
15 end

```

1.4 Hamiltonian Flow Monte Carlo

Hamiltonian Flow Monte Carlo (HFMC) methods came from statistical physics, where the computation of macroscopic properties requires sampling phase space configurations distributed according to some probability measure. One of its first application in sampling from high-dimensional distribution was proposed in [15], and its different extensions and converging properties in [2], [12] and [6]. Sampling on Riemann manifolds using generalised HFMC was introduced in [14].

We consider the system of particles described by a position and momentum X and P respectively, that are modelled by a Hamiltonian energy $\mathcal{H}(X, P)$. In statistical physics, the macroscopic properties can be obtained by averaging of some function A with respect to a probability measure \mathbb{P} describing the state of the system of particles :

$$\mathbb{E}^{\mathbb{P}} [A(X)] = \int_{\mathcal{X}} A(x, p) \mathbb{P}(dx, dp) \quad (1.36)$$

In most cases we approximate the above expectation numerically, such that as number of iteration increases, the microscopic sequences converge to the target distribution \mathbb{P} :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} A(X_n, P_n) = \mathbb{E}^{\mathbb{P}} [A(X)] \quad \mathbb{P} - a.s. \quad (1.37)$$

One of the simplest examples are so called canonical measures that have the following form :

$$\mathbb{P}(dx, dp) = \frac{1}{\mathcal{Z}} e^{-\beta \mathcal{H}(x, p)} dx dp, \quad \mathcal{Z} = \int_{\mathcal{X}} e^{-\beta \mathcal{H}(x, p)} dx dp \quad (1.38)$$

The measure \mathbb{P} is called Gibbs measure. Since we can separate potential and kinetic energies in the Hamiltonian, we can consider positions sampling, by using projection operator. Define the projection operator as $proj \circ \mathbb{P}(dx, dp) = \mathbb{P}(dx)$, and consider the canonical measure, ν , which is a projection of \mathbb{P} .

$$\nu(dx) = \frac{1}{\mathcal{Z}} e^{-\beta \Psi(x)} dx, \quad \mathcal{Z} = \int_{\mathcal{X}} e^{-\beta \Psi(x)} dx \quad (1.39)$$

If the Hamiltonian is separable, then the measure \mathbb{P} takes tensorized form, and each element of momenta follows independent Gaussian distributions. The main difficulty, that one encounters in computational statistics, biology is sampling of potential energy. The approaches to sample new configurations are as follows :

- "Brute force" sampling methods, such as rejection sampling
- MCMC techniques, when we accept or reject new proposals using MH algorithm
- Markovian stochastic dynamics, when we generate new samples using Langevin or generalized Langevin equations
- Deterministic methods on the extended state space

1.4.1 Sampling with HFMC

In chapter 3 we apply HFMC algorithms to estimate rare events. There are number of challenges to compute ensemble averages in eq. (1.37), for example the dynamics of X_n is not ergodic with respect to the Gibbs measure, i.e. paths are not sampled from ν . The system can contain high energy barriers that will prevent fast sampling.

A large deviation theory defines rare event, when the dynamics has to overcome a potential barrier $\Delta\Psi$, such that the exit time scales like

$$\tau \propto \exp(-\beta\Delta\Psi) \quad (1.40)$$

We can interpret τ in the sense that sampling from the Gibbs measure takes exponentially long time. There is the following law of large numbers.

Theorem 1.4.1 [30] *Given τ sufficiently small, let the numerical flow Ξ^τ be symmetric and symplectic. Then*

$$X_{n+1} = (\text{proj} \circ \Xi)(X_n, P_n), \quad P_n \propto e^{-\beta\mathcal{H}(X_n, \cdot)} \quad (1.41)$$

$$\mathbb{P}(X_{n+1} = x_{n+1}) = 1 \wedge e^{-\beta\Delta\mathcal{H}_{n+1}} \quad (1.42)$$

defines an irreducible Markov process $\{X_0, X_1, \dots\} \subset X$ with unique invariant probability measure π and the property

$$\frac{1}{N} \sum_{n=1}^N f(X_n) \rightarrow \int_{\mathbb{E}} f d\pi \quad \text{a.s.}, \quad \forall X_0 \in X \quad (1.43)$$

One can simply understand the irreducibility property as the probability of reaching any point on the configuration space is nonzero, i.e.

$$\mathbb{P}(X_{n+1} \in B(x_\tau) | X_n = x_0) > 0 \quad (1.44)$$

holds true $\forall x_0, x_\tau \in X$ and any Borel set $B = \mathcal{B}(X)$.

In algorithm 3, we show how we can sample using Hamiltonian dynamics, which consist of sampling new momenta proposals P according to the kinetic part of the canonical measure; performing L steps of numerical integration Ξ_n , which is the discretized version of Hamiltonian system, to obtain a new configuration (x^*, P^*) ; and finally computing $\text{prob} = \exp(-\beta(\mathcal{H}(x^*, P^*) - \mathcal{H}(x_n, P_n)))$ and accepting the new position X_{n+1} with probability

$a = \min(1, prob)$.

Algorithm 4 : Hamiltonian flow Monte Carlo algorithm

```

1 Initialization :  $n_S$  - #(simulations),  $n_t$  - #(time steps)
2 for  $n = 1, \dots, n_t$  do
3     for  $m = 1, \dots, n_S$  do
4         Generate  $X_n^{(m)}$  from prior  $\tilde{p}(X_0^{(m)}, \cdot)$ ;
5         Simulate initial momentum  $P_1^{(m)} \sim \mathcal{N}(0, I_M)$ , set  $x_H^{(m)} = X_n^{(m)}$  and run
           Hamiltonian flow;
6         for  $l_f = 1, \dots, L - 1$  do
7              $P_H^{(m)}((l_f + \frac{1}{2})\delta) = P_H^{(m)}(l_f) - \frac{\delta}{2} \frac{\partial \Psi}{\partial x_H}(x_H^{(m)}(l_f))$ 
            $x_H^{(m)}((l_f + 1)\delta) = x_H^{(m)}(l_f) + \delta P_H^{(m)}((l_f + \frac{1}{2})\delta) I_M^{-1}$ 
            $P_H^{(m)}((l_f + 1)\delta) = P_H^{(m)}((l_f + \frac{1}{2})\delta) - \frac{\delta}{2} \frac{\partial \Psi}{\partial x_H}(x_H^{(m)}((l_f + 1)\delta))$ 
8         end
9         Calculate acceptance probability and set  $x^* = x_H^{(m)}(L)$ ,  $P^* = P_H^{(m)}(L)$  :
           
$$a = 1 \wedge e^{(-\mathcal{H}(x^*, P^*) + \mathcal{H}(x_H^{(m)}, P_H^{(m)})) \Delta t} \tag{1.45}$$

           Draw  $u \sim \mathcal{U}nif(0, 1)$ ;
10        if  $u < a$  then
11            Set  $X_{n+1}^{(m)} = x^*$ ;
12        else
13            Reject, and set  $X_{n+1}^{(m)} = x_H^{(m)}$ 
14        end
15    end
16 end
17 end

```

As a simple example, consider sampling from a bivariate Gaussian distribution $p(x) = \mathcal{N}(\mu, \Sigma)$ with mean and covariance matrix

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \tag{1.46}$$

Then the potential energy function and partial derivatives are given by

$$\Psi(x) = -\log(p(x)) = \frac{x^T \Sigma x}{2}, \quad \frac{\partial \Psi}{\partial x_i} = x_i \tag{1.47}$$

Figure 1.1 shows samples from bivariate Gaussian distribution. We see that HFMC fastly explores the state space and samples from a target distribution.

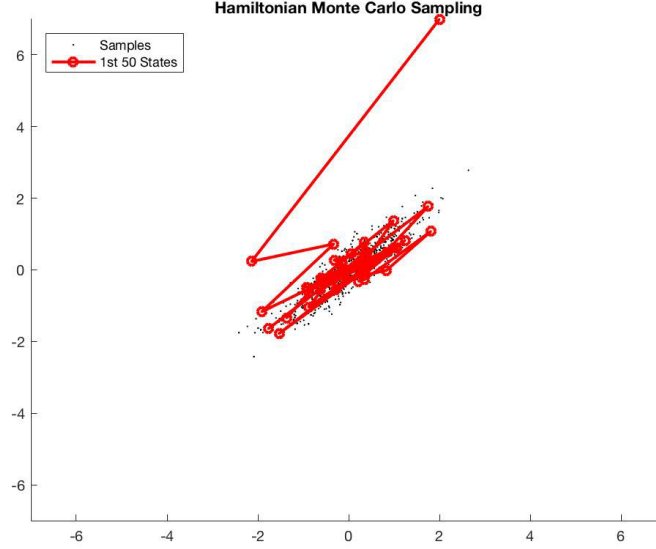


FIGURE 1.1 – HMC sampling from Bivariate Gaussian distribution

1.5 Filtering by Optimal Transport

In previous sections we showed how particle filters can be applied to approximate likelihood function by a set of weighted particles. It was shown in [1], [14], [15] that if the dimension of filtering problem increases, then the collapse of particles grows super - exponentially. In [29], authors introduced a new approach to update system's measurement, where the homotopy is formed, that gradually transforms an initial prior into a posterior density as scale parameter λ increases from 0 to 1. The idea is to choose a parameterized density and minimize the deviation between this density and homotopy, according to the measure of deviation. The approach can be interpreted as the optimal transport problem, which is defined as :

$$\begin{aligned} \inf_{\mathcal{T}} \mathbb{E}[|\mathcal{T}(X) - X|^2] \\ \text{s. th. } \mathbb{Q} = \mathcal{T}_{\#}\mathbb{P} \end{aligned} \quad (1.48)$$

That means that we try to approximate some importance measure \mathbb{Q} by an optimal transport map \mathcal{T} , that minimize the deviation between particles transported using the homotopy and posterior distribution. Instead of approximating the likelihood function by a set of particles sampled through importance resampling algorithm, we find a transport map that moves random variables using the following homotopy :

$$\psi(X_{t,\lambda} | \mathbb{Y}_t) = \frac{1}{\mathcal{Z}_\lambda} p(X_{t,\lambda} | \mathbb{Y}_{t-1}) \rho(Y_t | X_{t,\lambda})^\lambda \quad (1.49)$$

The homotopy continuously deforms a prior distribution $p(X_t | \mathbb{Y}_{t-1})$ into an unnormalized posterior $p(X_t | \mathbb{Y}_{t-1}) \rho(Y_t | X_t)$ as λ approaches to 1 :

$$\psi(X_{t,\lambda} | \mathbb{Y}_t) \xrightarrow{\lambda(0 \rightarrow 1)} \psi(X_t | \mathbb{Y}_t) \quad (1.50)$$

The transportation of particles in n_X dimensional state space is performed by a particle flow. The idea is to find a flow of particles that correspond to the flow of probability density defined by a homotopy. We suppose that flow of particles for Bayes rule follows the following dynamics in pseudo time λ :

$$dX_{t,\lambda} = g(X_{t,\lambda}) d\lambda + \eta(X_{t,\lambda}) dW_\lambda \quad (1.51)$$

where $g(X_{t,\lambda}) = \frac{dX_{t,\lambda}}{d\lambda}$.

In chapter 4 we will show that the flow $g(x, \lambda)$ can be found as a solution to Fokker-Plank partial differential equations, given that its diffusion matrix sums to zero, the flow is given by :

$$g(X_{t,\lambda}) = \left[\frac{\partial^2 \Psi(X_{t,\lambda})}{\partial X_{t,\lambda}^2} \right]^{-1} \left[\frac{\partial L(X_{t,\lambda})}{\partial X_{t,\lambda}} \right]^T \quad (1.52)$$

Algorithm 5 : Homotopy Transport Algorithm

