THÈSE

présentée à

L'UNIVERSITÉ BORDEAUX 1

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE

par **Peng Hu**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : **Mathématiques Appliquées**

****

Méthodes particulaires et applications en finance

****

Soutenue le 21 juin 2012 à l'Institut de Mathématiques de Bordeaux

Devant la commission d'examen composée de :

M. Bernard Bercu	Professeur, Université Bordeaux 1	Examinateur
M. Bruno Bouchard	Professeur, Université Paris 9	Examinateur
M. Rama Cont	Directeur de recherche, CNRS	Rapporteur
M. Pierre Del Moral	Directeur de Recherche, INRIA	Directeur de thèse
M. Benjamin Jourdain	Professeur, ENPC	Examinateur
Mme. Nadia Oudjane	Ingénieur chercheur, EDF R&D Clamart	Examinateur
M. Sumeetpal S. Singh	Professeur, University of Cambridge	Examinateur
M. Frederi Viens	Professeur, Purdue University	Rapporteur

INRIA Bordeaux - Sud-Ouest et
Institut de Mathématiques de Bordeaux
Université Bordeaux 1
352, cours de la Libération -F 33405
TALENCE cedex

École doctorale de Mathématiques
et Informatique de Bordeaux
U.F.R de Mathématiques Informatique
Bat A33, 1er étage, 352, cours de la
Libération -F 33405 TALENCE cedex

Méthodes particulaires et applications en finance

Résumé :

Cette thése est consacrée à l'analyse de ces modèles particulaires pour les mathématiques financières.

Le manuscrit est organisé en quatre chapitres. Chacun peut être lu séparément.

Le premier chapitre présente le travail de thèse de manière globale, définit les objectifs et résume les principales contributions. Le deuxième chapitre constitue une introduction générale à la théorie des méthodes particulaire, et propose un aperçu de ses applications aux mathématiques financières. Nous passons en revue les techniques et les résultats principaux sur les systèmes de particules en interaction, et nous expliquons comment ils peuvent être appliqués à la solution numérique d'une grande variété d'applications financières, telles que l'évaluation d'options compliquées qui dépendent des trajectoires, le calcul de sensibilités, l'évaluation d'options américaines ou la résolution numérique de problèmes de contrôle et d'estimation avec observation partielle.

L'évaluation d'options américaines repose sur la résolution d'une équation d'évolution à rebours, nommée l'enveloppe de Snell dans la théorie du contrôle stochastique et de l'arrêt optimal. Les deuxième et troisième chapitres se concentrent sur l'analyse de l'enveloppe de Snell et de ses extensions à différents cas particuliers. Un ensemble de modèles particulaires est alors proposé et analysés numériquement.

Mots-clés : Pricing d'option américaine, enveloppe de Snell, arrêt optimal, arbre génétique, évènement rare, système de particules en interaction, inégalités de concentration exponentielles.

Particle methods with applications in finance

Abstract:

This thesis is concerned with the analysis of these particle models for computational finance.

The manuscript is organized in four chapters. Each of them could be read separately.

The first chapter provides an overview of the thesis, outlines the motivation and summarizes the major contributions. The second chapter gives a general introduction to the theory of interacting particle methods, with an overview of their applications to computational finance. We survey the main techniques and results on interacting particle systems and explain how they can be applied to the numerical solution of a variety of financial applications; to name a few: pricing complex path dependent European options, computing sensitivities, pricing American options, as well as numerically solving partially observed control and estimation problems.

The pricing of American options relies on solving a backward evolution equation, termed Snell envelope in stochastic control and optimal stopping theory. The third and fourth chapters focus on the analysis of the Snell envelope and its variation to several particular cases. Different type of particle models are proposed and studied.

Keywords : Pricing of American option, Snell envelope, optimal stopping, genealogical trees, rare events, interacting particle system, exponential concentration inequalities.

Contents

Notations

Ω	Sample space
$\emptyset$	Empty set
$\sigma(F)$	Sigma-algebra generated by F
$\mathbb{E}$	Expectation operator
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^n$	n-dimensional Euclidean space
$\mathbb{N}$	Non-negative integers
$\mathbb{N}^+$	Positive integers
$\mathbb{P}$	Probability measure
$p(\cdot)$	Probability density
$\delta_x(\cdot)$	Dirac delta function centered on x
1_A	Indicator function of the set A
$\lvert\cdot\rvert$	Cardinality of a set
$\lVert\cdot\rVert_{L_p}$	$\mathbb{L}_P$ norm
E	General measurable space
$\mathcal{M}(E)$	Set of all finite positive measures on a measurable space
$\mathcal{P}(E)$	Set of all probability measures on a measurable space
$\mathcal{B}(E)$	Space of all bounded and measurable real-valued functions
$\mathcal{F}(E)$	Space of finite subsets of E
$\mathcal{C}(E)$	Collection of closed subsets of E
$\lVert\mu\rVert_{\text{tv}}$	Total variation norm
$\text{osc}(f)$	Oscillation of a measurable function f
$\mu(f)$	Lebesgue integral of f w.r.t. to the measure μ
f^-	The infimum of a function f
f^+	The supremum of a function f

Chapter 1

Introduction

1.1 Motivation

The history of mathematical and numerical finance starts in 1900, with the seminal thesis of Louis Bachelier, *Théorie de la Spéculation*, which introduced Brownian motion in order to model stock price movements and evaluate options. Not only did this remarkable work modeled the randomness of stock prices in a mathematical framework germane to the popular Nobel Prize in Economics winning solution proposed by Fischer Black, Myron Scholes and Robert Merton in 1973, but it also laid the foundation for some key concepts of stochastic analysis.

The celebrated Black-Scholes-Merton pricing paradigm which took the financial industry by storm, is not limited to the Samuelson's geometric Brownian motion model. However, it is based on a series of unrealistic assumptions, including Gaussian return fluctuations, constant volatility, risk-free interest rates, full liquidity, absence of frictions, no price impact from large or frequent trades, ..., and the list could go on. Furthermore, the original pricing arguments do not directly apply to derivatives with non-European exercises such as American options, without another level of sophistication and approximation.

The last two decades have seen a rapid development of increasingly realistic and sophisticated stochastic models and methods for pricing, hedging and risk management in rapidly growing markets, with more unfathomable financial products. Modern finance is becoming increasingly technical, requiring the use of complicated mathematical models, and involving numerical techniques based on theoretical results from subfields of mathematics ranging from stochastic analysis, dynamical system theory, nonlinear integro-differential equations, game theory, optimal control and dynamic programming, to statistical learning and information theory. Situated at the confluence of applied mathematics, computer sciences and economics,

quantitative finance distinguishes itself through its wide range of themes, and its interaction with a broad spectrum of scientific domains.

Research in financial mathematics is the driver for a great variety of numerical applications: parameter estimation, calibration of valuation models, derivative pricing, sensitivity analysis, hedging in incomplete markets, credit risk, risk and uncertainty quantification, portfolio optimization.... Resisting the temptation to cover as broad a range of applications as possible, we choose to focus on one particularly interesting issue, namely pricing of instruments with exercises of the American type. Our choice is motivated by the fact that optimal stopping problems offer a unique test-bed for research which goes beyond the edges of mathematical finance, at the crossroad of stochastic control and operations research. As a result, the numerical methods developed for these specific problems can in general be extended to a wide range of other stochastic control problems.

In the following sections, we will summarize the main results of this thesis. For more precise statements and detailed discussion, we refer the read to the related chapters and the references therein. As a service to the reader we also provide precise pointers to their location within each section of this manuscript.

Section 1.2 discusses the Snell envelope problem in the pricing of American options, and presents some mains results of chapter 3, such as a robustness lemma for the analysis of the approximation of Snell envelope, a particle algorithm based on genealogical tree, the non asymptotic estimates, as well as some numerical examples. This result has been published as a journal article [50] in *SIAM Journal on Financial Mathematics*.

Some of my works during my thesis are published as a book volume [51] of *Foundations and trends in Machine Learning*. This book presents some new concentration inequalities for Feynman-Kac particle processes. We analyze different types of stochastic particle methods, including particle profile occupation measures, genealogical tree based evolution methods, particle free energies, as well as backward Markov chain particle methods. Since we can not integrate the whole book in this thesis, we have chosen to select some interesting results and present them in section 1.3, and refer interested readers to [51]. In the end of section 1.3, we take the options' sensitivity computation as an example to introduce the chapter 2, which surveys the applications of particle methods in finance.

Following these particle methods, section 1.4 adapts the particle algorithm of section 1.2 to some specific cases. We give an overview of the chapter 4 with its mains results of convergence estimates and some numerical examples.

1.2 Pricing of American option

American option is a contract which allows the holder of the option the right to exercise the option at any point in time up to maturity. This kind of options are the most commonly traded products in the market. In discrete time setting, these problems are mathematically defined in terms of given real valued stochastic process $(Z_k)_{0\leq k\leq n}$, adapted to some increasing filtration $\mathcal{F} = (\mathcal{F}_k)_{0\leq k\leq n}$ that represents the available information at any time $0 \leq k \leq n$. For any $k \in \{0, \ldots, n\}$, we let $\mathcal{T}_k$ be the set of all stopping times τ taking values in $\{k, \ldots, n\}$. The Snell envelope of $(Z_k)_{0\leq k\leq n}$, is the stochastic process $(Y_k)_{0\leq k\leq n}$ defined for any $0 \leq k < n$ by the following backward equation

$$Y_k = Z_k \vee \mathbb{E}(Y_{k+1}|\mathcal{F}_k)$$

with the terminal condition $Y_n = Z_n$. The main property of this stochastic process is that

$$Y_k = \sup_{\tau\in\mathcal{T}_k} \mathbb{E}(Z_\tau|\mathcal{F}_k) = \mathbb{E}(Z_{\tau_k^*}|\mathcal{F}_k) \quad \text{with} \quad \tau_k^* = \min\{k \leq j \leq n \ : \ Y_j = Z_j\} \in \mathcal{T}_k$$

At this level of generality, in the absence of any additional information on the sigma-fields $\mathcal{F}_n$, or on the terminal random variable Z_n, no numerical computation of the Snell envelop is available. To get one step further, we assume that $(\mathcal{F}_n)_{n\geq 0}$ is the natural filtration associated with some Markov chain $(X_n)_{n\geq 0}$ taking values in some sequence of measurable state spaces $(E_n, \mathcal{E}_n)_{n\geq 0}$. We let $\eta_0 = \text{Law}(X_0)$ be the initial distribution on E_0, and we denote by $M_n(x_{n-1}, dx_n)$ the elementary Markov transition of the chain from E_{n-1} into E_n. We also assume that $Z_n = f_n(X_n)$, for some collection of non negative measurable functions f_n on E_n. In this situation, the computation of the Snell envelope amounts to solving the following backward functional equation

$$u_k = f_k \vee M_{k+1}(u_{k+1}) \tag{1.2.1}$$

for any $0 \leq k < n$, with the terminal value $u_n = f_n$. In the above displayed formula, $M_{k+1}(u_{k+1})$ stands for the measurable function on E_k defined for any $x_k \in E_k$ by the conditional expectation formula

$$\begin{aligned} M_{k+1}(u_{k+1})(x_k) &= \int_{E_{k+1}} M_{k+1}(x_k, dx_{k+1})\ u_{k+1}(x_{k+1}) \\ &= \mathbb{E}\left(u_{k+1}(X_{k+1})|X_k = x_k\right) \end{aligned} \tag{1.2.2}$$

1.2.1 A robustness lemma

Even it looks innocent, numerical solving numerically the recursion (1.2.1) often requires extensive calculations. The central problem is to compute the conditional

expectations $M_{k+1}(u_{k+1})$ on the whole state space E_k, at every time step $0 \leq k < n$. For Markov chain models taking values in some finite state spaces (with a reasonably large cardinality), the above expectations can be easily computed by a simple backward inspection of the whole realization tree that lists all possible outcomes and every transition of the chain. In more general situations, we need to resort to some approximation strategy. Most of the numerical approximation schemes amount to replacing the pair of functions and Markov transitions $(f_k, M_k)_{0\leq k\leq n}$ by some approximation model $(\widehat{f}_k, \widehat{M}_k)_{0\leq k\leq n}$ on some possibly reduced measurable subsets $\widehat{E}_k \subset E_k$. We let $\widehat{u}_k$ be the Snell envelope of $\widehat{f}_k$ associated with the sequence of operators $\widehat{M}_k$ from $\widehat{E}_{k-1}$ into $\widehat{E}_k$. Using the elementary inequality

$$|(a \vee a') - (b \vee b')| \leq |a - a'| + |b - b'|$$

which is valid for any $a, a', b, b' \in \mathbb{R}$, one readily obtains, for any $0 \leq k < n$

$$\begin{aligned} |u_k - \widehat{u}_k| &\leq |f_k - \widehat{f}_k| + |M_{k+1}u_{k+1} - \widehat{M}_{k+1}\widehat{u}_{k+1}| \\ |M_{k+1}u_{k+1} - \widehat{M}_{k+1}\widehat{u}_{k+1}| &\leq |(M_{k+1} - \widehat{M}_{k+1})u_{k+1}| + \widehat{M}_{k+1}|u_{k+1} - \widehat{u}_{k+1}|. \end{aligned}$$

Iterating the argument, one finally arrives at the following robustness lemma.

Lemma 1.2.1. *For any $0 \leq k < n$, on the state space $\widehat{E}_k$, we have that*

$$|u_k - \widehat{u}_k| \leq \sum_{l=k}^{n} \widehat{M}_{k,l}|f_l - \widehat{f}_l| + \sum_{l=k}^{n-1} \widehat{M}_{k,l}|(M_{l+1} - \widehat{M}_{l+1})u_{l+1}|$$

This lemma provides a simple and natural way to analyze the robustness properties of the Snell equation with respect to the pair parameters (f_k, M_k).

We emphasize that this non asymptotic robustness analysis also allows to combine in a natural way several approximation methods. For instance, under appropriate tightness conditions, cut-off techniques can be used to reduce the numerical analysis of (1.2.1) to compact state spaces $\widehat{E}_n$ and bounded functions $\widehat{f}_n$. In the same line of ideas, in designing any type of Monte Carlo approximation methods, we can suppose that the transitions of the chain X_n is known based on a preliminary analysis of Euler type approximation methods.

To illustrate the generality and robustness of this framework, we state the analyse of Broadie-Glasserman method in chapter 3 as an example. This importance sampling method was introduced in 1997 by Broadie and Glasserman. They assume that there exist a sequence of measures$(\eta_k)_{0\leq k\leq n}$ which are equivalent to the underling asset dynamic transition 1.2.2 such that

$$M_k(x_{k-1}, \cdot) \sim \eta_k.$$

We further construct $\widehat{\eta}_k = \frac{1}{N}\sum_{i=1}^N(\delta_{\xi_k^i})$ the occupation measure associated with a sequence of independent random variables $\xi_k := (\xi_k^i)_{1\le i\le N}$ with common distribution η_k. In this context, we can construct the approximation model $\widehat{M}_{k+1}$ with the following change of measure:

$$\widehat{M}_{k+1}(x_k, dx_{k+1}) := \widehat{\eta}_{k+1}(dx_{k+1})R_{k+1}(x_k, x_{x_{k+1}}),$$

where

$$R_{k+1}(x_k, x_{x_{k+1}}) := \frac{dM_{k+1}(x_k,\cdot)}{d\eta_{k+1}}(x_{k+1}).$$

By replacing the transition M_{k+1} in the recursion 1.2.1, we provide an approximation of the Snell envelope:

$$\widehat{u}_k = f_k \vee \widehat{M}_{k+1}(u_{k+1}), \tag{1.2.3}$$

for $0 \le k \le n$, with the terminal condition $\widehat{u}_n = f_n$. By Khintchine's inequality, we can easily provide an estimate of the local approximation error:

$$\sqrt{N}\ \widehat{\mathbb{E}}_{\eta_0}\left(\left|\left[M_{l+1} - \widehat{M}_{l+1}\right](f)(x_l')\right|^p\right)^{\frac{1}{p}} \le 2\ a(p)\ \eta_{l+1}\left[(R_{l+1}(x_l,\cdot)f)^p\right]^{\frac{1}{p}}\ .$$

where the constants $a(p)$ are defined as, for any non-negative integer r:

$$a(2r)^{2r} = (2r)_r\ 2^{-r} \quad \text{and} \quad a(2r+1)^{2r+1} = \frac{(2r+1)_{r+1}}{\sqrt{r+1/2}}\ 2^{-(r+1/2)}\ , \tag{1.2.4}$$

with the notation $(q)_p = q!/(q-p)!$, for any $1 \le p \le q$.

Here, we recall the approximation error bounds in the robustness lemma:

$$|u_k - \widehat{u}_k| \ \le\ \sum_{l=k}^{n-1} \widehat{M}_{k,l}|(M_{l+1} - \widehat{M}_{l+1})u_{l+1}|$$

Combining the above inequality and the local error, we are ready to state the following theorem.

Theorem 1.2.1. *For any integer $p \ge 1$, we denote by p' the smallest even integer greater than p. Then for any time horizon $0 \le k \le n$, and any $x_k \in E_k$, we have*

$$\begin{aligned}&\sqrt{N}\widehat{\mathbb{E}}_{\eta_0}\left(|u_k(x_k) - \widehat{u}_k(x_k)|^p\right)^{\frac{1}{p}} \\ &\qquad \le 2a(p)\sum_{k\le l<n}\left\{\int M_{k,l}(x_k, dx_l)\eta_{l+1}\left[(R_{l+1}(x_l,\cdot)u_{l+1})^{p'}\right]\right\}^{\frac{1}{p'}}\ .\end{aligned} \tag{1.2.5}$$

As mentioned previously, this framework can be easily applied to different approximation methods. In chapter 3, we apply this result to a series of approximation methods, including interpolation methods, cut-off type approximations, Euler time

schemes, quantification tree methods, and the Monte Carlo method of Broadie-Glasserman. In each situation, we provide non asymptotic convergence estimates, including $\mathbb{L}_p$-mean error bounds and exponential concentration inequalities. Without any doubt, the theory of empirical processes and measure concentration is one of the most powerful mathematical tools to analyze the deviations of Monte Carlo based approximations. Most $\mathbb{L}_p$-mean error bounds presented in this thesis have a form like 1.2.5. We present here a exponential concentration inequality lemma to better understand these non asymptotic convergence estimates:

Lemma 1.2.2. *Suppose the estimates have the following form:*

$$\sqrt{N} \sup_{x\in E_k} \mathbb{E}\left(|u_k(x) - \widehat{u}_k(x)|^p\right)^{\frac{1}{p}} \leq a(p)b_k(n),$$

where $b_k(n)$ are some finite constants whose values do not depend on the parameter p and $a(p)$ is a collection of constants defined in theorem 1.2.1 Then we deduce the following exponential concentration inequality

$$\sup_{x\in E_k} \mathbb{P}\left(|u_k(x_k) - \widehat{u}_k(x_k)| > \frac{b_k(n)}{\sqrt{N}} + \epsilon\right) \leq exp\left(-N\epsilon^2/(2b_k(n)^2)\right) \ . \tag{1.2.6}$$

Considering again the example of the Broadie-Glasserman method, this lemma implies that the probability of making some level of approximation error in this method is exponentially small.

1.2.2 A genealogical tree based particle method

In the final part of chapter 3, we propose a genealogical tree based algorithm based on a mean field approximation of the reference Markov process in terms of a neutral type genetic method. In contrast to Broadie-Glasserman Monte Carlo methods, the computational cost of this new stochastic particle approximation is proportional to the size of the random particle samples.

To have a idea of this particle method, we present here the pseudo algorithm in avoiding the extensive mathematics notations. We refer the reader to section 3.5 for the rigorous analysis.

The initial particle system $\xi_0^{(N)} = \left(\xi_0^{(i,N)}\right)_{0\leq i\leq N_0}$, is a sequence of N i.i.d. random copies of X_0. To simplify the presentation, when there is no confusion we suppress the population size parameter N, and we write ξ_k and ξ_k^i instead of $\xi_k^{(N)}$ and $\xi_k^{(i,N)}$. By construction, ξ_k is a genetic type method with a neutral selection transition and a mutation type exploration

$$\xi_k \in E_k^N \xrightarrow{Selection} \widehat{\xi}_k := \left(\widehat{\xi}_k^i\right)_{1\leq i\leq N} \in E_k^N \xrightarrow{Mutation} \xi_{k+1} \in E_{k+1}^N \ . \tag{1.2.7}$$

During the selection transition, we select randomly N path-valued particles $\widehat{\xi}_k := \left(\widehat{\xi}_k^i\right)_{1\leq i\leq N}$ among the N path-valued particles $\xi_k = (\xi_k^i)_{1\leq i\leq N}$. Sometimes, this elementary transition is called a neutral selection transition in the literature on genetic population models. During the mutation transition $\widehat{\xi}_k \rightsquigarrow \xi_k$, every selected path valued individual $\widehat{\xi}_k^i$ evolves randomly to a new path valued individual $\xi_{k+1}^i = (\widehat{\xi}_k^i, Y^i)$, by adding the variable $Y^i = x$ randomly chosen with the distribution $M_{k+1}(\widehat{\xi}_{k,k}^i, x)$, with $1 \leq i \leq N$. In previous display, every particle is a path-valued random variable defined by

$$\begin{array}{rcl} \xi_k^i & := & \left(\xi_{0,k}^i, \xi_{1,k}^i, \ldots, \xi_{k,k}^i\right) \\ \widehat{\xi}_k^i & := & \left(\widehat{\xi}_{0,k}^i, \widehat{\xi}_{1,k}^i, \ldots, \widehat{\xi}_{k,k}^i\right) \in (E_0 \times \ldots \times E_k) \ . \end{array}$$

By definition of the transition in path space, we also have that

$$\begin{array}{rcl} \xi_{k+1}^i & = & \left(\underbrace{\left(\xi_{0,k+1}^i, \xi_{1,k+1}^i, \ldots, \xi_{k,k+1}^i\right)}_{||}, \xi_{k+1,k+1}^i \right) \\ & = & \left(\overbrace{\left(\widehat{\xi}_{0,k}^i, \ \widehat{\xi}_{1,k}^i, \ldots, \ \widehat{\xi}_{k,k}^i\right)}, \ \xi_{k+1,k+1}^i \right) = \left(\widehat{\xi}_k^i, \xi_{k+1,k+1}^i\right) \ , \end{array}$$

where $\xi_{k+1,k+1}^i$ is a random variable with distribution $M_{k+1}(\widehat{\xi}_{k,k}^i, \cdot)$. In other words, the mutation transition $\widehat{\xi}_k^i \rightsquigarrow \xi_{k+1}^i$ simply consists in extending the selected path $\widehat{\xi}_k^i$ with an elementary move $\widehat{\xi}_{k,k}^i \rightsquigarrow \xi_{k+1,k+1}^i$ of the end point of the selected path.

The simulation mentioned above of the particle system can be summarized by the following pseudo algorithm:

Initialization At time step $k = 0$, generate N i.i.d. random copies of X_0 and set $\xi_0 = (\xi_0^i)_{0\leq i\leq N}$.

At each time step $k = 1, \cdots, n$

1. **Selection**: For each $i = 1, \cdots, N$, generate independently an indice $I_i \in \{1, \cdots, N\}$ with probability $\mathbb{P}(I_i = j) = 1/N$. Then set $\hat{\xi}_{k-1}^i = \xi_{k-1}^{I_i}$.
2. **Mutation**: For each $i = 1, \cdots, N$, generate independently N i.i.d. random variables $(\xi_{k,k}^i)_{0\leq i\leq N}$ according to the transition kernel $M_k(\hat{\xi}_{k-1,k-1}^i, \cdot)$. Then set $\xi_k^i = (\hat{\xi}_{k-1}^i, \xi_{k,k}^i)$.

With this particle system, we can construct the approximation of the Snell envelope 1.2.1 in the following backward algorithm:

Initialization At time step $k = n$, for all $i = 1, \cdots, N$, set $\hat{u}_n(\xi_{n,n}^i) = f(\xi_{n,n}^i)$.

At each time step $k = n-1, \cdots, 0$, for all $i = 1, \cdots, N$ set

$$\hat{u}_k(\xi^i_{k,n}) = f_k(\xi^i_{k,n}) \vee \frac{\sum_{j=1}^N \hat{u}_{k+1}(\xi^j_{k+1,n})\, 1_{\xi^j_{k,n}=\xi^i_{k,n}}}{\sum_{j=1}^N 1_{\xi^j_{k,n}}} \ .$$

In this context, we can provide by applying again the robustness lemma the non-asymptotic estimate of this approximation method:

Theorem 1.2.2. *For any $p \geq 1$, and $0 \leq i \leq N$ we have the following uniform estimate*

$$\sup_{0\leq k\leq n} \left\| (u_k - \widehat{u}_k)(\xi^i_{k,n}) \right\|_p \leq c_p(n)/\sqrt{N} \ , \tag{1.2.8}$$

with some collection of finite constants $c_p(n) < \infty$ whose values only depend on the parameters p and n.

In addition to above non asymptotic estimate, some bias analysis are also provided:

Theorem 1.2.3. *For any $0 \leq k \leq n$ and any $i \in \{1, \ldots, N\}$, we have*

$$\mathbb{E}\left(\widehat{u}_k(\xi^i_{k,n})|\xi_{k,n}\right) \geq u_k(\xi^i_{k,n}) \ .$$

As shown in above theorem, the estimator is always biased upward. In general, it is useful to know that the bias is positive. Once we compute the approximation of the Snell envelope, we can extract the exercise policy. Based on this stopping strategy, another estimate can be established with a negative bias. In practice, it is convenient that the biases of the two estimators have opposite signs.

We end this section with the numerical simulations in chapter 3. The numerical examples are taken from Bouchard and Warin [14], who provided precise approximations of option values in their examples. The asset prices are modeled by a d-dimensional Markov process $(\tilde{X}_t)$ such that each component (i.e. each asset) follows a geometric Brownian motion under the risk-neutral measure, that is, for assets $i = 1, \cdots, d$,

$$\frac{d\tilde{X}_t(i)}{\tilde{X}_t(i)} = rdt + \sigma_i dz^i_t \ , \tag{1.2.9}$$

where z^i, for $i = 1, \cdots, d$ are independent standard Brownian motions. The interest rate r is set to 5% annually. We also assume that for all $i = 1, \cdots, d$, $\tilde{X}_{t_0}(i) = 1$ and $\sigma_i = 20\%$ annually.
We consider two different Bermudan options with maturity $T = 1$ year and 11 equally distributed exercise opportunities at dates $t_k = kT/n$ with $k = 0, 1, \cdots, n = 10$, associated with two different payoffs:

1. a geometric average put option with strike $K = 1$ and payoff $(K - \prod_{i=1}^d \tilde{X}_T(i))_+$,

2. an arithmetic average put option with strike $K = 1$ and payoff $(K - \frac{1}{d}\sum_{i=1}^{d} \tilde{X}_T(i))_+$.

We report in Table 1.1 the benchmark option values computed in [14], for both the geometric and arithmetic put options (by using respectively the one dimensional PDE method and the least squares regression method with $8 \times 10^6 \times d^2$ simulations and ten basis functions for each direction).

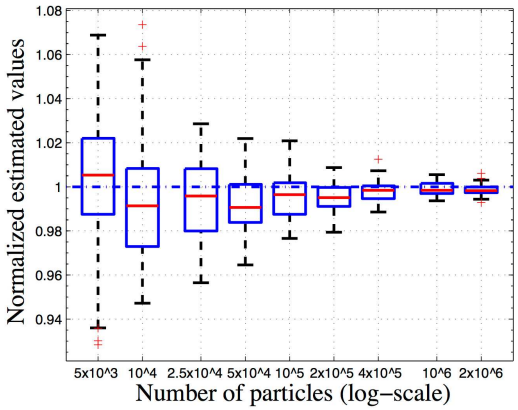
Number of assets	1	2	3	4	5	6
Geometric Payoff	0.06033	0.07815	0.08975	0.09837	0.10511	0.11073
Arithmetic Payoff	0.06033	0.03882	0.02947	0.02403	0.02046	0.01830

Table 1.1: Benchmark values for the geometric and arithmetic put options (taken from [14]).

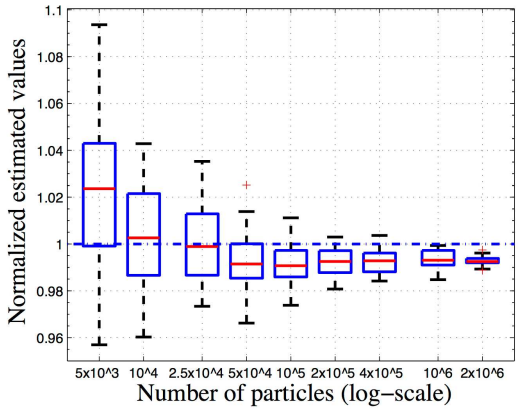
The genealogical tree algorithm is designed for finite state spaces. Hence, before applying it to the aforementioned continuous space examples, we have to approximate the continuous state space Markov chain solution of (4.7.1) by a Markov chain with a finite state space. To this end, one can first discretize the state space using either a random tree, or a stochastic mesh, or a Binomial tree or a quantization approach ... The state aggregation technique we use in this thesis is a quantization-like approach. This step is quiet technical, we refer the reader to section 3.5.5 for more details.

Simulations results are reported in Figure 1.1 for the geometric put payoff and in Figure 1.2 for the arithmetic put payoff. First notice that these results are consistent with theorem 1.2.3, and that the estimates have a positive bias in most cases (Notice that one can observe on the graph that for $d = 2$ or 3 the bias of our estimator can be negative. The negative bias arises in the discretization of the state space). Also notice that our algorithm has been implemented without any control variate technique. Moreover, our implementation has not been optimized. In particular, we have not investigated in this thesis any parallel implementations of our algorithm. Thus, it seems not relevant to report any running time measurements. We refer the reader to section 3.5.5 for the algorithm complexity, which gives a good indication of the number of operations required by our algorithm.

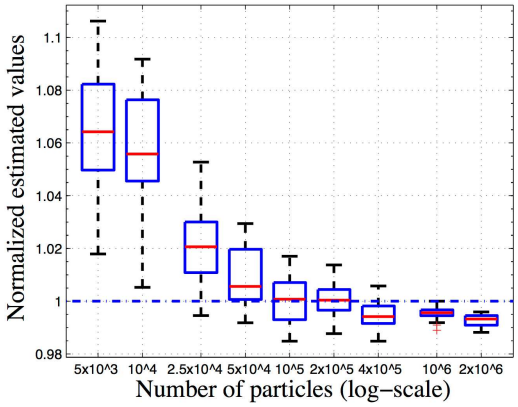
Hence, to compare the estimation errors of the *backward estimate* provided by our algorithm to a corresponding approach, we have reported, in Table 1.2, the estimation errors obtained with the genealogical algorithm using $N = 25000$ particles and $N^{\frac{d}{d+2}}$ sites in the space discretization step, in valuing the geometric put (on the first line) and the arithmetic put (on the second line) and, within parenthesis, the performances of the *backward estimate* provided by the quantization approach [4] implemented in [14], with 25600 quantization points for the same options. One can observe that both algorithms achieve similar performances for approximately the



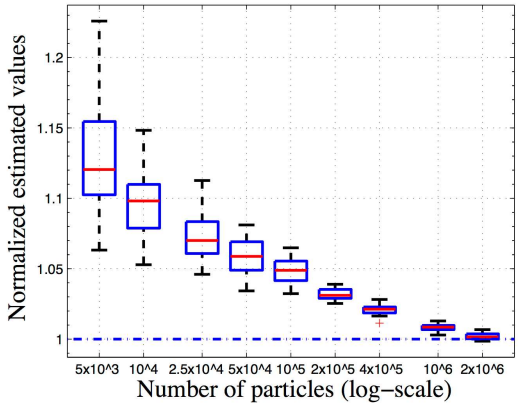
(a) $d = 1$



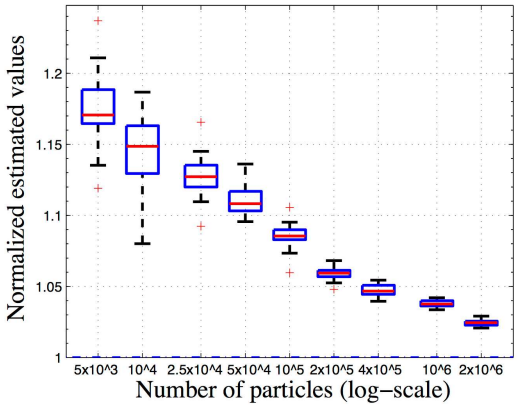
(b) $d = 2$



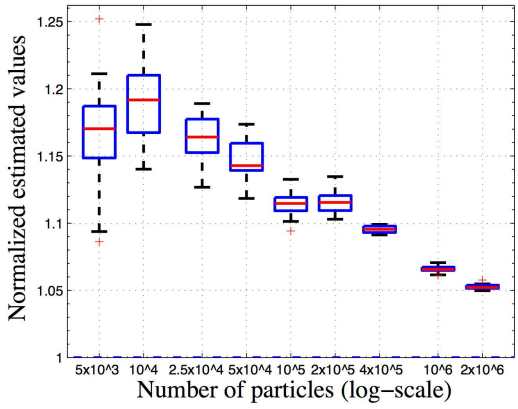
(c) $d = 3$



(d) $d = 4$

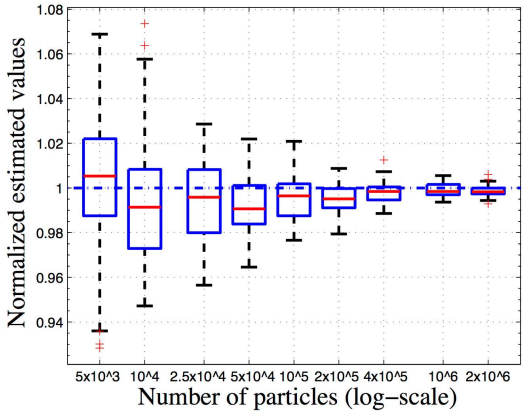


(e) $d = 5$

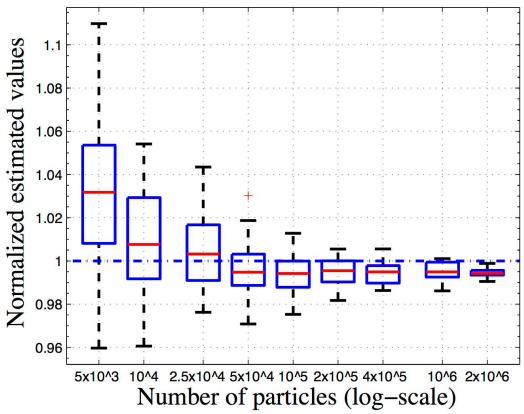


(f) $d = 6$

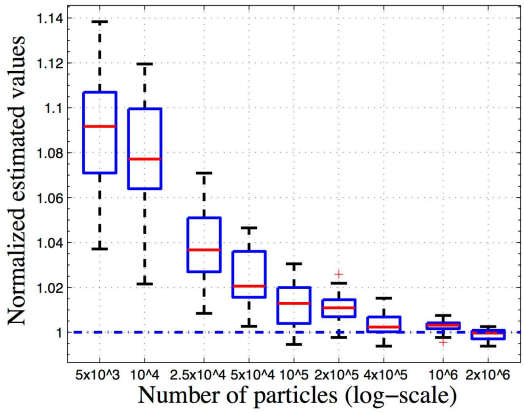
Figure 1.1: Boxplots for estimated option values (divided by the benchmark values) as a function of the number of particles for the **geometric** put-payoff. The box stretches from the 25th percentile to the 75th percentile, the median is shown as a line across the box, the whiskers extend from the box out to the most extreme data value within 1.5 IQR (Interquartile Range) and red crosses indicates outliers.



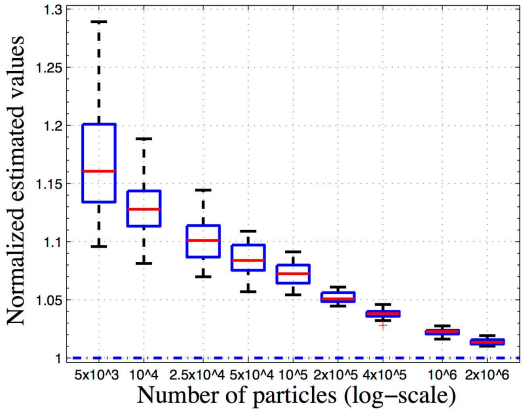
(a) $d = 1$



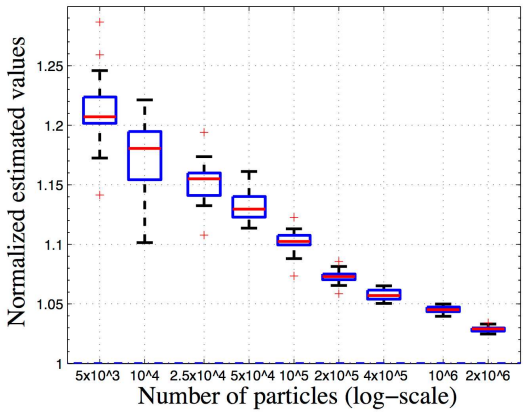
(b) $d = 2$



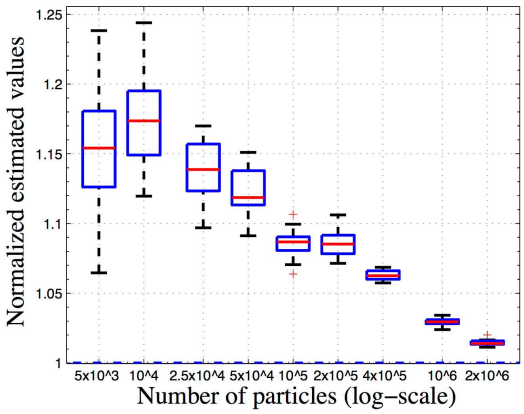
(c) $d = 3$



(d) $d = 4$



(e) $d = 5$



(f) $d = 6$

Figure 1.2: Boxplots for estimated option values (divided by the benchmark values) as a function of the number of particles for the **arithmetic** put-payoff. The box stretches from the 25th percentile to the 75th percentile, the median is shown as a line across the box, the whiskers extend from the box out to the most extreme data value within 1.5 IQR (Interquartile Range) and red crosses indicates outliers.

same number N of quantization points (for the quantization algorithm) and particles (for the genealogical algorithm).
Now, we can see in section 3.5.5 that the complexity (per time step) of the genealogical algorithm is of order $N^{\frac{2d+2}{d+2}}$ for the construction of the genealogical tree and of order N for the backward induction on the prices, which is slightly smaller than the complexity of the quantization approach of order N^2 for the backward induction on prices (without taking into account the complexity related to the construction of the quantization tree and to the computation of the transition probabilities). Hence, we can conclude that our new algorithm is competitive with respect to comparable algorithms.

Number of assets	$d = 3$	$d = 4$	$d = 5$	$d = 6$
Geometric Put error (in % of the option value)	2 (2)	7 (8)	14 (15)	17 (22)
Arithmetic Put error (in % of the option value)	3.5 (3.5)	10 (8)	15 (16)	14 (17)

Table 1.2: Error (in % of the option value in Table 1.1) of the genealogical algorithm with $N = 25000$ particles and $N' = N^{\frac{d}{d+2}}$ sites, and within parenthesis of the quantization algorithm with $N = 25600$ quantization points, (taken from [14]) for the geometric and arithmetic put options.

1.3 Particle methods

The numerical technique we use previously is based on the interacting particle systems. Though these particle methods are known as powerful numerical tools, these algorithms have rarely been applied to computational finance fields. In chapter 2, we present an overview of these particle techniques and their applications in finance. Here, we summarize this presentation with some interesting financial examples in this section.

Stochastic particle methods are increasingly used to solve a variety of problems, including nonlinear filtering equations, data assimilation problems, rare event sampling, hidden Markov chain parameter estimation, stochastic control problems and financial mathematics. To illustrate these methods, we start with a toy model of barrier option pricing as example. We consider the underlying asset as a Markov chain X_k taking values in $\mathbb{R}^d$, we will denote by $M = (M_{p,n})_{p,n}$ its transition probability

$$M_{p,n}(x, dy) := \mathbb{P}(X_n \in dy | X_p = x), \qquad x \in \mathbb{R}^d. \tag{1.3.1}$$

With a sequence of barrier sets $(A_k)_{0\leq k<n}$ and the payoff function f, we are then interested in the price of the barrier options given by:

$$\mathbb{E}\left(f(X_n)\prod_{p=1}^{n-1}1_{A_p}(X_p)\right).$$

which is, in most cases, difficult to compute. One natural way to resolve this estimation problem is to work with its associated conditional expectation:

$$\mathbb{E}\left(f(X_n)|\, X_p\in A_p, 0\leq p<n\right).$$

To compute above expectation, we can use the genealogical tree model associated with a genetic type interacting particle model. This genetic algorithm is defined with mutation transitions according to 1.3.1, and proportional selections with regard to (w.r.t.) the potential functions $(A_k)_{0\leq k<n}$. The occupation measures of the corresponding genealogical tree provides an approximation of the desired conditional distributions of the underlying asset. More generally, for any function F on the path space we have

$$\lim_{N\uparrow\infty}\frac{1}{N}\sum_{1}^{N}F(line_n(i))\simeq_{N\uparrow\infty}\mathbb{E}(F(X_0,\ldots,X_n)|X_p\in A_p, 0\leq p<n) \qquad (1.3.2)$$

where $\text{line}_n(i)$ stands for the $i-$th ancestral line of the genealogical tree, at time n.

1.3.1 Feynman-Kac measures

Interacting particle systems aim to design the interacting particle approximation of the Feynman-Kac measures, which represent the distribution of the paths of a Markov process, weighted by a collection of potential functions. These functional models encapsulate traditional changes of probability measures, commonly used in importance sampling, posterior distributions in Bayesian statistics, and the optimal filter in nonlinear filtering problems.

These stochastic models are defined in terms of only two ingredients:

A Markov chain X_n, with Markov transition M_n on some measurable state spaces $(E_n,\mathcal{E}_n)$ with initial distribution η_0, and a sequence of $(0,1]$-valued potential functions G_n on the set E_n.

The Feynman-Kac path measure associated with the pairs (M_n,G_n) is the probability measure $\mathbb{Q}_n$ on the product state space

$$\mathbf{E}_n:=(E_0\times\ldots\times E_n)$$

defined by the following formula

$$d\mathbb{Q}_n:=\frac{1}{\mathcal{Z}_n}\left\{\prod_{0\leq p<n}G_p(X_p)\right\}d\mathbb{P}_n \qquad (1.3.3)$$

where $\mathcal{Z}_n$ is a normalizing constant and $\mathbb{P}_n$ is the distribution of the random paths

$$\mathbf{X}_n = (X_0, \ldots, X_n) \in \mathbf{E}_n$$

of the Markov process X_p from the origin $p = 0$ with initial distribution η_0, up to the current time $p = n$. We also denote by

$$\Gamma_n = \mathcal{Z}_n \, \mathbb{Q}_n \tag{1.3.4}$$

its unnormalized version.

Several probabilistic interpretations in the domain of quantitative finance of this model can be found in chapter 2. To have an idea of these models, we take again the barrier option example presented in the beginning of this section:

$$X_n^A \in A_n := E_n \xrightarrow{\text{absorption } \sim A_n^c} \widehat{X}_n^A \xrightarrow{\text{exploration } \sim M_{n+1}} X_{n+1}^A. \tag{1.3.5}$$

The chain X_n^A starts at some initial state X_0^A randomly chosen with distribution η_0. During the selection stage, we set $\widehat{X}_n^A = X_n^A$ if $X_n^A \in A_n$, otherwise we put the particle in an auxiliary cemetery set A_n^c. When the particle $\widehat{X}_n^A$ is still alive (that is, if we have $\widehat{X}_n^c \in A_n$), it performs an elementary move $\widehat{X}_n^A \rightsquigarrow X_{n+1}^A$ according to the Markov transition M_{n+1}. Otherwise, the particle is absorbed and we set $X_p^A = \widehat{X}_n^A$, for any time $p > n$.

If we let T be the first time $\widehat{X}_n^c \notin A_n$, then we have the Feynman-Kac representation formulae

$$\mathbb{Q}_n = \mathrm{Law}((X_0^A, \ldots, X_n^A) \mid T \geq n) \quad \text{and} \quad \mathcal{Z}_n = \mathrm{Proba}\,(T \geq n)\,.$$

We also denote by η_n and γ_n, the n-th time marginal of $\mathbb{Q}_n$ and Γ_n. It is a simple exercise to check that

$$\gamma_n = \gamma_{n-1} Q_n \quad \text{and} \quad \eta_{n+1} = \Phi_{n+1}(\eta_n) := \Psi_{G_n}(\eta_n) M_{n+1} \tag{1.3.6}$$

with the positive integral operator

$$Q_n(x, dy) = G_{n-1}(x) \; M_n(x, dy)$$

and the Boltzmann-Gibbs transformation

$$\Psi_{G_n}(\eta_n)(dx) = \frac{1}{\eta_n(G_n)} \; G_n(x) \; \eta_n(dx). \tag{1.3.7}$$

In addition, the normalizing constants $\mathcal{Z}_n$ can be expressed in terms of the flow of marginal measures η_p, from the origin $p = 0$ up to the current time n, with the following multiplicative formulae:

$$\mathcal{Z}_n := \gamma_n(\mathbb{1}) = \mathbb{E}\left(\prod_{0 \leq p < n} G_p(X_p)\right) = \prod_{0 \leq p < n} \eta_p(G_p). \tag{1.3.8}$$

This multiplicative formula is easily checked using the induction

$$\gamma_{n+1}(1) = \gamma_n(G_n) = \eta_n(G_n)\ \gamma_n(1).$$

The abstract formulae discussed above are more general than they may appear. For instance, they can be used to analyze, without further work, path spaces models, including historical processes or transition space models, as well as finite excursion models. These functional models also encapsulate quenched Feynman-Kac models, Brownian type bridges and linear Gaussian Markov chains conditioned on starting and end points.

When the Markov transitions M_n are absolutely continuous with respect to some measures λ_n on E_n, for any $(x, y) \in (E_{n-1} \times E_n)$ we have

$$H_n(x, y) := \frac{dM_n(x, .)}{d\lambda_n}(y) > 0. \tag{1.3.9}$$

We also have the following backward formula

$$\mathbb{Q}_n(d(x_0, \ldots, x_n)) = \eta_n(dx_n)\ \prod_{q=1}^{n} \mathbb{M}_{q,\eta_{q-1}}(x_q, dx_{q-1}) \tag{1.3.10}$$

with the the collection of Markov transitions defined by

$$\mathbb{M}_{n+1,\eta_n}(x, dy)\ \propto\ G_n(y)\ H_{n+1}(y, x)\ \eta_n(dy). \tag{1.3.11}$$

1.3.2 Interacting particle systems

In this subsection we explain how to design an interacting particle approximation of the Feynman-Kac measures introduced in the previous paragraphs. These particle methods can be interpreted in different ways, depending on the application domain in which they are considered.

In the filtering example presented at the beginning of this chapter, these particle algorithms can be seen as a stochastic adaptive fixed approximation of the filtering equations. From a purely statistical point of view, these algorithms can also be seen as a sophisticated acceptance-rejection technique with an interacting recycling transition.

The particle model is defined as follows:

We start with a population of N possible candidate solutions $(\xi_0^1, \ldots, \xi_0^N)$ randomly chosen w.r.t. some distribution η_0.

The coordinates ξ_0^i are also called individuals or phenotypes, with $1 \leq N$. The random evolution of the particles is decomposed into two main steps : the free exploration and the adaptive selection transition.

During the updating-selection stage, multiple individuals in the current population $(\xi_n^1, \ldots, \xi_n^N)$ at time $n \in \mathbb{N}$ are stochastically selected based on the fitness

function G_n. In practice, we choose a random proportion B_n^i of an existing solution ξ_n^i in the current population with a mean value $\propto G_n(\xi_n^i)$ to breed a brand new generation of "improved" solutions $(\widehat{\xi}_n^1, \ldots, \widehat{\xi}_n^N)$. For instance, for every index i, with a probability $\epsilon_n G_n(\xi_n^i)$, we set $\widehat{\xi}_n^i = \xi_n^i$, otherwise we replace ξ_n^i with a new individual $\widehat{\xi}_n^i = \xi_n^j$ randomly chosen from the whole population with a probability proportional to $G_n(\xi_n^j)$. The parameter $\epsilon_n \geq 0$ is a tuning parameter that must satisfy the constraint $\epsilon_n G_n(\xi_n^i) \leq 1$, for every $1 \leq i \leq N$. During the mutation stage, every selected individual $\widehat{\xi}_n^i$ moves to a new solution $\xi_{n+1}^i = x$ randomly chosen in E_{n+1}, with a distribution $M_{n+1}(\widehat{\xi}_n^i, dx)$.

If we interpret the updating-selection transition as a birth and death process, then the important notion of the ancestral line of a current individual arises. More precisely, when a particle $\widehat{\xi}_{n-1}^i \longrightarrow \xi_n^i$ evolves to a new location ξ_n^i, we can interpret $\widehat{\xi}_{n-1}^i$ as the parent of ξ_n^i. Looking backwards in time and recalling that the particle $\widehat{\xi}_{n-1}^i$ has selected a site ξ_{n-1}^j in the configuration at time $(n-1)$, we can interpret this site ξ_{n-1}^j as the parent of $\widehat{\xi}_{n-1}^i$ and therefore as the ancestor denoted $\xi_{n-1,n}^i$ at level $(n-1)$ of ξ_n^i. Running backwards in time we may trace the whole ancestral line as

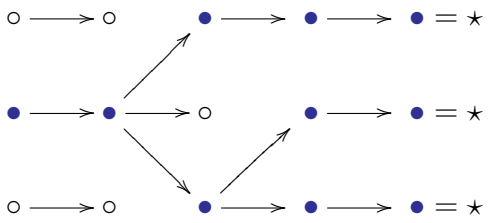
$$\xi_{0,n}^i \longleftarrow \xi_{1,n}^i \longleftarrow \ldots \longleftarrow \xi_{n-1,n}^i \longleftarrow \xi_{n,n}^i = \xi_n^i. \tag{1.3.12}$$

In the interacting particle system literature, most of the terminology we have used is drawn from filtering and genetic evolution theories. In this thesis we adapt these notions to the filed of quantitative finance.

In the barrier option example presented in the beginning of this chapter, the former particle model is dictated by the two-steps mutation-selection learning equations of the conditional distributions of the underlying asset X_k, given their survival zones A_k. In this setting, the potential functions represent the absorption w.r.t. the barrier sets A_k, while the free exploration transitions are related to the Markov transitions of the underlying asset. More formally, using the notation we used in the barrier option example, we have:

$$\mathbb{P}(X_k = x_k | X_{k-1} = x_{k-1}) = M_k(x_{k-1}|x_k) \quad \text{and} \quad 1_{A_k}(x_k) = G_k(x_k).$$

The different types of particle approximation measures associated with the genetic type particle model described above are summarized in the following synthetic picture corresponding to the case $N = 3$.



1.3.3 Concentration inequalities

In this subsection, we give an overview of the four particle approximation measures that can be extracted from the interacting population evolution model described above. Our concentration inequalities will basically be stated. The proofs of these results are quite subtle and are not provided in the this thesis. We encourage the reader to explore the book [51] for details and the references therein. In the further development of the next subsections, c_1 stands for a finite constant related to *the bias* of the particle model, while c_2 is related to *the variance* of the scheme. The value of these constants may vary from one line to another, but in all the situations they do not depend on the time parameter.

The precise form of the constants in these exponential inequalities depends on the contraction properties of Feynman-Kac flows. Our stochastic analysis requires us to combine the stability properties of the nonlinear semigroup of the Feynman-Kac distribution flow η_n, with the deep convergence results of empirical processes theory associated with interacting random samples.

Last population models

The occupation measures of the current population, represented by the stars in the above figure

$$\eta_n^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^i}$$

converge to the n-th time marginals η_n of the Feynman-Kac measures $\mathbb{Q}_n$. We shall measure the performance of these particle estimates through several concentration inequalities, with a special emphasis on uniform inequalities w.r.t. the time parameter. Our results will basically be stated as follows.

1) For any time horizon $n \geq 0$, any bounded function f, any $N \geq 1$, and for any $x \geq 0$, the probability of the event

$$\left[\eta_n^N - \eta_n\right](f) \leq \frac{c_1}{N}\ \left(1 + x + \sqrt{x}\right) + \frac{c_2}{\sqrt{N}}\ \sqrt{x}$$

is greater than $1 - e^{-x}$.

We have already mentioned one important consequence of these uniform concentration inequalities for time homogeneous Feynman-Kac models. Under some regularity conditions, the flow of measures η_n tends to some fixed point distribution η_∞, in the sense that

$$\|\eta_n - \eta_\infty\|_{tv} \leq c_3\ e^{-\delta n} \tag{1.3.13}$$

for some finite positive constants c_3 and δ. In the above display $\|\nu - \mu\|_{tv}$ stands for the total variation distance. As a direct consequence of the above inequalities,

we find that for any $x \geq 0$, the probability of the following events is greater than $1 - e^{-x}$:

$$\left[\eta_n^N - \eta_\infty\right](f) \leq \frac{c_1}{N}\ \left(1 + x + \sqrt{x}\right) + \frac{c_2}{\sqrt{N}}\ \sqrt{x} + c_3\ e^{-\delta n}.$$

2) For any $x = (x_i)_{1 \leq i \leq d} \in E_n = \mathbb{R}^d$, we set $(-\infty, x] = \prod_{i=1}^d (-\infty, x_i]$ and we consider the repartition functions

$$F_n(x) = \eta_n\left(1_{(-\infty,x]}\right) \quad \text{and} \quad F_n^N(x) = \eta_n^N\left(1_{(-\infty,x]}\right).$$

The probability of the following event

$$\sqrt{N}\ \left\|F_n^N - F_n\right\| \leq c\ \sqrt{d\ (x+1)}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some universal constant $c < \infty$ that does not depend on the dimension, nor on the time parameter. In the above display $\|F\| = \sup_x |F(x)|$ stands for the uniform norm. Furthermore, under the stability properties (1.3.13), if we set

$$F_\infty(x) = \eta_\infty\left(1_{(-\infty,x]}\right)$$

then, the probability of the following event

$$\left\|F_n^N - F_\infty\right\| \leq \frac{c}{\sqrt{N}}\ \sqrt{d\ (x+1)} + c_3\ e^{-\delta n}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some universal constant $c < \infty$ that does not depend on the dimension.

