HAL
open science

Fractional regularity and stochastic analysis of discretizations; Adaptive simulation algorithm in credit risk

Azmi Makhlouf

To cite this version:

Azmi Makhlouf. Fractional regularity and stochastic analysis of discretizations; Adaptive simulation algorithm in credit risk. Mathematics [math]. Institut National Polytechnique de Grenoble - INPG, 2009. English. NNT : . tel-00460269

HAL Id: tel-00460269

https://theses.hal.science/tel-00460269v1

Submitted on 26 Feb 2010

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

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

INSTITUT POLYTECHNIQUE DE GRENOBLE

N° attribué par la bibliothèque
|__|__|__|__|__|__|__|__|__|__|

T H E S E

pour obtenir le grade de

DOCTEUR de L'Institut Polytechnique de Grenoble

Spécialité : Mathématiques Appliquées

préparée au **laboratoire** LJK (Laboratoire Jean Kuntzmann)

dans le cadre de **l'Ecole Doctorale** « Mathématiques, Sciences et Technologies de l'Information, Informatique »

présentée et soutenue publiquement par

Azmi MAKHLOUF

le 27 Novembre 2009

Régularité fractionnaire et analyse stochastique de discrétisations ; Algorithme adaptatif de simulation en risque de crédit

DIRECTEUR DE THESE : Emmanuel GOBET

JURY

M.	Anestis ANTONIADIS,	Président
M.	Bruno BOUCHARD,	Rapporteur
M.	Stefan GEISS,	Rapporteur
M.	Emmanuel GOBET,	Directeur de thèse
M.	Philippe BRIAND,	Examinateur
M.	Saïd HAMADENE,	Examinateur
Mme	Marie-Claire QUENEZ,	Examinatrice

Titre

Régularité fractionnaire et analyse stochastique de discrétisations; Algorithme adaptatif de simulation en risque de crédit.

Résumé

Cette thèse concerne trois sujets de probabilités numériques et de mathématiques financières. D'abord, nous étudions le module de régularité $\mathbf{L}_2$ en temps de la composante Z d'une EDSR markovienne à coefficients lipschitziens, mais dont la fonction terminale g est irrégulière. Ce module est lié à l'erreur d'approximation par schéma d'Euler. Nous montrons, de façon optimale, que l'ordre de convergence est explicitement lié à la régularité fractionnaire de g. Ensuite, nous proposons une méthode de Monte-Carlo séquentielle pour le calcul efficace du prix d'une tranche de CDO, basée sur des variables de contrôle séquentielles, dans un cadre où les taux de recouvrement sont aléatoires et i.i.d. Enfin, nous analysons l'erreur de couverture associée à la stratégie en Delta-Gamma. La régularité fractionnaire de la fonction payoff joue un rôle crucial dans le choix des dates de rebalancement, afin d'atteindre des vitesses de convergence optimales.

Mots-clés

Probabilités numériques, mathématiques financières, équations différentielles stochastiques rétrogrades, régularité fractionnaire, Monte-Carlo séquentiel, erreur de couverture

---

Title

Fractional regularity and stochastic analysis of discretizations; Adaptive simulation algorithm in credit risk.

Abstract

This thesis deals with three issues from numerical probability and mathematical finance. First, we study the $\mathbf{L}_2$-time regularity modulus of the Z-component of a Markovian BSDE with Lipschitz-continuous coefficients, but with irregular terminal function g. This modulus is linked to the approximation error of the Euler scheme. We show, in an optimal way, that the order of convergence is explicitly connected to the fractional regularity of g. Second, we propose a sequential Monte-Carlo method in order to efficiently compute the price of a CDO tranche, based on sequential control variates. The recoveries are supposed to be i.i.d. random variables. Third, we analyze the tracking error related to the Delta-Gamma hedging strategy. The fractional regularity of the payoff function plays a crucial role in the choice of the trading dates, in order to achieve optimal rates of convergence.

Keywords

Numerical probability, mathematical finance, backward stochastic differential equations, fractional regularity, sequential control variates, tracking error

Contents

1 Introduction (in English) 7

1.1 $\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions 7

1.1.1 What is a BSDE ? 7

1.1.2 Framework and challenges 8

1.1.3 Main results and tools 10

1.2 Efficient evaluation of CDOs with stochastic recoveries 13

1.2.1 Modeling and valuation of Collateralized Debt Obligations (CDOs) 13

1.2.2 Our model 14

1.2.3 Our contribution 15

1.2.4 The price function 15

1.2.5 Computation of the price function using SCV 16

1.2.6 Numerical results 18

1.3 Delta-Gamma hedging 19

1.3.1 Delta hedging 19

1.3.2 Delta-Gamma hedging 19

1.3.3 Framework 20

1.3.4 Main results and tools 21

2 Introduction (en français) 25

2.1 Régularité $\mathbf{L}_2$ en temps d'EDSRs à fonction terminale irrégulière . . 25

2.1.1 Qu'est ce qu'une EDSR? 25

2.1.2 Cadre et défis 26

2.1.3 Principaux résultats et outils 28

2.2 Calcul efficace des prix de tranches de CDO 31

2.2.1 Modélisation et pricing des CDO 31

2.2.2 Notre modèle 32

2.2.3 Notre apport 33

2.2.4 La fonction prix 34

2.2.5 Calcul de la fonction prix en utilisant VCS 34

2.2.6 Résultats numériques 37

2.3 La couverture en Delta-Gamma 37

2.3.1 La couverture en Delta 37

2.3.2 La couverture en Delta-Gamma 38

2.3.3 Cadre 38
2.3.4 Principaux résultats et outils 39

I $\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions 43

3 Background results for BSDEs and motivations 45
3.1 Existence and uniqueness results 45
3.2 Applications 47
3.2.1 Finance 47
3.2.2 Semilinear PDEs 48
3.3 $\mathbf{L}_2$-time regularity of Z 50
3.4 Application to time discretization of BSDEs 51
3.4.1 The key role of the $\mathbf{L}_2$-time regularity of Z 52
3.4.2 Other related results 56
3.5 Numerical computations of the dynamic programming equation (3.9) 56
3.6 Approximation of the terminal condition 58

4 Fractional regularity 61
4.1 First results on irregular terminal conditions 61
4.2 Generalization using fractional regularity 62
4.3 Other related works on sensitivities involving fractional regularity . . 65
4.4 Conclusion 66

5 $\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions 67
5.1 Abstract 67
5.2 Framework 67
5.3 A brief account on numerical methods for BSDEs 68
5.4 Known results on the $\mathbf{L}_2$-regularity of Z 69
5.5 Summary of our results 69
5.6 Preliminaries 70
5.7 The BSDE with null generator ($f = 0$) and bounded terminal condition g 73
5.7.1 Main results 73
5.7.2 Optimality of the time net 79
5.7.3 Proofs 80
5.8 A representation and an estimate of $Z_t - z_t$ when the terminal condition g is bounded and belongs to $\mathbf{L}_{2,\alpha}$ 84
5.8.1 The main result 84
5.8.2 Proof of Theorem 5.8.1 86
5.9 $\mathbf{L}_2$-regularity of Z_t when the terminal condition $g \in \mathbf{L}_{2,\alpha}$, but is not necessarily bounded 92
5.9.1 The main results 92
5.9.2 Proof of Theorem 5.9.1 94

5.9.3 Proof of Theorem 5.9.2 97
5.A Appendix 97
5.A.1 Malliavin weights expressions 97
5.A.2 An estimate for linear BSDEs 98

II Efficient evaluation of CDOs with stochastic recoveries 99

6 A brief account on CDO pricing 101
6.1 Collateralized Debt Obligations (CDOs) 101
6.2 The main loss modeling approaches 102
6.3 Numerical valuation of CDOs 102
6.3.1 In bottom-up models 102
6.3.2 In top-down models 103
6.3.3 Our contribution 104

7 Sequential control variates for linear Markov processes 105
7.1 Introduction 105
7.2 Framework 106
7.3 Sequential control variates 106

8 Valuation of CDOs with stochastic recoveries using sequential control variates 109
8.1 The model 109
8.1.1 The portfolio loss and default processes 109
8.1.2 The choice of the parameters 110
8.2 The price function 113
8.3 Computation of the price function using SCV 115
8.3.1 Notations 115
8.3.2 Interpolation operator and points 116
8.3.3 Algorithm 117
8.3.4 Simulation of the loss and the defaults proportion processes . 118
8.3.5 Computation of $\mathcal{T}_{i,j,k}$ and $\mathcal{AT}_{i,j,k}$ 120
8.4 Numerical results 121
8.4.1 Preliminary tests 121
8.4.2 Numerical results for CDO tranches pricing 123

III Delta-Gamma hedging 129

9 Delta and Delta-Gamma hedging errors 131
9.1 Delta hedging 131
9.1.1 The $\mathbf{L}_2$-tracking error 131
9.1.2 The weak convergence 134
9.1.3 Transaction costs 135

9.2 Delta-Gamma hedging 136
9.3 Our contribution 138

10 The tracking error rate of the Delta-Gamma hedging strategy 139
10.1 Abstract 139
10.2 Statement of the problem 139
10.3 Objectives 141
10.4 Organization of the paper 142
10.5 Preliminaries 142
10.5.1 Asset stochastic model, payoff, additional hedging instruments 142
10.5.2 Greeks and martingales 145
10.5.3 Other notations 146
10.6 The Delta-Gamma hedging strategy 147
10.6.1 Decomposition of the tracking error as multiple Itô integrals . 147
10.6.2 Main results 150
10.7 Numerical Results 153
10.8 Proofs 157
10.8.1 Proof of Theorem 10.6.1. 157
10.8.2 Proof of Proposition 10.6.1. 159
10.A Appendix 167
10.A.1 Margrabe formula 167
10.A.2 Proof of Lemma 10.5.1 168

Bibliography 171

Chapter 1

Introduction (in English)

This thesis consists of three parts dealing with three issues from numerical probability and mathematical finance : $\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions, efficient computation of CDO tranches prices and the tracking error related to Delta-Gamma hedging strategy.