```

1 Initialization :  $i = 1, \dots, n_X$  - #(simulations),  $t = 1, \dots, N$  - #(time steps)
2 Draw  $\{X_0^{(i)}\}_{i=1}^{n_X}$  from the prior  $p_0(x)$ .
3 Set  $\{\omega_0^{(i)}\}_{i=1}^{n_X} = \frac{1}{n_X}$ 
4 for  $t = 1, \dots, N$  do
5     for  $i = 1, \dots, n_X$  do
6         Propagate particles using state equation  $X_t^{(i)} = f(X_{t-1}^{(i)}, Y_{t-1}^{(i)}, \epsilon_t)$ ;
7         Measurement update :  $Y_t = h(X_t^{(i)}, Y_{t-1}^{(i)}, \eta_t)$ ;
8         Initialize pseudo-time  $\lambda = 0$ ;
9         Set  $X_{t,\lambda}^{(i)} = X_{t|n-1}^{(i)}$ ;
10        while  $\lambda < 1$  do
11            Compute SCM  $\hat{S}_M$ ;
12            Calculate an estimate :  $X_{t,\lambda} = \frac{1}{n_X} \sum_i X_{t,\lambda}^{(i)}$ 
13            Compute the matrix  $\hat{H} = \frac{\partial h(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}}$ ;
14            Update the time :  $\lambda = \lambda + \Delta\lambda$ ;
15            Calculate the flow  $\frac{dX_{t,\lambda}^{(i)}}{d\lambda} = - \left[ \frac{\partial^2 \Psi(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}^2} \right]^{-1} \left[ \frac{\partial L(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}} \right]^T$ ;
16            Transport particles according to its flow :  $X_{t,\lambda}^{(i)} = X_{t,\lambda}^{(i)} + \Delta\lambda \frac{dX_{t,\lambda}^{(i)}}{d\lambda}$ ;
17        end
18        Update state estimate :
19         $\tilde{X}_t = \frac{1}{n_X} \sum_{i=1}^{n_X} X_{t,\lambda}^{(i)}$ 
20    end
21 end

```

1.6 Organization of the thesis and main contributions

Current thesis consists of three articles, uploaded to HAL and Arxiv that we plan to submit. A brief description and main contributions are presented below.

Chapter 2. (Article) *Particle Lévy tempered α -stable credit risk model simulation with application to CVA with WWR computation in multi-factor BGM setting* (R. Douady, Sh. Miryusupov) This chapter presents a Lévy driven α -stable stochastic intensity model to estimate default probability in CVA computation. We use the fact that under certain regular conditions the Lévy process's small jump part could be approximated by a Brownian motion in order to correlate market and credit risks, hence to take into account the WWR. Since default probabilities are rare events, we adapt IPS methods to simulate Lévy stochastic intensity. We use three discretization schemes : Euler with constant time steps, Euler and Milstein jump-adapted time steps.

Chapter 3. (Article) *Hamiltonian flow Simulation of Rare Events* (R. Douady, Sh. Miryusupov) In this chapter we present an algorithm to estimate rare events using Hamiltonian dynamics. In this approach we generate Markov Chains that converges to an invariant measure. This approach allows to have rare-event estimates that have small variability. We compare our approach with IPS on the example of Barrier option.

Chapter 4. (Article) *Optimal Transport Filtering with Particle Reweighting in Finance* (R. Douady, Sh. Miryusupov) For the last portion of this thesis, we will move to partially-observed models in order to estimate stochastic volatility. We show the adaptation of particle filter to estimate the hidden parameter, which is not observed. The optimal transport allows the transportation of particles from prior distribution into posterior by using homotopy that gradually transforms the likelihood function, when we move in pseudo time from 0 to 1. In order to improve we propose reweighted particle transportation.

Bibliographie

- [1] Bergomi, L., Stochastic Volatility Modeling, CRC/Chapman & Hall, 2016.
- [2] Basel III : A global regulatory framework for more resilient banks and banking systems, BCBS, 2011
- [3] Black, Fischer & John C. Cox (1976), Valuing corporate securities : Some effects of bond indenture provisions, Journal of Finance 31, 351367.
- [4] P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions, Institute of Mathematical Statistics Collections, vol. 3, pp. 318329, 2008.
- [5] Canabarro, E., Counterparty Credit Risk, Risk Books, 2010. Recovering Volatility from Option Prices by Evolutionary Optimization
- [6] R Cont, Sana Ben Hamida, Recovering Volatility from Option Prices by Evolutionary Optimization, Journal of Computational Finance, Vol. 8, No. 4, Summer 2005
- [7] E Cances, F Legoll, G Stoltz ESAIM : Mathematical Modelling and Numerical Analysis 41 (2), 351-389
- [8] Cappe, O., Moulines, E., and Rydeen, T. (2005). Inference in Hidden Markov Models. Springer.
- [9] Daum, F, & Huang, J. (2013). Particle flow with non-zero diffusion for nonlinear filters. In Proceedings of spie conference (Vol. 8745).
- [10] Daum, F, & Huang, J. (2011). Particle degeneracy : root cause and solution. In Proceedings of spie conference (Vol. 8050).
- [11] Daum, F, & Huang, J. (2015). Renormalization group flow in k-space for nonlinear filters, Bayesian decisions and transport.
- [12] Del Moral, P : Nonlinear Filtering : Interacting Particle Solution(1996). Markov Processes and Related Fields 2 (4), 555-580
- [13] Del Moral, P : Feynman-Kac Formulae : Genealogical and Interacting Particle Systems with Applications. Probability and Applications. Springer, New York (2004).
- [14] Del Moral, P, Doucet, A., and Jasra, A. (2012). On adaptive resampling strategies for sequential Monte Carlo methods. Bernoulli, 18(1) :252 278.
- [15] Del Moral, P : Mean field simulation for Monte Carlo integration. CRC Press (2013)
- [16] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. Ann. Statist., 36(5) :2344 2376.
- [17] Duane, S, Kennedy, AD, Pendleton, BJ, and Roweth, D. Hybrid monte carlo. Physics letters B, 1987.
- [18] Durmus, A., Moulines, E. and Saksman, E. On the convergence of Hamiltonian Monte Carlo, arXiv preprint arXiv :1705.00166 (2017)

- [19] Duffie, Darrell & Kenneth J. Singleton, 'Modeling term structures of defaultable bonds', *Review of Financial Studies* 12, 687720, 1999.
- [20] Chen, Tianqi, Emily B. Fox, and Carlos Guestrin. "Stochastic Gradient Hamiltonian Monte Carlo." *ICML*. 2014.
- [21] Giesecke, Kay, *Credit Risk Modeling and Valuation : An Introduction* (June 23, 2004). Available at SSRN.
- [22] Gregory, J., *Counterparty Credit Risk and Credit Value Adjustment : A Continuing Challenge for Global Financial Markets*, Wiley, 2012.
- [23] Girolami, M. and Calderhead, B. (2011), Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73 : 123214. doi :10.1111/j.1467-9868.2010.00765.x
- [24] Gordon, N., Salmond, D., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F, Radar Signal Process.*, 140 :107 113.
- [25] Hairer, E. and Söderlind, G. Explicit, Time Reversible, Adaptive Step Size Control. *SIAM Journal on Scientific Computing*, 2005, Vol. 26, No. 6 : pp. 1838-1851
- [26] Jarrow, Robert A. & Stuart M. Turnbull, 'Pricing derivatives on financial securities subject to credit risk', *Journal of Finance*, 1995, 50(1), 5386.
- [27] Liu, J. and Chen, R. (1995). Blind deconvolution via sequential imputations. *J. Am. Statist. Assoc.*, 90(420) :567 576
- [28] Lando, David (1998), 'On cox processes and credit risky securities', *Review of Derivatives Research* 2, 99120.
- [29] U.D. Hanebeck, K. Briechle and A. Rauh, *Progressive Bayes : a New Framework for Nonlinear State Estimation*, in *Proc. SPIE 2003, Multi- source Information Fusion : Architectures, Algorithms, and Applications*, B.V. Dasarathy, Ed., vol. 5099, Orlando, FL, April 23, 2003, pp. 256267.
- [30] Hartmann, C. *J Stat Phys* (2008) 130 : 687.
- [31] El Moselhy, Tarek A. and Marzouk, Youssef M.(2012). Bayesian inference with optimal maps. *Journal of Computational Physics*. (Vol. 231)
- [32] MacDonald, I. and Zucchini, W. (2009). *Hidden Markov models for time series : an introduction using R*. CRC Press
- [33] Stein, Elias M, and Jeremy C Stein. 1991. Stock Price Distributions with Stochastic Volatility : An Analytic Approach. *Review of Financial Studies* 4 : 727-752.
- [34] Rubino, G., Tuffin, B. : *Rare event simulation using Monte Carlo methods*. Wiley (2009)
- [35] Beiglböck, M., Henry-Labordère, P. & Penkner, F. *Finance Stoch* (2013) 17 : 477. doi :10.1007 /s00780-013-0205-8
- [36] Neal, Radford M. *MCMC using Hamiltonian dynamics*. *Handbook of Markov Chain Monte Carlo*, January 2010.
- [37] Villani, C. : *Topics in optimal transportation*, Graduate studies in Mathematics AMS, Vol 58.
- [38] M. Pykhtin, D. Rosen, Pricing Counterparty Risk at the Trade Level and CVA Allocations. *Journal of Credit Risk*, 6(4), 2010, pages 3-38
- [39] Pykhtin, M. and Zhu, S., *A Guide to Modelling Counterparty Credit Risk*, GARP Risk Review, July/August 2007, 1622.

- [40] Prisco, B. and Rosen, D., Modeling Stochastic Counterparty Credit Exposures for Derivatives Portfolios, in Pykhtin, M. (Ed.), Counterparty Credit Risk Modelling : Risk Management, Pricing and Regulation, Risk Books, 2005.
- [41] Rachev, S. T. and Ruschendorf, L. : Mass Transportation Problems. In Vol. 1 : Theory. Vol. 2 : Applications. Springer, Berlin, 1998.
- [42] Risken, H. (1989). The Fokker-Planck Equation, second edn, Springer, Berlin, Heidelberg, New York.
- [43] S. Meyn and R. Tweedie. Markov Chains and Stochastic Stability. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
- [44] Tierney, Luke. Markov Chains for Exploring Posterior Distributions. Ann. Statist. 22 (1994), no. 4, 1701–1728. doi :10.1214/aos/1176325750. [http ://projecteuclid.org/euclid.aos/1176325750](http://projecteuclid.org/euclid.aos/1176325750).
- [45] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, Obstacles to high-dimensional particle filtering, Monthly Weather Review, vol. 136, no. 12, pp. 4629-4640, 2008.
- [46] F. Septier and G. W. Peters, An Overview of Recent Advances in Monte-Carlo Methods for Bayesian Filtering in High-Dimensional Spaces, in Theoretical Aspects of Spatial-Temporal Modeling, G. W. Peters and T. Matsui, Eds. SpringerBriefs - JSS Research Series in Statistics, 2015.

Chapitre 2

Particle Levy tempered α -stable credit risk model simulation with application to CVA with WWR computation in multi-factor BGM setting

« "Stabilité première condition du bonheur publique. Comment s'accommode-t-elle avec la perfectibilité indéfinie?" »

A.S. Pouchkine

« "Stability - the first condition of public happiness. How does it fit with indefinite perfectibility?" »

A.S.Puchkin

Sommaire

2.1 Introduction	24
2.2 Problem Formulation and Definitions	24
2.3 Market Risk and Exposure Profiles	25
2.3.1 BGM Model	25
2.3.2 Expected Exposure	26
2.3.3 Multi-curve Interest Rate Modelling	27
2.4 Credit Risk	27
2.4.1 Lévy driven OU Process	28
2.4.2 Simulation of α -Stable Processes and Numerical Schemes	32
2.5 WWR Modelling	37
2.6 Particle Interpretations	37
2.6.1 Interacting Particle System for Rare Events Estimation	37
2.6.2 Mean Field IPS Approximations	38
2.6.3 IPS Simulation of the Stochastic Intensity	41

*CHAPITRE 2. PARTICLE LEVY TEMPERED α -STABLE CREDIT RISK MODEL
SIMULATION WITH APPLICATION TO CVA WITH WWR COMPUTATION IN
MULTI-FACTOR BGM SETTING*

2.7 Numerical Analysis and Discussion	44
2.7.1 Numerical Schemes Error Estimates	44
2.7.2 CVA Computation	48
2.7.3 CVA. Particle Monte Carlo	50
2.7.4 Numerical Results	50
2.8 Conclusion	52

Abstract

Since the beginning of crisis in 2007 banks observed anomaly in the behaviour of the credit value adjustments (CVA) that was due to the Wrong Way Risk (WWR). The WWR is known as a statistical dependence between exposure and a counterparty credit risk (CCR). Most of approaches of dealing with the WWR are computationally intensive that makes it hard to use in the banks.

In this article we model CVA with the WWR, where a stochastic intensity follows a Levy-driven Ornstein-Uhlenbeck dynamics with jump marks that have α -stable distribution. We use the fact that small jumps of a Lévy process under certain regulatory conditions could be approximated by a Brownian motion with a drift. A Lévy process decomposition into a Brownian motion and a compound Poisson process allows naturally embed the WWR into the CVA values through the correlation of the Brownian motion in the dynamics of exposure and the one in the Levy processs Gaussian approximation.

To reduce the samples' variability, we used mean field interacting particle system (IPS), that allowed to reduce the variability of our CVA estimates on the one hand, and reduce the number of simulated paths on the other hand. This approach also allowed to reduce errors of sample paths using Euler with constant time step, Euler and Milstein scheme with jump-adapted time steps to several order of magnitudes compared to a crude Monte Carlo.

Our results show that the WWR risk has a huge impact on the values of CVA. We hope that our results will be a message for banks and regulators in CVA computation.

2.1 Introduction

After the crisis banks observed anomaly in the behaviour of CVA that was due to the WWR. The WWR is known as a statistical dependence between exposure and counterparty credit risk (CCR). Most of approaches of dealing with the WWR are computationally intensive that makes it hard to use in the banks. The WWR is a negative statistical dependency between exposure and a counterparty's credit quality. The CVA estimation with the WWR poses a major challenge with respect to both computational burden and tractability, consequently it is not clearly accounted in Basel III regulation.

There are different approaches that have been proposed to assess the WWR, for example in [13] they add jump into the exposure process. In [5] authors proposed the change of measure, that models the presence of the WWR. [15] proposed adjusted default probability in the dependent CVA formula and in [22] used copula method to model the dependence between default time and exposures. We focus on the dependence between the counterparty default and general market risk factors and in particular we use the properties of α -stable process to correlation the approximation of small jumps in the Lévy intensity and exposure processes.

In the following sections we focus on two issues : CVA with WWR estimation and IPS interpretation of α -stable process. One the one hand we propose α -stable intensity process for default probability estimation, on the other hand we would like to see the impact of WWR/RWR on the values of CVA.

Not surprisingly, we find that WWR depends on the correlation between counterparty default risk and the credit spread of the underlying asset, in particular, the regulatory ratio of 1.4 appears to be underestimated, when the correlation is negative.

This paper contributes as follows : first, we construct tempered- α stable stochastic intensity model for probability of default estimation ; second, we apply interacting particle system to estimate default probability in the rare events estimation context ; third, the IPS framework that we develop for simulation of α - stable process demonstrated reduced variability of its estimates compared to a crude Monte Carlo estimators ; forth, we apply IPS to estimate CVA with right/wrong way risk. Our results show that the WWR has a big impact on the values of CVA, taking into account that BCBSs regulatory CCR capital charges assume that WWR risk increasing by a constant factor value of CVA, our results show that this assumption has to be reviewed.

2.2 Problem Formulation and Definitions

The general formula for CVA computation is given by :

$$\text{CVA} := \text{LGDE}[B(0, \tau)V_{\tau}^{+} \mathbb{1}_{\tau \leq T}] \quad (2.1)$$

where $\text{LGD} = (1 - R)$, and LGD =loss given default, R is a recovery rate, $B(0, t)$ is a discount factor for maturity t . V_t is a market exposure at time t , $x^{+} = \max(x, 0)$ and τ is counterparty's default time. The expectation is under the risk neutral measure \mathbb{Q} .

Definition The counterparty positive exposure of a derivative security V_t^{+} , is the non-negative part of the difference between a security value S_t minus a collateral C_t :

$$V_t^{+} = \max(S_t - C_t, 0) \quad (2.2)$$

The expected positive exposure at each time instant t :

$$\text{EPE}_t = \mathbb{E}[V_t^{+} | \mathcal{F}_0] \quad (2.3)$$

We assume that $C_t = 0$.

Definition The default distribution function up to time t is given by :

$$\mathbb{P}(\tau \leq t) := \mathbb{E}[\mathbb{1}_{\tau \leq t}] \quad (2.4)$$

The density of the first default time is given by :

$$f_\tau(t) := \frac{\partial}{\partial t} \mathbb{P}(\tau \leq t) \quad (2.5)$$

Given that the default time τ and discounted positive exposures $B_t V_t^+$ are independent, CVA at time t can be computed as :

$$\text{CVA}^{ind} = \text{LGD} \int_0^T \mathbb{E}[B_t V_t^+] f_\tau(t) dt \quad (2.6)$$

BCBS in [33] defined the following formula for a potential mark-to-market losses associated with a deterioration in the credit worthiness of a counterparty :

$$\text{CVA} \approx \text{LGD}_{mkt} \underbrace{\sum_{i=1}^T \max(0; e^{-\frac{s_{i-1} t_{i-1}}{\text{LGD}_{mkt}}} - e^{-\frac{s_i t_i}{\text{LGD}_{mkt}}})}_{\text{Default probability term}} \underbrace{\left(\frac{\mathbb{E}E_{i-1} D_{i-1} + \mathbb{E}E_i D_i}{2} \right)}_{\text{Exposure term}} \quad (2.7)$$

Unilateral CVA

It is a good question, whether (2.7) makes any sense, since you have two different probability measures in one formula, on the one hand CDS spreads, that are calculated using risk-neutral probability measure, on the other hand you have expected exposure, which is computed using historical risk measure.

Next question, how the WWR is treated within banking regulation? In [16] and [15] it is indicated that Basel II defines $\alpha = 1.4$, but gives banks to estimate their own α , subject to a floor of 1.2. So CVA WWR is approximated using the following formula :

$$\text{CVA}^{\text{WWR,BIII}} = a \times \text{CVA} \quad (2.8)$$

In this article we model CVA with WWR, where stochastic intensity has Levy-driven Ornstein-Uhlenbeck dynamics. Since for a different level of α , the jump magnitude changes, we can see its impact on the CVA and, in particular, how it impacts CVA with WWR. The specific form of Levy process decomposition allows us to correlate two Brownian motions, the first one in the stochastic intensity and another one in the exposure process.

To reduce the variance and the number of generated path of MC simulations, we will use particle methods. To validate this approach we will verify the mean absolute deviation(MAD) and $\|\cdot\|_{L^2}$ norm of simulated paths' errors using Euler scheme with constant time steps, Euler with jump-adapted and Milstein with jump-adapted time steps using MC and PMC algorithms.

2.3 Market Risk and Exposure Profiles

2.3.1 BGM Model

We give some basic notions on BGM model, and remind the way one could compute the exposure profiles for multifactor BGM model [26], [4] and on the example of two factor model we will demonstrate expected exposure and expected positive exposure.

Denote by $P(t, T)$ the discount factor at time t with maturity T . The Libor rate is defined as :

$$L(t, T_i) = \frac{1}{\delta} \left(\frac{1}{P(t, T_i)} - 1 \right) \quad (2.9)$$

The forward rate from T_{i-1} to T_i , set at time t , where $t \leq T_{i-1} \leq T_i$ is given by :

$$L_i(t) := L(t, T_i, T_{i-1}) = \frac{1}{\delta} \left(\frac{P(t, T_{i-1})}{P(t, T_i)} - 1 \right) \quad (2.10)$$

From this we can derive a money market rate :

$$P(t, T_i) = \frac{P(t, T_{i-1})}{1 + \delta L_i(t)} = \prod_{j=1}^i \frac{1}{1 + \delta L_j(t)} \quad (2.11)$$

Since the forward rate is a martingale under the measure \mathbb{Q}^{T_i} , we have

$$\mathbb{E}^{\mathbb{Q}^{T_i}} [L(t, T_i) - L_i(t)] = 0 \quad (2.12)$$

and thus follows the dynamics

$$dL_i(t) = L_i(t) \zeta_i(t) dW_t^i \quad (2.13)$$

where W_t^i is a Brownian motion, σ_i is a bounded deterministic function. The forward measure dynamics, under a forward-adjusted measure \mathbb{Q}^{T_i} is given by :

$$dL_i(t) = \sum_{j=g(t)}^i \frac{\delta_j L_j(t) L_i(t) \zeta_i(t) \zeta_j(t)}{1 + \delta_j L_j(t)} dt + L_i(t) \zeta_i(t) dW_{t,q}^i \quad (2.14)$$

The interest rate swap value with the strike K is given by :

$$S_b(t_i) = \sum_{k=i}^{b-1} P(t_i, t_{k+1}) \delta_k (L_k(t_i) - K) \quad (2.15)$$

A multifactor BGM ([4], [1]) is given by the following formula :

$$dL_i(t) = \sum_{j=g(t)}^i \frac{\delta_j L_j(t) L_i(t) \sum_{q=1}^d \zeta_{i,q}(t) \zeta_{j,q}(t)}{1 + \delta_j L_j(t)} dt + L_i(t) \sum_{q=1}^d \zeta_{i,q}(t) dW_{t,q}^i \quad (2.16)$$

where d is the number of factors.

2.3.2 Expected Exposure

We show the exposure profiles computation on the example of a two factor BGM model for LIBOR rates, by simulating exposure profiles of LIBOR swaps with 5 year maturity, which coincides with the last exercise date and lockout after 3 months and tenor spacing of three months.

Using Euler's discretization scheme [29], we have :

$$L_i(t_{n+1}) = L_i(t_n) + \mu_i(L(t_n), t_n) L_i(t_n) \Delta t + L_i(t_n) \sqrt{\Delta t} \sum_{q=1}^2 \zeta_{i,q}(t_n) \epsilon_{n+1,q} \quad (2.17)$$

where $\Delta t = t_{n+1} - t_n$ with the drift term

$$\mu_i(L(t_n), t_n) = \sum_{j=g(t_n)}^i \frac{\delta_j L_j(t_n) \sum_{q=1}^2 \zeta_{i,q}(t_n) \zeta_{j,q}(t_n)}{1 + \delta_j L_j(t_n)} \quad (2.18)$$

Then for $m = 1, \dots, M$, at time t_l the positive (PE) and negative (NE) exposures are given by

$$V_i^{(m),+}(t_l) = \max(S_l^{(m)}(t_l), 0) \quad V_i^{(m),-}(t_l) = \min(S_l^{(m)}(t_l), 0) \quad (2.19)$$

The expected positive exposure (EPE) at each time instant t_l :

$$\text{EPE}_{t_l} = \frac{1}{M} \sum_{m=1}^M V_l^{(m),+}(t_l) \quad (2.20)$$

In figure 2.1. you can see the exposure profiles of 1000 different scenarios.

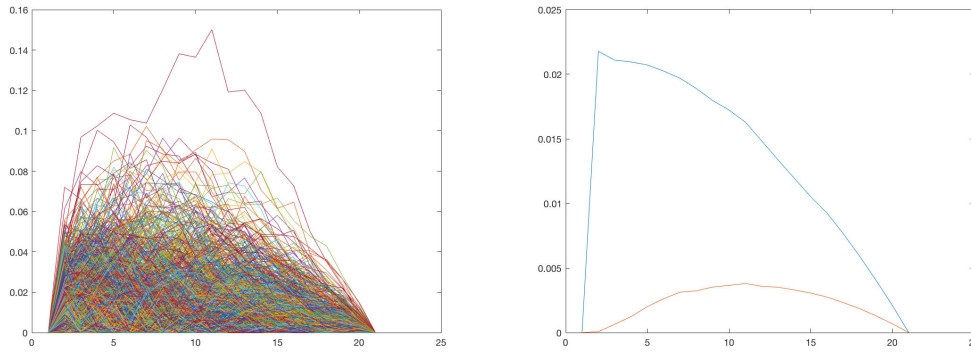


FIGURE 2.1 – Left : Positive exposure profiles. Right : EPE(Blue), ENE(Red)

2.3.3 Multi-curve Interest Rate Modelling

Before the crisis of 2007-2009 Euribor and Eonia overnight index swap (OIS) rates (a fixed swap rate that is paid/received in exchange for a floating rate tied to the Euro overnight rate Eonia), when both rates were assumed to be risk-free and followed each other closely. But during the crisis the situation radically changed and the spread between Euribor and Eonia OIS rates increased for all tenures. Since Euribor/Libor is not risk-free, many financial institutions moved away from the traditional discounting, instead Eonia is considered now as risk-free rate.

The previous assumption that one curve could be used for discounting and funding is not true anymore :

$$L_i(t) \neq \mathbb{E}^{Q^T_i} [L(t, T_i)] \quad (2.21)$$

2.4 Credit Risk

In this section we develop a Lévy driven α -stable stochastic intensity model. For the sake of completeness, we give some notions on Lévy processes, for more detailed treatment, we refer to [3],[17] and [14]. In the next sections, we provide some well-known properties and definitions on α -stable processes, then we show sufficient conditions for

a Lévy process to be decomposed into the sum of a drift, a Brownian motion and a compound Poisson process, that will be very important in imposing correlation structure of CVA with WWR.

In credit risk modelling there are two main approaches to model default probabilities. First one is the first passage approach, which is very similar to the first cross of a given barrier, the technique is widely used in a barrier option pricing and structural models developed by Merton. Second approach is a stochastic intensity model, which will be used for default probabilities computation in the next sections.

The first jump of the process can be simulated, conditional on the filtration \mathcal{F} as $\Lambda(\tau) =: \xi = \mathcal{E}xp(1)$, and by inversion we define the default time as : $\tau = \Lambda(\xi)^{-1}$.

The survival probability is given by :

$$P(\tau > t) = \mathbb{E}[e^{-\int_0^t \lambda_u du}] \quad (2.22)$$

The density of the first default time is given by :

$$f_\tau(t) = \mathbb{E}[\lambda_t e^{-\int_0^t \lambda_u du}] \quad (2.23)$$

As a simple example consider CIR process that describes the dynamics of the stochastic intensity λ_t

$$d\lambda_t = k(\phi - \lambda_t)dt + \sigma_\lambda \sqrt{\lambda_t} dW_t \quad (2.24)$$

The figure 1.7, shows the path of a CIR stochastic intensity and a survival function.

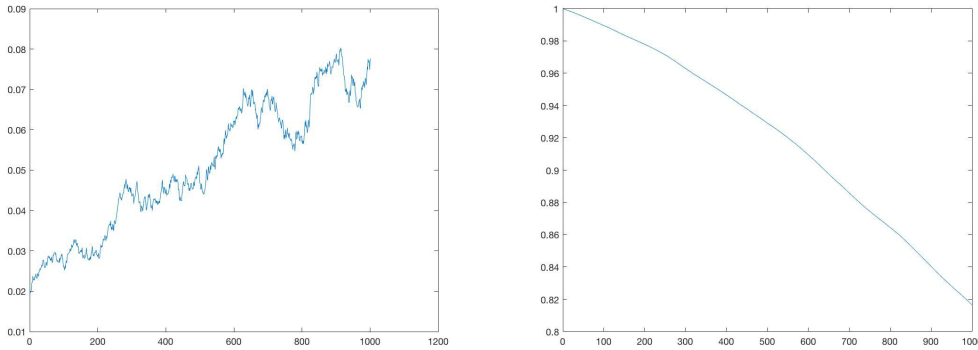


FIGURE 2.2 – The paths of CIR process (left) and survival function (right)

2.4.1 Lévy driven OU Process

Given that X is a Lévy process $X = \{X_t : t \geq 0\}$ in \mathbb{R}^d , then it is uniquely determined by its characteristic function in the Lévy-Khintchine form :

$$\mathbb{E}[e^{i\langle u, X_t \rangle}] = \exp \left(t \left[i\langle \mu, u \rangle + \int (e^{i\langle u, x \rangle} - 1 - i\langle u, x \rangle \mathbb{1}_{\|x\| \leq 1}) \nu(dx) \right] \right) \quad (2.25)$$

Let us consider the following Lévy driven OU process :

$$d\lambda_t = k(\phi - \lambda_t)dt + \int h_t(\lambda_{t-}, x) \rho(dx, dt) \quad (2.26)$$

where : $\rho(dx, dt)$ is a Poisson random measure that generates a sequence of pairs of random variables $\{\tau_i, \xi_i, i = 1, 2, \dots, \rho(T)\}$. The set of jump times is given by $\{\tau_i : \Omega \rightarrow \mathbb{R}_+, i =$

$1, \dots, \varrho(T)$ with intensity ll and jump sizes or marks are represented by the following set :
 $\{\xi_i : \Omega \rightarrow E, i = 1, \dots, \varrho(T)\}$.

If we assume that $h_t(\lambda_{t-}, x) = x$, then using Ito's formula, we can find the solution of (2.26) :

$$\lambda_t = \lambda_0 e^{-kt} + \phi(1 - e^{-kt}) + \int_0^t \int e^{k(s-t)} h_s(\lambda_{s-}, x) \varrho(dx, ds) \quad (2.27)$$

Equivalently we can rewrite it as :

$$\lambda_t = \lambda_0 e^{-kt} + \phi(1 - e^{-kt}) + \sum_{i=1}^{\varrho(t)} h(\lambda_{t-}, \xi_i) \quad (2.28)$$

α -Stable Processes

Definition X is α -stable if, $\forall n \in \mathbb{N}$, i.i.d. r.v. of X , $\sum_{i=1}^n X_i = bX + c$ for some constants $b = b(n) > 0$ and $c = c(n) \in \mathbb{R}^d$.

Definition A stable X is called α -stable, $\alpha \in]0, 2]$, whenever

$$X^{*r} \stackrel{d}{=} r^{1/\alpha} X + c \text{ and some constant } c = c(r) \in \mathbb{R}^d \quad (2.29)$$

Theorem 2.4.1 Let X be a non-trivial Lévy process in \mathbb{R} with generating triplet (μ, σ, ν) . Then X is α -stable for some $\alpha > 0$ iff exactly one of these conditions holds :

1. $\alpha = 2$ and $\nu = 0$.
2. $\alpha \in]0, 2[$, $\mu = 0$, and

$$\nu(dx) = (c_1 |x|^{-1-\alpha} \mathbb{1}_{x < 0} + c_2 |x|^{-1-\alpha} \mathbb{1}_{x > 0}) dx \text{ on } \mathbb{R} \text{ for some } c_1, c_2 > 0. \quad (2.30)$$

Definition A Lévy process X with Lévy triplet (μ, σ, ν) is said to be tempered α -stable if the Lévy measure ν is given by

$$\nu(dx) = ||x||^{-1-\alpha} \nu(x) dx \quad (2.31)$$

Gaussian Approximation of Lévy Driven Process

One of the main challenges to simulate Lévy processes roots in the fact that there is no general way to simulate its innovations. In this subsection we present the approach developed in [27] and [28], where necessary and sufficient conditions were established under which the small jumps of a one-dimensional and multidimensional Lévy process could be approximated by a Brownian motion. As we mentioned in the first section, this allow us to decompose a Lévy process into a Brownian motion and a compound Poisson process. Indeed, we will show that simulated paths of the Levy process resemble a Brownian motion for large values of stability index α and a compound Poisson process for smaller values.

Suppose that for $\epsilon \in]0, 1]$ we can decompose the Lévy measure ν as :

$$\nu = \nu_\epsilon + \nu^\epsilon \quad (2.32)$$

where

$$\int ||x||^2 \nu_\epsilon(dx) < \infty \text{ and } \nu^\epsilon(\mathbb{R}^d) < \infty \quad (2.33)$$

Assume that corresponding Lévy X process could be decomposed as :

$$X = X^{1,\epsilon} + \mathbf{P}^\epsilon + \mu_\epsilon \quad (2.34)$$

where $X^{1,\epsilon}$ is a Lévy process determined by :

$$\mathbb{E}[e^{i\langle u, X_t^{1,\epsilon} \rangle}] = \exp \left(t \left[i\langle \mu, u \rangle + \int (e^{i\langle u, x \rangle} - 1 - i\langle u, x \rangle \mathbb{1}_{\|x\| \leq 1}) \nu_\epsilon(dx) \right] \right) \quad (2.35)$$

$\mathbf{P}^\epsilon = \{\mathbf{P}_t^\epsilon : t \geq 0\}$ is a compound Poisson process with jump measure ν^ϵ and $\mu_\epsilon = \{t\mu_\epsilon : t \geq 0\}$ is a drift, i.e.

$$\mu_\epsilon = \mu + \int_{\|x\| \geq 1} x \nu_\epsilon(dx) - \int_{\|x\| \leq 1} x \nu^\epsilon(dx) \quad (2.36)$$

$X^{1,\epsilon}$ is a "small-jump" part of the process X . Next theorem says that, the small-jump part $X^{1,\epsilon}$ can be approximated by $\sqrt{\sigma_\epsilon}W$, where $W \in \mathbb{R}^d$ is a Brownian motion, independent of \mathbf{P}^ϵ . So, we have :

$$X \stackrel{\mathcal{D}}{=} \mu_\epsilon + \sqrt{\sigma_\epsilon}W + \mathbf{P}^\epsilon := X^\epsilon \quad (2.37)$$

Next results from [2](Theorem 2.1) gives necessary and sufficient conditions that guarantees Gaussian approximation of small jumps.

Theorem 2.4.2 $\sigma(\epsilon)^{-1}X^{1,\epsilon} \xrightarrow{\mathcal{D}} W$ as $\epsilon \rightarrow 0$, if and only if for each $\kappa > 0$,

$$\sigma(\kappa\sigma(\epsilon) \wedge \epsilon) \sim \sigma(\epsilon), \text{ as } \epsilon \rightarrow 0 \quad (2.38)$$

If we do not consider toy examples, condition (2.38) is difficult to verify. The next proposition from [2] gives a sufficient condition for eq.(2.38).

Proposition 2.4.3 Condition (2.38) is implied by

$$\lim_{\epsilon \rightarrow 0} \frac{\sigma(\epsilon)}{\epsilon} = \infty \quad (2.39)$$

Example Consider a symmetric α -stable process X that satisfies conditions in theorem 4.1 with Lévy measure given by :

$$\nu(dx) = \left(\frac{c_1}{|x|^{1+\alpha}} \mathbb{1}_{x < 0} + \frac{c_2}{x^{1+\alpha}} \mathbb{1}_{x > 0} \right) dx \quad (2.40)$$

where $c_1 + c_2 > 0$, then

$$\sigma(\epsilon)^2 = \sigma_\epsilon^2 = \int_{|x| \leq \epsilon} x^2 \nu(dx) = \frac{c_1 + c_2}{2 - \alpha} \epsilon^{2-\alpha} \quad (2.41)$$

Now we can check the condition (2.39) :

$$\frac{\sigma(\epsilon)}{\epsilon} = \sqrt{\frac{c_1 + c_2}{2 - \alpha}} \epsilon^{-\alpha/2} \xrightarrow{\epsilon \rightarrow 0} \infty \quad (2.42)$$

So, this condition is satisfied.

Example Another example, that we apply in the next sections, is exponentially tempered α -stable process with Lévy triplet $(0, 0, \nu)$, where Lévy measure ν is given by :

$$\nu(dx) = \left(\frac{e^{-|x|}}{|x|^{1+\alpha}} \mathbb{1}_{]-\infty, 0]} + \frac{e^{-x}}{x^{1+\alpha}} \mathbb{1}_{]0, \infty[} \right) dx \quad (2.43)$$

$$\mu_\epsilon = 0 - \int_{\epsilon < |x| < 1} x \nu(dx) \quad (2.44)$$

The variance of small jumps $\sigma(\epsilon)^2 = \sigma_\epsilon^2$.

$$\sigma(\epsilon)^2 = \sigma_\epsilon^2 = \int_{|x| < \epsilon} x^2 \nu(dx) = \int_{-\epsilon}^0 x^2 \frac{e^x}{(-x)^{1+\alpha}} dx + \int_0^\epsilon x^2 \frac{e^{-x}}{x^{1+\alpha}} dx = 2 \int_0^\epsilon x^{1-\alpha} e^{-x} dx \quad (2.45)$$

Next, using Karamata's integral theorem, which says that the integrals of regularly varying functions are again regularly varying[31]. For $\alpha > -1$,

$$\int_0^\epsilon x^{1-\alpha} e^{-x} dx \sim \frac{\epsilon^{2-\alpha}}{2-\alpha} \quad \text{for } \epsilon \rightarrow \infty \quad (2.46)$$

So, we have

$$\sigma(\epsilon) \sim \sqrt{\frac{2}{2-\alpha}} \epsilon^{1-\alpha/2} \quad (2.47)$$

and

$$\frac{\sigma(\epsilon)}{\epsilon} = \sqrt{\frac{2}{2-\alpha}} \epsilon^{-1/2-\alpha/2} \xrightarrow{\epsilon \rightarrow 0} \infty \quad (2.48)$$

Consequently, condition (2.39) is satisfied.

Jump-Diffusion Approximation of OU Lévy Process

A Lévy process decomposition allows us to approximate Lévy driven SDE by a jump-diffusion SDE. In this section we get "almost exact" solution that allows to compare errors of different numerical schemes under Monte Carlo and IPS approximations. Consider the following OU SDE :

$$d\lambda_t = \mu(\lambda_t) dt + h(\lambda_{t-}) dX_t \quad (2.49)$$

where X is a Lévy process with a triplet $(0, 0, \nu)$. Let $\mu(\lambda_t) = k(\phi - \lambda_t)$, $\sigma(\lambda_s) = 0$ and $h(\lambda_{s-}) = 1$, then the solution of (2.49) is

$$\lambda_t = \lambda_0 e^{-kt} + \phi(1 - e^{-kt}) dt + \int_0^t e^{k(s-t)} dX_s \quad (2.50)$$

$$X_t = \int_0^t \int_{\|x\| \leq 1} x \tilde{\varrho}(dx, ds) + \int_0^t \int_{\|x\| \geq 1} x \varrho(dx, ds) \quad (2.51)$$

The approximation of a Lévy process X_t is given by

$$X_t^\epsilon = \int_0^t \int_{\|x\| \leq \epsilon} \sigma_\epsilon \tilde{\varrho}^\epsilon(dx, ds) + \int_0^t \int_{\|x\| \geq \epsilon} x \varrho^\epsilon(dx, ds) \quad (2.52)$$

and the approximated solution of eq. (2.49) is

$$\lambda_t^\epsilon = \lambda_0 e^{-kt} + \phi(1 - e^{-kt}) dt + \int_0^t e^{k(s-t)} dX_s^\epsilon \quad (2.53)$$

Then the approximated version of the Lévy-driven SDE have the following form

$$\lambda_t^\epsilon = \lambda_0 e^{-kt} + \left(\phi(1 - e^{-kt}) + \mu_\epsilon \right) dt + \int_0^t e^{k(s-t)} \sqrt{\sigma_\epsilon} dW_s + \int_0^t \int_{\|x\| \geq \epsilon} e^{k(s-t)} x \mathbb{Q}^\epsilon(dx, ds) \quad (2.54)$$

The drift of the Lévy approximation $\mu_\epsilon = 0$, i.e.

$$\mu_\epsilon = 0 - \int_{\epsilon < |x| \leq 1} x \nu(dx) = - \int_{-1}^{-\epsilon} x \frac{e^x}{(-x)^{1+\alpha}} dx + \int_{\epsilon}^1 x \frac{e^{-x}}{x^{1+\alpha}} dx = 0 \quad (2.55)$$

2.4.2 Simulation of α -Stable Processes and Numerical Schemes

In previous sections we saw that any Lévy process X_t with a generating triplet $(\mu, 0, \nu)$ can be decomposed into the sum of two independent Levy processes :

$$X_t = X_t^{1,\epsilon} + P_t^\epsilon \quad (2.56)$$

The Gaussian approximation of "small-jump" part $X_t^{1,\epsilon}$ allows us to get the following approximation of Lévy process in equation (2.56) :

$$X_t \stackrel{\mathcal{D}}{=} P_t^\epsilon + \sigma_\epsilon W_t \quad (2.57)$$

To get the best approximation of X , we need to choose ϵ experimentally and find an efficient way for generating a compound Poisson process P^ϵ as $\epsilon \rightarrow 0$. That can be done using a series representation. Rosiński showed that Lévy process X_t with generating triplet $(\mu, 0, \nu)$ can be represented almost surely as a uniform in t convergent series of the following form :

$$X_t = \sum_{j=1}^{\infty} \left[H(\Gamma_j, \xi_j) \mathbb{1}_{u_j \leq t} - t \mu_j \right], \quad 0 \leq t \leq T \quad (2.58)$$

where $\{\Gamma_j\}$ are arrival times in a Poisson process with rate one, $\{\xi_j\}$ are i.i.d. jump marks taking values in some Euclidean space, $\{u_j\}$ are uniform i.i.d. r.v. on the interval $[0, T]$ and the sequences $\{\Gamma_j\}$, $\{\xi_j\}$ and $\{u_j\}$ are independent from each other. The function H is a jointly measurable real-valued function, such that the map $r \mapsto |H(r, \nu)|$ is nonincreasing and $\mu_j \in \mathbb{R}$. According to [27], the choice of H and ξ is not unique, but they have to satisfy to condition :

$$\int_{\mathbb{R}} f(x) \nu(dx) = T^{-1} \int_0^\infty \mathbb{E}[f(H(r, \xi))] dr \quad (2.59)$$

for any nonnegative Borel function f with $f(0) = 0$. For the case of P^ϵ we have the following series representation :

$$P_t^{\epsilon, \tau} = \sum_{j=\Gamma_j \leq \tau} \left[H(\Gamma_j, \xi_j) \mathbb{1}_{u_j \leq t} - t \check{\mu}(\tau) \right], \quad 0 \leq t \leq T \quad (2.60)$$

where

$$\check{\mu} = \int_0^s \int_{\|x\| \leq 1} x \vartheta(r, dx) dr, \quad s \geq 0 \quad (2.61)$$

for $B \in \mathcal{B}(\mathbb{R})$, the measure ϑ is defined by :

$$\vartheta(r, B) = \mathbb{P}(H(r, \xi_i) \in B), \quad r > 0, \quad \text{and} \quad \nu(B) = \int_0^\infty \vartheta(r, B) dr \quad (2.62)$$

α - Stable Process Simulation

Series representation of a Lévy processes provides a uniform approximation of a Lévy process along the sample paths. Next step is the simulation of the Lévy process based on this representation. A detailed treatment of stable processes can be found in [28], we use a series representation in the case of tempered- α stable processes. Let $\Gamma_j, U_j, \Upsilon_j, \Xi_j$ and ξ_j be sequences of i.i.d. random variables, such that $\mathbb{P}(\xi_j = \pm 1) = 0.5$, Υ_j are exponentially distributed with rate one, Ξ_j are uniformly distributed on $]0, 1[$, then we have

$$P_t^e = \sum_{j=1}^{\infty} \xi_j \left(\left[\frac{\alpha \Gamma_j}{2\kappa T} \right]^{-\frac{1}{\alpha}} \wedge \Upsilon_j \Xi_j^{\frac{1}{\alpha}} \right) \mathbb{1}_{U_j \leq t}, \quad 0 \leq t \leq T \quad (2.63)$$

For the compound Poisson process $P^{e,\tau}$, we simulate it using the following series :

$$P_t^{e,\tau} = \sum_{j=\Gamma_j \leq \tau} \xi_j \left(\left[\frac{\alpha \Gamma_j}{2\kappa T} \right]^{-\frac{1}{\alpha}} \wedge \Upsilon_j \Xi_j^{\frac{1}{\alpha}} \right) \mathbb{1}_{U_j \leq t}, \quad 0 \leq t \leq T \quad (2.64)$$

Details are given in the Algorithm 6.

Algorithm 6 : α -stable process generation using series representation