Particle free energy models

Mimicking the multiplicative formula (1.3.8), we set

$$\mathcal{Z}_n^N = \prod_{0 \leq p < n} \eta_p^N(G_p) \quad \text{and} \quad \gamma_n^N(dx) = \mathcal{Z}_n^N \times \eta_n^N(dx). \tag{1.3.14}$$

We have already mentioned that these rather complex particle models provide an unbiased estimate of the unnormalized measures. That is, we have that

$$\mathbb{E}\left(\eta_n^N(f_n) \prod_{0 \leq p < n} \eta_p^N(G_p)\right) = \mathbb{E}\left(f_n(X_n) \prod_{0 \leq p < n} G_p(X_p)\right). \tag{1.3.15}$$

The concentration properties of the *unbiased* particle free energies $\mathcal{Z}_n^N$ around their limiting values $\mathcal{Z}_n$ are developed in [51]. Our results will basically be stated as follows.

For any $N \geq 1$, and any $\epsilon \in \{+1, -1\}$, the probability of each of the following events

$$\frac{\epsilon}{n} \log \frac{\mathcal{Z}_n^N}{\mathcal{Z}_n} \leq \frac{c_1}{N}\ \left(1 + x + \sqrt{x}\right) + \frac{c_2}{\sqrt{N}}\ \sqrt{x}$$

is greater than $1 - e^{-x}$.

Genealogical tree model

The occupation measure of the N-genealogical tree model represented by the lines linking the solid circles converges as $N \to \infty$ to the distribution $\mathbb{Q}_n$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{(\xi^i_{0,n},\xi^i_{1,n},\ldots,\xi^i_{n,n})} = \mathbb{Q}_n. \tag{1.3.16}$$

In this notation, we have, for any $n \geq 0$, any bounded function $\mathbf{f}_n$ on the path space $\mathbf{E}_n$, such that (s.t.) $\|\mathbf{f}_n\| \leq 1$, and any $N \geq 1$, the probability of each of the following events

$$\left[\tfrac{1}{N}\sum_{i=1}^{N} \mathbf{f}_n(\xi^i_{0,n}, \xi^i_{1,n}, \ldots, \xi^i_{n,n}) - \mathbb{Q}_n(\mathbf{f}_n)\right]$$

$$\leq c_1 \ \frac{n+1}{N} \ \left(1 + x + \sqrt{x}\right) + c_2 \ \sqrt{\frac{(n+1)}{N}} \ \sqrt{x}$$

is greater than $1 - e^{-x}$.

The concentration properties of genealogical tree occupation measures can be derived more or less directly from those of the last population models. This rather surprising assertion comes from the fact that the n-th time marginal η_n of a Feynman-Kac measure associated with a reference historical Markov process has the same form as in the measure (1.3.3).

Using these properties, we prove concentration properties for interacting empirical processes associated with genealogical tree models. Our concentration inequalities will basically be stated as follows. We let $\mathcal{F}_n$ be the set of product functions of cell indicators in the path space $\mathbf{E}_n = \left(\mathbb{R}^{d_0} \times \ldots, \times \mathbb{R}^{d_n}\right)$, for some $d_p \geq 1$, $p \geq 0$. We also denote by η^N_n the occupation measure of the genealogical tree model. In this notation, the probability of the following event

$$\sup_{\mathbf{f}_n \in \mathcal{F}_n} \left|\eta^N_n(\mathbf{f}_n) - \mathbb{Q}_n(\mathbf{f}_n)\right| \leq c \ (n+1) \ \sqrt{\frac{\sum_{0\leq p\leq n} d_p}{N} \ (x+1)}$$

is greater than $1 - e^{-x}$, for any $x \geq 0$, for some universal constant $c < \infty$ that does not depend on the dimension.

Complete genealogical tree models

Mimicking the backward model (1.3.10) and the above formulae, we set

$$\Gamma^N_n = \mathcal{Z}^N_n \times \mathbb{Q}^N_n \tag{1.3.17}$$

with

$$\mathbb{Q}^N_n(d(x_0, \ldots, x_n)) = \eta^N_n(dx_n) \ \prod_{q=1}^{n} \mathbb{M}_{q,\eta^N_{q-1}}(x_q, dx_{q-1}).$$

Notice that the computation of sums w.r.t. these particle measures are reduced to summations over the particle locations ξ_n^i. It is therefore natural to identify a population of individuals $(\xi_n^1, \ldots, \xi_n^N)$ at time n to the ordered set of indexes $\{1, \ldots, N\}$. In this case, the occupation measures and the functions are identified with the following line and column vectors

$$\eta_n^N := \left[\frac{1}{N}, \ldots, \frac{1}{N}\right] \quad \text{and} \quad \mathrm{f}_n := \begin{pmatrix} f_n(\xi_n^1) \\ \vdots \\ f_n(\xi_n^N) \end{pmatrix}$$

and the matrices $\mathbb{M}_{n,\eta_{n-1}^N}$ by the $(N \times N)$ matrices

$$\mathbb{M}_{n,\eta_{n-1}^N} := \begin{pmatrix} \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^1, \xi_{n-1}^1) & \cdots & \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^1, \xi_{n-1}^N) \\ \vdots & \vdots & \vdots \\ \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^N, \xi_{n-1}^1) & \cdots & \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^N, \xi_{n-1}^N) \end{pmatrix} \tag{1.3.18}$$

with the (i, j)-entries

$$\mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^i, \xi_{n-1}^j) = \frac{G_{n-1}(\xi_{n-1}^j) H_n(\xi_{n-1}^j, \xi_n^i)}{\sum_{k=1}^N G_{n-1}(\xi_{n-1}^k) H_n(\xi_{n-1}^k, \xi_n^i)}.$$

For instance, the $\mathbb{Q}_n$-integration of normalized additive linear functionals of the form

$$\mathbf{f}_n(x_0, \ldots, x_n) = \frac{1}{n+1} \sum_{0 \le p \le n} f_p(x_p) \tag{1.3.19}$$

is given the particle matrix approximation model

$$\mathbb{Q}_n^N(\mathbf{f}_n) = \frac{1}{n+1} \sum_{0 \le p \le n} \eta_n^N \mathbb{M}_{n,\eta_{n-1}^N} \mathbb{M}_{n-1,\eta_{n-2}^N} \ldots \mathbb{M}_{p+1,\eta_p^N}(f_p).$$

These type of additive functionals arise in the calculation of the sensitivity measures. Its application in options' sensitivity computation is discussed in section 2.5.

For any $n \ge 0$, any normalized additive functional of the form (1.3.19), with $\max_{0 \le p \le n} \|f_p\| \le 1$, and any $N \ge 1$, the probability of each of the following events

$$\left[\mathbb{Q}_n^N - \mathbb{Q}_n\right](\bar{\mathbf{f}}_n) \le c_1 \, \frac{1}{N} \, (1 + (x + \sqrt{x})) + c_2 \, \sqrt{\frac{x}{N(n+1)}}$$

is greater than $1 - e^{-x}$.

For any $a = (a_i)_{1 \le i \le d} \in E_n = \mathbb{R}^d$, we denote by C_a the cell

$$C_a := (-\infty, a] = \prod_{i=1}^d (-\infty, a_i]$$

and $\mathbf{f}_{a,n}$ the additive functional

$$\mathbf{f}_{a,n}(x_0,\ldots,x_n) = \frac{1}{n+1}\sum_{0\leq p\leq n} 1_{(-\infty,a]}(x_p).$$

The probability of the following event

$$\sup_{a\in\mathbb{R}^d}\left|\mathbb{Q}_n^N(\mathbf{f}_{a,n}) - \mathbb{Q}_n(\mathbf{f}_{a,n})\right| \leq c\,\sqrt{\frac{d}{N}(x+1)}$$

is greater than $1-e^{-x}$, for any $x\geq 0$, for some constant $c<\infty$ that does not depend on the dimension, nor on the time horizon.

1.3.4 Example of applications in finance

In chapters 3 and 4, the numerical methods we use for pricing of American options are based on the interacting particle system. Recently, these particle techniques also have been applied to some other areas of finance; to name a few: the pricing of complex path dependent European options, sensitivities computing, as well as the numerical solution of partially observed control and estimation problems. In chapter 2, we present these applications with some detailed analysis. Here, we take the option sensitivity computation as an example to illustrate the application of particle methods.

We let $\theta\in\mathbb{R}^d$ be a parameter that may represent the volatility of some asset price movements, or any other kinetic parameter. We assume that the evolution of the risky asset price $S_k^{(\theta)}$ associated to some value of the parameter θ, is given by a one-step probability transition of the form

$$M_k^{(\theta)}(s,ds') := \text{Proba}\left(S_k^{(\theta)}\in ds' | S_{k-1}^{(\theta)}=s\right) = H_k^{(\theta)}(s,s')\ \lambda_k(ds')\ ,$$

for some positive density functions $H_k^{(\theta)}(s,s')$ and some reference measure λ_k. We also consider a collection of functions $G_k^{(\theta)}(s) = e^{-r_k^{(\theta)}(s)}$ that depend on θ. We also assume that the gradient and the Hessian of the logarithms of these functions with respect to the parameter θ are well defined.

In this situation, we denote a general form of the option price by

$$\Gamma_n^\theta(F_n) := \mathbb{E}\left(F_n(S_0^{(\theta)},\ldots,S_n^{(\theta)})\prod_{0\leq p<n} G_p^{(\theta)}\left(S_p^{(\theta)}\right)\right)\ . \tag{1.3.20}$$

Simple derivations show that the first and second order derivatives of the option value with respect to θ is given by

$$\begin{aligned}\nabla\Gamma_n^{(\theta)}(F_n) &= \Gamma_n^{(\theta)}(F_n\Lambda_n^{(\theta)})\\ \nabla^2\Gamma_n^{(\theta)}(F_n) &= \Gamma_n^{(\theta)}\left[F_n(\nabla\mathbb{L}_n^{(\theta)})'(\nabla\mathbb{L}_n^{(\theta)}) + F_n\nabla^2\mathbb{L}_n^{(\theta)}\right]\end{aligned}$$

with $\Lambda_n^{(\theta)} := \nabla \mathbb{L}_n^{(\theta)}$ and

$$\mathbb{L}_n^{(\theta)}(x_0, \ldots, x_n) := \sum_{p=1}^{n} \log \left(G_{p-1}^{(\theta)}(s_{p-1}) H_p^{(\theta)}(s_{p-1}, s_p) \right) .$$

These quantities can be approximated by the unbiased particle models

$$\begin{aligned} \nabla_N \Gamma_n^{(\theta)}(F_n) &:= \Gamma_n^{(\theta,N)}(F_n \Lambda_n^{(\theta)}) \\ \nabla_N^2 \Gamma_n^{(\theta)}(F_n) &= \Gamma_n^{(\theta,N)} \left[F_n (\nabla \mathbb{L}_n^{(\theta)})' (\nabla \mathbb{L}_n^{(\theta)}) + F_n \nabla^2 \mathbb{L}_n^{(\theta)} \right] . \end{aligned}$$

For each value of the parameter θ, we denote by $\Gamma_n^{(\theta,N)}$ the N-particle approximation measures associated with a given value of the parameter θ and defined in (1.3.17). The analysis for the approximation error of $\nabla \Gamma_n^{(\theta)}(F_n)$ would be a direct consequence of the concentration inequalities stated in the previous subsection.

1.4 More sophisticated American options

After the study of the Feynman-Kac measures, it is important to remark that the particle method we propose in section 1.2 can be easily adapted to pricing of some sophisticated American options. Instead of the uniform selection, we can design some Feynman-Kac measures based on certain potential information $(G_k)_{0 \leq k \leq n}$. One direct application in mind is the pricing of *deep Out of The Money (OTM) options*. In this situation, the payoff function is localized in a small region of the space. As in classic rare events problem, standard Monte Carlo simulations usually fail, and the main variance reduction technique is importance sampling. But in general, desirable changes of measure favoring sample paths realizing rare events are highly unlikely to lead to explicit formula. In this case importance sampling is no longer an option. A natural alternative is then interacting particles methods. Based on different choices of the potential function of $(G_k)_{0 \leq k \leq n}$, the particle techniques can usually send the simulation samples to desirable regions to achieve the variance reduction.

With this technique, we can consider even more complicate situation: the maximization of $\mathbb{E}(f_\tau(X_\tau) \prod_{k=0}^{\tau-1} B_k(X_k))$ for a given class of functions $(B_k)_{0 \leq k \leq n}$ modeling an obstacle. For instance in the case of *barrier options*, $(B_k)_{0 \leq k \leq n}$ take the form of indicator functions. In chapter 4, we can prove that the computations of deep OTM options and of barrier options are mathematically equivalent in our algorithm if we take $B_k = G_k$ in above display. In this context, in the further development of this section we will unify the notation and focus on the analysis of the Snell envelope for $\sup_\tau \mathbb{E}(f_\tau(X_\tau) \prod_{k=0}^{\tau-1} G_k(X_k))$.

1.4.1 Snell envelope with small probability event criteria

In chapter 4, we can see the Snell envelope for such options will be reduced to :

$$\left\{ \begin{array}{rcl} v_n(x_n) & = & f_n(x_n) \\ v_k(x_k) & = & f_k(x_k) \vee \left[G_k(x_k) M_{k+1}(v_{k+1})(x_k) \right] \ , \ \forall \ 0 \leq k \leq n-1 \ . \end{array} \right. \tag{1.4.1}$$

The above recursion implies that it is not relevant to compute precisely the conditional expectation $M_{k+1}(v_{k+1})(x_k)$ when the value of the criteria $G_k(x_k)$ is zero or very small, or when the gain function f_k is zero or very small. Hence from a variance reduction point of view, when approximating the conditional expectation $M_{k+1}(v_{k+1})(x_k)$ by a Monte Carlo method, it seems relevant to concentrate the simulations in the regions of E_{k+1} where G_{k+1} and/or f_{k+1} reach high values. Hence, to avoid the potential rare events G, we consider a change of measure on the measurable product space $(E_0 \times \cdots \times E_n, \mathcal{E}_0 \times \cdots \times \mathcal{E}_n)$, with the following form

$$d\mathbb{Q}_n = \frac{1}{Z_n} \left[\prod_{k=0}^{n-1} G_k(X_k) \right] d\mathbb{P}_n \ , \quad \text{with} \quad Z_n = \mathbb{E} \left(\prod_{k=0}^{n-1} G_k(X_k) \right) = \prod_{k=0}^{n-1} \eta_k(G_k) \ , \tag{1.4.2}$$

where η_k is the probability measure defined on E_k such that, for any measurable function f on E_k

$$\eta_k(f) := \frac{\mathbb{E}\Big(f(X_k) \prod_{p=0}^{k-1} G_p(X_p) \Big)}{\mathbb{E}\Big(\prod_{p=0}^{k-1} G_p(X_p) \Big)} \ .$$

The measures $(\eta_k)_{0 \leq k \leq n}$ defined above can be seen as the laws of random states $(\bar{X}_k)_{0 \leq k \leq n}$ under the probability measures $(\mathbb{Q}_k)_{0 \leq k \leq n}$. More interestingly, in Section 4.4 we will see that the sequence of random states $(\bar{X}_k)_{0 \leq k \leq n}$ forms a nonlinear Markov chain with transitions $\bar{X}_k \rightsquigarrow \bar{X}_{k+1}$ that depends on the current distribution η_k, at time k. The behavior of this chain is dictated by the potential functions $(G_k)_{0 \leq k \leq n}$ and the Markov transitions $(M_k)_{\leq k \leq n}$ of the reference process $(X_k)_{0 \leq k \leq n}$. Regions with high G_k-values are visited more likely.

To illustrate this remark, we examine the situation where $G_k(x_k) := 1_{A_k}(x_k)$ with $A_k \subset E_k$. In this situation, $law(X_k | X_p \in A_p, \ p < k) = law(\bar{X}_k) = \eta_k$ is the conditional distribution of X_k given the fact that $X_p \in A_p$, for any $p < k$. In this special case, the process $(\bar{X}_k)_{0 \leq k \leq n}$ is restricted to regions related to the choice of the sequence $(A_k)_{0 \leq k \leq n}$. This change of measure is known as the optimal twisted measure for sampling a Markov chain restricted to the subset regions A_k. More general change of measure are addressed in section 4.6 of chapter 4. These models are direct extension of 1.4.2 to potential functions that depend on the transition of the reference Markov chain.

Furthermore, it is also important to observe that, for any measurable function f on E_k

$$\eta_k(f) = \frac{\eta_{k-1}(G_{k-1} M_k(f))}{\eta_{k-1}(G_{k-1})} \ . \tag{1.4.3}$$

We denote the recursive relation between η_k and η_{k-1} by introducing the operators Φ_k such that, for all $1 \leq k \leq n$

$$\eta_k = \Phi_k(\eta_{k-1}) \ . \tag{1.4.4}$$

Let us now introduce the integral operator Q_k such that, for all $1 \leq k \leq n$

$$Q_k(f)(x_{k-1}) := \int G_{k-1}(x_{k-1}) M_k(x_{k-1}, dx_k) f(x_k) \ . \tag{1.4.5}$$

By definition, we can rewrite the recursion 1.4.1 as:

$$v_k(x_k) = f_k(x_k) \vee Q_{k+1}(v_{k+1})(x_k) = f_k(x_k) \vee \Phi_{k+1}(\eta) \left(\frac{dQ_{k+1}(x_k, \cdot)}{d\Phi_{k+1}(\eta)} v_{k+1} \right) \ ,$$

for any measure η on E_k.

To extend the discussion in previous sections on the particle models, we state here the analysis in chapter 4. To construct a particle method to sample the random variables according to above distributions, we remark by definition (1.4.4) that:

$$\Phi_k(\eta_{k-1}) = \eta_{k-1} K_{k,\eta_{k-1}} = \eta_{k-1} S_{k-1,\eta_{k-1}} M_k = \Psi_{G_{k-1}}(\eta_{k-1}) M_k \ . \tag{1.4.6}$$

Where $K_{k,\eta_{k-1}}$, $S_{k-1,\eta_{k-1}}$ and $\Psi_{G_{k-1}}$ are defined as follows:

$$\begin{cases} K_{k,\eta_{k-1}}(x_{k-1}, dx_k) &= (S_{k-1,\eta_{k-1}} M_k)(x_{k-1}, dx_k) \\ &= \int S_{k-1,\eta_{k-1}}(x_{k-1}, dx'_{k-1}) M_k(x'_{k-1}, dx_k) \ , \\ S_{k-1,\eta_{k-1}}(x, dx') &= \epsilon G_{k-1}(x) \delta_x(dx') + (1 - \epsilon G_{k-1}(x)) \Psi_{G_{k-1}}(\eta_{k-1})(dx') \\ \Psi_{G_{k-1}}(\eta_{k-1})(dx) &= \frac{G_{k-1}(x)}{\eta_{k-1}(G_{k-1})} \eta_{k-1}(dx) \ , \end{cases}$$

where the real ϵ is such that ϵG takes its values $[0, 1]$.

More generally, the operations Ψ and S can be expressed as $\Psi_G(\eta)(f) = \frac{\eta(Gf)}{\eta(G)} = \eta S_\eta(f)$ with $S_\eta(f) = \epsilon G f + (1 - \epsilon G) \Psi_G(\eta)(f)$. We recall from [40] that $\eta_k = law(\bar{X}_k)$, where $\bar{X}_{k-1} \rightsquigarrow \bar{X}_k$ is a Markov chain with transitions $K_{k,\eta_{k-1}}$ defined above.

The particle approximation provided in the present chapter is defined in terms of a Markov chain $\xi_k^{(N)} = (\xi_k^{(i,N)})_{1 \leq i \leq N}$ on the product state spaces E_k^N, where the given integer N is the number of particles sampled in every instant. The initial particle system, $\xi_0^{(N)} = \left(\xi_0^{(i,N)} \right)_{1 \leq i \leq N}$, is a collection of N i.i.d. random copies of X_0. We let $\mathcal{F}_k^N$ be the sigma-field generated by the particle approximation model from the origin, up to time k. To simplify the presentation, when there is no confusion we suppress the population size parameter N, and we write ξ_k and ξ_k^i instead of $\xi_k^{(N)}$

and $\xi_k^{(i,N)}$. By construction, ξ_k is a particle model with a selection transition and a mutation type exploration i.e. the evolution from ξ_k to ξ_{k+1} is composed by two steps:

$$\xi_k \in E_k^N \xrightarrow[S_{k,\eta_k^N}]{Selection} \widehat{\xi}_k := \left(\widehat{\xi}_k^i\right)_{1\leq i\leq N} \in E_k^N \xrightarrow[M_{k+1}]{Mutation} \xi_{k+1} \in E_{k+1}^N \ . \tag{1.4.7}$$

Then we define $\widehat{\eta}_k^N$ as the occupation measure after the mutation steps. More precisely,

$$\widehat{\eta}_k^N := \frac{1}{N}\sum_{1\leq i\leq N} \delta_{\widehat{\xi}_k^i} \ .$$

During the selection transition S_{k,η_k^N}, for $0 \leq i \leq N$ with a probability $\epsilon G_k(\xi_k^i)$ we decide to skip the selection step i.e. we leave $\widehat{\xi}_k^i$ stay on particle ξ_k^i, and with probability $1-\epsilon G_k(\xi_k^i)$ we decide to do the following selection: $\widehat{\xi}_k^i$ randomly takes the value in ξ_k^j for $0 \leq j \leq N$ with distribution $\frac{G_k(\xi_k^j)}{\sum_{l=1}^N G_k(\xi_k^l)}$. Note that when $\epsilon G_k \equiv 1$, the selection is skipped (i.e. $\widehat{\xi}_k = \xi_k$) so that the model corresponds exactly to the Broadie-Glasserman type model analysed by P. Del Moral and P. Hu et al. [50]. Hence, the factor ϵ can be interpreted as a level of selection against the rare events. During the mutation transition $\widehat{\xi}_k \rightsquigarrow \xi_{k+1}$, every selected individual $\widehat{\xi}_k^i$ evolves randomly to a new individual $\xi_{k+1}^i = x$ randomly chosen with the distribution $M_{k+1}(\widehat{\xi}_k^i, dx)$, for $1 \leq i \leq N$.

Now, we can construct the approximation scheme to estimate the Snell envelope $(v_k)_{0\leq k\leq n}$ with the following backward recursion:

$$\left\{\begin{array}{lcl} \widehat{v}_n & = & f_n \\ \widehat{v}_k(x_k) & = & f_k(x_k) \vee \eta_{k+1}^N\left(\frac{dQ_{k+1}(x_k,\cdot)}{d\Phi_{k+1}(\eta_k^N)}\widehat{v}_{k+1}\right) \quad \text{for all } 0 \leq k < n \ , \end{array}\right. \tag{1.4.8}$$

Following again the robust lemma we proposed in section 1.2, we can estimate the approximation error by the following theorem:

Theorem 1.4.1. *For any $0 \leq k \leq n$, any integer $p \geq 1$, we have*

$$\sup_{x\in E_k} \|(\widehat{v}_k - v_k)(x)\|_{L_p} \leq \frac{2\ a(p)}{\sqrt{N}} \sum_{k<l<n} q_{k,l}\ \left[Q_{k,l+1}(h_{l+1}^{p'-1} v_{l+1}^{p'})(x)\right]^{\frac{1}{p'}} \ ,$$

where p' is the smallest even integer greater than p, we refer the reader to 4.5 for more details on other coefficients in above display.

1.4.2 Numerical results

In the end of chapter 4, some numerical examples are provided to test this new algorithm. Those are very good examples to have an idea of the choice of potential

functions. In further development, we name our algorithm *Stochastic Mesh with Change of Measure (**SMCM**)* and compare the results with the standard *Stochastic Mesh (**SM**)* algorithm without change of measure.

In these numerical tests we have considered a simple Black-Scholes price model. However, notice that both algorithms (*SM* and *SMCM*) can be applied in a general Markovian framework. The asset prices are modeled by a d-dimensional Markov process (S_t) such that each component (i.e. each asset) follows a geometric Brownian motion under the risk-neutral measure, that is, for assets $i = 1, \cdots, d$,

$$dS_t(i) = S_t(i)(rdt + \sigma dz_t^i) \ , \tag{1.4.9}$$

where z^i, for $i = 1, \cdots, d$ are independent one dimensional standard Brownian motions. Unless otherwise specified, the interest rate r is set to 10% annually and the volatility is supposed to be the same for all assets, $\sigma = 20\%$ annually. The starting prices of the assets are for all $i = 1, \cdots, d$, $S_{t_0}(i) = 1$. We consider two types of Bermudan options with maturity $T = 1$ year and 11 equally distributed exercise opportunities at dates $t_k = kT/n$ with $k = 0, 1, \cdots, n = 10$, associated with two different payoffs:

1. Geometric average put option with payoff $(K - \prod_{i=1}^{d} S_T(i))_+$,
2. Arithmetic average put option with payoff $(K - \frac{1}{d}\sum_{i=1}^{d} S_T(i))_+$,

The benchmark values of these options are reported on Table 1.3.

Strike	$K = 0.95$	$K = 0.85$	$K = 0.75$
Option value	0.0279	0.0081	0.0015

Table 1.3: Benchmark values for the geometric put option obtained by using the Stochastic Mesh method with 10000 particles. $n = 11$ exercise opportunities, $T = 1$, $S_0 = 1$ and $r = 10\%/d$, $\sigma_i = 20\%/\sqrt{d}$ for the geometric payoff and $r = 10\%$, $\sigma_i = 20\%$ for the arithmetic payoff.

We consider the Markov chain $(X_k)_{0\leq k\leq n}$, taking values on $E_k = \mathbb{R}^{+d}$, obtained by discretization of the time-continuous process S defined by (4.7.1) at times of exercise opportunities, $0 = t_0 < \cdots < t_n = T$, such that for all $k = 0, \cdots, n$, $X_k = S_{t_k}$.
Now, we can introduce the sequence of positive functions $(G_k)_{1\leq k\leq n}$, defining the change of measure (4.3.1), as follows:

$$\begin{cases} G_0(x_1) = (f_1(x_1) \vee \varepsilon)^\alpha \ , \\ \\ G_k(x_k, x_{k+1}) = \frac{(f_{k+1}(x_{k+1})\vee\varepsilon)^\alpha}{(f_k(x_k)\vee\varepsilon)^\alpha} \ , \quad \text{for all } k = 1, \cdots, n-1 \ , \end{cases} \tag{1.4.10}$$

where f_k are the payoff functions and $\alpha \in (0, 1]$ and $\varepsilon > 0$ are parameters fixed in our simulations to the values $\alpha = 1/5$ and $\varepsilon = 10^{-7}$.

For each example, we have performed the algorithm for different numbers of mesh points $N = 100\,,\ 200\,,\ 400\,,\ 800\,,\ 1600\,,\ 3200\,,\ 6400$. 1000 runs of both algorithms (*Stochastic Mesh (SM)* and *Stochastic Mesh with Change of Measure (SMCM)*) were performed to compute the mean and confidence intervals of each estimate.

Simulation results are reported in Figure 1.3, 1.4 and 1.5 for the geometric and arithmetic put payoff, with strikes corresponding to standard out of the money puts to deep out of the money puts: $K = 0.95$, $K = 0.85$ and $K = 0.75$.

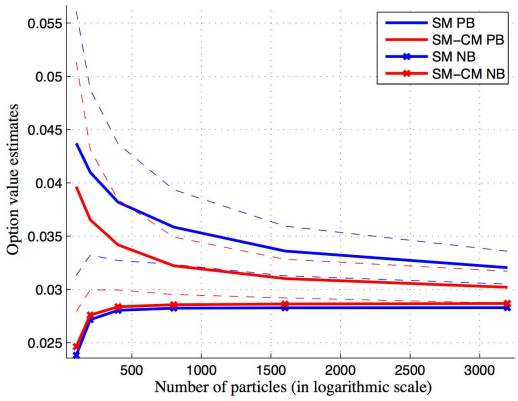
Notice that both algorithms (the Stochastic Mesh algorithm with and without Change of Measure) have been implemented without any standard variance reduction technique (control variate, stratification, ...). In term of complexity, the Stochastic Mesh algorithm with Change of Measure is equivalent to the standard Stochastic Mesh algorithm: the complexity is in both cases quadratic with the number of mesh points $O(N^2)$ since the number of operations required to operate the change of measure is negligible.

We have reported on our graphs two types of estimates:

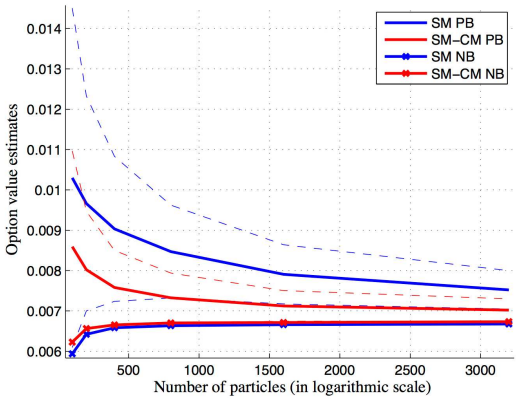
- the *Positively-biased estimator* provided by the backward induction on the value function;

- the *Negatively-biased estimator* provided by the associated optimal exercise policy. This estimate is obtained via a two-step procedure: first, the optimal policy is approximated in the backward induction on the value function, then the policy is evaluated using the standard forward Monte Carlo procedure. Note that the resulting estimator is known to provide a lower bound (in average) to the option price. In our simulation, we have used $N_{forward} = 10000$ Monte Carlo forward simulations.

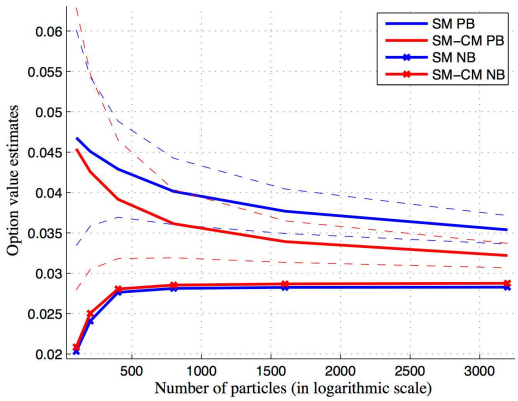
As expected, one can observe on Table 4.2, that the *SMCM* algorithm allows to obtain an estimate, $\hat{v}_{SMCM}$, with the same complexity but with a smaller variance than the standard *SM* algorithm estimate, $\hat{v}_{SM}$, especially for deep out the money options.
More surprisingly, one can observe on Table 4.2 and Figure 1.3, 1.4 and 1.5 that the *SMCM* algorithm also allows to reduce significantly the estimator bias which is known to compose the growing part of the error when the number of underlying assets increases. For instance, one can notice that the *SMCM* algorithm achieves the convergence in average of the Positively-biased estimate to the Negatively-biased estimate for a number of mesh points much smaller than for the *SM* algorithm. Hence, the *SMCM* could also be a way to deal with high dimensional optimal stop-



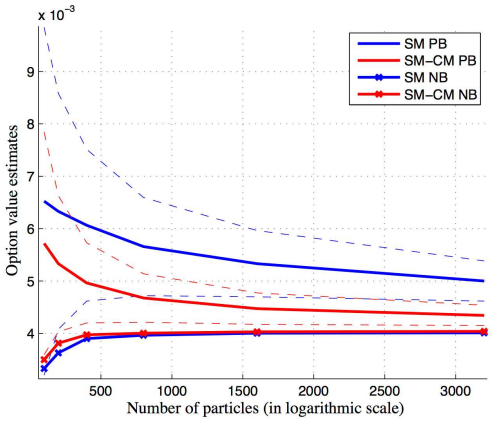
(a) Geometric Put with $d = 3$ assets



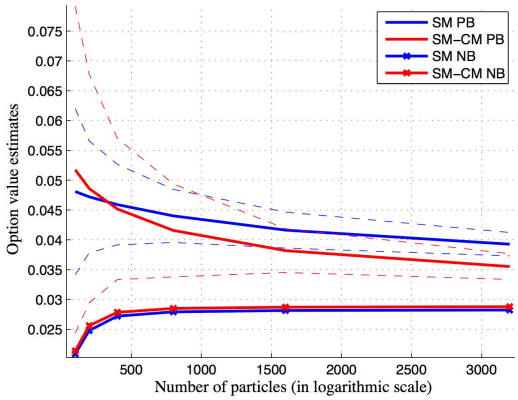
(b) Arithmetic Put with $d = 3$ assets



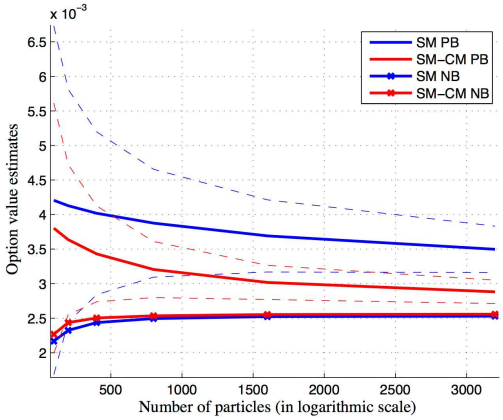
(c) Geometric Put with $d = 4$ assets



(d) Arithmetic Put with $d = 4$ assets

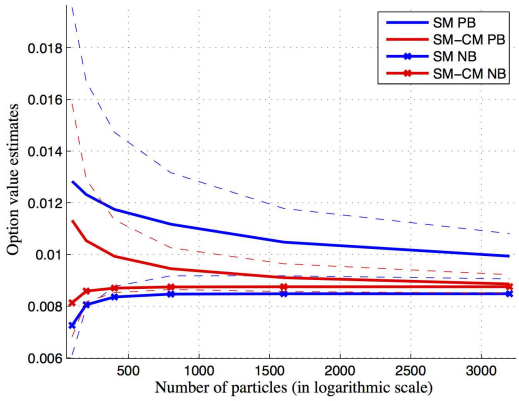


(e) Geometric Put with $d = 5$ assets

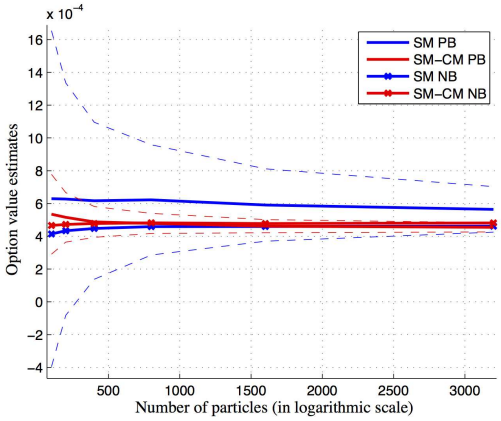


(f) Arithmetic Put with $d = 5$ assets

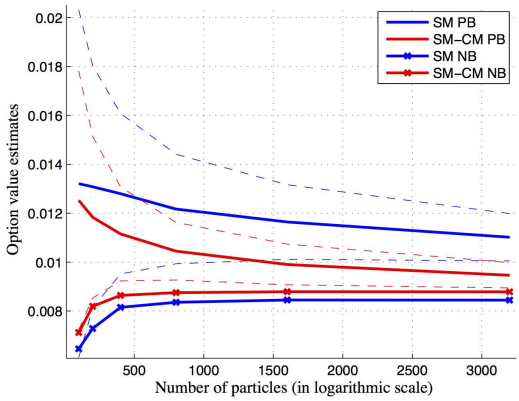
Figure 1.3: Positively-biased option values estimates (average estimates with 95% confidence interval computed over 1000 runs) and Negatively-biased option values estimates (average estimates over the 1000 runs each forward estimate being evaluated over 10000 forward Monte Carlo simulations), computed by the *SM* algorithm (in blue line) and the *SMCM* algorithm (in red line), as a function of the number of mesh points for geometric (on the left column) and arithmetic (on the right column) put options with **strike** $K = 0.95$.



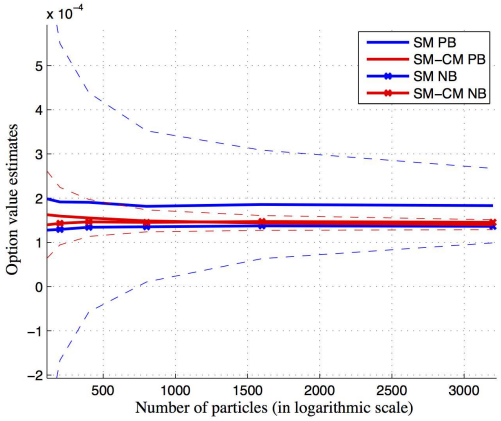
(a) Geometric Put with $d = 3$ assets



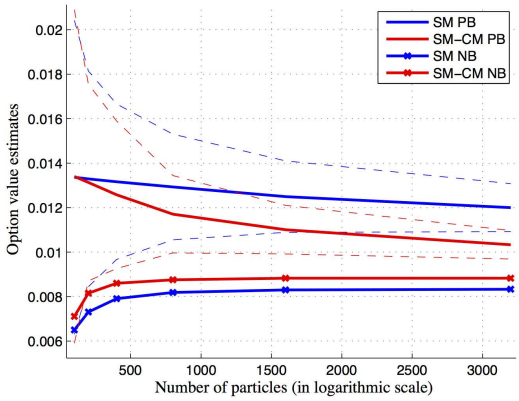
(b) Arithmetic Put with $d = 3$ assets



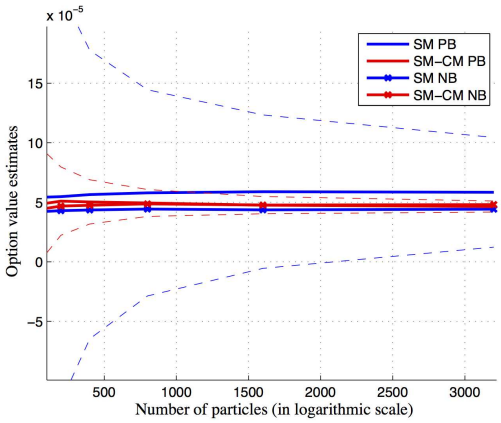
(c) Geometric Put with $d = 4$ assets



(d) Arithmetic Put with $d = 4$ assets

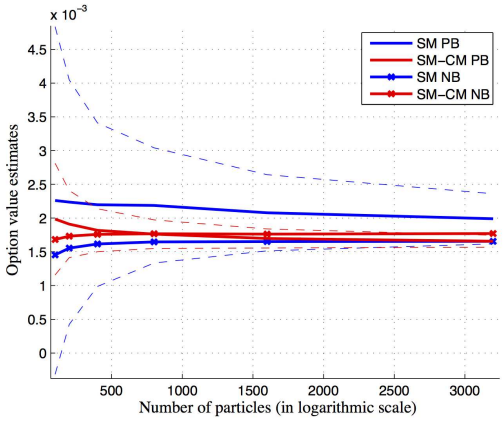


(e) Geometric Put with $d = 5$ assets

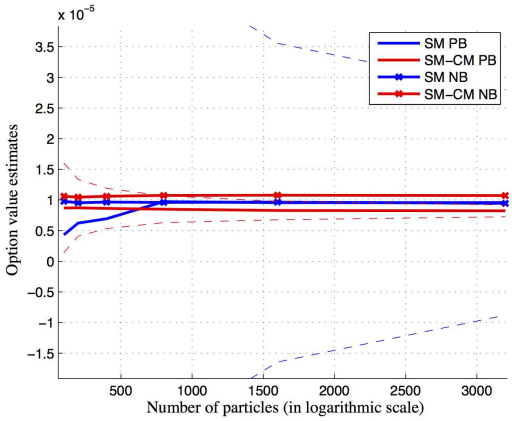


(f) Arithmetic Put with $d = 5$ assets

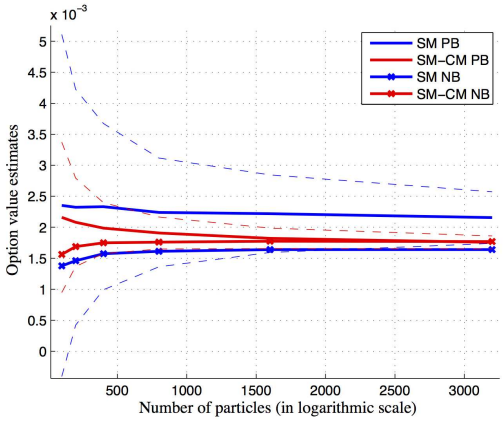
Figure 1.4: Positively-biased option values estimates (average estimates with 95% confidence interval computed over 1000 runs) and Negatively-biased option values estimates (average estimates over the 1000 runs each forward estimate being evaluated over 10000 forward Monte Carlo simulations), computed by the *SM* algorithm (in blue line) and the *SMCM* algorithm (in red line), as a function of the number of mesh points for geometric (on the left column) and arithmetic (on the right column) put options with **strike** $K = 0.85$.



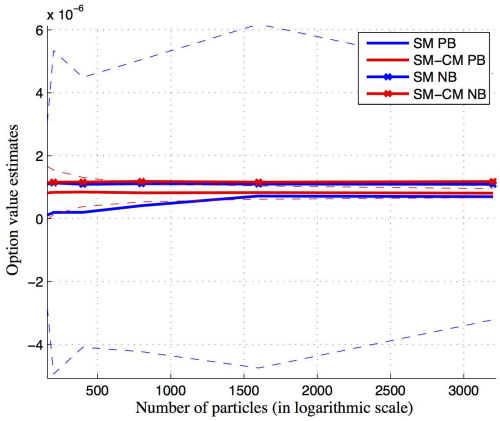
(a) Geometric Put with $d = 3$ assets



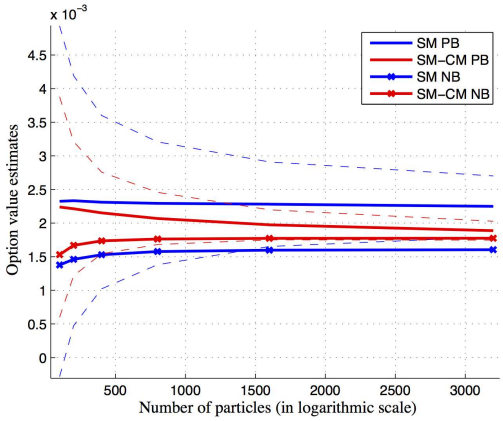
(b) Arithmetic Put with $d = 3$ assets



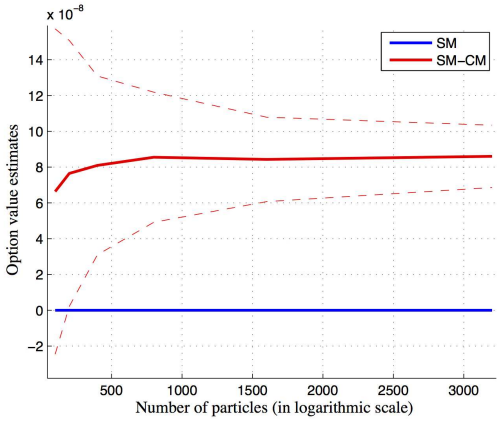
(c) Geometric Put with $d = 4$ assets



(d) Arithmetic Put with $d = 4$ assets



(e) Geometric Put with $d = 5$ assets



(f) Arithmetic Put with $d = 5$ assets

Figure 1.5: Positively-biased option values estimates (average estimates with 95% confidence interval computed over 1000 runs) and Negatively-biased option values estimates (average estimates over the 1000 runs each forward estimate being evaluated over 10000 forward Monte Carlo simulations), computed by the *SM* algorithm (in blue line) and the *SMCM* algorithm (in red line), as a function of the number of mesh points for geometric (on the left column) and arithmetic (on the right column) put options with **strike** $K = 0.75$. (For the clarity of the graph (f), the Negatively-biased estimate is not reported, the associated variance (for 10 000 forward Monte Carlo simulations) being relatively strong).

ping problems since the algorithm complexity remains insensitive to the dimension whereas the convergence rate is not significantly reduced.

Payoff	K	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
Geometric	0.95	1 (1%)	1 (3%)	1 (6%)	1 (9%)	1 (10%)
Put	0.85	5 (2%)	8 (6%)	6 (11%)	4 (14%)	3 (14%)
	0.75	18 (6%)	28 (11%)	18 (17%)	16 (18%)	11 (16%)
Arithmetic	0.95	1 (1%)	3 (2%)	3 (7%)	4 (13%)	5 (18%)
Put	0.85	5 (2%)	13 (6%)	24 (19%)	56 (24%)	100 (20%)
	0.75	18 (6%)	71 (15%)	363 (14%)	866 (16%)	− (−)

Table 1.4: Variance ratio ($\frac{Var(\hat{v}_{SM})}{Var(\hat{v}_{SMCM})}$) and Bias ratio ($\frac{\mathbb{E}(\hat{v}_{SM})-\mathbb{E}(\hat{v}_{SMCM})}{E(\hat{v}_{SM})}$) (within parentheses) computed over 1000 runs for $N = 3200$ mesh points. (For the arithmetic put, when $d = 5$ and $K = 0.75$, the 1000 estimates provided by the standard SM algorithm were all equal to zero, hence the associated variance ratio has not been reported).

1.5 Organization of the thesis and contributions

This thesis is organised four chapters.

Chapter 1 provides an overview of the thesis, outlines the motivation and summarizes the major contributions.

Chapter 2 is aimed to give a general introduction to the theory of interacting particle methods, and an overview of its applications to computational finance. This result has been published as a chapter in *Numerical methods in Finance* of Springer Proceedings in Mathematics.

- Sections 2.1-2.3 survey the main techniques and results on interacting particle systems.
- Sections 2.4-2.7 explain how these techniques can be applied to the numerical solution of a variety of financial applications such as pricing complex path dependent European options, computing sensitivities, pricing American options or numerically solving partially observed control and estimation problems.

Chapter 3 presents the optimal control problem in the pricing of American options, and analyzes the robustness properties of the Snell envelope equations. This chapter provides a general framework to analyse different approximation methods. This result has been published as a journal article in *SIAM Journal on Financial Mathematics*.

- Sections 3.3 and 3.4 consider a series of approximation schemes, including cut-off type approximations, Euler discretization schemes, interpolation methods, quantization tree methods, and the Stochastic Mesh method of Broadie-Glasserman. In each situation, we provide non asymptotic convergence estimates, including $\mathbb{L}_p$-mean error bounds and exponential concentration inequalities. We deduce these estimates from a single and general robustness property of Snell envelope semigroups. In particular, this analysis allows us to recover existing convergence results for the quantization tree method and to improve significantly the rates of convergence obtained for the Stochastic Mesh estimator of Broadie-Glasserman.
- Section 3.5 provides a new approach based on a genealogical tree approximation method of the reference Markov process in terms of a neutral type genetic model. In contrast to Broadie-Glasserman Monte Carlo method, the computational cost of this new stochastic approximation is linear in the number of random samples. Some simulation results are provided and confirm the interest of this new algorithm.

Chapter 4 analyzes and computes the Snell envelope in the specific case where the criterion to optimize is associated with a small probability or a rare event.

- Sections 4.1-4.3 present and analyze the Snell envelope with small probability criteria. We resolve the high dimensionality problem arisen in the small probability criteria.
- Sections 4.4-4.7 provide a new approach which combines the Stochastic Mesh approach of Broadie and Glasserman with a particle approximation scheme based on a specific change of measure designed to concentrate the computational effort in regions pointed out by the criteria. The theoretical analysis of this new algorithm provides non asymptotic convergence estimates. Finally, the numerical tests confirm the practical interest of this approach.

1.6 Publications

The following papers have been published or submitted based on works contained in this dissertation:

1. R. Carmona, P. Del Moral, P. Hu and N. Oudjane (eds.)
   Numerical Methods in Finance
   in *Springer Proceedings in Mathematics*, Vol 12 (2012).

2. R. Carmona, P. Del Moral, P. Hu and N. Oudjane
   An introduction to particle methods in finance
   in *Numerical Methods in Finance, Vol 12, pp. 3–50, Springer-Verlag* (2012).

3. P. Del Moral, P. HU and N. Oudjane
   Snell envelope with small probability criteria
   preprint inria-00507794 [submitted], 2010.

4. P. Del Moral, P. HU, N. Oudjane and B. Rémillard
   On the Robustness of the Snell Envelope
   in *SIAM J. Finan. Math., Vol. 2, pp. 587-626* (2011).

5. P. Del Moral, P. Hu and L. Wu
   On the concentration properties of Interacting particle processes
   in *Foundations and Trends in Machine Learning, vol. 3, nos. 3–4, pp. 225–389* (2012).

6. P. Hu and N. Oudjane
   Variance reduction techniques for thermal asset pricing
   Rapport consulting EDF R&D, 2012.

7. P. Del Moral, P. Hu and D. Weng
   Méthodes de Monte Carlo pour le pricing d'options américaines
   Lot 1. Comparaisons d'algorithmes. Contrat EDF OSIRIS-INRIA, 2010.

8. P. Del Moral and P. Hu
   Méthodes de Monte Carlo pour le pricing d'options américaines
   Lot 2. Confection et analyse de nouveaux algorithmes de Monte Carlo avancés. Contrat EDF OSIRIS-INRIA, 2010.

9. P. Del Moral and P. Hu
   Méthodes de Monte Carlo pour le pricing d'options américaines
   Lot 3. Confection et analyse de nouveaux algorithmes de Monte Carlo avancés. Contrat EDF OSIRIS-INRIA, 2010.

Chapter 2

A survey on Particle Methods with Financial Applications

2.1 Introduction

The growing field of Feynman-Kac expectations and related particle models is one of the most active contact points between probability theory and practical applications. The particle simulation techniques they suggest are also called sequential Monte Carlo methods in Bayesian statistics, and particle or genetic type filters in advanced signal processing. They are used to approximate a flow of probability measures with an increasing level of complexity. This class of probabilistic models includes conditional distributions of signals with respect to noisy and partial observations, non absorption probabilities in Feynman-Kac-Schrödinger models, Boltzmann-Gibbs measures, as well as conditional distributions of stochastic processes in critical regimes. For a thorough discussion on the application domains of interacting particle algorithms, we refer the reader to the first rigorous study of particle filters [39], the review article [55], the monograph [40], and the references therein.

Recently, these interacting particle techniques have been applied in several areas of finance. For instance, using the rare event interpretation of particle methods, R. Carmona, J. P. Fouque and D. Vestal proposed in [24] an interacting particle algorithm for the computation of the probabilities of simultaneous defaults in large credit portfolios. These developments for credit risk computation were then improved in the subsequent paper [21] by R. Carmona and S. Crépey, and by P. Del Moral and F. Patras in [57].

Following the particle filtering approach which is already widely used to estimate hidden Markov models, V. Genon-Catalot, T. Jeantheau and C. Laredo [73] in-

troduced particle methods for the estimation of stochastic volatility models. This approach has been applied for filtering nonlinear and non-Gaussian models by R. Casarin [27], R. Casarin and C. Trecroci [28], and V. Rossi and J. P. Vila[105]. More recently, M. S. Johannes, N. G. Polson and J.R. Stroud [79] used a similar approach to filter latent variables such as the jump times and sizes in jump diffusion price models. Particle techniques can also be used for stochastic optimization as demonstrated by S. Ben Hamida and R. Cont who provide in [10] a new calibration algorithm allowing for the existence of multiple global minima. Finally, in [50], interacting particle methods were used to estimate backward conditional expectation for American option pricing.

In this chapter, we survey the main ideas behind the *particle technology* alluded to above, with illustrations from recent applications in computational finance. We tried to provide a synthetic picture of particle solutions to some estimation problems arising in mathematical finance. We adopted an informal style of presentation, focusing on the ideas rather than on their detailed rigorous mathematical justification.

This chapter is organized as follows. In the following section, we highlight the natural link between option prices and Feynman-Kac formula. Then in the third section, the main principles and results related to particle methods are recalled. Finally, we dedicate the last sections of this chapter to the application of these particle techniques to some specific financial problems: credit risk analysis, sensitivity computation, American option pricing and control and estimation of partially observed models.

2.2 Option prices and Feynman-Kac formula

The numerical pricing of European-style options has been extensively studied in the mathematical finance literature. It would be foolish to try to cover this subject in the present chapter. We refer the reader to I. Karatzas and S.E. Shreve's book [80], and the more focused account by Y. Achdou and O. Pironneau [1] for a sample of texts relevant to the present discussion. European option pricing is a standard numerical problem in finance, well suited to our interpretation of option prices in terms of Feynman-Kac formula.

2.2.1 Discrete time models

We first consider discrete time models (often called multi-period models by economists). Option prices are often given by Feynman-Kac formulas of the form

$$Q_{p,n}(f_n)(X_p) := \mathbb{E}\left(f_n(X_n) \prod_{p \le q < n} G_q(X_q) \,|X_p \right) , \tag{2.2.1}$$

with the terminal condition $Q_{n,n}(f_n)(x) = f_n(X_n)$. Here p is the time at which the price is computed, and n is the time of maturity of the option, $f = (f_n)_n$ is a space-time function, i.e. a function of time n and a variable in the state space E_n at time n of the possible values of the underlying interest $X = (X_n)_n$, $f_n(x)$ giving the payoff of the option at time n if the underlying interest has the value x, i.e. $X_n = x$. We will assume throughout that $X = (X_n)_n$ is a Markov chain, and we will denote by $M = (M_{p,n})_{p,n}$ its transition probability

$$M_{p,n}(x, dy) := \mathbb{P}(X_n \in dy | X_p = x), \qquad x \in E_p. \tag{2.2.2}$$

We assume that the state space E_n of the chain at time n can change with n. We shall use the simpler notation M when the Markov chain is time homogeneous, in which case $M_{p,n} = M_{n-p} = M^{n-p}$, and the state space E does not change with time. The major ingredients in the above equation are the Markov chain X and the space-time potential $G = (G_n)_n$, given for each n, by a non-negative measurable function on E_n. The chain X is usually constructed from random factors evolving in time and price series $S^j = (S^j_n)_n$ which gives the time evolution of the risky asset prices. We give some simple example below. To conform with the terminology of the particle models used to understand theoretically and implement numerically the Feynman-Kac formulas of the above type, we shall sometimes call X_n a *particle* at time n. Depending on the application under consideration, the role of the potential G_n will be to capture the discounting necessary in the computation of the price, or some constraints (like barriers) present in the indenture of the option, or the risk premium in the form of a pricing kernel, or even to force the particle to visit some parts of the space-time domain where rare events occur, in which case its financial interpretation will not be possible.