1.1 $\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions

1.1.1 What is a BSDE ?

In few words, a backward stochastic differential equation (BSDE for short) is a kind of stochastic differential equation (SDE) for which the terminal value is specified instead of the initial one. The main point is to select a solution that is adapted to the underlying filtration.
BSDEs have been first introduced in 1973 by Bismut in [Bis73] for the linear case, in order to give a probabilistic interpretation of Pontryagin maximum principle, in an optimal stochastic control framework.
In 1990, Pardoux and Peng gave in [PP90] the first general result: they proved an existence and uniqueness result for nonlinear BSDEs, defined by

$$\begin{cases} -dY_t &= f(t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T &= \xi, \end{cases} \tag{1.1}$$

under the following assumptions: W is a standard Brownian motion, f (a random *generator* or *driver*) is Lipschitz continuous in the space variables y and z, and both the *terminal condition* ξ and the process $\{f(t,0,0)\}_{t\in[0,T]}$ are square integrable. The solution is a pair of adapted processes (Y, Z) such that Y is continuous, Z is predictable, and

$$\mathbb{E}\big[\sup_{t\in[0,T]} |Y_t|^2\big] + \mathbb{E}\big[\int_0^T |Z_s|^2 ds\big] < \infty.$$

Here the filtration is the augmented natural filtration of the Brownian motion W. Notice that, in the case where $f = 0$, the solution (Y, Z) is simply given by $Y_t = \mathbb{E}(\xi|\mathcal{F}_t)$, and Z is given by the predictable representation theorem.
Later, existence and uniqueness results have been developed under weaker assumptions and in different spaces (see Chapter 3, section 3.1). In 1997, El Karoui, Peng and Quenez [KPQ97b] have studied some other related subjects, namely the comparison theorem (in the one-dimensional case), concave BSDEs, continuity and differentiablity of Markovian BSDEs and differentiation of BSDEs in Malliavin sense. Besides, they have described some applications of BSDEs in finance. For instance, in a complete market, and within the Black-Scholes model, the price process of a positive contingent claim ξ, with maturity time T, solves a linear BSDE (i.e. the driver f is linear w.r.t. y and z). Non-linear BSDEs appear when the trading strategy is subject to additional constraints, or in incomplete market models. Recursive utilities can be also represented by BSDEs.
BSDEs have other applications in stochastic zero-sum differential games (see [HL95]), and in the switching problem (called also the starting and stopping problem), which involves reflected BSDEs (RBSDEs for short), see [HJ07]. RBSDEs have been introduced by El Karoui *et al.* in [KKP+97], and their development has been motivated by pricing American contingent claims (see [KPQ97a]).
Besides, a BSDE can be viewed as a stochastic representation of a non-linear parabolic PDE (see paragraph 1.1.2).

1.1.2 Framework and challenges

We will focus in our work on Markovian BSDEs with a multidimensional forward component and a Lipschitz-continuous driver:

$$\begin{cases} X_0 &= x_0, \\ dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ -dY_t &= f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T &= g(X_T), \end{cases} \tag{1.2}$$

where $g(X_T) \in \mathbf{L}_2$ and W is a standard q-dimensional Brownian motion. The $\mathbb{R}^d$-valued process X_t is the *forward component*, and $(Y_t, Z_t) \in \mathbb{R} \times \mathbb{R}^{1\times q}$ is the *backward component*. The functions b, σ, f and g are deterministic.
Just as SDEs are connected to linear parabolic PDEs via the Feynman-Kac formula, BSDE (1.2) is linked to a semilinear parabolic PDE: if, for fixed (t, x), one defines $(Y_s^{t,x}, Z_s^{t,x})_{s\in[t,T]}$ to be the solution of

$$\begin{cases} X_s^{t,x} &= x, \quad 0 \le s \le t, \\ \mathrm{d}X_s^{t,x} &= b(s, X_s^{t,x})\mathrm{d}s + \sigma(s, X_s^{t,x})\mathrm{d}W_s, \\ -\mathrm{d}Y_s^{t,x} &= f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\mathrm{d}s - Z_s^{t,x}\mathrm{d}W_s, \\ Y_T^{t,x} &= g(X_T^{t,x}), \end{cases} \tag{1.3}$$

and u to be the solution (in some sense) of

$$\begin{cases} \partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),\nabla_x u(t,x)\sigma(t,x)) = 0 \text{ for } t < T, \\ u(T,x) = g(x), \end{cases} \tag{1.4}$$

then

$$(Y_s^{t,x}, Z_s^{t,x}) = (u(s, X_s^{t,x}), \nabla_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x})).$$

In particular,

$$(u(t,x), \nabla_x u(t,x)\sigma(t,x)) = (Y_t^{t,x}, Z_t^{t,x}).$$

For more details, see paragraph 3.2.2 of Chapter 3.
Therefore, this part of the thesis lies within the framework of the probabilistic methods for solving non-linear PDEs. Indeed, the challenge behind the stochastic representation of PDEs (suitable for Monte-Carlo methods) is to remedy the curse of (high) dimensionality from which the deterministic methods suffer.
More precisely, we are interested in some discretization error associated to the BSDE (1.2). It is what we will call the $\mathbf{L}_2$-*time regularity* of the Z-component, and is defined by

$$\mathcal{E}(Z,\pi) := \sum_{i=0}^{N-1} \mathbb{E} \int_{t_i}^{t_{i+1}} |Z_t - \bar{Z}_{t_i}|^2 dt,$$

where $\bar{Z}_{t_i} = \frac{1}{t_{i+1}-t_i}\mathbb{E}^{\mathcal{F}_{t_i}} \int_{t_i}^{t_{i+1}} Z_s ds$, and $\pi = \{0 = t_0 < \cdots < t_i < \cdots < t_N = T\}$ is a given (deterministic) partition of the time interval $[0,T]$ with a fixed number of subintervals (equal to N), and with size $|\pi| := \sup_{0\le i<N}(t_{i+1} - t_i)$. $\mathcal{E}(Z,\pi)$ is the square of the $\mathbf{L}_2$-distance between Z and its best adapted $\mathbf{L}_2$-piecewise constant approximation.
The quantity $\mathcal{E}(Z,\pi)$ appears in the study of the rate of convergence of some numerical schemes used in solving the BSDE (1.2), in particular the one defined by the following *dynamic programming equation* (for $i < N$), in explicit form,

$$\begin{cases} Y_{t_i}^\pi &= \mathbb{E}^{\mathcal{F}_{t_i}}(Y_{t_{i+1}}^\pi + (t_{i+1} - t_i) f(t_i, X_{t_i}^\pi, Y_{t_{i+1}}^\pi, Z_{t_i}^\pi)), \\ Z_{t_i}^\pi &= \frac{1}{(t_{i+1}-t_i)}\mathbb{E}^{\mathcal{F}_{t_i}}(Y_{t_{i+1}}^\pi (W_{t_{i+1}} - W_{t_i})^*), \end{cases} \tag{1.5}$$

and $Y_{t_N}^\pi = g(X_T^\pi)$ (* denotes the transposition). The forward component X is approximated by X^π (using an Euler scheme for example).
Indeed, the order of convergence of (1.5) reduces to that of $\mathcal{E}(Z,\pi)$, up to the approximation error of the terminal condition and to the time net size $|\pi|$ (see paragraph 3.4.1 of Chapter 3). In other terms, obtaining a good rate for $\mathcal{E}(Z,\pi)$ implies a good approximation of the BSDE (1.2).
It is well known that $\mathcal{E}(Z,\pi)$ is of order $|\pi|$ when the terminal function g is Lipschitz-continuous ([Zha04]). Our goal is to study the rate of convergence of $\mathcal{E}(Z,\pi)$ when g is irregular. This is the case, for instance, where g is the indicator function of a domain (corresponding to binary options in finance). The rate of convergence has to be expressed according to some fractional regularity of the terminal function g and to the time mesh size $|\pi|$.

1.1.3 Main results and tools

Inspired by the works of Geiss *et al.* ([GG04][GH07]) and Gobet and Munos ([GM05b]), we define the fractional regularity of g by the order $\alpha \in (0,1]$ of decreasing of the expected conditional variance of $g(X_T)$ given X_t as t goes to T. We introduce the space $\mathbf{L}_{2,\alpha}$, which is a kind of stochastic equivalent of the fractional Sobolev space, defined by

$$\mathbf{L}_{2,\alpha} := \{g \text{ s.t. } \mathbb{E}(g(X_T)^2) + \sup_{0\le t<T} \frac{V_{t,T}(g)}{(T-t)^\alpha} < +\infty\}, \tag{1.6}$$

equipped with the squared norm

$$K^\alpha(g) := \mathbb{E}|g(X_T)|^2 + \sup_{t\in[0,T)} \frac{V_{t,T}(g)}{(T-t)^\alpha}, \tag{1.7}$$

where

$$V_{t,T}(g) := \mathbb{E}\left|g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T))\right|^2. \tag{1.8}$$

We will make regularity and boundedness assumptions for the coefficients b and σ, as well as the uniform ellipticity assumption for σ.
Our proofs rely on stochastic analysis techniques combining PDEs, martingales, Itô calculus and BSDEs in $\mathbf{L}_p$ ($p \in (1,2]$).
We mention that usually with these tools, g is supposed to have a polynomial growth, ensuring that $g(X_T)$ is in any $\mathbf{L}_p$ for $p > 0$. Here we stress the fact that we only assume $g(X_T) \in \mathbf{L}_2$ which is the minimal condition to discuss the existence and uniqueness of the solution of (1.2) in $\mathbf{L}_2$ spaces.
Moreover, our main results hold for $g \in \bigcup_{\alpha\in(0,1]} \mathbf{L}_{2,\alpha}$. We emphasize that this space is very large: it contains not only regular functions like the Hölder-continuous ones, but also irregular functions. For instance, the indicator function of a smooth domain belongs to $\mathbf{L}_{2,\frac{1}{2}}$.
We first study the case of null driver ($f \equiv 0$), for which the Z-component is denoted by z. We establish the link between $\mathcal{E}(z,\pi)$ and the second derivative of the (linear) PDE solution $u(t,x) := \mathbb{E}(g(X_T)|X_t = x)$ by showing that

$$\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2\, ds \le \frac{C}{N} + C\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} (t_{k+1}-r)\mathbb{E}|D^2u(r,X_r)|^2 dr$$

(and we have equivalence if $(X_t)_t$ is the Brownian motion).
We characterize the regularity of g in terms of some $\mathbf{L}_2$-estimates of the derivatives of u:

Lemma 1.1.1. *Let $\alpha \in (0,1]$. Then the three following assertions are equivalent:*
(i) $g \in \mathbf{L}_{2,\alpha}$.
(ii) $\exists C^\alpha(g) > 0$, such that, $\forall t \in [0,T)$,

$$\int_0^t \mathbb{E}\left|D^2u(s,X_s)\right|^2 ds \le \frac{C^\alpha(g)}{(T-t)^{1-\alpha}}.$$

(iii) $\exists C^\alpha(g) > 0$, *such that,* $\forall t \in [0,T)$,

$$\mathbb{E}\left|\nabla_x u(t,X_t)\right|^2 \leq \frac{C^\alpha(g)}{(T-t)^{1-\alpha}}.$$

And, if $g \in \mathbf{L}_{2,\alpha}$, *one can take* $C^\alpha(g) = CK^\alpha(g)$ *in (ii) and (iii).*
If $\alpha < 1$ *(resp.* $\alpha = 1$*), the previous three assertions are also equivalent to (resp. lead to) the following one:*
(iv) $\exists C^\alpha(g) > 0$, *such that,* $\forall t \in [0,T)$,

$$\mathbb{E}\left|D^2 u(t,X_t)\right|^2 \leq \frac{C^\alpha(g)}{(T-t)^{2-\alpha}}$$

(and one can take $C^\alpha(g) = CK^\alpha(g)$*).*

Thanks to these equivalences, Lemma 1.1.1 is crucial (although its proof is not technical) to show the optimality of our results. It states that the assumption "$g \in \mathbf{L}_{2,\alpha}$" is natural in our framework, if we want a rate of convergence for $\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2\, ds$ which is polynomial with respect to the time-step $|\pi|$. Indeed, Geiss and Hujo [GH07] (Theorem 1.3 and Theorem 2.1) showed that, if $\hat{X}$ is the Brownian motion or the geometric Brownian motion, then the following assertions are equivalent ($\alpha \in (0,1)$):
(GH-i) $\exists C > 0, \forall t \in [0,T)$,

$$\mathbb{E}\left|\nabla_x u(t,\hat{X}_t)\right|^2 \leq \frac{C}{(T-t)^{1-\alpha}}.$$

(GH-ii) $\exists C > 0$ such that, for all time-nets $\pi = (t_k)_{k=0\ldots N}$,

$$\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2\, ds \leq \frac{C}{N^\alpha}.$$

And Lemma 1.1.1 shows that the assertion (GH-i) is equivalent to the assertion "$g \in \mathbf{L}_{2,\alpha}$" (even for $\alpha = 1$). For related works with one-dimensional time-homogeneous diffusions, see [GG04].
Owing to Lemma 1.1.1, we prove that $\mathcal{E}(Z,\pi)$ is of order $|\pi|^\alpha$ when π is the uniform grid.
However, for non-equidistant time nets, we state the following theorem.