```

1 Initialization : M - #(simulations), N - #(time steps), j = 0, T^n = {T_1, ..., T_n}
2 s = -log(rand);
3 while s < tau do
4     j = j + 1;
5     tau(j) = -log(rand);
6     s = s + tau(j);
7     Poisson arrival times t(j) = s; Simulate jump times t_j(j) = rand;
8     Simulate random signs for symmetricity :
9     r_s = rand;
10    if r_s < 0.5 then
11        | dr(j) = 1;
12        | else
13        | | dr(j) = -1;
14        | end
15    end
16    Simulate interarrival times : Y(j) = -1/lambda log(rand);
17    Xi(j) = rand;
18 end
19 for j = 1 : T_n do
20     while t_j(j) < T_j do
21         | Compute sum_{n=1}^j xi_n ( [ alpha(n) / (2*kappa*T) ] ^ (-1/alpha) wedge Y_n Xi_n^(1/alpha) );
22     end
23 end

```

Numerical Schemes

In this section we show the way SDE in eq.(2.49) could be solved numerically using three discretization schemes : Euler with a constant time step, Euler and Milstein with

jump-adapted time steps. Now consider the following Lévy-driven SDE :

$$d\lambda_t = k(\phi - \lambda_t)dt + \sigma_\epsilon dW_t + \int_{\|x\| \geq \epsilon} x \varrho^\epsilon(dx, ds) \quad (2.65)$$

The simplest way to discretize it is the *Euler scheme*.

$$\begin{aligned} \lambda_{t_{n+1}} &= \lambda_{t_n} + k(\phi - \lambda_{t_n})\Delta t + \sigma_\epsilon \Delta W_n + \int_{t_n}^{t_{n+1}} \int_{|x| > \epsilon} x \varrho(dx, ds) = \\ &= \lambda_{t_n} + k(\phi - \lambda_{t_n})\Delta t + \sigma_\epsilon \Delta W_n + \sum_{k=\varrho(t_n)+1}^{\varrho(t_{n+1})} \xi_k \end{aligned} \quad (2.66)$$

where $\Delta t = t_{n+1} - t_n$ and $\Delta W_n = W_{t_{n+1}} - W_{t_n} \sim \mathcal{N}(0, \Delta t)$ is n -th increment of the Brownian motion process.

A *jump-adapted Euler* scheme was introduced in [21]. Define $\lambda_{t_{n+1}-} = \lim_{s \uparrow t_{n+1}} \lambda_s$ as the left limit of the process λ at time t_{n+1} .

Consider the following equidistant discretization of the interval $[0, T] : 0 = t_0, \dots, < t_n = T$ and let the jump times to be denoted as $\tau_1, \dots, \tau_{\varrho(T)}$, then the jump-adapted discretization be $\{t_0, t_1, \dots, t_{\varrho(T)}\} = \{t_0, t_1, \dots, t_N\} \cup \{\tau_0, \tau_1, \dots, \tau_{\varrho(T)}\}$.

The process before jump is given by

$$\lambda_{t_{n+1}-} = \lambda_{t_n} + \mu \Delta t_n + \sigma \Delta W_{t_n} \quad (2.67)$$

and the process after the jump

$$\lambda_{t_{n+1}} = \lambda_{t_{n+1}-} + \int h(\lambda_{t_{n+1}-}, x) \varrho(dx, \{t_{n+1}\}) \quad (2.68)$$

Note that the diffusion approximation is given by (2.67), and the impact of jump is simulated in the eq. (2.68). We can represent the two equations in a simple form :

$$\begin{cases} \lambda_{t_{n+1}} = \lambda_{t_{n+1}-} + h(\lambda_{t_{n+1}-}, \xi_{\varrho(t_{n+1})}) & \text{if jumps at } t_{n+1} \\ \lambda_{t_{n+1}} = \lambda_{t_{n+1}-} & \text{otherwise} \end{cases} \quad (2.69)$$

where $\int h(\lambda_{t_{n+1}-}, x) \varrho(dx, \{t_{n+1}\}) = 0$, if t_{n+1} is not a jump time.

We can achieve a strong order - 1 convergence, if we use a Milstein scheme. The process after the jump has the same form as in (2.68).

$$\lambda_{t_{n+1}-} = \lambda_{t_n} + \mu \Delta t_n + \sigma \Delta W_{t_n} + \frac{\sigma^2}{2} ((\Delta W_{t_n})^2 - \Delta t_n) \quad (2.70)$$

In the figures 3 and 4 we see the sample paths of α -stable process with values of $\alpha = 0.8$ and $\alpha = 1.5$ and survival probabilities. We simulated so-called almost exact solution given in eq. (2.54) and compared with three discretization schemes : Euler with constant discretization steps Δt , Euler and Milstein jump-adapted versions. You can observe that for larger values of α , the process resembles Brownian motion and compound Poisson process for low values, that gives the intuition behind the decomposition in eq. (2.56).

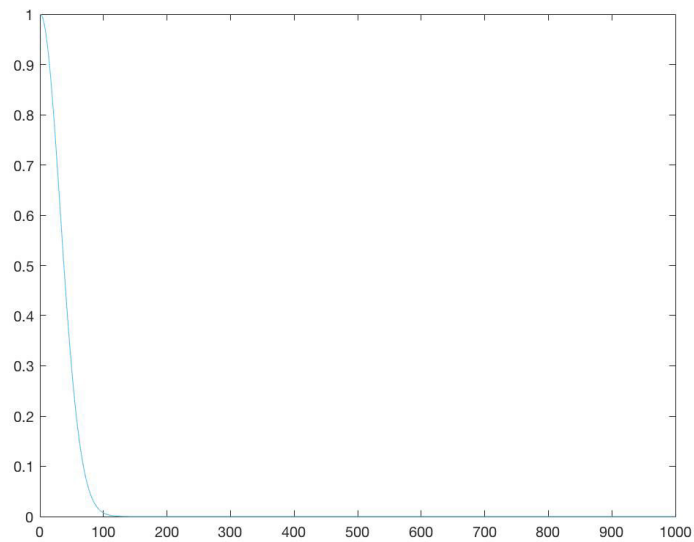
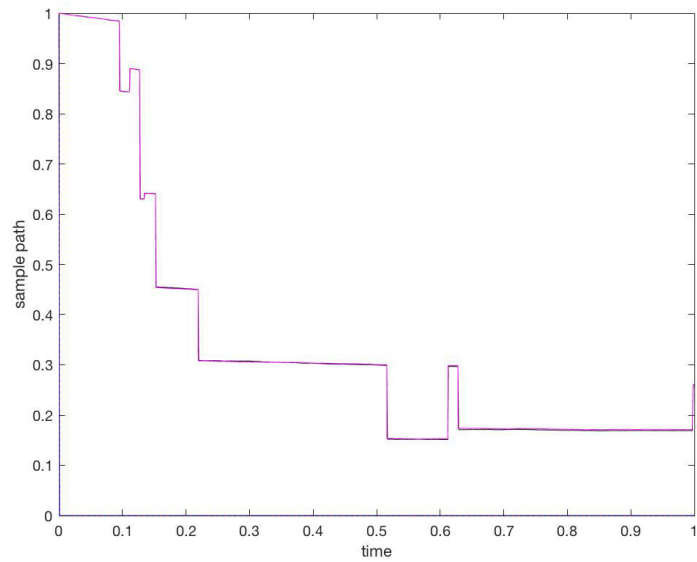


FIGURE 2.3 – OU Jump-diffusion process with exponentially tempered α -stable jumps (left), Survival probability (right) with $\alpha = 0.8$

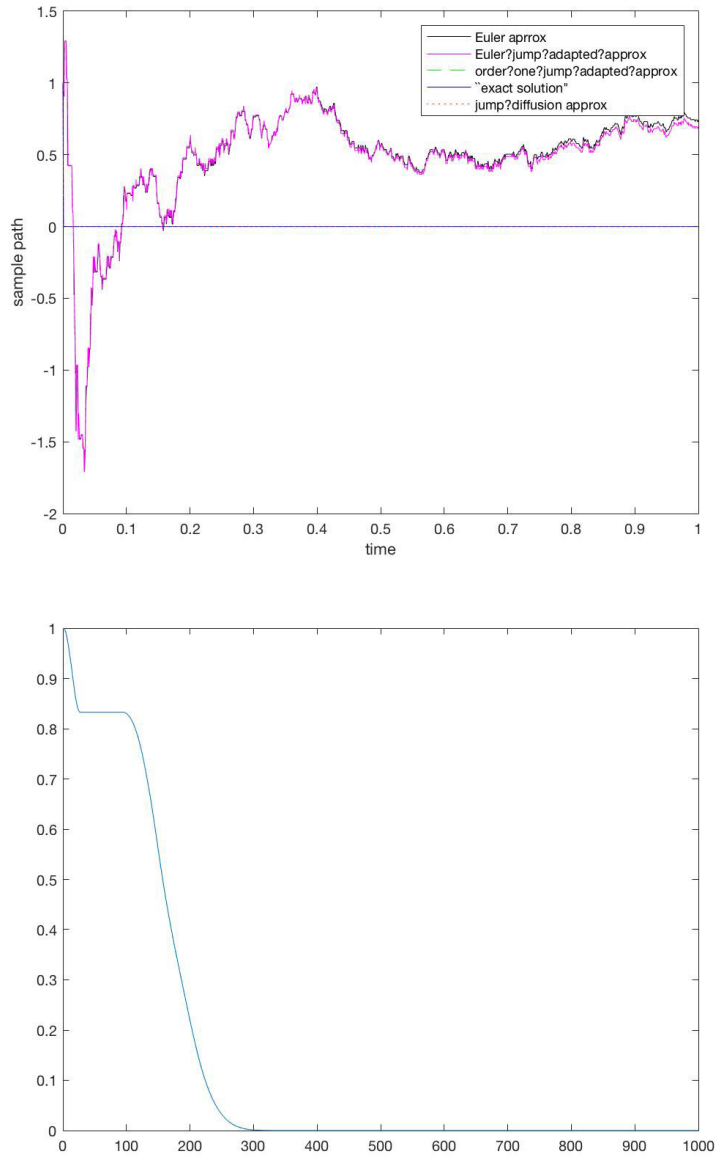


FIGURE 2.4 – OU Jump-diffusion process with exponentially tempered α -stable jumps (left), Survival probability (right) with $\alpha = 1.5$

2.5 WWR Modelling

One of the most interesting cases of CVA valuation is a dependence between the EPE and the default probability of a counterparty. When the dependence is negative, i.e. exposure tends to increase, when the probability of default of our counterparty increases or the credit quality worsens, we have the WWR. When the dependence is positive, we have Right Way Risk. There are a number of different approaches to correlate the exposure and default probability. One can use copulas, or correlate two Brownian motions, i.e. $\langle dW_t, dW_t^\lambda \rangle = \rho dt$, where ρ is a correlation parameter.

Another approach proposed in ([10], [23],[24] and [25]), models WWR by adding jumps in the exposure process (jump-at-default) modelling to reflect the effect of drastic drops in credit ratings to the underlying risk factors.

The following proposition, inspired by [30], allows to incorporate WWR into CVA.

Proposition 2.5.1 *Given the expected positive exposure V_τ^+ , the default time τ , the CVA with wrong-way risk is given by*

$$\text{CVA}^{\text{WWR}} = \text{LGDE}[B_\tau V_\tau^+ \mathbb{1}_{\tau < T}] = \text{LGD} \int_0^T \mathbb{E}[B_t V_t^+ f_\tau(t)] dt \quad (2.71)$$

Proof By using the tower property of the conditional expectation and Fubini theorem, we have the following :

$$\text{LGDE}[B_\tau V_\tau^+ \mathbb{1}_{\tau < T}] = \text{LGDE}[\mathbb{E}[B_\tau V_\tau^+ \mathbb{1}_{\tau < T} | \mathcal{G}_t]] = \text{LGDE} \left[\int_0^T \mathbb{E}[B_\tau V_\tau^+ | \mathcal{G}_t, \tau = t] f_\tau(t) dt \right] = \quad (2.72)$$

$$= \text{LGDE} \left[\int_0^T B_t V_t^+ f_\tau(t) dt \right] = \text{LGD} \int_0^T \mathbb{E}[B_t V_t^+ f_\tau(t)] dt \quad (2.73)$$

As we mentioned in previous sections, α stable process X_t allows Gaussian approximation of small jumps, so that we can express it as :

$$X_t \stackrel{\mathcal{D}}{=} \sigma_\epsilon W_t + P_t^\epsilon \quad (2.74)$$

and correlate the Brownian motion W_t in eq.(2.74) with another Brownian motion W_t^e in the exposure process, i.e.

$$Z_t = \rho W_t + \sqrt{1 - \rho^2} W_t^e \quad (2.75)$$

2.6 Particle Interpretations

2.6.1 Interacting Particle System for Rare Events Estimation

Interacting particle system is efficient method to calculate rare-event probabilities. For example, usually we formulate the problem in the following way. Given a sequence of rare event sets $A_n \subset \dots \subset A_1$, we would like to compute the following expectation after some Radon-Nikodym transformation, that turns a rare event into a regular one :

$$\mathbb{E}^{\mathbb{P}} [g(\lambda_N) \prod_{n=1}^N \mathbb{1}_{\lambda_n \in A_n}] = \mathbb{E}^{\mathbb{Q}} [g(\lambda_N) \frac{d\mathbb{P}}{d\mathbb{Q}} \prod_{n=1}^N \mathbb{1}_{\lambda_n \in A_n}] \quad (2.76)$$

In practice we do not have an explicit form of the likelihood ratio $\frac{d\mathbb{P}}{d\mathbb{Q}}$, so it becomes unfeasible unless one considers some simple examples. One of the solutions was interacting particle system(IPS), which was proposed by Del Moral and Garnier [8], where they

proposed to generate particles(samples) in two steps, i.e. particle mutation and selection. The idea is to approximate the ratio of \mathbb{P} with respect to some importance measure \mathbb{Q} by choosing a weight function ω that approximates the Radon-Nikodym derivative $\frac{d\mathbb{P}}{d\mathbb{Q}}$. If we assume that \mathbb{P} and \mathbb{Q} have density function p and q respectively, for M particles $\{\lambda_n^{(m)}\}_{m=1}^M$ we can define the weight function iteratively by :

$$\prod_{n=1}^N \omega_n(\lambda_n^{(m)}) \propto \frac{d\mathbb{P}}{d\mathbb{Q}}(\lambda_N^{(m)}) = \prod_{n=1}^N \frac{p_n(\lambda_n^{(m)}, \lambda_{n-1}^{(m)})}{q_n(\lambda_n^{(m)}, \lambda_{n-1}^{(m)})} \quad (2.77)$$

Since two measures \mathbb{P} and \mathbb{Q} form a Markov chain, the Radon-Nykodim derivative can be decomposed into the product of ratio of the transition density $p_n(\cdot, \lambda_{n-1}^{(m)})$ to the transition density $q_n(\cdot, \lambda_{n-1}^{(m)})$ with respect to the measure \mathbb{Q} .

The normalized importance weight function is given by :

$$W_n(\lambda_n^{(m)}) = \frac{\omega_n(\lambda_n^{(m)})}{\frac{1}{M} \sum_{j=1}^M \omega_n(\lambda_n^{(j)})} \quad (2.78)$$

The IPS estimate of an expectation (2.76) has the following form :

$$C^{\text{IPS}} = \frac{\mathbb{E} [f(\lambda_N) \prod_{n=1}^{N-1} \omega_n(\lambda_n) \mathbb{1}_{\lambda_n \in A_n}]}{\mathbb{E} [\prod_{n=1}^{N-1} \omega_n(\lambda_n) \mathbb{1}_{\lambda_n \in A_n}]} \quad (2.79)$$

In our experiments we choose a potential function (an unnormalized importance weight) of the form :

$$\omega_n(\lambda_N^{(m)}) = \prod_{n=1}^N e^{\delta(\lambda_n^{(m)} - \lambda_{n-1}^{(m)})} \quad (2.80)$$

where δ is an exponential tilting parameter.

2.6.2 Mean Field IPS Approximations

In this subsection we give a brief overview of interacting particle methods, that is very important to understand the mechanism behind the sequential importance resampling scheme that we will use in the next sections. For a thorough survey we refer to [5],[6].

We can express eq.(2.76) through the linear semigroups associated to the sequence of non-negative measures γ_n for any test function g that have the following form :

$$C^{\text{IPS}} = \gamma_N(g) = \eta_N(g) \gamma_N(\mathbb{1}) \quad (2.81)$$

η_n represent normalized measures. The flow of Feynman-Kac measures (γ_N, η_N) for any measurable and bounded function g is given by

$$\gamma_N(g) = \mathbb{E}[g(\lambda_N) \prod_{n=0}^{N-1} \omega_n(\lambda_n)] \text{ and } \eta_n(g) = \frac{\gamma_n(g)}{\gamma_n(\mathbb{1})} \quad (2.82)$$

where the sequence of measures γ_n satisfy a linear recursion

$$\gamma_n(g) = \gamma_{n-1}(Q_n(g)) \quad (2.83)$$

with integral operator Q_n :

$$Q_n(g)(x) = \mathbb{E}[g(\lambda_n) | \lambda_{n-1} = x] = \int Q_n(x, dy) g(y) \quad (2.84)$$

where the importance measure $Q_n(\cdot, dy)$

$$Q_n(x, dy) = \omega_{n-1}(x) \mathcal{M}_n(x, dy) \quad (2.85)$$

The sequence of measures η_n satisfy the nonlinear equation :

$$\eta_n = \Phi_n(\eta_n) = \Psi_{n-1}(\eta_{n-1}) \cdot \mathcal{M}_n \quad (2.86)$$

where $\Phi_n : E_n \rightarrow E_{n+1}$ and the operator $\Psi_n : E_n \rightarrow E_n$ is

$$\Psi_n(\eta_n)(x) = \frac{\omega_n(x)}{\eta_n(\omega_n)} \eta_n(x) \quad (2.87)$$

Observe, that eq.(2.86) represents very important class of McKean models. The main problem of Monte Carlo simulation of this type of models stems in the fact that it can not be computed exactly, since its Markov transitions \mathcal{M} depends on transitions that are unknown([7], [5]). The IPS allows to approximate the solution by a discrete-time process.

Assume that the Markov transition kernels \mathcal{M}_n have a sequence of densities \mathfrak{M}_n , and there exist some reference probability measures \varkappa_n , such that $\mathcal{M}_n \ll \varkappa_n$, i.e. \mathcal{M}_n are absolutely continuous with respect to \varkappa_n :

$$\mathcal{M}_n(x, dy) = \mathfrak{M}_n(x, y) \varkappa_n(dy) \quad (2.88)$$

Radon-Nikodym derivative given that $Q_n(x, \cdot) \ll \varkappa_n$:

$$\frac{dQ_n(x, \cdot)}{d\varkappa_n}(y) = \omega_{n-1}(x) \mathfrak{M}_n(x, y) \quad (2.89)$$

The following result shows the approximation of importance measure Q_n by a flow of non-negative measures η_n . The lemma is using techniques, that were adapted from [5] and [6].

Lemma 2.6.1 *Assume that*

$$\sup_{x, y \in E_n} \frac{\mathfrak{M}_n(x, z)}{\mathfrak{M}_n(y, z)} := h_n(z) \text{ s.th. } \|h_n\| < \infty \text{ and } \sup_{x, y \in E_n} \frac{\omega_n(x)}{\omega_n(y)} < \infty, \forall n \geq 0 \quad (2.90)$$

then the integral operator (2.84) is given by

$$Q_n(g)(x) = \eta_{n-1}(\omega_{n-1}) \int \frac{\omega_{n-1}(x) \mathfrak{M}_n(x, y)}{\eta_{n-1}(\omega_{n-1}) \mathfrak{M}_n(\cdot, y)} g(y) \eta_n(dy) \quad (2.91)$$

Proof Using eq.(2.86) we have :

$$\eta_n(dy) = \Psi_{n-1}(\eta_{n-1}) \cdot \mathcal{M}_n(dy) = \frac{\eta_{n-1}(\omega_{n-1}) \mathfrak{M}_n(\cdot, y) \varkappa_n(dy)}{\eta_{n-1}(\omega_{n-1})} \quad (2.92)$$

So the importance density takes the following form :

$$Q_n(x, dy) = \eta_{n-1}(\omega_{n-1}) \frac{\omega_{n-1}(x) \mathfrak{M}_n(x, y)}{\eta_{n-1}(\omega_{n-1}) \mathfrak{M}_n(\cdot, y)} \eta_n(dy) \quad (2.93)$$

The statement is obtained by applying eq.(2.84).

IPS could be understood as a stochastic linearization of the nonlinear recursive equation (2.86). The algorithm involves construction of E_n^M valued M-tuple Markov Chain $\mathfrak{X}_n = \{\mathfrak{X}_n^{(i)}\}_{i=1}^M \in E_n^M$ with transitions given by

$$\mathbb{P}(\mathfrak{X}_n \in d(x^{(1)}, \dots, x^{(M)}) | \mathfrak{X}_{n-1}) = \prod_{i=1}^M \Phi_n(\eta_n^M)(dx^{(i)}) \quad (2.94)$$

to approximate measure η_n :

$$\lim_{n \rightarrow \infty} \eta_n^M(g) = \eta_n(g) \quad \text{where} \quad \eta_n^M = \frac{1}{M} \sum_{i=1}^M \delta_{\mathfrak{X}^{(i)}} \quad (2.95)$$

The above could be proven using the law of large numbers under certain regularity conditions.

Now we can obtain the approximation \widehat{Q}_n of the importance measure Q_n by replacing the measures η_n by their M-tuple approximation η_n^M , which is the definition of mean field IPS approximation scheme :

$$\begin{aligned} \widehat{Q}_n(g)(x) &= \eta_{n-1}^M(\omega_{n-1}) \int \frac{\omega_{n-1}(x) \mathfrak{M}_n(x, y)}{\int \eta_{n-1}^M(dz) (\omega_{n-1}(z) \mathfrak{M}_n(z, y))} g(y) \eta_n^M(dy) = \\ &= \eta_{n-1}^M(\omega_{n-1}) \sum_{i=1}^M \frac{\omega_{n-1}(x) \mathfrak{M}_n(x, \mathfrak{X}^{(i)})}{\sum_{j=1}^M \eta_{n-1}^M(dz) (\omega_{n-1}(z) \mathfrak{M}_n(z, \mathfrak{X}^{(j)}))} g(\mathfrak{X}^{(i)}) \end{aligned} \quad (2.96)$$

Particle Monte Carlo Convergence Properties

There is a lot of research on the convergence of particle methods. Now, given M particles $\{\mathfrak{X}^{(i)}\}_{i=1}^M$, we cite from [8] the following result :

$$\frac{1}{M} \sum_{i=1}^M g(x) \delta_{\mathfrak{X}_N^{(i)}}(x) \xrightarrow{M \rightarrow \infty} \mathbb{E}[g(\lambda_N)] \quad a.s. \quad (2.97)$$

where δ is a Dirac function. The following proposition shows that the random sequence $\{\gamma_n^N(\mathbb{1})\}_{n=1}^k$ converges in probability to a deterministic sequence $\{\gamma_n(\mathbb{1})\}_{n=1}^k$ as its number M goes to infinity.

Proposition 2.6.2 *The Particle estimator is unbiased estimator, so that for any $p \geq 1$ and for all bounded and measurable functions g on some measurable space (E_k, \mathcal{E}_k) with $\|g\| \leq 1$, we have*

$$\mathbb{E}[\gamma_k^N(g)] = \gamma_k(g) \quad (2.98)$$

and moreover

$$\sup_{M \geq 1} \|\gamma_k^N(g) - \gamma_k(g)\|_{L^p} \leq \frac{c_p(k)}{\sqrt{M}} \quad (2.99)$$

where $c_p(k)$ is constant, whose values depends on the time final time horizon $1 \leq n \leq k$.

Proof We refer to [8].

Algorithm

The algorithm can be described by the following scheme, $\forall m = 1, \dots, M$:

$$\mathfrak{X}_n^{(i)} \in E_n \xrightarrow{\text{Selection}} \widehat{\mathfrak{X}}_n^{(i)} \in E_n \xrightarrow{\text{Sampling}} \mathfrak{X}_{n+1}^{(i)} \in E_{n+1} \quad (2.100)$$

At each time step $n = 1, \dots, N$ we draw M independent random variables from the density $q_n(\cdot, \mathfrak{X}_{n-1}^{(i)})$ to construct M particles $\widehat{\mathfrak{X}}_n^{(m)} = (\widehat{\mathfrak{X}}_n^{(1)}, \dots, \widehat{\mathfrak{X}}_n^{(M)})$. Given generated particles, we select independently M particles $\widehat{\mathfrak{X}}_n = (\widehat{\mathfrak{X}}_n^{(1)}, \dots, \widehat{\mathfrak{X}}_n^{(M)})$ with replacement of rejected particles according to their probability weights :

$$W_n(\mathfrak{X}_n^{(i)}) = \frac{\omega_n(\mathfrak{X}_n^{(i)})}{\frac{1}{M} \sum_{j=1}^M \omega_n(\mathfrak{X}_n^{(j)})} \quad (2.101)$$

Algorithm 7 : IPS algorithm