We give some simple examples to illustrate the versatility of formula (2.2.1).

European barrier option.

In this case, we assume that there is only one underlying stock whose time evolution is given by a Markov chain $S = (S_n)_n$, that $X_n = S_n$ is the price of this underlying stock at time n, that f_n gives the payoff function if the maturity is n, and that $K > 0$ is the strike of the option. If we assume that stochastic interest rates are given by a non-negative space-time function $r = (r_n)_n$ of the chain, and if we denote by $A = (A_n)$ the sequence of barrier sets A_p, then the price of the barrier option is given by the Feynman-Kac formula (2.2.1) with

$$f_n(X_n) = (X_n - K)_+ \quad \text{and} \quad G_q(X_q) = \mathbb{1}_{A_q}(X_q)\ e^{-r_q(X_q)}\ . \tag{2.2.3}$$

Asian option

This example is important because it allows us to illustrate the use of the Feynman-Kac formula (2.2.1) when the chain X evolves on *path space*. Indeed, if we assume

that $S = (S_n)_n$ is a Markov chain in a state space E giving the time evolution of the stock price on which the Asian option is written, at each time n we define X_n as the path from time $p = 0$ up to the current time $p = n$ of the underlying Markov chain. In other words:

$$X_n := (S_0, \ldots, S_n) \in E_n := E^{n+1}$$

and the payoff of the option can be written in the form

$$f_n(X_n) = (H_n(X_n) - K)_+ \ , \tag{2.2.4}$$

where $K > 0$ is the strike of the option and where, in the case of the one dimensional fixed strike Asian option ($E = \mathbb{R}$):

$$H_n(X_n) = \frac{1}{n+1} \sum_{p=0}^{n} S_p \ . \tag{2.2.5}$$

Notice that this formalism for the Asian option includes the case of plain European options if we take $H_n(X_n) = f_n(S_n)$. Notice also that, if we choose $K = 0$ in (2.2.4) and

$$H_n(X_n) = \frac{1}{n+1} \sum_{p=0}^{n} S_p - S_n \ ,$$

then we have the *floating strike* Asian option with a null price at the origin. Many other payoff functions on path space can be considered, including geometric means, *better-off* or *worse-off* lookback options related to the maximum or the minimum values of the historical asset prices.

Remark 1. Notice that Importance Sampling models can also be encapsulated in the Feynman-Kac formula (2.2.1). These stochastic sampling methods are *simple* change of probability measures. They are often used in rare event simulation to make events with small occurrence probability less rare [40, 47].

2.2.2 Continuous time models

In continuous time finance, the stochastic factors and the underlying stock prices are often given by diffusion models, and the reference Markov chain sequence S or X often results from a discretization procedure, such as those given by Euler or Milshtein schemes. For instance, let us suppose we are given an $\mathbb{R}^d$-valued Itô stochastic differential equation

$$dS_t^c = b(S_t^c)\ dt + \sigma(S_t^c)\ dW_t \ , \tag{2.2.6}$$

with some initial random vector $S_0^c \in \mathbb{R}^d$ with distribution $\eta_0 = \mathrm{Law}(S_0^c)$. Here, $W = (W_t)_{t \geq 0}$ is a standard d-dimensional Wiener process, and for any $x \in \mathbb{R}^d$,

$\sigma(x) = (\sigma_{i,j}(x))_{1\le i,j\le d}$ is a $d \times d$ symmetric nonnegative definite matrix, and $b(x) = (b_i(x))_{1\le i\le d}$ a d-dimensional vector. The Euler discretization scheme over the regular time subdivision (also called time grid) $(t_n)_{n\ge 0}$, with the mesh $(t_n - t_{n-1}) = \Delta > 0$ is given by

$$S_n - S_{n-1} = b\left(S_{n-1}\right)\ \Delta + \sigma\left(S_{n-1}\right)\ \left(W_{t_n} - W_{t_{n-1}}\right)\ . \tag{2.2.7}$$

The elementary Markov transition

$$M(x, dy) := \mathbb{P}\left(S_n \in dy | S_{n-1} = x\right)$$

(the time subscripts are not needed because of the time homogeneity of the chain) can alternatively be defined in the integral form on bounded test functions as below

$$M(f)(x) := \int\ M(x, dy)\ f(y) = \mathbb{E}\left(f\left(x + b(x)\Delta + \sigma(x)\sqrt{\Delta}\ Y\right)\right)\ , \tag{2.2.8}$$

where $Y = (Y^i)_{1\le i\le d}$ is a sequence of independent and centred Gaussian random variables with unit variance.

In the same vein, suppose that the evolution of the underlying prices is given by a jump type Markov process S^c which evolves between jumps times T_n as in (2.2.6) the jump times T_n being defined in terms of a sequence $(e_n)_{n\ge 1}$ of independent and identically exponentially distributed random variables with unit parameter by the following recursion

$$T_n = \inf\left\{t \ge T_{n-1}\ :\ \int_{T_{n-1}}^{t}\ \lambda(S_u)\ du \ge e_n\right\}\ , \tag{2.2.9}$$

with $T_0 = 0$ and some non negative function λ. At the time T_n of a jump, the process jumps from $S^c_{T_n-}$ to a new location $S^c_{T_n}$ randomly chosen with distribution $P(S^c_{T_n-}, dy)$ where $P(x, dy)$ is a given Markovian transition kernel.

A discrete time approximation model S_n is defined as above by replacing the transition M in (2.2.8), by the Markov transition MJ such that

$$(MJ)(x, dz) := \int M(x, dy)\ J(y, dz)\ ,$$

with the geometric jump type Markov transition

$$J(y, dz) = e^{-\lambda(y)\Delta}\ \delta_y(dz) + \left(1 - e^{-\lambda(y)\Delta}\right)\ P(y, dz)\ .$$

If we revisit the example of the barrier option for the sake of illustration, for time homogeneous barrier regions $A_n = A$, and non-negative stochastic interest rates $(R(S_t))_{t\ge 0}$ given by a function R on $\mathbb{R}$, if we set $r_n(x) = R_{t_n}(x)\Delta$ and $X = S$ in (2.2.3), then formula (2.2.1) gives a Δ-approximation of the continuous time model

$$\mathbb{E}\left(f_{t_n}(S^c_{t_n})\ \mathbb{1}_{T\ge t_n}\ \exp\left\{-\int_{t_p}^{t_n}\ R_s(S^c_u)du\right\}\Big| S^c_{t_p} = x\right)\ ,$$

where T stands for the first time the process S gets out of the barrier region A.

2.3 Interacting particle approximations

In this section, we present a brief introduction to interacting particle methods as they pertain to the computation of the Feynman-Kac expectations discussed in the previous section. These advanced stochastic techniques are becoming increasingly popular in economics as well as in finance. A detailed survey to this field can be found in [36, 54].

2.3.1 Feynman-Kac semigroups

First, we notice that the integral operators $Q_{p,n}$ defined in (2.2.1) can be interpreted as the linear semigroup associated with the flow of non negative measures γ_n whose values on test functions f_n are given by:

$$\gamma_n(f_n) := \int \gamma_n(dx)\ f_n(x) = \mathbb{E}\left(f_n(X_n) \prod_{0\leq q<n} G_q(X_q)\right) . \tag{2.3.1}$$

The operators $Q_{p,n}$ were defined in (2.2.1) through their action on functions. Letting them act on measures by duality we get:

$$\gamma_n(dy) = (\gamma_p Q_{p,n})(dy) := \int \gamma_p(dx)\ Q_{p,n}(x, dy) ,$$

and for $0 \leq p \leq q \leq n$ we have the semigroup property

$$Q_{p,n}(x, dz) = (Q_{p,q}Q_{q,n})(x, dz) := \int Q_{p,q}(x, dy)\ Q_{q,n}(y, dz) .$$

Using these formulas in numerical implementations requires extensive calculations due to the fact that the total mass of the measures γ_n obtained by choosing the constant function $f_n(x) \equiv 1$ in (2.3.1) is very costly to compute with a reasonable precision. To illustrate this assertion, let us suppose that $G_p = \mathbb{1}_A$, for any $p \leq n$. Then, the total mass $\gamma_n(\mathbb{1})$ coincides with the probability that the trajectories of the Markov chain X stay in the set A for all times:

$$\gamma_n(\mathbb{1}) = \mathbb{E}\left(\prod_{0\leq q<n} G_q(X_q)\right) = \mathbb{P}\left(X_p \in A \ \ 0 \leq p < n\right),$$

which is, in most cases, difficult to compute. One natural way to resolve this estimation problem is to work with the normalized distributions η_n defined by:

$$\eta_n(f_n) := \gamma_n(f_n)/\gamma_n(\mathbb{1}) . \tag{2.3.2}$$

which should be a reasonable alternative since the original unnormalized measures can be recovered from the normalized ones with the following easily checked multiplicative formula:

$$\gamma_n(f_n) = \eta_n(f_n) \times \prod_{0\leq p<n} \eta_p(G_p) . \tag{2.3.3}$$

The second key observation is that the normalized distributions η_n satisfy the following recursive equation giving a nonlinear transition in η_{n-1}:

$$\eta_n(dy) = \left(\eta_{n-1}K_{n,\eta_{n-1}}\right)(dy) = \int K_{n,\eta_{n-1}}(x,dy)\eta_{n-1}(dx) \ , \tag{2.3.4}$$

where for each probability measure η on E_{n-1}, the Markovian transition kernel $K_{n,\eta}$ on E_{n-1} is defined by

$$K_{n,\eta}(x,dz) = \int \mathcal{S}_{n-1,\eta}(x,dy)M_{n-1,n}(y,dz) \ , \tag{2.3.5}$$

where in the above displayed formula, $\mathcal{S}_{n-1,\eta}$ is the selection-jump type Markov transition defined by

$$\mathcal{S}_{n-1,\eta}(x,dy) = G_{n-1}(x)\ \delta_x(y) + (1-G_{n-1}(x))\ \ \Psi_{G_{n-1}}(\eta)(dy) \ , \tag{2.3.6}$$

with the Boltzmann-Gibbs transformation

$$\Psi_g(\eta)(dy) = \frac{g(y)}{\eta(g)}\ \eta(dy) \ . \tag{2.3.7}$$

Remark 2. It is instructive, and in fact crucial given the use of the above result in the next subsection, to understand the effect of this Boltzmann-Gibbs transformation (2.3.7) in the case of point measures. Indeed, in this case:

$$\eta = \sum_{i=1}^{N} \alpha_i \delta_{x_i} \ \hookrightarrow \ \psi_g(\eta) = \sum_{i=1}^{N} \beta_i \delta_{x_i}$$

where the new weights β_i are given by:

$$\beta_i = \frac{\alpha_i g(x_i)}{\sum_{j=1}^{N} \alpha_j g(x_j)}, \qquad\qquad i = 1, \cdots, N$$

which shows that de facto, the Boltzmann-Gibbs transformation is a *resampling with replacement* of the x_i's according to the weights $\alpha_i g(x_i)$ given by the original weights and the function g. This interpretation will be extremely important for Monte Carlo implementation purposes.

Remark 3. Formulas (2.3.4) and (2.3.5) show that the passage from η_{n-1} to η_n is done in two steps. The individual particles $x \in E_{n-1}$ distributed as η_{n-1}, are first moved into dy according to the transition $\mathcal{S}_{n-1,\eta_{n-1}}(x,dy)$. This is a *selection* since (2.3.6) says that the particle remains at x with probability $G_{n-1}(x)$, and with probability $1-G_{n-1}(x)$ it is chosen at random (independently of its current position $x \in E_{n-1}$) according to the distribution $\Psi_{\eta_{n-1}}(\eta_{n-1})(dy)$. The resulting particles $y \in E_{n-1}$ are then *mutated* into particle $z \in E_n$ according to the transition $M_{n-1,n}$ of the original Markov chain $X = (X_n)_n$. The interpretation of the selection step

will be crystal clear when we implement it in for probability distributions with finite supports which we will interpret as empirical distributions of particle systems.

Remark 4. Note that the above interpretation is not limited to $[0,1]$-valued potential functions G as long as G is non-negative and bounded, and as long as we replace G_n by $\epsilon_n G_n$ in (2.3.6) and (2.3.7) with ϵ_n such that $\epsilon_n G_n \in [0,1]$.

2.3.2 Interacting particle methodologies

We now revisit the measure flows of the previous subsection in the case of point measures given by the empirical distribution of a fixed but large number N of *particles*. Let $\underline{\xi} := (\xi_n)_{n\geq 0}$ be a Markov chain with product E_n^N as state space at time n. So at each time n, ξ_n is an N-tuple $\xi_n := (\xi_n^i)_{1\leq i\leq N}$. We assume that the transition probability of this chain is given by:

$$\mathbb{P}\left(\xi_n \in dx^1 \times \cdots \times x^N \,|\xi_{n-1}\right) = \prod_{1\leq i\leq N} K_{n,\eta_{n-1}^N}(\xi_{n-1}^i, dx^i) \;, \tag{2.3.8}$$

where η_{n-1}^N denotes the empirical measure of the components of ξ_{n-1}:

$$\eta_{n-1}^N := \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n-1}^i} \;.$$

We assume that the initial law η_0^N is a product distribution of the form $\eta_0^N = (\eta_0 \times \cdots \times \eta_0)$, or in other words that the initial system $\xi_0 = (\xi_0^i)_{1\leq i\leq N}$ consists of N independent and identically distributed random variables ξ_0^i with common law η_0. The transition mechanism of the chain $\underline{\xi}$ depends only upon the empirical distribution of the components of its state, not the actual values of these components. Indeed, given the empirical distribution η_{n-1}^N of the ξ_{n-1}^i's, these ξ_{n-1}^i's evolve independently of each other, each ξ_{n-1}^i moving according to the transition kernel K_{n,η_{n-1}^N}. So the interaction between the N particles is highly symmetric, and only through the empirical distribution of the particles. For this reasons, the name *mean field* particle system is used, still as a reference to the particle physics models for which they were introduced.

An interacting particle implementation of the measure flow introduced in the previous subsection is done via the flow of measures $(\eta_n^N)_n$ viewed as an approximation of the flow $(\eta_n)_n$. The rationale behind this approximation is that since η_n^N is the empirical distribution of N independent random variables with distributions $K_{n,\eta_{n-1}^N}(\xi_{n-1}^i, x)$, we expect that when η_{n-1}^N is a good approximation of η_{n-1} then in view of (2.3.8), η_n^N should be a good approximation of η_n. We define the approximation error (which is stochastic because of the randomness of the particles ξ_n^i) in terms of a sequence of centered random fields V_n^N defined by:

$$V^N = \sqrt{N}(\eta_n - \eta_n^N) = \sqrt{N}(\eta_{n-1}^N K_{n,\eta_{n-1}^N} - \eta_{n-1}K_{n,\eta_{n-1}}). \tag{2.3.9}$$

Then, under rather weak regularity conditions, one can prove that $(V_n^N)_{n\geq 0}$ converges in law as $N \to \infty$, toward a sequence of independent centered Gaussian fields $\underline{V} = (V_n)_{n\geq 0}$ with a variance function that can be explicitly expressed in terms of the Markov transitions $K_{n,\eta_{n-1}}$. This can be checked by induction [39] on the time parameter, or using martingale decompositions in terms of local sampling random fields [40, 55].

Using formula (2.3.3) as rationale, the unnormalized measures γ_n are approximated by the unbiased particle (unnormalized) measures γ_n^N defined by their actions on test functions by:

$$\gamma_n^N(f_n) = \eta_n^N(f_n) \times \prod_{0\leq p<n} \eta_p^N(G_p),$$

and the weak consistency results

$$\lim_{N\to\infty} \gamma_n^N(f_n) = \gamma_n(f_n)$$

for each fixed test function f_n is proven by an elementary argument.

The stochastic perturbation analysis discussed above is developed in some details in [40, 51, 55, 59], and in the recent book [43]. Under some appropriate regularity conditions on the flow of measures η_n, for any bounded measurable function f, any time horizon n, any $N \geq 1$, and any λ, the probability to have any of the following estimates is greater that $1 - 2e^{-\lambda}$

$$\left|\eta_n^N(f) - \eta_n(f)\right| \leq (1 + \sqrt{2\lambda})\ c/\sqrt{N}$$

and

$$\left|1 - \gamma_n^N(\mathbb{1})/\gamma_n(\mathbb{1})\right| \leq n\left(c_1\left(1 + 2(\lambda + \sqrt{\lambda})\right)/N + \sqrt{c_2\lambda/N}\right)\ ,$$

for some constants $c, c_1, c_2 < \infty$, whose values do not depend upon time.

By construction, the flow of Feynman-Kac measures evolves according to the two-step updating/prediction transitions,

$$\eta_n \xrightarrow{\mathcal{S}_{n,\eta_n}} \widehat{\eta}_n = \eta_n \mathcal{S}_{n,\eta_n} = \Psi_{G_n}(\eta_n) \xrightarrow{M_{n+1}} \eta_{n+1} = \widehat{\eta}_n M_{n+1}\ . \tag{2.3.10}$$

In the corresponding N-mean field particle model, this pair of recursions is replaced by a two-step selection/mutation transition in product spaces

$$\xi_n \in E^N \xrightarrow{\text{selection}} \widehat{\xi}_n \in E^N \xrightarrow{\text{mutation}} \xi_{n+1} \in E^N\ . \tag{2.3.11}$$

The genetic type evolution of the system is summarized by the following synthetic diagram:

$$\begin{bmatrix} \xi_n^1 \\ \vdots \\ \xi_n^i \\ \vdots \\ \xi_n^N \end{bmatrix} \xrightarrow{\mathcal{S}_{n,\eta_n^N}} \begin{bmatrix} \widehat{\xi}_n^1 & \xrightarrow{M_{n+1}} & \xi_{n+1}^1 \\ \vdots & & \vdots \\ \widehat{\xi}_n^i & \longrightarrow & \xi_{n+1}^i \\ \vdots & & \vdots \\ \widehat{\xi}_n^N & \longrightarrow & \xi_{n+1}^N \end{bmatrix}$$

with the selection Markov transition:

$$\mathcal{S}_{n,\eta_n^N}(\xi_n^i, x) := G_n(\xi_n^i)\ \mathbb{1}_{\xi_n^i}(x) + \left(1 - G_n(\xi_n^i)\right) \sum_{1\leq j\leq N} \frac{G_n(\xi_n^j)}{\sum_{1\leq k\leq N} G_n(\xi_n^k)} \mathbb{1}_{\xi_n^j}(x)\ . \tag{2.3.12}$$

For general non necessarily $[0,1]$-valued potential functions G, we replace the acceptance rate $G_n(\xi_n^i)$ by $G_n(\xi_n^i)/\max_j G_n(\xi_n^j)$.

2.3.3 Path space models

We now work in the path space set up introduced earlier in our discussion of the Asian option example. In other words, we assume that the reference Markov chains X and the potential function G in (2.3.1) are defined on path spaces:

$$X_n := (S_0, \cdots, S_n) \in E_n = E^{n+1} \quad \text{and} \quad G_n(X_n) := G_n\left(S_0, \ldots, S_n\right)\ . \tag{2.3.13}$$

Genealogical tree based algorithms

The abstract Feynman-Kac formulae discussed above are more general than it may appear. They can be used to analyze path spaces models, including historical processes or transition space models, as well as finite excursion models. These stochastic models also encapsulate quenched Feynman-Kac models with respect to some parameter, island type coarse grained particle algorithms, Brownian type bridges and linear Gaussian Markov chains conditioned on starting and end points. For n extensive discussion on these path space models, we refer the interested reader to Section 2.4, Section 2.6, and Chapters 11-12 in the monograph [40], as well as Section 2.6 of the lecture notes [51], and Section 2.3.4 and Section 2.7.3 of the present chapter.

In the situation of this subsection, γ_n is a measure on E^{n+1} defined by

$$\gamma_n(f_n) = \mathbb{E}\left(f_n(S_0, \ldots, S_n) \prod_{0\leq q<n} G_q(S_0, \ldots, S_q)\right)\ .$$

Its mean field particle approximation is defined as before, but now, a particle at time n is a path of length $n+1$. The selection transition consists in selecting a path-particle with high potential value, while the mutation transition simply consists in extending the path with an elementary move according to the auxiliary process $X'_n = S_n$, with Markov transitions M'_n on the state space E. When the potential functions only depend upon the terminal value of the paths

$$G_n(X_n) := G'_n(S_n)\ ,$$

for some G'_n which we sometimes call *fitness function*, we can check that the path particle model gives the evolution of the genealogical tree model associated with the

time evolution of the individuals ξ_n^i evolving with M'_n-mutations and G'_n-selections. In this situation, if

$$\xi_n^i := (\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i)$$

stands for the i-th ancestral line of the current individual $\xi_{n,n}^i$ after the n-th mutation, then for any function f_n on E_n, we have that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N f_n\left(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i\right) = \frac{\mathbb{E}\left(f_n(S_0,\ldots,S_n)\ \prod_{0\le p<n} G'_p(S_p)\right)}{\mathbb{E}\left(\prod_{0\le p<n} G'_p(S_p)\right)} \ . \tag{2.3.14}$$

In addition, we also have the unbiased unnormalized estimates in the sense that:

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^N f_n\left(\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i\right) \times \prod_{0\le p<n} \frac{1}{N}\sum_{i=1}^N G'_p(\xi_{p,p}^i) \\ &\simeq_{N\uparrow\infty} \mathbb{E}\left(f_n(S_0,\ldots,S_n)\ \prod_{0\le p<n} G'_p(S_p)\right) \ . \end{aligned} \tag{2.3.15}$$

If we look at what this particle algorithm gives in the case of the one dimensional fixed strike Asian option (2.2.5) with stochastic interest rates $r_p(X'_p) > 0$ and time homogeneous barrier set $A_p = A$, we have the unbiased estimates:

$$\begin{aligned} &\mathbb{E}\left(\left(\frac{1}{n+1}\sum_{p=0}^n S_p - K\right)_+ \ 1\!\!1_{T\ge n}\ \exp\left\{\sum_{0\le q<n} r_q(S_q)\right\}\right) \\ &\simeq_{N\uparrow\infty} \frac{1}{N}\sum_{i=1}^N \left(\frac{1}{n+1}\sum_{p=0}^n \xi_{p,n}^i - K\right)_+ \times \prod_{0\le p<n} \frac{1}{N}\sum_{i=1}^N e^{-r_p(\xi_{p,p}^i)} 1\!\!1_A(\xi_{p,p}^i) \ , \end{aligned}$$

where T stands for the first exit time of the process S outside the barrier A. The approximation of European barrier call option prices with strike $K > 0$, stochastic interest rates $r_p(S_p)$, and time homogeneous barrier set A is even simpler. It is given by the unbiased estimates:

$$\begin{aligned} &\mathbb{E}\left((S_n - K)_+ \ 1\!\!1_{T\ge n}\ \exp\left\{\textstyle\sum_{0\le q<n} r_q(S_q)\right\}\right) \\ &\simeq_{N\uparrow\infty} \frac{1}{N}\sum_{i=1}^N \left(\xi_{n,n}^i - K\right)_+ \times \prod_{0\le p<n} \frac{1}{N}\sum_{i=1}^N e^{-r_p(\xi_{p,p}^i)} 1\!\!1_A(\xi_{p,p}^i) \ . \end{aligned}$$

If we use as before the notations η_n^N and η_n for the occupation measures of the ancestral lines and its limiting measures defined in (2.3.15), using the concentration analysis of mean field particle models developed in [59], the following exponential estimate was proved in [51]. Under some natural regularity conditions on the flow of

the n-th time marginal measures, for any bounded measurable function f_n on path space, time horizon n, $N \geq 1$, and λ, the following estimate:

$$\left|\eta_n^N(f_n) - \eta_n(f_n)\right| \leq \left(c_1\ (n+1)\left(1+2(\lambda+\sqrt{\lambda})\right)/N + c_2\ \sqrt{\lambda(n+1)/N}\right)\ ,$$

holds with probability greater that $1-2e^{-\lambda}$. Here c_1 and c_2 are finite constants whose values do not depend on the time parameter.

Backward Markov chain model

To distinguish path space measures and their finite time marginals, we denote by Γ_n and $\mathbb{Q}_n$ the measures on path space defined for any function F_n on E_n by

$$\Gamma_n(F_n) = \mathbb{E}\left(F_n(S_0,\ldots,S_n)\prod_{0\leq q<n} G_q(S_q)\right) \quad \text{and} \quad \mathbb{Q}_n(F_n) = \Gamma_n(F_n)/\Gamma_n(\mathbb{1})\ , \tag{2.3.16}$$

for some Markov chain $S=(S_n)_n$ on the state space E with initial distribution η_0, and some space time potential $G=(G_n)_n$. We also denote by γ_n and η_n the n-th marginal measure defined for any function f_n on E by

$$\gamma_n(f_n) = \mathbb{E}\left(f_n(S_n)\prod_{0\leq q<n} G_q(S_q)\right) \quad \text{and} \quad \eta_n(F_n) = \gamma_n(F_n)/\gamma_n(\mathbb{1})\ .$$

We observe that

$$\Gamma_n(d(s_0,\ldots,s_n)) := \left\{\prod_{0\leq q<n} G_q(s_q)\right\}\ \mathbb{P}_n(ds_0\times\cdots\times ds_n)\ , \tag{2.3.17}$$

with the probability measure $\mathbb{P}_n$ on the path space E_n defined by

$$\mathbb{P}_n(ds_0\times\cdots\times ds_n) = \eta_0(ds_0)M_1(s_0,ds_1)\ldots M_n(s_{n-1},ds_n)\ .$$

We further assume that the Markov transitions $M_n(s,ds')$ of the reference Markov chain S has a density $H_n(s,s')$ with respect to some measure $\lambda_n(ds')$:

$$M_n(s,ds') = H_n(s,s')\ \lambda_n(ds')\ .$$

In this case, one easily derives the following backward representation:

$$\begin{aligned}&\mathbb{Q}_n(d(s_0,\ldots,s_n))\\ &:= \eta_n(ds_n)\times\ \mathbb{M}_{n,\eta_{n-1}}(s_n,ds_{n-1})\ \cdots\ \mathbb{M}_{2,\eta_1}(s_2,ds_1)\times\mathbb{M}_{1,\eta_0}(s_1,ds_0)\ ,\end{aligned} \tag{2.3.18}$$

with the time reversal Markov transitions $\mathbb{M}_{n,\eta_{n-1}}(s_n,ds_{n-1})$ defined by

$$\mathbb{M}_{n,\eta_{n-1}}(s_n,ds_{n-1}) := \frac{\eta_{n-1}(ds_{n-1})G_{n-1}(s_{n-1})H_n(s_{n-1},s_n)}{\eta_{n-1}\left(G_{n-1}H_n(\cdot,s_n)\right)}\ .$$

We refer the interested reader to the article [45] for a detailed discussion on these Markov transitions. Mimicking formula (2.3.18) an alternative particle approximation of the measures $\mathbb{Q}_n$ by the following estimates

$$\begin{aligned}
&\mathbb{Q}_n^N(d(s_0,\ldots,s_n))\\
&:= \eta_n^N(ds_n)\times \mathbb{M}_{n,\eta_{n-1}^N}(s_n,ds_{n-1})\cdots\mathbb{M}_{2,\eta_1^N}(s_2,ds_1)\times\mathbb{M}_{1,\eta_0^N}(s_1,ds_0) \qquad (2.3.19)\\
&\to_{N\uparrow\infty}\ \mathbb{Q}_n(d(s_0,\ldots,s_n))
\end{aligned}$$

and the unbiased unnormalized estimates

$$\begin{aligned}
&\Gamma_n^N(d(s_0,\ldots,s_n))\\
&:= \gamma_n^N(\mathbb{1})\times\mathbb{Q}_n^N(d(s_0,\ldots,s_n))\\
&= \gamma_n^N(ds_n)\times \mathbb{M}_{n,\eta_{n-1}^N}(s_n,ds_{n-1})\cdots\mathbb{M}_{2,\eta_1^N}(s_2,ds_1)\times\mathbb{M}_{1,\eta_0^N}(s_1,ds_0) \qquad (2.3.20)\\
&\to_{N\uparrow\infty}\ \Gamma_n(d(s_0,\ldots,s_n))\ .
\end{aligned}$$

Notice also that the computation of sums with respect to these particle measures are reduced to summations over the particles locations ξ_n^i. It is therefore natural to identify a population of individual $(\xi_n^1,\ldots,\xi_n^N)$ at time n to a specific ordering of the set $\{1,\ldots,N\}$ of indexes. In this case, the occupation measures and the functions are identified with the following row and column vectors

$$\eta_n^N := \left[\frac{1}{N},\ldots,\frac{1}{N}\right] \quad \text{and} \quad \mathrm{f}_n := \begin{bmatrix} f_n(\xi_n^{\mathbb{1}})\\ \vdots\\ f_n(\xi_n^N)\end{bmatrix}$$

and the matrices $\mathbb{M}_{n,\eta_{n-1}^N}$ by the $N\times N$ matrices

$$\mathbb{M}_{n,\eta_{n-1}^N} := \begin{bmatrix} \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^{\mathbb{1}},\xi_{n-1}^{\mathbb{1}}) & \cdots & \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^{\mathbb{1}},\xi_{n-1}^N)\\ \vdots & \vdots & \vdots\\ \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^N,\xi_{n-1}^{\mathbb{1}}) & \cdots & \mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^N,\xi_{n-1}^N)\end{bmatrix},$$

with the (i,j)-entry $\mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^i,\xi_{n-1}^j)$ defined by:

$$\mathbb{M}_{n,\eta_{n-1}^N}(\xi_n^i,\xi_{n-1}^j) = \frac{G_{n-1}(\xi_{n-1}^j)H_n(\xi_{n-1}^j,\xi_n^i)}{\sum_{k=1}^N G_{n-1}(\xi_{n-1}^k)H_n(\xi_{n-1}^k,\xi_n^i)}\ .$$

For instance, the $\mathbb{Q}_n$-integration of normalized additive linear functionals of the form

$$\overline{F}_n(s_0,\ldots,s_n) = \frac{1}{n+1}\sum_{0\le p\le n} f_p(s_p) \qquad (2.3.21)$$

is given in the particle matrix approximation model by:

$$\mathbb{Q}_n^N(\overline{F}_n) = \frac{1}{n+1} \sum_{0\leq p\leq n} \eta_n^N \mathbb{M}_{n,\eta_{n-1}^N} \mathbb{M}_{n-1,\eta_{n-2}^N} \ldots \mathbb{M}_{p+1,\eta_p^N}(f_p) \ . \tag{2.3.22}$$

Several non asymptotic convergence estimates have been developed in [45], distinguishing the bias error

$$\sup_{n\geq 0} \left| \mathbb{E}\left(\mathbb{Q}_n^N(\overline{F}_n)\right) - \mathbb{Q}_n^N(\overline{F}_n) \right| \leq \frac{c}{N} \ ,$$

and the mean quadratic error

$$\mathbb{E}\left|\mathbb{Q}_n^N(\overline{F}_n) - \mathbb{Q}_n(\overline{F}_n)\right|^2 \leq \frac{c}{N}\left(\frac{1}{n+1} + \frac{1}{N}\right) \quad , \quad \text{for all } n \geq 0 \ ,$$

where c is some finite constant that does not depend on the time parameter n. Thus, for any large time horizon $n \geq N$, the upper bound on the mean square error given in the above right hand side is of the order $1/N^2$. More recently, a different estimate was proven in [51] using the concentration methodology developed in [59]. Under some appropriate regularity conditions on the flow of the n-th time marginal measures, for any sequence of bounded measurable functions f_n, any time horizon n, any $N \geq 1$, and any λ, the estimate

$$\left|\mathbb{Q}_n^N(\overline{F}_n) - \mathbb{Q}_n(\overline{F}_n)\right| \leq \left(c_1\left(1 + 2(\lambda + \sqrt{\lambda})\right)/N + c_2\sqrt{\lambda/(N(n+1))}\right) \ ,$$

for some constants $c_1, c_2 < \infty$ whose values do not depend upon n, holds with probability greater that $1 - 2e^{-\lambda}$.

Additive functionals of the form (2.3.21) arise in many applications in finance. For instance, in the context of continuous Asian option, this approach could allow to improve seriously the trade-off between the bias induced by the discrete approximation of the continuous integral payoff and the variance of the Monte Carlo method approximating the expectation. We refer to [71] for a survey of numerical methods for this type of options. As an example we consider the case of continuous Asian options with payoff functions of the following general form

$$F_T\left((S_t^c)_{0\leq t\leq T}\right) = \left(\frac{1}{T}\int_0^T f(u, S_u^c)\, du - K\right)^+ \ .$$

The strategy coming out of the above discussion suggests to first estimate the payoff by the following arithmetic average

$$\overline{F}_n(S_{t_0}^c, \cdots, S_{t_n}^c) = \frac{1}{n+1} \sum_{0\leq p\leq n} f(t_p, S_{t_p}) \ ,$$

with a time discretization $t_{p+1} - t_p = T/(n+1)$, inducing an approximation error of order $1/n$. Then the backward Markov chain scheme (2.3.22) can be used to estimate

the expectation with $N \geq n$ particles, inducing the same order of approximation error $1/n$. Of course, usual variance reduction techniques as control variate can be applied in addition to that approach.

This type of additive functionals will be used in Section 2.5 in our discussion of sensitivity measure computations.

2.3.4 Parallel island particle models

Island genetic models are powerful parallel computational techniques used to speed up interacting genetic search algorithms. These coarse grained parallel procedures are very popular in genetic algorithms literature (see for instance [35, 112],[90], and the references therein).

In our context, we run in parallel several genetic type interacting particle algorithms on a collection of islands. At a geometric stochastic rate, the populations between islands interact according to some selective migration processes. The island selection mechanism is defined in terms of the averaged fitness of the individuals in the island population.

To define these island particle models more precisely, we observe that the unbiased properties of the unnormalized Feynman-Kac measures γ_n^N can be rewritten as follows

$$\mathbb{E}\left(f_n(S_n)\prod_{0\leq q<n} G_q(S_q)\right) = \mathbb{E}\left(\mathcal{F}_n(\mathcal{X}_n)\prod_{0\leq p<n}\mathcal{G}_p(\mathcal{X}_p)\right)\ , \tag{2.3.23}$$

with the Markov chain $\mathcal{X}_n = (\xi_n^i)_{1\leq i\leq N}$ on the product spaces $\mathcal{E}_n = E_n^N$, an the empirical functionals $\mathcal{F}_n$, and $\mathcal{G}_n$ defined by

$$\mathcal{F}_n(\mathcal{X}_n) = \eta_n^N(f_n) = \frac{1}{N}\sum_{i=1}^N f_n(\xi_n^i) \quad \text{and} \quad \mathcal{G}_n(\mathcal{X}_n) = \eta_n^N(G_n) = \frac{1}{N}\sum_{i=1}^N G_n(\xi_n^i)\ .$$

Now, it is important to notice that the r.h.s. term in formula (2.3.23) has exactly the same mathematical form as the Feynman-Kac measures γ_n introduced in (2.3.1). Thus, applying the particle methodologies developed in Section 2.3.2 to these models, we define an N-interacting island particle model with a mutation and a selection transition on the space of islands $\mathcal{E}_n$.

During the mutation stage, the population in each island evolve independently one another according to the genetic type Markov transitions of the chain $\mathcal{X}_n$. In other words, we run in parallel the selection mutation transitions of N genetic particle models (2.3.11). During the selection stage, we evaluate the $\mathcal{G}_n$-potential value of each island. As in (2.3.12), at a geometric rate we select the island populations using the empirical potential function $\mathcal{G}_n$.

We observe that the island version of the acceptance ratio in the selection transition (2.3.12) discussed in the end of Section 2.3.2 tends to 1, as the number of

individuals in each island tends to infinity. In other words, the independence degree between the islands increases with respect to the size of their populations. As proposed in the recent article [2], an alternative sampling approach is to use an independent Metropolis-Hasting model with a target measure defined by the r.h.s. term in formula (2.3.23) (up to some normalizing constant). One again, the unbiased property (2.3.23) ensures that the limiting target measure coincides with the desired Feynman-Kac measures η_n, as well as the measure $\mathbb{Q}_n$ for the path space version of these island models. These island particle models will be used in Section 2.7.3 dedicated to fixed parameter estimation in Hidden Markov chain models.

2.4 Application in credit risk analysis

The simulation of credit events with remarkably small probabilities is a key issue for regulatory and risk management purposes, as well as for the pricing of credit derivatives. The main variance reduction technique used in Monte Carlo computations of rare events is importance sampling. However in general multi-name credit models, desirable changes of measure favoring sample paths realizing rare events are highly unlikely to lead to explicit formula. In this case importance sampling is no longer an option. A natural alternative is then interacting particles methods.

Though interacting particle systems are known to provide very efficient variance reductions in Monte Carlo approximations of rare events, these algorithms have only appeared recently in the credit risk literature with for instance the articles of Carmona, Fouque and Vestal [24] and Carmona and Crepey [21]. In Chapter 21 of [17], the authors provide an overview of the main techniques and results of the application of interacting particle systems to credit risk analysis. We also refer to [57] for some recent applications of these techniques in the financial risk area. All these results show the strengths of IPS based Monte Carlo computations of small default probabilities, especially when other methods fail. A systematic comparison with importance sampling is provided in [21].

2.4.1 Change of measure for rare events and Feynman-Kac formula

We consider a Markov chain $S = (S_n)_{0\leq n\leq T}$ representing at each time n, d correlated risky sources $S_n = (S_n^1, \ldots, S_n^d) \in E$. We are interested in understanding the asymptotic behavior of probabilities of rare events of the form $\{V_T(S_T) \geq K\}$ or more generally $\{V_T(S_0, \ldots, S_T) \geq K\}$, where V_T is some real positive function whose value can be thought of as a risk measure.

To compute $\mathbb{P}(V_T(S_T) \geq K)$, standard Monte Carlo simulations usually fail, because of the difficulty to ensure that enough simulation samples realize the rare

event. A partial remedy amounts to providing a reasonably tight upper-bound based on large deviations ideas. Indeed, for any $\lambda \geq 0$, we have:

$$\mathbb{P}(V_T(S_T) \geq K) = (\mathbb{E}\left(\mathbf{1}_{V_T(S_T)\geq K} e^{\lambda V_T(S_T)} e^{-\lambda V_T(S_T)}\right) \leq e^{-\lambda K}\mathbb{E}\left(\mathbf{1}_{V_T(S_T)\geq K} e^{\lambda V_T(S_T)}\right),$$

and if we denote by $\mathbb{E}^{(\lambda)}$ the expectation under the probability $\mathbb{P}^{(\lambda)}$ defined by

$$d\mathbb{P}^{(\lambda)} \propto e^{\lambda V_T(S_T)} d\mathbb{P} \ ,$$

we have

$$\mathbb{E}\left(\mathbf{1}_{V_T(S_T)\geq K} e^{\lambda V_T(S_T)}\right) = \mathbb{E}^{(\lambda)}\left(\mathbf{1}_{V_T(S_T)\geq K}\right)\mathbb{E}\left(e^{\lambda V_T(S_T)}\right),$$

and using the fact that:

$$\mathbb{E}^{(\lambda)}\left(\mathbf{1}_{V_T(S_T)\geq K}\right) \leq 1,$$

we get:

$$\mathbb{P}(V_T(S_T) \geq K) \leq e^{-\sup_{\lambda\geq 0}(\lambda K - \Lambda(\lambda))} \ , \tag{2.4.1}$$

where $\Lambda(\lambda)$ is defined by Fenchel transformation as $\log(\mathbb{E}(\lambda V_T(S_T)))$.

From the above argument we see that we can approximate the desired probability by searching a proper λ. This large deviation type approach is widely used, but in the form of (2.4.1), it requires extensive calculations in order to obtain a reasonable approximation of the desired probability.

Del Moral and Garnier provide in [47] a zero-bias estimate with interacting particle systems. The idea is to construct a genealogical tree based model as mentioned in Section 2.3.3 instead of the large deviation type inequality used above.

Using again the same change of measure from $\mathbb{P}$ to $\mathbb{P}^{(\lambda)}$, we remark that the target probability

$$\mathbb{P}(V_T(S_T) \geq K) = \mathbb{E}\left(\mathbf{1}_{V_T(S_T)\geq K} e^{\lambda V_T(S_T)} e^{-\lambda V_T(S_T)}\right)$$

can be written as

$$\mathbb{E}^{(\lambda)}\left(\mathbf{1}_{V_T(S_T)\geq K} e^{-\lambda V_T(S_T)}\right)\mathbb{E}\left(e^{\lambda V_T(S_T)}\right) = \mathbb{E}^{(\lambda)}\left(f_T(S_T)\right)\mathbb{E}\left(e^{\lambda V_T(S_T)}\right) \ ,$$

with $f_T(S_T) := \mathbf{1}_{V_T(S_T)\geq K} e^{-\lambda V_T(S_T)}$. It is also important to notice that, with the convention $V_0 = 0$, we have the following decomposition

$$e^{\lambda V_T(S_T)} \equiv \prod_{p=1}^{T} e^{\lambda(V_p(S_p) - V_{p-1}(S_{p-1}))} \ .$$

By using the notation $\mathcal{X}_k = (S_k, S_{k+1})$ for $0 \leq k < T$, the above product can be defined as

$$\prod_{p=1}^{T} G_{p-1}(\mathcal{X}_{p-1}) \ , \quad \text{where} \quad G_{p-1}(\mathcal{X}_{p-1}) := e^{\lambda(V_p(S_p) - V_{p-1}(S_{p-1}))} \ .$$

Using the notation $F_T(\mathcal{X}_T) = f_T(S_T)$, we see that we need to simulate the same formulae as in Section 2.3.3.

$$\mathbb{E}^{(\lambda)}\left(f_T(S_T)\right) = \frac{\mathbb{E}\left(F_T(\mathcal{X}_T)\prod_{p=1}^{T} G_p(\mathcal{X}_P)\right)}{\mathbb{E}\left(\prod_{p=1}^{T} G_p(\mathcal{X}_P)\right)} := \eta_T(F_T) \ .$$

The general discussion of the previous section shows that these quantities can be approximated efficiently by interacting particle systems.

In practice, we are interested in the conditional law $\mathcal{L}\Big((S_0,\ldots,S_T)|V_T(S_T) \geq K\Big)$. By modifying the function F_T in the above analysis, the same framework can be applied directly. Particle models are very flexible, but the choice of the space-time potential function on path space can become very tricky and the performance of the algorithm can deteriorate with a poor choice of this potential. The particular choice

$$G_p(s_0,\ldots,s_p) = \mu e^{\lambda(V_p(s_0,\ldots,s_p)-V_{p-1}(s_0,\ldots,s_{p-1}))} \tag{2.4.2}$$

was proposed and analyzed in [47] where μ is chosen so that $G_p \leq 1$ and λ can be fine-tuned to the given rare event set.

2.4.2 On the choice of the potential functions

As mentioned earlier, the choice of suitable space-time potential functions G is a key ingredient in the ability of interacting particle systems to tackle rare events problems. In the recent work of Carmona, Fouque and Vestal [24], the authors propose a choice of the potential functions that departs from the one given above in (2.4.2). Their construction illustrates the flexibility of the particle methods regarding the crucial point of choice of the potential functions. In the case of large credit portfolios, typically with $d = 125$, we write the dynamics of the various assets values as a Markov chain S_n^i, with time $n = 1,\cdots,T$ and $i = 1,\ldots,d$ associated to

$$G_n = \exp\left(-\alpha\sum_{i=1}^{d}\log\frac{\min\limits_{0\leq l\leq n} S_l^i}{\min\limits_{0\leq l\leq n-1} S_l^i}\right) \ ,$$

where the parameter α has to be fine-tuned to the particular class of rare events of interest. Numerical performance of this technique is dicussed in [24] where examples are provided under a structural model with stochastic volatility. The authors demonstrate the efficiency of this method, especially in situations where importance sampling is not possible or numerically unstable.

In a similar vein, a fast algorithm without requirement of fine-tuned parameters has been recently developed for multiple defaults models by setting the potential function

$$G_p(x) = 1 - \mathbf{1}_{\{c\}}(x) \ , \tag{2.4.3}$$

where the c stands for a cemetery state under a multilevel splitting approach introduced in Chapter 12 in [40]. Let $S = (S_n)_{0 \leq n \leq T}$ be a Markov chain on a sequence of state spaces $E = (E_n)_{0 \leq n \leq T}$ and X_n (resp. $F_n = E_0 \times \cdots \times E_n$) the corresponding path space Markov chain (resp. sequence of path state spaces). We assume that a sequence of subsets $U_1, \ldots, U_T,\ U_p \in F_p$ is fixed. We are typically interested in the probability $1 - \mathbb{P}(X_1 \notin U_1, \ldots, X_T \notin U_T)$ that the trajectory does enter at least one of these subsets. The key idea is to introduce a series of intermediate events interpolating between the series of the full state space $E_1, \ldots, E_T$ of the path space and the target rare event series $U_1, \ldots, U_T$. Then we assume that such a series is given:

$$\forall\ p \leq T,\ U_p = U_p^{(k)} \subset U_p^{(k-1)} \subset \cdots \subset U_p^{(1)} \subset U_p^{(0)} = F_p\ .$$

Then the state $\{c\}$ appearing in (2.4.3) is defined in the construction of a new Markov chain in constant state space $F := F_0 \cup \cdots \cup F_T \cup \{c\}$. With a series of stopping times:

$$\tau_j := (T+1) \wedge \inf\{p, X_p \in U_p^j\}$$

with the convention that $X_{T+1} := c$. Then the process $Z_0 := X_0,\ Z_1 := X_{\tau_1}, \ldots, Z_k := X_{\tau_k}$ is a Markov chain on F. In this context, the potential functions (2.4.3) consist in, roughly speaking, killing the trajectories at some point of the recursion of the particle algorithm when they reach some of the intermediate rare event sets associated to c.

2.5 Sensitivity computation

Partial derivatives of financial option values allow traders to determine how sensitive the values of options are to small changes in the set of parameters on which they depend, such as the volatility parameter, the risk free stochastic interest rates or prices of assets related to the option. The computation of these *sensitivities*, often called *Greeks* (because they are traditionally denoted by Greek letters) is a central problem in computational finance that must be addressed for risk analysis applications. Besides, in the specific case of sensitivities with respect to assets prices, (called *delta* and *gamma* for the first and second order derivatives) the practical issue is even more crucial since they are the basic ingredients of dynamic hedging strategies.

There are mainly three approaches to compute sensitivities. We refer to the survey paper of Kohatsu-Higa and Montero [81], for a detailed presentation and comparison of those methods. The most natural and simple approach to compute sensitivities is the usual finite difference method. It is easily implemented but known to necessitate large computing budgets (requiring for instance two option calculations in the case of a first order sensitivity) and unstable with a subtle trade-off

between bias and variance. We focus here on the two alternative approaches introduced in the pioneering paper of Broadie and Glasserman [18], namely the *likelihood ratio* method and the *pathwise*, or *tangent process* method. In this section, these techniques are presented in terms Feynman-Kac formula, showing in some specific examples how particle methods can be used.

2.5.1 Likelihood ratio: application to dynamic parameter derivatives

This technique introduced in [18] requires that the underlying interest on which a European option is written admits a sufficiently regular density with respect to Lebesgue measure, also known as state price density. The main idea is to interchange differentiation and integration and whenever the derivative with respect to a variable not appearing in the payoff function, to apply the differentiation on the density of the distribution. The advantage of this approach is that it does not require any regularity assumption on the payoff function, allowing for kinks and discontinuities. This approach has been generalized by Fournié, Lasry, Lebuchoux, Lions and Touzi [69] to path space using Malliavin integration-by-parts argument, allowing for a wide class *Greek weights*.

In this subsection, we focus on the computation of the sensitivity of an option to dynamic parameters related to the risky asset evolution or to the risk free rate variations.

We let $\theta \in \mathbb{R}^d$ be a parameter that may represent the volatility of some asset price movements, or any other kinetic parameter. We assume that the evolution of the risky asset price $S_k^{(\theta)}$ associated to some value of the parameter θ, is given by a one-step probability transition of the form

$$M_k^{(\theta)}(s, ds') := \text{Proba}\left(S_k^{(\theta)} \in ds' | S_{k-1}^{(\theta)} = s\right) = H_k^{(\theta)}(s, s')\ \lambda_k(ds')\ ,$$

for some positive density functions $H_k^{(\theta)}(s, s')$ and some reference measure λ_k. We also consider a collection of functions $G_k^{(\theta)}(s) = e^{-r_k^{(\theta)}(s)}$ that depend on θ. We also assume that the gradient and the Hessian of the logarithms of these functions with respect to the parameter θ are well defined.

In this situation, following the Feynman-Kac representation (2.2.1) or (2.3.16), a general form of the option price on path space is provided by

$$\Gamma_n^{\theta}(F_n) = \mathbb{E}\left(F_n(S_0^{(\theta)}, \ldots, S_n^{(\theta)}) \prod_{0 \leq p < n} G_p^{(\theta)}\left(S_p^{(\theta)}\right)\right)\ . \tag{2.5.1}$$

For each value of the parameter θ, we denote by $\Gamma_n^{(\theta,N)}$ the N-particle approximation measures associated with a given value of the parameter θ and defined in (2.3.20).

Simple derivations, show that the first order derivative of the option value with respect to θ is given by

$$\begin{aligned}\nabla\Gamma_n^{(\theta)}(F_n) &= \Gamma_n^{(\theta)}(F_n\Lambda_n^{(\theta)}) \\ \nabla^2\Gamma_n^{(\theta)}(F_n) &= \Gamma_n^{(\theta)}\left[F_n(\nabla\mathbb{L}_n^{(\theta)})'(\nabla\mathbb{L}_n^{(\theta)}) + F_n\nabla^2\mathbb{L}_n^{(\theta)}\right]\end{aligned}$$

with $\Lambda_n^{(\theta)} := \nabla\mathbb{L}_n^{(\theta)}$ and

$$\mathbb{L}_n^{(\theta)}(x_0,\ldots,x_n) := \sum_{p=1}^{n}\log\left(G_{p-1}^{(\theta)}(s_{p-1})H_p^{(\theta)}(s_{p-1},s_p)\right) .$$

These quantities can be approximated by the unbiased particle models

$$\begin{aligned}\nabla_N\Gamma_n^{(\theta)}(F_n) &:= \Gamma_n^{(\theta,N)}(F_n\Lambda_n^{(\theta)}) \\ \nabla_N^2\Gamma_n^{(\theta)}(F_n) &= \Gamma_n^{(\theta,N)}\left[F_n(\nabla\mathbb{L}_n^{(\theta)})'(\nabla\mathbb{L}_n^{(\theta)}) + F_n\nabla^2\mathbb{L}_n^{(\theta)}\right] .\end{aligned}$$

We illustrate the above discussion with the computation of the Vega of the option, i.e. the sensitivity to changes in the diffusion volatility coefficient of the stochastic equation (2.2.7), with $d = 1$. We suppose $X_n^{(\theta)} = S_n^{(\theta)}$ satisfies equation

$$S_n^{(\theta)} - S_{n-1}^{(\theta)} = b\left(S_{n-1}^{(\theta)}\right)\ \Delta + \left[\sigma\left(S_{n-1}^{(\theta)}\right) + \theta\ \sigma'\left(S_{n-1}^{(\theta)}\right)\right]\ \left(W_{t_n} - W_{t_{n-1}}\right) ,$$

for some function σ' such that $\sigma + \theta\ \sigma' > 0$ for any $\theta \in [0,1]$. In this situation, we have

$$\frac{\partial}{\partial\theta}\sum_{p=1}^{n}\log\left(H_p^{(\theta)}(s_{p-1},s_p)\right)$$

$$= \sum_{p=1}^{n}\frac{\sigma'(s_{p-1})}{\sigma(s_{p-1}) + \theta\sigma'(s_{p-1})}\ \left[\left(\frac{(s_p - s_{p-1}) - b(s_{p-1})\Delta}{(\sigma(s_{p-1}) + \theta\sigma'(s_{p-1}))\sqrt{\Delta}}\right)^2 - 1\right] .$$

To compute the rho of the option, i.e. the sensitivity to changes in the drift of the stochastic equation (2.2.7), with $d = 1$, we assume that $X_n^{(\theta)}$ satisfies equation

$$S_n^{(\theta)} - S_{n-1}^{(\theta)} = \left[b\left(S_{n-1}^{(\theta)}\right) + \theta b'\left(S_{n-1}^{(\theta)}\right)\right]\ \Delta + \sigma\left(S_{n-1}^{(\theta)}\right)\ \left(W_{t_n} - W_{t_{n-1}}\right) ,$$

for some function b'. In this situation, we have

$$\frac{\partial}{\partial\theta}\sum_{p=1}^{n}\log\left(H_p^{(\theta)}(s_{p-1},s_p)\right)$$

$$= \sum_{p=1}^{n}\left[(s_p - s_{p-1}) - \left[b(s_{p-1}) + \theta b'(s_{p-1})\right]\Delta\right] \times b(s_{p-1})/\sigma^2(s_{p-1}) .$$

Finally, if we assume that changes in the stochastic interest rates are given by the space-time potential function

$$G_n(x) = \exp\left(-\left[r_n(x) + \theta r'_n(x)\right]\right) ,$$

for some non negative functions r_n and r'_n, then we have:

$$\frac{\partial}{\partial\theta} \sum_{0\leq p<n} \log\left(G_p^{(\theta)}(s_p)\right) = -\sum_{0\leq p<n} r'_p(s_p) .$$

We illustrate these particle models with an European option associated with a risky asset $S_n^{(\theta)} = S_n$ whose values do not depend on θ, and payoff function $f_n(S_n)$. In this situation, the option price is given by the formula

$$\gamma_n^{(\theta)}(f_n) = \mathbb{E}\left(f_n(S_n)\ \exp\left\{-\sum_{0\leq q<n} r_q^{(\theta)}(S_q)\right\}\right) .$$

Then using the backward Markov chain model developed in Section 2.3.3, we obtain the following unbiased particle matrix approximation for the sensitivity with respect to the interest rate r:

$$\begin{aligned}\nabla\gamma_n^{(\theta)}(f_n) \quad &= \quad -\sum_{0\leq p<n} \gamma_n^{(\theta)}\left(f_n\ \mathbb{M}_{n,\eta_{n-1}^{(\theta)}} \ldots \mathbb{M}_{p+1,\eta_p^{(\theta)}}\left(\nabla r_p^{(\theta)}\right)\right)\\ &\simeq_{N\uparrow\infty} \quad -\sum_{0\leq p<n} \gamma_n^{(\theta,N)}\left(f_n\ \mathbb{M}_{n,\eta_{n-1}^{(\theta,N)}} \ldots \mathbb{M}_{p+1,\eta_p^{(\theta,N)}}\left(\nabla r_p^{(\theta)}\right)\right) .\end{aligned}$$

2.5.2 Tangent process: application to initial state derivatives

We review the tangent process approach introduced by Broadie and Glasserman in [18], and focus on the computation of the sensitivity of an option price to perturbations of the initial value of the underlying asset price – this sensitivity is usually called the *delta* of the option – which is in general more complex than in the case of the sensitivity measures with respect to the dynamic parameters parameters. Efficient numerical schemes for the implementation of the method we are about to discuss can be found in Giles and Glasserman [74].