Theorem 1.1.1. *Assume that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0,1]$.
Take $\beta = 1$, *if* $\alpha = 1$, *and* $\beta < \alpha$ *otherwise. Then,* $\exists C > 0$ *such that, for any time net* $\pi = \{t_k, k = 0\ldots N\}$,

$$\mathcal{E}(z,\pi) \leq C|\pi| + C \sup_{k=0\ldots N-1} \left(\frac{t_{k+1}-t_k}{(T-t_k)^{1-\beta}}\right).$$

Now, we state that one can reach the optimal bound $\frac{C}{N}$ by using the time net defined for $\beta \in (0,1]$ by

$$\pi^{(\beta)} := \{t_k^{(N,\beta)} := T - T\big(1 - \frac{k}{N}\big)^{\frac{1}{\beta}}, 0 \leq k \leq N\} \tag{1.9}$$

(note that $\pi^{(1)} = (t_k^{(N,1)})$ coincides with the equidistant net, and that, for $\beta < 1$, the points in $\pi^{(\beta)}$ are more concentrated near T).

Corollary 1.1.1. *Assume that $g \in \mathbf{L}_{2,\alpha}$, for some $\alpha \in (0,1]$. Let β be as in Theorem 1.1.1. Then,*

$$\mathcal{E}(z, \pi^{(\beta)}) = \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k^{(N,\beta)}}^{t_{k+1}^{(N,\beta)}} \left| z_s - \bar{z}_{t_k^{(N,\beta)}} \right|^2 ds = \mathcal{O}(\frac{1}{N}).$$

We mention that, with the equidistant time net, the order $\frac{1}{N}$ cannot be achieved for some functions $g \in \mathbf{L}_{2,\alpha}$ with $\alpha < 1$; the optimal order in this case is equal to $\frac{1}{N^\alpha}$, as showed in [GT01] in some examples.
Then, for the non-null driver case, we show that $Z_t - z_t = U_t \sigma(t, X_t)$, where U is itself solution to a linear BSDE (see Chapter 5, section 5.8). The idea is to use the representation of the Z-component as a Malliavin derivative of the Y-component. However, because of the blow-up of the derivatives of u at time T, we proceed by time localization, and we state some technical stability lemmas. After estimating U, it turns out that the non-null driver case is just a perturbation of the null driver one, which means that $\mathcal{E}(Z, \pi)$ behaves like $\mathcal{E}(z, \pi)$ (up to the time net size):

Theorem 1.1.2. *Assume that $g \in \mathbf{L}_{2,\alpha}$, for some $\alpha \in (0,1]$. Then, there is a positive constant C such that, for any time net $\pi = \{t_k : k = 0...N\}$*

$$\mathcal{E}(Z, \pi) = \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left| Z_s - \bar{Z}_{t_k} \right|^2 ds \leq C\mathcal{E}(z, \pi) + C|\pi|.$$

This readily leads to our main result, which states that $\mathcal{E}(Z, \pi)$ is of order $\frac{1}{N^\alpha}$ if π is the equidistant time net, and $\frac{1}{N}$ if π is equal to a suitable $\pi^{(\beta)}$.

Theorem 1.1.3. *Assume that $g \in \mathbf{L}_{2,\alpha}$, for some $\alpha \in (0,1]$. Then,*

a) for the choice of the equidistant time net $\pi^{(1)}$,

$$\mathcal{E}(Z, \pi^{(1)}) = \mathcal{O}(\frac{1}{N^\alpha});$$

b) for the choice of $\pi^{(\beta)}$, with β as in Corollary 1.1.1,

$$\mathcal{E}(Z, \pi^{(\beta)}) = \mathcal{O}(\frac{1}{N}).$$

The significance of our results is threefold. First, we have succeeded, through a good choice of the time net, to obtain an order of convergence for $\mathcal{E}(Z, \pi)$ that is equal to $\frac{1}{N}$, in spite of the irregularity of the terminal function g. Second, this holds for the general framework of multidimensional forward component of a BSDE with Lipschitz-continuous driver. Third, the results are optimal.

1.2 Efficient evaluation of CDOs with stochastic recoveries

1.2.1 Modeling and valuation of Collateralized Debt Obligations (CDOs)

A *CDO tranche*, based on an underlying multi-name porfolio of assets (corporate bonds, bank loans, credit swaps, ...), is a bilateral contract consisting of a protection against the default losses in the portfolio, within an interval $[K_1, K_2]$, over some time period $[0, T]$. The (fixed) values K_1 and K_2 are respectively called the *lower attachment point* (or *attachment point*) and the *upper attachment point* (or *detachment point*) of the tranche.
When a name q of the underlying pool defaults, only a part of its notional ν_q can be recovered, and we denote by R_q the corresponding recovery rate (which may be non deterministic). Thus, the loss caused by the default of the name q, called the *loss given default*, is equal to $l_q := \nu_q(1 - R_q)$, and the total portfolio loss over a time period $[0, t]$ is equal to

$$L_t := \sum_q l_q \mathbb{1}_{\tau_q \leq t} = \sum_q \nu_q (1 - R_q) \mathbb{1}_{\tau_q \leq t},$$

where τ_q denotes the default time of the name q.
It is known that the valuation of the CDO tranche amounts to evaluating the *expected tranche notionals* $\mathbb{E}((L_T - K)^+)$, for different values of T and of K. For more details, see e.g. [Das05].
Two main approaches have been developed for the valuation of credit derivatives: the so-called *bottom-up* and *top-down* approaches (see [Gie08]).
In the bottom-up models, one starts from the description of each individual name dynamics (default time, default intensity, loss given default,...), as well as the possible dependencies between them; then, the accumulated loss dynamics implicitely follows. On the contrary, within the top-down approach, one first explicitely models the portfolio loss process (without reference to the underlying constituents), and then tries to come back up to the dynamics of the individual names. Note that, within both models, some *auxiliary factor processes* may be introduced in order to take into account some additional information.
Each of the two approaches has its advantages and drawbacks, depending on the purpose for which it is developed. In practice, the bottom-up models are the best suited for hedging issues, whereas the top-down models are the best suited for fast valuation and calibration. We refer to [Gie08] and [BCJ08] for more detailed comparison and critic of both approaches.

1.2.2 Our model

The portfolio loss and default processes

We consider a credit portfolio composed of n_0 names (that range from 50 to 150 in practice), and we assume that each name (firm) q may default at most once in the time interval $[0, T]$, at a random time τ_q.
We define the (normalized) *portfolio loss process* $(L_t)_{t\in[0,T]}$ by

$$L_t := \frac{1}{n_0}\sum_{q=1}^{n_0}(1 - R_q)\mathbb{1}_{\tau_q\leq t},$$

where the (stochastic) *recoveries* $(R_q)_{1\leq q\leq n_0}$ are i.i.d. random variables (also independent of the default times $(\tau_q)_q$) taking values in [0,1] (which implies that $L_t \in [0, 1]$). The names notionals (ν_q) are supposed to be equal.
In order to take into account a possible default contagion, we introduce, as a factor process, the *default process* $(N_t)_{t\in[0,T]}$, where N_t represents the number of defaults occuring before time t:

$$N_t := \sum_{q=1}^{n_0}\mathbb{1}_{\tau_q\leq t}.$$

The intensity process λ

We will assume that the compensator of the (non-decreasing and adapted) process N is time-differentiable. We denote its time derivative, which represents the *default intensity process*, by $(\lambda_t)_{t\in[0,T]}$ (this means that $(N_t - \int_0^t \lambda_s \mathrm{d}s)_t$ is a local martingale).
Different models for the intensity process λ have been suggested in the literature, attempting to take into account some contagion phenomena that may appear in the markets reality. The main models that express the intensity process λ_t as a function of N_t are the linear model ($\lambda_t := c_0 + c_1 N_t$, see Frey and Backhaus [FB08]) and the exponential model ($\lambda_t := c_0 e^{c_1 N_t}$, see Davis and Lo [DL01]).
We choose a kind of intermediate model between the linear description and the exponential one, that has been introduced by Frey and Backhaus in [FB08]. It is defined by

$$\lambda_t = \lambda(t, L_t, N_t) := c_0 + \frac{c_1}{c_2}\left(e^{c_2(N_t-\mu(t))^+} - 1\right),$$

where c_0, c_1, c_2 are positive constants, and μ represents the number of defaults one can *a priori* expect. This means that there is no contagion if the observed number of defaults is less than the expected one (if $N_t < \mu$, $\lambda_t = c_0$), otherwise, the contagion grows exponentially with respect to the number of defaults (in a catastrophic scenario, for example). Note also the flexibility of this model, since one can make it almost linear by choosing c_2 "small" (indeed, when $c_2 \simeq 0$, $\lambda_t \simeq c_0 + c_1(N_t - \mu)^+$).
In addition, we assume that $\lambda(t, l, n)$ vanishes as soon as $n \geq n_0$ (since each of the n_0 names is supposed to default at most once).