```

1 Initialization : M - #(simulations), N - #(time steps),  $\lambda_0$  - initial value
2 for  $n = 1, \dots, N$  do
3   for  $m = 1, \dots, M$  do
4     Generate  $\mathfrak{X}_n^{(m)}$  from  $p(\cdot, \mathfrak{X}_{n-1}^{(m)})$  and set  $\widehat{\mathfrak{X}}_n^{(m)} = (\widehat{\mathfrak{X}}_n^{(m)}, \mathfrak{X}_{n-1}^{(m)})$ ;
5     if  $\widehat{\mathfrak{X}}_n^{(m)} \in A$  then
6        $\widehat{\mathfrak{X}}_n^{(m)} = 0$ 
7     else
8       Compute the weight :  $\omega_n(\widehat{\mathfrak{X}}_n^{(m)})$ .
9     end
10    end
11  end
12  if  $n < N$  then
13    Resample using probability weight :  $W_n(\widehat{\mathfrak{X}}_n^{(m)}) = \frac{\omega_n(\widehat{\mathfrak{X}}_n^{(m)})}{\frac{1}{M} \sum_{j=1}^M \omega_n(\widehat{\mathfrak{X}}_n^{(j)})}$  to sample
14     $\mathfrak{X}_n^{(m)}$ .
15  end

```

2.6.3 IPS Simulation of the Stochastic Intensity

In this section we formulate the problem of a rare event estimation and show on the example of a stochastic intensity the application of IPS to estimate survival probabilities.

When we deal with credit risks most of the problems are formulated as occurrences of rare events, in the sense that $\mathbb{P}(\tau > t)$ is very small, when t is large. If one can transform from a measure \mathbb{P} to the measure \mathbb{Q} , using Radon-Nikodym derivative transformation, so that in the new measure \mathbb{Q} , t is not too large, then one can use central limit theorem approximation.

$$f_\tau(t) = \mathbb{E}^{\mathbb{Q}}[\lambda_t e^{-\int_0^t \lambda_u du}] = \mathbb{E}^{\mathbb{Q}}[\lambda_t e^{-\int_0^t \lambda_u du} \frac{d\mathbb{P}}{d\mathbb{Q}}] \quad (2.102)$$

Interacting particle system, described in previous sections allows us to approximate a rare event probability by sequentially regenerating particles according to their probability weights.

Define the cumulated intensity as $\Lambda(t) = \int_0^t \lambda_u du$ and its discrete version $\Lambda(t_N) = \int_0^{t_N} \lambda_u du$ for the time interval $0 = t_0 < \dots t_N = t$. From large deviation analysis, we know that $\mathbb{P}(\xi > \Lambda(t_N)) \rightarrow 0$, as $\Lambda(t_N) \rightarrow \infty$, i.e.

$$\mathbb{P}(\tau > t_N) = \mathbb{P}(\xi > \Lambda(t_N)) = \mathbb{E}^{\mathbb{P}}[\mathbb{1}_{\xi > \Lambda(t_N)}] = \mathbb{E}^{\mathbb{P}}[\mathbb{1}_{\xi > \Lambda(t)} e^{\alpha\xi - \alpha\Lambda(t)}] \leq e^{-(\alpha\Lambda(t_N) - \Gamma(\alpha))}$$

with $\Gamma(\alpha) = \log(\mathbb{E}^{\mathbb{P}}[e^{\alpha\xi}])$ (2.103)

Unnormalized particle approximations are given by :

$$\mathbb{E}^{\mathbb{P}}[\mathbb{1}_{\xi > \Lambda(t_N)}] = \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\xi > \Lambda(t_N)} e^{-\alpha\Lambda(t_N)}] = \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\xi > \Lambda(t_N)} \prod_{l=1}^N e^{\alpha(\Lambda(t_l) - \Lambda(t_{l-1}))}] \quad (2.104)$$

Now we can give particle interpretation of the default probability. Assuming that initial values are equal, i.e. $p_0 = \mu_0$ and for any bounded measurable function $g : \mathbb{E} \rightarrow \mathbb{R}$, under new measure \mathbb{Q} we have the following representation of the original expectation under measure \mathbb{P} :

$$f_{\tau}(t_N) = \frac{\mathbb{E}^{\mathbb{Q}}[\lambda_{t_N} e^{-\int_0^{t_N} \lambda_u du} \prod_{l=1}^{N-1} \tilde{\omega}_l(\lambda_l)]}{\mathbb{E}^{\mathbb{Q}}[\prod_{l=1}^{N-1} \omega_l(\lambda_l)]} \quad (2.105)$$

In figure 5, we can see the simulated paths of three discretization schemes with constant and jump-adapted time steps to compare with almost-exact solution.

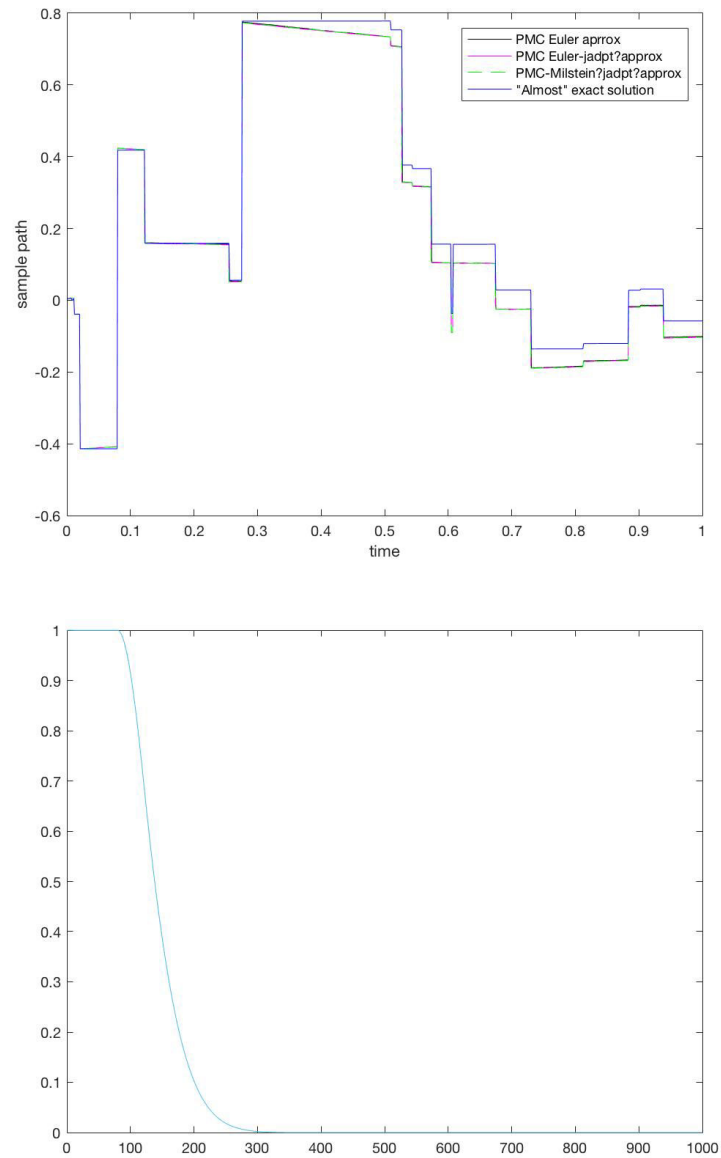


FIGURE 2.5 – PMC simulation of OU Jump-diffusion process with exponentially tempered α -stable jumps (left), Survival probability (right) with $\alpha = 0.8$

2.7 Numerical Analysis and Discussion

2.7.1 Numerical Schemes Error Estimates

We split errors into two categories : $\|\cdot\|_{L^2}$ and $\|\cdot\|_{L^1}$ errors. Strong errors are given in terms of $\|\cdot\|_{L^2}$ norm. Given the estimates \widehat{X}_t^{ll} , where $ll = \{MC_{Eu}, MC_{Eu-jadp}, MC_{mIs-jadp}\}$ set of numerical schemes, of the process X_t , we define a strong error estimates as :

$$err_s = (\mathbb{E} [|\widehat{X}_t - X_t|^2])^{1/2} \approx \sqrt{\frac{1}{M} \sum_{i=1}^M |\widehat{X}_t^{(i)} - X_t^{(i)}|^2} \quad (2.106)$$

$\|\cdot\|_{L^1}$ errors are defined as :

$$err_w = \mathbb{E} [|\widehat{X}_t - X_t|] \approx \frac{1}{M} \sum_{i=1}^M |\widehat{X}_t^{(i)} - X_t^{(i)}| \quad (2.107)$$

Table 1 and 2 shows $\|\cdot\|_{L^2}$ and $\|\cdot\|_{L^1}$ errors respectively of 5000 simulated Monte Carlo paths and 1000 particles. Level of stability is chosen to be $\alpha = 0.8$, when most of jumps are large. We observe that errors are sufficiently minimized when we use IPS. The same demonstrates figure 7. Figure 6 shows that the particles variability and errors are converging between -7 and -8 for three numerical schemes as we increase the number of time steps.

TABLEAU 2.1 – $\|\cdot\|_{L^2}$ error estimates of numerical schemes using MC and PMC density approximations, $\alpha = 0.8$

Δt	MC _{eu}	t	MC _{eujadp}	t	MC _{mIsjadp}	t	PMC _{eu}	t	PMC _{eujadp}	t	PMC _{mIsjadp}	t
2^{-6}	2.54e-01	0.22	2.62e-01	0.15	2.62e-01	0.14	7.10e-02	0.03	2.16e-01	0.02	2.15e-01	0.02
2^{-7}	1.55e-01	0.27	1.67e-01	0.16	1.67e-01	0.17	6.05e-02	0.27	8.23e-02	0.16	8.10e-02	0.17
2^{-8}	1.85e-01	0.32	3.43e-02	0.22	1.85e-01	0.25	7.88e-02	0.06	1.08e-01	0.05	1.09e-01	0.05
2^{-9}	2.74e-01	0.28	2.91e-01	0.19	2.92e-01	0.20	6.63e-02	0.12	1.30e-01	0.10	1.30e-01	0.10
2^{-10}	3.19e-01	0.29	3.30e-01	0.20	3.29e-01	0.20	7.14e-02	0.23	7.90e-02	0.21	7.89e-02	0.21

TABLEAU 2.2 – $\|\cdot\|_{L^1}$ error estimates of numerical schemes using MC and PMC density approximations . $\alpha = 0.8$

Δt	MC _{eu}	t	MC _{eujadp}	t	MC _{mIsjadp}	t	PMC _{eu}	t	PMC _{eujadp}	t	PMC _{mIsjadp}	t
2^{-6}	2.04e-01	0.22	2.09e-01	0.15	2.09e-01	0.14	5.48e-02	0.03	1.656e-01	0.02	1.652e-01	0.02
2^{-7}	1.03e-01	0.27	1.12e-01	0.16	1.12e-01	0.17	5.25e-02	0.04	6.28e-02	0.03	6.23e-02	0.03
2^{-8}	1.3e-01	0.32	1.36e-01	0.22	1.36e-01	0.25	6.88e-02	0.06	8.32e-02	0.05	8.34e-02	0.05
2^{-9}	2.03e-01	0.28	2.1e-01	0.19	2.1e-01	0.20	5.91e-02	0.12	7.95e-02	0.1	7.95e-02	0.1
2^{-10}	1.95e-01	0.29	2.1e-01	0.20	2.1e-01	0.20	5.86e-02	0.23	6.32e-02	0.21	6.36e-02	0.21

TABLEAU 2.3 – $\|\cdot\|_{L^2}$ error estimates of numerical schemes using MC and PMC density approximations, $\alpha = 1.5$

Δt	MC _{eu}	t	MC _{eujadp}	t	MC _{mIsjadp}	t	PMC _{eu}	t	PMC _{eujadp}	t	PMC _{mIsjadp}	t
2^{-6}	3.15e-01	0.28	9.89e-01	0.44	9.85e-01	0.49	2.3e-01	0.4	5.23e-01	1.42	5.24e-01	1.36
2^{-7}	3.626e-01	0.13	6.964e-01	0.33	6.922e-01	0.33	1.823e-01	0.54	3.861e-01	3.21	3.840e-01	1.27
2^{-8}	3.230e-01	0.1	4.575e-01	0.25	4.563e-01	0.29	2.036e-01	1.5	2.837e-01	7.6	2.673e-01	1.9
2^{-9}	3.264e-01	1.05	6.083e-01	2.55	6.011e-01	0.75	2.066e-01	4.01	3.212e-01	10.24	3.232e-01	2.29
2^{-10}	3.34e-01	0.29	3.78e-01	0.44	4.11e-01	0.42	2.10e-01	15.9	3.17e-01	21.9	3.04e-01	4.47

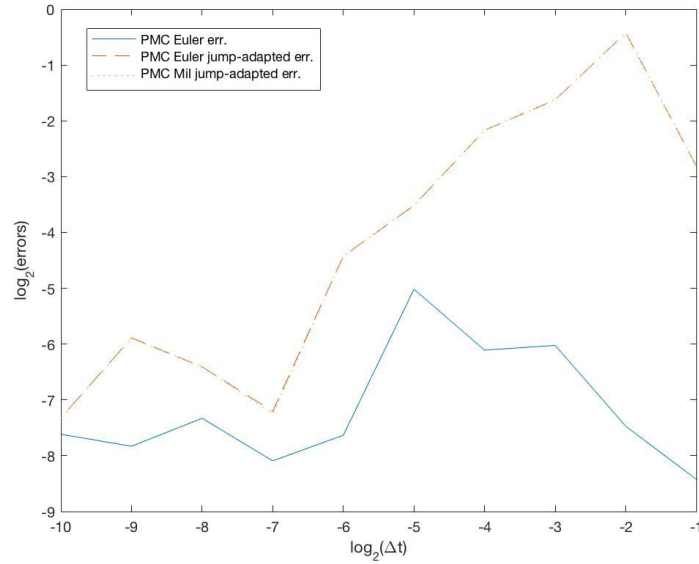


FIGURE 2.6 – PMC Euler, Euler jump-adapted, Milstein jump-adapted strong errors, $\alpha = 0.8$, $\#(\text{particles})=1000$.

TABLEAU 2.4 – $\|\cdot\|_{L^1}$ error estimates of numerical schemes using MC and PMC density approximations . $\alpha = 1.5$

Δt	MC_{eu}	t	$MC_{eu_{jdp}}$	t	$MC_{mils_{jdp}}$	t	PMC_{eu}	t	$PMC_{eu_{jdp}}$	t	$PMC_{mils_{jdp}}$	t
2^{-6}	2.51e-01	0.28	8.1e-01	0.44	8.01e-01	0.49	1.82e-01	0.4	4.34e-01	1.42	4.33e-01	1.36
2^{-7}	3.548e-01	0.13	6.962e-01	0.33	6.922e-01	0.33	1.694e-01	0.54	3.857e-01	3.21	3.840e-01	1.27
2^{-8}	3.203e-01	0.1	4.572e-01	0.25	4.563e-01	0.29	1.966e-01	1.59	2.831e-01	7.6	2.673e-01	1.9
2^{-9}	3.243e-01	1.05	6.082e-01	2.55	6.011e-01	0.75	2.037e-01	4.01	3.206e-01	10.24	3.232e-01	2.29
2^{-10}	3.33e-01	0.29	3.77e-01	0.44	4.11e-01	0.42	2.08e-01	15.9	3.16e-01	21.9	3.04e-01	4.47

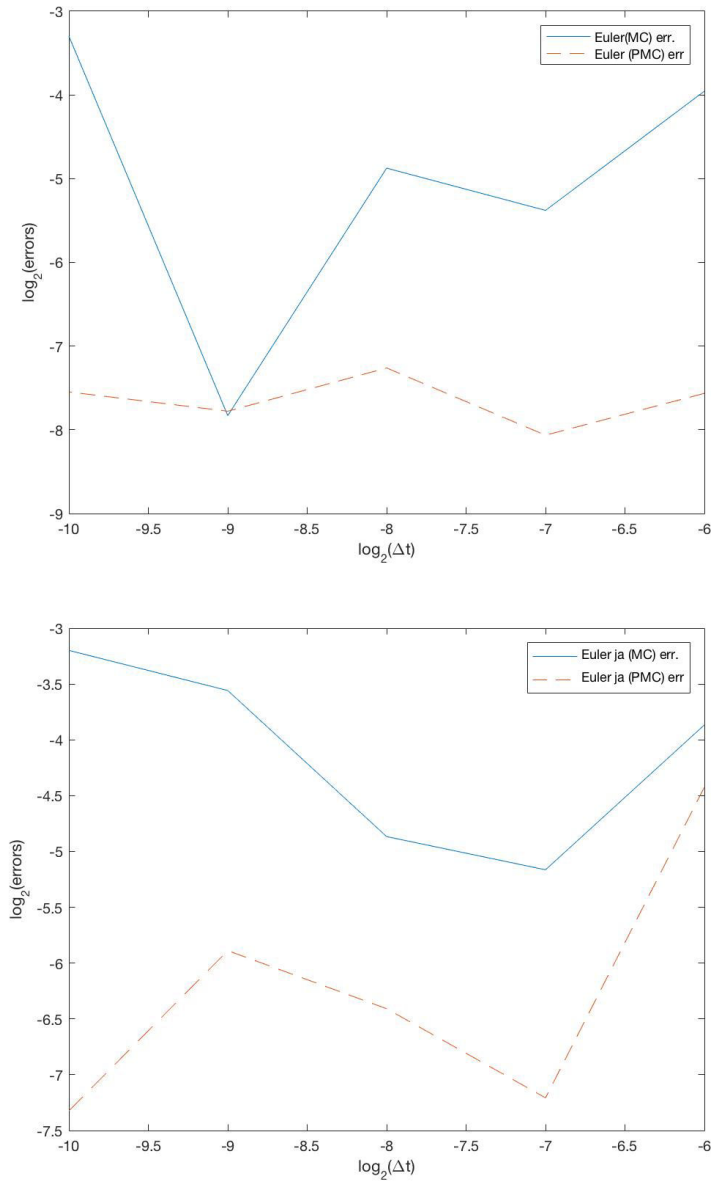


FIGURE 2.7 – MC vs PMC Euler (left) and jump-adapted (right) strong errors. $\alpha = 0.8$, $\#(\text{particles}) = 1000$, $\#(\text{simulated paths}) = 5000$

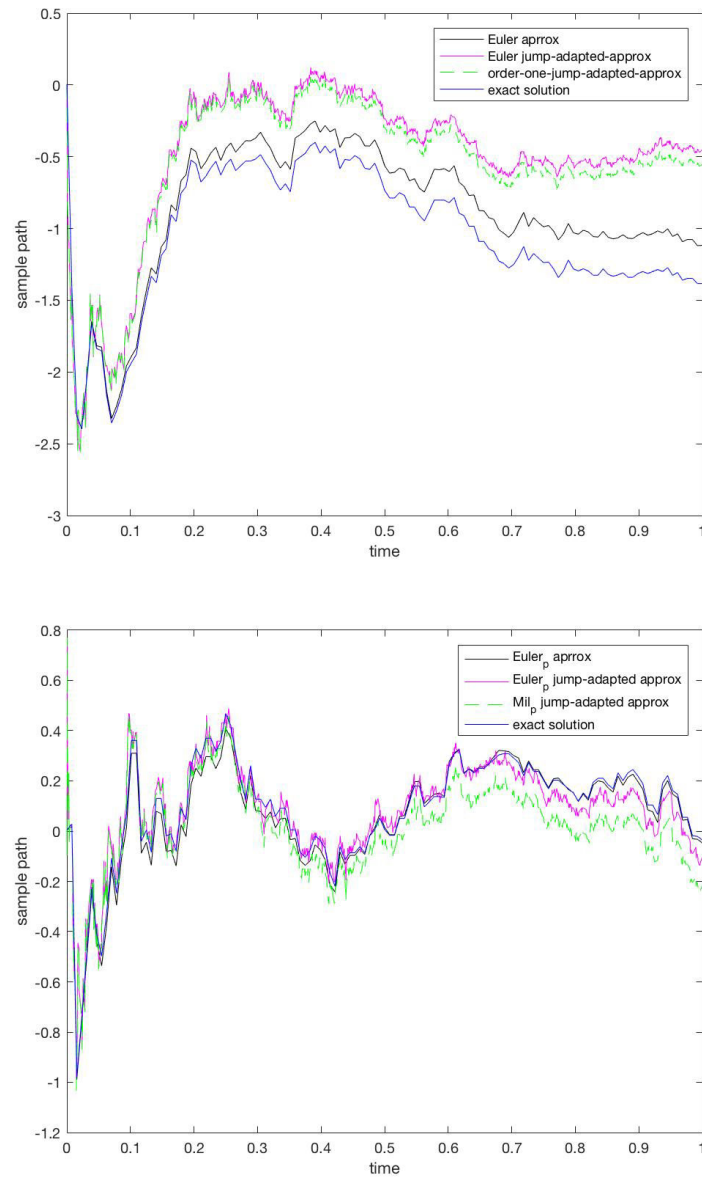


FIGURE 2.8 – MC vs PMC Euler/Milstein (left) and jump-adapted (right) paths. $\alpha = 1.5$, #(particles)=1000, #(simulated paths)=5000

2.7.2 CVA Computation

In this subsection we present the computation methodology for CVA estimation. We consider two cases of CVA modelling :

- CVA^{ind}, when there is no correlation between default probability and the exposure process.
- CVA^{wwr}, when we assume a negative correlation between market risk and credit risk.

First we outline two factor BGM model that will be used for expected exposure V_t simulation. Second, depending on the case of presence or absence of the WWR, we simulate the Levy driven OU stochastic intensity process.

Exposure Profiles Specification

We consider a two-factor BGM to model LIBOR interest rates :

$$dL_i(t) = \sum_{j=g(t)}^i \frac{\delta_j L_j(t) L_i(t) \sum_{q=1}^2 \zeta_{i,q}(t) \zeta_{j,q}(t)}{1 + \delta_j L_j(t)} dt + L_n(t) \sum_{q=1}^2 \zeta_{i,q}(t) dW_t \quad (2.108)$$

Euler discretization scheme gives us :

$$L_i(t_{n+1}) = L_i(t_n) + \mu_n(L_i(t_n), t_n) L_i(t_n) \Delta t + L_i(t_n) \sqrt{\Delta t} \sigma_i(t_n) \epsilon_{n+1} \quad (2.109)$$

where $\Delta t = t_{n+1} - t_n$ and drift μ_n is given by

$$\mu_i(L_i(t_n), t_n) = \sum_{j=g(t)}^i \frac{\delta_j L_j(t) L_i(t) \sum_{q=1}^2 \zeta_{i,q}(t) \zeta_{j,q}(t)}{1 + \delta_j L_j(t)} \quad (2.110)$$

CVA

Using the process (2.109) simulate independent interest rate path realizations $L_1^{(m)}, \dots, L_N^{(m)}$ to get exposure profiles from eq.(2.19). Simulate stochastic intensity λ_t using on of the discretization schemes (Euler, Euler jump-adapted and Milstein jump-adapted). Then, the unbiased estimator of CVA value in eq.(2.6) is

$$\widehat{CVA}^{ind} = \text{LGD} \int_0^T \mathbb{E}[B_t V_t^+] f(t) dt = \text{LGD} \sum_{l=1}^N \int_{T_l}^{T_{l+1}} \mathbb{E}[B_t V_t^+] f(t) dt = \quad (2.111)$$

$$= \text{LGD} \sum_{l=1}^N B_l \sum_{m=1}^M V_{t_l}^{(m),+} \int_{T_l}^{T_{l+1}} f_{\tau}(t) dt = \quad (2.112)$$

$$= \text{LGD} \sum_{i=1}^N B_l \sum_{m=1}^M V_{t_l}^{(m),+} \sum_{j=1}^{M^j} \lambda_l^{(j)} e^{-\sum_{k=1}^l \lambda_k^{(j)} \Delta k} \Delta t \quad (2.113)$$

where $\Delta t = T_{l+1} - T_l$, $B_l = B(0, T_{l+1})$ and $\Delta k = t_k - t_{k-1}$.

In the case of CVA with WWR we specify different levels of correlation ρ between two Brownian motions in (2.75). Monte Carlo estimator of CVA with WWR is

$$\widehat{CVA}^{WWR} = \text{LGD} \int_0^T B_t \mathbb{E}[V_t^+ f_{\tau}(t)] dt = \text{LGD} \sum_{l=1}^N \int_{T_l}^{T_{l+1}} B_t \mathbb{E}[B_t V_t^+ f_{\tau}(t)] dt = \quad (2.114)$$

$$= \text{LGD} \sum_{l=1}^N \int_{T_l}^{T_{l+1}} B_t \sum_{m=1}^M V_{t_l}^{(m),+} f_{\tau}^{(m)}(t) dt \quad (2.115)$$

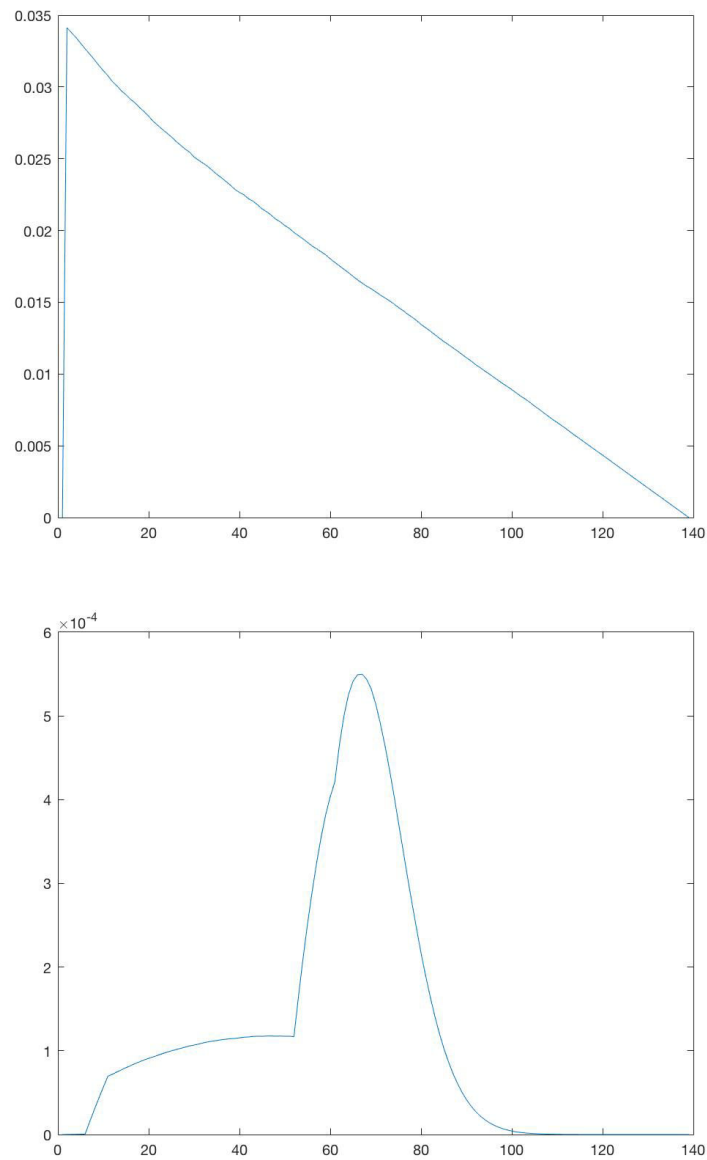


FIGURE 2.9 – EPE profile of the LIBOR swap (left), CVA process (right)

2.7.3 CVA. Particle Monte Carlo

Another unbiased estimator CVA^P of the expectation in eq.(2.6) can be obtained by means of IPS.

The exposure profiles are simulated similarly as in the previous section. The simulation of a survival probability, as it is described in algorithm 2, consists of two steps : sampling (mutation) and survival (selection). Define M initial particles $\{\mathcal{X}_n^{(m)}\}_{m=1}^M$. The selection step consists of randomly choosing a set of M particles from the selection transition, which, for example, could be chosen as :

$$\mathcal{S}_n = \omega_n(\mathcal{X}_n^{(m)}) \mathbb{1}_{\mathcal{X}_n^{(m)}}(x) + (1 - \omega_n(\mathcal{X}_n^{(m)})) \sum_i \frac{\omega_n(\mathcal{X}_n^{(i)})}{\sum_j \omega_n(\mathcal{X}_n^{(j)})} \mathbb{1}_{\mathcal{X}_n^{(i)}}(x) \quad (2.116)$$

Define selected particles as λ_n . The mutation step from λ_n to λ_{n+1} : every survived particle $\lambda_n^{(i)}$ propagates randomly to a new particle $\lambda_{n+1}^{(i)} = y$ randomly chosen from the transition kernel $\mathcal{M}(\lambda_n^{(i)}, dy)$.

At the final time N, the CVA estimate is

$$\widehat{CVA}^{P,ind} = \text{LGD} \sum_{l=1}^N \frac{1}{M} \sum_{i=1}^M B_l V_l^{(m),+} \times \left[\left(\prod_{l=1}^{N-1} \frac{1}{M} \sum_{m=1}^M \omega_l(\lambda_l^{(m)}) \right) \times \sum_{m=1}^M f_\tau(\lambda_N^{(m)}) \right] \quad (2.117)$$

CVA with WWR estimate is given by

$$\widehat{CVA}^{P,WWR} = \text{LGD} \sum_{l=1}^N \frac{1}{M} \sum_{i=1}^M B_l V_l^{(m),+} f_\tau(\lambda_l^{(m)}) \times \left(\prod_{l=1}^{N-1} \frac{1}{M} \sum_{m=1}^M \omega_l(\lambda_l^{(m)}) \right) \quad (2.118)$$

We assume that parameters of the underlying stochastic processes are already calibrated. We simulate 20 times CVA values :

$$C\bar{V}A = \sum_{l=1}^{M_s} \widehat{CVA}_l \quad (2.119)$$

where M_s is the number of repeated simulations.

2.7.4 Numerical Results

We simulate CVA values using two numerical methods : first, EPE using standard MC approach, second, EPE using particle methods ; We compare CVA without WWR and CVA with WWR on the one hand, and compare with Basel III recommendations on the other.

Table 5. shows that Basel III's approximation is overvalued we use the formula (2.118) for 10^4 and 10^5 scenarios.

Next, we are interested in CVA values with WWR, when we correlate two Brownian motions, the one is the exposure and another one in Gaussian small jump approximation of Levy process. Table 6. shows CVA with RWR values for different levels initial intensity and correlation. Table 7. gives CVA with WWR values, we can see that Basel III approximation is overvalued for initial intensity values lower than 0.03 and for 0.04.

CHAPITRE 2. PARTICLE LEVY TEMPERED α -STABLE CREDIT RISK MODEL
SIMULATION WITH APPLICATION TO CVA WITH WWR COMPUTATION IN
MULTI-FACTOR BGM SETTING

TABLEAU 2.5 – CVA WWR/free for LIBOR SWAP. PMC approach. $\alpha = 0.8$, $\#(scenarios) := \#(sc)$

λ_0	CVA (st.dev)	CVA WWR (st.dev)	CVA WWR(BIII) (st.dev)	Θ_{sim}^{BIII}	CVA (st.dev)	CVA WWR (st.dev)	CVA WWR(BIII) (st.dev)	Θ_{sim}^{BIII}
$\#(sc)$	10^4	10^4	10^4	10^4	10^5	10^5	10^5	10^5
0	1.02e-04 (7.71e-05)	1.29e-04 (9.93e-05)	1.43e-04 -	1.11	8.32e-05 (7.7e-05)	1.19e-04 (1e-04)	1.16e-04 -	0.84
0.01	1.03e-04 (8.03e-05)	1.36e-04 (1.02e-04)	1.45e-04 -	1.07	1.25e-04 (6.5e-05)	1.2e-04 (1.1e-04)	1.75e-04 -	1.46
0.02	9.95e-04 (7.55e-05)	1.06e-04 (1.02e-04)	1.39e-04 -	1.31	1.24e-04 (7.78e-05)	1.63e-04 (9.12e-05)	1.74e-04 -	1.07
0.03	8.62e-05 (8.16e-05)	9.82e-05 (1.07e-04)	1.2e-04 -	1.22	1.07e-04 (8.46e-05)	1.2e-04 (1.1e-04)	1.5e-04 -	1.25
0.04	8.85e-05 (7.63e-05)	1.44e-04 (9.63e-05)	1.24e-04 -	1.37	1.15e-04 (7.79e-05)	1.46e-04 (9.29e-05)	1.61e-04 -	1.1

TABLEAU 2.6 – LIBOR SWAP CVA WWR/free for different values of ρ and λ_0 , $\alpha = 0.8$. PMC approach.

$\rho \backslash \lambda_0$	0	0.01	0.02	0.03	0.04
0.1	8.6e-05	9.1e-05	1.22e-04	7.52e-05	8.54e-05
0.2	6.9e-05	8e-05	1.3e-04	8.93e-05	1.14e-04
0.3	6.8e-05	9.4e-05	7.69e-05	1.13e-04	1.04e-04
0.4	1e-04	6.2e-05	9.07e-05	9.36e-05	1.01e-04
0.5	8.2e-05	7.0e-05	9.81e-05	1.15e-04	9.98e-05
CVA WWR(B III)	1.43e-04	1.45e-04	1.39e-04	1.12e-04	1.97e-04
CVA	1.02e-04	1.03e-04	9.95e-05	8.62e-05	1.4e-04

TABLEAU 2.7 – LIBOR SWAP CVA WWR/free for different values of ρ and λ_0 , $\alpha = 0.8$. PMC approach.

$\rho \backslash \lambda_0$	0	0.01	0.02	0.03	0.04