To simplify our presentation, we only consider European-style options with smooth payoff functions $f - (f_n)_n$.

As before, the strategy is to interchange the differentiation and expectation operations. However, in the present situation. this requires regularity of the payoff function, so discontinuous payoff profiles will have to be regularized using Gaussian kernel convolution type techniques, or any related smoothing method. For instance, we can approximate the call option (2.2.3) by the following smoothed payoff profile

$$f_\epsilon(x) = \frac{1}{2}\left[(x-K) + \sqrt{(K-x)^2+\epsilon}\right] \to_{\epsilon\downarrow 0} f(x) = (x-K)_+ .$$

We assume that the stochastic dynamics of the underlying stock price $S = (S_n)_n$ are given by an induction equation of the type:

$$S_{n+1} := F_n(S_n) = (F_n \circ F_{n-1} \circ \cdots \circ F_0)(S_0) \tag{2.5.2}$$

starting at some random state S_0, where the random functions F_n are of the form of the form

$$F_n(x) = \mathcal{F}_n(x, W_n) \ , \tag{2.5.3}$$

for functions $\mathcal{F}_n$

$$\mathcal{F}_n \ : \mathbb{R}^{d+d'} \ni (x, w) \hookrightarrow \mathcal{F}_n(x, w) \in \mathbb{R}^d$$

and some independent random variables W_n taking values in $\mathbb{R}^{d'}$, with $d' \geq 1$. We also assume that these random variables are also independent of S_0. Under these assumptions, the prices of European options are given by the semigroup of the Markov chain S defined for any regular function f and initial state x by

$$P_{n+1}(f)(x) := \mathbb{E}\left(f(S_{n+1}) \mid S_0 = x\right) = \mathbb{E}\left(f(S_{n+1}(x))\right) \ ,$$

with the random flows $\left(S_n(\,\cdot\,)\right)_{n\geq 0}$ defined for any $n \geq 0$ and $x \in E$ by:

$$S_{n+1}(x) = F_n(S_n(x)) \ ,$$

with the initial condition $S_0(x) = x$. By the chain rule, for any $1 \leq i, j \leq d$ and any $x \in \mathbb{R}^d$ we have

$$\frac{\partial S^i_{n+1}}{\partial x^j}(x) = \sum_{1\leq k\leq d} \frac{\partial F^i_n}{\partial x^k}(S_n(x)) \ \frac{\partial S^k_n}{\partial x^j}(x) \ . \tag{2.5.4}$$

Interchanging derivations and expectations in the definition of the semigroup we get:

$$\frac{\partial P_{n+1}(f)}{\partial x^j}(x) = \mathbb{E}\left(\sum_{1\leq i\leq d} \frac{\partial f}{\partial x^i}(S_{n+1}(x)) \ \frac{\partial S^i_{n+1}}{\partial x^j}(x)\right) \ . \tag{2.5.5}$$

Let us denote by $V_n = (V_n^{(i,j)})_{1\leq i,j\leq d}$ and $A_n = (A_n^{(i,j)})_{1\leq i,j\leq d}$ the random $d \times d$ matrices whose entries are given by:

$$V_n^{(i,j)}(x) = \frac{\partial S^i_n}{\partial x^j}(x)$$

and

$$A_n^{(i,j)}(x) = \frac{\partial F^i_n}{\partial x^j}(x) = \frac{\partial \mathcal{F}^i_n(\boldsymbol{.}, W_n)}{\partial x^j}(x) := \mathcal{A}_n^{(i,j)}(x, W_n) \ .$$

With this notation in hand, equation (2.5.4) can be rewritten in terms of the following random matrix formulae

$$\begin{aligned} V_{n+1}(x) &= A_n(S_n(x)) \ V_n(x) \\ &= A_n(S_n(x))A_{n-1}(S_{n-1}(x)) \cdots A_1(S_1(x))A_0(x) := \prod_{p=0}^{n} A_p(S_p(x)) \end{aligned} \tag{2.5.6}$$

with a product $\prod_{p=0}^{n} A_p$ of non commutative random elements A_p taken in the order A_n, A_{n-1},..., A_0. Using equation (2.5.5) with the payoff function $f = f_{n+1}$, we get:

$$\begin{aligned} \nabla P_{n+1}(f_{n+1})(x) &= \mathbb{E}\left(\nabla f_{n+1}(S_{n+1}(x))\ V_{n+1}(x)\right) \\ &= \mathbb{E}\left(\nabla f_{n+1}(S_{n+1}) \prod_{0\leq p\leq n} A_p(S_p) \mid S_0 = x\right) , \end{aligned} \tag{2.5.7}$$

which is, except for the fact that we are dealing with products of non-commuting random matrices, of the form of the Feynman-Kac formulas studied in this chapter.

For one dimensional models of the form

$$S_{n+1} = S_n + b\left(S_n\right)\ \Delta + \sigma\left(S_n\right)\ \sqrt{\Delta}\ W_n\ , \tag{2.5.8}$$

with a sequence of independent and and identically distributed mean zero Gaussian random variables W_n, it is readily checked that

$$A_n(x) = \mathcal{A}_n(x, W_n) = \left(1 + \frac{\partial b}{\partial x}(x)\ \Delta + \frac{\partial \sigma}{\partial x}(x)\ \sqrt{\Delta}\ W_n\right)$$

and therefore

$$\begin{aligned} V_{n+1}(x) &= \prod_{p=0}^{n}\left(1 + \frac{\partial b}{\partial x}(S_p)\ \Delta + \frac{\partial \sigma}{\partial x}(S_p)\ \sqrt{\Delta}\ W_p\right) \\ &\simeq_{\Delta\downarrow 0} \exp\sum_{0\leq p\leq n}\left(\frac{\partial b}{\partial x}(S_p)\ \Delta + \frac{\partial \sigma}{\partial x}(S_p)\ \sqrt{\Delta}\ W_p\right) . \end{aligned}$$

As already mentioned, for non smooth payoff functions we can use the following Gaussian regularization kernel

$$P_{n+1,\epsilon}(f_{n+1})(x) := \mathbb{E}\left(f_{n+1}(S_{n+1}(x) + \epsilon Y)\right) \simeq_{\epsilon\downarrow 0} P_{n+1,\epsilon}(f_{n+1})(x)\ , \tag{2.5.9}$$

for some auxiliary Gaussian variable, independent of S_n and W_n. In this case, we have the following formula

$$\frac{\partial}{\partial x} P_{n+1,\epsilon}(f_{n+1})(x) = \mathbb{E}\left(\epsilon^{-1}\left[f_{n+1}(S_{n+1}(x) + \epsilon Y) - f_{n+1}(S_{n+1}(x))\right]\ Y\ V_{n+1}(x)\right)\ .$$

In the particular case $d = 1$, the particle interpretation developed in Section 2.3.2 applies directly. W

Remark 5. As an aside, we also mention that these expansions are closely related to the time discretization of the stochastic integrals arising in exponential weights of the Feynman-Kac interpretation of the Kushner-Stratonovitch filtering equation [38]. In this interpretation, the particle interpretations of the Feynman-Kac formulae (2.5.7) coincide with the particle filters developed in the last referenced article.

Before getting into multi-dimensional models, let us pause for a while to discuss the connexions of the above methodology with the existing literature. Firstly, we observe that the Gaussian regularization formula (2.5.9) can be interpreted as the addition of an extra Gaussian move. This suggests that we can alternatively use the last transition to regularize the model.

$$P_{n+1}(f_{n+1})(x) = \mathbb{E}\left(\mathbb{E}\Big(f_{n+1}\big(S_{n+1}(x)\big)\Big|S_n(x)\Big)\right) \ .$$

Letting $H_{n+1}(x_n, x_{n+1})$ be the density of the Markov transition $S_n = x_n \rightsquigarrow S_{n+1}$ with respect to the Lebesgue measure, arguing as above we find that

$$\frac{\partial}{\partial x} P_{n+1}(f_{n+1})(x) = \mathbb{E}\left(f_{n+1}(S_{n+1}(x))\ dH_{n+1}(S_n(x), S_{n+1}(x))\ V_n(x)\right) \ ,$$

with the weight function

$$\begin{aligned} & dH_{n+1}(x_n, x_{n+1}) \\ &= \frac{\partial}{\partial x_n} \log H_{n+1}(x_n, x_{n+1}) \\ &= \left(\left(\tfrac{(x_{n+1}-x_n)-b(x_n)\Delta}{\sigma(x_n)\sqrt{\Delta}}\right)^2 - 1\right) \frac{\partial}{\partial x} \log \sigma(x_n) - \left(\tfrac{(x_{n+1}-x_n)-b(x_n)\Delta}{\sigma(x_n)\sqrt{\Delta}}\right) \ \frac{1 + \frac{\partial b}{\partial x}(x_n)\Delta}{\sigma(x_n)\sqrt{\Delta}} \ . \end{aligned}$$

These formulae and the corresponding conventional weighted Monte Carlo approximations have been recently proposed by N. Chen and P. Glasserman [31] as an alternative to the Malliavin calculus computation of the Greeks introduced by E. Fournié, J.M. Lasry, J. Lebuchoux, P.L. Lions, and N. Touzi in their groundbreaking articles [69, 70]. If $P_{s,t}$ denotes the semigroup associated with the (continuous time) diffusion equation (2.2.6) (recall that $d = 1$ in the present discussion):

$$P_{s,t}(f)(S_s) = \mathbb{E}\left(f(S_t) \mid S_s\right) \ ,$$

one easily checks that, for any $0 \leq s \leq t$ it holds:

$$P_{s,t}(f)(S_s) = P_{0,t}(f)(S_0) + \int_0^s \frac{\partial P_{r,t}(f)}{\partial x}(S_r)\ \sigma(S_r)\ dW_r \ ,$$

and if we set $s = t$ in the above equation, then we find that

$$\mathbb{E}\left[f(S_t(x))\ \int_0^t \frac{\partial S_s}{\partial x}(x)\ \sigma^{-1}(S_s(x)))\ dW_s\right] \ = \ \mathbb{E}\left[\int_0^t \frac{\partial P_{s,t}(f)}{\partial x}(S_s(x))\ \frac{\partial S_s}{\partial x}(x)\ ds\right] \ ,$$

whenever σ is a smooth positive function bounded away from 0. Recalling that

$$\frac{\partial}{\partial x} P_{0,t}(f)(x) = \frac{\partial}{\partial x}\mathbb{E}\left[P_{s,t}(f)(S_s(x))\right] = \mathbb{E}\left[\frac{\partial P_{s,t}(f)}{\partial x}(S_s(x))\ \frac{\partial S_s}{\partial x}(x)\right] \ ,$$

we arrive at a *Malliavin formulation* of the semigroup derivatives

$$\frac{\partial}{\partial x} P_{0,t}(f)(x) = \mathbb{E}\left[f(S_t(x))\ \frac{1}{t}\int_0^t \sigma^{-1}(S_s(x))\ \frac{\partial S_s}{\partial x}(x)\ dW_s\right] \ .$$

A rigorous derivation of the above equations is provided in [69, 70]. We also refer the reader to the contribution of B. Bouchard and X. Warin in the present volume.

The Euler time discretization scheme justifies using the discrete time approximate model:

$$S_{(n+1)\Delta} - S_{n\Delta} = b\left(S_{n\Delta}\right)\ \Delta + \sigma\left(S_{n\Delta}\right)\ \sqrt{\Delta}\ Y_n\ , \tag{2.5.10}$$

for a sequence of independent mean zero Gaussian random variables Y_n. We thus have the approximation model

$$\tfrac{\partial}{\partial x} P_{0,(n+1)\Delta}(f)(x) \simeq_{\Delta\downarrow 0} \frac{1}{(n+1)\sqrt{\Delta}} \sum_{0\leq p\leq n} \mathbb{E}\left(f(S_{(n+1)\Delta}(x))\mathcal{Z}_p(x)\right)\ , \tag{2.5.11}$$

with the random weights

$$\mathcal{Z}_p(x) := \varphi\left(S_{p\Delta}(x), Y_p\right) \prod_{0\leq q<p} G_q(S_{q\Delta}(x), Y_q)$$

$$\varphi\left(x,y\right) = \sigma^{-1}(x)\ y \qquad \text{and} \quad G_q(x,y) = 1 + \frac{\partial b}{\partial x}(x)\ \Delta + \frac{\partial \sigma}{\partial x}(x)\ \sqrt{\Delta}\ y\ .$$

The ratio $1/\sqrt{\Delta}$ in the right hand side of (2.5.11) may induce numerical degeneracies. One way to overcome this problem and *to remove* this term from the numerical scheme is to use the following formula

$$\mathbb{E}\left(f(S_{(n+1)\Delta}(x))\mathcal{Z}_p(x)\right) = \mathbb{E}\left(\Upsilon_{p+1,n+1}(f)\left[S_{p\Delta}(x), Y_p\right] \times \mathcal{Z}_p(x)\right)\ ,$$

with the function

$$\begin{array}{l}\Upsilon_{p+1,n+1}(f)[x,y] \\ = P_{(p+1)\Delta,(n+1)\Delta}(f)\left(x + b(x)\Delta + \sigma(x)\sqrt{\Delta}y\right) - P_{(p+1)\Delta,(n+1)\Delta}(f)\left(x + b(x)\Delta\right)\ .\end{array}$$

Under some appropriate regularity conditions, we notice that

$$\begin{array}{l}\Upsilon_{p+1,n+1}(f)[x,y] \\ \simeq_{\Delta\downarrow 0}\ P_{p\Delta,(n+1)\Delta}(f)\left(x + b(x)\Delta + \sigma(x)\sqrt{\Delta}y\right) - P_{p\Delta,(n+1)\Delta}(f)\left(x + b(x)\Delta\right) \\ \simeq_{\Delta\downarrow 0}\ \frac{\partial P_{p\Delta,(n+1)\Delta}(f)}{\partial x}\left(x\right)\ \sigma(x)\sqrt{\Delta}\ y\ ,\end{array}$$

which implies that

$$\begin{array}{l}\tfrac{\partial}{\partial x} P_{0,(n+1)\Delta}(f)(x) \\ \simeq_{\Delta\downarrow 0}\ \frac{1}{(n+1)} \sum_{0\leq p\leq n} \mathbb{E}\left(\frac{\partial P_{p\Delta,(n+1)\Delta}(f)}{\partial x}\left(S_{p\Delta}(x)\right)\ Y_p^2 \prod_{0\leq q<p} G_q(S_{q\Delta}(x), Y_q)\right)\ .\end{array}$$

In higher dimensions, the calculations are more involved. To analyze these models, we design a Feynman-Kac interpretation of the distributions of product of random matrices.

Using the notation $\|.\|$ for a fixed norm in $\mathbb{R}^d$, we assume that for any state U_0 in the unit sphere $\mathcal{S}^{d-1}$, we have

$$\left\| \left[\prod_{0 \leq p \leq n} A_p(S_p) \right] \; U_0 \right\| > 0 \; .$$

In this situation, we have the multiplicative formulae

$$\left[\nabla f_{n+1}(S_{n+1}) \prod_{0 \leq p \leq n} A_p(S_p) \right] \; U_0 = \left[\nabla f_{n+1}(S_{n+1}) \; U_{n+1} \right] \prod_{0 \leq p \leq n} \| A_p(S_p) \; U_p \| \; ,$$

with the well defined $\mathcal{S}^{d-1}$-valued Markov chain defined by

$$U_{n+1} = \frac{A_n(S_n) U_n}{\| A_n(S_n) U_n \|} = \frac{\left[\prod_{0 \leq p \leq n} A_p(S_p) \right] \; U_0}{\left\| \left[\prod_{0 \leq p \leq n} A_p(S_p) \right] \; U_0 \right\|} \; .$$

If we choose $U_0 = u_0$, then we obtain the following Feynman-Kac interpretation of the gradient of a semigroup

$$\nabla P_{n+1}(f_{n+1})(x) \; u_0 = \mathbb{E} \left(F_{n+1}(\mathcal{X}_{n+1}) \prod_{0 \leq p \leq n} G_p \left(\mathcal{X}_p \right) \right) \; .$$

In the above display, $\mathcal{X}_n$ is the *multivariate* Markov chain sequence

$$\mathcal{X}_n := (S_n, U_n, W_n)$$

and the functions F_{n+1} and G_n are defined by

$$F_{n+1}(x, u, w) := \nabla f_{n+1}(x) \; u \quad \text{and} \quad G_n \left(x, u, w \right) := \| \mathcal{A}_n(x, w) \; u \| \; .$$

In physics literature, the mean field particle approximations of these non commutative Feynman-Kac models are often referred as *Resampled Monte Carlo methods* [110].

2.6 American-style option pricing

2.6.1 Description of the model

Optimal stopping problems are at the heart of the theory of stochastic control. Their importance in quantitative finance is due to the large number of financial instruments with American exercises, sometimes called Bermudan exercises in the framework of discrete time models.

In this section, n stands for a fixed final time horizon, and for each $k \in \{0, \dots, n\}$, we let $\mathcal{F}_k$ denote the set of events known at time k and $\mathcal{T}_k$ the set of $\mathcal{F}_k$ - stopping times τ taking values in $\{k, \dots, n\}$. These stopping times are used to model the decision by the holder of the option to exercise it at a given time of his or her choice. The payoff is given by an adapted (*no crystal ball* can be used in this model!) stochastic process $Z = (Z_k)_{0 \le k \le n}$. For each $k \in \{0, \dots, n\}$, Z_k represents the reward to the holder for exercising the option at time k. To recast the problem in the framework used so far, we assume that the filtration $\mathcal{F} = (\mathcal{F}_k)_{0 \le k \le n}$ is generated by a Markov chain $X = (X_k)_{0 \le k \le n}$ in some measurable state space E, and that $Z_k = F_k(X_0, \dots, X_k)$ for some known deterministic functions F_k on E^{k+1}. As usual, we shall use the notation $M = (M_k)_k$ to denote the transition probability of the Markov chain X.

The Snell envelope of $(Z_k)_{0 \le k \le n}$, is the stochastic process $(U_k)_{0 \le k \le n}$ defined for any $0 \le k < n$ by the following backward equation

$$U_k = Z_k \vee \mathbb{E}(U_{k+1} | (X_0, \dots, X_k)) \ ,$$

with the terminal condition $U_n = Z_n$. The main property of this stochastic process is that

$$U_k = \sup_{\tau \in \mathcal{T}_k} \mathbb{E}(Z_\tau | (X_0, \dots, X_k)) = \mathbb{E}(Z_{\tau_k^*} | (X_0, \dots, X_k)) \qquad (2.6.1)$$

$$\text{with} \qquad \tau_k^* = \min \{k \le l \le n \ : \ U_l = Z_l\} \in \mathcal{T}_k \ .$$

Notice that $U_k \ge Z_k$, for any $0 \le k \le n$ and τ_k^* is given by the following backward formula

$$\tau_k^* = k \ \mathbf{1}_{Z_k \ge U_k} + \tau_{k+1}^* \ \mathbf{1}_{Z_k < U_k} \quad \text{with} \quad \tau_n^* = n \ .$$

To get one step further, we let $\eta_0 = \text{Law}(X_0)$ be the initial distribution on E, and we denote by $M_k(x, y)$ the elementary Markov transition of the chain X_k from E into itself.

To be more specific we also assume that

$$Z_k = F_k(X_0, \dots, X_k) := f_k(X_k) \prod_{0 \le l < k} G_l(X_l), \qquad 0 \le k \le n,$$

for some non negative space-time functions $f = (f_k)_k$ and $G = (G_k)_k$ on $\{0, 1, \cdots, n\} \times E$. In this situation, the Snell envelope process is given in terms of deterministic functions u_k through

$$U_k = u_k(X_k) \prod_{0 \le p < k} G_p(X_p),$$

where the functions u_k are given inductively by the backward functional equation

$$u_k = \mathcal{H}_{k+1}(u_{k+1}), \qquad 0 \le k < n, \qquad (2.6.2)$$

with the terminal value $u_n = f_n$, and the functional transformations

$$\mathcal{H}_{k+1}(u_{k+1}) := f_k \vee Q_{k+1}(u_{k+1}) \quad \text{with} \quad Q_{k+1}(x, dy) = G_k(x)\ M_{k+1}(x, dy)\ .$$

In the above displayed formula, $Q_{k+1}(u_{k+1})$ stands for the measurable function on E defined for any $x_k \in E$ by the conditional expectation:

$$Q_{k+1}(u_{k+1})(x) \ = \ G_k(x)\ \mathbb{E}\left(u_{k+1}(X_{k+1})|X_k = x\right) = \int\ Q_{k+1}(x, dy)\ u_{k+1}(y)\ .$$

For a detailed derivation of these formulae, we refer the interested reader to the article [49].

We let $\mathcal{H}_{k,l} = \mathcal{H}_{k+1} \circ \mathcal{H}_{k+1,l}$, with $k \leq l \leq n$, be the nonlinear semigroups associated with the backward equations (4.2.1). We use the convention $\mathcal{H}_{k,k} = I$, the identity operator, so that $u_k = \mathcal{H}_{k,l}(u_l)$, for any $k \leq l \leq n$.

If for any given sequence of bounded integral operators $(Q_k)_k$ from some state space E into itself, we denote by $Q_{k,l}$ the iterated composition operator defined by

$$Q_{k,l} := Q_{k+1} Q_{k+2} \cdots Q_l,$$

for any $k \leq l$, with the convention $Q_{k,k} = I$, then one can check that a necessary and sufficient condition for the existence of the Snell envelope $(u_k)_{0 \leq k \leq n}$ is that $Q_{k,l} f_l(x) < \infty$ for any $1 \leq k \leq l \leq n$, and any state $x \in E$. To check this claim, we simply notice that

$$f_k \leq u_k \leq f_k + Q_{k+1} u_{k+1} \qquad \forall\ 1 \leq k \leq n$$

implies that

$$f_k \leq u_k \leq \sum_{k \leq l \leq n} Q_{k,l} f_l \qquad \forall\ 1 \leq k \leq n\ . \tag{2.6.3}$$

From the readily proved Lipschitz property $|\mathcal{H}_k(u) - \mathcal{H}_k(v)| \leq Q_{k+1}\left(|u - v|\right)$, for any functions u, v on E, we also have that

$$|\mathcal{H}_{k,l}(u) - \mathcal{H}_{k,l}(v)| \leq Q_{k,l}\left(|u - v|\right)\ , \tag{2.6.4}$$

for any functions u, v on E, and any $k \leq l \leq n$.

2.6.2 A perturbation analysis

Even if it may look innocent at first, solving numerically the recursion (4.2.1) often requires extensive calculations. The major issue is to compute the conditional expectations $M_{k+1}(u_{k+1})$ on the whole state space E, at every time step $0 \leq k < n$.

For Markov chain models taking values in some finite state spaces, the above expectations can be computed by systematic backward inspection of the realization

tree that lists all possible outcomes and every transition of the chain. For excessively large state spaces, or more general situations, we need to resort to approximation strategies.

Over the last two decades, several approximation methodologies have been proposed, including Carriere-Longstaff-Schwartz's functional regression style methods [26, 32, 88, 108], refined singular values decomposition strategies [14], Monte Carlo simulation methods [19, 58, 49, 50, 87], and the quantization grid technology developed by Pagès and his co-authors [93, 94, 95, 92, 96].

Most of the numerical approximation schemes amount to replacing the pair $(f_k, Q_k)_{0\leq k\leq n}$ by some approximation model $(\widehat{f}_k, \widehat{Q}_k)_{0\leq k\leq n}$ on some possibly reduced finite subsets $\widehat{E} \subset E$. We let $\widehat{u}_k$ be the Snell envelope on $\widehat{E}_k$ associated with the functions $\widehat{f}_k$ and the sequence of transition operators $\widehat{M}_k$ from $\widehat{E}$ into itself.

$$\widehat{u}_k = \widehat{\mathcal{H}}_{k+1}(\widehat{u}_{k+1}) := \widehat{f}_k \vee \widehat{Q}_{k+1}(\widehat{u}_{k+1}) \ . \tag{2.6.5}$$

Let also $\widehat{\mathcal{H}}_{k,l} = \widehat{\mathcal{H}}_{k+1} \circ \widehat{\mathcal{H}}_{k+1,l}$ with $k \leq l < n$ be the nonlinear semigroups associated with the backward equations (3.2.5) so that $\widehat{u}_k = \widehat{\mathcal{H}}_{k,l}(\widehat{u}_l)$ for any $k \leq l \leq n$. Using the elementary inequality $|a \vee a' - b \vee b'| \leq |a - b| + |a' - b'|$ which is valid for any $a, a', b, b' \in \mathbb{R}$, for any $0 \leq k < n$ and for any functions u on E_{k+1} one readily obtains the local approximation inequality

$$\left|\mathcal{H}_{k+1}(u) - \widehat{\mathcal{H}}_{k+1}(u)\right| \leq |f_k - \widehat{f}_k| + |(Q_{k+1} - \widehat{Q}_{k+1})(u)| \ . \tag{2.6.6}$$

To transfer these local estimates to the semigroups $\mathcal{H}_{k,l}$ and $\widehat{\mathcal{H}}_{k,l}$ we use a perturbation analysis. The difference between the approximate and the exact Snell envelope can be written as a telescoping sum

$$u_k - \widehat{u}_k = \sum_{l=k}^{n} \left[\widehat{\mathcal{H}}_{k,l}(\mathcal{H}_{l+1}(u_{l+1})) - \widehat{\mathcal{H}}_{k,l}(\widehat{\mathcal{H}}_{l+1}(u_{l+1}))\right] \ ,$$

setting for simplicity $\mathcal{H}_{n+1}(u_{n+1}) = u_n$ and $\widehat{\mathcal{H}}_{n+1}(u_{n+1}) = \widehat{u}_n$, for $l = n$. Combining the Lipschitz property (3.2.4) of the semigroup $\widehat{\mathcal{H}}_{k,l}$ with the local estimate (3.2.6), one gets the final estimates:

$$|u_k - \widehat{u}_k| \ \leq \ \sum_{l=k}^{n} \widehat{Q}_{k,l}|f_l - \widehat{f}_l| + \sum_{l=k}^{n-1} \widehat{Q}_{k,l}|(Q_{l+1} - \widehat{Q}_{l+1})u_{l+1}| \ .$$

The perturbation analysis of nonlinear semigroups discussed above is a natural and fundamental tool for the analysis of the Snell envelope approximations. It can be used sequentially, and without further work, to obtain non asymptotic estimates for models combining several levels of approximations. In the same vein, and whenever possible, it can also be used as a technical tool to reduce the analysis of Snell approximation models on compact state spaces or even on finite but possibly large

quantization trees or Monte Carlo grids. This perturbation analysis is clearly not new, it has been used with success in [41, 48, 85, 109] in the context of nonlinear filtering semigroups and particle approximation models. In the context of optimal stopping problems and numerical quantization schemes, these techniques were also used for instance in the papers of Egloff [64] and Gobet, Lemor and Warin [76] or Pagès [92]. To the best of our knowledge, the general and abstract formulation given above has first been presented in the recent article [50].

2.6.3 Particle approximations

In this subsection, we focus on a type of Monte Carlo importance sampling scheme which is a version called *average density* of the Stochastic Mesh schemes proposed by Broadie and Glasserman in [19]. The formulation of this algorithm in terms of interacting particles was crucial to derive precise convergence results in [50].

We let η_n be the normalized Feynman-Kac measures defined in (2.3.2). By (2.3.10), we have that

$$\eta_{k+1} = \Psi_{G_k}(\eta_k)\, M_{k+1} \ .$$

Now, we assume that the Markov transitions M_k have a density H_k with respect to some reference measure λ_k

$$M_{k+1}(x, dy) = H_k(x, y)\ \lambda_k(dy) \ .$$

Under this assumption, we can rewrite $Q_{k+1}(u_{k+1})(x)$ as follows

$$Q_{k+1}(u_{k+1})(x) = \eta_k(G_k) \int \eta_{k+1}(y)\ \frac{G_k(x)H_{k+1}(x,y)}{\int \eta_k(dz)G_k(z)H_{k+1}(z,y)}\ u_{k+1}(y) \ , \qquad (2.6.7)$$

and as before, we let

$$\eta_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i} \to_{N\to\infty} \eta_n$$

be the particle approximation of the measures η_n defined in Section 2.3.2 . We denote by $\widehat{Q}_{k+1}$ the matrix obtained by replacing the measures η_k by their N-particle approximations:

$$\begin{aligned}
\widehat{Q}_{k+1}(f)(x) &:= \eta_k^N(G_k) \int \eta_{k+1}^N(y)\ \frac{G_k(x)H_{k+1}(x,y)}{\int \eta_k^N(dz)G_k(z)H_{k+1}(z,y)}\ f(y) \\
&= \eta_k^N(G_k) \sum_{j=1}^N \frac{G_k(x)H_{k+1}(x,\xi_{k+1}^j)}{\sum_{j'=1}^N G_k(\xi_k^{j'})H_{k+1}(\xi_k^{j'},\xi_{k+1}^j)}\ f(\xi_{k+1}^j) \ ,
\end{aligned}$$

for any test function f on E. Notice that these expressions are easily computed (with computational cost N^2) at any state ξ_k^i of the k-th population when the

values $f(\xi_{k+1}^j)$ of the function f are known

$$\widehat{Q}_{k+1}(f)(\xi_k^i) \;=\; \eta_k^N(G_k)\;\sum_{j=1}^N\;\frac{G_k(\xi_k^i)H_{k+1}(\xi_k^i,\xi_{k+1}^j)}{\sum_{j'=1}^N G_k(\xi_k^{j'})H_{k+1}(\xi_k^{j'},\xi_{k+1}^j)}\;f(\xi_{k+1}^j)\;.$$

By (3.2.5) the corresponding backward particle approximation of the Snell envelope is given by the following equations, for $i=1,\cdots N$,

$$\widehat{u}_k(\xi_k^i)=f_k(\xi_k^i)\vee\left(\eta_k^N(G_k)\;\sum_{j=1}^N\;\frac{G_k(\xi_k^i)H_{k+1}(\xi_k^i,\xi_{k+1}^j)}{\sum_{j'=1}^N G_k(\xi_k^{j'})H_{k+1}(\xi_k^{j'},\xi_{k+1}^j)}\;\widehat{u}_{k+1}(\xi_{k+1}^j)\right)\;.$$

Also notice that the values $\widehat{u}_k(x)$ on any state x can be computed using the formula

$$\widehat{u}_k(x)=f_k(x)\vee\left(\eta_k^N(G_k)\;\sum_{j=1}^N\;\frac{G_k(x)H_{k+1}(x,\xi_{k+1}^j)}{\sum_{j'=1}^N G_k(\xi_k^{j'})H_{k+1}(\xi_k^{j'},\xi_{k+1}^j)}\;\widehat{u}_{k+1}(\xi_{k+1}^j)\right)\;.$$

For a thorough discussion on these particle models, their convergence analysis, and a variety of related approximation grid type models, we refer the reader to the pair of articles [49, 50]. In particular, this formalization allows to prove that the L_p mean error induced by this version of Stochastic Mesh approximation vanishes, under mild assumptions, with a rate $1/\sqrt{N}$. Also, a new Monte Carlo approximation scheme is proposed in [50] using simulations of a genealogical tree with neutral selections and mutations associated with a discrete-space Markov chain approximating the price dynamics. The main advantage of this new scheme is the fact that the computational effort of the algorithm is linear in the number of sampled points, as opposed to quadratic as for the Stochastic Mesh scheme.

2.7 Pricing models with partial observation models

Managing large portfolios and pricing financial instruments under partial observations are quite common problems in quantitative finance. See for instance the series of articles [83, 98, 100, 101, 111], and references therein. The case of stochastic volatility models is the epitome of these situations: one can *more or less* observe stock prices but not the evolution of the stochastic volatility.

2.7.1 Abstract formulation and particle approximation

We work in discrete time and we recast the dynamical financial model in the framework of hidden Markov models. The basic object is a pair process $(X,Y)=$

$((X_n, Y_n))_n$ forming a Markov chain on some product space $E^X \times E^Y$ with elementary transitions given

$$\mathbb{P}\left((X_n, Y_n) \in d(x,y) \mid (X_{n-1}, Y_{n-1})\right) = M_n(X_{n-1}, dx) \times g_n(x,y)\ \lambda_n(dy)\ , \quad (2.7.1)$$

for some positive likelihood function g_n, and some reference probability measure λ_n on E^Y. According to our setup throughout the paper, the marginal process $X = (X_n)_n$ is also assumed to be a Markov chain and as usual, we denote by M_n its transition probability. We can think of X_n as a vector of prices and random factors (instantaneous volatility could be one of them), and Y_n a vector of observations of quantities derived from the components of X_n. We can also consider X_n as a stochastic volatility model, and Y_n the stock price observations. For the sake of definiteness, we choose to illustrate the particle methods on a pricing problem, so we assume that we are given a European payoff function $f_n(X_n, Y_n)$ for each time $n \geq 0$. The price of the contingent claim is given at time $p \leq n$ by:

$$V_{p,n}(f_n) := \mathbb{E}\left[f_n(X_n, Y_n) \prod_{p \leq q < n} G'_q(X_q, Y_q) \mid (Y_0, \ldots, Y_p) \right] .$$

for some non negative functions G'_p related to barrier sets or stochastic interest rates, as explained in Section 2.2.3. It is important to observe that the conditional expectations

$$U_{p,n}(f_n)(x,y) := \mathbb{E}\left[f_n(X_n, Y_n) \prod_{p \leq q < n} G'_q(X_q, Y_q) \mid (X_p, Y_p) = (x,y) \right]$$

have the same form as the Feynman-Kac definitions of the measures introduced in (2.3.1), with the reference Markov chain (X_q, Y_q), from the initial time $q = p$, starting from $(X_p, Y_p) = (x,y)$ at time p. For any starting point $(X_p, Y_p) = (x,y)$, these unnormalized distributions can be approximated by running an N-particle model on $(E^X \times E^Y)$, with selection potential functions G'_q. We denote by $U^N_{p,n}(f_n)(x,y)$ the corresponding *unbiased* particle approximation. Fix an observation sequence $Y = y$, and consider the Feynman-Kac models (2.3.1) associated with the likelihood potential functions:

$$G_p(x) := g_p(x, y_p) \qquad 0 \leq p \leq n\ .$$

To emphasize the dependence of the Feynman-Kac measures on the observation sequence, we use the notations

$$\eta_{n+1}^{[y_0,\ldots,y_n]} \quad \text{and} \quad \gamma_{n+1}^{[y_0,\ldots,y_n]} \quad (2.7.2)$$

for the normalized and unnormalized measures associated with the series of observations $Y_p = y_p$, for $0 \leq p \leq n$. These conditional distributions can be approximated

using an N-particle model on E^X, with selection potential functions G_q. We denote by

$$\eta_{n+1}^{([y_0,\ldots,y_n],N)} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_{n+1}^{([Y_0,\ldots,Y_{n-1}],i)}}$$

the empirical measures providing the particle approximation. Notice that

$$\mathrm{Law}(X_n \mid Y_p = y_p,\ 0 \le p < n) = \eta_n^{[y_0,\ldots,y_{n-1}]} \tag{2.7.3}$$

and by the Bayes rule

$$\mathrm{Law}(X_n \mid Y_p = y_p,\ 0 \le p \le n) := \Psi_{g_n(\bullet,y_n)}\left(\eta_n^{[y_0,\ldots,y_{n-1}]}\right) = \widehat{\eta}_n^{[y_0,\ldots,y_n]}\ , \tag{2.7.4}$$

and by construction, we have:

$$V_{p,n}(f_n) = \int \Psi_{g_p(\bullet,Y_p)}\left(\eta_n^{[Y_0,\ldots,Y_{n-1}]}\right)(dx)\ U_{p,n}(f_n)(x,Y_p)\ ,$$

and these quantities can be approximated combining the particle estimates defined above. Indeed, we have that

$$V_{p,n}(f_n) \simeq_{N\uparrow\infty} V_{p,n}^N(f_n)\ ,$$

with

$$\begin{aligned} V_{p,n}^N(f_n) &:= \int \Psi_{g_p(\bullet,Y_p)}\left(\eta_n^{([Y_0,\ldots,Y_{n-1}],N)}\right)(dx)\ U_{p,n}^N(f_n)(x,Y_p) \\ &= \sum_{i=1}^{N} \frac{g_p(\xi_p^{([Y_0,\ldots,Y_{p-1}],i)},Y_p)}{\sum_{j=1}^N g_p(\xi_p^{([Y_0,\ldots,Y_{p-1}],j)},Y_p)}\ U_{p,n}^N(f_n)(\xi_p^{([Y_0,\ldots,Y_{p-1}],i)},Y_p)\ . \end{aligned}$$

2.7.2 Optimal stopping with partial observation

We work with the setup of a pair (X_n, Y_n) Markov chain model introduced in the previous section. According to our discussion in Section 2.6.1, the Snell envelop associated with an American option with finite maturity n, payoffs $Z_k = f_k(X_k, Y_k)$ is given by

$$U_k := \sup_{\tau\in\mathcal{T}_k^Y} \mathbb{E}(f_\tau(X_\tau,Y_\tau)|(Y_0,\ldots,Y_k))\ ,$$

where $\mathcal{T}_k^Y$ stands for the set of all $\mathcal{F}_k^Y$ - stopping times τ taking values in $\{k,\ldots,n\}$, where the filtration is know given by the sigma fields $\mathcal{F}_k^Y$ generated by the observation sequence Y_p, from $p=0$ up to the time k. We denote by $\eta_n^{[y_0,\ldots,y_{n-1}]}$ and $\widehat{\eta}_n^{[y_0,\ldots,y_n]}$ the conditional distributions defined in (2.7.3) and (2.7.4). With these notations, for any $0 \le k \le n$ we have that

$$\begin{aligned} \mathbb{E}(f_\tau(X_\tau,Y_\tau)|(Y_0,\ldots,Y_k)) &= \mathbb{E}\left(\sum_{p=k}^{n} \mathbf{1}_{\tau=p}\ \mathbb{E}\left(f_p(X_p,Y_p) \mid (Y_0,\ldots,Y_p)\right) \mid (Y_0,\ldots,Y_k)\right) \\ &= \mathbb{E}\left(F_\tau\left(Y_\tau,\widehat{\eta}_\tau^{[Y_0,\ldots,Y_\tau]}\right) \mid (Y_0,\ldots,Y_k)\right)\ , \end{aligned} \tag{2.7.5}$$

with the conditional payoff function

$$F_p\left(Y_p, \widehat{\eta}_p^{[Y_0,\ldots,Y_p]}\right) = \int \widehat{\eta}_p^{[Y_0,\ldots,Y_p]}(dx)\ f_p(x, Y_p)\ .$$

It is well known that

$$\mathcal{X}_p := \left(X_p, Y_p, \widehat{\eta}_p^{[Y_0,\ldots,Y_p]}\right)$$

is a Markov chain with elementary transitions defined by, for any integrable function F on product space $E^X \times E^Y \times \mathcal{P}(E^X)$,

$$\mathbb{E}\left[F\left(X_p, Y_p, \widehat{\eta}_p^{[Y_0,\ldots,Y_p]}\right)\ \middle|\ \left(X_{p-1}, Y_{p-1}, \widehat{\eta}_{p-1}^{[Y_0,\ldots,Y_{p-1}]}\right) = (x, y, \mu)\right]$$

$$= \int\int \lambda_p(dy_p)\ M_p\left(x, dx_p\right) g_p(x_p, y_p)\ F\left(x_p, y_p, \Psi_{g_p(\bullet, y_p)}\left(\mu M_p\right)\right)\ .$$

A proof of this assertion can be found in any textbook on advanced stochastic filtering. For instance, the book of W. Runggaldier and L. Stettner [106] provides a detailed treatment of discrete time partially observed models, their non linear filtering, and related partially observed control problems.

Roughly speaking, using Bayesian notation, we have

$$\begin{aligned}
\eta_p^{[y_0,\ldots,y_{p-1}]}(dx_p) &= dp_p(x_p \mid (y_0, \ldots, y_{p-1})) \\
&= \int dp_p(x_p \mid x_{p-1}) \times p_n(x_{p-1} \mid (y_0, \ldots, y_{p-1})) \\
&= \widehat{\eta}_{p-1}^{[y_0,\ldots,y_{p-1}]} M_p(dx_p)
\end{aligned}$$

and

$$\Psi_{g_p(\bullet, y_p)}\left(\widehat{\eta}_{p-1}^{[y_0,\ldots,y_{p-1}]} M_p\right)(dx_p)$$

$$= \frac{p(y_p|x_p)}{\int p_p(y_p \mid x'_p)\ dp_p(x'_p \mid (y_0, \ldots, y_{p-1}))}\ dp_p(x_p \mid (y_0, \ldots, y_{p-1}))$$

$$= dp_p(x_p \mid (y_0, \ldots, y_{p-1}, y_p))\ ,$$

from which we can prove that

$$\begin{aligned}
\mu M_p(g_p(\bullet, y_p)) &= \int p_p(y_p \mid x_p)\ dp_p(x_p \mid (y_0, \ldots, y_{p-1})) \\
&= p_p(y_p \mid (y_0, \ldots, y_{p-1}))
\end{aligned}$$

and

$$\Psi_{g_p(\bullet, y_p)}\left(\mu M_p\right) = \widehat{\eta}_p^{[y_0,\ldots,y_p]}\ ,$$

as long as $\mu = \widehat{\eta}_{p-1}^{[y_0,\ldots,y_{p-1}]}$ $\left(\Rightarrow \mu M_p = \eta_p^{[y_0,\ldots,y_{p-1}]}\right)$.

The above discussion suggests the following interpretation. We can rewrite (2.7.5) as the Snell envelop of a fully observed *augmented* Markov chain:

$$\mathbb{E}(f_\tau(X_\tau, Y_\tau)|(Y_0, \ldots, Y_k)) = \mathbb{E}\left(F_\tau\left(\mathcal{X}_\tau\right) \mid (\mathcal{X}_0, \ldots, \mathcal{X}_k)\right) .$$

This Markov chain $\mathcal{X}_n$ takes values in an infinite dimensional state space, and it can rarely be sampled without some addition level of approximation. Therefore, most of the grid or Monte Carlo simulation based techniques for solving these models require the introduction of a specific grid approximation of conditional distributions, or judicious approximation sampling schemes. The particle methodology advocated in this chapter provides a natural strategy. Using the particle approximations discussed in Section 2.3.2, we can replace the chain $\mathcal{X}_n$ by the N-particle approximation defined by

$$\mathcal{X}_n^N := \left(Y_p, \widehat{\eta}_p^{([Y_0,\ldots,Y_p],N)}\right) ,$$

where

$$\widehat{\eta}_p^{([Y_0,\ldots,Y_p],N)} := \Psi_{g_p(\,\cdot\,,Y_p)}\left(\widehat{\eta}_{p-1}^{([Y_0,\ldots,Y_{p-1},N)]}\right)$$

stands for the updated measure associated with the particle scheme associated with the likelihood selection functions $g_p(\,\cdot\,, Y_p)$. The corresponding N-particle approximation of the Snell envelop is now given by

$$\mathbb{E}(f_\tau(X_\tau, Y_\tau)|(Y_0, \ldots, Y_k)) \simeq_{N\uparrow\infty} \mathbb{E}\left(F_\tau\left(\mathcal{X}_\tau^N\right) \mid (\mathcal{X}_0^N, \ldots, \mathcal{X}_k^N)\right) .$$

In this setup, the approximated optimal stopping problem requires the computation of the quantities

$$U_k^N := \sup_{\tau \in \mathcal{T}_k^N} \mathbb{E}\left(F_\tau\left(\mathcal{X}_\tau^N\right) \mid (\mathcal{X}_0^N, \ldots, \mathcal{X}_k^N)\right) ,$$

where $\mathcal{T}_k^N$ stands for the set of $\mathcal{F}_k^{\mathcal{X}_k^N}$ - stopping times τ taking values in $\{k, \ldots, n\}$, where the filtration is formed by the sigma fields generated by the Markov chain random variables $\mathcal{X}_k^N$, from $p = 0$ up to the current time k.

We close this section with an alternative representation in terms of the unnormalized filters $\gamma_n^{[Y_0,\ldots,Y_{n-1}]}$ defined in (2.7.2). We let $\mathbb{P}_n$ be the probability distribution of a Markov chain $(X_p, Y_p)_{0\leq p\leq n}$ defined in (2.7.1), and $\mathbb{P}_n^{(0)}$ the probability distribution of the Markov chain $(X_p, Y_p)_{0\leq p\leq n}$ with independent random observations Y_p with distribution λ_p with $p \leq n$, also assume to be *independent of the chain* $(X_p)_{0\leq p\leq n}$,. By construction, $\mathbb{P}_n$ is absolutely continuous with respect to $\mathbb{P}_n^{(0)}$, and its Radon-Nykodym derivative is given by:

$$\frac{d\mathbb{P}_n}{d\mathbb{P}_n^{(0)}} = \prod_{0\leq p\leq n} g_p(X_p, Y_p) .$$

Now, for any $\tau \in \mathcal{T}_0^Y$ we observe that

$$\mathbb{E}(f_\tau(X_\tau, Y_\tau)) = \mathbb{E}^{(0)}\left(\sum_{p=0}^{n} \mathbf{1}_{\tau=p}\, \mathbb{E}^{(0)}\left(f_p(X_p, Y_p) \prod_{0\le q\le p} g_q(X_q, Y_q) \mid (Y_0, \ldots, Y_p)\right)\right)$$

and

$$\begin{aligned}
&\mathbb{E}^{(0)}\left(f_p(X_p, Y_p)\ \textstyle\prod_{0\le q\le p} g_q(X_q, Y_q) \mid (Y_0, \ldots, Y_p)\right) \\
&= \gamma_p^{[Y_0,\ldots,Y_{p-1}]}\left(f_p(\boldsymbol{.}, Y_p) g_p(\,\cdot\,, Y_p)\right) \\
&= \eta_p^{[Y_0,\ldots,Y_{p-1}]}\left(f_p(\boldsymbol{.}, Y_p) g_p(\,\cdot\,, Y_p)\right) \times \textstyle\prod_{0\le q<n} \eta_q^{[Y_0,\ldots,Y_{q-1}]}\left(g_q(\,\cdot\,, Y_q)\right)\ .
\end{aligned}$$

The last assertion is a direct consequence of the multiplicative formula (2.3.3) for unnormalized Feynman-Kac measures. Arguing as above, we introduce the Markov chain

$$\mathcal{X}_n := \left(Y_n, \eta_n^{[Y_0,\ldots,Y_{n-1}]}\right)$$

the payoff and the potential functions

$$F_n(\mathcal{X}_n) := \eta_n^{[Y_0,\ldots,Y_{n-1}]}\left(f_n(\boldsymbol{.}, Y_n) g_n(\boldsymbol{.}, Y_n)\right) \quad \text{and} \quad \mathcal{G}_n(\mathcal{X}_n) := \eta_n^{[Y_0,\ldots,Y_{n-1}]}\left(g_n(\boldsymbol{.}, Y_n)\right)\ .$$

By construction, we have

$$\mathbb{E}\left(f_\tau(X_\tau, Y_\tau)\right) = \mathbb{E}^{(0)}\left(F_\tau(\mathcal{X}_\tau) \prod_{0\le p<\tau} \mathcal{G}_p(\mathcal{X}_p)\right)\ .$$

We have now reduced the optimal stopping problem with partial observations to a conventional optimal stopping problem of a measure valued Markov chain $\mathcal{X}_n$ with *stochastic potential functions* $\mathcal{G}_p(\mathcal{X}_p)$, and independent random observations sequences. Once more, using the particle approximation models discussed in Section 2.3.2, we can replace the chain $\mathcal{X}_n$ by the N-particle approximation model defined by

$$\mathcal{X}_n^N := \left(Y_p, \eta_p^{([Y_0,\ldots,Y_p],N)}\right)\ .$$

Here again, we have turned a complex optimal stopping problem under partial observations into an almost equivalent optimal stopping problem of an *easy to sample* Markov chain sequence of the same form as the one discussed in Section 2.6.1. These particle transformations can also be used for more general stochastic control problems with partial observations. We refer the reader to [12, 11, 106, 60, 89, 91, 99] for a more thorough discussion on this subject.

2.7.3 Parameter estimation in hidden Markov chain models

In many economic and financial applications, the parameters are unknown and must be estimated from partial and noisy observations. This situation is typical of hidden Markov chain problems which arise in a variety of domains, ranging from signal

processing, medical Bayesian inference, communication and information theory. For an overview of some of the problems occurring in finance and econometrics we refer the reader to [8, 33, 34, 37, 82], to mention only a few.

As in (2.7.1) these models are framed in terms of a signal-observation type pair Markov chain $(X_n, Y_n)_n$ with a collection of transition probabilities $M_{\theta,n}$ and likelihood functions $g_{\theta,n}$ that depend upon the realization of a random parameter $\Theta = \theta$ taking values in some state space S, equipped with a probability measure μ. We also denote by $\eta_{\theta,0}$ the conditional distribution of X_0 given $\Theta = \theta$.

The example we have in mind is the quintessential calibration problem in partially observed models arising in computational finance. One way to set up a stochastic volatility model as a filtering problem in discrete time is to choose $(X_n, Y_n) = (\sigma_n, S_n)$. In this case $X_n = \sigma_n$ represents the instantaneous stochastic volatility, and the observation $Y_n = S_n$ is given by the price of the asset. In most practical applications, the evolution of these quantities is given by a parametric model of the form:

$$\begin{aligned} \sigma_k &= F^1_{\theta,n}(\sigma_{n-1}, S_{n-1}, W^1_n) \\ S_k &= F^2_{\theta,n}(\sigma_{n-1}, S_{n-1}, W^2_n), \end{aligned}$$

where $F^1_{\theta,n}$ and $F^2_{\theta,n}$ are functions depending upon some unknown parameter θ. The objective is to compute the conditional distribution $\mathrm{Law}(\theta | S_0, \ldots, S_n)$ of θ given the observations of the price. To be more specific, we can precise our illustration by choosing the popular Heston's stochastic volatility model. In our framework, this model is given by:

$$\begin{aligned} F^1_{\theta,n}(\sigma_{n-1}, S_{n-1}, W^1_n) &= (ab + (1-a)\sigma_{n-1})\Delta t + c\sqrt{\sigma_{n-1}}\Delta W^1_n \\ F^2_{\theta,n}(\sigma_{n-1}, S_{n-1}, W^2_n) &= S_{k-1}(1+d)\Delta t + S_{k-1}\sqrt{\sigma_{n-1}}\Delta W^2_n, \end{aligned}$$

where $\theta = (a, b, c, d)$ is the collection of parameters to calibrate, and $(W^i_n)_{i=1,2}$ are independent Brownian motions.

Using the notations of Section 2.3.3, the conditional distribution of the random path $(X_0, \ldots, X_n)$, given $\Theta = \theta$, and the sequence $(Y_0, \ldots, Y_{n-1}) = (y_0, \ldots, y_{n-1})$ of observations is given by the Feynman-Kac measures

$$\mathbb{Q}_{\theta,n}(dx_0 \times \cdots \times dx_n) = \frac{1}{\mathcal{Z}_n(\theta)} \left\{ \prod_{0 \le q < n} G_{\theta,q}(x_q) \right\} \mathbb{P}_{\theta,n}(dx_0 \times \cdots \times dx_n) \ ,$$

with the potential functions $G_{\theta,q}(x_q) = g_{\theta,n}(x_q, y_q)$, and the conditional distribution $\mathbb{P}_{\theta,n}$ of the random path $(X_0, \ldots, X_n)$ given $\Theta = \theta$. As in (2.3.3), the normalizing constants $\mathcal{Z}_n(\theta)$ are given by the multiplicative formula

$$\mathcal{Z}_n(\theta) = \prod_{0 \le p < n} \eta^{[y_0, \ldots, y_{p-1}]}_{\theta,p}(G_{\theta,p}) \ ,$$

with the p-th marginal distributions $\eta_{\theta,p}^{[y_0,\ldots,y_{p-1}]}$ of the measure $\mathbb{Q}_{\theta,n}$ i.e. the conditional distribution of the random variable X_p given $\Theta = \theta$, and the sequence $(Y_0,\ldots,Y_{p-1}) = (y_0,\ldots,y_{p-1})$ of observations. In the Bayesian literature, the normalizing constants $\mathcal{Z}_{n+1}(\theta)$ are often called the likelihood functions of the parameter θ, given the observation data $(y_0,\ldots,y_n)$, and they are denoted by $p(y_0,\ldots,y_n \mid \theta)$ to emphasize that they are given by the conditional density of the observations given the unknown parameter.

Above analysis are based on the fact that W_n^1 and W_n^2 are independent, otherwise the conditional distribution can be given by approximate Bayesian computation techniques (see, for instance, [46, 52, 53] and section 2.5.2 in [51]). The authors prove that the approximate Bayesian computation models are also under the framework of Feynman-Kac measures.

In this section, the observation sequence $Y = y$ is fixed, so in order to streamline the notations we suppress the superscript $^{[y_0,\ldots,y_n]}$and write $\eta_{\theta,p}$ and $\widehat{\eta}_{\theta,p} = \Psi_{G_{\theta,n}}(\eta_{\theta,p})$ for the one step predictor $\eta_{\theta,p}^{[y_0,\ldots,y_{p-1}]}$ and the optimal filter $\widehat{\eta}_{\theta,p}^{[y_0,\ldots,y_p]}$.