The recoveries $(R_q)_q$

In the literature devoted to the valuation and hedging of CDOs, the recoveries $(R_q)_q$ are often supposed to be deterministic, for the sake of simplicity, and equal to a constant (40%, for example). However, this assumption may not be realistic, and, as pointed out in [Ron04], the impact of the recovery rate on the defaults distribution may be very important. It is also showed in [DG01] that the recoveries distribution may differ from one type of debt to another.
In our study, the recoveries $(R_q)_q$ are i.i.d. random variables, also independent of the default times $(\tau_q)_q$, and we assume, as in [Ron04], that they follow a beta distribution on $(0,1)$, with parameters $a > 0$ and $b > 0$. The density of this distribution is defined by

$$\beta^{a,b}(z) = \frac{1}{B(a,b)} z^{a-1}(1-z)^{b-1}\mathbb{1}_{(0,1)}(z),$$

where $B(a,b) = \int_0^1 z^{a-1}(1-z)^{b-1}\mathrm{d}z = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, and Γ is the gamma function.
This provides a versatile family of distributions with bounded support and different shapes. Besides, the main advantage of taking a beta distribution for the recoveries is to be able to calibrate the parameters a and b from the knowledge of the mean $\mu := \mathbb{E}(R)$ and the variance $\sigma^2 := \mathbb{V}ar(R)$ of the recoveries.

1.2.3 Our contribution

Within a top-down Markovian approach, and allowing the recoveries to be non-deterministic (we only suppose that they are i.i.d random variables), we propose a sequential Monte-Carlo method in order to efficiently compute the CDO tranche price. We apply the *sequential control variates* (SCV for short) algorithm, introduced by Gobet and Maire [GM05a] for the computation of solutions of second order linear PDEs that have a Feynman-Kac-type representation (see Chapter 7).

1.2.4 The price function

We define the *defaults proportion process* $(P_t)_{t\in[0,T]}$ (taking values in $[0,1]$) by

$$P_t := \frac{N_t}{n_0} = \frac{1}{n_0}\sum_{q\geq 1}\mathbb{1}_{\tau_q\leq t}.$$

The defaults proportion process and the loss process started resp. from p and l at time t are denoted resp. by $(P_s^{t,l,p})_{t\leq s\leq T}$ and $(L_s^{t,l,p})_{t\leq s\leq T}$.
We denote by X the Markov process (L,P):

$$X_t := (L_t, P_t), \quad X_s^{t,l,p} := (L_s^{t,l,p}, P_s^{t,l,p}).$$

Let $\bar{D} := [0,T]\times[0,1]\times[0,1]$. We assume now that $\lambda_t = \lambda(t, L_t, P_t)$ with λ uniformly bounded on $\bar{D}$, and that $\lambda(t,l,p) = 0$ when $p > 1 - \frac{1}{n_0}$ (i.e. when $n > n_0 - 1$) or when $l > p$ (which is natural, since $L_t \leq P_t$).

For generic uniformly bounded functions $\tilde{f}$ and $\tilde{g}$, defined on $\bar{D}$, and for $f(t,l,p) := 0$ and $g(l,p) := (l-K)^+$, we set (for $(t,l,p) \in \bar{D}$)

$$\begin{aligned} \Psi(\tilde{f},\tilde{g},X^{t,l,p}) &:= \tilde{g}(X_T^{t,l,p}) - \int_t^T \tilde{f}(s,X_s^{t,l,p})\mathrm{d}s, \\ v(t,l,p) &:= \mathbb{E}[\Psi(\tilde{f},\tilde{g},X^{t,l,p})], \\ u(t,l,p) &:= \mathbb{E}[\Psi(f,g,X^{t,l,p})] = \mathbb{E}[(L_T-K)^+|L_t=l,P_t=p]. \end{aligned} \tag{1.10}$$

We show (see Theorem 8.2.1) that v is the *classic solution* (i.e. uniformly bounded w.r.t. (t,l,p) and differentiable w.r.t. to t) of

$$\begin{cases} \mathcal{A}v &= \tilde{f} \quad \text{when } t<T, \\ \mathcal{B}v &= \tilde{g} \quad \text{when } t=T, \end{cases} \tag{1.11}$$

where

$$\begin{aligned} \mathcal{A}v(t,l,p) &:= \partial_t v(t,l,p) + \lambda(t,l,p)\int_0^1 \left(v(t,l+\frac{y}{n_0},p+\frac{1}{n_0}) - v(t,l,p)\right)\beta^{b,a}(y)dy, \\ \mathcal{B}v(t,l,p) &:= v(t,l,p), \end{aligned}$$

and so is the price function u we aim to compute, with $\tilde{f}=f$ and $\tilde{g}=g$. This will enable us to use the SCV method in order to compute u.

1.2.5 Computation of the price function using SCV

The well-known Monte-Carlo method is especially convenient for high-dimensional problems, but it suffers from the slow convergence. Two major variance-reduction techniques have been used in order to minimize the convergence error: the *control variates* approach and the *importance sampling* technique (see e.g. Halton [Hal70] and Glasserman [Gla04]). Moreover, sequential Monte-Carlo algorithms, using iteratively the two previous techniques, have been developed (see, for instance, [Hal62] and [Hal70]).
Gobet and Maire have introduced in [GM04] the *sequential control variate* algorithm in order to solve the Poisson equation with Dirichlet boundary conditions over square domains. Then, in [GM05a], they have extended their approach to general Markov processes (including diffusion and jump processes) related to linear elliptic and parabolic second order PDEs, with linear boundary conditions.
The idea of the SCV algorithm is to approximate the solution u by a sequence $(u_n)_n$ such that, at the nth step,

$$u_n(x_i) = u_{n-1}(x_i) + \bar{y}_n(x_i),$$

where $(x_i)_{1\le i\le N}$ are some points of $\bar{D}$ (interpolation knots, for example), and $\bar{y}_n(x_i)$ is a Monte-Carlo evaluation of the correction term $y_n(x_i) := u(x_i) - u_{n-1}(x_i)$, represented as an expectation: indeed, from (1.10) and (1.11), and from the linearity of

the functional Ψ, one has

$$y_n(x) = \mathbb{E}[\Psi(f - \mathcal{A}u_{n-1}, g - \mathcal{B}u_{n-1}, X(x))].$$

A global approximation based on the values on the points $(x_i)_i$ is then obtained by

$$u_n = u_{n-1} + \mathcal{P}\bar{y}_n,$$

where $\mathcal{P}$ is a linear approximation operator (an interpolation operator, for example) based on the points $(x_i)_i$, i.e., for any function φ and for some basis functions $(\mathcal{C}_i)_i$,

$$\mathcal{P}\varphi(.) = \sum_{i=1}^{N} \varphi(x_i)\mathcal{C}_i(.).$$

The operator $\mathcal{P}$ is assumed to have the stability property:

$$\mathcal{P}(\mathcal{P}\varphi) = \mathcal{P}\varphi.$$

This property is satisfied by the interpolation and the projection operators.
Here is the algorithm.

Algorithm 1.2.1 (SCV)**.**

1. *Initialization:* $u_0 := 0$.
2. *At iteration n: for each x_i, simulate M independent paths $(X^{n,m}(x_i))_{1\leq m\leq M}$ of $X(x_i)$ (independently of anything else also), and set*

$$\bar{y}_n(x_i) := \frac{1}{M}\sum_{m=1}^{M} \Psi(f - \mathcal{A}u_{n-1}, g - \mathcal{B}u_{n-1}, X^{n,m}(x_i)).$$

3. *Always at iteration n: build the global solution using the linear approximation of the correction term:*

$$u_n := u_{n-1} + \mathcal{P}\bar{y}_n = u_{n-1} + \sum_{i=1}^{N} \bar{y}_n(x_i)\mathcal{C}_i.$$

We mention that, in their paper, Gobet and Maire have presented the general case where the process $X(x)$ is obtained by a discretization procedure $X^{\Delta}(x)$, with Δ being the discretization parameter (for example, the time step for the Euler procedure). In our framework, the simulation of $X = (L, P)$ is exact.
Besides, Gobet and Maire have studied the bias $m_n := \sup_{1\leq i\leq N} |\mathbb{E}(u_n(x_i) - u(x_i))|$ and the variance $v_n := \sup_{1\leq i\leq N} |\mathbb{V}ar(u_n(x_i))|$. They have mainly established that these two errors converge geometrically with the number of iterations of the algorithm, up to an error term related to the quality of approximation of u by the operator $\mathcal{P}$ (we do not detail the required assumptions on the basis functions $(\mathcal{C}_i)_i$):

Theorem 1.2.1. *(Theorems 3.1 and 3.2 in [GM05a])*

i)

$$m_n \leq \rho_m m_{n-1} + \mathcal{E}_m(\mathcal{P}u - u),$$

where $\rho_m < 1$ (and depends on Δ, N and $\mathcal{P}$) when Δ is small enough, and $\mathcal{E}_m(\mathcal{P}u - u)$ is some norm (depending on Δ and on N) related to the approximation error $\mathcal{P}u - u$. And, consequently,

$$\limsup_n m_n \leq \frac{1}{1-\rho_m}\mathcal{E}_m(\mathcal{P}u - u).$$

ii)

$$v_n \leq \rho_v v_{n-1} + \frac{1}{M}\mathcal{E}_v(\mathcal{P}u - u),$$

where $\rho_v < 1$ (and depends on Δ, N, M and $\mathcal{P}$) when Δ is small enough and M large enough, and $\mathcal{E}_v(\mathcal{P}u - u)$ is some norm (depending on Δ and on N) related to the approximation error $\mathcal{P}u - u$. And, consequently,

$$\limsup_n v_n \leq \frac{1}{1-\rho_v}\frac{1}{M}\mathcal{E}_v(\mathcal{P}u - u).$$

Remark 1.2.1. *If u belongs to the chosen approximation space, i.e. $\mathcal{P}u = u$, then $\mathcal{E}_m(\mathcal{P}u - u) = \mathcal{E}_v(\mathcal{P}u - u) = 0$, and the bias and the variance converge geometrically to 0 (when Δ is small enough and M is large enough).*

The above estimates show that we do not need to take M large to have an accurate evaluation, since small values of M may be compensated by small approximation error $\mathcal{P}u - u$. This is the key feature of SCV method.

1.2.6 Numerical results

We first give exact simulation methods for the process $X = (L, P)$ and for the recoveries (with beta distribution). For the interpolation procedure, we use a Gauss-Lobatto-Tchebychev interpolation w.r.t. each of the three variables t, l and p, in order to include the boundaries of the respective intervals $[0, T]$, $[0, 1]$ and $[0, 1]$ (since we want to compute the value of $u(0, 0, 0)$).
Through numerical results, we first show, with $K = 0$ and λ constant (which is a case where the exact solution is explicitely known), that the SCV algorithm turns to be about 10 times more efficient than the crude Monte-Carlo method. Then, within our general model, we show that SCV is faster and more accurate than the crude Monte-Carlo method when the attachment point K is not too much bigger than the average loss (see Chapter 8, Section 8.4).