-0.5	1.25e-04	1.11e-04	9.91e-05	1.21e-04	1.27e-04
-0.6	8.67e-05	8.76e-05	1.07e-04	1.21e-04	1.18e-04
-0.7	1.02e-04	1.06e-04	1.16e-04	1.16e-04	1.64e-04
-0.8	8.12e-05	9.58e-04	1.49e-04	1.56e-04	1.29e-04
-0.9	1.37e-05	1.25e-04	1.29e-04	1.59e-04	1.79e-04
CVA WWR(B III)	1.43e-04	1.45e-04	1.39e-04	1.12e-04	1.44e-04
CVA	1.02e-04	1.03e-04	9.95e-05	8.62e-05	8.85e-04

2.8 Conclusion

In this paper we adapted IPS approach to the simulation of α -stable process. Using mean field algorithms we reduced $\|\cdot\|_{L^2}$ and $\|\cdot\|_{L^1}$ errors for three discretization scheme : Euler with constant time step, Euler and Milstein with jump-adapted time steps compared to a vanilla Monte Carlo algorithm. IPS allowed reduce the number of simulation paths and reduce the variability of estimates.

We also developed α -stable intensity process to estimate probability of default. We showed how WWR could be naturally embedded into the CVA values. Importantly, our results demonstrated that the Basel III approximation of CVA with WWR far from being adequate. As an extension, in the next article we will use OIS spreads in the exposure valuation in the case of other products, such as Bermudan swaptions.

Bibliographie

- [1] Leif Andersen. A simple approach to the pricing of Bermudan swaptions in the multi-factor Libor market model. *Journal of Computational Finance*, 3 :532, 2000.
- [2] S. Asmussen and J. Rosiński. Approximation of small jumps of Lévy processes with view towards simulation. *J. Appl. Probab.*, 38 :482483, 2001.
- [3] Bertoin, J. (1996). *Lévy Processes*, Cambridge University Press.
- [4] Brigo, D. & Mercurio, F. (2001), *Interest Rate Models : Theory and Practice with Smile, Inflation and Credit*, Heidelberg : Springer Verlag
- [5] D. Brigo, F. Vrina. Disentangling wrong-way risk : pricing CVA via change of measures and drift adjustment. Available on arxiv.
- [6] Marco Bianchetti ; *Interest Rates After The Credit Crunch : Multiple-Curve Vanilla Derivatives and SABR*, 2012
- [7] M. BOSSY and D. TALAY. A stochastic particle method for some one-dimensional non-linear PDE. *Mathematics and Computer in Simulation*, 38 :43-50, 1995.
- [8] Carmona, Rene, Jean-Pierre Fouque & Douglas Vestal (2009), Interacting particle systems for the computation of rare credit portfolio losses, *Finance and Stochastics* 13(4), 613633.
- [9] Carmona, Rene, Stephane Crepey (2010), Particle methods for the estimation of Markovian credit portfolio loss distributions, *International Journal of Theoretical and Applied Finance* 13(4), 577602.
- [10] Ehlers P., Schönbucher, P. (2006). The influence of FX risk on credit spreads. *Default-risk.com*.
- [11] Fabio Mercurio, QFR, Bloomberg ; *Interest Rates and The Credit Crunch : New Formulas and Market Models*, 2009.
- [12] Fabio Mercurio, *LIBOR Market Models with Stochastic Basis*, Bloomberg Education and Quantitative Research Paper No. 2010-05-FRONTIERS.
- [13] Fabio Mercurio, Minqiang Li, *Jumping with default : wrong-way risk modelling for CVA*, Credit riks, 2015
- [14] Andreas E. Kyprianou : *Fluctuations of Lévy Processes with Applications. Introductory Lectures*. Springer, Universitext, Heidelberg 2014.
- [15] John C. Hull, Alan White, CVA and Wrong-Way Risk, *Financial Analysts Journal*, Vol. 68, No. 5, 2012
- [16] Juan Carlos Garcia Cespedes, Juan Antonio de Juan Herrero, Dan Rosen and David Saunders. Effective modeling of wrong way risk, counterparty credit risk capital and alpha in Basel II, *Journal of Risk Model Validation*, Volume 4, Number 1 (MARCH 2010), 71-98

-
- [17] Cont, R. & Tankov, P. (2004). *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, Boca Raton
- [18] Del Moral, P. : *Mean field simulation for Monte Carlo integration*. CRC Press (2013)
- [19] P. Del Moral, J. Garnier. *Genealogical Particle Analysis of Rare events*. *Annals of Applied Probability*, vol. 15, no. 4, 24962534 (2005).
- [20] Del Moral, P. : *Feynman-Kac Formulae : Genealogical and Interacting Particle Systems with Applications*. *Probability and Applications*. Springer, New York (2004).
- [21] E. Platen, *An approximation method for a class of Itô processes with jump component*, *Liet. Mat. Rink.* 22 (2) (1982) 124136.
- [22] M. Pykhtin, D. Rosen, *Pricing Counterparty Risk at the Trade Level and CVA Allocations*. *Journal of Credit Risk*, 6(4), 2010, pages 3-38
- [23] Li, A. (2007). *A jump diffusion model for contingent CDS valuation*. Available at SSRN.
- [24] Li, A. (2013). *Valuation of credit contingent options with applications to quanto CDS*. Available at SSRN.
- [25] Li, E. (2013). *CVA modeling for foreign and emerging market counterparties*. *Fixed Income Analytics*, Wells Fargo Bank.
- [26] M. Musiela, M. Rutkowski. *Martingale Methods in Financial Modelling*. Publisher, Springer, 1998.
- [27] J. Rosiński. *Series representations of Lévy processes from the perspective of point processes*. In *Lévy processes : Theory and Applications*, O.E. Barndorff-Nielsen and T. Mikosch and S. I. Resnick (eds), 2001.
- [28] Rosiński, J. (2007). *Tempering stable processes*, *Stochastic Processes and their Applications* 117, 677-707.
- [29] E. Platen, N. Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, Springer 2010
- [30] Samim Ghamami, Lisa R. Goldberg, *Stochastic Intensity Models of Wrong Way Risk : Wrong Way CVA Need Not Exceed Independent CVA*, FEDS Working Paper 2014-54.
- [31] W. Whitt, *Stochastic Process-Limits : an introduction to stochastic-process limits and their application to queues*, Springer 2002
- [32] arXiv :1402.4102
- [33] *Basel III : A global regulatory framework for more resilient banks and banking systems*, BCBS, 2011

Chapitre 3

Hamiltonian flow Simulation of Rare Events

« "Die Welt ist alles, was der Fall ist" »

Ludwig Wittgenstein

« "The world is all that is the case" »

Ludwig Wittgenstein

Sommaire

3.1 Introduction	57
3.2 Monte Carlo and Interacting Particle System	57
3.2.1 Problem Formulation	57
3.2.2 Algorithm	59
3.3 Hamiltonian flow Monte Carlo	59
3.3.1 Markov Chains on a Phase Space	59
3.3.2 Hamiltonian Flow's Integrator and Properties	61
3.3.3 Hamiltonian Flow Monte Carlo on Rare Events Sets	62
3.3.4 Algorithm	64
3.4 Convergence Analysis	65
3.4.1 IPS convergence	65
3.4.2 HFMC Convergence	66
3.5 Applications and Numerical Results	67
3.6 Conclusion and Further Research	69

Abstract

In this article we present a sampling method, Hamiltonian flow Monte Carlo, which is based on the simulation of Hamiltonian system to estimate rare events. The classical way to deal with this problem is the change of measure, so that our samples can reach rare events area, however, in practise it is hard to find an optimal importance density and the Monte Carlo simulation performs terribly, when the rare-event probability goes deep into the tail. The application of Hamiltonian dynamics allows to estimate rare events and sample from target distributions defined as the change of measures. Our approach uses a physical simulation of moving particles with momentum under the impact of the energy function to propose Markov chain transitions, that allows rapidly explore a state space. Its fast exploration can be explained by the fact that it extends the state space by an auxilliary momentum variables, P , and then runs a physical simulation to move long distances along probability contours in the extended state space.

The estimates demonstrated a variance reduction of the presented algorithm and its efficiency with respect to a standard Monte Carlo and interacting particle based system (IPS). We tested the algorithm on the case of the barrier option pricing.

3.1 Introduction

Hamiltonian flow based Monte Carlo simulations originates from physics have been used in many applications in statistics, engineering for a number of years. However these methods are not widely used in the estimation of rare events and in the financial option pricing practise.

This paper proposes Hamiltonian flow Monte Carlo technique for an efficient estimation of the rare event probability. Similarly to an importance sampling technique this method involves a change of probability measure. The random variables are sampled according to a modified probability measure that differs from a reference measure.

A rare event is the probability $\mathbb{P}(f(x_t) > a_t)$ for large values of x . One way to deal with this problem is to change an original measure, so that a_k is not too large in the new measure. Define the following set :

$$A_t = \{x \in \mathbb{R}^M, f(x_t) > a_t\} \quad (3.1)$$

$$\mathbb{E}^{\mathbb{P}}[\mathbb{1}_{A_t}] = \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{A_t} L_t] \quad (3.2)$$

where $L = \frac{d\mathbb{P}}{d\mathbb{Q}}$ is a Radon-Nykodim derivative.

Hamiltonian approach in the Monte Carlo context was developed by Duane et al. [11], R. Neal [15] where they proposed an algorithm for sampling probability distributions with continuous state spaces. The advantage of Hamiltonian based Monte Carlo is in the fact that we can extend the state space by including a momentum variable that will force in our context to move long distances in the state space in a single update step. We use this property of Hamiltonian dynamics to explore rarely-visited areas of the state space and efficiently estimate rare-event probability. Algorithm consists of two parts : simulation of Hamiltonian dynamics and Metropolis-Hastings test, that removes the bias and allows large jumps in the state space. We will show the performance of the algorithm on the Down-Out Barrier option technique with low level of a barrier.

HFMC can be considered within the optimal transportation problem, which was posed back in the 18-th century. Like HFMC, other simulation based approaches such as particle methods [5],[8] or the transportation using a homotopy [10] allow to move a set of particles from the measure \mathbb{P} to the measure \mathbb{Q} , by minimizing the transportation cost. We will show how the rare events estimation can be computed using interacting particle systems [6].

The paper is organized as follows. Section 2 introduces state of the art approach to estimate rare events : MC and PMC and formulates the problem within the context of Barrier option pricing. Section 3. describes Hamiltonian flow Monte Carlo algorithm. Section 4 presents the law of large numbers and convergence for HFMC. An adaptation of Hamiltonian flow Algorithm in the case of a Barrier option, numerical results and discussion is presented in Section 5. Section 6 concludes.

3.2 Monte Carlo and Interacting Particle System

3.2.1 Problem Formulation

Barrier option pricing is one of the cases when we encounter to the case of rare events. Consider a sequence of random variables $\{X_n\}_{n \geq 0}$, which in the financial context can be interpreted as asset prices, which forms Markov Chain on the space \mathbb{R}^{ns} . Given some stochastic process $\{X_t\}_{t \geq 0}$, for any test function f , we would like to compute the following expectation :

$$C = \mathbb{E}^{\mathbb{P}} [f(\mathbf{X}_T) \mathbb{1}_{X_{t \in [0, T]} \in A_t}] \quad (3.3)$$

One of the most popular ways to deal with this problem is importance sampling, when we replace the original statistical measure \mathbb{P} by an importance measure \mathbb{P}^δ . Then for $\mathbf{X}_n = (X_1, \dots, X_n)$.

$$C = \int f(\mathbf{X}_{n_t}) \frac{d\mathbb{P}}{d\mathbb{P}^\delta}(\mathbf{X}_{n_t}) d\mathbb{P}^\delta(\mathbf{X}_{n_t}) \prod_{n=1}^{n_t} \mathbb{1}_{X_n \in A_n} = \mathbb{E}^{\mathbb{P}^\delta} \left[f(\mathbf{X}_{n_t}) \frac{d\mathbb{P}}{d\mathbb{P}^\delta} \prod_{n=1}^{n_t} \mathbb{1}_{X_n \in A_n} \right] \quad (3.4)$$

where the rare event set A_n is given by :

$$A_n = \{X_n \in \mathbb{R}^{n_s}, f(X_n) > a_n\} \quad (3.5)$$

In practise we don't have an explicit form of the likelihood ratio $\frac{d\mathbb{P}}{d\mathbb{P}^\delta}$, so it becomes unfeasible unless one considers very simple toy examples. One of the solutions was interacting particle system (IPS), which was proposed by Del Moral and Garnier [8], where they proposed to generate particles (samples) in two steps, i.e. particle mutation and selection. The idea is to approximate the ratio of \mathbb{P} with respect to some importance measure \mathbb{P}^δ by choosing a weight function ω_n that approximates the Radon-Nikodym derivative $\frac{d\mathbb{P}}{d\mathbb{P}^\delta}$. If we assume that \mathbb{P} and \mathbb{P}^δ have density function p and \tilde{p} respectively, for n_s particles $\{X_n^{(m)}\}_{m=1}^{n_s}$ we can define the weight function iteratively by :

$$\prod_{n=1}^{n_t} \omega_n(\mathbf{X}_n^{(m)}) \propto \frac{d\mathbb{P}}{d\mathbb{P}^\delta}(\mathbf{X}_{n_t}^{(m)}) = \prod_{n=1}^{n_t} \frac{p_n(X_n^{(m)}, X_{n-1}^{(m)})}{p_n^\delta(X_n^{(m)}, X_{n-1}^{(m)})} \quad (3.6)$$

Since two measures \mathbb{P} and \mathbb{P}^δ form Markov chain, the Radon-Nykodim derivative can be decomposed into the product of ratio of the transition density $p_n(\cdot, X_{n-1}^{(m)})$ to the transition density with respect to the measure \mathbb{P}^δ .

The normalized importance weight function is given by :

$$W_n(\mathbf{X}_n^{(m)}) = \frac{\omega_n(\mathbf{X}_n^{(m)})}{\frac{1}{n_s} \sum_{j=1}^{n_s} \omega_n(\mathbf{X}_n^{(j)})} \quad (3.7)$$

The IPS estimate of an expectation (3.3) will have the following form :

$$\hat{C}^{\text{IPS}} = \frac{\mathbb{E} \left[f(\mathbf{X}_{n_t}) \prod_{n=1}^{n_t-1} \omega_n(\mathbf{X}_n) \mathbb{1}_{X_n \in A_n} \right]}{\mathbb{E} \left[\prod_{n=1}^{n_t-1} \omega_n(\mathbf{X}_n) \mathbb{1}_{X_n \in A_n} \right]} \quad (3.8)$$

In our experiments we choose a potential function (an unnormalized importance weight) of the form :

$$\omega_n(\mathbf{X}_{n_t}^{(m)}) = \prod_{n=1}^{n_t} e^{\delta(X_n^{(m)} - X_{n-1}^{(m)})} \quad (3.9)$$

where δ is an exponential tilting parameter. One issue with this approach is in the fact that an optimal choice of tilting parameter δ has to be judiciously chosen by running simulations and, in fact, it is fixed across all time steps $n = 1, \dots, n_t$.

3.2.2 Algorithm

The algorithm can be described by the following scheme, $\forall m = 1, \dots, n_S$:

$$X_n^{(m)} \xrightarrow{\text{Sampling}} \widehat{X}_{n+1}^{(m)} \xrightarrow{\text{Selection}} \Phi(\widehat{X}_{n+1}^{(m)}) = X_{n+1}^{(m)} \quad (3.10)$$

At each time step $n = 1, \dots, n_t$ we draw n_S independent random variables from the density $p_n^\delta(\cdot, X_{n-1}^{(m)})$ to construct n_S particles, $\widehat{\mathbf{X}}_n^{(m)} = (\widehat{X}_n^{(m)}, \widehat{\mathbf{X}}_{n-1}^{(m)})$. Given generated particles, we select, or draw independently n_S particles $\mathbf{X}_n^{(m)} = (X_0^{(m)}, \dots, X_n^{(m)})$ with replacement of rejected particles according to their probability weights :

$$W_n(\widehat{\mathbf{X}}_n^{(m)}) = \frac{\omega_n(\widehat{\mathbf{X}}_n^{(m)})}{\frac{1}{n_S} \sum_{j=1}^{n_S} \omega_n(\widehat{\mathbf{X}}_n^{(j)})} \quad (3.11)$$

And at time step n_t , we get the following IPS estimator :

$$\mathbb{E}[f(X_{n_t}) \prod_{n=1}^{n_t} \mathbb{1}_{X_n \in A_n}] \approx \frac{1}{n_S} \sum_{m=1}^{n_S} \left(f(\widehat{X}_{n_t}^{(m)}) \prod_{n=1}^{n_t} \omega_{n-1}(\mathbf{X}_{n-1}^{(m)}) \mathbb{1}_{\widehat{\mathbf{X}}_n^{(m)} \in A_n} \right)$$

Algorithm 8 : IPS algorithm