From the previous discussion, it should be clear that the conditional distributions of the parameter Θ with respect to the sequence of observations $(Y_0,\ldots,Y_{n-1}) = (y_0,\ldots,y_{n-1})$ is given by the measures

$$\mu_n(d\theta) := \frac{1}{\mathcal{Z}_n}\left(\prod_{0\le p<n} h_p(\theta)\right)\mu(d\theta) \quad \text{with the functions} \quad h_p(\theta) = \eta_{\theta,p}(G_{\theta,p}) \tag{2.7.6}$$

for some normalizing constant $\mathcal{Z}_n$. In the Bayesian literature, the likelihood functions $h_p(\theta)$ are often denoted by $p(y_p \mid (y_0,\ldots,y_{p-1}),\theta)$. In some instances, such as classical linear-Gaussian models for example, the local likelihood functions $h_p(\theta)$ can be computed explicitly in terms of Gaussian densities and optimal one-step predictors given by the Kalman recursions. In this case, we can use a dedicated Monte Carlo Markov Chain model (MCMC for short) algorithm to sample from the Boltzmann-Gibbs measures (2.7.6). One can also turn this MCMC algorithm into an interacting MCMC model. This is done by letting K_n be a MCMC transition with target measure $\mu_n = \mu_n K_n$. By definition of the Boltzmann-Gibbs transformation (2.3.7), we readily see that

$$\mu_{n+1} = \Psi_{h_n}(\mu_n) \Rightarrow \mu_{n+1} = \Psi_{h_n}(\mu_n)K_{n+1},$$

which shows that μ_n is given by the normalized Feynman-Kac measure defined for any measurable function f on S, by the following equation

$$\mu_n(f) \propto \mathbb{E}\left(f(\Theta_n)\prod_{0\le p<n} h_p(\Theta_p)\right),$$

where Θ_n is a Markov chain on S with initial distribution $\mu_0 = \mu$ and Markov transitions K_n. The interacting particle approximation

$$\mu_n^N = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\Theta_n^i}$$

of the measures μ_n (and their normalizing constants) is a genetic type particle model on the product space S^N

$$\Theta_n = \left(\Theta_n^i\right)_{1 \leq i \leq N} \in S^N \xrightarrow{\text{selection}} \widehat{\Theta}_n = \left(\widehat{\Theta}_n^i\right)_{1 \leq i \leq N} \in S^N \xrightarrow{\text{mutation}} \Theta_{n+1} \in S^N \ . \tag{2.7.7}$$

The mutation transitions are given by the MCMC transitions K_n, and the selection transitions are obtained from the selection potential functions h_n. The complete conditional distribution of the random sequence $(\Theta, (X_0, \ldots, X_n))$ given the sequence of observations $(Y_0, \ldots, Y_{n-1}) = (y_0, \ldots, y_{n-1})$ is given by the Feynman-Kac measures

$$\mu_n(d\theta) \times \mathbb{Q}_{\theta,n}(d(x_0, \ldots, x_n)) \simeq_{N \uparrow \infty} \mu_n^N(d\theta) \times \mathbb{Q}_{\theta,n}^N(d(x_0, \ldots, x_n)) \ .$$

The measures $\mathbb{Q}_{\theta,n}^N$ appearing in the above right hand side stand for the particle backward model defined in (2.3.19). Alternatively, we can also use the genealogical tree approximation discussed in (2.3.3).

For linear-Gaussian models, we emphasize that the measure $\mathbb{Q}_{\theta,n}$ can be computed explicitly. More precisely, the backward Markov chain formula (2.3.18) can be computed using the updating transition of the Kalman filter, with the Gaussian likelihood density function $H_{\theta,n+1}$ of the transition $M_{\theta,n+1}$. In this case, (2.3.18) is the backward product of the Gaussian transitions given below

$$\mathbb{M}_{\theta,n+1,\eta_{\theta,n}}(x_{n+1}, dx_n) := \frac{H_{\theta,n+1}(x_n, x_{n+1})}{\widehat{\eta}_{\theta,n}\left(H_{\theta,n+1}(\cdot, x_{n+1})\right)} \ \widehat{\eta}_{\theta,n}(dx_n) \ .$$

When the local likelihood functions h_n are not known, we need to add another approximation level. To this end, we also consider the probability distribution $P(\theta, d\xi)$ of the N-particle model

$$\xi_\theta := (\xi_{\theta,0}, \xi_{\theta,1}, \ldots, \xi_{\theta,T}) \ ,$$

on the interval $[0, T]$, with mutation transitions $M_{\theta,n}$, and potential selection functions $G_{\theta,n}$, with $n \leq T$. We fix a large time horizon T, and for any $0 \leq n \leq T$, we set

$$\overline{\mu}_n(d(\xi, \theta)) = \frac{1}{\overline{\mathcal{Z}}_n} \left\{ \prod_{0 \leq p < n} \overline{h}_p(\xi, \theta) \right\} \overline{\mu}(d(\xi, \theta)) \ , \tag{2.7.8}$$

for some normalizing constants $\overline{\mathcal{Z}}_n$, the reference measure $\overline{\mu}$ being given by

$$\overline{\mu}(d(\xi, \theta)) = \mu(d\theta) \ P(\theta, d\xi),$$

and the potential functions h_n on the product space $\left(\left(\prod_{0\leq p\leq T} E_p^N\right)\times S\right)$ defined by

$$\overline{h}_n(\xi,\theta)=\frac{1}{N}\sum_{1\leq i\leq N} G_{\theta,n}(\xi^i_{\theta,n})=\eta^N_{\theta,n}\left(G_{\theta,n}\right)\in(0,\infty)\ .$$

Firstly, we observe that these target measures have the same form as the Boltzmann-Gibbs measures (2.7.6). Thus, they can be sampled using the MCMC or the interacting MCMC methodologies discused above. For a detailed discussion these types of sophisticated serial MCMC methodologies, we refer the reader to the recent article [2].

More interestingly, using the unbiased property of the unnormalized particle models presented in (2.3.23), we clearly have $\overline{\mathcal{Z}}_n=\mathcal{Z}_n$ and

$$\begin{aligned}\int P(\theta,d\xi)\ \left\{\prod_{0\leq p<n}\overline{h}_p(\xi,\theta)\right\} &= \mathbb{E}\left(\prod_{0\leq p<n}\eta^N_{\theta,p}\left(G_{\theta,p}\right)\right)\\ &= \prod_{0\leq p<n}\eta_{\theta,p}(G_{\theta,p})\ ,\end{aligned}$$

from which we conclude that the Θ-marginal of $\overline{\mu}_n$ coincides with the desired target measure

$$\left(\overline{\mu}_n\circ\Theta^{-1}\right)(d\theta)=\mu_n(d\theta)=\frac{1}{\mathcal{Z}_n}\ \left\{\prod_{0\leq p<n}\eta_{\theta,p}(G_{\theta,p})\right\}\ \nu(d\theta)\ .$$

Chapter 3

On the robustness of the Snell envelope

3.1 Introduction

The evaluation of optimal stopping time of random processes, based on a given optimality criterion, is one of the major problems in stochastic control and optimal stopping theory, particularly in financial mathematics with American options pricing and hedging. The present chapter is restricted to the case of the discrete time optimal stopping problem corresponding in finance to the case of Bermudan options.

It is well known that the price of Bermudan options giving the opportunity to exercise a payoff f_k at discrete dates $k = 0, \cdots, n$, can be calculated by a backward dynamic programming formula. This recursion consists in comparing at each time step k the immediate payoff f_k and the expectation of the future gain (or the so-called continuation value), which precisely involves the Markov transition M_{k+1} of the underlying assets process (X_k).

The first objective of this chapter is to provide a simple framework to analyze in unison most of the numerical schemes currently used in practice to approximate the Snell envelope, which are precisely based on the approximation of the dynamic programming recursion. The idea is to analyze the related approximation error in terms of robustness properties of the Snell envelope with respect to the pair parameters (f_k, M_k). Hence, we include in our analysis approximation schemes which are defined in terms of some *approximate* pairs of functions and transitions $(\widehat{f}_k, \widehat{M}_k)_{k\geq 0}$. After stating the robustness Lemma 4.2.1 in the preliminary Section 3.2, we deduce from it non asymptotic convergence theorems, including $\mathbb{L}_p$-mean error bounds and related exponential inequalities for the deviations of Monte Carlo type approximation models.

In Section 3.3, that approach allows us to derive non asymptotic error bounds for deterministic approximation schemes such as *cut-off techniques, Euler type discrete time approximations, quantization tree models, interpolation type approximations*, then recovering or improving some existing results or in some cases providing new bounds. We emphasize that this non asymptotic robustness analysis also allows to combine in a natural way several approximation models. For instance, under appropriate tightness conditions, cut-off techniques can be used to reduce the numerical analysis of the Snell envelope to compact state spaces and bounded functions $\widehat{f}_n$. In the same line of ideas, in designing any type of Monte Carlo approximation models, we can suppose that the transitions of the chain X_n are known, based on a preliminary analysis of Euler type approximation models.

In Section 3.4, we focus on two kinds of Monte Carlo importance sampling approximation schemes. The first one is the Stochastic Mesh method introduced by M. Broadie and P. Glasserman in their seminal paper [19] (see also [87], for some recent refinements). The principal idea of that methodology is to operate a change of measure to replace conditional expectations by simple expectations involving Markov transition densities with respect to some reference measures. The number of sampled points with respect to the reference measures η_n required by this model can be constant in every exercise date. This technique avoids the explosion issue of the naive Monte Carlo method. As any full Monte Carlo type technique, the main advantage of their approach is that it applies to high dimensional Bermudan options with a finite but possibly large number of exercise dates. In [19], the authors provided a set of conditions under which the Monte Carlo importance scheme converges as the computational effort increases. However, the computing time grows quadratically with the number of sampled points in the stochastic mesh. In this context, in Section 3.4.2, we provide new non asymptotic estimates, including $\mathbb{L}_p$-mean error bounds and exponential concentration inequalities. Our analysis allows us to derive Theorem 3.4.7 improving significantly existing convergence results (see [19] or [3]). The second type of Monte Carlo importance sampling scheme discussed in this chapter is another version of the Broadie-Glasserman model, called *average density* in their original article. The main advantage of this strategy comes from the fact that the sampling distribution η_n can be chosen as the distribution of the random states X_n of the reference Markov chain, even if the Radon-Nikodym derivatives $R_n(x,y) = \frac{dM_n(x,\cdot)}{d\eta_n}(y)$ are not known explicitly. Here, we only assume that the Markov transitions $M_n(x,\cdot)$ are absolutely continuous with respect to some measures λ_n. We can then approximate these functions with empirical measures. In this situation, we can recover a similar approximation to the original stochastic mesh method, except that the Radon-Nikodym derivatives $R_{k+1}(\xi_k^i, \xi_{k+1}^j)$ are replaced by approximations. The stochastic analysis of this particle model is provided in the second part of Section 3.4.2 and follows essentially the same line of arguments as

the one of the Broadie-Glasserman model.

In the final part of the chapter, Section 3.5, we present a new Monte Carlo approach based on the genealogical tree evolution model associated with a neutral genetic model with mutations given by the Markov transitions M_n. The main advantage of this new strategy comes from the fact that the computational effort of the algorithm is now linear in the number of sampled points. We recall that a neutral genetic model is a Markov chain with a selection/mutation transition. During the mutation phase, the particles explore the state space independently according to the Markov transitions while the selection step induces interactions between the various particles. This type of model is frequently used in biology, and genetic algorithms literature (see for instance [56], and references therein).

An important observation concerns the genealogical tree structure of the genetic particle model that we consider. The main advantage of this path particle model comes from the fact that the occupation measure of the ancestral tree model converges in some sense to the distribution of the path of the reference Markov chain. It is also well known that the Snell envelope associated with a Markov chain evolving on some finite state space is easily computed using the tree structure of the chain evolution. Therefore, replacing the reference distribution $\mathbb{P}_n$ by its N-approximation $\mathbb{P}_n^N$, we define an N-approximated Markov model whose evolutions are described by the genealogical tree model defined above. We can then construct the approximation $\widehat{u}_k$ as the Snell envelope associated with this N-approximated Markov chain. Several estimates of convergence are provided in Section 3.5. Finally, some numerical simulations are performed, illustrating the interest of our new algorithm.

3.2 Preliminaries

In a discrete time setting, the problem is related to the pricing of Bermuda options and is defined in terms of a given real valued stochastic process $(Z_k)_{0\leq k\leq n}$, adapted to some increasing filtration $\mathcal{F} = (\mathcal{F}_k)_{0\leq k\leq n}$ that represents the available information at any time $0 \leq k \leq n$. For any $k \in \{0, \ldots, n\}$, let $\mathcal{T}_k$ be the set of all stopping times τ taking values in $\{k, \ldots, n\}$. The Snell envelope of $(Z_k)_{0\leq k\leq n}$, is the stochastic process $(U_k)_{0\leq k\leq n}$ defined for any $0 \leq k < n$ by the following backward equation

$$U_k = Z_k \vee \mathbb{E}(U_{k+1}|\mathcal{F}_k) \ ,$$

with the terminal condition $U_n = Z_n$, where $a \vee b = \max(a, b)$. The main property of this stochastic process is that

$$\begin{aligned} & U_k = \sup_{\tau\in\mathcal{T}_k} \mathbb{E}(Z_\tau|\mathcal{F}_k) = \mathbb{E}(Z_{\tau_k^*}|\mathcal{F}_k) \qquad (3.2.1) \\ \text{with} \qquad & \tau_k^* = \min\{k \leq l \leq n \ : \ U_l = Z_l\} \in \mathcal{T}_k. \end{aligned}$$

At this level of generality, in the absence of any additional information on the filtration $\mathcal{F}$, or on the terminal random variable Z_n, no numerical computation of the Snell envelope is available. To get one step further, we assume that $(\mathcal{F}_n)_{n\geq 0}$ is the natural filtration associated with some Markov chain $(X_n)_{n\geq 0}$ taking values in some sequence of measurable state spaces $(E_n, \mathcal{E}_n)_{n\geq 0}$. Let $\eta_0 = \mathrm{Law}(X_0)$ be the initial distribution on E_0, and define by $M_n(x_{n-1}, dx_n)$ the elementary Markov transition of the chain from E_{n-1} into E_n. We also assume that $Z_n = f_n(X_n)$, for some collection of non-negative measurable functions f_n on E_n. In this situation, the computation of the Snell envelope amounts to solving the following backward functional equation

$$u_k = \mathcal{H}_{k+1}(u_{k+1}) = f_k \vee M_{k+1}(u_{k+1}), \tag{3.2.2}$$

for any $0 \leq k < n$, with the terminal value $u_n = f_n$. In the above displayed formula, $M_{k+1}(u_{k+1})$ stands for the measurable function on E_k defined for any $x_k \in E_k$ by the conditional expectation formula

$$\begin{aligned} M_{k+1}(u_{k+1})(x_k) &= \int_{E_{k+1}} M_{k+1}(x_k, dx_{k+1})\; u_{k+1}(x_{k+1}) \\ &= \mathbb{E}\left(u_{k+1}(X_{k+1}) | X_k = x_k\right) \; . \end{aligned}$$

Let $\mathcal{H}_{k,l} = \mathcal{H}_{k+1} \circ \mathcal{H}_{k+1,l}$, with $k \leq l \leq n$, be the nonlinear semigroups associated with the backward equation (4.2.1). We use the convention $\mathcal{H}_{k,k} = Id$, the identity operator, so that $u_k = \mathcal{H}_{k,l}(u_l)$, for any $k \leq l \leq n$. Given a sequence of bounded integral operators M_k from some state space E_{k-1} into another E_k, let us denote by $M_{k,l}$ the composition operator such that $M_{k,l} := M_{k+1} M_{k+2} \cdots M_l$, for any $k \leq l$, with the convention $M_{k,k} = Id$, the identity operator. With this notation, one can check that a necessary and sufficient condition for the existence of the Snell envelope $(u_k)_{0\leq k\leq n}$ is that $M_{k,l} f_l(x) < \infty$ for any $1 \leq k \leq l \leq n$, and any state $x \in E_k$. To check this claim, we simply notice that

$$f_k \leq u_k \leq f_k + M_{k+1} u_{k+1} \, , \; \forall \, 1 \leq k \leq n \implies f_k \leq u_k \leq \sum_{k\leq l\leq n} M_{k,l} f_l \, , \; \forall \, 1 \leq k \leq n \, . \tag{3.2.3}$$

From the readily proved Lipschitz property $|\mathcal{H}_k(u) - \mathcal{H}_k(v)| \leq M_{k+1}(|u - v|)$, for any functions u, v on E_k, we also have that

$$|\mathcal{H}_{k,l}(u) - \mathcal{H}_{k,l}(v)| \leq M_{k,l}(|u - v|) \, , \tag{3.2.4}$$

for any functions u, v on E_l, and any $k \leq l \leq n$.

Even if it looks simple, the numerical solving of the recursion (4.2.1) often requires extensive computations. The central problem is to compute the conditional expectation $M_{k+1}(u_{k+1})$ on the whole state space E_k, at every time step $0 \leq k < n$.

For Markov chain models taking values in some finite state spaces (with a reasonably large cardinality), the above expectations can be easily computed by a simple backward inspection of the whole realization tree that lists all possible outcomes and every transition of the chain. In more general situations, we need to resort to some approximation strategy. Most of the numerical approximation schemes amount to replacing the pair of functions and Markov transitions $(f_k, M_k)_{0\leq k\leq n}$ by some approximation model $(\widehat{f}_k, \widehat{M}_k)_{0\leq k\leq n}$ on some possibly reduced measurable subsets $\widehat{E}_k \subset E_k$. Let $\widehat{u}_k$ be the Snell envelope on $\widehat{E}_k$ associated with the functions $\widehat{f}_k$ and the sequence of integral operators $\widehat{M}_k$ from $\widehat{E}_{k-1}$ into $\widehat{E}_k$. As in (4.2.1), the computation of the Snell envelope $\widehat{u}_k$ amounts to solving the following backward functional equation

$$\widehat{u}_k \quad = \quad \widehat{\mathcal{H}}_{k+1}(\widehat{u}_{k+1}) = \widehat{f}_k \vee \widehat{M}_{k+1}(\widehat{u}_{k+1}) \ . \tag{3.2.5}$$

Let $\widehat{\mathcal{H}}_{k,l} = \widehat{\mathcal{H}}_{k+1} \circ \widehat{\mathcal{H}}_{k+1,l}$, with $k \leq l \leq n$, be the nonlinear semigroups associated with the backward equations (3.2.5), so that $\widehat{u}_k = \widehat{\mathcal{H}}_{k,l}(\widehat{u}_l)$, for any $k \leq l \leq n$. Using the elementary inequality $|(a \vee a') - (b \vee b')| \leq |a - b| + |a' - b'|$, which is valid for any $a, a', b, b' \in \mathbb{R}$, for any $0 \leq k < n$ and for any functions u on E_{k+1} one readily obtains the local approximation inequality

$$\left|\mathcal{H}_{k+1}(u) - \widehat{\mathcal{H}}_{k+1}(u)\right| \leq |f_k - \widehat{f}_k| + |(M_{k+1} - \widehat{M}_{k+1})(u)|. \tag{3.2.6}$$

To transfer these local estimates to the semigroups $\mathcal{H}_{k,l}$ and $\widehat{\mathcal{H}}_{k,l}$ we use the same perturbation analysis as in [41, 48, 85, 109] in the context of nonlinear filtering semigroups and particle approximation models. The difference between the approximate and the exact Snell envelope can be written as a telescoping sum

$$u_k - \widehat{u}_k = \sum_{l=k}^{n} \left[\widehat{\mathcal{H}}_{k,l}(\mathcal{H}_{l+1}(u_{l+1})) - \widehat{\mathcal{H}}_{k,l}(\widehat{\mathcal{H}}_{l+1}(u_{l+1}))\right] ,$$

setting for simplicity $\mathcal{H}_{n+1}(u_{n+1}) = u_n$ and $\widehat{\mathcal{H}}_{n+1}(u_{n+1}) = \widehat{u}_n$, for $l = n$. Combining the Lipschitz property (3.2.4) of the semigroup $\widehat{\mathcal{H}}_{k,l}$ with the local estimate (3.2.6), one finally gets the following robustness lemma, which is a natural and fundamental tool for the analysis of the Snell envelope approximations.

Lemma 3.2.1. *For any $0 \leq k < n$, on the state space $\widehat{E}_k$, we have that*

$$|u_k - \widehat{u}_k| \quad \leq \quad \sum_{l=k}^{n} \widehat{M}_{k,l}|f_l - \widehat{f}_l| + \sum_{l=k}^{n-1} \widehat{M}_{k,l}|(M_{l+1} - \widehat{M}_{l+1})u_{l+1}| \ .$$

The perturbation analysis of nonlinear semigroups described above, and the resulting robustness lemma are not really new. As mentioned previously, it is a rather standard tool in approximation theory and numerical probability. More precisely,

these Lipschitz type estimates are often used by induction or as an intermediate technical step in the proof of a convergence theorem of some particular approximation scheme.

In the context of optimal stopping problems, similar induction arguments are developped to prove the convergence of some specific approximation models, for instance in the papers of Egloff [64] and Gobet, Lemor and Warin [76] or Pagès [92]. However, to the best of our knowledge, the general and abstract semigroup formulation given above and its direct application to different approximation models seems to be the first result of this type for that class of models.

Besides the fact that the convergence of many Snell approximation schemes results from a single robustness property, Lemma 4.2.1 can be used sequentially and without further work to obtain non asymptotic estimates for models combining several levels of approximations. In the same vein, and whenever it is possible, Lemma 4.2.1 can also be used as a technical tool to reduce the analysis of Snell approximation models on compact state spaces or even on finite but possibly large quantization trees or Monte Carlo type grids.

We end this section with an exponential inequality that can be readily deduced from the $\mathbb{L}_p$-mean error bounds presented in this chapter. For a more thorough discussion on the connexion between Khintchine style $\mathbb{L}_p$-mean error bounds and concentration inequalities, we refer the reader to [40], [44, 45], and the more recent article on the concentration properties of mean field type particle models [59].

Lemma 3.2.2. *Suppose the estimates have the following form:*

$$\sqrt{N} \sup_{x\in E_k} \mathbb{E}\left(|u_k(x)-\widehat{u}_k(x)|^p\right)^{\frac{1}{p}} \leq a(p)b_k(n),$$

where $b_k(n)$ are some finite constants whose values do not depend on the parameter p and $a(p)$ is a collection of constants such that for all non-negative integer r:

$$a(2r)^{2r} = (2r)_r\ 2^{-r} \quad \textit{and} \quad a(2r+1)^{2r+1} = \frac{(2r+1)_{r+1}}{\sqrt{r+1/2}}\ 2^{-(r+1/2)}\ , \tag{3.2.7}$$

with the notation $(q)_p = q!/(q-p)!$, for any $1 \leq p \leq q$. Then we deduce the following exponential concentration inequality

$$\sup_{x\in E_k} \mathbb{P}\left(|u_k(x_k)-\widehat{u}_k(x_k)| > \frac{b_k(n)}{\sqrt{N}}+\epsilon\right) \leq exp\left(-N\epsilon^2/(2b_k(n)^2)\right)\ . \tag{3.2.8}$$

Proof. This result is a direct consequence of the fact that for any non-negative random variable U, if there exists a bounded positive real b such that

$$\forall\, r \geq 1, \qquad \mathbb{E}\left(U^r\right)^{\frac{1}{r}} \leq a(r)b\ ,$$

where $a(r)$ is defined by (3.2.7), then

$$\mathbb{P}\left(U \geq b+\epsilon\right) \leq \exp\left(-\epsilon^2/(2b^2)\right)\ .$$

To check this implication, we first notice that

$$\mathbb{P}\left(U \geq b+\epsilon\right) \leq \inf_{t\geq 0}\{e^{-t(b+\epsilon)}\mathbb{E}[e^{tU}]\} \ .$$

Then developing the exponential and using the moments boundedness assumption, one obtains that for all $t \geq 0$,

$$\mathbb{E}\left(e^{tU}\right) \leq \exp\left(\frac{(bt)^2}{2} + bt\right) \ .$$

As a result,

$$\mathbb{P}\left(U \geq b+\epsilon\right) \leq \exp\left(-\sup_{t\geq 0}\left(\epsilon t - \frac{(bt)^2}{2}\right)\right) \ .$$

■

3.3 Some deterministic approximation models

In this section, we analyze the robustness of the Snell envelope with respect to some deterministic approximation schemes that are parts of many algorithms proposed to approximate the Snell envelope. Hence, the non asymptotic error bounds provided in this section can be applied and combined to derive convergence rates for such algorithms. We recover or improve previous results and in some cases, state new error bounds.

3.3.1 Cut-off type models

It is often useful, when computing the Snell envelope, to approximate the state space by a compact set. Indeed, Glasserman and Yu (2004) [75] showed that for standard (unbounded) models (like Black-Scholes), the Monte Carlo estimation requires samples of exponential size in the number of variables of the value function, whereas the bounded state space assumption enables to estimate the Snell envelope from samples of polynomial size in the number of variables. For instance, in [65], the authors proposed a new algorithm that first requires a cut-off step which consists in replacing the price process by another process killed at first exit from a given bounded set. However, no bound is provided for the error induced by this cut-off approximation. In this section, we formalize a general cut-off model and provide some bounds on the error induced on the Snell envelope.

We suppose that for each n, E_n is a topological space with σ-fields $\mathcal{E}_n$ that contains the Borel σ-field on E_n. Our next objective is to find conditions under which we can reduce the backward functional equation (4.2.1) to a sequence of compact sets $\widehat{E}_n$.

To that end, we further assume that the initial measure η_0 and the Markov transition M_n of the chain X_n satisfy the following tightness property: For every sequence of positive numbers $\epsilon_n \in [0,1[$, there exists a collection of compact subsets $\widehat{E}_n \subset E_n$ s.t.

$$(\mathcal{T}) \qquad \eta_0(\widehat{E}_0^c) \leq \epsilon_0 \quad \text{and} \quad \forall n \geq 0 \quad \sup_{x_n \in \widehat{E}_n} M_{n+1}(x_n, \widehat{E}_{n+1}^c) \leq \epsilon_{n+1} \ .$$

For instance, this condition is clearly met for regular Gaussian type transitions on the Euclidean space, for some collection of increasing compact balls.

In this situation, a natural cut-off consists in considering the Markov transitions $\widehat{M}_k$ restricted to the compact sets $\widehat{E}_k$

$$\forall x \in \widehat{E}_{k-1} \qquad \widehat{M}_k(x, dy) := \frac{M_k(x, dy)\ 1_{\widehat{E}_k}}{M_k(1_{\widehat{E}_k})(x)} \ .$$

These transitions are well defined as soon as $M_k(x, \widehat{E}_k) > 0$, for any $x \in \widehat{E}_{k-1}$. Using the decomposition

$$\begin{aligned}
[\widehat{M}_k - M_k](u_k) &= \widehat{M}_k(u_k) - M_k(1_{\widehat{E}_k} u_k) - M_k(1_{\widehat{E}_k^c} u_k) \\
&= \left(1 - \frac{1}{M_k(1_{\widehat{E}_k})}\right) M_k(u_k 1_{\widehat{E}_k}) - M_k(1_{\widehat{E}_k^c} u_k) \\
&= \frac{M_k(1_{\widehat{E}_k^c})}{M_k(1_{\widehat{E}_k})} M_k(u_k 1_{\widehat{E}_k}) - M_k(1_{\widehat{E}_k^c} u_k) \ ,
\end{aligned}$$

then using Lemma 4.2.1 yields

$$\begin{aligned}
\|u_k - \widehat{u}_k\|_{\widehat{E}_k} &:= \sup_{x \in \widehat{E}_k} |u_k(x) - \widehat{u}_k(x)| \\
&\leq \sum_{l=k+1}^{n} \left[\left\| \frac{M_l(1_{\widehat{E}_l^c})}{M_l(1_{\widehat{E}_l})} \right\|_{\widehat{E}_{l-1}} \|M_l(u_l 1_{\widehat{E}_l})\|_{\widehat{E}_{l-1}} + \|M_l(u_l 1_{\widehat{E}_l^c})\|_{\widehat{E}_{l-1}} \right] \\
&\leq \sum_{l=k+1}^{n} \left[\frac{\epsilon_l}{1-\epsilon_l} \|M_l(u_l)\|_{\widehat{E}_{l-1}} + \|M_l(u_l^2)\|_{\widehat{E}_{l-1}}^{1/2} \epsilon_l^{1/2} \right] \ .
\end{aligned}$$

We summarize the above discussion with the following result.

Theorem 3.3.1. *We assume that the tightness condition $(\mathcal{T})$ is met, for every sequence of positive numbers $\epsilon_n \in [0,1[$, and for some collection of compact subsets $\widehat{E}_n \subset E_n$. In this situation, for any $0 \leq k \leq n$, we have that*

$$\|u_k - \widehat{u}_k\|_{\widehat{E}_k} \leq \sum_{l=k+1}^{n} \frac{\epsilon_l^{1/2}}{1-\epsilon_l^{1/2}} \|M_l(u_l^2)\|_{\widehat{E}_{l-1}}^{1/2} \ .$$

Note that

$$u_k \leq \sum_{l=k}^{n} M_{k,l}(f_l),$$

and therefore

$$\left\|M_k(u_k^2)\right\|_{\widehat{E}_{k-1}} \leq (n-k+1) \sum_{l=k}^{n} \left\|M_{k-1,l}(f_l)^2\right\|_{\widehat{E}_{k-1}}.$$

Consequently, one can find sets $(\widehat{E}_l)_{k<l\leq n}$ so that $\|u_k - \widehat{u}_k\|_{\widehat{E}_k}$ is as small as one wants as soon as $\|M_{k,l}(f_l)^2\|_{\widehat{E}_k} < \infty$, for any $0 \leq k < l \leq n$. A similar cut-off approach was introduced and analyzed in Bouchard and Touzi [13], but the cut-off was operated on some regression functions and not on the transition kernels.

3.3.2 Euler approximation models

In several application model areas, the discrete time Markov chain $(X_k)_{k\geq 0}$ is often given in terms of an $\mathbb{R}^d$-valued and continuous time process $(X_t)_{t\geq 0}$ given by a stochastic differential equation of the following form

$$dX_t \;=\; a(X_t)dt + b(X_t)dW_t, \qquad law(X_0) \;=\; \eta_0, \tag{3.3.1}$$

where η_0 is a known distribution on $\mathbb{R}^d$, a, b are known functions, and W is a d-dimensional Wiener process. Except in some particular instances, the time homogeneous Markov transitions $M_k = M$ are usually unknown, and we need to resort to an Euler approximation scheme.

In this situation, any approximation of the Snell envelope, which is based on simulations of the price process will be impacted by the error induced by the Euler scheme used in simulations. We propose here to provide bounds for that error. Notice that in this setting, the exercise dates are discrete and fixed, so that our results are not comparable with those from Dupuis and Wang (2004) [62] who analyzed the convergence of the discrete time optimal stopping problem to the continuous time optimal stopping problem when the frequency of exercise dates increases to infinity. Similarly, for numerical approximations of Backward Stochastic Differential Equations (BSDE), [13] and [76] also analysed the case where the number of exercise opportunities grows to infinity.

The discrete time approximation model with a fixed time step $1/m$ is defined by the following recursive formula

$$\begin{aligned} \widehat{\xi}_0(x) &= x \\ \widehat{\xi}_{\frac{(i+1)}{m}}(x) &= \widehat{\xi}_{\frac{i}{m}}(x) + a\left(\widehat{\xi}_{\frac{i}{m}}(x)\right)\frac{1}{m} + b\left(\widehat{\xi}_{\frac{i}{m}}(x)\right)\frac{1}{\sqrt{m}}\,\epsilon_i\,. \end{aligned}$$

where the ϵ_i's are i.i.d. centered and $\mathbb{R}^d$-valued Gaussian vectors with unit covariance matrix. The chain $(\widehat{\xi}_k)_{k\geq 0}$ is an homogeneous Markov with a transition kernel which we denote by $\widehat{M}$.

We further assume that the functions a and b are twice differentiable, with bounded partial derivatives of orders 1 and 2, and the matrix $(bb^*)(x)$ is uniformly non-degenerate.

In this situation, the integral operators M and $\widehat{M}$ admit densities, denoted by p and $\widehat{p}$. According to Bally and Talay [5, 6], we have that

$$[p \vee \widehat{p}] \ \leq \ c\, q \quad \text{and} \quad m\,|\widehat{p} - p| \ \leq \ c\, q\ , \tag{3.3.2}$$

with the Gaussian density $q(x,x') := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}|x-x'|^2}$, and a pair of constants (c,σ) depending only on the pair of functions (a,b). Let Q, be the Markov integral operator on $\mathbb{R}^d$ with density q. We consider a sequence of functions $(f_k)_{0\leq k\leq n}$ on $\mathbb{R}^d$. Let $(u_k)_{0\leq k\leq n}$ and $(\widehat{u}_k)_{0\leq k\leq n}$ be the Snell envelope on $\mathbb{R}^d$ associated to the pair (M, f_k) and $(\widehat{M}, f_k)$. Using Lemma 4.2.1, we readily obtain the following estimate

$$|u_k - \widehat{u}_k| \ \leq \ \sum_{l=k}^{n-1} \widehat{M}^{l-k} |(M - \widehat{M}) u_{l+1}| \leq \frac{c}{m} \ \sum_{l=k}^{n-1} \widehat{M}^{l-k} Q |u_{l+1}|\ .$$

Rather crude upper bounds that do not depend on the approximation kernels $\widehat{M}$ can be derived using the first inequality in (3.3.2)

$$|u_k - \widehat{u}_k| \leq \frac{1}{m} \sum_{l=1}^{n-k} c^l\ Q^l |u_{l+k}|\ .$$

Recalling that $u_{l+k} \leq \sum_{l+k\leq l'\leq n} M^{l'-(l+k)} f_{l'}$, we also have that

$$\begin{aligned} |u_k - \widehat{u}_k| \ &\leq \ \frac{1}{m} \sum_{l=1}^{n-k} c^l\ Q^l \sum_{l+k\leq l'\leq n} c^{l'-(l+k)}\ Q^{l'-(l+k)} f_{l'} \\ &\leq \ \frac{1}{m} \sum_{l=1}^{n-k} \sum_{l+k\leq l'\leq n} c^{l'-k}\ Q^{l'-k} f_{l'} = \frac{1}{m} \sum_{1\leq l\leq n-k} l\ c^l\ Q^l f_{k+l}\ . \end{aligned}$$

We summarize the above discussion with the following theorem.

Theorem 3.3.2. *Suppose the functions $(f_k)_{0\leq k\leq n}$ on $\mathbb{R}^d$ are chosen such that $Q^l f_{k+l}(x) < \infty$, for any $x \in \mathbb{R}^d$, and $1 \leq k+l \leq n$. Then, for any $0 \leq l \leq n$, we have the inequalities*

$$|u_k - \widehat{u}_k| \leq \frac{c}{m} \ \sum_{l=k}^{n-1} \widehat{M}^{l-k} Q |u_{l+1}| \leq \frac{1}{m} \sum_{1\leq l\leq n-k} l\ c^l\ Q^l f_{k+l}\ .$$

3.3.3 Interpolation type models

Most algorithms proposed to approximate the Snell envelope provide discrete approximations $\hat{u}_k^i$ at some discrete (potentially random) points ξ_k^i of E_k. However, for several purposes, it can be interesting to consider approximations $\hat{u}_k$ of functions u_k on the whole space E_k. One motivation to do so is, for instance, to be able to define a new (low biased) estimator, $\bar{U}_k$, using a Monte Carlo approximation of (3.2.1), with a stopping rule $\hat{\tau}_k$ associated with the approximate Snell envelope $\hat{u}_k$, by replacing u_k by $\hat{u}_k$ in the characterization of the optimal stopping time τ_k^* (3.2.1), i.e.

$$\bar{U}_k = \frac{1}{M}\sum_{i=1}^{M} f_{\hat{\tau}_k^i}(X^i_{\hat{\tau}_k^i}) \quad \text{with} \quad \hat{\tau}_k^i = \min\left\{k \le l \le n \ : \ \hat{u}_l(X_l^i) = f_l(X_l^i)\right\} . \quad (3.3.3)$$

where $X^i = (X_1^i, \cdots, X_n^i)$ are i.i.d. path according to the reference Markov chain dynamic.

In this section, we analyze non asymptotic errors of some specific approximation schemes providing such interpolated estimators $\hat{u}_k$ of u_k on the whole state E_k. Let $\widehat{M}_{k+1} = \mathcal{I}_k \widetilde{M}_{k+1}$ be the composition of the Markov transition $\widetilde{M}_{k+1}$ from a finite set S_k into the whole state space E_{k+1}, with an auxiliary interpolation type and Markov operator $\mathcal{I}_k$ from E_k into S_k, so that

$$\forall x_k \in S_k \qquad \mathcal{I}_k(x_k, ds) = \delta_{x_k}(ds) \ ,$$

and such that the integrals

$$x \in E_k \mapsto \mathcal{I}_k(\varphi_k)(x) = \int_{S_k} \mathcal{I}_k(x, ds)\ \varphi_k(s) \ ,$$

of any function φ_k on S_k are easily computed starting from any point x_k in E_k. We further assume that the finite state spaces S_k are chosen so that

$$\|f - \mathcal{I}_k f\|_{E_k} \le \epsilon_k(f, |S_k|) \to 0 \quad \text{as} \quad |S_k| \to \infty \ , \quad (3.3.4)$$

for continuous functions f_k on E_k. An example of interpolation transition $\mathcal{I}_k$ is provided hereafter. Let $\widehat{M}_k = \mathcal{I}_{k-1}\widetilde{M}_k$ be the composition operator on the state spaces $\widehat{E}_k = E_k$.

The approximation models $\widetilde{M}_k$ are non necessarily deterministic. In [**?**], the authors examined the situation where

$$\forall s \in S_k \qquad \widetilde{M}_k(s, dx) = \frac{1}{N_k}\sum_{1 \le i \le N_k} \delta_{X_k^i(s)}(dx) \ ,$$

where $X_k^i(s)$ stands for a collection of N_k independent random variables with common law $M_k(s, dx)$.

Theorem 3.3.3. *We suppose that the Markov transitions M_k are Feller, in the sense that $M_k(C(E_k)) \subset C(E_{k-1})$, where $C(E_k)$ stands for the space of continuous functions on the E_k. Let $(u_k)_{0\leq k\leq n}$, and respectively $(\widehat{u}_k)_{0\leq k\leq n}$ be the Snell envelope associated with the functions $f_k = \widehat{f}_k$, and the Markov transitions M_k, and respectively $\widehat{M}_k = \mathcal{I}_{k-1}\widetilde{M}_k$ on the state spaces $\widehat{E}_k = E_k$. Then*

$$\|u_k - \widehat{u}_k\|_{E_k} \leq \sum_{l=k}^{n-1} \left[\epsilon_l\left(M_{l+1}u_{l+1}, |S_l|\right) + \|(M_{l+1} - \widetilde{M}_{l+1})u_{l+1}\|_{S_l}\right] .$$

The proof of the theorem is a direct consequence of Lemma 4.2.1 combined with the following decomposition

$$\begin{aligned} &\|u_k - \widehat{u}_k\|_{E_k} \\ &\quad \leq \sum_{l=k}^{n-1} \left[\|(Id - \mathcal{I}_l)M_{l+1})u_{l+1}\|_{E_l} + \|\mathcal{I}_l(M_{l+1} - \widetilde{M}_{l+1})u_{l+1}\|_{E_l}\right] . \end{aligned} \tag{3.3.5}$$

We illustrate these results in the typical situation where the space E_k are the convex hull generated by the finite sets S_k. Firstly, we present the definition of the interpolation operators. Let $\P = \{\P^1, \ldots, \P^m\}$ be a partition of a convex and compact space E into simplexes with disjoint non empty interiors, so that $E = \cup_{1\leq i\leq m}\P_i$. We denote by $\delta(\P)$ the refinement degree of the partition $\P$

$$\delta(\P) := \sup_{1\leq i\leq m} \sup_{x,y\in\P_i} \|x - y\| .$$

Let $S = \mathcal{V}(\P)$ be the set of vertices of these simplexes. We denote by $\mathcal{I}$ be the interpolation operator defined by $\mathcal{I}(f)(s) = f(s)$, if $s \in S$, and if x belongs to some simplex $\P^j$ with vertices $\{x_1^j, \ldots, x_{d_j}^j\}$

$$\mathcal{I}(f)(\sum_{1\leq i\leq d_j} \lambda_i\ x_j^i) = \sum_{1\leq i\leq d_j} \lambda_i\ f(x_i^j) ,$$

where the barycenters $(\lambda_i)_{1\leq i\leq d_j}$ are the unique solution of

$$x = \sum_{1\leq i\leq d_j} \lambda_i\ x_i^j \quad \text{with} \quad (\lambda_i)_{1\leq i\leq d_j} \in [0,1]^{d_j} \quad \text{and} \quad \sum_{1\leq i\leq d_j} \lambda_i = 1 .$$

The Markovian interpretation is that starting from x, one chooses the " closest simplex" and then one chooses one of its vertices x_i with probability λ_i.

For any $\delta > 0$, let $\omega(f, \delta)$ be the δ-modulus of continuity of a function $f \in C(E)$

$$\omega(f, \delta) := \sup_{(x,y)\in E:\|x-y\|\leq\delta} |f(x) - f(y)| .$$

The following technical Lemma provides a simple way to check condition (3.3.4) for interpolation kernels.

Lemma 3.3.1. *Then for any $f, g \in C(E)$,*

$$\sup_{x\in E} |f(x) - \mathcal{I}g(x)| \leq \max_{x\in S} |f(x) - g(x)| + \omega(f, \delta(\P)) + \omega(g, \delta(\P)) \ . \tag{3.3.6}$$

In particular, we have that

$$\sup_{x\in E} |f(x) - \mathcal{I}f(x)| \leq \omega(f, \delta(\P)) \ .$$

Proof. Suppose x belongs to some simplex $\P^j$ with vertices $\{x_1^j, \ldots, x_{d_j}^j\}$, and let $(\lambda_i)_{1\leq i\leq d_j}$ be the barycenter parameters $x = \sum_{1\leq i\leq d_j} \lambda_i \, x_j^i$. Since we have $\mathcal{I}g(x_i^j) = g(x_i^j)$, and $\mathcal{I}g(x_i^j) = g(x_i^j)$ for any $i \in \{1, \ldots, d_j\}$, it follows that

$$\begin{aligned}
|f(x) - \mathcal{I}g(x)| &\leq \sum_{i=1}^{d_j} \lambda_i |(f(x) - f(x_i^j)| + \sum_{i=1}^{d_j} \lambda_i |f(x_i^j) - \mathcal{I}g(x_i^j)| \\
&\qquad + \sum_{i=1}^{d_j} \lambda_i |\mathcal{I}g(x_i^j) - g(x)| \\
&= \sum_{i=1}^{d_j} \lambda_i |(f(x) - f(x_i^j)| + \sum_{i=1}^{d_j} \lambda_i |f(x_i^j) - g(x_i^j)| \\
&\qquad + \sum_{i=1}^{d_j} \lambda_i |g(x_i^j) - g(x)| \ .
\end{aligned}$$

This implies that

$$\sup_{x\in \P^j} |f(x) - \mathcal{I}g(x)| \leq \max_{x\in \P^j} |f(x) - g(x)| + \omega(f, \delta(\P^j)) + \omega(g, \delta(\P^j)) \ ,$$

with

$$\omega(f, \delta(\P^j)) = \sup_{\|x-y\|\leq \delta(\P^j)} |f(x) - f(y)| \quad \text{and} \quad \delta(\P^j) := \sup_{x,y\in \P^j} \|x - y\| \ .$$

The end of the proof is now clear. ∎

Combining (3.3.5) and (3.3.6), we obtain the following result.

Proposition 3.3.2. *Let $\P_k = \{\P_k^1, \ldots, \P_k^{m_k}\}$ be a partition of a convex and compact space E_k into simplexes with disjoint non empty interiors, so that $E_k = \cup_{1\leq i\leq m_k} \P_i$. Let $S_k = \mathcal{V}(\P_k)$ be the set of vertices of these simplexes. Let $(\widehat{u}_k)_{0\leq k\leq n}$, be the Snell envelope associated with the functions $\widehat{f}_k = f_k$ and the Markov transitions $\widehat{M}_k = \mathcal{I}_{k-1}\widetilde{M}_k$ on the state spaces $E_k = \widehat{E}_k$.*

$$\|u_k - \widehat{u}_k\|_{E_k} \leq \sum_{l=k}^{n-1} \left[\omega(M_{l+1}u_{l+1}, \delta(\P_l)) + \|(M_{l+1} - \widetilde{M}_{l+1})u_{l+1}\|_{S_l}\right] \ .$$

To illustrate the results of Theorem 3.3.3 and Proposition 3.3.2, we have derived the effective convergence rate induced by the interpolation in a specific example. Following the previous Section, let us consider the $\mathbb{R}^d$-valued Markov chain $(\hat{\xi}_k)_{0\leq k\leq n}$ defined as the Euler time discretization of the stochastic differential equation (3.3.1), with a time step $\Delta t = 1$, i.e.

$$\begin{aligned} \hat{\xi}_0 &= x \\ \hat{\xi}_{k+1} &= \hat{\xi}_k + a(\hat{\xi}_k)\Delta t + b(\hat{\xi}_k)\sqrt{\Delta t}\epsilon_k \ , \end{aligned} \tag{3.3.7}$$

where ϵ_k are i.i.d. centered Gaussian vectors on $\mathbb{R}^d$ with unit covariance matrix. Let $\mathrm{L}ip(\mathbb{R}^d)$ be the set of all Lipschitz functions f on $\mathbb{R}^d$, and we set

$$L(f) = \sup_{x,y\in\mathbb{R}^d, x\neq y} \frac{\|f(x) - f(y)\|}{\|x - y\|} \ , \qquad f \in \mathrm{L}ip(\mathbb{R}^d). \tag{3.3.8}$$

We assume that $a : \mathbb{R}^d \to \mathbb{R}$ and $b : \mathbb{R}^d \to \mathcal{M}(d, d)$ are Lipschitz continuous functions. Then, we can prove that the time homogeneous Markov transitions $M_k = M$ associated to the Markov chain $(\hat{\xi}_k)_{0\leq k\leq n}$ is such that for any Lipschitz continuous function f on $\mathbb{R}^d$,

$$|M(f)(x) - M(f)(y)| \leq (1 + \alpha)L(f)\|x - y\| \ , \tag{3.3.9}$$

with $\alpha := \alpha(L(a), L(b), \Delta t) := L(a)\Delta t + dL(b)\sqrt{\Delta t} \geq 0$. Hence, we observe that $M_k(\mathrm{L}ip(\mathbb{R}^d)) \subset \mathrm{L}ip(\mathbb{R}^d)$. We also observe that

$$\begin{array}{c} \big(f_k \quad \text{and} \quad u_{k+1} \in \mathrm{L}ip(\mathbb{R}^d)\big) \\ \Downarrow \\ \big(u_k \in \mathrm{L}ip(\mathbb{R}^d) \quad \text{with} \quad L(u_k) \leq L(f_k) \vee L(M_{k+1}(u_{k+1}))\big) \end{array} \tag{3.3.10}$$

Moreover, assume that the payoff function $f_k = f$ for all $k = 0, \cdots, n$. Using (3.3.9) together with (3.3.10) implies

$$L(u_k) \leq (1 + \alpha)^{n-k} L(f) \ .$$

Using again (3.3.9) yields

$$\omega\big(M_{l+1}u_{l+1}, \delta(\mathcal{P}_l)\big) \leq (1 + \alpha)^{n-l} L(f)\delta(\mathcal{P}_l) \ .$$

Finally, in the specific case of model (3.3.7), with payoff functions $f_k = f$ and some refinement degrees of the partitions $\delta(\mathcal{P}_k) \leq \delta$, we obtain the following bound for the convergence of our interpolation model

$$\|u_k - \widehat{u}_k\|_{E_k} \leq \frac{(1 + \alpha)^{n-k+1}}{\alpha} L(f)\delta + \sum_{l=k}^{n-1} \|(M_{l+1} - \widetilde{M}_{l+1})u_{l+1}\|_{S_l} \ .$$

3.3.4 Quantization tree models

Quantization tree models belong to the class of deterministic grid approximation methods. The basic idea consists in choosing finite space grids

$$\widehat{E}_k = \left\{x_k^1, \ldots, x_k^{m_k}\right\} \subset E_k = \mathbb{R}^d \ ,$$

and some neighbourhoods measurable partitions $(A_k^i)_{1\leq k\leq m_k}$ of the whole space E_k such that the random state variable X_k is suitably approximated, as $m_k \to \infty$, by discrete random variables of the following form

$$\widehat{X}_k := \sum_{1\leq i\leq m_k} x_k^i \ 1_{A_k^i}(X_k) \simeq X_k \ .$$

The numerical efficiency of these quantization methods heavily depends on the choice of these grids. There exists various criteria to choose judiciously these objects, including minimal $\mathbb{L}_p$-quantization errors, that ensure that the corresponding Voronoi type quantized variable $\widehat{X}_k$ minimizes the $\mathbb{L}_p$ distance to the real state variable X_k. For further details on this subject, we refer the interested reader to the pioneering article of G. Pagès [92], and the series of articles of V. Bally, G. Pagès, and J. Printemps [4], G. Pagès and J. Printems [96], as well as [7, 20, 76, 95], and references therein. The second approximation step of these quantization model consists in defining the coupled distribution of any pair of variables $(\widehat{X}_{k-1}, \widehat{X}_k)$ by setting

$$\mathbb{P}\left(\widehat{X}_k = x_k^j \ , \ \widehat{X}_{k-1} = x_{k-1}^i\right) = \mathbb{P}\left(X_k \in A_k^j \ , \ X_{k-1} \in A_{k-1}^i\right) \ ,$$

for any $1 \leq i \leq m_{k-1}$, and $1 \leq j \leq m_k$. This allows to interpret the quantized variables $(\widehat{X}_k)_{0\leq k\leq n}$ as a Markov chain taking values in the states spaces $(\widehat{E}_k)_{0\leq k\leq n}$ with Markov transitions

$$\widehat{M}_k(x_{k-1}^i, x_k^j) := \mathbb{P}\left(\widehat{X}_k = x_k^j \mid \widehat{X}_{k-1} = x_{k-1}^i\right) = \mathbb{P}\left(X_k \in A_k^j \mid X_k \in A_{k-1}^i\right) \ .$$

Using the decompositions

$$\begin{aligned} M_k(f)(x_{k-1}^i) &= \sum_{j=1}^{m_k} \int_{A_k^j} f(y) \ \mathbb{P}(X_k \in dy \mid X_{k-1} = x_{k-1}^i) \\ &= \sum_{j=1}^{m_k} \int_{A_k^j} f(y) \ \mathbb{P}(X_k \in dy \mid X_{k-1} \in A_{k-1}^i) \\ &\quad + \int \left[M(f)(x_{k-1}^i) - M(f)(x)\right] \mathbb{P}(X_{k-1} \in dx | X_{k-1} \in A_{k-1}^i) \ , \end{aligned}$$

and

$$\widehat{M}_k(f)(x_{k-1}^i) = \sum_{j=1}^{m_k} \int_{A_k^j} f(x_k^j) \ \mathbb{P}(X_k \in dy \mid X_{k-1} \in A_{k-1}^i) \ ,$$

we find that

$$\begin{aligned}&[M_k - \widehat{M}_k](f)(x^i_{k-1})\\ &= \sum_{j=1}^{m_k} \int_{A^j_k} [f(y) - f(x^j_k)]\ \mathbb{P}(X_k \in dy \mid X_{k-1} \in A^i_{k-1})\\ &+ \int \left[M(f)(x^i_{k-1}) - M(f)(x)\right]\ \mathbb{P}(X_{k-1} \in dx \mid X_{k-1} \in A^i_{k-1})\ .\end{aligned}$$

We further assume that $M_k(\mathrm{L}ip(\mathbb{R}^d)) \subset \mathrm{L}ip(\mathbb{R}^d)$. From previous considerations, we find that

$$\begin{aligned}|[M_k - \widehat{M}_k](f)(x^i_{k-1})| \ &\leq\ L(f)\ \mathbb{E}\left[|X_k - \widehat{X}_k|^p \mid \widehat{X}_{k-1} = x^i_{k-1})\right]^{\frac{1}{p}}\\ &\quad + L(M_k(f))\ \mathbb{E}(|X_{k-1} - \widehat{X}_{k-1}|^p \mid \widehat{X}_{k-1} = x^i_{k-1})^{\frac{1}{p}}\ .\end{aligned}$$

This clearly implies that

$$\begin{aligned}\widehat{M}_{k,l}|(M_{l+1} - \widehat{M}_{l+1})f|(x^i_k) \ &\leq\ L(f)\ \left[\mathbb{E}(|X_{l+1} - \widehat{X}_{l+1}|^p \mid \widehat{X}_k = x^i_k)\right]^{\frac{1}{p}}\\ &\quad + L(M_{l+1}(f))\ \mathbb{E}(|X_l - \widehat{X}_l|^p \mid \widehat{X}_k = x^i_k)^{\frac{1}{p}}\ .\end{aligned}$$

Using (3.3.10), we also obtain that $L(u_k) \leq L(f_k) \vee L(M_{k+1}(u_{k+1}))$. Using Lemma 4.2.1, we readily arrive at the following Proposition similar to Theorem 2 in [4].