1.3 Delta-Gamma hedging

1.3.1 Delta hedging

The contrast between the time-continuous nature of the models in finance and the obvious time-discrete nature of the transactions leads to an inevitable *tracking error* when trying to replicate a derivative security (with a self-financing strategy).
Within the Black-Scholes model, the well-known Delta hedging strategy (DHS for short) consists in holding a fixed number δ_{t_i} of assets S between two rebalancing times t_i and t_{i+1}, for $0 \leq i \leq N-1$, where N is the number of the time intervals. The quantity δ_{t_i} is equal to $\partial_S u(t_i, S_{t_i})$, where u is the price function. At the maturity time T, the terminal value $V_T^{\Delta,N}$ of the portfolio constructed in this way may differ from the payoff to hedge $g(S_T)$, inducing the tracking error $\mathcal{E}_N^\Delta := V_T^{\Delta,N} - g(S_T)$. Under the risk-neutral measure, the discounted trackning error $\bar{\mathcal{E}}_N^\Delta := e^{-rT}\mathcal{E}_N^\Delta$ is a martingale which writes

$$\bar{\mathcal{E}}_N^\Delta = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (\delta_{t_i} - \delta_t) \mathrm{d}\bar{S}_t, \tag{1.12}$$

where r is the (constant) interest rate and $\bar{S}_t := e^{-rT} S_t$ (which is also a martingale under the risk-neutral probability).
Zhang [Zha99] showed that, if g is Lipschitz-continuous, the $\mathbf{L}_2$ error $(\mathbb{E}|\mathcal{E}_N^\Delta|^2)^{\frac{1}{2}}$ has a convergence rate $N^{1/2}$ with uniform rebalancing dates $t_i = iT/N$. In Gobet and Temam [GT01], it is shown that the rate of convergence actually strongly depends on the smoothness of the payoff function g: for instance for digital options, the order of convergence becomes $1/4$ instead of $1/2$. Geiss and coauthors deeply investigated these features: in a series of papers initiated by Geiss [Gei02], they paved the way to connect the fractional regularity of the payoff and the rate of convergence of the related discrete time delta-hedging strategy. Moreover, they showed that for a given payoff, a suitable non-uniform grid with N dates can be chosen to achieve the rate $N^{1/2}$: the more irregular the payoff, the more concentrated near T the points. These ideas have been extended in this thesis (Part I) to BSDEs and to multidimensional diffusion models.

1.3.2 Delta-Gamma hedging

In all the above works (and also from expression (1.12)), one can see that estimating $\bar{\mathcal{E}}_N^\Delta$ amounts to estimating a suitable weighted integral of the so-called Gamma $(\partial_S^2 u(t, S_t))_{0 \leq t < T}$ of the option. Thus, in order to reduce the tracking error, one can think of reducing the *p*ortfolio Gamma by incorporating in the hedging portfolio another instrument, the price of which is $(C(t, S_t))_{0 \leq t \leq T}$: in practice, it is a liquid vanilla option (At-The-Money Call/Put). The resulting strategy is called Delta-Gamma hedging strategy (DGHS in short) and is well-known from practionners (see Hull [Hul09], p.371). It is obtained by equalizing the second sensitivity of the hedging portfolio and that of the option value. It follows that the numbers of options C

and assets to hold at time t_i are equal to (in the one-dimensional setting)

$$\delta_{t_i}^C := \frac{\partial_S^2 u(t_i, S_{t_i})}{\partial_S^2 C(t_i, S_{t_i})}, \qquad \delta_{t_i} := \partial_S u(t_i, S_{t_i}) - \frac{\partial_S^2 u(t_i, S_{t_i})}{\partial_S^2 C(t_i, S_{t_i})} \partial_S C(t_i, S_{t_i}).$$

This specification of piecewise-constant strategy on each interval $[t_i, t_{i+1}[$ leads to a final wealth process V_T^N hopefully closer to $g(S_T)$ compared with that of the simple DHS. The purpose of our work is to quantify how much the resulting tracking error $\mathcal{E}_N^{\Delta\Gamma} := V_T^N - g(S_T)$ is reduced with respect to the number of rebalancing dates. More precisely, we aim at connecting the $\mathbf{L}_2$-convergence rate of $\mathcal{E}_N^{\Delta\Gamma}$ to the payoff regularity, regarding the choice of rebalancing dates.

1.3.3 Framework

Within a multidimensional setting, we only consider geometric Brownian motions for the modeling of $S = (S^1, ..., S^d)$ (however, we believe that our results hold for a wider class of SDE model on S). It writes for $1 \leq j \leq d$:

$$\begin{cases} S_0^j &= s_0^j, \\ \mathrm{d}S_t^j &= \mu_j S_t^j \mathrm{d}t + \sigma_j S_t^j \mathrm{d}\hat{W}_t^j, \end{cases}$$

where S_t^j, s_0^j and σ_j are positive, and $\hat{W} = (\hat{W}^1, ..., \hat{W}^d)$ is a d-dimensional Brownian motion, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is the historical probability and $(\mathcal{F}_t)_{0 \leq t \leq T}$ is the natural filtration of $\hat{W}$, augmented with $\mathbb{P}$-null sets. We assume that the components of $\hat{W}$ are correlated, with $\langle \hat{W}^j, \hat{W}^k \rangle_t = \rho_{j,k} t$, and that the matrix $(\rho_{j,k})_{1 \leq j,k \leq d}$ has a full rank.
As hedging instruments, we propose to use Calls to handle the individual gammas and exchange options (spread option with strike zero) to handle the cross gammas, with common maturity time $T_2 > T$. The price function of each of these instruments writes

$$C^{j,k}(t, S^j, S^k) := \mathbb{E}_{\mathbb{Q}} \left[e^{-r(T_2 - t)} (S_{T_2}^k - K_{j,k} S_{T_2}^j)_+ | S_t^j = S^j, S_t^k = S^k \right],$$

where $\mathbb{Q}$ is the risk-neutral measure. Since the models for S are correlated geometric Brownian motions, one can take advantage of the closed Black-Scholes and Margrabe formulas for such options (see Appendix 10.A.1). The number of each of these intruments that has to be handled between times t_i and t_{i+1} will be denoted by $\delta_{t_i}^{j,k}$.
In the spirit of section 1.1, we want to use the fractional regularity of $g \in \mathbf{L}_{2,\alpha}$ (with $\mathbf{L}_{2,\alpha}$, $K^\alpha(g)$ and $V_{t,T}(g)$ defined by (1.6), (1.7) and (1.8)) in order to estimate the error $(\mathbb{E}|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g, \pi)|^2)^{1/2}$ and to find a time grid that achieves an optimal rate. Besides, we will suppose that $\mathbb{E}_{\mathbb{P}} |g(S_T)|^{2p_0} < \infty$ for some $p_0 > 1$, and we set

$$K^{\alpha, p_0}(g) := \left(\mathbb{E}_{\mathbb{P}} |g(S_T)|^{2p_0} \right)^{\frac{1}{p_0}} + \sup_{t \in [0,T)} \frac{V_{t,T}(g)}{(T-t)^\alpha}.$$

1.3.4 Main results and tools

A key tool in our work is to use the martingale property (under the risk-neutral measure) of the discounted price process (denoted by $\bar{u}(t)$) and its derivatives (denoted, up to multiplicative constants, by $\bar{u}_l^{(1)}(t)$ for the first derivative w.r.t. S^l, $\bar{u}_{l,m}^{(2)}(t)$ for the second derivative w.r.t. (S^l, S^m), and $\bar{u}_{l,m,n}^{(3)}(t)$ for the third derivative w.r.t. (S^l, S^m, S^n)).
We begin by decomposing the (discounted) tracking error $\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi)$ as a sum of triple stochastic integrals w.r.t. W, where $(W_t^j := \hat{W}_t^j + \lambda_j t)_{1\le j\le d}$ is a $\mathbb{Q}$-Brownian motion (the correlations remaining unchanged). This decomposition consists of a principal part, related to the third derivatives of the option price function, and some residual terms.

Theorem 1.3.1. *Assume that* $\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0} < \infty$ *for some* $p_0 > 1$. *We have*

$$\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi) = -\sum_{i=0}^{N-1}\sum_{l,m,n=1}^{d}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\left(\bar{u}_{l,m,n}^{(3)}(r) + R_{l,m,n}^{i,(3)}(r)\right)\mathrm{d}W_r^n\mathrm{d}W_s^m\mathrm{d}W_t^l, \tag{1.13}$$

where

$$\begin{aligned}
R_{l,m,n}^{i,(3)}(t) = &\left(\sigma_m\bar{u}_{l,m}^{(2)}(t)\mathbb{1}_{n=m} + \sigma_l\bar{u}_{l,m}^{(2)}(t)\mathbb{1}_{n=l} + \sigma_l\bar{u}_{l,n}^{(2)}(t)\mathbb{1}_{m=l}\right)\\
&+\sigma_l^2\left(\bar{u}_l^{(1)}(t) - \frac{\bar{S}_t^l}{\bar{S}_{t_i}^l}\bar{u}_l^{(1)}(t_i)\right)\mathbb{1}_{n=m=l}\\
&-\sum_{0\le j<k\le d}\delta_{t_i}^{j,k}\bar{C}_{l,m,n}^{j,k,(3)}(t)\\
&-\sum_{0\le j<k\le d}\delta_{t_i}^{j,k}\left(\sigma_m\bar{C}_{l,m}^{j,k,(2)}(t)\mathbb{1}_{n=m} + \sigma_l\bar{C}_{l,m}^{j,k,(2)}(t)\mathbb{1}_{n=l} + \sigma_l\bar{C}_{l,n}^{j,k,(2)}(t)\mathbb{1}_{m=l}\right)\\
&-\sigma_l^2\sum_{0\le j<k\le d}\delta_{t_i}^{j,k}\left(\bar{C}_l^{j,k,(1)}(t) - \frac{\bar{S}_t^l}{\bar{S}_{t_i}^l}\bar{C}_l^{j,k,(1)}(t_i)\right)\mathbb{1}_{n=m=l}.
\end{aligned} \tag{1.14}$$

Let us formally comment the consequences of (1.13). Set $T_{l,m,n}^{i,(3)}(r) := \bar{u}_{l,m,n}^{(3)}(r) + R_{l,m,n}^{i,(3)}(r)$. At first sight, if $T_{l,m,n}^{i,(3)}(r)$ had a $\mathbb{Q}$-$\mathbf{L}_2$ moment uniformly bounded in r, we would deduce that

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi)|^2 &\le \sum_{i=0}^{N-1}\sum_{l,m,n=1}^{d}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\sup_{0\le r\le T}|T_{l,m,n}^{i,(3)}(r)|_{\mathbf{L}_2}^2\mathrm{d}r\mathrm{d}s\mathrm{d}t\\
&\le C\sup_{i,l,m,n;\ 0\le r\le T}|T_{l,m,n}^{i,(3)}(r)|_{\mathbf{L}_2}^2|\pi|^2,
\end{aligned}$$

i.e. the tracking error would have a $\mathbb{Q}$-$\mathbf{L}_2$ norm of the order of the time step $|\pi|$. In particular, N regular trading dates would lead to an optimal rate of convergence.