```

1 Initialization :  $n_S$  - #(simulations),  $n_t$  - #(time steps),  $X_0$  - initial value
2 for  $n = 1, \dots, n_t$  do
3   for  $m = 1, \dots, n_S$  do
4     Generate  $X_n^{(m)}$  from  $p(\cdot, X_{n-1}^{(m)})$  and set  $\widehat{\mathbf{X}}_n^{(m)} = (\widehat{X}_n^{(m)}, \mathbf{X}_{n-1}^{(m)})$ ;
5     if  $\widehat{X}_n^{(m)} \in A$  then
6        $\widehat{X}_n^{(m)} = 0$ 
7     else
8       Compute the weight :  $\omega_n(\widehat{\mathbf{X}}_n^{(m)})$ .
9     end
10    end
11  end
12  if  $n < N$  then
13    Resample using probability weight :  $W_n(\widehat{\mathbf{X}}_n^{(m)}) = \frac{\omega_n(\widehat{\mathbf{X}}_n^{(m)})}{\frac{1}{n_S} \sum_{j=1}^{n_S} \omega_n(\widehat{\mathbf{X}}_n^{(j)})}$  to sample  $\mathbf{X}_n^{(m)}$ .
14  end
15 end

```

3.3 Hamiltonian flow Monte Carlo

3.3.1 Markov Chains on a Phase Space

From section 2 we know that one of the ways to deal with rare event probabilities is to change a measure :

$$\mathbb{E}^{\mathbb{P}}[f(X_{n_t}) \prod_{n=1}^{n_t} \mathbb{1}_{X_n \in A_n}] = \mathbb{E}^{\mathbb{Q}}[f(X_{n_t}) \frac{d\mathbb{P}}{d\mathbb{Q}} \prod_{n=1}^{n_t} \mathbb{1}_{X_n \in A_n}] \quad (3.12)$$

To approximate Radon-Nikodym derivative $\frac{d\mathbb{P}}{d\mathbb{Q}}$ we will generate Markov Chain that will converge to an ergodic distribution. Let us introduce a random process X_u in a pseudo-time u and consider the following SDE, which is a gradient flow disturbed by a noise :

$$dX_u = -\nabla\psi(X_u)du + 2\sqrt{\beta^{-1}}dW_u \quad (3.13)$$

where $\psi(X) := -\log(p(X))$ is a potential. Under the assumption of ergodicity, the auto-correlated path X_u asymptotically, i.e. $u \rightarrow \infty$ draws samples from a stationary distribution :

$$\pi(X) = \frac{1}{\mathcal{Z}} \exp(\psi(X)) \quad (3.14)$$

where a normalizing constant \mathcal{Z} :

$$\mathcal{Z} = \int_{\mathbb{R}^n} e^{-\beta\psi(x)} dx \quad (3.15)$$

This can be seen as a unique solution of the following Fokker-Plank equation, given that ψ satisfies to some growth condition :

$$\frac{\partial p(t, x)}{\partial t} = \text{div}(\nabla(\psi(x)p)) + \beta^{-1} \Delta p \quad (3.16)$$

When we mentioned ergodicity, we meant, that for a class of regular functions $\phi : \mathbb{R}^X \rightarrow \mathbb{R}$ and x_0 a.s., the Markov Chain satisfies :

$$\frac{1}{L} \sum_{l=1}^L \phi(x_l) \rightarrow \int_{\mathbb{R}^X} \phi(x) \pi(dx) = \mathbb{E}^\pi[\phi(X)] \quad (3.17)$$

Observe that eq. (3.13) is a reversible process, which is interesting from theoretical point of view, but in practise the speed of convergence is not optimal. One of the ways to improve the convergence is to add a divergence-free drift b and consider the following modified SDE :

$$dX_u = (-\nabla\psi(X_u) + b(X_u))du + 2\sqrt{\beta^{-1}}dW_u \quad (3.18)$$

in order to satisfy detailed balance condition, we assume that $\nabla(b e^{-\psi}) = 0$.

Another way to improve the convergence is to consider a generalized Langevin SDE :

$$\ddot{X}_u^Y = -\nabla\psi(X_u^Y) - \dot{X}_u^Y + \sqrt{2\beta^{-1}}\dot{W}_u \quad (3.19)$$

We can rewrite it as :

$$\begin{cases} dX_u = P_u du \\ dP_u = -\nabla\psi(X_u)du - P_u du + \sqrt{\frac{2}{\beta}}dW_u \end{cases} \quad (3.20)$$

where the pair (X, P) is a kinetic process with X is the position and $P = \frac{dX}{du}$ is the velocity, that acts as an instantaneous memory.

The invariant function of the Markov process $\{x, P\}$, if it exists, is given by :

$$\pi_0(x, P) = \frac{1}{\mathcal{Z}} e^{-\beta\mathcal{H}(x, P)}, \quad \mathcal{Z} = \int_{\mathbb{R}^2} e^{-\beta\mathcal{H}(x, P)} dP dx \quad (3.21)$$

where

$$\mathcal{H}(x, P) = \frac{1}{2} P \mathcal{M}^{-1} P + \psi(x) \quad (3.22)$$

is a Hamiltonian function on \mathbb{R}^2 .

We will use Hamiltonian system to generate Markov Chain and approximate a Radon-Nikodym derivative $\frac{d\mathbb{P}}{d\mathbb{Q}}$. Hamiltonian flow Monte Carlo uses a physical simulation of moving particles with momentum under the impact of the energy function to propose Markov chain transitions, that allows rapidly explore state space. Its fast exploration can be explained by the fact that it extends the state space by an auxiliary momentum variables, P , and then runs a physical simulation to move long distances along probability contours in the extended state space.

We remind that, given Markov Chain $\{X_l\}_{l \geq 0}$, Birkhoff theorem says that

$$\frac{1}{n_S} \sum_{l=1}^{n_S} f(X_l) \xrightarrow{n_S \rightarrow \infty} \int f(x) d\pi(dx) = \varrho \text{ a.s.} \quad (3.23)$$

where ϱ is the expectation of $f(X)$ with respect to the unique invariant distribution π of Markov Chain.

3.3.2 Hamiltonian Flow's Integrator and Properties

We will use a configuration space \mathbf{M} with periodic boundary conditions. Each point on \mathbf{M} will be a set of n_S particles : $X^{(1)}, \dots, X^{(n_S)}$ and a generic momentum space \mathbb{R}^{n_S} , in this case the cotangent space is given by $T^* \mathbf{M} = \mathbb{R}^{n_S} \times \mathbb{R}^{n_S}$.

$$\begin{aligned} \Xi_u : T^* \mathbf{M} &\rightarrow T^* \mathbf{M} \\ (X, P) &\rightarrow \Xi_u(X, P) \end{aligned}$$

$\Xi_u(X_0, P_0)$ is the solution to the Hamilton's equation :

$$\begin{cases} dX_u = \mathfrak{M}^{-1} P_u du \\ dP_u = -\nabla \Psi(X_u) du \end{cases} \quad (3.24)$$

Hamiltonian system has three main properties : reversibility, conservation of energy and volume preservation.

Symplectic Integration of Hamiltonian Equations

In most cases we can not compute Hamiltonian flow in closed form and that is why we need to discretize the system (3.24). To make sure that we can preserve symplecticness and reversibility, we will discretize using leap-frog integrator, which is a symplectic integrator of Hamiltonian system.

Split Hamiltonian (3.22) into 3 parts :

$$\mathcal{H}_1 = \frac{1}{2} \Psi(X), \quad \mathcal{H}_2 = \frac{1}{2} \langle P, \mathfrak{M}^{-1} P \rangle, \quad \mathcal{H}_3 = \frac{1}{2} \Psi(X) \quad (3.25)$$

Taking each of these terms separately to be Hamiltonian function of Hamiltonian system gives rise to equations of motion with trivial dynamics.

$$\begin{cases} P_n(u + \frac{\Delta u}{2}) = P_n(u) - \frac{\Delta u}{2} \frac{\partial \Psi}{\partial x_n}(X_n(u)) \\ X_n(u + \Delta u) = X_n(u) + \Delta u P_n(u + \frac{\Delta u}{2}) \mathfrak{M}^{-1} \\ P_n(u + \frac{\Delta u}{2}) = P_n(u + \frac{\Delta u}{2}) - \frac{\Delta u}{2} \frac{\partial \Psi}{\partial x_n}(X_n(u + \Delta u)) \end{cases} \quad (3.26)$$

where Δu is the discretization size of Hamiltonian.

Consider a concatenation of three maps :

$$\Xi_n = \Xi_{\Delta u, \mathcal{H}_3} \circ \Xi_{\Delta u, \mathcal{H}_2} \circ \Xi_{\Delta u, \mathcal{H}_1} \quad (3.27)$$

where $\Xi_{\Delta u, \mathcal{H}_1} : (X(0), P(0)) \rightarrow (X(\Delta u), P(\Delta u))$. Similarly, $\Xi_{\Delta u, \mathcal{H}_1} = \Xi_{\Delta u, \mathcal{H}_3}$, and $\Xi_{\Delta u, \mathcal{H}_2}$ is calculated to be position update. Since the energy is preserved by the flow, the trajectories evolve on the submanifold of constant energy :

$$T^* \mathbf{M}(E_0) = \{(X, P) \in T^* \mathbf{M}; (\mathcal{H}(X, P) = E_0)\} \quad (3.28)$$

where $E_0 = \mathcal{H}(X_0, P_0)$ is the energy of the initialized data.

3.3.3 Hamiltonian Flow Monte Carlo on Rare Events Sets

Let $\mathcal{H}(X, P)$ be a Hamiltonian function on \mathbb{R}^{2ns} , where X is a potential, and P is a momentum variable of Hamiltonian system. The algorithm consists of two steps, first sampling from prior distribution values for potential and momentum and then a physical simulation of Hamiltonian dynamics. To make sure that at the end of each physical simulation of time step $n + 1$ we will have a probability measure, i.e. values will not exceed 1, we will use a Metropolis-Hastings test α_{n+1} , by choosing the minimum between 1 and the ratio of generated values of potential at time steps $n + 1$ and n , which is an acceptance probability of potential simulated by Hamiltonian dynamics. If we extend the state space $X = \{X_1, \dots, X_n\}$ and denote the extended space as $\tilde{X} = \{X_1, \dots, X_n, P_1, \dots, P_n\}$, we can denote the acceptance probability as :

$$\alpha_{n+1}(\tilde{X}_n, \tilde{X}_{n+1}) = 1 \wedge e^{(-\mathcal{H}(X_{n+1}, P_{n+1}) + \mathcal{H}(X_n, P_n)) \Delta t} \quad (3.29)$$

If we assume that the importance measure \mathbb{Q} admits the following importance distribution with a kernel \mathcal{K} :

$$q(d\tilde{X}_{n+1}) = \int_{\mathbb{R}^{2M}} p(\tilde{X}_n) \mathcal{K}(\tilde{X}_n, d\tilde{X}_{n+1}) d\tilde{X}_n \quad (3.30)$$

Then, the associated Radon-Nikodym derivative will have the following form :

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(\tilde{X}_{n+1}) = \frac{d\mathbb{P}(\tilde{X}_{n+1})}{\int_{\mathbb{R}^{2M}} p(\tilde{X}_n) \mathcal{K}(\tilde{X}_n, d\tilde{X}_{n+1}) d\tilde{X}_n} \quad (3.31)$$

Assume that at each times step n we have n_s sample of r.v. $\{X_n^{(m)}\}_{m=1}^{n_s}$. Now we can define a transition kernel \mathcal{K} as follows.

Definition Consider a mapping $\Xi_n : \tilde{X}_n^{(m)} \rightarrow \tilde{X}_{n+1}^{(m)}$, which is a transformation in \mathbb{R}^{2ns} , $u \sim \text{Unif}[0, 1]$. Then a transition kernel $\mathcal{K}(\cdot, d\tilde{X}_{n+1}^{(m)})$ is given by :

$$\mathcal{K}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)}) = \mathbb{1}_{u \leq \alpha_{n+1}} \Xi_n(\tilde{X}_n^{(m)}) d\tilde{X}_{n+1}^{(m)} + \mathbb{1}_{u > \alpha_{n+1}} \tilde{X}_n^{(m)} \delta_{\tilde{X}_n^{(m)}}(d\tilde{X}_{n+1}^{(m)}) \quad (3.32)$$

This kernel can be interpreted as the probability to move from the point $\tilde{X}_n^{(m)}$ to a new proposed point $\tilde{X}_{n+1}^{(m)}$, which is simulated through a discretized Hamiltonian flow $\Xi_n(\cdot)$. If the proposed step is not accepted, then next step is the same as the current step, i.e. $\tilde{X}_{n+1}^{(m)} = \tilde{X}_n^{(m)}$. This procedure allows as to leave the joint distribution of $X_n^{(m)}$ and $P_n^{(m)}$ invariant. Volume preservation means that the determinant of the Jacobian matrix of a transformation Ξ_n is equal to one.

We will need basic property of symplectic integrators, i.e. reversibility.

Lemma 3.3.1 *The integrator Ξ_n is reversible.*

We refer to [13] for the proof of this result.

Assumption 3.3.2 — *The potential $\Psi \in \mathcal{C}^1$ is bounded from above;*
 — *The gradient $\nabla\Psi$ is a globally Lipschitz function.*

Lemma 3.3.3 *If the potential Ψ satisfies to the assumption 3.3.2, then the kernel \mathcal{K} is irreducible and Markov Chain satisfies*

$$\forall x \in \mathbf{M}, \forall B \in \mathcal{B}(\mathbf{M}), \mu^{\text{Leb}}(B) > 0, \mathcal{K}(x, B) > 0 \quad (3.33)$$

Proof We refer to [2].

Proposition 3.3.4 *Given that the assumption 3.3.2 holds, then for $n = 1, \dots, n_S$, the irreducible Markov Chain defined by a transformation Ξ_n is reversible under the distribution π :*

$$\pi(d\tilde{X}_n^{(m)}) \mathcal{K}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)}) = \pi(d\tilde{X}_{n+1}^{(m)}) \mathcal{K}(\tilde{X}_{n+1}^{(m)}, d\tilde{X}_n^{(m)}) \quad (3.34)$$

Thus $\pi(x)$ is the invariant distribution of Markov Chain $\{\tilde{X}_n\}_{n=1}^{n_S}$.

Proof Rewrite the kernel \mathcal{K} as :

$$\mathcal{K}(x, dy) = \alpha(x, y) \Xi_n(x) dy + \wp(x) \delta_x(dy) \quad (3.35)$$

where

$$\wp(x) = 1 - \int \alpha(x, z) \Xi_n(x) dz \quad (3.36)$$

$$\begin{aligned} \int \mathcal{K}(x, B) \pi(x) dx &= \int \left[\int_B \alpha(x, y) \Xi_n(x) dy \right] \pi(x) dx + \int \wp(x) \delta_x(B) \pi(x) dx = \\ &= \int_B \left[\int \pi(x) \alpha(x, y) \Xi_n(x) dx \right] dy + \int_B \wp(x) \pi(x) dx = \\ &= \int_B \left[\int \pi(y) \alpha(y, x) \Xi_n(y) dx \right] dy + \int_B \wp(x) \pi(x) dx \\ &= \int_B \pi(y) (1 - \wp(y)) dy + \int_B \wp(x) \pi(x) dx = \int_B \pi(y) dy \quad (3.37) \end{aligned}$$

Corollary 3.3.5 *The kernel \mathcal{K} satisfies reversibility condition with an indicator function of the rare event set :*

$$\pi(x) \mathcal{K}(x, y) \mathbb{1}_{x \in A} = \pi(y) \mathcal{K}(y, x) \mathbb{1}_{y \in A} \quad (3.38)$$

Now we can define rare event transitions through the kernel \mathcal{M} .

Definition Assume that the assumption 3.3.2 holds and consider Markov Chain $(X_n^{(m)})_{n \geq 1}$ with an initial prior $p_1(X_1)$ and define the following transition kernel $p(\tilde{X}_{n+1}^{(m)} \in d\tilde{X}_{n+1}^{(m)} | \tilde{X}_n^{(m)}) = \mathcal{M}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)})$.

$$\mathcal{M}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)}) = \mathcal{K}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)}) \mathbb{1}_{\mathcal{K}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)}) \in A_n} + \tilde{X}_n^{(m)} \mathbb{1}_{\mathcal{K}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)}) \notin A_n} \quad (3.39)$$

It means that the point $\tilde{X}_n^{(m)}$ moves to a new point $\tilde{X}_{n+1}^{(m)}$ only if it is inside a rare event set A_n , otherwise we stay at point $\tilde{X}_n^{(m)}$.

Proposition 3.3.6 Let $n = 1, \dots, n_t$. Markov chain X_n is invariant under the kernel $\mathcal{M}(\cdot, dX_{n+1})$.

Proof

$$\begin{aligned} \int \pi(dx) \mathcal{M}(x, dy) \mathbb{1}_{x \in A} &= \int \pi(dx) [\mathbb{K}(x, y) \mathbb{1}_{x \in A} + \mathbb{K}(x, A^c) \delta_x(dy)] \mathbb{1}_{x \in A} = \\ &= \int \int \pi(dx) \mathbb{K}(x, dz) [\mathbb{1}_{z \in A} \delta_z(dy) + \mathbb{1}_{A^c}(z) \delta_x(dy)] \mathbb{1}_{x \in A} = \\ &= \int \pi(dx) \mathbb{K}(x, dy) \mathbb{1}_{y \in A} \mathbb{1}_{x \in A} + \int \pi(dy) \mathbb{K}(y, dz) \mathbb{1}_{A^c}(y) \mathbb{1}_{x \in A} = \pi(dy) \mathbb{1}_{y \in A} \end{aligned} \quad (3.40)$$

Invariance of $\tilde{X}_n^{(m)}$ says that for any bounded and measurable function f , the distribution of $f(\mathcal{M}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)}))$ and $f(\tilde{X}_n^{(m)})$ is the same.

$$\mathbb{E}[f(\mathcal{M}(\tilde{X}_n^{(m)}, d\tilde{X}_{n+1}^{(m)}))] = \mathbb{E}[f(\tilde{X}_n^{(m)})] \quad (3.41)$$

Under the kernel \mathcal{M} of \tilde{X}_n , the final HFMC estimate is given by :

$$\hat{C}^{\text{HFMC}} = \frac{1}{n_S} \sum_{m=1}^{n_S} f(X_{n_t}^{(m)}) \mathbb{1}_{\{X_{n+1}^{(m)}, X_n^{(m)} \in A_n\}} \quad (3.42)$$

3.3.4 Algorithm

Hamiltonian function is defined by $\mathcal{H}(X, P) = \Psi(X) + \frac{1}{2} P^T \mathfrak{M}^{-1} P$, where $\Psi(X)$ - is a potential energy function, and the second term is a kinetic energy function with a momentum variable P and mass matrix \mathfrak{M} . Usually one sets a mass matrix \mathfrak{M} to be an identity matrix I . The proposed samples are obtained by a physical simulation of Hamiltonian dynamics :

$$\begin{cases} dX_u = \mathfrak{M}^{-1} P_u du \\ dP_u = -\nabla \Psi(X_u) du \end{cases} \quad (3.43)$$

We start by simulating M random variables from a prior $X_1 = p_0(\cdot, X_0^{(m)})$, which is the density of the underlying SDE and generating M random variables from gaussian distribution for momentum $\{P_0^{(m)}\}_{m=1}^M$.

For each step $n = 1, \dots, N$ we set $x_H^{(m)} = X_n^{(m)}$, $P_H^{(m)} = P_n^{(m)}$. The proposed new candidates are obtained after L -leapfrog steps of the simulation of Hamiltonian dynamics and they are defined by $x^* = x_H^{(m)}(L)$ and $P^* = P_H^{(m)}(L)$. These new set of proposed candidates are then accepted according to the following Metropolis-Hastings test. First generate uniformly distributed random variable $u \sim \mathcal{U}ni f(0, 1)$, then compute α :

$$\alpha = 1 \wedge e^{(-\mathcal{H}(x^*, P^*) + \mathcal{H}(x_H^{(m)}, P_H^{(m)})) \Delta t}; \quad (3.44)$$

If proposed candidates (x^*, P^*) are accepted, i.e. $\alpha > u$ we set $X_{n+1}^{(m)} = x^*$, and if they are rejected, i.e. $\alpha \leq u$, we set $X_{n+1}^{(m)} = x_H^{(m)}$. At the end, calculate estimator in (3.42). The main steps of the algorithm are summarized in Algorithm 2.

The Metropolis-Hastings test insures a volume preservation. That explains the fact that we don't need to compute a normalizing constant in our algorithm. Volume preservation means that the absolute value of the Jacobian matrix of the leapfrog integrator is

equal to one, this is because candidates are proposed through simulation of Hamiltonian flow.

Algorithm 9 : Hamiltonian flow Monte Carlo in Rare event setting

```

1 Initialization :  $n_S$  - #(simulations),  $n_t$  - #(time steps)
2 for  $n = 1, \dots, n_t$  do
3   for  $m = 1, \dots, n_S$  do
4     Generate  $X_n^{(m)}$  from prior  $\tilde{p}(X_0^{(m)}, \cdot)$ ;
5     Simulate initial momentum  $P_1^{(m)} \sim \mathcal{N}(0, I_M)$ , set  $x_H^{(m)} = X_n^{(m)}$  and run
     Hamiltonian flow :
6     for  $l_f = 1, \dots, L - 1$  do
7        $P_H^{(m)}((l_f + \frac{1}{2})\delta) = P_H^{(m)}(l_f) - \frac{\delta}{2} \frac{\partial \Psi}{\partial x_H}(x_H^{(m)}(l_f))$ 
        $x_H^{(m)}((l_f + 1)\delta) = x_H^{(m)}(l_f) + \delta P_H^{(m)}((l_f + \frac{1}{2})\delta) I_M^{-1}$ 
        $P_H^{(m)}((l_f + 1)\delta) = P_H^{(m)}((l_f + \frac{1}{2})\delta) - \frac{\delta}{2} \frac{\partial \Psi}{\partial x_H}(x_H^{(m)}((l_f + 1)\delta))$ 
8     end
9     Calculate acceptance probability and set  $x^* = x_H^{(m)}(L)$ ,  $P^* = P_H^{(m)}(L)$  :
           
$$a = 1 \wedge e^{(-\mathcal{H}(x^*, P^*) + \mathcal{H}(x_H^{(m)}, P_H^{(m)})) \Delta t}; \tag{3.45}$$