Proposition 3.3.3. *Assume that $(f_k)_{0\leq k\leq n} \in Lip(\mathbb{R}^d)^{n+1}$, and $M_k(Lip(\mathbb{R}^d)) \subset Lip(\mathbb{R}^d)$, for any $1 \leq k \leq n$. In this case, we have $(u_k)_{0\leq k\leq n} \in Lip(\mathbb{R}^d)^{n+1}$, and for any $0 \leq k \leq n$, we have the almost sure estimate*

$$\begin{aligned}|u_k - \widehat{u}_k|(\widehat{X}_k) \ &\leq\ L(M_{k+1}(u_{k+1}))\ |X_k - \widehat{X}_k|\\ &\quad + \sum_{l=k+1}^{n-1} (L(u_l) + L(M_{l+1}(u_{l+1})))\ \mathbb{E}(|X_l - \widehat{X}_l|^p \mid \widehat{X}_k)^{\frac{1}{p}}\\ &\quad + L(f_n)\ \left[\mathbb{E}(|X_n - \widehat{X}_n|^p \mid \widehat{X}_k)\right]^{\frac{1}{p}}\ .\end{aligned}$$

Proof. Using the decomposition

$$\widehat{u}_k(\widehat{X}_k) - u_k(X_k) = [\widehat{u}_k(\widehat{X}_k) - u_k(\widehat{X}_k)] + [u_k(\widehat{X}_k) - u_k(X_k)]\ ,$$

we have that

$$|u_k(\widehat{X}_k) - u_k(X_k)| \leq L(u_k)\ |\widehat{X}_k - X_k|\ .$$

Then the proof is completed by the following inequality:

$$\begin{aligned}|\widehat{u}_k(\widehat{\xi}_k) - u_k(X_k)| \ &\leq\ L(f_n)\ \left[\mathbb{E}(|X_n - \widehat{X}_n|^p \mid \widehat{X}_k)\right]^{\frac{1}{p}}\\ &\quad + \sum_{l=k}^{n-1} (L(u_l) + L(M_{l+1}(u_{l+1})))\ \mathbb{E}(|X_l - \widehat{X}_l|^p \mid \widehat{X}_k)^{\frac{1}{p}}\ .\end{aligned}$$

■

In contrast with [4] which focuses on optimizing deterministic grids, we remark that the independent applications of Lemma 4.2.1 in this model and in the previous examples illustrate the generality of our framework.

3.4 Monte Carlo importance sampling approximation schemes

3.4.1 Path space models

The choice of non homogeneous state spaces E_n is not without consequences. In several applications, the underlying Markov model is a path-space Markov chain

$$X_n = (X'_0, \ldots, X'_n) \in E_n = (E'_0 \times \ldots \times E'_n) \ . \tag{3.4.1}$$

The elementary prime variables X'_n represent an elementary Markov chain with Markov transitions $M'_k(x_{k-1}, dx'_k)$ from E'_{k-1} into E'_k. In this situation, the historical process X_n can be seen as a Markov chain with transitions given for any $x_{k-1} = (x'_0, \ldots, x'_{k-1}) \in E_{k-1}$ and $y_k = (y'_0, \ldots, y'_k) \in E_k$ by the following formula

$$M_k(x_{k-1}, dy_k) = \delta_{x_{k-1}}(dy_{k-1}) \ M'_k(y'_{k-1}, dy'_k) \ .$$

This path space framework is, for instance, well suited when dealing with path dependent options as Asian options.

Besides, this path space framework is also well suited for the analysis of Snell envelope under different probability measures. To fix the ideas, we associate with the latter a canonical Markov chain $\left(\Omega, \mathcal{F}, (X'_n)_{n\geq 0}, \mathbb{P}'_{\eta'_0}\right)$ with initial distribution η'_0 on E'_0, and Markov transitions M'_n from E'_{n-1} into E'_n. We use the notation $\mathbb{E}'_{\eta'_0}$ to denote the expectations with respect to $\mathbb{P}'_{\eta'_0}$. We further assume that there exists a sequence of measures $(\eta_k)_{0\leq k\leq n}$ on the state spaces $(E'_k)_{0\leq k\leq n}$ such that

$$\eta'_0 \sim \eta_0 \quad \text{and} \quad M'_k(x'_{k-1}, .) \sim \eta_k \ , \tag{3.4.2}$$

for any $x'_{k-1} \in E'_{k-1}$, and $1 \leq k \leq n$. Let $(\Omega, \mathcal{F}, (X'_n)_{n\geq 0}, \mathbb{P}_{\eta_0})$ be the canonical space associated with a sequence of independent random variables X'_k with distribution η_k on the state space E'_k, with $k \geq 1$. Under the probability measure $\mathbb{P}_{\eta_0}$, the historical process $X_n = (X'_0, \ldots, X'_n)$ can be seen as a Markov chain with transitions

$$M_k(x_{k-1}, dy_k) = \delta_{x_{k-1}}(dy_{k-1}) \ \eta_k(dy'_k) \ .$$

By construction, for any integrable function f'_k on E'_k, we have

$$\mathbb{E}_{\eta'_0}(f'_n(X'_n)) = \mathbb{E}_{\eta_0}\left(f_n(X_n)\right) \ ,$$

with the collection of functions f_k on E_k given for any $x_k = (x'_0, \ldots, x'_k) \in E_k$ by

$$f_k(x_k) = f'_k(x'_k) \times \frac{d\mathbb{P}'_k}{d\mathbb{P}_k}(x_k) \quad \text{with} \qquad \frac{d\mathbb{P}'_k}{d\mathbb{P}_k}(x_k) = \frac{d\eta'_0}{d\eta_0}(x'_0) \prod_{1\leq l\leq k} \frac{dM'_l(x'_{l-1},.)}{d\eta_l}(x'_l) \ . \tag{3.4.3}$$

Proposition 3.4.1. *The Snell envelope u_k and u'_k associated with the pairs (f'_k, M'_k) and (f_k, M_k) are given, for any $0 \leq k < n$, by the backward recursions*

$$u'_k = f'_k \vee M'_{k+1}(u'_{k+1}) \quad \textit{and} \quad u_k = f_k \vee M_{k+1}(u_{k+1}) \quad \textit{with} \quad (u'_n, u_n) = (f'_n, f_n) \ .$$

These functions are connected by the formula

$$\forall\, 0 \leq k \leq n, \quad \forall\, x_k = (x'_0, \ldots, x'_k) \in E_k, \qquad u_k(x_k) = u'_k(x'_k) \times \frac{d\mathbb{P}'_k}{d\mathbb{P}_k}(x_k) \ . \tag{3.4.4}$$

Proof. The first assertion is a simple consequence of the definition of a Snell envelope, and formula (3.4.4) is easily derived using the fact that

$$u'_k(x'_k) \ = \ f'_k(x'_k) \vee \left(\int_{E'_{k+1}} \eta_{k+1}(dx'_{k+1}) \ \frac{dM'_{k+1}(x'_k,.)}{d\eta_{k+1}}(x'_{k+1}) \ u'_{k+1}(x'_{k+1}) \right) \ .$$

That completes the proof of the proposition. ∎

Under condition (3.4.2), the above proposition shows that the computation of the Snell envelope associated with a given pair of functions and Markov transitions (f'_k, M'_k) reduces to that of the path space models associated with sequence of independent random variables with distributions η_n. More formally, the restriction $\mathbb{P}_{\eta_0,n}$ of reference measure $\mathbb{P}_{\eta_0}$ to the σ-field $\mathcal{F}_n$ generated by the canonical random sequence $(X'_k)_{0\leq k\leq n}$ is given by the the tensor product measure $\mathbb{P}_{\eta_0,n} = \otimes_{k=0}^n \eta_k$. Nevertheless, under these reference distributions, the numerical solving of the backward recursion stated in the above proposition still involves integrations with respect to the measures η_k. These equations can be solved if we replace these measures by some sequence of (possibly random) measures $\widehat{\eta}_k$ with finite support on some reduced measurable subset $\widehat{E}'_k \subset E'_k$, with a reasonably large and finite cardinality. We extend $\widehat{\eta}_k$ to the whole space E'_k by setting $\widehat{\eta}_k(E'_k - \widehat{E}'_k) = 0$.

Let $\widehat{\mathbb{P}}_{\widehat{\eta}'_0}$ be the distribution of a sequence of independent random variables $\widehat{\xi}'_k$ with distribution $\widehat{\eta}_k$ on the state space $\widehat{E}'_k$, with $k \geq 1$. Under the probability measure $\widehat{\mathbb{P}}_{\widehat{\eta}'_0}$, the historical process $X_n = (X'_0, \ldots, X'_n)$ can now be seen as a Markov chain taking values in the path spaces

$$\widehat{E}_k := \left(\widehat{E}'_0 \times \ldots \times \widehat{E}'_k\right) \ ,$$

with Markov transitions given for any $x_{k-1} = (x'_0, \ldots, x'_{k-1}) \in \widehat{E}_{k-1}$ and $y_k = (y'_0, \ldots, y'_k) \in \widehat{E}_k$ by the following formula

$$\widehat{M}_k(x_{k-1}, dy_k) = \delta_{x_{k-1}}(dy_{k-1})\ \widehat{\eta}_k(dy'_k)\ .$$

Notice that the restriction $\widehat{\mathbb{P}}_{\widehat{\eta}'_0,n}$ of these approximated reference measure $\widehat{\mathbb{P}}_{\widehat{\eta}'_0}$ to the σ-field $\mathcal{F}_n$ generated by the canonical random sequence $(X'_k)_{0\leq k\leq n}$ is now given by the the tensor product measure $\widehat{\mathbb{P}}_{\widehat{\eta}'_0,n} = \otimes_{k=0}^n \widehat{\eta}_k$.

Let $\widehat{u}_k$ be the Snell envelope on the path space $\widehat{E}_k$, associated with the pair $(\widehat{f}_k, \widehat{M}_k)$, with the sequence of functions $\widehat{f}_k = f_k$ given in (3.4.3). By construction, for any $0 \leq k \leq n$, and any path $x_k = (x'_0, \ldots, x'_k) \in \widehat{E}_k$, we have

$$\widehat{u}_k(x_k) = \widehat{u}'_k(x'_k) \times \frac{d\mathbb{P}'_k}{d\mathbb{P}_k}(x_k)\ ,$$

with the collection of functions $(\widehat{u}'_k)_{0\leq k\leq n}$ on the state spaces $(E'_k)_{0\leq k\leq n}$ given by the backward recursions

$$\widehat{u}'_k(x'_k) = f'_k(x'_k) \vee \left(\int_{\widehat{E}'_{k+1}} \widehat{M}'_{k+1}(x'_k, dx'_{k+1})\ \widehat{u}'_{k+1}(x'_{k+1}) \right)\ , \tag{3.4.5}$$

with the random integral operator $\widehat{M}'_k$ from E'_k into $\widehat{E}'_{k+1}$ defined below

$$\widehat{M}'_{k+1}(x'_k, dx'_{k+1}) = \widehat{\eta}_{k+1}(dx'_{k+1})\ R_{k+1}(x'_k, x'_{k+1})\ ,$$

with the Radon-Nikodym derivatives $R_{k+1}(x'_k, x'_{k+1}) = \frac{dM'_{k+1}(x'_k,.)}{d\eta_{k+1}}(x'_{k+1})$.

3.4.2 Broadie-Glasserman models

We consider the path space models associated to the change of measures presented in Section 3.4.1. We use the same notation. We further assume that $\widehat{\eta}_k = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_k^i}$ is the occupation measure associated with a sequence of independent random variables $\xi_k := (\xi_k^i)_{1\leq i\leq N}$ with common distribution η_k on $\widehat{E}'_k = E'_k$. We further assume that $(\xi_k)_{0\leq k\leq n}$ are independent. This Monte Carlo type model has been introduced in 1997 by Broadie and Glasserman (see for instance [19], and references therein). Let $\widehat{\mathbb{E}}$ be the expectation operator associated with this additional level of randomness, and we set $\widehat{\mathbb{E}}_{\eta_0} := \widehat{\mathbb{E}} \otimes \mathbb{E}_{\mathbb{P}_{\eta_0}}$.

In this situation, we observe that

$$(M'_{k+1} - \widehat{M}'_{k+1})(x'_k, dx'_{k+1}) = \frac{1}{\sqrt{N}}\ \widehat{V}_{k+1}(dx'_{k+1})\ R_{k+1}(x'_k, x'_{k+1})\ ,$$

with the random fields $\widehat{V}_{k+1} := \sqrt{N}\ [\eta_{k+1} - \widehat{\eta}_{k+1}]$. From these observations, we readily prove that the approximation operators $\widehat{M}'_{k+1}$ are unbiased, in the sense that

$$\forall 0 \leq k \leq l \quad \forall x'_l \in E_l \qquad \widehat{\mathbb{E}}_{\eta_0}\left(\widehat{M}'_{k,l}(f)(x'_l)\,|\mathcal{F}_k \right) = M'_{k,l}(f)(x'_l)\ , \tag{3.4.6}$$

for any bounded function f on E_{l+1}. Furthermore, for any even integer $p \geq 1$, we have

$$\sqrt{N}\,\widehat{\mathbb{E}}_{\eta_0}\left(\left|\left[M'_{l+1}-\widehat{M}'_{l+1}\right](f)(x'_l)\right|^p\right)^{\frac{1}{p}} \leq 2\;a(p)\;\eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)f)^p\right]^{\frac{1}{p}}\;.$$

The above estimate is valid as soon as the r.h.s. in the above inequality is well defined.

We are now in position to state and prove the following theorem.

Theorem 3.4.1. *For any integer $p \geq 1$, we denote by p' the smallest even integer greater than p. Then for any time horizon $0 \leq k \leq n$, and any $x'_k \in E'_k$, we have*

$$\begin{aligned}\sqrt{N}\widehat{\mathbb{E}}_{\eta_0}\left(|u'_k(x'_k)-\widehat{u}'_k(x'_k)|^p\right)^{\frac{1}{p}} \qquad (3.4.7)\\ \leq 2a(p)\sum_{k\leq l<n}\left\{\int M'_{k,l}(x'_k,dx'_l)\eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)u'_{l+1})^{p'}\right]\right\}^{\frac{1}{p'}}\;.\end{aligned}$$

Note that, as stated in the introduction, this result implies exponential rate of convergence in probability. Hence, this allows to improve noticeably existing convergence results stated in [19], where there was no rate of convergence, and in [3], where the rate of convergence in probability was polynomial.

Proof. For any even integers $p \geq 1$, any $0 \leq k \leq l$, any measurable function f on E_{l+1}, and any $x_k \in E'_k$, using the generalized Minkowski inequality we find that

$$\begin{aligned}\sqrt{N}\;\widehat{\mathbb{E}}_{\eta_0}\left(\left|\widehat{M}'_{k,l}\left|\left[M'_{l+1}-\widehat{M}'_{l+1}\right](f)\right|(x'_k)\right|^p|\mathcal{F}_l\right)^{\frac{1}{p}}\\ \leq 2a(p)\;\int\;\widehat{M}'_{k,l}(x'_k,dx'_l)\;\eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)f)^p\right]^{\frac{1}{p}}\;.\end{aligned}$$

By the zero-bias property (3.4.6), we conclude that

$$\begin{aligned}\sqrt{N}\;\widehat{\mathbb{E}}_{\eta_0}\left(\left|\widehat{M}'_{k,l}\left|\left[M'_{l+1}-\widehat{M}'_{l+1}\right](f)\right|(x'_k)\right|^p\right)^{\frac{1}{p}}\\ \leq 2a(p)\;\left\{\int\;M'_{k,l}(x'_k,dx'_l)\;\eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)f)^p\right]\right\}^{1/p}\;.\end{aligned}$$

For odd integers $p=2q+1$, with $q \geq 0$, we use the fact that

$$\mathbb{E}(Y^{2q+1})^2 \leq \mathbb{E}(Y^{2q})\;\mathbb{E}(Y^{2(q+1)}) \quad \text{and} \quad \mathbb{E}(Y^{2q}) \leq \mathbb{E}(Y^{2(q+1)})^{\frac{q}{q+1}}\;,$$

for any non negative random variable Y and

$$(2(q+1))_{q+1} = 2\;(2q+1)_{q+1} \quad \text{and} \quad (2q)_q = (2q+1)_{q+1}/(2q+1)\;,$$

so that

$$a(2q)^{2q}a(2(q+1))^{2(q+1)} \leq 2^{-(2q+1)}(2q+1)^2_{q+1}/(q+1/2) = \left(a(2q+1)^{2q+1}\right)^2\;,$$

and

$$N\ \widehat{\mathbb{E}}_{\eta_0}\left(\left|\widehat{M}'_{k,l}\left|\left[M'_{l+1}-\widehat{M}'_{l+1}\right](f)\right|(x'_k)\right|^{2q+1}\right)^2$$

$$\leq \left(2^{(2q+1)}a(2q+1)^{2q+1}\right)^2\ \int\ M'_{k,l}(x'_k,dx'_l)\ \eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)f)^{2(q+1)}\right]^{\frac{q}{q+1}}$$

$$\times \int\ M'_{k,l}(x'_k,dx'_l)\ \eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)f)^{2(q+1)}\right]\ .$$

Using the fact that $\mathbb{E}(Y^{\frac{q}{q+1}}) \leq \mathbb{E}(Y)^{\frac{q}{q+1}}$, we prove that the r.h.s. term in the above display is upper bounded by

$$\left(2^{(2q+1)}a(2q+1)^{2q+1}\right)^2\left\{\int M'_{k,l}(x'_k,dx'_l)\eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)f)^{2(q+1)}\right]\right\}^{2\left(1-\frac{1}{2(q+1)}\right)}\ ,$$

from which we conclude that

$$\sqrt{N}\ \widehat{\mathbb{E}}_{\eta_0}\left(\left|\widehat{M}'_{k,l}\left|\left[M'_{l+1}-\widehat{M}'_{l+1}\right](f)\right|(x'_k)\right|^{2q+1}\right)^{\frac{1}{2q+1}}$$

$$\leq 2a(2q+1)\ \left\{\int\ M'_{k,l}(x'_k,dx'_l)\ \eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)f)^{2(q+1)}\right]\right\}^{\frac{1}{2(q+1)}}\ .$$

That complete the proof of the theorem. ∎

The $\mathbb{L}_p$-mean error estimates stated in Theorem 3.4.1 are expressed in terms of $\mathbb{L}_{p'}$ norms of Snell envelope functions and Radon-Nikodym derivatives. The terms in the r.h.s. of (3.4.7) have the following interpretation:

$$\int\ M'_{k,l}(x'_k,dx'_l)\ \eta_{l+1}\left[(R_{l+1}(x'_l,\cdot)u_{l+1})^{p'}\right]$$

$$= \mathbb{E}\left[\left(R_{l+1}(X'_l,\xi^1_{l+1})u_{l+1}(\xi^1_{l+1})\right)^{p'}|X'_k=x'_k\right]\ .$$

In the above display, $\mathbb{E}(\cdot)$ stands for the expectation with respect to some reference probability measure under which X'_l is a Markov chain with transitions M'_l, and ξ^1_{l+1} is an independent random variable with distribution η_{l+1}. Loosely speaking, the above quantities can be very large when the sampling distributions η_{l+1} are far from the distribution of the random states X'_{l+1} of the reference Markov chain at time $(l+1)$. Next, we provide an original strategy that allows for instance to take $\eta_{l+1}=\mathrm{law}(X'_{l+1})$ as the sampling distribution, even if R_{l+1} is not known (i.e. cannot be evaluated at any point of E_{l+1}). In the sequel, we consider N independent copies $(\xi^i_0,\cdots\xi^i_n)_{1\leq i\leq N}$ of the Markov chain $(X'_0,X'_1,\cdots X'_n)$, from the origin $k=0$, up to the final time horizon $k=n$. Then, for all $k=0,\cdots n$, we define the associated occupation measure $\widehat{\eta}_k=\frac{1}{N}\sum_{i=1}^N\delta_{\xi^i_k}$. For all $k=0,\cdots n$, let $\mathcal{F}_k$ be the sigma field generated by the random sequence $(\xi_l)_{0\leq l\leq k}$.

We also assume that the Markov transitions $M'_n(x'_{n-1}, dx'_n)$ are absolutely continuous with respect to some measures $\lambda_n(dx'_n)$ on E'_n and we have

$$(H)_0 \qquad \forall (x'_{n-1}, x'_n) \in \left(E'_{n-1} \times E'_n\right) \quad H_n(x'_{n-1}, x'_n) = \frac{dM'_n(x'_{n-1}, .)}{d\lambda_n}(x'_n) > 0 \ ,$$

where H_n is supposed to be known up to a normalizing constant. In this situation, we have $\eta_{k+1} \ll \lambda_{k+1}$, with the Radon-Nikodym derivative given below

$$\eta_{k+1}(dx'_{k+1}) = \eta_k M'_{k+1}(dx'_{k+1}) = \eta_k \left(H_{k+1}(\cdot, x'_{k+1})\right) \ \lambda_{k+1}(dx'_{k+1}) \ .$$

Also notice that the backward recursion of the Snell envelope u'_k can be rewritten as

$$\begin{aligned} u'_k(x'_k) &= f'_k(x'_k) \vee \left(\int_{E'_{k+1}} \eta_{k+1}(dx'_{k+1}) \ \frac{dM'_{k+1}(x'_k, .)}{d\eta_{k+1}}(x'_{k+1}) \ u'_{k+1}(x'_{k+1}) \right) \\ &= f'_k(x'_k) \vee \left(\int_{E'_{k+1}} \eta_{k+1}(dx'_{k+1}) \ \frac{H_{k+1}(x'_k, x'_{k+1})}{\eta_k(H_{k+1}(\cdot, x'_{k+1}))} \ u'_{k+1}(x'_{k+1}) \right) \ . \end{aligned}$$

Arguing as in (3.4.5), we define the approximated Snell envelope $(\widehat{u}'_k)_{0 \leq k \leq n}$ on the state spaces $(E'_k)_{0 \leq k \leq n}$ by setting

$$\widehat{u}'_k(x'_k) = f'_k(x'_k) \vee \left(\int_{\widehat{E}'_{k+1}} \widehat{M}'_{k+1}(x'_k, dx'_{k+1}) \ \widehat{u}'_{k+1}(x'_{k+1}) \right) \ ,$$

with the random integral operator $\widehat{M}'$ from E_k into $\widehat{E}_{k+1}$ defined below

$$\begin{aligned} \widehat{M}'_{k+1}(x'_k, dx'_{k+1}) &= \widehat{\eta}_{k+1}(dx'_{k+1}) \ \frac{dM'_{k+1}(x'_k, .)}{d\widehat{\eta}_k M'_{k+1}}(x'_{k+1}) \\ &= \widehat{\eta}_{k+1}(dx'_{k+1}) \ \frac{H_{k+1}(x'_k, x'_{k+1})}{\widehat{\eta}_k(H_{k+1}(\cdot, x'_{k+1}))} \ . \end{aligned}$$

By construction, these random approximation operators $\widehat{M}'_{k+1}$ satisfy the zero-bias property stated in (3.4.6), and we have

$$(M'_{k+1} - \widehat{M}'_{k+1})(x'_k, dx'_{k+1}) = \frac{1}{\sqrt{N}} \ \widehat{V}_{k+1}(dx'_{k+1}) \ \widehat{R}_{k+1}(x'_k, x'_{k+1}) \ ,$$

with the random fields $\widehat{V}_{k+1}$ and the $\mathcal{F}_k$-measurable random functions $\widehat{R}_{k+1}$ defined below

$$\widehat{V}_{k+1} := \sqrt{N} \ [\widehat{\eta}_k M'_{k+1} - \widehat{\eta}_{k+1}] \quad \text{and} \quad \widehat{R}_{k+1}(x'_k, x'_{k+1}) := \frac{H_{k+1}(x'_k, x'_{k+1})}{\widehat{\eta}_k(H_{k+1}(\cdot, x'_{k+1}))} \ .$$

Furthermore, for any even integer $p \geq 1$, and any measurable function f on E_l we have

$$\sqrt{N} \ \widehat{\mathbb{E}}_{\eta_0} \left(\left| \left[M'_{l+1} - \widehat{M}'_{l+1} \right] (f)(x'_l) \right|^p | \mathcal{F}_l \right)^{\frac{1}{p}} \leq 2 \ a(p) \ \widehat{\eta}_l M'_{l+1} \left[(\widehat{R}_{l+1}(x'_l, \cdot) f)^p \right]^{\frac{1}{p}} \ .$$

The above estimate is valid as soon as the r.h.s. in the above inequality is well defined. For instance, assuming that

$$(H)_1 \qquad \|M'_{l+1}(u_{l+1}^{2p})\| < \infty$$
$$\text{and} \qquad \sup_{x'_l, y'_l \in E'_l} \frac{H_{l+1}(x'_l, x'_{l+1})}{H_{l+1}(y'_l, x'_{l+1})} \leq h_{l+1}(x'_{l+1}) \text{ with } \|M'_{l+1}(h_{l+1}^{2p})\| < \infty \ ,$$

we find that

$$\sqrt{N}\ \mathbb{E}\left(\left|\left[M'_{l+1} - \widehat{M}'_{l+1}\right](u'_{l+1})(x'_l)\right|^p |\mathcal{F}_l\right)^{\frac{1}{p}}$$
$$\leq 2\ a(p)\ \left(\|M'_{l+1}(h_{l+1}^{2p})\|\ \|M'_{l+1}((u'_{l+1})^{2p})\|\right)^{\frac{1}{2p}} \ .$$

Rephrasing the proof of Theorem 3.4.1, we just proved the following result.

Theorem 3.4.2. *Under the conditions $(H)_0$ and $(H)_1$ stated above, for any even integer $p > 1$, any $0 \leq k \leq n$, and $x'_k \in E'_k$, we have*

$$\sqrt{N}\ \mathbb{E}\left(|u'_k(x'_k) - \widehat{u}'_k(x'_k)|^p\right)^{\frac{1}{p}} \tag{3.4.8}$$
$$\leq 2a(p) \sum_{k \leq l < n} \left(\|M'_{l+1}(h_{l+1}^{2p})\|\ \|M'_{l+1}((u'_{l+1})^{2p})\|\right)^{\frac{1}{2p}} \ .$$

In the end, recovering and extending results from [19], it is interesting to point out that both the Broadie-Glasserman estimator and this new BG type adapted estimator have positive bias.

Proposition 3.4.2. *For any $0 \leq k \leq n$ and any $x'_k \in E'_k$*

$$\mathbb{E}\left(\widehat{u}'_k(x'_k)\right) \geq u'_k(x'_k) \ . \tag{3.4.9}$$

Proof. This inequality can be proved easily by a simple backward induction. The terminal condition $\widehat{u}'_n = u'_n$ implies directly the inequality on instant n. Assuming the inequality holds true in instant k, then Jensen's inequality implies that

$$\begin{aligned} \mathbb{E}\left(\widehat{u}'_k(x'_k)\right) &\geq f_k(x'_k) \vee \mathbb{E}\left(\widehat{M}'_{k+1}(\widehat{u}'_{k+1})(x'_k)\right) \\ &\geq f_k(x'_k) \vee M_{k+1}u'_{k+1}(x'_k) = u'_k(x'_k) \ , \end{aligned}$$

completing the proof of the proposition. ∎

3.5 A genealogical tree based model

3.5.1 Neutral genetic models

Using the notations of Section 3.4.1, set

$$X_n = (X'_0, \ldots, X'_n) \in E_n = (E'_0 \times \ldots \times E'_n) \ .$$

Further assume that the state spaces E'_n are finite, and denote by η_k the distribution of the path-valued random variable X_k on E_k, with $0 \leq k \leq n$.

Further let M'_k be the Markov transition from X'_{k-1} to X'_k, and let M_k be the Markov transition from X_{k-1} to X_k. In Section 3.4.1, we have seen that

$$M_k((x'_0,\ldots,x'_{k-1}),d(y'_0,\ldots,y'_k)) = \delta_{(x'_0,\ldots,x'_{k-1})}(d(y'_0,\ldots,y'_{k-1}))\ M'_k(y'_{k-1},dy'_k)\ .$$

In the further development, we fix the final time horizon n, and for any $0 \leq k \leq n$, we denote by π_k the k-th coordinate mapping

$$\pi_k\ :\ x_n = (x'_0,\ldots,x'_n) \in E_n = (E'_0 \times \ldots \times E'_n) \mapsto \pi_k(x_n) = x'_k \in E'_k\ .$$

In this notation, for any $0 \leq k < n$, $x'_k \in E'_k$ and any function $f \in \mathcal{B}(E'_{k+1})$, we have

$$\eta_n = \mathrm{Law}(X'_0,\ldots,X'_n) \quad \text{and} \quad M'_{k+1}(f)(x) := \frac{\eta_n((1_x \circ \pi_k)\ (f \circ \pi_{k+1}))}{\eta_n((1_x \circ \pi_k))}\ . \tag{3.5.1}$$

By construction, it is also readily checked that the flow of measure $(\eta_k)_{0\leq k\leq n}$ also satisfies the following equation

$$\eta_k := \Phi_k\left(\eta_{k-1}\right)\ , \qquad \forall\, 1 \leq k \leq n, \tag{3.5.2}$$

with the linear mapping $\Phi_k\left(\eta_{k-1}\right) := \eta_{k-1}M_k$.

The genealogical tree based particle approximation associated with these recursion is defined in terms of a Markov chain $\xi_k^{(N)} = (\xi_k^{(i,N)})_{1\leq i\leq N_k}$ in the product state spaces $E_k^{N_k}$, where $N = (N_k)_{0\leq k\leq N}$ is a given collection of integers.

$$\mathbb{P}\left(\xi_k^{(N)} = (x_k^1,\ldots,x_k^{N_k})\mid \xi_{k-1}\right) = \prod_{1\leq i\leq N_k} \Phi_k\left(\frac{1}{N_{k-1}}\sum_{1\leq i\leq N_{k-1}} \delta_{\xi_{k-1}^i}\right)\left(x_k^i\right)\ . \tag{3.5.3}$$

The initial particle system $\xi_0^{(N)} = \left(\xi_0^{(i,N)}\right)_{0\leq i\leq N_0}$, is a sequence of N_0 i.i.d. random copies of X_0. Let $\mathcal{F}_k^N$ be the sigma-field generated by the particle approximation model from the origin, up to time k.

To simplify the presentation, when there is no confusion we suppress the population size parameter N, and we write ξ_k and ξ_k^i instead of $\xi_k^{(N)}$ and $\xi_k^{(i,N)}$. By construction, ξ_k is a genetic type model with a neutral selection transition and a mutation type exploration

$$\xi_k \in E_k^{N_k} \xrightarrow{\text{Selection}} \widehat{\xi}_k := \left(\widehat{\xi}_k^i\right)_{1\leq i\leq \widehat{N}_k} \in E_k^{\widehat{N}_k} \xrightarrow{\text{Mutation}} \xi_{k+1} \in E_{k+1}^{N_{k+1}}\ , \tag{3.5.4}$$

with $\widehat{N}_k := N_{k+1}$.

During the selection transition, we select randomly N_{k+1} path valued particles $\widehat{\xi}_k := \left(\widehat{\xi}_k^i\right)_{1\leq i\leq N_{k+1}}$ among the N_k path valued particles $\xi_k = (\xi_k^i)_{1\leq i\leq N_k}$. Sometimes,

this elementary transition is called a neutral selection transition in the literature on genetic population models. During the mutation transition $\widehat{\xi}_k \rightsquigarrow \xi_k$, every selected path valued individual $\widehat{\xi}_k^i$ evolves randomly to a new path valued individual $\xi_{k+1}^i = x$ randomly chosen with the distribution $M_{k+1}(\widehat{\xi}_k^i, x)$, with $1 \leq i \leq \widehat{N}_k$. By construction, every particle is a path-valued random variable defined by

$$\begin{array}{rcl} \xi_k^i & := & \left(\xi_{0,k}^i, \xi_{1,k}^i, \ldots, \xi_{k,k}^i\right) \\ \widehat{\xi}_k^i & := & \left(\widehat{\xi}_{0,k}^i, \widehat{\xi}_{1,k}^i, \ldots, \widehat{\xi}_{k,k}^i\right) \in E_k := (E_0' \times \ldots \times E_k') \ . \end{array}$$

By definition of the transition in path space, we also have that

$$\begin{array}{rcl} \xi_{k+1}^i & = & \left(\underbrace{\left(\xi_{0,k+1}^i, \xi_{1,k+1}^i, \ldots, \xi_{k,k+1}^i\right)}_{||}, \xi_{k+1,k+1}^i \right) \\ & = & \left(\overbrace{\left(\widehat{\xi}_{0,k}^i, \ \widehat{\xi}_{1,k}^i, \ldots, \ \widehat{\xi}_{k,k}^i\right)}, \ \xi_{k+1,k+1}^i \right) = \left(\widehat{\xi}_k^i, \xi_{k+1,k+1}^i\right) \ , \end{array}$$

where $\xi_{k+1,k+1}^i$ is a random variable with distribution $M_{k+1}'(\widehat{\xi}_{k,k}^i, \cdot)$. In other words, the mutation transition $\widehat{\xi}_k^i \rightsquigarrow \xi_{k+1}^i$ simply consists in extending the selected path $\widehat{\xi}_k^i$ with an elementary move $\widehat{\xi}_{k,k}^i \rightsquigarrow \xi_{k+1,k+1}^i$ of the end point of the selected path.

From these observations, it is easy to check that the terminal random population model $\xi_{k,k} = \left(\xi_{k,k}^i\right)_{1 \leq i \leq N_k}$ and $\widehat{\xi}_{k,k} = \left(\widehat{\xi}_{k,k}^i\right)_{1 \leq i \leq N_{k+1}}$ is again defined as a genetic type Markov chain defined as above by replacing the pair (E_k, M_k) by the pair (E_k', M_k'), with $1 \leq k \leq n$. The latter coincides with the mean field particle model associated with the time evolution of the k-th time marginals η_k' of the measures η_k on E_k'. Furthermore, the above path-valued genetic model coincide with the genealogical tree evolution model associated with the terminal state random variables.

Let η_k^N and $\widehat{\eta}_k^N$ be the occupation measures of the genealogical tree model after the mutation and the selection steps; that is, we have that

$$\eta_k^N := \frac{1}{N_k} \sum_{1 \leq i \leq N_k} \delta_{\xi_k^i} \quad \text{and} \quad \widehat{\eta}_k^N := \frac{1}{\widehat{N}_k} \sum_{1 \leq i \leq \widehat{N}_k} \delta_{\widehat{\xi}_k^i} \ .$$

In this notation, the selection transition $\xi_k, \rightsquigarrow \widehat{\xi}_k$ consists in choosing $\widehat{N}_k$ conditionally independent and identically distributed random paths $\widehat{\xi}_k^i$ with common distribution η_k^N. In other words, $\widehat{\eta}_k^N$ is the empirical measure associated with $\widehat{N}_k$ conditionally independent and identically distributed random paths $\widehat{\xi}_k^i$ with common distribution η_k^N. Also observe η_k^N is the empirical measure associated with N_k conditionally independent and identically distributed random paths ξ_k^i with common distribution $\eta_{k-1}^N M_k$.

In practice, we can take $N_0 = N_1 = \cdots = N_n = N$ when we do not have any information on the variance of X_k. In the case when we know the approximate variance of X_k, we can take a large N_k when the variance of X'_k is large. To clarify the presentation, In the further development of the chapter we further assume that the particle model has a fixed population size $N_k = N$, for any $k \geq 0$.

In the sequel, the simulation of the path valued particle system $(\xi_k)_{0\leq k\leq n}$ will be called the *Forward step* and is summarized in the following algorithm.

Forward algorithm

Initialization At time step $k = 0$, generate N i.i.d. random copies of X_0 and set $\xi_0 = (\xi_0^i)_{0\leq i\leq N}$.

At each time step $k = 1, \cdots, n$

1. **Selection**: For each $i = 1, \cdots, N$, generate independently an indice $I_i \in \{1, \cdots, N\}$ with probability $\mathbb{P}(I_i = j) = 1/N$. Then set $\hat{\xi}_{k-1}^i = \xi_{k-1}^{I_i}$.
2. **Mutation**: For each $i = 1, \cdots, N$, generate independently N i.i.d. random variables $(\xi_{k,k}^i)_{0\leq i\leq N}$ according to the transition kernel $M'_k(\hat{\xi}_{k-1,k-1}^i, \cdot)$. Then set $\xi_k^i = (\hat{\xi}_{k-1}^i, \xi_{k,k}^i)$.

3.5.2 Convergence analysis

For general mean field particle interpretation models (3.5.3), several estimates can be derived for the above particle approximation model (see for instance [40]). For instance, for any $n \geq 0$, $r \geq 1$, and any $f_n \in \mathrm{Osc}_1(E_n)$, and any $N \geq 1$, we have the unbiased and the mean error estimates:

$$\mathbb{E}\left(\eta_n^N(f_n)\right) = \eta_n(f_n) = \mathbb{E}\left(\widehat{\eta}_n^N(f_n)\right) \tag{3.5.5}$$

$$\text{and} \qquad \sqrt{N}\ \mathbb{E}\left(\left|\left[\eta_n^N - \eta_n\right](f_n)\right|^r\right)^{\frac{1}{r}} \leq 2\ a(r)\ \sum_{p=0}^{n} \beta(M_{p,n})\ ,$$

with the Dobrushin ergodic coefficients

$$\beta(M_{p,n}) := \sup_{(x_p, y_p \in E_p)} \left\|M_{p,n}(x_p, \cdot) - M_{p,n}(y_p, \cdot)\right\|_{tv}\ ,$$

and the collection of constants $a(p)$ defined in (3.2.7). Arguing as in (4.5.3), for time homogeneous population sizes $N_n = N$, for any functions $f \in \mathrm{Osc}_1(E_n)$, we conclude that

$$\mathbb{P}\left(\left|\left[\eta_n^N - \eta_n\right](f)\right| \geq \frac{b(n)}{\sqrt{N}} + \epsilon\right) \leq \exp\left(-\frac{N\epsilon^2}{2b(n)^2}\right) \tag{3.5.6}$$

$$\text{with} \quad b(n) := 2\ \sum_{p=0}^{n} \beta(M_{p,n})\ .$$

For the path space models (3.5.1), we have $\beta(M_{p,n}) = 1$ so that the estimates (4.5.1) and (3.5.6) takes the form

$$\sqrt{N}\ \mathbb{E}\left(\left|\left[\eta_n^N - \eta_n\right](f_n)\right|^r\right)^{\frac{1}{r}} \leq 2\ a(r)\ (n+1) \tag{3.5.7}$$

and

$$\mathbb{P}\left(\left|\left[\eta_n^N - \eta_n\right](f)\right| \geq \frac{2(n+1)}{\sqrt{N}} + \epsilon\right) \leq \exp\left(-\frac{N\epsilon^2}{8(n+1)^2}\right) .$$

In the next lemma we extend these estimates to unbounded functions.

Lemma 3.5.1. *For any $p \geq 1$, we denote by p' the smallest even integer greater than p. In this notation, for any $k \geq 0$ and any function f, we have the almost sure estimate*

$$\begin{aligned}\sqrt{N}\mathbb{E}\left(\left|[\eta_n^N - \eta_{k-1}^N M_{k-1,n}](f)\right|^p \left|\mathcal{F}_{k-1}^N\right.\right)^{\frac{1}{p}} \\ \leq 2a(p)\ \sum_{l=k}^{n}\ \left[\eta_{k-1}^N M_{k-1,l}(|M_{l,n}(f)|^{p'})\right]^{\frac{1}{p'}} .\end{aligned} \tag{3.5.8}$$

In particular, for any $f \in \mathbb{L}_{p'}(\eta_n)$, we have the non asymptotic estimates

$$\sqrt{N}\ \mathbb{E}\left(\left|[\eta_n^N - \eta_n](f)\right|^p\right)^{1/p} \leq 2\ a(p)\ \|f\|_{p',\eta_n}\ (n+1) . \tag{3.5.9}$$

Proof. Writing $\eta_{-1}^N M_0 = \eta_0$, for any $k \geq 0$, we have the decomposition

$$[\eta_n^N - \eta_{k-1}^N M_{k,n}] = \sum_{l=k}^{n} [\eta_l^N - (\eta_{l-1}^N M_l)] M_{l,n} ,$$

with the semigroup

$$M_{k,n} = M_{k+1} M_{k+2} \ldots M_n .$$

Using the fact that

$$\mathbb{E}\left(\eta_l^N(f)\left|\eta_{l-1}^N\right.\right) = (\eta_{l-1}^N M_l)(f) ,$$

we obtain that

$$\mathbb{E}\left(\left|[\eta_l^N - (\eta_{l-1}^N M_l)](f)\right|^p \left|\mathcal{F}_{l-1}^N\right.\right)^{\frac{1}{p}} \leq \mathbb{E}\left(\left|[\eta_l^N - \mu_l^N](f)\right|^p \left|\mathcal{F}_{l-1}^N\right.\right)^{\frac{1}{p}} ,$$

where $\mu_l^N := \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_l^i}$ stands for a independent copy of η_l^N given η_{l-1}^N. Using Khinchine's type inequalities, we have

$$\begin{aligned}\sqrt{N}\ \mathbb{E}\left(\left|[\eta_l^N - \mu_l^N](f)\right|^p \left|\mathcal{F}_{l-1}^N\right.\right)^{\frac{1}{p}} &\leq 2\ a(p)\ \mathbb{E}\left(\left|f\left(\xi_l^1\right)\right|^{p'} \mid \mathcal{F}_{l-1}^N\right)^{\frac{1}{p'}} \\ &= 2\ a(p)\ \left[\eta_{l-1}^N M_l(|f|^{p'})\right]^{\frac{1}{p'}} .\end{aligned}$$

Using the unbias property of the particle scheme, we have

$$\forall k \leq l \leq n \qquad \mathbb{E}\left(\eta_l^N(f)\left|\mathcal{F}_{k-1}^N\right.\right) = (\eta_{k-1}^N M_{k-1,l})(f) .$$

This implies that

$$\begin{aligned}\sqrt{N}\ \mathbb{E}\left(\left|[\eta_l^N-(\eta_{l-1}^N M_l)](f)\right|^p\left|\mathcal{F}_{k-1}^N\right.\right)^{\frac{1}{p}} &\leq 2\ a(p)\ \mathbb{E}\left(\eta_{l-1}^N M_l(|f|^{p'})\left|\mathcal{F}_{k-1}^N\right.\right)^{\frac{1}{p'}}\\ &= 2\ a(p)\ \left[\eta_{k-1}^N M_{k-1,l}(|f|^{p'})\right]^{\frac{1}{p'}}\ .\end{aligned}$$

The end of the proof of (3.5.8) is now a direct application of Minkowski's inequality, while the proof of (3.5.9) is a direct consequence of (3.5.8). ∎

3.5.3 Particle approximations of the Snell envelope

In Section 3.5.1, we have presented a genealogical based algorithm whose occupation measures η_n^N converge, as $N\uparrow\infty$, to the distribution η_n of the reference Markov chain $(X'_0,\ldots,X'_n)$ from the origin, up to the final time horizon n. Mimicking formula (3.5.1), we define the particle approximation of the Markov transitions M'_k as follows :

$$\widehat{M}'_{k+1}(f)(x):=\frac{\eta_n^N((1_x\circ\pi_k)\ (f\circ\pi_{k+1}))}{\eta_n^N((1_x\circ\pi_k))}:=\frac{\sum_{1\leq i\leq N}\ 1_x(\xi_{k,n}^i)\ f(\xi_{k+1,n}^i)}{\sum_{1\leq i\leq N}1_x(\xi_{k,n}^i)}\ ,$$

for every state x in the support $\widehat{E}_{k,n}$ of the measure $\eta_n^N\circ\pi_k^{-1}$. Note that $\widehat{E}_{k,n}$ coincides with the collection of ancestors $\xi_{k,n}^i$ at level k of the population of individuals at the final time horizon. This random set can alternatively be defined as the set of states $\xi_{k,k}^i$ of the particle population at time k such that $\eta_n^N((1_{\xi_{k,k}^i}\circ\pi_k))>0$; more formally, we have

$$\widehat{E}_{k,n}:=\cup_{1\leq i\leq N}\left\{\xi_{k,k}^i\ :\ \eta_n^N((1_{\xi_{k,k}^i}\circ\pi_k))>0\right\}\ . \tag{3.5.10}$$

It is interesting to observe that the random Markov transitions $\widehat{M}'_{k+1}$ coincides with the conditional distributions of the states X'_{k+1} given the current time states X'_k of a canonical Markov chain $X_n:=(X'_0,\ldots,X'_n)$ with distribution η_n^N on the path space $E_n:=(E'_0\times\ldots\times E'_n)$. Thus, the flow of k-th time marginal measures

$$\eta_{k,n}^N:=\frac{1}{N}\sum_{i=1}^N\ \delta_{\xi_{k,n}^i}\ ,$$

are connected by the following formula

$$\eta_{k,n}^N\widehat{M}'_{k,l}=\eta_{l,n}^N\ ,\qquad\forall\, k\leq l\leq n,$$

with the semigroup $\widehat{M}'_{k,l}$ associated with the Markov transitions $\widehat{M}'_{k+1}$ given by

$$\widehat{M}'_{k,l}(f)(x)=\widehat{M}'_{k+1}\widehat{M}'_{k+1}\ldots\widehat{M}'_l(f)(x)=\frac{\eta_n^N((1_x\circ\pi_k)\ (f\circ\pi_l))}{\eta_n^N((1_x\circ\pi_k))}\ , \tag{3.5.11}$$

for every state x in $\widehat{E}_{k,n}$. In connection with (3.5.10), we also have the following formula

$$\eta_{k,n}^N = \frac{1}{N}\sum_{i=1}^N \left(N\ \eta_n^N\left(1_{\xi_{k,k}^i}\circ\pi_k\right)\right)\delta_{\xi_{k,k}^i} = \sum_{i=1}^N \eta_n^N\left(1_{\xi_{k,k}^i}\circ\pi_k\right)\delta_{\xi_{k,k}^i}\ ,$$

with the proportion $\eta_n^N\left(1_{\xi_{k,k}^i}\circ\pi_k\right)$ of individuals at the final time horizon having the common ancestor $\xi_{k,k}^i$ at level k. It is also interesting to observe that

$$\begin{aligned}\mathbb{E}\left(\eta_{k,n}^N(f)\,\middle|\,\mathcal{F}_k^N\right) &= \sum_{i=1}^N \mathbb{E}\left(\eta_n^N\left(1_{\xi_{k,k}^i}\circ\pi_k\right)\middle|\,\mathcal{F}_k^N\right) f(\xi_{k,k}^i)\\ &= \sum_{i=1}^N \underbrace{\eta_k^N M_{k,n}\left(1_{\xi_{k,k}^i}\circ\pi_k\right)}_{=1/N} f(\xi_{k,k}^i) = \eta_k^N(f)\ .\end{aligned}$$

The Snell envelope associated with this particle approximation model is defined by the backward recursion:

$$\widehat{u}_k(x) = \begin{cases} f_k(x)\vee \widehat{M}'_{k+1}(u_{k+1})(x) & \forall x\in\widehat{E}_{k,n}\\ 0 & \text{otherwise}\ .\end{cases}$$

In terms of the ancestors at level k, this recursion takes the following form

$$\widehat{u}_k\left(\xi_{k,n}^i\right) = f_k\left(\xi_{k,n}^i\right)\vee \widehat{M}'_{k+1}(\widehat{u}_{k+1})\left(\xi_{k,n}^i\right)\ , \qquad \forall\, 1\le i\le N.$$

In the sequel, the computation of the Snell envelope approximation $(\hat{u}_k)_{0\le k\le n}$ will be called the *Backward step* and is summarized in the following algorithm.

Backward algorithm

Initialization At time step $k=n$, for all $i=1,\cdots,N$, set $\hat{u}_n(\xi_{n,n}^i) = f(\xi_{n,n}^i)$.

At each time step $k=n-1,\cdots,0$, for all $i=1,\cdots,N$ set

$$\hat{u}_k(\xi_{k,n}^i) = f_k(\xi_{k,n}^i)\vee \frac{\sum_{j=1}^N \hat{u}_{k+1}(\xi_{k+1,n}^j)\,1_{\xi_{k,n}^j=\xi_{k,n}^i}}{\sum_{j=1}^N 1_{\xi_{k,n}^j}}\ .$$

For later use in the further development of this section, we quote a couple of technical lemmas. The first one provides some $\mathbb{L}_p$ estimates of the normalizing quantities of the Markov transitions $\widehat{M}'_{k+1}$. The second one allows to quantify the deviations of $\widehat{M}'_{k+1}$ around its limiting values M'_{k+1}, as $N\to\infty$.

Lemma 3.5.2. *For any $p\ge 1$, and $0\le i\le N$ we have the following uniform estimate*

$$\sup_{N\ge 1}\sup_{0\le l\le k\le n}\left|\left|\eta_k^N(1_{\xi_{l,k}^i}\circ\pi_l)^{-1}\right|\right|_p < \infty\ . \tag{3.5.12}$$

Lemma 3.5.3. *For any $p \geq 1$, and $0 \leq i \leq N$ we have the following uniform estimate*

$$\sup_{0\leq l\leq n} \left\| \widehat{M}'_{l+1}(f)(\xi^i_{l,n}) - M'_{l+1}(f)(\xi^i_{l,n}) \right\|_p \leq c_p(n)/\sqrt{N} \ , \tag{3.5.13}$$

with some collection of finite constants $c_p(n) < \infty$ whose values only depend on the parameters p and n.

The proofs of these lemmas are rather technical, thus there are postponed to the appendix.

We are now in position to state and prove the main result of this section.

Theorem 3.5.1. *For any $p \geq 1$, and $0 \leq i \leq N$ we have the following uniform estimate*

$$\sup_{0\leq k\leq n} \left\| (u_k - \widehat{u}_k)(\xi^i_{k,n}) \right\|_p \leq c_p(n)/\sqrt{N} \ , \tag{3.5.14}$$

with some collection of finite constants $c_p(n) < \infty$ whose values only depend on the parameters p and n.

Proof. Firstly, we use the following decomposition

$$|u_k - \widehat{u}_k| 1_{\widehat{E}_{k,n}} \leq \sum_{k\leq l\leq n-1} \widehat{M}'_{k,l} |(\widehat{M}'_{l+1} - M'_{l+1})(u_{l+1})| \ 1_{\widehat{E}_{k,n}} \ . \tag{3.5.15}$$

By construction, we have

$$\widehat{M}'_{k,l} |(\widehat{M}'_{l+1} - M'_{l+1})(u_{l+1})| 1_{\widehat{E}_{(k,n)}} = \widehat{M}'_{k,l} |1_{\widehat{E}_{l,n}} (\widehat{M}'_{l+1} - M'_{l+1})(u_{l+1})| 1_{\widehat{E}_{k,n}} \ .$$

By (3.5.11), if we set

$$\widetilde{u}_{l+1} = |(\widehat{M}'_{l+1} - M'_{l+1})(u_{l+1})| \ ,$$

on the set $\widehat{E}_{l,n}$, then we have that

$$\widehat{M}'_{k,l}(\widetilde{u}_{l+1})(\xi^i_{k,n}) = \frac{\eta^N_n((1_{\xi^i_{k,n}} \circ \pi_k) \ (\widetilde{u}_{l+1} \circ \pi_l))}{\eta^N_n((1_{\xi^i_{k,n}} \circ \pi_k))} \ .$$

For any $p \geq 1$, we have

$$\begin{aligned} \left\| \widehat{M}'_{k,l}(\widetilde{u}_{l+1})(\xi^i_{k,n}) \right\|_p \leq & \left\| \eta^N_n((1_{\xi^i_{k,n}} \circ \pi_k))^{-1} \right\|_2^{1/p} \\ & \times \mathbb{E}\left(\eta^N_n((1_{\xi^i_{k,n}} \circ \pi_k) \ (\widetilde{u}_{l+1} \circ \pi_l)^{2p}) \right)^{1/(2p)} \ . \end{aligned}$$

This implies that

$$\left\| \widehat{M}'_{k,l}(\widetilde{u}_{l+1})(\xi^i_{k,n}) \right\|_p \leq \left\| \eta^N_n((1_{\xi^i_{k,n}} \circ \pi_k))^{-1} \right\|_2^{1/p} \times \sup_{1\leq j\leq N} \left\| \widetilde{u}_{l+1}(\xi^j_{l,n}) \right\|_{2p} \ .$$

The proof of (3.5.14) is now a clear consequence of Lemma 3.5.2 and Lemma 3.5.3. ∎

3.5.4 Bias analysis

To end this subsection, we will prove that just as the bias of the Broadie-Glasserman type estimators, the bias of the genealogical tree based estimator is always positive.

Note that, for any $0 \leq k \leq n$, function f on space E'_k and any $i \in \{1, \ldots, N\}$ we have

$$\mathbb{E}\left(f(\xi^i_{k+1,n})|\xi_{k,n}\right) = M_{k+1}f(\xi^i_{k,n}) \ . \tag{3.5.16}$$

This is because in the neutral genealogical tree model, the selection steps are independent of the mutations steps. Here, $\xi_{k,n}$ contains all the information on the construction of the tree plus the information on the values of the nodes on this tree at instant k. The equation (3.5.16) comes from the fact that given the information $\xi_{k,n}$ the particle $\xi^i_{k+1,n}$ follows the distribution $M'_{k+1}(\xi^i_{k,n}, \cdot)$.