Actually, this argumentation is not complete because the moments of $T^{i,(3)}_{l,m,n}(r)$ may explode as r goes to T (indeed, for non smooth payoff, the Greeks may go to $\pm\infty$ as the time to maturity $T-r$ shrinks to 0). It illustrates that a significant piece of work in the Delta-Gamma error analysis lies in the understanding of the behavior of $|T^{i,(3)}_{l,m,n}(r)|^2_{\mathbf{L}_2}$ as $r \to T$. This feature is tightly estimated through the fractional regularity of g (see Proposition 1.3.1 and Corollary 1.3.1).

Proposition 1.3.1. *The following estimates hold:*

- $$\mathbb{E}_{\mathbb{P}}\left|\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi)\right|^2 \leq C\sum_{i=0}^{N-1}\sum_{l,m,n=1}^{d}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(3)}_{l,m,n}(r)+R^{i,(3)}_{l,m,n}(r)\right|^2 \mathrm{d}r\mathrm{d}s\mathrm{d}t. \tag{1.15}$$

- *For $l,m,n = 1...d$ and $0 \leq t < T$,*
$$\mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(3)}_{l,m,n}(t)\right|^2 \leq C\frac{V_{t,T}(g)}{(T-t)^3}. \tag{1.16}$$

- *There exists a positive $\pi^{\text{threshold}}$ (depending on p_0, T and T_2) such that, if $|\pi| \leq \pi^{\text{threshold}}$, then, for $0 \leq t_i \leq t < t_{i+1} \leq T$,*
$$\mathbb{E}_{\mathbb{P}}\left|R^{i,(3)}_{l,m,n}(t)\right|^2 \leq C\frac{V_{t,T}(g)+\left(\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0}\right)^{\frac{1}{p_0}}}{(T-t)^2}. \tag{1.17}$$

A straightforward estimate is the following:

Corollary 1.3.1. *Let $\alpha \in (0,1]$. Suppose that $g \in \mathbf{L}_{2,\alpha}$ and that $\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0} < \infty$ for some $p_0 > 1$. Then, when $|\pi| \leq \pi^{\text{threshold}}$, and for $0 \leq t < T$,*
$$\mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(3)}_{l,m,n}(t)+R^{i,(3)}_{l,m,n}(t)\right|^2 \leq C\frac{K^{\alpha,p_0}(g)}{(T-t)^{3-\alpha}}. \tag{1.18}$$

In order to get the estimate (1.16), we use the usual representation of Greeks in terms of Malliavin calculus weights. One has just to express this representation under the historical probability $\mathbb{P}$.
However, the estimate (1.17) is more intricate. Indeed, one can see from the expression (1.14) that $\mathbf{L}_p$-estimates for quotients of the hedging instruments Greeks have to be made. If the numerator and the denominator are expressed at the same time point t, then their respective Gaussian exponentials parts cancel (using the Margrabe formula). The difficulty appears when the two times are different, as in $\frac{\bar{C}^{j,k,(2)}_{l,m}(t)}{\bar{C}^{j,k,(2)}_{l,m}(t_i)}$, for instance (with $t_i \leq t \leq t_{i+1}$). Indeed, if $t-t_i$ is too large, the quantity $\mathbb{E}_{\mathbb{P}}[|\bar{C}^{j,k,(2)}_{l,m}(t)|^p|\mathcal{F}_{t_i}]$ becomes too heavy-tailed in comparison with the denominator

$\bar{C}^{j,k,(2)}_{l,m}(t_i)$, and the quotient blows up. Thus, we care about the condition that the time net size $|\pi|$ must satisfy in order this explosure does not take place. This leads to an upper threshold size that we denote by $\pi^{\text{threshold}}$.
The estimates (1.15) and (1.18) enable us to establish our main result. The time net $\pi^{(\beta)}$ is still defined by (1.9).

Theorem 1.3.2. *Let $\alpha \in (0,1]$. Assume $g \in \mathbf{L}_{2,\alpha}$ and $\mathbb{E}_{\mathbb{P}}|g(S_T)|^{2p_0} < \infty$ for some $p_0 > 1$.*

1. Regular grid $\pi^{(1)}$. *For N sufficiently large to ensure $|\pi^{(1)}| = \frac{T}{N} \leq \pi^{\text{threshold}}$, one has*
$$\left(\mathbb{E}_{\mathbb{P}}\left|\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi^{(1)})\right|^2\right)^{1/2} = \mathcal{O}(\frac{1}{N^{\alpha/2}}). \tag{1.19}$$

2. Non regular grid $\pi^{(\beta)}$, $\beta \in (0,1)$. *Taking N sufficiently large to ensure $|\pi^{(\beta)}| \leq \pi^{\text{threshold}}$, one has*
$$\left(\mathbb{E}_{\mathbb{P}}\left|\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi^{(\beta)})\right|^2\right)^{1/2} = \begin{cases} \mathcal{O}(\frac{1}{N^{\frac{\alpha}{2\beta}}}) \text{ if } \beta \in (\frac{\alpha}{2},1), \\ \mathcal{O}(\frac{\sqrt{\log N}}{N}) \text{ if } \beta = \frac{\alpha}{2}, \\ \mathcal{O}(\frac{1}{N}) \text{ if } \beta \in (0,\frac{\alpha}{2}). \end{cases}$$

We mention that, for the regular grid, the estimate (1.19) of the rate of convergence $N^{\alpha/2}$ is in general tight for $\alpha < 1$ (because the estimate (1.18) is tight, see the discussions in Geiss and Hujo [GH07] and Gobet and Makhlouf [GM08]). In addition, it coincides with the rate of convergence of the Delta hedging strategy: in other words, for non-smooth payoffs, the DGHS used with regular grids does not improve the rate of convergence of the tracking error. The use of irregular grids crucially helps to increase the convergence rate of the tracking error, for any fractional regularity $\alpha \in (0,1]$, up to the rate N. This is confirmed by our numerical results (see Chapter 10, Section 10.7): in Figure 1.1, we illustrate, for a digital Call (whose regularity degree α is equal to 0.5), how the choice of the time net may improve the convergence order.
In conclusion, as an application, our results may be used in order to reduce the risk inherent in hedging European contingent claims.

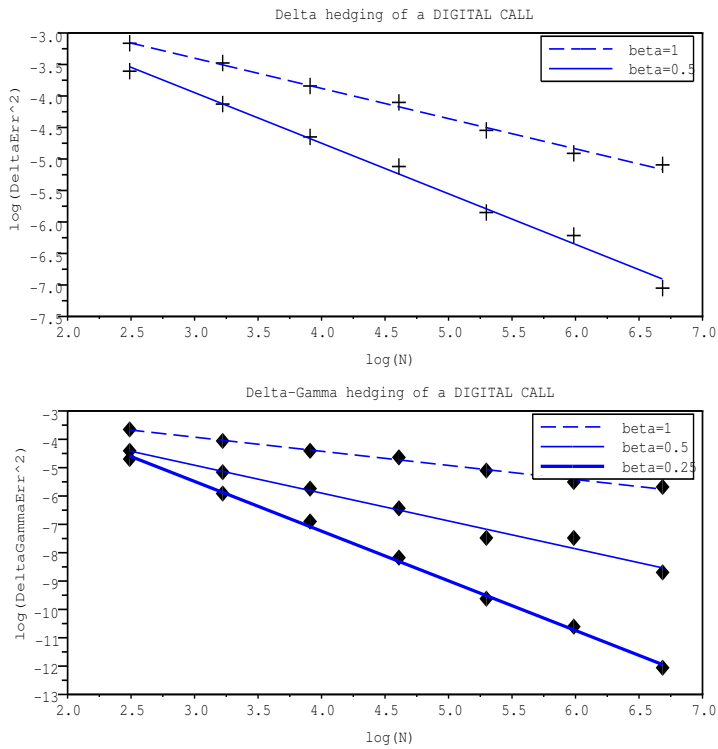

Figure 1.1: For a digital Call: at the top (Delta hedging strategy), $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N^{\Delta}(g,\pi^{(\beta)})|^2)$ vs $\log(N)$. At the bottom (Delta-Gamma hedging strategy), $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi^{(\beta)})|^2)$ vs $\log(N)$.

Chapter 2

Introduction (en français)

Cette thèse est composée de trois parties concernant trois problèmes issus des domaines des probabilités numériques et des mathématiques financières: la régularité $\mathbf{L}_2$ en temps d'EDSR à fonction terminale irrégulière, le calcul efficace des prix de tranches de CDO et l'*erreur de couverture* (*tracking error*) inhérente à la stratégie de couverture en Delta-Gamma.

2.1 Régularité $\mathbf{L}_2$ en temps d'EDSRs à fonction terminale irrégulière

2.1.1 Qu'est ce qu'une EDSR?

En quelques mots, une équation différentielle stochastique rétrograde (EDSR) est une sorte d'équation différentielle stochastique (EDS) où c'est la condition terminale qui est spécifiée à la place de la condition initiale, tout en veillant à choisir une solution qui soit adaptée à la filtration sous-jacente.
Les EDSR ont été initialement introduites en 1973 par Bismut [Bis73], dans le cas linéaire, dans le but de donner une interprétation probabiliste du principe du maximum de Pontryagin dans un cadre de contrôle stochastique optimal.
En 1990, Pardoux et Peng [PP90] ont énoncé le premier résultat général: ils ont prouvé l'existence et l'unicité de solution d'EDSR non linéaires, définies par

$$\begin{cases} -dY_t &= f(t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T &= \xi, \end{cases} \tag{2.1}$$

sous les hypothèses suivantes: W est un mouvement brownien standard, f (un *générateur* ou *driver* aléatoire) est lipschitzien par rapport aux variables d'espace y et z, et la *condition terminale* ξ, ainsi que le processus $\{f(t,0,0)\}_{t\in[0,T]}$, sont de carré intégrable. La solution est un couple de processus adaptés (Y, Z) tels que Y est continu, Z est prévisible, et

$$\mathbb{E}\big[\sup_{t\in[0,T]} |Y_t|^2\big] + \mathbb{E}\big[\int_0^T |Z_s|^2 ds\big] < \infty.$$

Ici, la filtration correspond à la filtration naturelle augmentée du mouvement brownien W.
Par exemple, dans le cas où $f = 0$, la solution (Y, Z) est donnée par $Y_t = \mathbb{E}(\xi|\mathcal{F}_t)$, et Z est donné par la propriété de représentation prévisible.
Plus tard, des résultats d'existence et d'unicité ont été énoncés sous des hypothèses plus faibles et dans des espaces différents (voir Chapitre 3, section 3.1). En 1997, El Karoui, Peng et Quenez [KPQ97b] ont étudié d'autres problématiques, telles que le théorème de comparaison (dans le cas unidimensionnel), les EDSR concaves, la continuité et la différentiabilité des EDSR markoviennes, et la différentiation des EDSR au sens de Malliavin. D'autre part, ils ont décrit quelques applications des EDSR en finance. Par exemple, dans un marché complet, et dans un modèle de type Itô, le processus de prix d'une option européenne de payoff ξ et de maturité T satisfait une EDSR linéaire (c.-à-d. le driver f est linéaire en (y, z)). Les EDSR non linéaires apparaîssent quand la stratégie d'exercice est sujette à des contraintes, ou dans les modèles de marchés incomplets. Les fonctions d'utilité récursives peuvent être aussi représentées par des EDSR.
Les EDSR possèdent d'autres applications en jeux différentiels stochastiques de somme nulle (cf. [HL95]) et pour le problème de "switching" où les EDSR réfléchies (EDSRR) interviennent (cf. [HJ07]). Les EDSRR ont été introduites par El Karoui *et al.* in [KKP$^+$97], et leur développement a été motivé par la valorisation des options américaines (cf. [KPQ97a]).
D'autre part, une EDSR peut être vue comme une représentation stochastique d'une EDP parabolique non linéaire (voir paragraphe 2.1.2).