       Draw  $u \sim \mathcal{U}nif(0, 1)$ ;
10    if  $u < a$  then
11      Set  $X_{n+1}^{(m)} = x^*$ ;
12    else
13      | Reject, and set  $X_{n+1}^{(m)} = x_H^{(m)}$ 
14    end
15  end
16  if  $X_n^m, X_{n+1}^m \in A$  then
17    | Set  $X_n^m = 0, X_{n+1}^m = 0$ 
18  end
19 end
20 end
21 Compute :
```

$$\widehat{C}^{\text{HFMC}} = \frac{1}{n_S} \sum_{m=1}^{n_S} \left(f(X_{n_t}^{(m)}) \prod_{n=1}^{n_t} e^{(-\mathcal{H}(X_{n+1}^{(m)}, P_{n+1}^{(m)}) + \mathcal{H}(X_n^{(m)}, P_n^{(m)})) \Delta t} \mathbb{1}_{X_n^{(m)}, X_{n+1}^{(m)} \in A_n} \right)$$

3.4 Convergence Analysis

3.4.1 IPS convergence

IPS convergence, and in particular the asymptotic behaviour as number of particles $n_S \rightarrow \infty$ was thoroughly studied in [6].

The following result given in [5] allows a non asymptotic control of variance of the rare event probability.

Assumption 3.4.1

$$\tilde{\delta}_n := \sup_{x, y} \frac{\omega_n(x)}{\omega_n(y)} < +\infty \tag{3.46}$$

Theorem 3.4.2 *When the assumption (3.4.1) is met for some $\tilde{\delta}_n$, we have the nonasymptotic estimates :*

$$\mathbb{E} \left[\left| \frac{C^{\text{IPC}}}{C} - 1 \right|^2 \right] \leq \frac{4}{n_S} \sum_{s=1}^{n_t} \frac{\widehat{\delta}_s^{(n_t)}}{p_k} \quad (3.47)$$

where $\widehat{\delta}_s^{(n_t)} = \prod_{s \leq k < s+n_t} \tilde{\delta}_k$

3.4.2 HFMC Convergence

LLN and Convergence Rate

Birkhoff ergodic theorem allows us have law of large numbers(LLN) like convergence. So, we are interested in a sigma-algebra \mathcal{G} of invariant events, in particular when \mathcal{G} is trivial.

From lemma 3.3 we know that Markov chain generated by HFMC is irreducible, and we can see that Markov Chain that we get from the rare event kernel \mathcal{M} satisfies irreducibility conditions due to the fact that the transition density is always positive. Applying the results by [17], we have :

Proposition 3.4.3 [17] *Suppose that \mathcal{M} is a π -irreducible Metropolis kernel. Then \mathcal{M} is a Harris recurrent.*

Proposition 3.4.4 [17] *If \mathcal{M} is positive Harris and aperiodic then for every initial distribution λ :*

$$\| \int \lambda(dx) (\mathcal{M})^l(x, \cdot) - \pi \|_{\text{TV}} \rightarrow 0, \quad l \rightarrow \infty \quad (3.48)$$

for π almost all x .

where $\|\cdot\|_{\text{TV}}$ is a total variation distance.

Geometric Ergodicity

To establish central limit theorem (CLT), we need a geometric ergodicity of the chain.

Definition A subset C of that state space $(\mathbb{R}^{n_s}, \mathcal{B}(\mathbb{R}^{n_s}))$ is petite if there exists a non-zero positive measure ν on the state space and subsampling distribution q such that

$$\mathcal{K}_q(x, A) \geq \nu(A), \quad \forall A \in \mathcal{B}(\mathbb{R}^{n_s}) \text{ and } x \in C \quad (3.49)$$

Definition A subset C of that state space $(\mathbb{R}^{n_s}, \mathcal{B}(\mathbb{R}^{n_s}))$ is small if there exists a non-zero positive measure ν on the state space and real-valued number $l \in \mathbb{R}$ such that

$$\mathcal{K}^l(x, A) \geq \nu(A), \quad \forall A \in \mathcal{B}(\mathbb{R}^{n_s}) \text{ and } x \in C \quad (3.50)$$

Observe, that every small set is petite.

Theorem 3.4.5 *Suppose for an irreducible, aperiodic Markov chain having transition probability kernel \mathcal{K} and a state space \mathbb{R}^{n_s} , there exists a petite set C areal valued function V , satisfying $\nu \geq 1$, and constants $b < \infty$ and $\lambda < 1$ such that*

$$\mathcal{K}V(x) \leq \lambda V(x) + b \mathbb{1}_C(x), \quad \forall x \in \mathbb{R}^{n_s} \quad (3.51)$$

holds. Then the chain is geometrically ergodic.

The function V is called a geometric drift. Take the expectation of the both sides of (3.51) and using the invariance of measure π with respect to the kernel \mathcal{K} :

$$\mathbb{E}^\pi[V(X)] \leq \frac{b\pi(C)}{(1-\lambda)} \quad (3.52)$$

In other words, for $\lambda \in (0, 1]$ a function satisfying (3.51) is always π -integrable.

Proposition 3.4.6 *Assume that there exist $\lambda \in [0, 1)$ and $b \in \mathbb{R}_+$ such that*

$$\mathcal{K}V \leq \lambda V + b \quad (3.53)$$

and

$$\limsup \mathcal{K}(x, \mathcal{B}(x) \cap \mathcal{B}(x)) = 0 \quad (3.54)$$

It was shown in [12] that under certain conditions, HFMC kernel is geometrically ergodic.

3.5 Applications and Numerical Results

We will test our algorithm on down-out(DOC) Barrier option pricing, and compare its estimate with a standard Monte Carlo and particle Monte Carlo methods. Lets consider a toy example and assume that our asset follows the following SDE :

$$dX_t = \mu X_t dt + \sigma X_t dW_t \quad (3.55)$$

where μ is a drift, σ is a constant volatility parameter. European DOC call Barrier option is a usual call option contract that pays a payoff $\max(S_T - K, 0)$, provided that the asset price S has not fallen below a barrier B during the lifetime of the option. If the pricing process ever reaches the barrier B , then the option becomes worthless.

We use Euler-Muruyama discretization scheme and we use the following notation $X_{t_n} := X_n$, so for a time discretization : $0 = t_0, t_1, \dots, t_{n_t} = T$, the solution of the SDE in (3.55) :

$$X_n = X_{n-1} e^{(\mu - 0.5\sigma^2)\Delta t + \sigma\Delta t\epsilon_n} \quad (3.56)$$

The DOC barrier call option price of a discretely monitored barrier at maturity T is :

$$C = e^{-(r-q)T} \mathbb{E}[g(X_{n_t}) \prod_{n=1}^{n_t} \mathbb{1}_{X_{t \in [t_{n-1}, t_n]} \in A_n}] \quad (3.57)$$

where r, q are respectively an interest and a dividend rates, $g(x) = (x - K)^+$ is a payoff function and the set A_n in the case of a DOC barrier call option :

$$A_n = \inf_{t_{n-1} \leq t \leq t_n} \{t : X_t > B\}$$

We use continuity correction that was proposed in [1] : $B = B \exp^{-0.5826\sigma\Delta t}$.

HFMC estimator to compute DOC call option is given by :

$$\hat{C}^{\text{HFMC}} = e^{-(r-q)T} \frac{1}{n_S} \sum_{m=1}^{n_S} \left(g(X_{n_t}^{(m)}) \prod_{n=1}^{n_t} e^{(-\mathcal{H}(X_{n+1}^{(m)}, P_{n+1}^{(m)}) + \mathcal{H}(X_n^{(m)}, P_n^{(m)}))\Delta t} \mathbb{1}_{X_n^{(m)}, X_{n+1}^{(m)} \in A_n} \right) \quad (3.58)$$

Monte Carlo estimate is given by :

$$\widehat{C}^{\text{MC}} = e^{-(r-q)T} \frac{1}{n_S} \sum_{m=1}^{n_S} \left(g(X_{n_t}^{(m)}) \prod_{n=1}^{n_t} \mathbb{1}_{X_n^{(m)} \in A_n} \right) \quad (3.59)$$

The IPS estimator is given by :

$$\widehat{C}^{\text{IPS}} = e^{-(r-q)T} \frac{1}{n_S} \sum_{m=1}^{n_S} \left(g(\widehat{X}_{n_t}^{(m)}) \prod_{t=1}^{n_t} W_{n-1}(\mathbf{X}_{n-1}^m) \mathbb{1}_{\widehat{\mathbf{X}}_n \in A_n} \right)$$

In the context of a rare event, we chose the barrier level at 65, with an initial price $X_0 = 100$, Strike $K = 100$, interest rate $r = 0.1$, volatility $\sigma = 0.3$, $T = 0.5$ and zero dividends $q = 0$. In the table 1 and 2, Hamiltonian flow MC, MC and IPS are presented. We used 50000 and 75000 particles with 750 equally spaced time steps in Table 1 and Table 2.

It is very important to choose the number and the size of leapfrog steps. We chose them such that the acceptance probability α is bigger than 0.8.

We compare each approach by estimating the standard deviations, root mean squared error (RMSE), bias, relative mean squared error(RRMSE), time required to compute each estimate and the figure of merit (FOM). We run 20 MC experiments. The RMSE estimator is given by :

$$\text{RMSE} = \sqrt{\frac{1}{M_s} \sum_{l=1}^{M_s} \|C - \widehat{C}_l\|^2} \quad (3.60)$$

where C is price computed analytically, \widehat{C}_l are Monte Carlo estimates and M_s is the number of Monte Carlo experiments.

The RRMSE is computed using the following formula :

$$\text{RRMSE} = \frac{\text{RMSE}}{\widehat{C}} \quad (3.61)$$

To measure the efficiency of each method presented in the article, we will use the figure of merit(FOM) :

$$\text{FOM} = \frac{1}{R^2 \times \text{CPU}_t} \quad (3.62)$$

where CPU_t is CPU time need to compute the estimator and R is a relative error, which is the measure of a statistical precision :

$$R = \frac{\text{St.dev}}{\widehat{C}} \propto \frac{1}{\sqrt{n_S}} \quad (3.63)$$

where $\bar{C} = \sum_{l=1}^{M_s} \widehat{C}_l$.

We run 20 independent Monte Carlo experiments for each estimate. Since IPS and the simulation of Hamiltonian dynamics requires more time to compute an estimate, we use the figure of merit to compare three approaches. From the table 1 and 2 we can observe that HFMC demonstrates standard deviations, bias and relative RMSE.

[H]

TABLEAU 3.1 – DOC Barrier option estimates statistics. $B = 65, X_0 = 100, K = 100, r = 0.1, \sigma = 0.3, T = 1/2$, and $div = 0; \delta = 0.0001, \#(\text{Leap frog step}) : 35$. True price : 10.9064, $n_S = 50000, n_t = 750$

Stat	MC	PMC	HFMC
St. dev.	0.088518965	0.08562686	0.065318495
RMSE	0.007011127	0.008004332	0.0143
RRMSE	0.001298078	0.000292621	1.87148E-05
CPU time	3.7251	4.8432	5.90675
FOM	4097.9	3387.2	4737.6

[H]

TABLEAU 3.2 – DOC Barrier option estimates statistics. $B = 65, X_0 = 100, K = 100, r = 0.1, \sigma = 0.3, T = 1/2$, and $div = 0; \delta = 0.0009, \#(\text{Leap frog step}) : 40$. True price : 10.9064, $n_S = 75000, n_t = 750$

Stat	MC	PMC	HFMC
St. dev.	0.062385996	0.044259477	0.038039517
RMSE	0.037561882	0.051285344	0.037561882
RRMSE	0.000355199	0.000240548	0.000129293
CPU time	2.2626	6.0322	7.6832
FOM	13475.2	10117.7	10711.0

3.6 Conclusion and Further Research

We proposed an importance sampling algorithm based on the simulation of Hamiltonian system, that generates Markov Chain that follows along the gradient of the target distributions over large distances of the state space, while producing low-variance samples.

From the simulated results we saw that HFMC allows efficiently estimate rare event probabilities, which we tested on the case of DOC Barrier options. Its estimates show lower variance and bias than that of MC and IPS.

It will interesting to adapt a stochastic gradient Hamiltonian Monte Carlo algorithm [6], when one can avoid computing the gradient at each simulations. Taking into account the big data problem and the necessity of online estimations, we can get sufficient improvements. Another extension is the adaptation to the Riemann Manifold Hamiltonian Monte Carlo [14], when we can create a statistical manifold and tune HFMC by computing explicitly the mass matrix M in the kinetic energy of the algorithm.

In the next article we will show the performance of mixed IPS and Hamiltonian flow Monte Carlo. It will allow faster explore the state space on the one hand, and push trajectories into rare event area on the other hand. By resampling we can reduce the correlation between generated from Hamiltonian system Markov chains.

Bibliographie

- [1] M. Broadie, P. Glasserman, and S.G. Kou. A Continuity Correction for Discrete Barrier Options. *Mathematical Finance*, 7 :325-349, 1997.
- [2] E Cances, F Legoll, G Stoltz *ESAIM : Mathematical Modelling and Numerical Analysis* 41 (2), 351-389
- [3] Carmona, Rene, Jean-Pierre Fouque & Douglas Vestal (2009), Interacting particle systems for the computation of rare credit portfolio losses, *Finance and Stochastics* 13(4), 613633.
- [4] Carmona, Rene, Stephane Crepey (2010), Particle methods for the estimation of Markovian credit portfolio loss distributions, *International Journal of Theoretical and Applied Finance* 13(4), 577602.
- [5] Cerou, E; Del Moral, P; Guyader, A. A nonasymptotic theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. H. Poincaré Probab. Statist.* 47 (2011), no. 3, 629–649. doi :10.1214/10-AIHP358. <http://projecteuclid.org/euclid.aihp/1308834852>.
- [6] Chen, Tianqi, Emily B. Fox, and Carlos Guestrin. "Stochastic Gradient Hamiltonian Monte Carlo." *ICML*. 2014.
- [7] Del Moral, P. : Mean field simulation for Monte Carlo integration. CRC Press (2013)
- [8] P. Del Moral, J. Garnier. Genealogical Particle Analysis of Rare events. *Annals of Applied Probability*, vol. 15, no. 4, 24962534 (2005).
- [9] Del Moral, P. : Feynman-Kac Formulae : Genealogical and Interacting Particle Systems with Applications. *Probability and Applications*. Springer, New York (2004).
- [10] Douady, Raphael, and Miryusupov, Shohruh. "Optimal Transport Filtering with Particle Reweighting in Finance." *arXiv preprint arXiv :1704.07698* (2017).
- [11] Duane, S, Kennedy, AD, Pendleton, BJ, and Roweth, D. Hybrid monte carlo. *Physics letters B*, 1987.
- [12] Durmus, A., Moulines, E. and Saksman, E. On the convergence of Hamiltonian Monte Carlo, *arXiv preprint arXiv :1705.00166* (2017)
- [13] Hairer, E. and Söderlind, G. Explicit, Time Reversible, Adaptive Step Size Control. *SIAM Journal on Scientific Computing*, 2005, Vol. 26, No. 6 : pp. 1838-1851
- [14] Girolami, M. and Calderhead, B. (2011), Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73 : 123214. doi :10.1111/j.1467-9868.2010.00765.x
- [15] Neal, Radford M. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, January 2010.
- [16] Risken, H. (1989). *The Fokker-Planck Equation*, second edn, Springer, Berlin, Heidelberg, New York.

- [17] S. Meyn and R. Tweedie. Markov Chains and Stochastic Stability. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
- [18] Tierney, Luke. Markov Chains for Exploring Posterior Distributions. *Ann. Statist.* 22 (1994), no. 4, 1701–1728. doi :10.1214/aos/1176325750. [http ://projecteuclid.org/euclid.aos/1176325750](http://projecteuclid.org/euclid.aos/1176325750).

Chapitre 4

Optimal Transport Filtering with Particle Reweighing in Finance

« "Comme est pauvre la langue de la joie ! Tout ce qui fut sera encore et seul est doux l'instant de la reconnaissance" »

Ossip Mandelstam

« "How poor is the language of happiness ! Everything's happened before and will happen again, but still the moment of each meeting is sweet" »

Ossip Mandelstam

Sommaire

4.1 Introduction	75
4.2 Particle Filtering	75
4.2.1 Problem formulation	75
4.2.2 Particle Filtering Algorithm	78
4.3 Homotopy Transport	79
4.3.1 Homotopy Transport Algorithm	82
4.4 Homotopy Transport with Particle Reweighing	83
4.4.1 PF-Enhanced Homotopy Transport Algorithm	84
4.5 Numerical Applications and Results	85
4.6 Conclusions and Further Research	88

Abstract

In this paper we show the application of an optimal transportation problem to estimate stochastic volatility process by using a flow that optimally transports the set of particles from the prior to a posterior distribution. Importantly, it is well-known that a naive particle filter has curse of dimensionality, when the quality of weights deteriorates as the number of iteration increases. Homotopy transportation approach allows efficiently estimate stochastic volatility in high-dimensional problems. We also show how to direct the flow to a rarely visited areas of the state space by using a particle method (a mutation and a reweighing mechanism). The efficiency of our approach is demonstrated on the simple example of the European option price under the Stein-Stein stochastic volatility model for which a closed form formula is available. Both homotopy and reweighted homotopy methods show a lower variance, root-mean squared errors and a bias compared to other filtering schemes recently developed in the signal-processing literature, including particle filter techniques.

4.1 Introduction

Optimal transport problem that was formulated by Monge in the XVIII century, then reformulated fifty years ago by Kantorovich and it has recently been rediscovered by C. Villani [12]. This problem then was applied in different contexts, for example in option pricing [11].

Particle methods which were extensively researched by P. Del Moral in [6], [5] and [7] allow to find so-called "optimal transport". For this purpose a set of discrete weighted samples, i.e. particles, is used to approximate an importance measure, and then to predict a posterior distribution by propagating the set of particles until we get an estimate.

Another approach has been proposed by Daum's et al. [2], [3] that allows the reduction of the number of particles we need in order to get a tolerable level of errors in the filtering problem. The main idea behind this method is the evolution in homotopy parameter λ (a "pseudotime") from prior to the target density. They introduced a particle flow, in which particles are gradually transported without the necessity to randomly sample from any distribution. This approach as an optimal transport problem allows optimally move the set of particles according to Bayes' rule. In other words, the particles are progressively transported according to their flow. One can in this way reduce the number of needed samples, since the variance and bias of the estimator is lower and as a result reduce the computational burden in both the estimation and the prediction steps.

In this paper we adapt homotopy transport in Stein-Stein stochastic volatility model [9] to price a European option and extend Daum's et al. method by reweighing the generated particles' trajectories that allows to efficiently transport the particles from a prior transition density to a posterior one under the measurement impact. The idea of transportation and reweighing mechanism is to transport particles through the sequence of densities that move the least during the synthetic time until they reach the posterior distribution. By regenerating particles according to their weight at each time step we are able to direct the flow and further minimize the variance of the estimates. The transportation of particles can be understood as a geodesic flow in a convex subset of a Euclidean space.

We show that homotopy transport allows to significantly reduce the variance compared to a particle filtering technique. Path reweighing allows further reduce both the variance and the bias of estimators.

The rest of the article is organized as follows. Section 2 formulates the problem of computing the expectation when we have partially observed variables and shows the solution using particle filter method. Section 3 formulates the problem defined in section 2 in the context of optimal transport and presents the homotopy transport approach to solve the problem. Section 4 shows the mixture of homotopy transport and path reweighing approach and, actually, extends the method proposed in section 3. Section 5 provides numerical results. Section 6 concludes.

4.2 Particle Filtering

4.2.1 Problem formulation

Many problems arises in financial applications when one has to compute expectations with partially observed information. A simple example is an option pricing with hidden volatility dynamics. Assume that we denote by $\{Y_t\}_{t \geq 0} \in \mathbb{R}^{n_Y}$ asset returns, which are observed from the dynamics of prices, while the hidden factor $\{X_t\}_{t \geq 0} \in \mathbb{R}^{n_X}$ is unobservable. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and the set of observed data at time t be (\mathcal{F}_t)

is a filtration generated by a process $(Y_t)_{t \geq 0}$.

The classical problem, where particle filtering is applied, is to extract a sequence of hidden variables X_t . It is formalized in the following way, given an initial \mathbb{R}^{n_x} -dimensional random variables x_0 with distribution \mathbb{P}_{x_0} , then for $t \in \mathbb{N}$:

$$\begin{cases} X_t = f(X_{t-1}, \epsilon_t) \\ Y_t = h(X_t, Y_{t-1}, \eta_t) \end{cases} \quad (4.1)$$

where the first equation is the hidden process, with $\epsilon_t : \Omega \rightarrow \mathbb{R}^{n_x}$ are i.i.d random variables, the map $f : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x}$ is $\mathcal{B}(\mathbb{R}^{n_x})$ -measurable. The second equation is called a measurement model with $\eta_t : \Omega \rightarrow \mathbb{R}^{n_y}$ are i.i.d. random variables and the map $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_y}$ is $\mathcal{B}(\mathbb{R}^{n_x}) \otimes \mathcal{B}(\mathbb{R}^{n_y})$ -measurable.

Given above stochastic dynamical system, we would like to compute the following conditional expectation :

$$\mathbb{E}[z(X_t) | \mathcal{F}_t] = \frac{1}{\mathcal{Z}} \int v(dx_{0:t}) \rho_t(x_{0:t}, Y_{1:t}) z(x_t) \quad (4.2)$$

with a distribution of $X_{0:t}$:

$$v(dx_{0:t}) = p_0(dx_0) \prod_{l=0}^t k_l(X_{l-1}, X_l) \mu(dx_l) \quad (4.3)$$

and normalizing constant \mathcal{Z} :

$$\mathcal{Z} = \int v(dx_{0:t}) \rho_t(x_{0:t}, Y_{1:t}) \quad (4.4)$$

where $(X_t)_{t \geq 0}$ forms a Markov Chain in $(\mathbb{R}^{n_x}, \mathcal{B}(\mathbb{R}^{n_x}))$ with transition density $k_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \rightarrow \mathbb{R}_+^{n_x}$ with respect to the measure $\mu(dx)$. The random variables $(Y_t)_{t \geq 0}$ in $(\mathbb{R}^{n_y}, \mathcal{B}(\mathbb{R}^{n_y}))$ are conditionally independent given $(X_t)_{t \geq 0}$ with transition density (likelihood) $\rho_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \rightarrow \mathbb{R}_+^{n_y}$ with reference measure γ .

Intuitively, one can think that we can use naive Monte Carlo technique to approximate (4.2) :

$$\mathbb{E}[z(X_t) | \mathcal{F}_t] \approx \int (\mathbb{M}^{n_x} \mathbb{P})(dx_{0:t}) z(x_t) = \frac{1}{n_x} \sum_{i=1}^{n_x} f(X_t^{(i)}) \quad (4.5)$$

where the sampling operator $\mathbb{M}^{n_x} v = \frac{1}{n_x} \sum_{i=1}^{n_x} \delta_{X^{(i)}}$, $X \in \mathbb{R}^{n_x}$ and $\forall i = 1, \dots, n_x$, $X^{(i)}$ are i.i.d. draws from v .

The problem with naive Monte Carlo Sampling lies in the fact that we don't know how to sample from conditional distribution $\mathbb{P}(d(x_{0:t})) = \mathbb{P}(X_{0:t} \in dx_{0:t} | Y_{1:t})$. Moreover, computation of normalization constant \mathcal{Z} is a big challenge.

There is a lot of research made to tackle this problem, for example [6], where the problem is transformed from a partially observed to a fully observed, by introducing a so called filtering distribution, that links observed and latent variables and recursively updates it.

Definition Conditional probability (filtering distribution) $\Xi_t = \mathbb{P}(X_t \in \cdot | Y_1, \dots, Y_t)$ with prior $X_0 \sim p_0$ can be computed sequentially :

$$\Xi_t z = \frac{\int \Xi_{t-1}(dx_{t-1}) k_t(X_{t-1}, x) \mu(dx) \rho_t(x, Y_t) z(x)}{\int \Xi_{t-1}^\mu(dx_{t-1}) k_t(X_{t-1}, x) \mu(dx) \rho_t(x, Y_t)} \quad (4.6)$$

with $\Xi_0 = p_0$ and $\Xi_t z = \int \mathbb{P}(x_{0:t} \in dx_{0:t} | Y_1, \dots, Y_t) z(x_{0:t})$

Denote the corresponding values of the hidden process as (X_0, \dots, X_t) and the values of the measurement process as (Y_0, \dots, Y_t) . If there exists an absolutely continuous probability measure $\mathbb{P} \ll \mathbb{Q}$, than for $t = 0, \dots, N$ we have :

$$\mathbb{E}^{\mathbb{P}}[z(X_t)|\mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[z(X_t) \frac{d\mathbb{P}}{d\mathbb{Q}}(X_{0:t})|\mathcal{F}_t] \quad (4.7)$$

An importance measure \mathbb{Q} can be chosen arbitrarily as soon as the continuity of the measure is preserved. But usually in a sequential importance sampling literature it is common to see the approximation of \mathbb{Q} , given that there exists an absolutely continuous importance kernel \tilde{K}_t , such that for $K \ll \tilde{K}_t$ as :

$$\mathbb{Q}(B) = \sum_{i=1}^M \omega_t^{(i)} \tilde{K}_t(X_{t-1}^{(i)}, A_i), \quad B \in \mathcal{B}(\mathbb{R}^{n_x}) \quad (4.8)$$

where $A_i = \{X_t \in \mathbb{R}^{n_x} | \mathbb{1}_B(X_{t-1}^{(i)}, X_t) = 1\}$, $(\omega_{t-1}^{(i)})_{i=1}^M$ is the weight function, and for $i = 1, \dots, n_x$, $(X_0^{(i)}, \dots, X_t^{(i)})$ are independent trajectory realizations. Now assume that the prior and sampling kernels K_t and \tilde{K}_t have densities k_t and \tilde{k}_t with respect to the measure μ , $\forall t = 1, \dots, T$.

For $0 < \dots < t$, the Radon-Nikodym derivative in (4.7) is :

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(X_{0:t}) = \frac{1}{\mathcal{Z}} \rho_1(X_1, Y_1) \frac{k_1(X_0, X_1)}{\tilde{k}_1(X_0, X_1)} \dots \rho_t(X_t, Y_t) \frac{k_t(X_{t-1}, X_t)}{\tilde{k}_t(X_{t-1}, X_t)} \quad (4.9)$$

where the importance measure is given by :

$$\mathbb{Q}(dx_{0:t}) = p_0(dx_0) \tilde{k}_1(x_0, x_1) \mu(dx_1) \dots \tilde{k}_t(x_{t-1}, x_t) \mu(dx_t) \quad (4.10)$$

Observe, that we still can not compute a normalization constant Z , otherwise to compute the filtering distribution Ξ_t will not be a problem, so we will need to apply normalized operator \mathbb{M}^{n_x} to approximate filtering distribution :

$$\mathbb{E}^{\mathbb{P}}[z(X_{0:t})|\mathcal{F}_t] \approx \int \mathbb{M}^{n_x} \mathbb{P}(dx_{0:t}) z(x_t) = \sum_{i=1}^{n_x} \tilde{\omega}_t^{(i)} z(X_t^{(i)}) \delta_{X_t^{(i)}}(dx_t) \quad (4.11)$$

where the normalized importance weight function :

$$\tilde{\omega}_t^{(i)}(X_t^{(i)}) = \frac{\omega_t^{(i)}(X_t^{(i)})}{\sum_{j=1}^M \omega_t^{(j)}(X_t^{(j)})} \quad (4.12)$$

and an unnormalized weight is given by :

$$\omega_t^{(i)}(X_t^{(i)}) = \prod_{l=1}^t \rho_l(X_l^{(i)}, Y_l) \frac{k_l(X_{l-1}^{(i)}, X_l^{(i)})}{\tilde{k}_l(X_{l-1}^{(i)}, X_l^{(i)})} \quad (4.13)$$

Observe that importance weights $\{\tilde{\omega}_t^{(i)}\}_{i=1}^{n_x}$ are positive and $\sum_{i=1}^{n_x} \tilde{\omega}_t^{(i)} = 1$.

Since Particle filters showed weight degeneracy as number of time steps increased, Gordon et al. (1993) proposed a resampling step to the algorithm, which can be described by the following nonlinear equation :

$$\Xi_t = \Phi_t \Xi_{t-1}^{\#} \quad \text{with} \quad \Xi_0 = p_0 \quad (4.14)$$

where the nonlinear operator Φ_t is given by :

$$(\Phi_t \mathbf{v})_z = \frac{\int v(dx_{t-1}) k_t(X_{t-1}, x) \mu(dx) \rho_t(x, Y_t) z(x)}{\int v(dx_{t-1}) k_t(X_{t-1}, x) \mu(dx) \rho_t(x, Y_t)} \quad (4.15)$$

The action of the operator Φ can be schematically described as :

$$\Xi_{t-1} \xrightarrow{\text{Mutation}} \mathcal{M} \Xi_{t-1} \xrightarrow{\text{Reweighing}} \Omega_t \mathcal{M} \Xi_{t-1} \quad (4.16)$$

where the mutation operator \mathcal{M} is given by

$$(\mathcal{M} \mathbf{v})(z) = \int v(dx_{t-1}) p(x_{t-1}, x) \mu(dx) z(x) \quad (4.17)$$

and the reweighing operator Ω_t has the form

$$\Omega_t(\mathbf{v})_z = \frac{\int v(dx) \rho_t(x, Y_t) f(x)}{\int v(dx) g(x, Y_t)} \quad (4.18)$$

After the reweighing step we get the following approximation of the filtering distribution Ξ_{t-1} :

$$\hat{\Xi}_{t-1} = \sum_{i=1}^{n_X} \tilde{\omega}_{t-1}^{(i)} \delta_{X_{t-1}^{(i)}} \quad (4.19)$$

where $\{X_{t-1}^{(i)}\}_{i=1}^{n_X} \sim \mathcal{M} \hat{\Xi}_{t-2}$. We see from above equations that n_X particles are sampled from an empirical distribution $\hat{\Xi}_t$, i.e. it is itself defined through n_X particles.

Let us give the intuition behind the reweighing step. The idea behind it is in the fact, that at this step particles with low weights have lower probability to be sampled compared with particles with high importance weights. Consequently, in this step particles with low weights will be neglected, while particles with high weights will be sampled more frequently.

4.2.2 Particle Filtering Algorithm

The algorithm allows to approximate Ξ_{t-1}^μ by the empirical distribution $\hat{\Xi}_{t-1}^\mu$ compute by the following recurrence equations :

$$\hat{\Xi}_t = \hat{\Phi}_t \hat{\Xi}_{t-1} \quad \text{with} \quad \hat{\Xi}_0 = p_0 \quad (4.20)$$

where $\hat{\Phi}_t := \Omega_t \mathbb{M}^{n_X} \mathcal{M}$. It consists of three steps :

$$\hat{\Xi}_{t-1} \xrightarrow{\text{Mutation}} \mathcal{M} \hat{\Xi}_{t-1} \xrightarrow{\text{Sampling}} \mathbb{M}^{n_X} \mathcal{M} \hat{\Xi}_{t-1} \xrightarrow{\text{Reweighing}} \Omega_t \mathbb{M}^{n_X} \mathcal{M} \hat{\Xi}_{t-1} \quad (4.21)$$

At time $t = 0$, we generate M i.i.d. random variables from the prior distribution. For $t = 1, \dots, N - 1$ we propagate $X_t \in \mathbb{R}^{n_X}$ according to the dynamics of the hidden process, update the measurement, to get a couple of random vectors (X_{t+1}, Y_{t+1}) in the first step. Resample particles according to their probability weights $\hat{\omega}_{t+1}(X_{t+1})$ and set resampled particles \hat{X}_t . At the final time step t compute the estimate of (4.6) :

$$\hat{C}^{\text{PF}} = \frac{1}{n_X} \sum_{i=1}^{n_X} z(\hat{X}_t^{(i)}) \hat{\omega}_{t-1}^{(i)}(X_{t-1}^{(i)}) \delta_{X_{t-1}^{(i)}} \quad (4.22)$$

where $\{X_{t-1}^{(i)}\}_{i=1}^{n_X} \sim \mathcal{M}_{\widehat{\Xi}_{t-2}}$.

Algorithm 10 : PF Algorithm

```

1 Initialization :  $i = 1, \dots, n_X - \#(\text{simulations}), t = 1, \dots, T - \#(\text{time steps})$ 
2 Draw  $\{X_0^{(i)}\}_{i=1}^{n_X}$  from the prior  $p_0(x)$ . Set  $\{\omega_0^{(i)}\}_{i=1}^{n_X} = \frac{1}{n_X}$ ;
3 for  $t = 1, \dots, N$  do
4   for  $i = 1, \dots, n_X$  do
5     Propagate particles using state equation  $X_t^{(i)} = f(X_{t-1}^{(i)}, Y_{t-1}^{(i)}, \epsilon_t)$ ;
6     Measurement update :  $Y_t = h(X_t^{(i)}, Y_{t-1}^{(i)}, \eta_t)$ ;
7     Compute effective sample size  $M_{eff} : \frac{1}{\sum_{i=1}^M (\omega_t^{(i)})^2}$ ;
8     if  $M_{eff} < M$  or  $k < N$  then
9       Resample using weight  $\widehat{\omega}_t^{(i)}(X_t^{(i)}) = \frac{\omega_t^{(i)}(X_t^{(i)})}{\frac{1}{n_X} \sum_{j=1}^M \omega_t^{(j)}(X_t^{(j)})}$ 
10    end
11  end
12  Set resampled particles as  $\widehat{X}_t^{(i)}$ 
13 end

```

Despite the advantage of sampling from highly non-linear and non-gaussian filtering distributions, we need to mention its limitations. In fact, today we have to deal with high-dimensional data, as it was shown in [1], [14], [15], the collapse of weights occurs unless the sample size grows super-exponentially. Homotopy transport allows us to sample efficiently in high-dimensional framework, while avoiding the explosion of the sample size.

4.3 Homotopy Transport

The classical optimal transport problem is to find over all maps $\mathcal{T} : \mathbb{R}^{n_X} \rightarrow \mathbb{R}^{n_X}$, such that for $X \sim \mathbb{P}$, $\mathcal{T}(X) \sim \mathbb{Q}$ and $\mathcal{T} \in \mathcal{C}^1$; which optimizes the following criterion :

$$\begin{aligned} \inf_{\mathcal{T}} \mathbb{E}[\|\mathcal{T}(X) - X\|^2] \\ \text{s.t. } \mathbb{Q} = \mathcal{T}_\# \mathbb{P} \end{aligned} \quad (4.23)$$

In other words, we would like to find a continuous transformation that minimizes the distance between measure \mathbb{P} and measure \mathbb{Q} among all these that pushforward a prior measure \mathbb{P} towards a measure \mathbb{Q} . In the context of filtering problem we would like to find a transformation \mathcal{T} , that transport particles from a sampling measure \mathbb{P} to \mathbb{Q} :

$$\mathbb{E}^{\mathbb{Q}}[z(X_t) \frac{d\mathbb{P}}{d\mathbb{Q}}(X_{0:t}) | \mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[z(\mathcal{T}(X_t)) | \mathcal{F}_t] \quad (4.24)$$

One can solve this problem using variational methods [8].

For the sake of exposition we represent posterior distribution, presented in the form of a normalized importance weight in the following way :

$$\psi(X_t | \mathbf{Y}_t) = \frac{1}{\mathcal{Z}_t} p(X_t | \mathbf{Y}_{t-1}) \rho(Y_t | X_t) \quad (4.25)$$

where $\mathbf{Y}_t = (Y_0, \dots, Y_t)$, the prior is $p(X_t | \mathbf{Y}_{t-1})$, the likelihood is $\rho(Y_t | X_t)$ and \mathcal{Z}_t is a normalization factor : $\mathcal{Z}_t = \int p(X_t | \mathbf{Y}_{t-1}) \rho(Y_t | X_t) dX_t$. Actually, the equation (4.25) is equivalent

to the normalized importance weight in the eq. (4.12). Now, if we consider a continuous map $\mathcal{T} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x}$, then :

$$\psi(\mathcal{T}(X_t)|Y_t) = \frac{1}{\mathcal{Z}_t} p(\mathcal{T}(X_t)|Y_{t-1}) \rho(Y_t|\mathcal{T}(X_t)) \quad (4.26)$$

Homotopy gradually modifies the prior density into the posterior density, as a scaling parameter $\lambda \in [0, 1]$ increases from 0 to 1. In other words, by iterating we will transport homotopy $\psi(X_{t,\lambda}|Y_t)$ to a true posterior $\psi(X_t|Y_t)$, while minimizing the cost of transport. There are several conditions that homotopy has to satisfy. First, at λ_0 we should have our prior, i.e. $\psi(x_{t,\lambda_0}|Y_t) = p(X_t)$ and at some point $\lambda_{0 \rightarrow 1}$, we will get approximation of our posterior density. Define a new set of density functions : $\psi(X_{t,\lambda}|Y_t) := \psi(X_t|Y_t)$, $p(X_{t,\lambda}|Y_{t-1}) := p(X_t|Y_{t-1})$, $\rho(Y_t|X_{t,\lambda})^\lambda := \rho(Y_t|X_{t,\lambda})$ and $\mathcal{Z}_\lambda := \int p(X_{t,\lambda}|Y_{t-1}) \rho(Y_t|X_{t,\lambda})^\lambda dx_\lambda$, so that homotopy is defined as :

$$\psi(X_{t,\lambda}|Y_t) = \frac{1}{\mathcal{Z}_\lambda} \underbrace{p(X_{t,\lambda}|Y_{t-1})}_{\text{prior}} \underbrace{\rho(Y_t|X_{t,\lambda})^\lambda}_{\text{likelihood}} \quad (4.27)$$

In order to simplify the calculation we take the logarithm of homotopy :

$$\Psi(X_{t,\lambda}|Y_t) = G(X_{t,\lambda}) + \lambda L(X_{t,\lambda}) - \log \mathcal{Z}_\lambda \quad (4.28)$$

where $\Psi(X_{t,\lambda}) = \log \psi(X_{t,\lambda}|Y_t)$, $G(X_{t,\lambda}) = \log p(X_{t,\lambda}|Y_{t-1})$, $L(X_{t,\lambda}) = \log \rho(Y_t|X_{t,\lambda})$. The dynamics of homotopy transport in the artificial time λ is known as *log-homotopy* [2]. In some sense, the dynamics of transport will be given by the flow movement in the artificial time λ , so we will look for a flow $\frac{dx}{d\lambda}$ that rules the movement of particles following log-homotopy.

If we assume that in pseudo-time λ , the flow $\frac{dx}{d\lambda}$ follows the following SDE :

$$dX_{t,\lambda} = g(X_{t,\lambda}) d\lambda + \eta(X_{t,\lambda}) dW_\lambda \quad (4.29)$$

where W_λ is a vector field that pushes forward particles from prior to posterior distribution.

We impose the following assumptions :

- I/ The densities $p(X_{t,\lambda}|Y_{t-1})$ and $\rho(Y_t|X_{t,\lambda})$ are twice differentiable with respect to $X_{t,\lambda}$;
- II/ The function that governs the particle transport $g(X_{t,\lambda})$ is differentiable with respect to $X_{t,\lambda}$;
- III/ The Hessian matrix of the density Ψ is non-singular ;

Now given the conditional probability density function (4.28), we can compute the function $g(X_{t,\lambda}) = \frac{dX_{t,\lambda}}{d\lambda}$ using the forward Kolmogorov equation :

$$\frac{\partial \psi(X_{t,\lambda})}{\partial \lambda} = -tr \left[\frac{\partial}{\partial X_{t,\lambda}} (g(X_{t,\lambda}) \psi(X_{t,\lambda})) \right] + \frac{1}{2} tr \left[\frac{\partial}{\partial X_{t,\lambda}} Q(X_{t,\lambda}) \frac{\partial \psi(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] \quad (4.30)$$

where $Q(X_{t,\lambda}) = \eta(X_{t,\lambda}) \eta^\top(X_{t,\lambda})$ is the diffusion tensor of the process, and $tr(\cdot)$ is a trace operator. The forward Kolmogorov equation is used to relate the flow of particles $\frac{dX_{t,\lambda}}{d\lambda}$ with the evolution of log-homotopy as $\lambda_{0 \rightarrow 1}$, under the diffusion process.

$$\begin{aligned} \frac{\partial \psi(X_{t,\lambda})}{\partial \lambda} &= -tr \left[\psi(X_{t,\lambda}) \frac{\partial g(X_{t,\lambda})}{\partial X_{t,\lambda}} + g(X_{t,\lambda})^\top \frac{\partial \psi(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] + \frac{1}{2} div \left[\frac{\partial}{\partial X_{t,\lambda}} Q(X_{t,\lambda}) \frac{\partial \psi(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] = \\ &= -\psi(X_{t,\lambda}) tr \left[\frac{\partial g(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] - g(X_{t,\lambda})^\top \frac{\partial \psi(X_{t,\lambda})}{\partial X_{t,\lambda}} + \frac{1}{2} div \left[\frac{\partial}{\partial X_{t,\lambda}} Q(X_{t,\lambda}) \frac{\partial \psi(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] \end{aligned} \quad (4.31)$$

where $div(\cdot)$ is a divergence operator. On the other hand if we take the derivative of equation (4.28) with respect to λ , we have :

$$\frac{\partial \Psi(X_{t,\lambda})}{\partial \lambda} = L(X_{t,\lambda}) - \frac{\partial}{\partial \lambda} \log \mathcal{Z}_\lambda \quad (4.32)$$

Since $\Psi(X_{t,\lambda})$ is a composition of two functions, we will need to use the chain rule :

$$\frac{\partial \Psi(X_{t,\lambda})}{\partial \lambda} = \frac{1}{\psi(X_{t,\lambda})} \frac{\partial \psi(X_{t,\lambda})}{\partial \lambda} \quad (4.33)$$

By substituting eq. (4.33) into (4.32) and rearranging the terms :

$$\frac{\partial \psi(X_{t,\lambda})}{\partial \lambda} = \psi(X_{t,\lambda}) \left[L(X_{t,\lambda}) - \frac{\partial}{\partial \lambda} \log \mathcal{Z}_\lambda \right] \quad (4.34)$$

Observe that (4.31) and (4.34) are identical, so by equating and dividing on $\psi(X_{t,\lambda})$ we get :

$$\begin{aligned} L(X_{t,\lambda}) - \frac{\partial}{\partial \lambda} \log \mathcal{Z}_\lambda = & -g(X_{t,\lambda})^T \frac{1}{\psi(X_{t,\lambda})} \frac{\partial \psi(X_{t,\lambda})}{\partial X_{t,\lambda}} - \\ & - tr \left[\frac{\partial g(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] + \frac{1}{2\psi(X_{t,\lambda})} div \left[\frac{\partial}{\partial X_{t,\lambda}} Q(X_{t,\lambda}) \frac{\partial \psi(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] \end{aligned} \quad (4.35)$$

In [4], authors propose to take the derivative of (4.35) with respect to $X_{t,\lambda}$ in order to find explicitly the equation of flow on the one hand, and to get rid of the normalization constant \mathcal{Z}_λ that lead to instabilities on the other hand.

$$\begin{aligned} \frac{\partial L(X_{t,\lambda})}{\partial X_{t,\lambda}} = & -g(X_{t,\lambda})^T \frac{\partial^2 \Psi(X_{t,\lambda})}{\partial X_{t,\lambda}^2} - \frac{\partial \Psi(X_{t,\lambda})}{\partial X_{t,\lambda}} \frac{\partial g(X_{t,\lambda})}{\partial X_{t,\lambda}} - \frac{\partial}{\partial X_{t,\lambda}} tr \left[\frac{\partial g(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] + \\ & + \frac{\partial}{\partial X_{t,\lambda}} \left(\frac{1}{2\psi(X_{t,\lambda})} div \left[Q(X_{t,\lambda}) \frac{\partial \psi(X_{t,\lambda})}{\partial X_{t,\lambda}} \right] \right) \end{aligned} \quad (4.36)$$

Observe that we get a highly nonlinear PDE. We use the solution found in [2] and [3], which states that if we can find a vector field $g(X_{t,\lambda})$ and diffusion tensor $Q(X_{t,\lambda})$, such that sum of the last three terms in (4.36) are equal to zero. The PDE, then simplifies to :

$$\frac{\partial L(X_{t,\lambda})}{\partial X_{t,\lambda}} = -g(X_{t,\lambda})^T \frac{\partial^2 \Psi(X_{t,\lambda})}{\partial X_{t,\lambda}^2} \quad (4.37)$$

Using the assumption III, i.e. the Hessian matrix $\frac{\partial^2 \Psi(X_{t,\lambda})}{\partial X_{t,\lambda}^2}$ is non-singular, we get explicitly the flow $g(X_{t,\lambda})$:

$$g(X_{t,\lambda}) = - \left[\frac{\partial^2 \Psi(X_{t,\lambda})}{\partial X_{t,\lambda}^2} \right]^{-1} \left[\frac{\partial L(X_{t,\lambda})}{\partial X_{t,\lambda}} \right]^T \quad (4.38)$$

4.3.1 Homotopy Transport Algorithm

Sampling from the prior. First we generate M i.i.d random variables $X_t^{(i)}$ from the prior density $p_0(x)$, initialize pseudo-time λ and set the state variables that will be transported as : $X_{t,\lambda}^{(i)} = X_{t|t-1}^{(i)}$.

Transportation Stage. For $t = 2, \dots, N - 1$, compute the derivative with respect to $X_{t,\lambda}$ of the measurement function. If h is non-linear, a second order Taylor expansion at $X_{t,\lambda}$ allows speeding up the calculation by linearizing the first derivative. After that, update the pseudo time by setting : $\lambda = \lambda + \Delta\lambda$.

Compute the flow $g(X_{t,\lambda}^{(i)})$. Note, that the first Hessian can be derived by twice differentiating a log-homotopy equation (4.28) :

$$\frac{\partial^2 \Psi(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}^2} = \frac{\partial^2 G(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}^2} + \lambda \frac{\partial^2 L(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}^2} \quad (4.39)$$

The first term in (4.39) $\frac{\partial^2 G(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}^2}$ is estimated by using a sample covariance matrix of t particles generated from the prior distribution :

$$\frac{\partial^2 G(X_{t,\lambda})}{\partial X_{t,\lambda}^2} \approx -\widehat{S}_{M_x}^{-1} \quad (4.40)$$

Compute the transportation of particles from the measure \mathbb{P} to the measure \mathbb{Q} :

$$X_{t,\lambda}^{(i)} = X_{t,\lambda}^{(i)} + \Delta\lambda g(X_{t,\lambda}^{(i)}) \quad (4.41)$$

And finally update the state parameter :

$$\check{X}_t = \frac{1}{n_X} \sum_{i=1}^{n_X} X_{t,\lambda}^{(i)} \quad (4.42)$$

Maturity.

At the final time interval $]N - 1, N]$ compute the estimator of (4.24) :

$$\widehat{C}^{\text{HT}} = \frac{1}{n_X} \sum_{i=1}^{n_X} z(\check{X}_t^{(i)}) \quad (4.43)$$

Algorithm 11 : Homotopy Transport Algorithm