Theorem 3.5.2. *For any $0 \leq k \leq n$ and any $i \in \{1, \ldots, N\}$, we have*

$$\mathbb{E}\left(\widehat{u}_k(\xi^i_{k,n})|\xi_{k,n}\right) \geq u_k(\xi^i_{k,n}) \ . \tag{3.5.17}$$

Proof. To prove this, we will use a simple induction argument.
For $l = n$, $\widehat{u}_n = u_n$, then we easily check that the following inequality is verified for all $i = 1, \ldots, N$,

$$\mathbb{E}\left(\widehat{u}_l(\xi^i_{l,n})|\xi_{l,n}\right) \geq u_l(\xi^i_{l,n}) \ . \tag{3.5.18}$$

Assume that (3.5.18) is verified for all $i = 1, \ldots, N$ and let us prove that the same inequality is valid for instant $l-1$.
With the elementary decomposition:

$$\begin{aligned}
\mathbb{E}\left(\widehat{M}'_l(\widehat{u}_l)(\xi^i_{l-1,n})|\xi_{l-1,n}\right) &= \mathbb{E}\left(\frac{\sum_{j=1}^N \widehat{u}_l(\xi^j_{l,n}) 1_{\xi^j_{l-1,n}=\xi^i_{l-1,n}}}{\sum_{j=1}^N 1_{\xi^j_{l-1,n}=\xi^i_{l-1,n}}}|\xi_{l-1,n}\right) \\
&= \frac{\sum_{j=1}^N \mathbb{E}\left(\widehat{u}_l(\xi^j_{l,n})|\xi_{l-1,n}\right) 1_{\xi^j_{l-1,n}=\xi^i_{l-1,n}}}{\sum_{j=1}^N 1_{\xi^j_{l-1,n}=\xi^i_{l-1,n}}} \ .
\end{aligned}$$

By assumption (3.5.18) and equation (3.5.16), we have

$$\begin{aligned}
\mathbb{E}\left(\widehat{u}_l(\xi^j_{l,n})|\xi_{l-1,n}\right) &\geq \mathbb{E}\left(u_l(\xi^j_{l,n})|\xi_{l-1,n}\right) \\
&= M_l u_l(\xi^j_{l-1,n}) \ .
\end{aligned}$$

Applying the preceding decomposition, it follows easily

$$\begin{aligned}
\mathbb{E}\left(\widehat{M}_l\widehat{u}_l(\xi^i_{l-1,n})|\xi_{l-1,n}\right) &\geq \frac{\sum_{j=1}^N M_l u_l(\xi^i_{l-1,n}) 1_{\xi^j_{l-1,n}=\xi^i_{l-1,n}}}{\sum_{j=1}^N 1_{\xi^j_{l-1,n}=\xi^i_{l-1,n}}} \\
&= M_l u_l(\xi^i_{l-1,n}) \ .
\end{aligned}$$

Then we can complete this proof by using Jensen's inequality, getting

$$\begin{aligned}\mathbb{E}\left(\widehat{u}_{l-1}(\xi^i_{l-1,n})|\xi_{l-1,n}\right) &\geq f_{l-1}(\xi^i_{l-1,n}) \vee \mathbb{E}\left(\widehat{M_l}\widehat{u}_l(\xi^i_{l-1,n})|\xi_{l-1,n}\right)\\ &\geq f_{l-1}(\xi^i_{l-1,n}) \vee M_l u_l(\xi^i_{l-1,n})\\ &= u_{l-1}(\xi^i_{l-1,n})\ .\end{aligned}$$

■

3.5.5 Numerical simulations

In this section, we give numerical examples to test the genealogical tree algorithm on two types of options from dimension 1 up to 6.

Prices dynamics and options model

Our numerical examples are taken from Bouchard and Warin [14], who provided precise approximations of option values in their examples. The asset prices are modeled by a d-dimensional Markov process $(\tilde{X}_t)$ such that each component (i.e. each asset) follows a geometric Brownian motion under the risk-neutral measure, that is, for assets $i = 1, \cdots, d$,

$$\frac{d\tilde{X}_t(i)}{\tilde{X}_t(i)} = rdt + \sigma_i dz^i_t\ , \tag{3.5.19}$$

where z^i, for $i = 1, \cdots, d$ are independent standard Brownian motions. The interest rate r is set to 5% annually. We also assume that for all $i = 1, \cdots, d$, $\tilde{X}_{t_0}(i) = 1$ and $\sigma_i = 20\%$ annually.
We consider two different Bermudan options with maturity $T = 1$ year and 11 equally distributed exercise opportunities at dates $t_k = kT/n$ with $k = 0, 1, \cdots, n = 10$, associated with two different payoffs:

1. a geometric average put option with strike $K = 1$ and payoff $(K - \prod_{i=1}^d \tilde{X}_T(i))_+$,
2. an arithmetic average put option with strike $K = 1$ and payoff $(K - \frac{1}{d}\sum_{i=1}^d \tilde{X}_T(i))_+$.

Note that the geometric average put payoff involves the process $\prod_{i=1}^d \tilde{X}(i)$ which can be identified to a one-dimensional non standard exponential Brownian motion. This trick was used in [14] to compute a precise benchmark option value by PDE techniques. We report in Table 4.1 the benchmark option values computed in [14], for both the geometric and arithmetic put options (by using respectively the one dimensional PDE method and the least squares regression method with $8 \times 10^6 \times d^2$ simulations and ten basis functions for each direction).

Number of assets	1	2	3	4	5	6
Geometric Payoff	0.06033	0.07815	0.08975	0.09837	0.10511	0.11073
Arithmetic Payoff	0.06033	0.03882	0.02947	0.02403	0.02046	0.01830

Table 3.1: Benchmark values for the geometric and arithmetic put options (taken from [14]).

State space discretization

The genealogical tree algorithm is designed for finite state spaces. Hence, before applying it to the aforementioned continuous space examples, we have to approximate the continuous state space Markov chain solution of (4.7.1) by a Markov chain with a finite state space. To this end, one can first discretize the state space using either a random tree, or a stochastic mesh, or a Binomial tree or a quantization approach ... In our numerical simulations, the quantization discretization seemed to be the most efficient.

State space partitioning Here, we propose to use a quantization-like approach for the space discretization step. We simulate a first set of M iid paths at each $n+1$ possible exercise dates $t_0, \cdots, t_n$, $(\tilde{X}^i_{t_k})^{i=1,\cdots,M}_{k=0,\cdots,n}$ according to dynamic (4.7.1). Assume now, that there exists two integers N' and P such that M can be written as the product $M = N'P$. Then, at each time step t_k, the particle set $\mathcal{S}_k = \{\tilde{X}^1_{t_k}, \cdots, \tilde{X}^M_{t_k}\}$ can be partitioned into N' localized subsets $\{\mathcal{S}^1_k, \cdots, \mathcal{S}^{N'}_k\}$ of P particles. Assume now that there exists d integers $(Q_1, \cdots, Q_d)$ such that N' can be written as the product $N' = Q_1 \cdots Q_d$. Assume for simplicity that $N' = Q^d$. One way to build this partition $\{\mathcal{S}^1_k, \cdots, \mathcal{S}^{N'}_k\}$ is then to apply the following procedure as in [14]:

1. sort the particles according to the first coordinate and split the sorted particles into Q subsets containing the same number of particles $Q^{d-1}P$;

2. if $d \geq 2$, for each subset, sort the particles according to the second coordinate and split the sorted particles into Q subsets containing the same number of particles $Q^{d-2}P$, which finally leads to Q^2 subsets containing the same number of particles $Q^{d-2}P$;

3. if $d \geq 3$, repeat this procedure recursively, in each direction $i = 3, \cdots, d$.

This operation is realized with a complexity $O(dM\log(M))$ and produces a partitions of $\mathcal{S}_k$ into $N' = Q^d$ subsets $\mathcal{S}^1_k, \cdots, \mathcal{S}^{N'}_k$ with the same number P of particles. Now, for each subset $\mathcal{S}^j_k$, for $j = 1, \cdots, N'$, we compute a representative state, S^j_k as the average particle over all the elements of $\mathcal{S}^j_k$. Then at each time step t_k for $k = 1, \cdots, n$, we will consider the finite state space $E_k = \{S^1_k, \cdots, S^{N'}_k\}$ and we set

$E_0 = \{X_{t_0}\}$. In the sequel, the discrete points $S_k^1, \cdots, S_k^{N'}$ will be referred to as the *sites*.

Finite state space Markov chain Assume now that a sequence of finite state spaces $E_k \subset \mathbb{R}^d$ is given for $k = 1, \cdots, n$ (for instance by the above procedure). We define a finite state space Markov chain $(X'_k)_{k=0,\cdots,n}$ such that $X'_0 = \tilde{X}_{t_0}$ and for all $k = 1, \cdots, n$,

- $X'_k \in E_k$;
- $\mathbb{P}\left(X'_k = S_k^j \,|\, X'_{k-1} = S_{k-1}^i\right) = \mathbb{P}\left(\tilde{X}_{t_k} \in V_k^j \,|\, \tilde{X}_{t_{k-1}} = S_{k-1}^i\right)$, where V_k^j denotes the Voronoi cell associated to the site S_k^j in the the discrete set E_k and $(\tilde{X}_{t_k})$ is the Markov process verifying (4.7.1) observed at the discrete times $t_0, \cdots, t_n$.

To simulate a transition of the Markov Chain $(X'_k)_{k=0,\cdots,n}$ from the state $S_{k-1}^i \in E_{k-1}$ at the time step $k-1$ to the time step k one can apply the following procedure:

1. Simulate a random variable $\tilde{X}_{t_k}$ according to $\tilde{M}_k(S_{k-1}^i, \cdot)$, where $\tilde{M}_k$ denotes the transition kernel of the continuous state space Markov chain verifying (4.7.1) from time t_{k-1} to t_k.
2. Set $X'_k = S_k^{i^*}$, where $S_k^{i^*}$ is the nearest neighbor of $\tilde{X}_{t_k}$ among the elements of E_k.

Complexity

In comparison with the quantization method proposed in [92], the genealogical algorithm based on the above space discretization only needs to simulate the finite state space Markov chain (X'_k) and avoids the time consuming computation of the transition probabilities.
In terms of complexity, the major part of the computing time is spent in the *forward step* described in Section 3.5.1 for simulating the discrete space Markov chain (X'_k). More precisely, for each transition, one has to compute a nearest neighbour among N' sites which finally leads to a complexity of order $O(NN')$ by time step, when considering the whole set of N particles.
In terms of approximation error, we can decompose the error induced by the whole procedure, on the Snell envelope approximation, into the sum of two terms:

1. The state space discretization error which can be upper bounded, according to [92] or Proposition 3.3.3 , by $\frac{c}{N'^{1/d}}$
2. The error induced by the genealogical tree algorithm, which could be upper bounded, according to the proof of Theorem 3.5.1, by $c\frac{N'^{\beta}}{N^{1/2}}$, for a given positive real $\beta > 0$.

Hence, to minimize the resulting upper bound on the global error, one has to choose judiciously the number of sites N' as a function of the number of particles such that $N' = 0(N^{\frac{d}{2\beta d+2}})$. With this choice, the complexity of the global procedure is of order $O(N^{\frac{(1+2\beta)d+2}{2\beta d+2}})$, with an approximation error bounded by $\frac{c}{N^{\frac{1}{2\beta d+2}}}$. In our numerical simulations, we have set $\beta = 1/2$ so that the complexity grows with the dimension from $N^{4/3}, N^{3/2}, N^{8/5}, \cdots, N^2$ for dimensions $d = 1, 2, 3, \cdots, \infty$.

On the other hand, in the *backward step*, (described in Section 3.5.3) consisting of computing the Snell envelope, our algorithm only requires a complexity which is linear in the number of particles, N. Hence, for a given underlying price process, our approach can rapidly approximate several Bermudan options with different payoff functions.

Numerical results

For each example, we have performed the algorithm for different numbers of particles for $N = 5 \times 10^3, 1 \times 10^4, 2.5 \times 10^4, 5 \times 10^4, 1 \times 10^5, 2 \times 10^5, 4 \times 10^5, 1 \times 10^6, 2 \times 10^6$. In each case, the sites were computed on the base of $M = \max(500000, 50 \times N') = \max(500000, 50 \times N^{\frac{d}{d+2}})$ simulations. Many runs of the algorithm were performed to build box plots for our estimates: 50 runs for $N < 10^6$ and 24 runs for $N = 1 \times 10^6$ and $N = 2 \times 10^6$.

Simulations results are reported in Figure 1.1 for the geometric put payoff and in Figure 1.2 for the arithmetic put payoff. First, notice that our algorithm has been implemented without any control variate technique. Moreover, our implementation has not been optimized. In particular, we have not investigated in this chapter any parallel implementations of our algorithm. Thus, it seems not relevant to report any running time measurements on the chapter. However, the algorithm complexity gives a good indication of the number of operations required by our algorithm. Moreover, the estimates reported on our graph correspond to the *backward estimate* provided by Algorithm A2 in Bouchard and Warin [14] and should be compared to that type of estimate. We could also obtain a *forward estimate* with our genealogical approach by applying the backward induction on the stopping times (just as in the Longstaff-Schwartz algorithm) with probably better performances than the *backward estimator*, but this is not the subject of the present chapter.
Hence, to compare the estimation errors of the *backward estimate* provided by our algorithm to a corresponding approach, we have reported, in Table 3.2, the estimation errors obtained with the genealogical algorithm using $N = 25000$ particles and $N' = N^{\frac{d}{d+2}}$ sites, in valuing the geometric put (on the first line) and the arithmetic put (on the second line) and, within parenthesis, the performances of the *backward estimate* provided by the quantization approach [4] implemented in [14], with 25600 quantization points for the same options. One can observe that both algorithms achieve similar performances for approximately the same number N of

quantization points (for the quantization algorithm) and particles (for the genealogical algorithm).
Now, notice that the complexity (per time step) of the genealogical algorithm is of order $NN' = N^{\frac{2d+2}{d+2}}$ for the construction of the genealogical tree and of order N for the backward induction on the prices, which is slightly smaller than the complexity of the quantization approach of order N^2 for the backward induction on prices (without taking into account the complexity related to the construction of the quantization tree and to the computation of the transition probabilities). Hence, we can conclude that our new algorithm is competitive with respect to comparable algorithms.

Number of assets	$d = 3$	$d = 4$	$d = 5$	$d = 6$
Geometric Put error (in % of the option value)	2 (2)	7 (8)	14 (15)	17 (22)
Arithmetic Put error (in % of the option value)	3.5 (3.5)	10 (8)	15 (16)	14 (17)

Table 3.2: Error (in % of the option value in Table 4.1) of the genealogical algorithm with $N = 25000$ particles and $N' = N^{\frac{d}{d+2}}$ sites, and within parenthesis of the quantization algorithm with $N = 25600$ quantization points, (taken from [14]) for the geometric and arithmetic put options.

Notice that one can observe on the graph, that for $d = 2$ or 3 the bias of our estimator can be negative. However, this is not in contradiction to Theorem 3.5.2. Indeed, recall that our estimator cumulates two kinds of approximations:

1. The first approximation is the discretization of the Markov chain which can induce a negative bias.

2. The second is the backward genealogical algorithm to compute the Snell envelope of the discrete Markov chain which (by theorem) induces a positive bias.

Looking into further applications, this algorithm is also well suited for Bermudan options with path dependent payoff. Indeed, by construction, the genealogical tree algorithm is defined in terms of the historical process, then it is able to compute conditional expectations with respect to the whole past of the process with no additional complexity.
In the same vein, we believe that this algorithm and the related convergence result could be extended, with slight modifications, to the more general case of reflected Backward Stochastic Differential Equations (BSDE) with non zero driver that does not depend on the z variable and which satisfies suitable regularity conditions.

Finally, in further research, it could also be interesting to extend this algorithm for the computation of price sensitivities for hedging purposes.

ACKNOWLEDGEMENTS: We are grateful to Laurent Plagne for his tremendous help in accelerating our numerical simulations implementation.

Chapter 4

Snell envelope with small probability criteria

4.1 Introduction

The Snell envelope is related to the calculation of the optimal stopping time of a random process based on a given optimality criteria. Several approximation schemes have been proposed recently to numerically compute the Snell envelope. In this chapter, we are interested in some specific optimality criteria related to the realization of a small probability or even rare events. In other words, given a random process $(X_k)_{0\leq k\leq n}$ and some *payoff* functions $(f_k)_{0\leq k\leq n}$, we want to maximize an expected gain $\mathbb{E}(f_\tau(X_\tau))$ by choosing τ on a set of random stopping times $\mathcal{T}$. When the payoff functions f_k are localized in a small region of the space, standard Monte Carlo simulations usually fail, because of the difficulty in ensuring enough simulation samples to realize the (relative-)rare events. For example, in finance, when $f(x) = (K - x)^+$, the so-called put option value is difficult to compute when K is much smaller than the initial asset price x_0. In even more complicated cases, we can consider the maximization of $\mathbb{E}(f_\tau(X_\tau)\prod_{k=0}^{\tau-1} B_k(X_k))$ for a given class of functions $(B_k)_{0\leq k\leq n}$ modeling an obstacle. For instance in the case of *barrier options*, $(B_k)_{0\leq k\leq n}$ take the form of indicator functions.

In this chapter, we propose a Monte Carlo algorithm to compute the Snell envelope, combining the Stochastic Mesh method introduced by M. Broadie and P. Glasserman [19] and a judicious interacting particle scheme which allows to concentrate the computational effort in the regions of interest w.r.t. the criteria. The principal idea of Broadie-Glasserman model is to operate a change of measure to replace conditional expectations by simple expectations. Besides, the change of measures can also be used with a variance reduction purpose to accelerate Monte

Carlo methods. However, in general, the choice of an efficient (in term of variance) change of measure, with an explicit Radon-Nikodym derivative, leading to an *easy-to-simulate* distribution is difficult. Precisely, the authors in [47] proposed an adaptive scheme based on an original interacting particle algorithm to approximate rare event expectations, allowing us to bypass the tricky steps of guessing a correct change of measure. In the present chapter, we extend this adaptive scheme for the recursive computation of the conditional expectations appearing in the context of optimal stopping problems. The main idea of the present chapter is then to mix the interacting particle algorithm in [47] with the Stochastic Mesh algorithm of Broadie and Glassserman [19].

This chapter is organized as follows. In Section 4.2, notations and generalities on the Snell envelope are presented. Moreover, some specific examples are outlined to motivate the scope of the chapter. In Section 4.3, we introduce a change of measure which allows to concentrate the computational effort in the regions of interest w.r.t. the criteria. In Section 4.4, we propose an interacting particle scheme to approximate the resulting (changed) measure. Section 4.5, is devoted to the theoretical analysis of this new Stochastic Mesh algorithm based on an interacting particle scheme. We provide non asymptotic convergence estimates and prove that the resulting estimator is positively biased. Finally, some numerical simulations are performed, in Section 4.7, showing the practical interest of the proposed algorithm.

4.2 Preliminary

For the convenience of the reader, we begin by introducing some notations and basic results that will be used all along the chapter.

4.2.1 Notations

We denote respectively by $\mathcal{P}(E)$, and $\mathcal{B}(E)$, the set of all probability measures on some measurable space $(E, \mathcal{E})$, and the Banach space of all bounded and measurable functions f equipped with the uniform norm $\|f\|$. We let $\mu(f) = \int \mu(dx)\ f(x)$, be the Lebesgue integral of a function $f \in \mathcal{B}(E)$, w.r.t. a measure $\mu \in \mathcal{P}(E)$.
We recall that a bounded integral kernel $M(x, dy)$ from a measurable space $(E, \mathcal{E})$ into an auxiliary measurable space $(E', \mathcal{E}')$ is an operator $f \mapsto M(f)$ from $\mathcal{B}(E')$ into $\mathcal{B}(E)$ such that the functions

$$x \mapsto M(f)(x) := \int_{E'} M(x, dy) f(y)$$

are $\mathcal{E}$-measurable and bounded, for any $f \in \mathcal{B}(E')$. In the above displayed formulae, dy stands for an infinitesimal neighborhood of a point y in E'. Sometimes, for indicator functions $f = 1_A$, with $A \in \mathcal{E}$, we also use the notation $M(x, A) :=$

$M(1_A)(x)$. The kernel M also generates a dual operator $\mu \mapsto \mu M$ from $\mathcal{M}(E)$ into $\mathcal{M}(E')$ defined by $(\mu M)(f) := \mu(M(f))$. A Markov kernel is a positive and bounded integral operator M with $M(1) = 1$. Given a pair of bounded integral operators (M_1, M_2), we let $(M_1 M_2)$ be the composition operator defined by $(M_1 M_2)(f) = M_1(M_2(f))$. Given a sequence of bounded integral operators M_n from some state space E_{n-1} into another E_n, we set $M_{k,l} := M_{k+1} M_{k+2} \cdots M_l$, for any $k \leq l$, with the convention $M_{k,k} = Id$, the identity operator. In the context of finite state spaces, these integral operations coincide with the traditional matrix operations on multidimensional state spaces.

We also assume that the reference Markov chain X_n with initial distribution $\eta_0 \in \mathcal{P}(E_0)$, and elementary transitions $M_n(x_{n-1}, dx_n)$ from E_{n-1} into E_n is defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}_{\eta_0})$, and we use the notation $\mathbb{E}_{\mathbb{P}_{\eta_0}}$ to denote the expectations w.r.t. $\mathbb{P}_{\eta_0}$. In this notation, for all $n \geq 1$ and for any $f_n \in \mathcal{B}(E_n)$, we have that

$$\mathbb{E}_{\mathbb{P}_{\eta_0}} \{f_n(X_n) | \mathcal{F}_{n-1}\} = M_n f_n(X_{n-1}) := \int_{E_n} M_n(X_{n-1}, dx_n)\ f_n(x_n)$$

with the σ-field $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ generated by the sequence of random variables X_p, from the origin $p = 0$ up to the time $p = n$. We also use the conventions $\prod_{\emptyset} = 1$, and $\sum_{\emptyset} = 0$.

4.2.2 Robustness Lemma

In the discrete time setting, the Snell envelope are defined in terms of a given Markov process $(X_k)_{k \geq 0}$ taking values in some sequence of measurable state spaces $(E_n, \mathcal{E}_k)_{k \geq 0}$ adapted to the natural filtration $\mathcal{F} = (\mathcal{F}_k)_{k \geq 0}$. We let $\eta_0 = \text{Law}(X_0)$ be the initial distribution on E_0, and we denote by $M_k(x_{k-1}, dx_k)$ the elementary Markov transition of the chain from E_{k-1} into E_k. For a given time horizon n and any $k \in \{0, \ldots, n\}$, we let $\mathcal{T}_k$ be the set of all stopping times τ taking values in $\{k, \ldots, n\}$. For a given sequence of non negative measurable functions f_k on E_k, we define a target process $Z_k = f_k(X_k)$. Then $(U_k)_{0 \leq k \leq n}$ the Snell envelope of process $(Z_k)_{0 \leq k \leq n}$ is defined by a recursive formula:

$$U_k = Z_k \vee \mathbb{E}(U_{k+1} | \mathcal{F}_k)$$

with terminal condition $U_n = Z_n$. The main property of the Snell envelope defined as above is

$$U_k = \sup_{\tau \in \mathcal{T}_k} \mathbb{E}(Z_\tau | \mathcal{F}_k) = \mathbb{E}(Z_{\tau_k^*} | \mathcal{F}_k) \ \text{ with } \ \tau_k^* = \min \{k \leq j \leq n \ :\ U_j = Z_j\} \in \mathcal{T}_k\ .$$

Then the computation of the Snell envelope $(U_k)_{0 \leq k \leq n}$ amounts to solving the following backward functional equation.

$$u_k = f_k \vee M_{k+1}(u_{k+1}) \tag{4.2.1}$$

for any $0 \leq k < n$ with the terminal condition $u_n = f_n$.

But at this level of generality, we can hardly have a closed solution of the function u_k. In this context, lots of numerical approximation schemes have been proposed. Most of them amount to replacing in recursion (4.2.1) the pair of functions and Markov transitions $(f_k, M_k)_{0\leq k\leq n}$ by some approximation model $(\widehat{f}_k, \widehat{M}_k)_{0\leq k\leq n}$ on some possibly reduced measurable subsets $\widehat{E}_k \subset E_k$. In paper [50], the authors provided the following robustness lemma to estimate the error related to the resulting approximation $\widehat{u}_k$ of the Snell envelope u_k, for several types of approximation models $(\widehat{f}_k, \widehat{M}_k)_{0\leq k\leq n}$.

Lemma 4.2.1. *For any $0 \leq k < n$, on the state space $\widehat{E}_k$, we have that*

$$|u_k - \widehat{u}_k| \leq \sum_{l=k}^{n} \widehat{M}_{k,l}|f_l - \widehat{f}_l| + \sum_{l=k}^{n-1} \widehat{M}_{k,l}|(M_{l+1} - \widehat{M}_{l+1})u_{l+1}| \ .$$

This lemma provides a natural way to compare and combine different approximation models. In the present chapter, this Lemma will be applied in the specific framework for the small probability criteria.

4.2.3 Motivations

The choice of nonhomogeneous state spaces $\mathbf{E}_n$ is not innocent. In several application areas the underlying Markov model is a path-space Markov chain:

$$\mathbf{X}_n = (X_0, \ldots, X_n) \in \mathbf{E}_n = (E_0 \times \ldots \times E_n) \ . \tag{4.2.2}$$

The elementary prime variables X_n represent an elementary Markov chain with Markov transitions $M_k(x_{k-1}, dx_k)$ from E_{k-1} into E_k. In this situation, the historical process $\mathbf{X}_n$ can be seen as a Markov chain with transitions given for any $\mathbf{x}_{k-1} = (x_0, \ldots, x_{k-1}) \in \mathbf{E}_{k-1}$ and $\mathbf{y}_k = (y_0, \ldots, y_k) \in \mathbf{E}_k$ by the following formula

$$\mathbf{M}_k(\mathbf{x}_{k-1}, d\mathbf{y}_k) = \delta_{\mathbf{x}_{k-1}}(d\mathbf{y}_{k-1}) \ M_k(y_{k-1}, dy_k) \ .$$

As we will see in this sequel, this path space framework is, for instance, well suited when dealing with path dependent options as Asian options or Barrier options. Besides, this path space framework is also well suited for the analysis of the Snell envelope under different probability measures.

The multiplicatively path dependent case Now come back to the multiplicatively path dependent Snell envelope that we mentioned in the introduction and formalize the the path space model. For a given collection of real valued functions $(f_k)_{0\leq k\leq n}$ and $(B_k)_{0\leq k\leq n}$, defined on $(E_k)_{0\leq k\leq n}$, we define a class of real valued functions $(F_k)_{0\leq k\leq n}$ defined on the path spaces $(\mathbf{E}_k)_{0\leq k\leq n}$ by

$$F_k(\mathbf{x}_k) := f_k(x_k) \prod_{0\leq p\leq k-1} B_p(x_p) \ , \quad \text{for all } 0 \leq k \leq n \ ,$$

for all $\mathbf{x}_k = (x_0, \cdots, x_k) \in \mathbf{E}_k$. Instead of $\mathbb{E}(f_\tau(X_\tau))$ we want to maximize the expected gain $\mathbb{E}(F_\tau(\mathbf{X}_\tau))$ w.r.t. τ in a set of random stopping times $\mathcal{T}$. In other words, one is interested in computing the Snell envelope $(\mathbf{u}_k)_{0\leq k\leq n}$ associated to the gain functions $(F_k)_{0<k\leq n}$; it satisfies the recursion:

$$\left\{ \begin{array}{l} \mathbf{u}_n(\mathbf{x}_n) = F_n(\mathbf{x}_n) \\ \mathbf{u}_k(\mathbf{x}_k) = F_k(\mathbf{x}_k) \vee \mathbf{M}_{k+1}(\mathbf{u}_{k+1})(\mathbf{x}_k), \forall\ 0 \leq k \leq n-1\ . \end{array} \right. \tag{4.2.3}$$

At this stage, two difficulties may arise. First, the above recursion seems to require the approximation of high dimensional conditional expectations, defined on the path spaces E_k, at each time step from $k = n-1$ up to $k = 0$. Second, when the optimality criteria B_p is localized in a specific region of E_p, for each p, then the product $\prod_{p=0}^{k-1} B_p(x_p)$ can be interpreted as a rare event. Hence, at first glance, the computation of Snell envelopes in the multiplicatively path dependent case seems to combine two additional numerical difficulties w.r.t. the standard case, related to the computation of conditional expectations in both high dimensional and rare event situations. The dimensionality problem is easily bypassed by considering an intermediate standard Snell envelope without path dependent criteria, which is directly related to the multiplicatively path dependent Snell envelope. Indeed, consider the standard (non path dependent) Snell envelope $(v_k)_{0\leq k\leq n}$ satisfying the following recursion:

$$\left\{ \begin{array}{rcl} v_n(x_n) & = & f_n(x_n) \\ v_k(x_k) & = & f_k(x_k) \vee \left[B_k(x_k) M_{k+1}(v_{k+1})(x_k)\right]\ ,\ \forall\ 0 \leq k \leq n-1\ . \end{array} \right. \tag{4.2.4}$$

For all $0 \leq k \leq n$, let us denote by $\mathbf{v}_k$ the real valued functions defined on $\mathbf{E}_k$, such that $\mathbf{v}_k(\mathbf{x}_k) := v_k(x_k) \prod_{p=0}^{k-1} G_p(x_p)$. By construction, one can easily check that for all $0 \leq k \leq n$, $\mathbf{u}_k \equiv \mathbf{v}_k$ and in particular $\mathbf{u}_0(\mathbf{x}_0) = v_0(x_0)$. Indeed, one can verify that $(\mathbf{v}_k)_{0\leq k\leq n}$ follow the same recursion (4.2.3) as $(\mathbf{u}_k)_{0\leq k\leq n}$ and have the same terminal condition. Now that we have underlined the link between $\mathbf{u}_k$ and v_k, the computation of the original Snell envelope $\mathbf{u}_k$ can be done by using one of the many approximation schemes developed for the standard (non path dependent) case.

Besides, to deal with the rare event problem, we propose a change of measure which allows to concentrate the computational effort in the regions of interest w.r.t. the criteria $(B_k)_{0\leq k\leq n-1}$.

Rare event associated with Payoff function Another Snell envelope problem associated with a small probability event comes from the payoff function when $f(X_n)$ is difficult to simulate. An example arises from the Bermudan put options when the strike K is much smaller than the initial price of the underlying asset. In this case, the standard Monte Carlo approach is not able to concentrate the computational effort in regions where the payoff function $x \mapsto f(x) = (K-x)^+$ does not vanish to

zero. In full generality, for a payoff function f concentrated in a relative small region of the space, the choice of an efficient change of measure for computing the recursive conditional expectations is difficult. This problem becomes even more tricky when the number of the underlying assets is greater than three. In the following section, we propose a simple adaptive scheme that allows to approximate an efficient change of measure without requiring any a priori information.

4.3 Snell envelope and change of measure

Now, recall the reduced Snell envelope for the multiplicatively path dependent case:

$$\left\{ \begin{array}{rcl} v_n(x_n) & = & f_n(x_n) \\ v_k(x_k) & = & f_k(x_k) \vee \left[B_k(x_k) M_{k+1}(v_{k+1})(x_k) \right] \ , \ \forall \ 0 \leq k \leq n-1 \ . \end{array} \right.$$

The above recursion implies that it is not relevant to compute precisely the conditional expectation $M_{k+1}(v_{k+1})(x_k)$ when the value of the criteria $B_k(x_k)$ is zero or very small, or when the gain function f_k is zero or very small. Hence from a variance reduction point of view, when approximating the conditional expectation $M_{k+1}(v_{k+1})(x_k)$ by a Monte Carlo method, it seems relevant to concentrate the simulations in the regions of E_{k+1} where B_{k+1} and/or f_{k+1} reach high values. Hence, to avoid the potential rare events B, we consider a change of measure on the measurable product space $(E_0 \times \cdots \times E_n, \mathcal{E}_0 \times \cdots \times \mathcal{E}_n)$, with the following form

$$d\mathbb{Q}_n = \frac{1}{Z_n} \left[\prod_{k=0}^{n-1} G_k(X_k) \right] d\mathbb{P}_n \ , \quad \text{with} \quad Z_n = \mathbb{E} \left(\prod_{k=0}^{n-1} G_k(X_k) \right) = \prod_{k=0}^{n-1} \eta_k(G_k) \ , \tag{4.3.1}$$

where $(G_k)_{0 \leq k < n}$ is a sequence of non-negative functions defined on $(E_k)_{0 \leq k < n}$ (typically $G_k := B_k$, and G_k is written instead of B_k in further development of this chapter) and η_k is the probability measure defined on E_k such that, for any measurable function f on E_k

$$\eta_k(f) := \frac{\mathbb{E}\Big(f(X_k) \prod_{p=0}^{k-1} G_p(X_p) \Big)}{\mathbb{E}\Big(\prod_{p=0}^{k-1} G_p(X_p) \Big)} \ .$$

The measures $(\eta_k)_{0 \leq k \leq n}$ defined above can be seen as the laws of random states $(\bar{X}_k)_{0 \leq k \leq n}$ under the probability measures $(\mathbb{Q}_k)_{0 \leq k \leq n}$. More interestingly, in Section 4.4 we will see that the sequence of random states $(\bar{X}_k)_{0 \leq k \leq n}$ forms a nonlinear Markov chain with transitions $\bar{X}_k \rightsquigarrow \bar{X}_{k+1}$ that depends on the current distribution η_k, at time k. The behavior of this chain is dictated by the potential functions $(G_k)_{0 \leq k \leq n}$ and the Markov transitions $(M_k)_{\leq k \leq n}$ of the reference process $(X_k)_{0 \leq k \leq n}$. Regions with high G_k-values are visited more likely.

To illustrate this remark, we examine the situation where $G_k(x_k) = B_k(x_k) := 1_{A_k}(x_k)$ with $A_k \subset E_k$. In this situation, $law(X_k|X_p \in A_p,\ p < k) = law(\bar{X}_k) = \eta_k$ is the conditional distribution of X_k given the fact that $X_p \in A_p$, for any $p < k$. In this special case, the process $(\bar{X}_k)_{0\leq k\leq n}$ is restricted to regions related to the choice of the sequence $(A_k)_{0\leq k\leq n}$. This change of measure is know as the optimal twisted measure for sampling a Markov chain restricted to the subset regions A_k. More general change of measure are addressed in section 4.6. These models are direct extension of 4.3.1 to potential functions that depend on the transition of the reference Markov chain.

When the rare event problem comes from the payoff, we can construct a collection of G_k to force the particle step by step to achieve the payoff. But in this case, there is no more explicit obstacle B_k to help us to construct such potential functions. A choice of G_k is provided in section 4.7.2. For further reading, readers are referred to [47]. The authors have proposed several choices to minimize the variance.

At this stage, it is important to emphasize that the analysis of the both case where the choice of G_k is explicit or not, are mathematically equivalent. The only difference comes from the fact that the recursion 4.2.4 has additional term B_k compared to 4.2.1. And the mathematical analysis of the later is easier and can be induced directly from the former (by deleting all the B_k appeared in the Snell envelope recursion in the analysis). So only the analysis of the multiplicatively path dependent case are provided in this chapter.

Furthermore, it is also important to observe that, for any measurable function f on E_k

$$\eta_k(f) = \frac{\eta_{k-1}(G_{k-1}M_k(f))}{\eta_{k-1}(G_{k-1})} \ . \tag{4.3.2}$$

We denote the recursive relation between η_k and η_{k-1} by introducing the operators Φ_k such that, for all $1 \leq k \leq n$

$$\eta_k = \Phi_k(\eta_{k-1}) \ . \tag{4.3.3}$$

Let us now introduce the integral operator Q_k such that, for all $1 \leq k \leq n$

$$Q_k(f)(x_{k-1}) := \int G_{k-1}(x_{k-1})M_k(x_{k-1}, dx_k)f(x_k) \ . \tag{4.3.4}$$

In further developments of this chapter, we suppose that $M_k(x_{k-1}, \cdot)$ are equivalent to some measures λ_k, for any $0 \leq k \leq n$ and $x_{k-1} \in E_{k-1}$, i.e. there exists a collection of positive functions H_k and measures λ_k such that:

$$M_k(x_{k-1}, dx_k) = H_k(x_{k-1}, x_k)\lambda_k(dx_k) \ . \tag{4.3.5}$$

Now, we are in a position to state the following Lemma.

Lemma 4.3.1. *For any measure η on E_k, recursion (4.2.4) defining v_k can be rewritten:*

$$v_k(x_k) = f_k(x_k) \vee Q_{k+1}(v_{k+1})(x_k) = f_k(x_k) \vee \Phi_{k+1}(\eta)\left(\frac{dQ_{k+1}(x_k,\cdot)}{d\Phi_{k+1}(\eta)} v_{k+1}\right) ,$$

for any $x_k \in E_k$, where

$$\frac{dQ_{k+1}(x_k,\cdot)}{d\Phi_{k+1}(\eta)}(x_{k+1}) = \frac{G_k(x_k)H_{k+1}(x_k,x_{k+1})\eta(G_k)}{\eta(G_k H_{k+1}(\cdot,x_{k+1}))} ,$$

for any $(x_k, x_{k+1}) \in E_k \times E_{k+1}$.

Proof. Under Assumption (4.3.5), we have immediately the following formula

$$M_{k+1}(x_k, dx_{k+1}) = H_{k+1}(x_k,x_{k+1})\frac{\eta_k(G_k)}{\eta_k(G_k H_{k+1}(\cdot,x_{k+1}))}\eta_{k+1}(dx_{k+1}) . \tag{4.3.6}$$

Now, note that the above equation is still valid for any measure η,

$$M_{k+1}(x_k, dx_{k+1}) = H_{k+1}(x_k,x_{k+1})\frac{\eta(G_k)}{\eta(G_k H_{k+1}(\cdot,x_{k+1}))}\Phi_{k+1}(\eta)(dx_{k+1}) . \tag{4.3.7}$$

Hence, the Radon Nikodym derivative of $M_{k+1}(x_k, dx_{k+1})$ w.r.t. $\Phi_{k+1}(\eta)$ is such that

$$\frac{dM_{k+1}(x_k,\cdot)}{d\Phi_{k+1}(\eta)}(x_{k+1}) = H_{k+1}(x_k,x_{k+1})\frac{\eta(G_k)}{\eta(G_k H_{k+1}(\cdot,x_{k+1}))} . \tag{4.3.8}$$

We end the proof by applying the arguments above to recursion (4.2.4).

4.4 A particle approximation scheme

In this section, we first propose a particle model to sample the random variables according to these distributions. This sample scheme is then combined with the Stochastic Mesh scheme to finally provide an original particle algorithm to approximate the Snell envelope $(v_k)_{0\le k\le n}$.
By definition (4.3.3) of Φ_{k+1}, we have the following formula

$$\Phi_k(\eta_{k-1}) = \eta_{k-1}K_{k,\eta_{k-1}} = \eta_{k-1}S_{k-1,\eta_{k-1}}M_k = \Psi_{G_{k-1}}(\eta_{k-1})M_k . \tag{4.4.1}$$

Where $K_{k,\eta_{k-1}}$, $S_{k-1,\eta_{k-1}}$ and $\Psi_{G_{k-1}}$ are defined as follows:

$$\begin{cases} K_{k,\eta_{k-1}}(x_{k-1}, dx_k) &= (S_{k-1,\eta_{k-1}}M_k)(x_{k-1}, dx_k) \\ &= \int S_{k-1,\eta_{k-1}}(x_{k-1}, dx'_{k-1})M_k(x'_{k-1}, dx_k) , \\ S_{k-1,\eta_{k-1}}(x, dx') &= \epsilon G_{k-1}(x)\delta_x(dx') + (1-\epsilon G_{k-1}(x))\Psi_{G_{k-1}}(\eta_{k-1})(dx') \\ \Psi_{G_{k-1}}(\eta_{k-1})(dx) &= \frac{G_{k-1}(x)}{\eta_{k-1}(G_{k-1})}\eta_{k-1}(dx) , \end{cases}$$

where the real ϵ is such that ϵG takes its values $[0,1]$.
More generally, the operations Ψ and S can be expressed as $\Psi_G(\eta)(f) = \frac{\eta(Gf)}{\eta(G)} = \eta S_\eta(f)$ with $S_\eta(f) = \epsilon Gf + (1-\epsilon G)\Psi_G(\eta)(f)$. We recall from [40] that $\eta_k = law(\bar{X}_k)$, where $\bar{X}_{k-1} \rightsquigarrow \bar{X}_k$ is a Markov chain with transitions $K_{k,\eta_{k-1}}$ defined above.

The particle approximation provided in the present chapter is defined in terms of a Markov chain $\xi_k^{(N)} = (\xi_k^{(i,N)})_{1\leq i\leq N}$ on the product state spaces E_k^N, where the given integer N is the number of particles sampled in every instant. The initial particle system, $\xi_0^{(N)} = \left(\xi_0^{(i,N)}\right)_{1\leq i\leq N}$, is a collection of N i.i.d. random copies of X_0. We let $\mathcal{F}_k^N$ be the sigma-field generated by the particle approximation model from the origin, up to time k. To simplify the presentation, when there is no confusion we suppress the population size parameter N, and we write ξ_k and ξ_k^i instead of $\xi_k^{(N)}$ and $\xi_k^{(i,N)}$. By construction, ξ_k is a particle model with a selection transition and a mutation type exploration i.e. the evolution from ξ_k to ξ_{k+1} is composed by two steps:

$$\xi_k \in E_k^N \xrightarrow[S_{k,\eta_k^N}]{Selection} \widehat{\xi}_k := \left(\widehat{\xi}_k^i\right)_{1\leq i\leq N} \in E_k^N \xrightarrow[M_{k+1}]{Mutation} \xi_{k+1} \in E_{k+1}^N \ . \tag{4.4.2}$$

Then we define η_k^N and $\widehat{\eta}_k^N$ as the occupation measures after the mutation and the selection steps. More precisely,

$$\eta_k^N := \frac{1}{N}\sum_{1\leq i\leq N} \delta_{\xi_k^i} \quad \text{and} \quad \widehat{\eta}_k^N := \frac{1}{N}\sum_{1\leq i\leq N} \delta_{\widehat{\xi}_k^i} \ .$$

During the selection transition S_{k,η_k^N}, for $0 \leq i \leq N$ with a probability $\epsilon G_k(\xi_k^i)$ we decide to skip the selection step i.e. we leave $\widehat{\xi}_k^i$ stay on particle ξ_k^i, and with probability $1-\epsilon G_k(\xi_k^i)$ we decide to do the following selection: $\widehat{\xi}_k^i$ randomly takes the value in ξ_k^j for $0 \leq j \leq N$ with distribution $\frac{G_k(\xi_k^j)}{\sum_{l=1}^N G_k(\xi_k^l)}$. Note that when $\epsilon G_k \equiv 1$, the selection is skipped (i.e. $\widehat{\xi}_k = \xi_k$) so that the model corresponds exactly to the Broadie-Glasserman type model analysed by P. Del Moral and P. Hu et al. [50]. Hence, the factor ϵ can be interpreted as a level of selection against the rare events. During the mutation transition $\widehat{\xi}_k \rightsquigarrow \xi_{k+1}$, every selected individual $\widehat{\xi}_k^i$ evolves randomly to a new individual $\xi_{k+1}^i = x$ randomly chosen with the distribution $M_{k+1}(\widehat{\xi}_k^i, dx)$, for $1 \leq i \leq N$.
It is important to observe that by construction, η_{k+1}^N is the empirical measure associated with N conditionally independent and identically distributed random individual ξ_{k+1}^i with common distribution $\Phi_{k+1}(\eta_k^N)$.

Now, we are in a position to describe precisely the new approximation scheme proposed to estimate the Snell envelope $(v_k)_{0\leq k\leq n}$. The main idea consists in taking $\eta = \eta_k^N$, in Lemma 4.3.1, then observing that Snell envelope $(v_k)_{0\leq k\leq n}$ is solution of

the following recursion, for all $0 \leq k < n$,

$$v_k(x_k) = f_k(x_k) \vee \Phi_{k+1}(\eta_k^N) \left(\frac{dQ_{k+1}(x_k, \cdot)}{d\Phi_{k+1}(\eta_k^N)} v_{k+1} \right) \ .$$

Now, if $\Phi_{k+1}(\eta_k^N)$ is well estimated by η_{k+1}^N, it is relevant to approximate v_k by $\widehat{v}_k$ defined by the following backward recursion

$$\left\{ \begin{array}{rcl} \widehat{v}_n & = & f_n \\ \widehat{v}_k(x_k) & = & f_k(x_k) \vee \eta_{k+1}^N \left(\dfrac{dQ_{k+1}(x_k, \cdot)}{d\Phi_{k+1}(\eta_k^N)} \widehat{v}_{k+1} \right) \quad \text{for all } 0 \leq k < n \ , \end{array} \right. \tag{4.4.3}$$

Note that in the above formula (4.4.3), the function v_k is defined not only on E_k^N but on the whole state space E_k.
To simplify notations, we set

$$\widehat{Q}_{k+1}(x_k, dx_{k+1}) = \eta_{k+1}^N(dx_{k+1}) \frac{dQ_{k+1}(x_k, \cdot)}{d\Phi_{k+1}(\eta_k^N)}(x_{k+1}) \ .$$

Finally, with this notation, the real Snell envelope $(v_k)_{0 \leq k \leq n}$ and the approximation $(\hat{v}_k)_{0 \leq k \leq n}$ are such that, for all $0 \leq k < n$,

$$\begin{array}{rcl} v_k & = & f_k \vee Q_{k+1}(v_{k+1}) \\ \widehat{v}_k & = & f_k \vee \widehat{Q}_{k+1}(\widehat{v}_{k+1}) \ . \end{array}$$

In the change of measure interpretation presented in section 4.3, the particle algorithm developed above can be seen as a stochastic acceptance-rejection technique with recycling transitions. This type of particle sampling model has been used in other contexts, including financial risk analysis in [24, 23]. For an overview of these novel particle algorithms in financial mathematics, we refer the interested reader to the book [22].

4.5 Convergence and bias analysis

By the previous construction, we can approximate $\Phi_{k+1}(\eta_k^N)$ by η_{k+1}^N. In this section, we will first analyze the error associated with that approximation and then derive an error bound for the resulting Snell envelope approximation scheme. To simplify notations, in further development, we consider the random fields V_k^N defined as

$$V_k^N := \sqrt{N} \ \left(\eta_k^N - \Phi_k(\eta_{k-1}^N) \right) \ .$$

The following lemma shows the conditional zero-bias property and mean error estimates for the approximation η_{k+1}^N of $\Phi_{k+1}(\eta_k^N)$.

Lemma 4.5.1. *For any integer $p \geq 1$, we denote by p' the smallest even integer greater than p. In this notation, for any $0 \leq k \leq n$ and any integrable function f on E_{k+1}, we have*

$$\mathbb{E}\left(\eta_{k+1}^N(f)|\mathcal{F}_k^N\right) = \Phi_{k+1}(\eta_k^N)(f)$$

and

$$\mathbb{E}\left(\left|V_k^N(f)\right|^p|\mathcal{F}_k^N\right)^{\frac{1}{p}} \leq 2\ a(p)\ \left[\Phi_{k+1}(\eta_k^N)(|f|^{p'})\right]^{\frac{1}{p'}}$$

with the collection of constants

$$a(2p)^{2p} = (2p)_p\ 2^{-p} \quad \textit{and} \quad a(2p+1)^{2p+1} = \frac{(2p+1)_{p+1}}{\sqrt{p+1/2}}\ 2^{-(p+1/2)}\ .$$

Proof : The conditional zero-bias property is easily proved as follows

$$\begin{aligned}\mathbb{E}\left(\eta_{k+1}^N(f)|\eta_k^N\right) &= \frac{1}{N}\sum_{i=1}^N \mathbb{E}(f(\xi_{k+1}^i)|\eta_k^N)\\ &= \frac{1}{N}\sum_{i=1}^N K_{k+1,\eta_k^N}(f)(\xi_k^i)\\ &= (\eta_k^N K_{k+1,\eta_k^N})(f) = \Phi_{k+1}(\eta_k^N)(f)\ .\end{aligned}$$

Then the above equality implies

$$\mathbb{E}\left(\left|\left[\eta_{k+1}^N - \Phi_{k+1}(\eta_k^N)\right](f)\right|^p|\mathcal{F}_k^N\right)^{\frac{1}{p}} \leq \mathbb{E}\left(\left|\left[\eta_{k+1}^N - \mu_{k+1}^N\right](f)\right|^p|\mathcal{F}_k^N\right)^{\frac{1}{p}}\ ,$$

where $\mu_{k+1}^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y_{k+1}^i}$ stands for an independent copy of η_{k+1}^N given η_k^N. Using Khintchine's type inequalities yields that

$$\begin{aligned}\sqrt{N}\ \mathbb{E}\left(\left|[\eta_{k+1}^N - \mu_{k+1}^N](f)\right|^p|\mathcal{F}_k^N\right)^{\frac{1}{p}} &\leq 2\ a(p)\ \mathbb{E}\left(\left|f\left(\xi_{k+1}^1\right)\right|^{p'}\mid\mathcal{F}_k^N\right)^{\frac{1}{p'}}\\ &= 2\ a(p)\ \left[\Phi_{k+1}(\eta_k^N)(|f|^{p'})\right]^{\frac{1}{p'}}\ .\end{aligned}$$

We end the proof by combining the above two inequalities.

■

A consequence of the zero-bias property proved in Lemma 4.5.1 is that

$$\mathbb{E}(\widehat{Q}_{k+1}(f)(x_k)|\eta_k^N) = Q_{k+1}(f)(x_k)\ .$$

To estimate the error between v_k and the approximation $\hat{v}_k$, it is useful to introduce the following random integral operator R_k^N such that for any measurable function on E_{k+1},

$$R_{k+1}^N(f)(x_k) = \sqrt{N}\left(\widehat{Q}_{k+1}(f)(x_k) - Q_{k+1}(f)(x_k)\right)\ .$$

Note that

$$R^N_{k+1}(f)(x_k) := \int V^N_{k+1}(dx_{k+1}) \ \frac{dQ_{k+1}(x_k,.)}{d\Phi_{k+1}(\eta^N_k)}(x_{k+1}) \ f(x_{k+1}) \ ,$$

then, applying again Lemma 4.5.1 implies the following Khintchine's type inequality

$$\begin{aligned}
&\mathbb{E}(\left|R^N_{k+1}(v_{k+1})(x_k)\right|^p |\eta^N_k)^{\frac{1}{p}} \\
&\quad \leq 2 \ a(p) \left[\int_{E_{k+1}} \Phi_{k+1}(\eta^N_k)(dx_{k+1}) \left(\frac{dQ_{k+1}(x_k,\cdot)}{d\Phi_{k+1}(\eta^N_k)}(x_{k+1}) v_{k+1}(x_{k+1})\right)^{p'}\right]^{\frac{1}{p'}}
\end{aligned}$$

Let $\widehat{Q}_{k,l} = \widehat{Q}_{k+1}\widehat{Q}_{k+2}\ldots\widehat{Q}_l$ for any $0 \leq k < l \leq n$, then it follows easily, by recursion, that

$$\mathbb{E}(\widehat{Q}_{k,l}(f)(x_k)|\eta^N_k) = Q_{k,l}(f)(x_k) \ .$$

Now, by Lemma 4.2.1, we conclude

$$\sqrt{N} \, |(v_k - \widehat{v}_k)| \leq \sum_{k<l<n} \widehat{Q}_{k,l}|(R^N_{l+1})(v_{l+1})| \ . \tag{4.5.1}$$

We are now in position to state the main result of this chapter.

Theorem 4.5.1. *For any $0 \leq k \leq n$ and any integer $p \geq 1$, we have*

$$\sup_{x\in E_k} \|(\widehat{v}_k - v_k)(x)\|_{L_p} \leq \sum_{k<l<n} \frac{2 \ a(p)}{\sqrt{N}} q_{k,l} \ \left[Q_{k,l+1}(h^{p'-1}_{l+1} v^{p'}_{l+1})(x)\right]^{\frac{1}{p'}} \ ,$$

with a collection of constants $q_{k,l}$ and functions h_k defined as

$$q_{k,l} := \left[\|h_{k+1}\| \prod_{m=k}^{l} \|G_m\|\right]^{\frac{p'-1}{p'}} \quad \text{and} \quad h_k(x_k) := \sup_{x,y\in E_{k-1}} \frac{H_k(x,x_k)}{H_k(y,x_k)} \ . \tag{4.5.2}$$

Proof : First, decomposition (4.5.1) yields

$$\sqrt{N} \, \|(\widehat{v}_k - v_k)(x)\|_{L_p} \leq \sum_{k<l<n} \left\|\widehat{Q}_{k,l}|(R^N_{l+1})(v_{l+1})|(x)\right\|_{L_p} \ , \quad \text{for all } x \in E_k \ .$$

Note that

$$\|\widehat{Q}_{k,l}(1)\| \leq b_{k,l} \ , \quad \text{where} \quad b_{k,l} := \|h_{k+1}\| \prod_{m=k}^{l-1} \|G_m\| \ .$$

Then it follows easily that for any integrable function f on E_l

$$(\widehat{Q}_{k,l}(f))^p \leq (b_{k,l})^{p-1} \widehat{Q}_{k,l}(f^p) \ .$$

This yields that

$$\left\|\widehat{Q}_{k,l}\left|(R^N_{l+1}))(v_{l+1})\right|(x)\right\|_{L_p} \leq (b_{k,l})^{\frac{p-1}{p}} \mathbb{E}\left(\widehat{Q}_{k,l}\left(\left|(R^N_{l+1}))(v_{l+1})\right|\right)^p(x)\right)^{\frac{1}{p}} \ .$$

Applying Lemma 4.5.1 to the right-hand side of the above inequality, we obtain for any $x_l \in E_l$

$$\mathbb{E}\left(\left|(R^N_{l+1}))(v_{l+1})(x_l)\right|^p |\eta^N_l\right)^{\frac{1}{p}}$$
$$\leq 2\ a(p)\left[\int_{E_{l+1}} \Phi_{l+1}(\eta^N_l)(dx_{l+1})\left(\frac{dQ_{l+1}(x_l,\cdot)}{d\Phi_{l+1}(\eta^N_l)}(x_{l+1})v_{l+1}(x_{l+1})\right)^{p'}\right]^{\frac{1}{p'}}$$

from which we find that

$$\mathbb{E}\left(\left|(R^N_{l+1}))(v_{l+1})(x_l)\right|^p |\eta^N_l\right)^{\frac{1}{p}}$$
$$\leq 2\ a(p)\left[\int_{E_{l+1}} Q_{l+1}(x_l,dx_{l+1})\left(\frac{dQ_{l+1}(x_l,\cdot)}{d\Phi_{l+1}(\eta^N_l)}(x_{l+1})\right)^{p'-1} v_{l+1}(x_{l+1})^{p'}\right]^{\frac{1}{p'}}$$

By definition (4.5.2) of functions h_{l+1} and in developing the Radon Nikodym derivative, we obtain

$$\frac{dQ_{l+1}(x_l,\cdot)}{d\Phi_{l+1}(\eta^N_l)}(x_{l+1}) = \frac{\eta^N_l(G_l)G_l(x_l)H_{l+1}(x_l,x_{l+1})}{\eta^N_l(G_lH_{l+1})(\cdot,x_{l+1})} \leq \|G_l\|h_{l+1}(x_{l+1})\ ,$$

which implies

$$\mathbb{E}\left(\left|(R^N_{l+1}))(v_{l+1})(x_l)\right|^p |\eta^N_l\right)^{\frac{1}{p}}$$
$$\leq 2\ a(p)\|B_l\|^{\frac{p'-1}{p'}}\left[\int_{E_{l+1}} Q_{l+1}(x_l,dx_{l+1})\,(h_{l+1}(x_{l+1}))^{p'-1}\,v_{l+1}(x_{l+1})^{p'}\right]^{\frac{1}{p'}}$$

Gathering the above arguments, we conclude that

$$\|(\widehat{v}_k - v_k)(x)\|_{L_p} \leq \sum_{k<l<n}\frac{2\ a(p)}{\sqrt{N}}q_{k,l}\ \left(Q_{k,l+1}(h^{p'-1}_{l+1}v^{p'}_{l+1})(x)\right)^{\frac{1}{p'}}\ .$$

■

Remarks : The constants $q_{k,l}$ could be largely reduced. In fact, $q_{k,l}$ comes from bounding $\|\prod_m \eta^N_m(G_m)\|_{L_p}$. In [29], the authors proved $\|\prod_m G_m\|_{L_2} + \frac{constant}{N}$ as a non asymptotic boundary for $\|\prod_m \eta^N_m(G_m)\|_{L_2}$. In most cases, the functions G take their values in $[0,1]$, then the boundary $\|\prod_m G_m\| \leq 1$ holds, but $\|\prod_m G_m\|_{L_2}$ is very small.