2.1.2 Cadre et défis

Nous nous intéressons aux EDSR markoviennes à composante progressive multidimensionnelle et à driver lipschitzien:

$$\begin{cases} X_0 &= x_0, \\ dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ -dY_t &= f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T &= g(X_T), \end{cases} \tag{2.2}$$

où $g(X_T) \in \mathbf{L}_2$ et W est un mouvement brownien standard de dimension q. La *composante progressive* X_t est à valeurs dans $\mathbb{R}^d$, et la *composante rétrograde* (Y_t, Z_t) est à valeurs dans $\mathbb{R} \times \mathbb{R}^{1\times q}$. Les fonctions b, σ, f et g sont déterministes.
De la même manière qu'une EDS est liée aux EDP paraboliques linéaires via la formule de Feynman-Kac, l'EDSR (2.2) est liée à une EDP parabolique semi-linéaire: si, pour (t, x) fixé, on définit $(Y_s^{t,x}, Z_s^{t,x})_{s\in[t,T]}$ comme la solution de

$$\begin{cases} X_s^{t,x} &= x, \quad 0 \le s \le t, \\ \mathrm{d}X_s^{t,x} &= b(s, X_s^{t,x})\mathrm{d}s + \sigma(s, X_s^{t,x})\mathrm{d}W_s, \\ -\mathrm{d}Y_s^{t,x} &= f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\mathrm{d}s - Z_s^{t,x}\mathrm{d}W_s, \\ Y_T^{t,x} &= g(X_T^{t,x}), \end{cases} \tag{2.3}$$

et u comme la solution (en un certain sens) de

$$\begin{cases} \partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),\nabla_x u(t,x)\sigma(t,x)) = 0 \text{ for } t < T, \\ u(T,x) = g(x), \end{cases} \tag{2.4}$$

alors

$$(Y_s^{t,x}, Z_s^{t,x}) = (u(s, X_s^{t,x}), \nabla_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x})).$$

En particulier,

$$(u(t,x), \nabla_x u(t,x)\sigma(t,x)) = (Y_t^{t,x}, Z_t^{t,x}).$$

Pour plus de détails, voir le paragraphe 3.2.2 du Chapitre 3.
Ainsi, cette partie de la thèse s'inscrit dans le cadre des méthodes probabilistes de résolution des EDP non linéaires. En effet, le défi que la représentation stochastique des EDPs (qui convient bien aux méthodes de Monte-Carlo) cherche à relever est de remédier au problème de la grande dimension qui se pose pour les méthodes déterministes.
Plus précisément, nous nous intéressons à une erreur de discrétisation associée à l'EDSR (2.2). C'est ce qu'on désigne par *régularité* $\mathbf{L}_2$ *en temps* de la composante Z, et qu'on définit par

$$\mathcal{E}(Z,\pi) := \sum_{i=0}^{N-1} \mathbb{E} \int_{t_i}^{t_{i+1}} |Z_t - \bar{Z}_{t_i}|^2 dt,$$

où $\bar{Z}_{t_i} = \frac{1}{t_{i+1}-t_i}\mathbb{E}^{\mathcal{F}_{t_i}} \int_{t_i}^{t_{i+1}} Z_s ds$, et $\pi = \{0 = t_0 < \cdots < t_i < \cdots < t_N = T\}$ est une subdivision (déterministe) de l'intervalle de temps $[0,T]$ en un nombre fixé de sous-intervalles (égal à N), de taille $|\pi| := \sup_{0\le i<N}(t_{i+1} - t_i)$. $\mathcal{E}(Z,\pi)$ est le carré de la distance $\mathbf{L}_2$ entre Z et sa meilleure approximation $\mathbf{L}_2$ par un processus adapté constant par morceaux.
La quantité $\mathcal{E}(Z,\pi)$ apparaît dans l'étude de la vitesse de convergence de certains schémas numériques d'approximation de l'EDSR (2.2), en particulier celui défini par l'*équation de programmation dynamique* suivante, dans une forme explicite (pour $i < N$):

$$\begin{cases} Y_{t_i}^\pi &= \mathbb{E}^{\mathcal{F}_{t_i}}(Y_{t_{i+1}}^\pi + (t_{i+1} - t_i) f(t_i, X_{t_i}^\pi, Y_{t_{i+1}}^\pi, Z_{t_i}^\pi)), \\ Z_{t_i}^\pi &= \frac{1}{(t_{i+1}-t_i)}\mathbb{E}^{\mathcal{F}_{t_i}}(Y_{t_{i+1}}^\pi (W_{t_{i+1}} - W_{t_i})^*), \end{cases} \tag{2.5}$$

et $Y_{t_N}^\pi = g(X_T^\pi)$ (* désigne la transposée). La composante progressive X est approchée par X^π (en utilisant un schéma d'Euler par exemple).
En effet, l'ordre de convergence de (2.5) se réduit à celui de $\mathcal{E}(Z,\pi)$, auquel s'ajoutent l'erreur d'approximation de la condition terminale et la taille $|\pi|$ de la grille en temps (voir le paragraphe 3.4.1 du Chapitre 3). En d'autres termes, obtenir une bonne vitesse de convergence pour $\mathcal{E}(Z,\pi)$ permet de bien approcher l'EDSR (2.2).
Il est déjà connu que $\mathcal{E}(Z,\pi)$ est d'ordre $|\pi|$ quand la fonction terminale g est lipschitzienne ([Zha04]). Notre objectif est d'étudier l'ordre de convergence de $\mathcal{E}(Z,\pi)$ quand g est irrégulière. C'est le cas, par exemple, de la fonction indicatrice d'un

domaine (qui correspond aux options digitales en finance). Nous cherchons à exprimer la vitesse de convergence en fonction d'une certaine régularité fractionnaire de la fonction terminale g et de la taille $|\pi|$ de la grille en temps.

2.1.3 Principaux résultats et outils

En s'inspirant des travaux de Geiss *et al.* ([GG04][GH07]) et de Gobet et Munos ([GM05b]), nous définissons la régularité fractionnaire de g par l'ordre $\alpha \in (0,1]$ de décroissance de la variance conditionnelle moyenne de $g(X_T)$ sachant X_t, quand t tend vers T. Nous introduisons l'espace $\mathbf{L}_{2,\alpha}$, qui est une sorte d'équivalent stochastique de l'espace de Sobolev fractionnaire, et nous le définissons par

$$\mathbf{L}_{2,\alpha} := \{g \text{ t.q. } \mathbb{E}(g(X_T)^2) + \sup_{0\le t<T} \frac{V_{t,T}(g)}{(T-t)^\alpha} < +\infty\}, \tag{2.6}$$

équipé d'une norme dont le carré est

$$K^\alpha(g) := \mathbb{E}|g(X_T)|^2 + \sup_{t\in[0,T)} \frac{V_{t,T}(g)}{(T-t)^\alpha}, \tag{2.7}$$

avec

$$V_{t,T}(g) := \mathbb{E}\left|g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T))\right|^2. \tag{2.8}$$

Nous nous placerons sous des hypothèses de régularité et de bornitude pour les coefficients b et σ, ainsi que d'ellipticité uniforme pour σ.
Nos preuves reposent sur des techniques d'analyse stochastique combinant EDP, martingales, calcul d'Itô et EDSR dans $\mathbf{L}_p$ ($p \in (1,2]$).
Nous signalons que généralement, avec ces outils, g est supposée de croissance polynômiale, assurant à $g(X_T)$ d'appartenir à tous les $\mathbf{L}_p$ pour $p > 0$. Ici, nous soulignons que nous supposons seulement $g(X_T) \in \mathbf{L}_2$, ce qui représente la condition minimale pour discuter de l'existence et de l'unicité de solution de (2.2) dans les espaces $\mathbf{L}_2$.
De plus, nos résultats principaux sont valables pour $g \in \bigcup_{\alpha\in(0,1]} \mathbf{L}_{2,\alpha}$. Nous signalons que cet espace est très large: il contient non seulement les fonctions régulières telles que les fonctions hölderiennes, mais aussi des fonctions irrégulières. Par exemple, la fonction indicatrice d'un domaine régulier appartient à $\mathbf{L}_{2,\frac{1}{2}}$.
Nous étudions d'abord le cas de driver nul ($f \equiv 0$), où on note la composante Z par z. Nous établissons le lien entre $\mathcal{E}(z,\pi)$ et la dérivée seconde de la solution $u(t,x) := \mathbb{E}(g(X_T)|X_t = x)$ de l'EDP linéaire correspondante, en montrant que

$$\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2\, ds \le \frac{C}{N} + C\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} (t_{k+1}-r)\mathbb{E}|D^2u(r,X_r)|^2 dr$$

(et il y a équivalence si $(X_t)_t$ est le mouvement brownien).
Ensuite, nous caractérisons la régularité de g en termes d'estimations $\mathbf{L}_2$ des dérivées de u:

Lemme 2.1.1. *Soit $\alpha \in (0,1]$. Les trois assertions suivantes sont équivalentes:*
(i) $g \in \mathbf{L}_{2,\alpha}$.
(ii) $\exists C^\alpha(g) > 0$, telle que, $\forall t \in [0,T)$,

$$\int_0^t \mathbb{E}\left|D^2 u(s,X_s)\right|^2 ds \leq \frac{C^\alpha(g)}{(T-t)^{1-\alpha}}.$$

(iii) $\exists C^\alpha(g) > 0$, telle que, $\forall t \in [0,T)$,

$$\mathbb{E}\left|\nabla_x u(t,X_t)\right|^2 \leq \frac{C^\alpha(g)}{(T-t)^{1-\alpha}}.$$

Et, si $g \in \mathbf{L}_{2,\alpha}$, on peut prendre $C^\alpha(g) = CK^\alpha(g)$ en (ii) et (iii).
Si $\alpha < 1$ (resp. $\alpha = 1$), les trois assertions précédentes sont aussi équivalentes à (resp. impliquent) la suivante:
(iv) $\exists C^\alpha(g) > 0$, telle que, $\forall t \in [0,T)$,

$$\mathbb{E}\left|D^2 u(t,X_t)\right|^2 \leq \frac{C^\alpha(g)}{(T-t)^{2-\alpha}}$$

(et on peut prendre $C^\alpha(g) = CK^\alpha(g)$).