```

1 Initialization :  $i = 1, \dots, n_X$  - #(simulations),  $t = 1, \dots, N$  - #(time steps)
2 Draw  $\{X_0^{(i)}\}_{i=1}^{n_X}$  from the prior  $p_0(x)$ .
3 Set  $\{\omega_0^{(i)}\}_{i=1}^{n_X} = \frac{1}{n_X}$ 
4 for  $t = 1, \dots, N$  do
5   for  $i = 1, \dots, n_X$  do
6     Propagate particles using state equation  $X_t^{(i)} = f(X_{t-1}^{(i)}, Y_{t-1}^{(i)}, \epsilon_t)$ ;
7     Measurement update :  $Y_t = h(X_t^{(i)}, Y_{t-1}^{(i)}, \eta_t)$ ;
8     Initialize pseudo-time  $\lambda = 0$ ;
9     Set  $X_{t,\lambda}^{(i)} = X_{t|n-1}^{(i)}$ ;
10    while  $\lambda < 1$  do
11      Compute SCM  $\widehat{S}_M$ ;
12      Calculate an estimate :  $X_{t,\lambda} = \frac{1}{n_X} \sum_i X_{t,\lambda}^{(i)}$ 
13      Compute the matrix  $\widehat{H} = \frac{\partial h(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}}$ ;
14      Update the time :  $\lambda = \lambda + \Delta\lambda$ ;
15      Calculate the flow  $\frac{dX_{t,\lambda}^{(i)}}{d\lambda} = - \left[ \frac{\partial^2 \Psi(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}^2} \right]^{-1} \left[ \frac{\partial L(X_{t,\lambda}^{(i)})}{\partial X_{t,\lambda}} \right]^T$ ;
16      Transport particles according to its flow :  $X_{t,\lambda}^{(i)} = X_{t,\lambda}^{(i)} + \Delta\lambda \frac{dX_{t,\lambda}^{(i)}}{d\lambda}$ ;
17    end
18    Update state estimate :
19     $\check{X}_t = \frac{1}{n_X} \sum_{i=1}^{n_X} X_{t,\lambda}^{(i)}$ 
20  end
21 end

```

4.4 Homotopy Transport with Particle Reweighting

Taking into account the difficulties one faces in non-Gaussian and high-dimensional problems, the idea of a particle transport without any use of sampling techniques is very elucidating. The next question that arises is whether we can direct the transportation by choosing those particles that have higher probability of reaching rarely visited areas of the state space? We propose a mixture of homotopy particle transport with a particle reweighting at each time step. The numerical test that we performed on the toy example of a Stein-Stein stochastic volatility model shows that we significantly reduce the variance and bias of our estimator.

The algorithm consists of two steps : first we transport particles according to its flow, and second, we choose those particles that have higher probability of faster exploring the state space.

$$\mathbb{E}^{\widetilde{\mathbb{Q}}} \left[z(X_t) \frac{d\mathbb{P}}{d\widetilde{\mathbb{Q}}} (X_{0:t}) | \mathcal{F}_t \right] = \mathbb{E}^{\mathbb{P}} \left[z(\mathcal{T}(X_t)) | \mathcal{F}_t \right] = \mathbb{E}^{\mathbb{Q}} \left[z(\mathcal{T}(X_t)) \frac{d\mathbb{P}}{d\mathbb{Q}} (X_{0:t}) | \mathcal{F}_t \right] \quad (4.44)$$

where \mathcal{T} is a flow of particles under the pseudotime λ described in the section 4.3.1.

By setting $\mathbf{X}_t = (X_0, \dots, X_t)$, we can express our Radon-Nikodym derivative in a product form :

$$\frac{d\mathbb{P}}{d\mathbb{Q}} (X_{0:t}) = \frac{d\mathbb{P}}{d\widetilde{\mathbb{Q}}} \times \frac{d\widetilde{\mathbb{Q}}}{d\mathbb{Q}} (X_{0:t}) \quad (4.45)$$

where the first Radon-Nikodym derivative denotes the transport of particles from a measure \mathbb{P} to a measure $\tilde{\mathbb{Q}}$, then we choose the particles that have high probability of reaching rare corners of the state space, using $\frac{d\tilde{\mathbb{Q}}}{d\mathbb{Q}}$ that allows us to reassess the weights of the particles.

As in the section 2, an importance measure \mathbb{Q} that will play a resampling to choose the trajectories with higher weight, given that there exists an importance kernel \tilde{K}_t , such that $K_t \ll \tilde{K}_t$, can be defined as :

$$\mathbb{Q}(B) = \sum_{i=1}^{n_X} \omega_t^{(i)} \tilde{K}_t(X_t^{(i)}, A_i), \quad B \in \mathcal{B}(\mathbb{R}^{n_X}) \quad (4.46)$$

where the set $A_i = \{\mathcal{F}(X_{t+1}) \in \mathbb{R}^{n_X} | \mathbb{1}_B(X_t^{(i)}, \mathcal{F}(X_{t+1})) = 1\}$. Assuming, that the prior and sampling kernels K_t and \tilde{K}_t have densities k_t and \tilde{k}_t respectively, then the Radon-Nikodym derivative is

$$\frac{d\tilde{\mathbb{Q}}}{d\mathbb{Q}}(X_{0:t}) = \prod_{l=0}^t \rho_l(\mathcal{F}(X_l), Y_l) \frac{\omega_{l-1}(X_{l-1}) k_l(X_{l-1}, \mathcal{F}(X_l))}{\omega_{l-1}(X_{l-1}) \tilde{k}_l(X_{l-1}, \mathcal{F}(X_l))} \quad (4.47)$$

such that $\omega_t(X_t) = \omega_t^{(i)}(X_t^{(i)})$ if $X_t = X_t^{(i)}$, and $\omega_t(X_t) = 1$ otherwise.

The an unnormalized weight is given by :

$$\omega_t^{(i)}(\mathcal{F}(X_t^{(i)})) = \prod_{l=1}^t \rho_l(\mathcal{F}(X_l^{(i)}), Y_l) \frac{k_l(X_{l-1}, \mathcal{F}(X_l^{(i)}))}{\tilde{k}_l(X_{l-1}, \mathcal{F}(X_l^{(i)}))} \quad (4.48)$$

So, now we have homotopy transport with particle reweighing estimator :

$$\hat{C}^{\text{TRW}} = \frac{1}{n_X} \sum_{i=1}^{n_X} z(\mathcal{F}(X_t^{(i)})) \hat{\omega}_{t-1}^{(i)}(\mathcal{F}(X_{t-1}^{(i)})) \quad (4.49)$$

4.4.1 PF-Enhanced Homotopy Transport Algorithm

The algorithm can be described by the following scheme, $\forall i = 1, \dots, n_X$:

$$X_t^{(i)} \xrightarrow{\text{Sampling}} X_{t+1}^{(i)} \xrightarrow{\text{Transportation}} \mathcal{F}(X_{t+1}^{(i)}) = \check{X}_{t+1}^{(i)} \xrightarrow{\text{Reweighing}} \Phi(\check{X}_{t+1}^{(i)}) = \hat{X}_{t+1}^{(i)} \quad (4.50)$$

where Φ is an operator that denotes the resampling mechanism of particles. If we assume that there is a continuous kernel \tilde{K}_t , such that $K_t \ll \tilde{K}_t$ with densities k_t and \tilde{k}_t respectively, then we can define a weight function $\omega_t^{(i)}$:

$$\omega_t^{(i)}(\check{X}_t^{(i)}) = \prod_{l=1}^t \rho_l(\check{X}_l^{(i)}, Y_l) \frac{k_l(\hat{X}_{l-1}^{(i)}, \check{X}_l^{(i)})}{\tilde{k}_l(\hat{X}_{l-1}^{(i)}, \check{X}_l^{(i)})} \quad (4.51)$$

Detailed Algorithm

Sampling from the prior. As in the section 4.3.1, we start with M particles sampled from the prior distribution p_0 , initialize pseudo-time λ and set the state variables that will be transported as : $X_{t,\lambda}^{(i)} = X_{t|n-1}^{(i)}$.

Transportation Stage. Follow steps 6-8 of the Algorithm 2 in the section 4.3.1.

Path Reweighing Stage. Compute the normalized importance weight :

$$\hat{\omega}_t^{(i)}(\check{X}_t^{(i)}) = \frac{\omega_t^{(i)}(\check{X}_t^{(i)})}{\sum_{j=1}^{n_X} \omega_t^{(j)}(\check{X}_t^{(j)})} \quad (4.52)$$

Maturity At the time interval $]N - 1, N]$ compute the final Homotopy transport re-weighted estimator :

$$\widehat{C}^{\text{TRW}} = \frac{1}{n_X} \sum_{i=1}^{n_X} z(\widehat{X}_t^{(i)}) \widehat{\omega}_{t-1}^{(i)} (\check{X}_{t-1}^{(i)}) \quad (4.53)$$

Algorithm 12 : Homotopy Transport with Particle Reweighting Algorithm

- 1 Initialization : $i = 1, \dots, n_X$ - #(simulations), $t = 1, \dots, T$ - #(time steps)
 - 2 Draw $\{X_0^{(i)}\}_{i=1}^{n_X}$ from the prior $p_0(x)$.
 - 3 Set $\{\omega_0^{(i)}\}_{i=1}^{n_X} = \frac{1}{n_X}$
 - 4 **for** $t = 1, \dots, N$ **do**
 - 5 **for** $i = 1, \dots, n_X$ **do**
 - 6 Follow steps 6-8 of the Algorithm 2 in the section 4.3.1.
 - 7 Follow steos 7-12 of the Algorithm 1 in the section 4.2.2
 - 8 **end**
 - 9 **end**
-

4.5 Numerical Applications and Results

As a toy example, we decided to test the algorithms on a Stein-Stein stochastic volatility model. We set log-returns as $Y_t = \log(S_t)$, then the model takes the following form :

$$\begin{cases} dY_t = (\mu - \frac{X_t^2}{2})dt + X_t dB_t \\ dX_t = \kappa(\theta - X_t)dt + \sigma dW_t \end{cases} \quad (4.54)$$

where X_t is a volatility process, Y_t the dynamics of log-returns, μ is a drift, θ is a long-term variance, κ - the rate of reversion, σ is the volatility of volatility, and B_t and W_t are two independent Brownian motions, in the sense that $\langle dB_t, dW_t \rangle = 0$.

Using the above presented stochastic volatility model, we would like to compute estimates for a European option. For a given interest rate r , maturity T , strike price K , and for a function $z(\cdot, x) = \max(x - K, 0)$, the call price of the option is given by :

$$C(X_t, Y_t) = B_{t,T} \mathbb{E}^{\mathbb{P}} [z(X_T, Y_T) | \mathcal{F}_t] \quad (4.55)$$

where $\mathcal{F}_t = \sigma\{(Y_0, \dots, Y_t)\}$.

We chose Euler-Muruyama discretization scheme, which gives :

$$\begin{cases} Y_t - Y_{t-1} = (\mu - \frac{X_{t-1}^2}{2})\Delta t + X_{t-1} \sqrt{\Delta t} \epsilon_t \\ X_t - X_{t-1} = \kappa(\theta - X_{t-1})\Delta t + \sigma \sqrt{\Delta t} \eta_t \end{cases} \quad (4.56)$$

where Δt is a discretization size, ϵ_t and η_t are independent Gaussian variates, $\mathcal{N}(0, 1)$.

We compare each approach by estimating the standard deviations, the root mean squared error (RMSE), the bias, the relative mean squared error(RRMSE), the time required to compute each estimate and the figure of merit (FOM). We run 20 Monte Carlo experiments. For $l = 1, \dots, M_s$ the RMSE estimator is given by :

$$\text{RMSE} = \sqrt{\frac{1}{M_s} \sum_{l=1}^{M_s} \|C - \widehat{C}_l\|^2} \quad (4.57)$$

where C is the price computed analytically, \widehat{C}_l are Monte Carlo estimates and $M_s = 20$ is the number of Monte Carlo experiments. As a reference price, we used the article by EM Stein [9].

$$Bias = \sqrt{RMSE^2 - St.dev^2} \quad (4.58)$$

where $St.dev$ are standard deviations of MC estimates. The RRMSE is computed using the following formula :

$$RRMSE = \frac{RMSE}{\hat{C}} \quad (4.59)$$

To measure the efficiency of each method presented in the article, we will use the figure of merit(FOM)[10] :

$$FOM = \frac{1}{R^2 \times CPU_t} \quad (4.60)$$

where CPU_t is the CPU time need to compute the estimator and R is a relative error, which is the measure of statistical precision :

$$R = \frac{St.dev}{\bar{C}} \propto \frac{1}{\sqrt{M}} \quad (4.61)$$

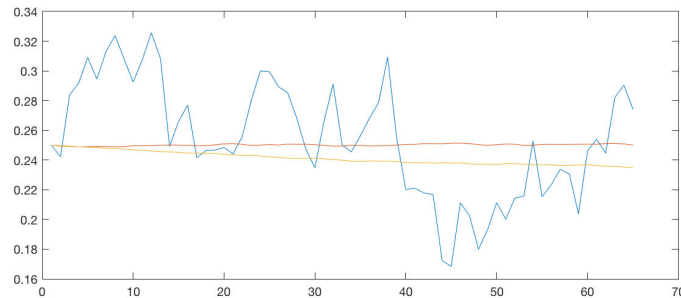
where $\bar{C} = \sum_{l=1}^{M_s} \hat{C}_l$

We used 20 000 and 40 000 simulations over 64 time intervals for our MC experiments. Table 1. shows that homotopy and reweighted(RW)-homotopy algorithms shows less statistical errors then traditional PF. If we compare homotopy and RW-homotopy, we can see that FOM says that the first is more efficient the latest, due to the fact that we need more time to reweight the paths. Meanwhile RW-homotopy shows less erros and st. deviations.

[H]

TABLEAU 4.1 – Stein-Stein Stochastic volatility option price estimates statistics. $S_0 = 100$, $K = 90$, $r = 0.0953$, $\sigma = 0.2$, $\kappa = 4$, $\theta = 0.25$, $V_0 = 0.25$, $T = 1/2$, and dividends $d = 0$ True price : 16.05, $t = 20000$, $M = 64$

Stat	MC	PF	Homotopy	RW-Homotopy
St. dev.	0.127495344	0.106264197	0.102775848	0.08360908
RMSE	0.148073563	0.115032508	0.105302932	0.084510606
Bias	0.075304165	0.044049955	0.022931037	0.012311146
RRMSE	0.00137298	0.000827032	0.000827032	0.000444367
CPU time	0.1327525	0.31177	0.179	0.38819
FOM	118181.69	72715.84	135692.61	95193.97



[H]

FIGURE 4.1 – Volatility dynamics, PF (Blue), Homotopy (Red), RW-homotopy(Yellow)

Despite the fact that Monte Carlo estimate showed higher FOM, than PF, due to the fact that it takes less time to compute Monte Carlo estimator. Whereas PF has lower RMSE and the bias.

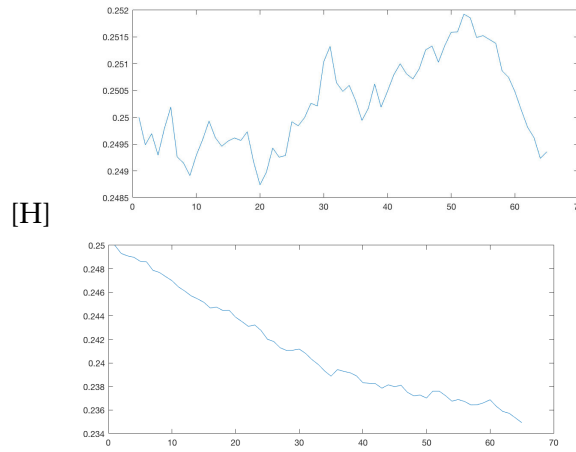


FIGURE 4.2 – Zoomed volatility dynamics. Homotopy (left), RW-homotopy (right)

[H]

TABLEAU 4.2 – Stein-Stein Stochastic volatility option price estimates statistics. $S_0 = 100$, $K = 90$, $r = 0.0953$, $\sigma = 0.2$, $\kappa = 4$, $\theta = 0.25$, $V_0 = 0.25$, $T = 1/2$, and dividends $d = 0$ True price : 16.05, $t = 40000$, $M = 64$

Stat	MC	PF	Homotopy	RW-Homotopy
St. dev.	0.070351719	0.060799052	0.048943672	0.045246118
RMSE	0.130446299	0.079273246	0.04921257	0.045762201
Bias	0.109849318	0.050869665	0.005137504	0.006853309
RRMSE	0.001067583	0.000392831	0.00015101	0.000130578
CPU time	0.278895	0.54737	0.26618	0.581495
FOM	184049.069	126479.8136	403391.758	216062.7397

4.6 Conclusions and Further Research

The estimation of latent variables has a lot of applications in engineering and finance. We provide homotopy based algorithm and its extension with reweighted trajectories that permits to solve the optimal transportation problem.

Numerical results that we applied in European option pricing with stochastic volatility demonstrated the efficiency of the proposed algorithms with respect to error, bias and other statistics. Both algorithms outperformed particle filtering. The path-reweighing allowed to reduce standard deviations, and in some cases the bias and the RMSE compared to the homotopy transport algorithm.

From our experiments we can observe the following :

- Homotopy transport is fast algorithm, which is spectacularly demonstrated in the figure of merit statistics.
- Efficiency of homotopy transport algorithm increases as the number of particles increases.
- Implementation of homotopy transport requires less effort than a vanilla Monte Carlo algorithm.
- Homotopy transport proved to be unbiased estimator.
- Homotopy with path reweighing proved to reduce the bias when the number of particles is small compared to homotopy transport without reweighing.

While reweighted homotopy transport approach showed the reduced RMSE and Bias in low-dimensions, the mixture of homotopy transport and bootstrap resampling, the importance weight can converge to unity in high-dimensional problems([15]) and potentially worsen. In the next article we plan to use the combination of MCMC and homotopy transport methods on real data. It will be also interesting to investigate the homotopy transport on non-gaussian examples.

Bibliographie

- [1] P. Bickel, B. Li, and T. Bengtsson, Sharp failure rates for the bootstrap particle filter in high dimensions, Institute of Mathematical Statistics Collections, vol. 3, pp. 318329, 2008.
- [2] Daum, F, & Huang, J. (2013). Particle flow with non-zero diffusion for nonlinear filters. In Proceedings of spie conference (Vol. 8745).
- [3] Daum, F, & Huang, J. (2011). Particle degeneracy : root cause and solution. In Proceedings of spie conference (Vol. 8050).
- [4] Daum, F, & Huang, J. (2015). Renormalization group flow in k-space for nonlinear filters, Bayesian decisions and transport.
- [5] Del Moral, P : Mean field simulation for Monte Carlo integration. CRC Press (2013)
- [6] Del Moral, P : Feynman-Kac Formulae : Genealogical and Interacting Particle Systems with Applications. Probability and Applications. Springer, New York (2004).
- [7] Del Moral, P : Nonlinear Filtering : Interacting Particle Solution(1996). Markov Processes and Related Fields 2 (4), 555-580
- [8] El Moselhy, Tarek A. and Marzouk, Youssef M.(2012). Bayesian inference with optimal maps. Journal of Computational Physics. (Vol. 231)
- [9] Stein, Elias M, and Jeremy C Stein. 1991. Stock Price Distributions with Stochastic Volatility : An Analytic Approach. Review of Financial Studies 4 : 727-752.
- [10] Rubino, G., Tuffin, B. : Rare event simulation using Monte Carlo methods. Wiley (2009)
- [11] Beiglböck, M., Henry-Labordère, P. & Penkner, F. Finance Stoch (2013) 17 : 477. doi :10.1007 /s00780-013-0205-8
- [12] Villani, C. : Topics in optimal transportation, Graduate studies in Mathematics AMS, Vol 58.
- [13] Rachev, S. T. and Ruschendorf, L. : Mass Transportation Problems. In Vol. 1 : Theory. Vol. 2 : Applications. Springer, Berlin, 1998.
- [14] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, Obstacles to high-dimensional particle filtering, Monthly Weather Review, vol. 136, no. 12, pp. 46294640, 2008.
- [15] F. Septier and G. W. Peters, An Overview of Recent Advances in Monte-Carlo Methods for Bayesian Filtering in High-Dimensional Spaces, in Theoretical Aspects of Spatial-Temporal Modeling, G. W. Peters and T. Matsui, Eds. SpringerBriefs - JSS Research Series in Statistics, 2015.

Appendix

Flow related computations

In the classical particle filtering approach the desired estimate is approximate by M particles :

$$p(X_t|Y_{1:n-1}) = \frac{1}{n_X} \sum_{i=1}^{n_X} p(X_t|X_{t-1} = X_{t-1}^{(i)}) \quad (4.62)$$

So that posterior at time t :

$$p(X_t|Y_{1:n}) = \frac{1}{\mathcal{Z}_t} \rho(Y_t|X_t) p(X_t|Y_{1:n-1}) \quad (4.63)$$

The transition density is given by :

$$p(X_t|X_{t-1}) = \mathcal{N}(X_t; \mu_X, \sigma_X) \quad (4.64)$$

where $\mu_X = X_{t-1} + \kappa(\mu_X - X_{t-1})\Delta t$ and $\sigma_X = \sigma_X^2 \Delta t$.

The likelihood, $p(Y_t|Y_{t-1}, X_{t-1})$:

$$m_{t,N}(X_{t-1}) = Y_{t-1} + \left(\mu - \frac{X_{t-1}^2}{2}\right)\Delta t \quad (4.65)$$

$$\sigma_{t,N}^p(X_{t-1}) = X_{t-1}^2 \Delta t \quad (4.66)$$

So,

$$\rho(Y_t|Y_{t-1}, X_{t-1}) = \prod_{t=1}^N \mathcal{N}(Y_t; m_{t,N}(X_{t-1}), \sigma_{t,N}^p(X_{t-1})) \quad (4.67)$$

The unnormalized posterior is given by :

$$\mathcal{P}_t = \rho(Y_t|Y_{t-1}, X_{t-1}) \quad (4.68)$$

$$X_t = \psi(X_t; \mathcal{P}_t) \quad (4.69)$$

Next,

$$\psi(X) = -\log(\mathcal{P}(X)) \quad (4.70)$$

by removing some constants that have no impact on posterior distribution, we have

$$\psi(X) = \sum_{t=1}^N \frac{(Y_t - m_{t,N}(X_{t-1}))^2}{2\sigma_{t,N}^p(X_{t-1})} + \frac{1}{2} \log(\sigma_{t,N}^p(X_{t-1})) \quad (4.71)$$

$$\frac{\partial \psi}{\partial x}(X) = \frac{1}{2} \sum_{t=1}^N \left(\frac{\nabla_X \sigma_{t,N}^p(X_{t-1})}{X_{t-1}} - \frac{(Y_t - m_{t,N})(2\sigma_{t,N}^p(X_{t-1})\nabla_X m_{t,N}(X_{t-1}) + (Y_t - m_{t,N})\nabla_X \sigma_{t,N}^p(X_{t-1}))}{\sigma_{t,N}^p(X_{t-1})^2} \right) \quad (4.72)$$

$$\nabla_X m_{t,N}(X_{t-1}) = -X_{t-1} \Delta t \quad (4.73)$$

$$\nabla_X \sigma_{t,N}^p(X_{t-1}) = 2X_{t-1} \Delta t \quad (4.74)$$

$$\frac{\partial \Psi}{\partial x}(\mathbf{X}) = \frac{1}{2} \sum_{t=1}^N (2\Delta t - \frac{(Y_t - m_{t,N})(2\sigma_{t,N}^p(X_{t-1})\nabla_X m_{t,N}(X_{t-1}) + (Y_t - m_{t,N})\nabla_X \sigma_{t,N}^p(X_{t-1}))}{\sigma_{t,N}^p(X_{t-1})^2}) \quad (4.75)$$

$$u = (Y_t - m_{t,N})(-2X_{t-1}^3\Delta t^2 + 2(Y_t - m_{t,N})X_{t-1}\Delta t) \quad (4.76)$$

$$u' = 2X_{t-1}^2\Delta t^2(X_{t-1}^2\Delta t - (Y_t - m_{t,N})) + (Y_t - m_{t,N})(-6X_{t-1}^2\Delta t^2 + 2\Delta t((Y_t - m_{t,N}))) \quad (4.77)$$

$$v = X_{t-1}^4\Delta t^2, \quad v' = 4X_{t-1}^3\Delta t^2 \quad (4.78)$$

$$\frac{\partial^2 \Psi}{\partial x^2} = \frac{u'v - v'u}{v^2} \quad (4.79)$$