When the function G vanishes in some regions of the state space, we also mention that the particle model is only defined up to the first time $\tau^N = k$ such that $\eta^N_k(G_k) = 0$. We can prove that the event $\{\tau^N \leq n\}$ has an exponentially small probability to occur, with the number of particles N. In fact, the estimates presented in the above theorems can be extended to this singular situation by replacing $\widehat{v}_k$ by the particle estimates $\widehat{v}_k 1_{\tau^N \geq n}$. The stochastic analysis of these singular models are

quite technical, for further details we refer the reader to section 7.2.2 and section 7.4 in the book [40].

It is also very natural to assume the functions $(v_k)_{0\leq k\leq n}$ are bounded by M in the sense that

$$\left(Q_{k,l+1}(v_{l+1}^p)(x)\right)^{\frac{1}{p}} < M$$

, for any integer p. Then a new weak boundary

$$\frac{2\ a(p)\ (n-k)}{\sqrt{N}} M\left(1 \vee \left(\|h\|^2 \|G\|^{n-k}\right)\right)$$

is provided to simplify the notations, where $\|h\| = \max_k \|h_k\|$ and $\|G\| = \max_k \|G_k\|$ To understand better the $\mathbb{L}_p$-mean error bounds in the theorem, we deduce the following exponential concentration inequality:

Proposition 4.5.2. *For any $0 \leq k \leq n$ and any $\epsilon > 0$, we have*

$$\sup_{x\in E_k} \mathbb{P}\left(|v_k(x) - \widehat{v}_k(x)| > \frac{c}{\sqrt{N}} + \epsilon\right) \leq exp\left(-N\epsilon^2/c^2\right) \ , \tag{4.5.3}$$

with constant $c = 2(n-k)M\left(1 \vee \left(\|h_k\|^2 \|G\|^{n-k}\right)\right)$.

Proof : This result is a direct consequence from the fact that for any non negative random variable U such that

$$\exists b < \infty \text{ s.t. } \forall r \geq 1 \qquad \mathbb{E}\left(U^r\right)^{\frac{1}{r}} \leq a(r)\ b \ \Rightarrow \mathbb{P}\left(U \geq b + \epsilon\right) \leq \exp\left(-\epsilon^2/(2b^2)\right) .$$

To check this claim, we develop the exponential and verify that

$$\forall t \geq 0 \ \ \mathbb{E}\left(e^{tU}\right) \leq \exp\left(\frac{(bt)^2}{2} + bt\right) \Rightarrow \mathbb{P}(U \geq b + \epsilon) \leq \exp\left(-\sup_{t\geq 0}(\epsilon t - \frac{(bt)^2}{2})\right)$$

■

Similarly to Broadie-Glasserman model, the following proposition shows that in this model we also over-estimate the Snell envelope.

Proposition 4.5.3. *For any $0 \leq k \leq n$ and any $x_k \in E_k$*

$$\mathbb{E}\left(\widehat{v}_k(x_k)\right) \geq v_k(x_k) \ . \tag{4.5.4}$$

Proof. We can easily prove this inequality with a simple backward induction. The terminal condition $\widehat{v}_n = v_n$ implies directly the inequality at instant n. Assuming the inequality at time $k+1$, then the Jensen's inequality implies

$$\begin{aligned}
\mathbb{E}\left(\widehat{v}_k(x_k)\right) &\geq f_k(x_k) \vee \mathbb{E}\left(\widehat{Q}_{k+1}\widehat{v}_{k+1}(x_k)\right) \\
&= f_k(x_k) \vee \mathbb{E}\left(\int_{E_{k+1}^N} \widehat{Q}_{k+1}(x_k, dx_{k+1}) \mathbb{E}\left(\widehat{v}_{k+1}(x_{k+1})|\mathcal{F}_{k+1}^N\right)\right) \ .
\end{aligned}$$

By the induction assumption at time $k+1$, we have

$$\begin{aligned}\mathbb{E}\left(\int_{E_{k+1}^N} \widehat{Q}_{k+1}(x_k, dx_{k+1})\mathbb{E}\left(\widehat{v}_{k+1}(x_{k+1})|\mathcal{F}_{k+1}^N\right)\right) &\geq \mathbb{E}\left(\widehat{Q}_{k+1}v_{k+1}(x_k)\right) \\ &= Q_{k+1}v_{k+1}(x_k)\ .\end{aligned}$$

Then the inequality still holds at time k, which completes the proof.

■

4.6 Applications and extensions

In this section, we apply the Feynman-Kac methodology developed in section 4.4 to two type of importance sampling Monte Carlo techniques. We start with some important observation related to potential functions on transitions spaces.

For potential functions $G_k(X_k, X_{k+1})$ depending on the local transitions (X_k, X_{k+1}) of the reference process, the change of measure has the same form as in 4.3.1, replacing X_k by the Markov chain $\mathcal{X}_k = (X_k, X_{k+1})$. In this situation, the Snell envelop $\mathbf{v}_k(x_0, \ldots, x_k)$ associated with the payoff functions given bellow:

$$F_k(x_0, \ldots, x_k) = f_k(x_k) \prod_{0\leq p<k} G_p(x_p, x_{p+1}),$$

has the form

$$\mathbf{v}_k(x_0, \ldots, x_k) = v_k(x_k) \prod_{0\leq p<k} G_p(x_p, x_{p+1}). \tag{4.6.1}$$

The sequence of functions $(u_k)_{0\leq k\leq n}$ satisfies the backward recursion:

$$\begin{aligned} u_n &= f_n \\ u_k(x_p) &= f_p(x_p) \vee \int M_{k+1}(x_k, dx_{k+1})G_k(x_k, x_{k+1})u_{k+1}(x_{k+1}). \end{aligned} \tag{4.6.2}$$

This equation has exactly the same form as 4.2.4, by replacing the function $B_k(x_k)$ by the function $G_k(x_k, x_{k+1})$.

We illustrate these properties in two situations.

The first one concerns the design of more general change of reference measure. For instance, let us suppose we are given a judicious Markov transition $M'_k(x_{k-1}, x_k)$ such that $M'_k(x_{k-1}, \cdot)$ is absolutely continuous w.r.t. $M_k(x_{k-1}, \cdot)$. In this situation, we have

$$\begin{aligned}&\mathbb{E}(f_n(X_n) \prod_{0\leq p<n} G_p(X_p)) \\ &\qquad = \mathbb{E}\left(f_n(X'_n) \prod_{0\leq p<n} \left[G_p(X'_p)\frac{dM_{p+1}(X'_p, \cdot)}{dM'_{p+1}(X'_p, \cdot)}(X'_{p+1})\right]\right),\end{aligned} \tag{4.6.3}$$

where $(X'_p)_{0\leq p\leq n}$ is a Markov chain with initial condition $\eta'_0 = \eta_0 = law(X_0)$, and Markov transitions M'_p. We can rewrite 4.6.3 as follows:

$$\mathbb{E}(f_n(X_n)\prod_{0\leq p<n} G_p(X_p)) = \mathbb{E}(f_n(X'_n)\prod_{0\leq p<n} G'_p(X'_p, X'_{p+1})),$$

with $G'_p(x_p, x_{p+1}) = G_p(x_p)\frac{dM_{p+1}(x_p,\cdot)}{dM'_{p+1}(x_p,\cdot)}(x_{p+1})$.

The second example concerns the design of an importance sampling strategy. Suppose we are given a sequence of positive payoff functions $(f_k)_{0\leq k\leq n}$, with $f_0 \equiv 1$. In this situation, we have

$$\mathbb{E}(f_n(X_n)) = \mathbb{E}(\prod_{0\leq p<n} G_p(X_p, X_{p+1}))$$

, with the potential function $G_p(x_p, x_{p+1}) = \frac{f_{p+1}(x_{p+1})}{f_p(x_p)}$. In this context, the Snell envelop 4.6.1 and 4.6.2 are given by the backward recursion:

$$\begin{aligned} u_n &= 1 \\ u_p(x_p) &= 1 \vee \int M_{p+1}(x_p, dx_{p+1}) G_p(x_p, x_{p+1}) u_{p+1}(x_{p+1}). \end{aligned}$$

4.7 Numerical simulations

In this section, we give numerical examples to test our new algorithm, the *Stochastic Mesh with Change of Measure (SMCM)*, on Bermudan options from dimension 1 up to 5, compared with the standard *Stochastic Mesh (SM)* algorithm without change of measure.

4.7.1 Prices dynamics and options model

In our numerical tests we have considered a simple Black-Scholes price model. However, notice that both algorithms (*SM* and *SMCM*) can be applied in a general Markovian framework. The asset prices are modeled by a d-dimensional Markov process (S_t) such that each component (i.e. each asset) follows a geometric Brownian motion under the risk-neutral measure, that is, for assets $i = 1, \cdots, d$,

$$dS_t(i) = S_t(i)(rdt + \sigma dz^i_t)\ , \tag{4.7.1}$$

where z^i, for $i = 1, \cdots, d$ are independent one dimensional standard Brownian motions. Unless otherwise specified, the interest rate r is set to 10% annually and the volatility is supposed to be the same for all assets, $\sigma = 20\%$ annually. The starting prices of the assets are for all $i = 1, \cdots, d$, $S_{t_0}(i) = 1$. We consider two types of Bermudan options with maturity $T = 1$ year and 11 equally distributed exercise opportunities at dates $t_k = kT/n$ with $k = 0, 1, \cdots, n = 10$, associated with two different payoffs:

1. Geometric average put option with payoff $(K - \prod_{i=1}^{d} S_T(i))_+$,
2. Arithmetic average put option with payoff $(K - \frac{1}{d}\sum_{i=1}^{d} S_T(i))_+$,

Note that the geometric average put payoff involves the process $\prod_{i=1}^{d} S(i)$ which can be identified to a one-dimensional non standard exponential Brownian motion. For this specific case of geometric put payoff, we chose to vary, in our simulations, the short term interest rate and the volatility with the number of underlying assets d, such that the option value remains the same for all d:

$$r(d) = r/d \ , \quad \text{and} \quad \sigma(d) = \sigma/\sqrt{d} \ . \tag{4.7.2}$$

Then, we chose as a benchmark value the estimate obtained by the standard Stochastic Mesh approach with $N = 6400$ mesh points for $d = 1$ asset. These benchmark values are reported on Table 4.1.

Strike	$K = 0.95$	$K = 0.85$	$K = 0.75$
Option value	0.0279	0.0081	0.0015

Table 4.1: Benchmark values for the geometric put option obtained by using the Stochastic Mesh method with 10000 particles. $n = 11$ exercise opportunities, $T = 1$, $S_0 = 1$ and $r = 10\%/d$, $\sigma_i = 20\%/\sqrt{d}$ for the geometric payoff and $r = 10\%$, $\sigma_i = 20\%$ for the arithmetic payoff.

4.7.2 Choice of potential functions

We consider the Markov chain $(X_k)_{0\leq k\leq n}$, taking values on $E_k = \mathbb{R}^{+d}$, obtained by discretization of the time-continuous process S defined by (4.7.1) at times of exercise opportunities, $0 = t_0 < \cdots < t_n = T$, such that for all $k = 0, \cdots, n$, $X_k = S_{t_k}$.
Now, we can introduce the sequence of positive functions $(G_k)_{1\leq k\leq n}$, defining the change of measure (4.3.1), as follows:

$$\begin{cases} G_0(x_1) = (f_1(x_1) \vee \varepsilon)^\alpha \ , \\ \\ G_k(x_k, x_{k+1}) = \frac{(f_{k+1}(x_{k+1})\vee\varepsilon)^\alpha}{(f_k(x_k)\vee\varepsilon)^\alpha} \ , \quad \text{for all } k = 1, \cdots, n-1 \ , \end{cases} \tag{4.7.3}$$

where f_k are the payoff functions and $\alpha \in (0, 1]$ and $\varepsilon > 0$ are parameters fixed in our simulations to the values $\alpha = 1/5$ and $\varepsilon = 10^{-7}$.

4.7.3 Numerical results

For each example, we have performed the algorithm for different numbers of mesh points $N = 100\,,\ 200\,,\ 400\,,\ 800\,,\ 1600\,,\ 3200\,,\ 6400$. 1000 runs of both algorithms (

Stochastic Mesh (SM) and *Stochastic Mesh with Change of Measure (SMCM)*) were performed to compute the mean and confidence intervals of each estimate.

Simulations results are reported in Figure 1.3, 1.4 and 1.5 for the geometric and arithmetic put payoff, with strikes corresponding to standard out of the money puts to deep out of the money puts: $K = 0.95$, $K = 0.85$ and $K = 0.75$. Notice that both algorithms (the Stochastic Mesh algorithm with and without Change of Measure) have been implemented without any standard variance reduction technique (control variate, stratification, ...). In term of complexity, the Stochastic Mesh algorithm with Change of Measure is equivalent to the standard Stochastic Mesh algorithm: the complexity is in both cases quadratic with the number of mesh points $O(N^2)$ since the number of operations required to operate the change of measure is negligible.

We have reported on our graphs to types of estimates:

- the *Positively-biased estimator* provided by the backward induction on the value function;
- the *Negatively-biased estimator* provided by the associated optimal exercise policy. This estimate is obtained via a two-step procedure: first, the optimal policy is approximated in the backward induction on the value function, then the policy is evaluated using the standard forward Monte Carlo procedure. Note that the resulting estimator is known to provide a lower bound (in average) to the option price. In our simulation, we have used $N_{forward} = 10000$ Monte Carlo forward simulations.

As expected, one can observe on Table 4.2, that the *SMCM* algorithm allows to obtain an estimate, $\hat{v}_{SMCM}$, with the same complexity but with a smaller variance than the standard *SM* algorithm estimate, $\hat{v}_{SM}$, especially for deep out the money options.
More surprisingly, one can observe on Table 4.2 and Figure 1.3, 1.4 and 1.5 that the *SMCM* algorithm also allows to reduce significantly the estimator bias which is known to compose the growing part of the error when the number of underlying assets increases. For instance, one can notice that the *SMCM* algorithm achieves the convergence in average of the Positively-biased estimate to the Negatively-biased estimate for a number of mesh points much smaller than for the *SM* algorithm. Hence, the *SMCM* could also be a way to deal with high dimensional optimal stopping problems since the algorithm complexity remains insensitive to the dimension whereas the convergence rate is not significantly reduced.

Payoff	K	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
Geometric	0.95	1 (1%)	1 (3%)	1 (6%)	1 (9%)	1 (10%)
Put	0.85	5 (2%)	8 (6%)	6 (11%)	4 (14%)	3 (14%)
	0.75	18 (6%)	28 (11%)	18 (17%)	16 (18%)	11 (16%)
Arithmetic	0.95	1 (1%)	3 (2%)	3 (7%)	4 (13%)	5 (18%)
Put	0.85	5 (2%)	13 (6%)	24 (19%)	56 (24%)	100 (20%)
	0.75	18 (6%)	71 (15%)	363 (14%)	866 (16%)	− (−)

Table 4.2: Variance ratio ($\frac{Var(\hat{v}_{SM})}{Var(\hat{v}_{SMCM})}$) and Bias ratio ($\frac{\mathbb{E}(\hat{v}_{SM})-\mathbb{E}(\hat{v}_{SMCM})}{E(\hat{v}_{SM})}$) (within parentheses) computed over 1000 runs for $N = 3200$ mesh points. (For the arithmetic put, when $d = 5$ and $K = 0.75$, the 1000 estimates provided by the standard SM algorithm were all equal to zero, hence the associated variance ratio has not been reported).

Appendices

Appendix A

Appendix

A.1 Proof of Lemma 3.5.2

Set

$$\delta_{l,n}(N) := \inf_{x \in E'_l} \eta_n^N(g_{l,x}) \ ,$$

with the function $g_{l,x}$ defined in (A.2.2). Note that

$$\mathbb{P}\left(\delta_{l,n}(N) = 0\right) \leq \sum_{x \in E'_l} \mathbb{P}\left(\eta_n^N(g_{l,x}) = 0\right) \ .$$

On the other hand, for any $\epsilon \in [0,1)$ we have

$$\mathbb{P}\left(\eta_n^N(g_{l,x}) = 0\right) \leq \mathbb{P}\left(\left|\eta_n^N(g_{l,x}) - \eta_n(g_{l,x})\right| > \epsilon \ \eta_n(g_{l,x})\right) \ .$$

Arguing as in (3.5.7), for any $x \in E'_l$ s.t. $\eta_n(g_{l,x})\,(= \mathbb{P}\left(X'_l = x\right)) > 0$ we prove that

$$\sqrt{N}\ \mathbb{E}\left(\left|\eta_n^N(g_{l,x}) - \eta_n(g_{l,x})\right|^r\right)^{\frac{1}{r}} \leq 2\ a(r)\ (n+1)\ \eta_n(g_{l,x})^{-1} \tag{A.1.1}$$

and therefore

$$\mathbb{P}\left(\left|\eta_n^N(g_{l,x}) - \eta_n(g_{l,x})\right| \geq \left(\frac{2(n+1)}{\sqrt{N}} + \epsilon\right) \eta_n(g_{l,x})\right) \leq \exp\left(-\frac{N\epsilon^2}{8(n+1)^2}\right) .$$

For any $N \geq (2(n+1)/(1-\epsilon))^2$, this implies that

$$\mathbb{P}\left(\delta_{l,n}(N) = 0\right) \leq \mathrm{Card}(E'_l)\ \exp\left(-\frac{N\epsilon^2}{8(n+1)^2}\right) .$$

If we choose, $\epsilon = 1/2$ and $N \geq (4(n+1))^2$, we conclude that

$$\mathbb{P}\left(\delta_{l,n}(N) = 0\right) \leq \mathrm{Card}(E'_l)\ \exp\left(-\frac{N}{32(n+1)^2}\right) .$$

On the other hand, by construction we have the almost sure estimate

$$\eta_n^N(g_{l,\xi_{l,n}^i}) = \sum_{x\in E_l'} \eta_n^N(g_{l,x})\ 1_{\xi_{l,n}^i=x} \geq \delta_{l,n}(N)\ 1_{\delta_{l,n}(N)>0} + \frac{1}{N}\ 1_{\delta_{l,n}(N)=0}\quad ,$$

from which we find that

$$\eta_n^N(g_{l,\xi_{l,n}^i})^{-1} \ \leq\ \delta_{l,n}(N)^{-1}\ 1_{\delta_{l,n}(N)>0} + N\ 1_{\delta_{l,n}(N)=0}\ .$$

Therefore, we have

$$\begin{aligned}\left|\left|\eta_n^N(g_{l,\xi_{l,n}^i})^{-1}\right|\right|_p \ &\leq\ ||\delta_{l,n}(N)^{-1}\ 1_{\delta_{l,n}(N)>0}||_p + N||1_{\delta_{l,n}(N)=0}||_p \\ &\leq\ \sum_{x\in E_l'} ||\eta_n^N(g_{l,x})^{-1}\ 1_{\eta_n^N(g_{l,x})>0}||_p + N\ \mathbb{P}(\delta_{l,n}(N)=0)^{1/p}\ .\end{aligned}$$

If we set $\overline{g}_{l,n}(x) = g_{l,x}/\eta_n(g_{l,x})$, using the fact that

$$\frac{1}{1-u} = 1+u+u^2+\frac{u^3}{1-u}\ ,$$

for any $u\neq 1$, and $\eta_n^N(\overline{g}_{l,x})^{-1}\ 1_{\eta_n^N(g_{l,x})>0} \leq N\ \eta_n(g_{l,x})$, we find that

$$\begin{aligned}\eta_n^N(\overline{g}_{l,x})^{-1}\ 1_{\eta_n^N(g_{l,x})>0} \leq 1 + \left|1-\eta_n^N(\overline{g}_{l,x})\right| + \left(1-\eta_n^N(\overline{g}_{l,x})\right)^2 \\ +N\ \eta_n(g_{l,x})\ \left|1-\eta_n^N(\overline{g}_{l,x})\right|^3\ .\end{aligned}$$

Combining this estimate with (A.1.1), for any $p\geq 1$ we prove the following upper bound

$$\begin{aligned}\|\eta_n^N(\overline{g}_{l,x})^{-1}\ 1_{\eta_n^N(g_{l,x})>0}\|_p \ &\leq\ 1+\frac{1}{\sqrt{N}}\ 2a(p)(n+1) + (2a(2p)(n+1))^2\frac{1}{N} \\ &\quad +\frac{1}{\sqrt{N}}(2a(3p)(n+1))^3\ ,\end{aligned}$$

from which we find the rather crude estimates

$$\|\eta_n^N(\overline{g}_{l,x})^{-1}\ 1_{\eta_n^N(g_{l,x})>0}\|_p \leq 1+\frac{3}{\sqrt{N}}\ a'(p)\ (n+1)^3\ ,$$

with the collection of finite constants $a'(p) := 2a(p)+(2a(2p))^2+(2a(3p))^3$. Using the above exponential inequalities, we find that

$$\left|\left|\eta_n^N(g_{l,\xi_{l,n}^i})^{-1}\right|\right|_p$$

$$\leq \textstyle\sum_{x\in E_l'}\ \frac{1}{\eta_n(g_{l,x})}\left[1+\frac{3}{\sqrt{N}}\ a'(p)\ (n+1)^3\right] + N\ \mathrm{Card}(E_l')^{1/p}\ \exp\left(-\frac{N}{32p(n+1)^2}\right)\ ,$$

completing the proof of the lemma. ∎

A.2 Proof of Lemma 3.5.3

By construction, we have

$$\forall x \in \widehat{E}_{l,n} \qquad M'_{l+1}(f)(x) = \frac{\eta_l^N M_{l,n}((1_x \circ \pi_l)\ (f \circ \pi_{l+1}))}{\eta_l^N M_{l,n}((1_x \circ \pi_l))} \ . \tag{A.2.1}$$

Thus, by (A.2.1) we have

$$\widehat{M}'_{l+1}(f)(x) - M'_{l+1}(f)(x) := \frac{\eta_n^N(g_{l,x} f_{l+1})}{\eta_n^N(g_{l,x})} - \frac{\eta_l^N M_{l,n}(g_{l,x} f_{l+1})}{\eta_l^N M_{l,n}(g_{l,x})} \ ,$$

for any $x \in \widehat{E}_{l,n}$, with the collection of functions

$$g_{l,x} := 1_x \circ \pi_l \quad \text{and} \quad f_{l+1} := f \circ \pi_{l+1} \ . \tag{A.2.2}$$

It is readily checked that

$$\widehat{M}'_{l+1}(f)(x) - M'_{l+1}(f)(x) \ = \ \frac{1}{\eta_n^N(\bar{g}_{l,x}^N)} \ \left[\eta_n^N(\bar{f}_{l+1,x}^N) - \eta_l^N M_{l,n}(\bar{f}_{l+1,x}^N)\right] \ ,$$

for any $x \in \widehat{E}_{l,n}$, with the pair of $\mathcal{F}_l^N$-measurable functions

$$\bar{f}_{l+1,x}^N := \frac{g_{l,x}}{\eta_l^N M_{l,n}(g_{l,x})} \left[f_{l+1} - \frac{\eta_l^N M_{l,n}(g_{l,x} f_{l+1})}{\eta_l^N M_{l,n}(g_{l,x})}\right] \quad \text{and} \quad \bar{g}_{l,x}^N = \frac{g_{l,x}}{\eta_l^N M_{l,n}(g_{l,x})} \ .$$

It is also important to observe as $g_{l,x}$ varies only on E'_l , then

$$\eta_l^N M_{l,n}(g_{l,x}) = \eta_l^N(g_{l,x}) \leq 1 \ .$$

In this notation, for any $0 \leq i \leq N$ and any $p \geq 1$, we have

$$\begin{aligned} &\left|\left|\widehat{M}'_{l+1}(f)(\xi_{l,n}^i) - M'_{l+1}(f)(\xi_{l,n}^i)\right|\right|_p \\ &\qquad \leq \left|\left|\eta_n^N(g_{l,\xi_{l,n}^i})^{-1}\right|\right|_{2p} \ \left|\left|\eta_n^N(\bar{f}_{l+1,\xi_{l,n}^i}^N) - \eta_l^N M_{l,n}(\bar{f}_{l+1,\xi_{l,n}^i}^N)\right|\right|_{2p} \ . \end{aligned} \tag{A.2.3}$$

The collection of random functions $\bar{f}_{l+1,\xi_{l,l}^j}^N$ are well defined and we have

$$\begin{aligned} &\left(\eta_n^N(\bar{f}_{l+1,\xi_{l,n}^i}^N) - \eta_l^N M_{l,n}(\bar{f}_{l+1,\xi_{l,n}^i}^N)\right)^\beta \\ &= \frac{1}{\eta_l^N\left(g_{l,\xi_{l,n}^i}\right)} \ \frac{1}{N}\sum_{j=1}^N \ \left[\eta_n^N(\bar{f}_{l+1,\xi_{l,l}^j}^N) - \eta_l^N M_{l,n}(\bar{f}_{l+1,\xi_{l,l}^j}^N)\right]^\beta \ 1_{\xi_{l,l}^j = \xi_{l,n}^i} \ , \end{aligned}$$

for any $\beta \geq 0$. Combining the above formula for $\beta = 2p$ and Holder's inequality, we prove that

$$\begin{aligned} &\left|\left|\eta_n^N(\bar{f}_{l+1,\xi_{l,n}^i}^N) - \eta_l^N M_{l,n}(\bar{f}_{l+1,\xi_{l,n}^i}^N)\right|\right|_{2p} \\ &\leq \left|\left|\eta_l^N\left(g_{l,\xi_{l,n}^i}\right)^{-1}\right|\right|_q^{1/(2p)} \times \sup_{1\leq j\leq N}\left|\left|\eta_n^N(\bar{f}_{l+1,\xi_{l,l}^j}^N) - \eta_l^N M_{l,n}(\bar{f}_{l+1,\xi_{l,l}^j}^N)\right|\right|_{2pq'} \ , \end{aligned}$$

for any $q, q' \geq 1$, with $\frac{1}{q} + \frac{1}{q'} = 1$.

We observe that, as $(\xi^j_{l,l}, (\xi^i_{l,l})_{0\leq i\leq N}, (\xi^i_{l,n})_{0\leq i\leq N})$ have the same distribution, for any $1 \leq j \leq N$, then for any function h and any $1 \leq j, j' \leq N$ we have:

$$\mathbb{E}\left(h(\xi^j_{l,l}, (\xi^i_{l,l})_{0\leq i\leq N}, (\xi^i_{l,n})_{0\leq i\leq N})\right) = \mathbb{E}\left(h(\xi^{j'}_{l,l}, (\xi^i_{l,l})_{0\leq i\leq N}, (\xi^i_{l,n})_{0\leq i\leq N})\right) \ ,$$

which implies that

$$\begin{aligned}\sup_{1\leq j\leq N} \left\|\eta^N_n(\bar{f}^N_{l+1,\xi^j_{l,l}}) - \eta^N_l M_{l,n}(\bar{f}^N_{l+1,\xi^j_{l,l}})\right\|_{2pq'} \\ = \left\|\eta^N_n(\bar{f}^N_{l+1,\xi^j_{l,l}}) - \eta^N_l M_{l,n}(\bar{f}^N_{l+1,\xi^j_{l,l}})\right\|_{2pq'} \ .\end{aligned}$$

As this equation works for any $1 \leq j \leq N$, in further development we take $j = 1$ to simplify the notation.

Using Lemma 4.5.1, and recalling that $\eta^N_l M_{l,n}(g_{l,x}) = \eta^N_l(g_{l,x})$, for any $1 \leq j \leq N$ we prove the almost sure estimate

$$\begin{aligned}&\sqrt{N}\ \mathbb{E}\left(\left|[\eta^N_n - \eta^N_l M_{l,n}](\bar{f}^N_{l+1,\xi^1_{l,l}})\right|^{2pq'} \left|\mathcal{F}^N_l\right.\right)^{\frac{1}{2pq'}} \\ &\leq 2\ a(2pq')(n-l)\ \left[\eta^N_l M_{l,n}\left(\left|\bar{f}^N_{l+1,\xi^1_{l,l}}\right|^{2pq'}\right)\right]^{\frac{1}{2pq'}} \\ &\leq 4\ a(2pq')(n-l)\ \|f_{l+1}\|\ \left(\eta^N_l M_{l,n}(g_{l,\xi^1_{l,l}})\right)^{\frac{1}{2pq'}-1} \ .\end{aligned}$$

This yields that

$$\begin{aligned}\sqrt{N}\mathbb{E}\left(\left|[\eta^N_n - \eta^N_l M_{l,n}](\bar{f}^N_{l+1,\xi^1_{l,l}})\right|^{2pq'} \left|\mathcal{F}^N_l\right.\right)^{\frac{1}{2pq'}} \\ \leq 4\ a(2pq')(n-l)\|f_{l+1}\|\ \eta^N_l(g_{l,\xi^1_{l,l}})^{-1} \ ,\end{aligned}$$

and therefore

$$\begin{aligned}&\sqrt{N}\left\|\eta^N_n(\bar{f}^N_{l+1,\xi^i_{l,n}}) - \eta^N_l M_{l,n}(\bar{f}^N_{l+1,\xi^i_{l,n}})\right\|_{2pq'} \\ &\leq 4\ a(2pq')(n-l)\|f_{l+1}\|\ \left\|\eta^N_l\left(g_{l,\xi^i_{l,n}}\right)^{-1}\right\|_q^{1/(2p)} \left\|\eta^N_l(g_{l,\xi^1_{l,l}})^{-1}\right\|_{2pq'} \ .\end{aligned}$$

Finally, by (A.2.3), we conclude that

$$\begin{aligned}&\sqrt{N}\ \left\|\widehat{M}'_{l+1}(f)(\xi^i_{l,n}) - M'_{l+1}(f)(\xi^i_{l,n})\right\|_p \\ &\leq 4\ a(2pq')(n-l)\|f_{l+1}\|\ \left\|\eta^N_n(g_{l,\xi^i_{l,n}})^{-1}\right\|_{2p} \left\|\eta^N_l\left(g_{l,\xi^i_{l,n}}\right)^{-1}\right\|_q^{1/(2p)} \\ &\qquad\qquad \times\ \left\|\eta^N_l(g_{l,\xi^1_{l,l}})^{-1}\right\|_{2pq'} \ .\end{aligned}$$

We prove (3.5.13), by taking $q = 1 + 2p$ and $q' = 1 + 1/(2p)$ so that $q = 2pq' \geq 2p$

$$\sqrt{N} \ \left|\left|\widehat{M}'_{l+1}(f)(\xi^i_{l,n}) - M'_{l+1}(f)(\xi^i_{l,n})\right|\right|_p$$

$$\leq 4 \ a(1+2p)(n-l)\|f_{l+1}\| \ \sup_{l\leq k\leq n} \left|\left|\eta^N_k(g_{l,\xi^1_{l,k}})^{-1}\right|\right|^{2+1/(2p)}_{1+2p} \ .$$

This end of proof is now a direct consequence of Lemma 3.5.2. ∎

Bibliography

[1] Y. Achdou and O. Pironneau. Computational methods for option pricing. *SIAM, Frontiers in Applied Mathematics series*, (2005). [cited at p. 36]

[2] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal Royal Statistical Society B*, vol. 72, no. 3, pp. 269-342 (2010). [cited at p. 50, 75]

[3] A. N. Avramidis and H. Matzinger, *Convergence of the Stochastic Mesh Estimator for Pricing Bermudan Options*, Journal of Comput. Finance, vol. 7, 73-91 (2004). [cited at p. 78, 96]

[4] V. Bally, G. Pagès and J. Printems, *A quantization tree method for pricing and hedging multidimensional American options*, Math. Finance, 15(1), pp. 119-168 (2005). [cited at p. 9, 91, 92, 93, 111]

[5] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function*, Probab. Th. Related Fields, pp. 104, 43-60, (1996). [cited at p. 86]

[6] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations: II. Approximation of the density*, Monte Carlo Methods Appl., 2, pp. 93-128, (1996). [cited at p. 86]

[7] O. Bardou, S. Bouthemy and G. Pagès. Optimal quantization for the pricing of swing options. *Applied Mathematical Finance*, vol. 16, Issue 2, pp. 183 - 217 (2009). [cited at p. 91]

[8] R. Bahr and S. Hamori. *Hidden Markov models : Applications to finance and economics*. Advanced studies in theoretical and applied econometrics. vol. 40, Kluwer Academic Publishers (2004). [cited at p. 72]

[9] C. Bender and J. Steiner. Least-squares Monte Carlo for backward SDEs. *Numerical Methods in finance, Springer* (2011). [cited at p. -]

[10] S. Ben Hamida and R. Cont. Recovering Volatility from Option Prices by Evolutionary Optimization. *Journal of Computational Finance*, Vol. 8, No. 4, pp. 43-76 (2005). [cited at p. 36]

[11] T. Björk, M. H A Davis and C. Landèn. Optimal investment under partial information *Mathematical Methods of Operations Research* (2010). [cited at p. 71]

[12] V. S. Borkar. Controlled diffusion processes. *Probability Surveys*, Vol. 2, pp. 213-244 (2005). [cited at p. 71]

[13] B. Bouchard and N. Touzi, *Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations*, Stochastic Process Appl. 111, pp. 175-206, 2004. [cited at p. 85]

[14] B. Bouchard and X. Warin. Monte-Carlo valuation of American options: facts and new algorithms to improve existing methods. *Numerical Methods in finance, Springer* (2011). [cited at p. 8, 9, 12, 64, 108, 109, 111, 112]

[15] J. P. Bouchaud and M. Potters. Back to basics: historical option pricing revisited. *Philosophical Transactions: Mathematical, Physical & Engineering Sciences*, 357(1758), pp. 2019-2028 (2002). [cited at p. -]

[16] P. P. Boyle and D. Emanuel. Discretely adjusted option hedges. *Journal of Financial Economics*, vol. 8, pp. 259-282 (1980). [cited at p. -]

[17] D. Brigo, T. Bielecki and F. Patras. *Credit Risk Frontiers*. Wiley–Bloomberg Press (2011). [cited at p. 50]

[18] M. Broadie and P. Glasserman. Estimating security prices using simulation. *Management Science*, 42, pp. 269–285 (1996). [cited at p. 54, 56]

[19] M. Broadie and P. Glasserman. A Stochastic Mesh Method for Pricing High-Dimensional American Options. *Journal of Computational Finance*, vol. 7, pp. 35–72, (2004). [cited at p. 64, 65, 78, 95, 96, 99, 115, 116]

[20] J. Bucklew and G. Wise, *Multidimensional Asymptotic Quantization Theory with rth Power distortion Measures*, IEEE Trans. on Information Theory, Special issue on Quantization, 28, no. 2, pp. 239-247 (1982). [cited at p. 91]

[21] R. Carmona and S. Crépey. Importance Sampling and Interacting Particle Systems for the Estimation of Markovian Credit Portfolios Loss Distribution. *International Journal of Theoretical and Applied Finance*, vol. 13, No. 4 (2010) 577 – 602. [cited at p. 35, 50]

[22] R. Carmona, P. Del Moral, P. Hu and N. Oudjane (eds.). *Numerical Methods in Finance* of Springer Proceedings in Mathematics, Vol. 12, Springer-Verlag (2012). [cited at p. 124]

[23] R. Carmona, P. Del Moral, P. Hu and N. Oudjane. An Introduction to Particle Methods with Financial Applications, in *Numerical Methods in Finance*, Vol 12, pp. 3–50, Springer-Verlag (2012). [cited at p. 124]

[24] R. Carmona, J.-P. Fouque and D. Vestal. Interacting Particle Systems for the Computation of Rare Credit Portfolio Losses. *Finance and Stochastics*, vol. 13, no. 4, 2009 pp. 613-633 (2009). [cited at p. 35, 50, 52, 124]

[25] R. Carmona and N. Touzi. Optimal Multiple Stopping and Valuation of Swing Options. *Mathematical Finance*, Vol. 18, Issue 2, pp. 239-268 (2008). [cited at p. -]

[26] J. F. Carrière. Valuation of the Early-Exercise Price for Options using Simulations and Nonparametric Regression. *Insurance : Mathematics and Economics*, 19, 19-30 (1996). [cited at p. 64]

[27] R. Casarin. Simulation Methods for Nonlinear and Non-Gaussian Models in Finance. *Premio SIE, Rivista Italiana degli Economisti*, vol. 2, pp. 341-345 (2005). [cited at p. 36]

[28] R. Casarin and C. Trecroci. Business Cycle and Stock Market Volatility: A Particle Filter Approach, *Cahier du CEREMADE N. 0610*, University Paris Dauphine (2006). [cited at p. 36]

[29] F. Cérou, P. Del Moral, A. Guyader *A non asymptotic variance theorem for unnormalized Feynman-Kac particle models*. HAL-INRIA RR-6716 (2008). To appear in the journal : Annales de l'Institut Henri Poincar (2010) [cited at p. 127]

[30] A. Cerny. Dynamic programming and mean variance hedging in discrete time. *Tanaka Business School Discussion Papers* TBS/DP04/15 London: Tanaka Business School (2004). [cited at p. -]

[31] N. Chen and P. Glasserman. Malliavin Greeks without Malliavin calculus. Stochastic Processes and their Applications 117, pp. 1689 – 1723 (2007). [cited at p. 59]

[32] E. Clément, D. Lamberton and P. Protter. An analysis of a least squares regression method for American option pricing. *Finance and Stochastics*, 6, 449-472, (2002). [cited at p. 64]

[33] R. Cont and P. Tankov. Non-parametric calibration of jump-diffusion option pricing models. *Journal of computational finance*, vol. 7, no. 3, pp. 1–49 (2004). [cited at p. 72]

[34] R. Cont and P. Tankov. Retrieving Levy processes from option prices: regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization*, vol. 45, no. 1, pp. 1–25 (2007). [cited at p. 72]

[35] A.L. Corcoran and R.L. Wainwright. A parallel island model genetic algorithm for the multiprocessor scheduling problem. *Proceedings of the 1994 ACM/SIGAPP Symposium on Applied Computing*, March 6-8, pp. 483-487, ACM Press (1994). [cited at p. 49]

[36] D. D. Creal. A survey of sequential Monte Carlo methods for economics and finance. *To appear in Econometric Reviews* (2011). [cited at p. 40]

[37] S. Crépey. Calibration of the local volatility in a trinomial tree using Tikhonov regularization. *Inverse Problems*, vol. 19, pp. 91–127 (2003) [cited at p. 72]

[38] D. Crisan, P. Del Moral and T. Lyons. Interacting Particle Systems Approximations of the Kushner Stratonovitch Equation. *Advances in Applied Probability*, vol.31, no. 3, 819-838 (1999). [cited at p. 58]

[39] Del Moral P. Non Linear Filtering: Interacting Particle Solution. *Markov Processes and Related Fields*, Volume 2 Number 4, 555–580 (1996). [cited at p. 35, 43]

[40] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, New York: Springer-Verlag (2004). [cited at p. 24, 35, 38, 43, 44, 53, 82, 102, 123, 128]

[41] P. Del Moral, Measure Valued Processes and Interacting Particle Systems. Application to Nonlinear Filtering Problems. *Annals of Applied Probabability*, vol. 8, no. 2, pp. 1254-1278 (1998). [cited at p. 65, 81]

[42] P. Del Moral, A. Doucet and A. Jasra. Sequential Monte Carlo samplers. *J. Royal Statist. Soc.* B, vol. 68, pp. 411–436 (2006). [cited at p. -]

[43] P. Del Moral and A. Doucet. *Sequential Monte Carlo and Genetic particle models. Theory and Practice*, in preparation, Chapman & Hall (2011). [cited at p. 43]

[44] P. Del Moral, A. Doucet and A. Jasra. On Adaptive Resampling Procedures for Sequential Monte Carlo Methods. HAL-INRIA RR-6700-2008. To appear in Bernoulli (2011). [cited at p. 82]

[45] P. Del Moral, A. Doucet and S. S. Singh. A Backward Particle Interpretation of Feynman-Kac Formulae, *M2AN*, vol 44, no. 5, pp. 947–976 (sept. 2010). [cited at p. 47, 48, 82]

[46] P. Del Moral, A. Doucet and S. S. Singh. An Adaptive Sequential Monte Carlo Method for Approximate Bayesian Computation , *Statistics and Computing*, vol. 1-12 (2011). [cited at p. 73]

[47] P. Del Moral and J. Garnier. Genealogical Particle Analysis of Rare events. *Annals of Applied Probability*, vol. 15, no. 4, 2496–2534 (2005). [cited at p. 38, 51, 52, 116, 121]

[48] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms, *Annales de l'Institut Henri Poincaré*, Vol. 37, No. 2, 155–194 (2001). [cited at p. 65, 81]

[49] P. Del Moral, P. Hu and N. Oudjane. Snell envelope with path dependent multiplicative optimality criteria *HAL-INRIA*, RR-7360 (2010). [cited at p. 63, 64, 66]

[50] P. Del Moral, P. Hu, N. Oudjane and Br. Rémillard. On the Robustness of the Snell envelope. in *SIAM Journal on Financial Mathematics*, Vol. 2, pp. 587-626 (2011). [cited at p. 2, 25, 36, 64, 65, 66, 118, 123]

[51] P. Del Moral, P. Hu and L. Wu. On the concentration properties of Interacting particle processes, in *Foundations and Trends in Machine Learning*, vol. 3, nos. 3–4, pp. 225–389, (2012). [cited at p. 2, 17, 18, 43, 44, 45, 48, 73]

[52] P. Del Moral, J. Jacod. The Monte-Carlo Method for filtering with discrete time observations. Central Limit Theorems. T he Fields Institute Communications, Numerical Methods and Stochastics, Ed. T.J. Lyons, T.S. Salisbury, American Mathematical Society (2002). [cited at p. 73]

[53] P. Del Moral, J. Jacod and P. Protter. The Monte Carlo method for filtering with discrete time observations. *Probability Theory and Related Fields*, vol. 120, no. 3, pp. 346–368 (2001). [cited at p. 73]

[54] P. Del Moral and A. Jasra. SMC for option pricing, *Stochastic Analysis and Applications* Volume 29, Issue 2, pp. 292–316 (2011). [cited at p. 40]

[55] P. Del Moral and L. Miclo. *Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering*. Séminaire de Probabilits XXXIV, Ed. J. Azma and M. Emery and M. Ledoux and M. Yor, Lecture Notes in Mathematics, Springer-Verlag Berlin, Vol. 1729, 1–145 (2000). [cited at p. 35, 43]

[56] P. Del Moral, L. Miclo, F. Patras and S. Rubenthaler, *The Convergence to Equilibrium of Neutral Genetic Models*, Stochastic Analysis Appl., Volume 28, Issue 1, pp. 123 - 143 (2010). [cited at p. 79]

[57] P. Del Moral and Fr. Patras. Interacting path systems for credit risk. *Credit Risk Frontiers*. D. Brigo, T. Bielecki, F. Patras Eds. *Wiley–Bloomberg Press*, (2011), 649–674. *Short announcement available as Interacting path systems for credit portfolios risk analysis. INRIA:RR-7196 (2010).* [cited at p. 35, 50]

[58] P. Del Moral, Br. Rémillard and S. Rubenthaler. Monte Carlo approximations of American options that preserve monotonicity and convexity. (2011). [cited at p. 64]

[59] P. Del Moral and E. Rio. Concentration Inequalities for Mean Field Particle Models. HAL-INRIA RR-6901, to appear in the *Annals of Applied Probability* (2011). [cited at p. 43, 45, 48, 82]

[60] A. Doucet, L. Montesano and A. Jasra. Optimal filtering for partially observed point processes. *Proceedings ICASSP* (2006). [cited at p. 71]

[61] A. Doucet, J. F. G. de Freitas and N. J. Gordon (eds.) *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, (2001). [cited at p. -]

[62] P. Dupuis and H. Wang, *On the convergence from discrete to continuous time in an optimal stopping problem*, Ann. Appl. Probab., 15, pp. 1339-1366 (2005). [cited at p. 85]

[63] P. Dybvig. Distributional analysis of portfolio choice. *The Journal of Business*, 61(3):369–393 (1988). [cited at p. -]

[64] D. Egloff. Monte Carlo algorithms for optimal stopping and statistical learning. *Annals of Applied Probability* , 15, pp. 1–37 (2005). [cited at p. 65, 82]

[65] D. Egloff, M. Kohler, and N. Todorovic, *A dynamic look-ahead Monte Carlo algorithm for pricing Bermudan options*, Ann. Appl. Probab., 17, pp. 1138-1171 (2007). [cited at p. 83]

[66] J. J. Florentin. Optimal observability and optimal control. *J. of Electronics and Control*. Vol. 13, pp. 263–279 (1962). [cited at p. -]

[67] I. Florescu and Fr. Viens. Stochastic Volatility: Option Pricing using a Multinomial Recombining Tree. *Applied Mathematical Finance*, vol. 15, Issue 2, pp.151–181 (2008). [cited at p. -]

[68] J. P. Fouque, G. Papanicolaou and R. Sircar. Derivatives in financial markets with stochastic volatility. *Cambridge University Press* (2000). [cited at p. -]

[69] E. Fournié, J. M. Lasry, J. Lebuchoux, P. L. Lions and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, vol. 3, pp. 391–412 (1999). [cited at p. 54, 59, 60]

[70] E. Fournié, J. M. Lasry, J. Lebuchoux and P. L. Lions. Applications of Malliavin calculus to Monte Carlo methods in finance. II. *Finance and Stochastics*, vol. 5, pp. 201–236 (2001). [cited at p. 59, 60]

[71] M. C. Fu, D. B. Madan and T. Wang. Pricing continuous Asian options: a comparison of Monte Carlo and Laplace transform inversion methods. *Journal of Computational Finance*, vol. 2, pp. 49–74, (1998). [cited at p. 48]

[72] D. Gasbarra. Particle filters for counting process observations. http://www.rni.helsinki.fi/∼dag/newpart2.ps, Research report Helsinki University, (2001). [cited at p. -]

[73] V. Genon-Catalot, Th. Jeantheau and C. Laredo. Conditional Likelihood Estimators for Hidden Markov Models and Stochastic Volatility Models. *Scandinavian Journal of Statistics*, Vol 30: 297–316 (2003). [cited at p. 35]

[74] M. Giles and P. Glasserman, Smoking adjoints: fast Monte Carlo Greeks. *Risk*, pp. 92–96 (2006). [cited at p. 56]

[75] P. Glasserman, and B. Yu, *Number of paths versus number of basis functions in American option pricing*, Ann. Appl. Probab., 14, pp. 1-30 (2004). [cited at p. 83]

[76] E. Gobet, J. P. Lemor and X. Warin. A regression-based Monte-Carlo method for backward stochastic differential equations. *Annals Applied Probability*, 15, pp. 2172–2202 (2005). [cited at p. 65, 82, 85, 91]

[77] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Lecture Notes in Math., No. 1730, Springer (2000). [cited at p. -]

[78] J.H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975. [cited at p. -]

[79] M. S. Johannes, N. G. Polson and J. R. Stroud. Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices. *Review of Financial Studies*, 22, Issue. 7, pp. 2759–2799 (2009). [cited at p. 36]

[80] I. Karatzas and S. Shreve. *Methods of Mathematical Finance*. Springer (1998). [cited at p. 36]

[81] A. Kohatsu-Higa and M. Montero. *Malliavin Calculus in Finance*. Handbook of Computational and Numerical Methods in Finance, Birkhauser, pp. 111–174 (2004). [cited at p. 53]

[82] R. Lagnado and S. Osher. A technique for calibrating derivative security pricing models: numerical solution of the inverse problem. *Journal of computational finance*, vol. 1, pp. 13–25 (1997). [cited at p. 72]

[83] J. M. Lasry and P. L. Lions. Contrôle stochastique avec informations partielles et applications à la Finance. Comptes Rendus de l'Académie des Sciences - Series I - Mathematics, vol. 328, issue 11, pp. 1003–1010 (1999). [cited at p. 66]

[84] F. Le Gland and N. Oudjane. A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic Processes and their Applications*, vol. 106, no. 2, pp. 279–316 (2003). [cited at p. -]

[85] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Annals Applied Probability*, Vol. 14, no. 1, 144-187 (2004). [cited at p. 65, 81]

[86] A. E. B. Lim. Mean variance hedging when there are jumps. *SIAM J. Control Optim.*. vol. 44, no.5, pp. 1893–1922 (2005). [cited at p. -]

[87] G. Liu and L. J. Hong. Revisit of stochastic mesh method for pricing American options. Operations Research Letters. 37(6), 411–414 (2009). [cited at p. 64, 78]

[88] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies* 14, pp. 113–147 (2001). [cited at p. 64]

[89] D. Lefèvre. An introduction to Utility Maximization with Partial Observation. *Finance*, vol. 23, (2002). [cited at p. 71]

[90] F. A. Matsen and J. Wakeley. Convergence to the Island-Model Coalescent Process in Populations With Restricted Migration. *Genetics*, vol. 172, pp. 701-708 (2006). [cited at p. 49]

[91] L. Mazliak. Approximation of a Partially Observable Stochastic Control Problem. *Markov Processes and Related Fields*, vol.5, pp. 477–486, (1999). [cited at p. 71]

[92] G. Pagès. A space vector quantization method for numerical integration. *Journal of Comput. Appl. Math.*, 89, pp. 1–38 (1997). [cited at p. 64, 65, 82, 91, 110]

[93] G. Pagès and B. Wilbertz. Optimal quantization methods for pricing American style options. *Numerical Methods in finance, Springer* (2011). [cited at p. 64]

[94] G. Pagès and H. Pham. Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli*, vol.11, pp. 893–932 (2005). [cited at p. 64]

[95] G. Pagès, H. Pham and J. Printems. Optimal quantization methods and applications to numerical problems in finance. *in Handbook of Computational and Numerical Methods in Finance.* Ed. S. T. Rachev, *Birkhuser, Boston*, 2004, pp. 253–297 (2004). [cited at p. 64, 91]

[96] G. Pagès and J. Printems. Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods and Appl.*, 11(4), pp. 407–446 (2005). [cited at p. 64, 91]

[97] N. Papageorgiou, Br. Rémillard and A. Hocquard . Replicating the properties of hedge fund returns. *Journal of Alternative Invesments*, vol. 11, pp. 8–38 (2008). [cited at p. -]

[98] H. Pham, W. Runggaldier and A. Sellami. Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo Methods and Applications*. Volume 11, Issue 1, pp. 57–81 (2005). [cited at p. 66]

[99] H. Pham. Mean-variance hedging under partial observation. *International Journal of Theoretical and Applied Finance*, 4, pp. 263–284 (2001). [cited at p. 71]

[100] H. Pham and M. C. Quenez. Optimal portfolio in partially observed stochastic volatility models. *Annals of Applied Probability*, 11, pp. 210–238 (2001). [cited at p. 66]

[101] H. Pham, M. Corsi, and W. Runggaldier, Numerical Approximation by Quantization of Control Problems in Finance Under Partial Observations. *Handbook of Numerical Analysis*, vol. 15, pp. 325-360 (2009). [cited at p. 66]

[102] B. R. Rambharat and A. E. Brockwell. Sequential Monte Carlo pricing of American-style options under stochastic volatility models. *The Annals of Applied Statistics*, vol. 4, no. 1, pp. 222–265 (2010). [cited at p. -]

[103] Br. Rémillard, A. Hocquard, H. Langlois and N. Papageorgiou. Optimal hedging of American options in discrete time. *Numerical Methods in finance, Springer* (2011). [cited at p. -]

[104] Br. Rémillard and S. Rubenthaler. Optimal hedging in discrete and continuous time. Technical Report G-2009-77, Cahiers du Gerad (2009). [cited at p. -]

[105] V. Rossi and J. P. Vila. Nonlinear filtering in discrete time : A particle convolution approach. Ann. I.SU.P., vol.50, no. 3, pp. 71–102 (2006). [cited at p. 36]

[106] W. Runggaldier and L. Stettner. Approximations of Discrete Time Partially Observed Control Problems Applied Mathematics Monographs CNR, Giardini Editori, Pisa (1994). [cited at p. 69, 71]

[107] M. Schweizer. Variance-optimal hedging in discrete time. *Math. Oper. Res.*, vol. 20, no. 1, pp. 1–32 (1995). [cited at p. -]

[108] J. N. Tsitsiklis and B. Van Roy. Regression Methods for Pricing Complex American-Style Options. *IEEE Transactions on Neural Networks*, Vol. 12, No. 4 (special issue on computational finance), pp. 694-703 (2001). [cited at p. 64]

[109] R. Van Handel. Uniform time average consistency of Monte Carlo particle filters, *Stoch. Proc. Appl.*, 119, pp. 3835–3861 (2009). [cited at p. 65, 81]

[110] J. Vanneste. Estimating generalized Lyapunov exponents for products of random matrices. *Phys. Rev. E*, vol. 81, 036701 (2010). [cited at p. 61]

[111] F. Viens. Portfolio optimization under partially observed stochastic volatility. *Proceedings of the 8th international conference on advances in communication and control : telecommunications / signal processing*, pp. 3–12, (2001). [cited at p. 66]

[112] D. Whitley, S. Rana and R.B. Heckendorn. The island Model Genetic algorithm: On separability, population size and convergence. *CIT. Journal of computing and information technology.* vol. 7, no. 1, pp. 33-47 (1999). [cited at p. 49]