Grâce à ces équivalences, le lemme 2.1.1 s'avère crucial (bien que sa preuve ne soit pas très technique) pour montrer l'optimalité de nos résultats. Avec ce lemme, l'hypothèse "$g \in \mathbf{L}_{2,\alpha}$" apparaît naturellement dans notre cadre de travail, si on veut un ordre de convergence de $\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2 ds$ qui soit polynômial par rapport au pas de discrétisation $|\pi|$. En effet, Geiss et Hujo [GH07] (Théorème 1.3 et Théorème 2.1) ont montré que, si $\hat{X}$ est le mouvement brownien ou le mouvement brownien géométrique, alors les assertions suivantes sont équivalentes ($\alpha \in (0,1)$):
(GH-i) $\exists C > 0, \forall t \in [0,T)$,

$$\mathbb{E}\left|\nabla_x u(t,\hat{X}_t)\right|^2 \leq \frac{C}{(T-t)^{1-\alpha}}.$$

(GH-ii) $\exists C > 0$ telle que, pour toutes les subdivisions $\pi = (t_k)_{k=0\ldots N}$,

$$\sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2 ds \leq \frac{C}{N^\alpha}.$$

Et Lemme 2.1.1 montre que l'assertion (GH-i) est équivalente à l'assertion "$g \in \mathbf{L}_{2,\alpha}$" (même pour $\alpha = 1$). Pour des travaux connexes, avec des diffusions scalaires et homogènes en temps, cf. [GG04].
Grâce à Lemme 2.1.1, nous montrons que $\mathcal{E}(Z,\pi)$ est d'ordre $|\pi|^\alpha$ lorsque π est la grille régulière.
Cependant, pour des grilles en temps quelconques, nous énonçons le théorème suivant.

Théorème 2.1.1. *Supposons que $g \in \mathbf{L}_{2,\alpha}$, pour un certain $\alpha \in (0,1]$. On prend $\beta = 1$, si $\alpha = 1$, et $\beta < \alpha$ sinon. Alors, $\exists C > 0$ telle que, pour toute subdivision $\pi = \{t_k, k = 0...N\}$,*

$$\mathcal{E}(z,\pi) \leq C|\pi| + C \sup_{k=0...N-1} \left(\frac{t_{k+1} - t_k}{(T - t_k)^{1-\beta}} \right).$$

Maintenant, nous pouvons affirmer que la borne optimale $\frac{C}{N}$ peut être atteinte avec l'utilisation de la subdivision définie, pour $\beta \in (0,1]$, par

$$\pi^{(\beta)} := \big\{ t_k^{(N,\beta)} := T - T\big(1 - \frac{k}{N}\big)^{\frac{1}{\beta}}, 0 \leq k \leq N \big\}. \tag{2.9}$$

Remarque: $\pi^{(1)} = (t_k^{(N,1)})$ coïncide avec la grille régulière, et pour $\beta < 1$, les points de $\pi^{(\beta)}$ sont plus concentrés près de T.

Corollaire 2.1.1. *Supposons que $g \in \mathbf{L}_{2,\alpha}$, pour un certain $\alpha \in (0,1]$. On prend β comme dans Théorème 2.1.1. Alors,*

$$\mathcal{E}(z,\pi^{(\beta)}) = \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k^{(N,\beta)}}^{t_{k+1}^{(N,\beta)}} \left| z_s - \bar{z}_{t_k^{(N,\beta)}} \right|^2 ds = \mathcal{O}(\frac{1}{N}).$$

Nous signalons que, avec la grille régulière, l'ordre $\frac{1}{N}$ ne peut pas être atteint avec certaines fonctions $g \in \mathbf{L}_{2,\alpha}$ de régularité $\alpha < 1$; l'ordre optimal dans ce cas vaut $\frac{1}{N^\alpha}$, comme montré dans [GT01] à travers quelques exemples.
Ensuite, dans le cas général de driver non nul, nous montrons que $Z_t - z_t = U_t \sigma(t, X_t)$, où U est lui-même solution d'une EDSR linéaire (voir Chapitre 5, section 5.8). L'idée est d'utiliser la représentation de la composante Z comme dérivée au sens de Malliavin de la composante Y. Cependant, à cause de l'explosion des dérivées de u en T, nous procédons par localisation en temps, et nous énonçons quelques lemmes techniques assurant la stabilité. Après avoir estimé U, le cas de driver non nul s'avère n'être qu'une perturbation du cas de driver nul, ce qui signifie que $\mathcal{E}(Z,\pi)$ se comporte comme $\mathcal{E}(z,\pi)$ (au pas de discrétisation près):

Théorème 2.1.2. *Supposons que $g \in \mathbf{L}_{2,\alpha}$, pour un certain $\alpha \in (0,1]$. Alors, il existe une constante positive C telle que, pour toute subdivision $\pi = \{t_k : k = 0...N\}$*

$$\mathcal{E}(Z,\pi) = \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left| Z_s - \bar{Z}_{t_k} \right|^2 ds \leq C\mathcal{E}(z,\pi) + C|\pi|.$$

Ceci nous permet d'énoncer notre résultat principal: $\mathcal{E}(Z,\pi)$ est d'ordre $\frac{1}{N^\alpha}$ si π est la grille régulière, et d'ordre $\frac{1}{N}$ si π est une subdivision $\pi^{(\beta)}$ convenable.

Théorème 2.1.3. *Supposons que $g \in \mathbf{L}_{2,\alpha}$, pour un certain $\alpha \in (0,1]$. Alors,*

a) avec le choix de la grille régulière $\pi^{(1)}$,

$$\mathcal{E}(Z, \pi^{(1)}) = \mathcal{O}(\frac{1}{N^{\alpha}});$$

b) avec le choix de $\pi^{(\beta)}$, *avec* β *comme dans Corollaire 2.1.1,*

$$\mathcal{E}(Z, \pi^{(\beta)}) = \mathcal{O}(\frac{1}{N}).$$

L'intérêt de nos résultats est triple. D'abord, nous avons réussi, grâce à un bon choix de la subdivision en temps, à obtenir un ordre de convergence de $\frac{1}{N}$ pour $\mathcal{E}(Z, \pi)$, malgré l'irrégularité de la fonction terminale g. Ensuite, ceci est valable dans le cadre général d'EDSR à composante progressive multi-dimensionnelle et à driver lipschitzien. Enfin, les résultats sont optimaux.

2.2 Calcul efficace des prix de tranches de CDO

2.2.1 Modélisation et pricing des CDO

Une *tranche de CDO* (Collateralized Debt Obligation), basée sur un portefeuille sous-jacent de noms (entreprises) regroupant plusieurs actifs (obligations, prêts, CDO,...) est un contrat bilatéral qui consiste en une protection contre les pertes occasionnées par des défauts dans le portefeuille, dans une tranche $[K_1, K_2]$, sur une période $[0, T]$. Les valeurs (fixées) K_1 et K_2 s'appellent respectivement le *point d'attachement* (ou *strike bas*) et le *point de détachement* (ou *strike haut*).
Lorsqu'un nom q du portefeuille fait défaut, une partie seulement de son nominal ν_q peut être récupérée, et on note R_q le taux de recouvrement correspondant (qui peut être non déterministe). Ainsi, la perte causée par le défaut du nom q est égale à $l_q := \nu_q(1 - R_q)$, et la perte totale du portefeuille sur une période $[0, t]$ vaut

$$L_t := \sum_q l_q \mathbb{1}_{\tau_q \leq t} = \sum_q \nu_q (1 - R_q) \mathbb{1}_{\tau_q \leq t},$$

où τ_q désigne le temps de défaut du nom q.
Il est connu que le calcul du prix d'une tranche de CDO se ramène au calcul de $\mathbb{E}((L_T - K)^+)$, pour différentes valeurs de T et de K. Pour plus de détails, voir, par exemple, [Das05].
Deux approches principales ont été développées pour modéliser la perte du portefeuille: ce sont les approches *bottom-up* et *top-down* (cf. [Gie08]).
Dans les modèles *bottom-up*, on part de la description de la dynamique individuelle de chaque nom (temps de défaut, intensité de défaut, perte, ...), ainsi que les dépendances éventuelles entre eux. La dynamique de la perte agrégée s'en déduit implicitement. Au contraire, dans l'approche *top-down*, on modélise d'abord le processus de perte du portefeuille (sans faire appel à ses constituants sous-jacents), pour ensuite restituer la dynamique individuelle de chaque nom. Dans chacun des deux

modèles, quelques *facteurs auxiliaires* peuvent être ajoutés dans le but de prendre en compte des informations supplémentaires.
Chacune des deux approches a ses avantages et ses inconvénients, en fonction de l'objectif pour lequel elle est suivie. En pratique, les modèles *bottom-up* conviennent plus à la couverture, alors que les modèles *top-down* conviennent plutôt au pricing et à la calibration rapides. On renvoie à [Gie08] et [BCJ08] pour une comparaison et une critique plus détaillées des deux approches.

2.2.2 Notre modèle

Les processus de défaut et de perte du portefeuille

On considère un portefeuille de crédit composé de n_0 noms (n_0 varie de 50 à 150 en pratique), et on suppose que chaque nom q peut faire défaut au plus une seule fois sur la période $[0, T]$, à un temps aléatoire τ_q.
On définit le *processus* $(L_t)_{t\in[0,T]}$ *de perte (normalisée) du portefeuille* par

$$L_t := \frac{1}{n_0}\sum_{q=1}^{n_0}(1 - R_q)\mathbb{1}_{\tau_q\leq t},$$

où les *taux de recouvrement* (stochastiques) $(R_q)_{1\leq q\leq n_0}$ sont des variables aléatoires i.i.d. (indépendantes aussi des temps de défauts $(\tau_q)_q$) à valeurs dans $[0, 1]$ (ce qui implique que $L_t \in [0, 1]$). Les nominaux (ν_q) sont supposés égaux.
Afin de prendre en compte un éventuel phénomène de contagion, on introduit, comme facteur auxiliaire, le *processus de défaut* $(N_t)_{t\in[0,T]}$, où N_t représente le nombre de défauts qui se sont produits avant la date t:

$$N_t := \sum_{q=1}^{n_0}\mathbb{1}_{\tau_q\leq t}.$$

L'intensité de défaut λ

On suppose que le compensateur du processus (croissant et adapté) N est dérivable en temps. On note sa dérivée, qui représente l'*intensité de défaut*, par $(\lambda_t)_{t\in[0,T]}$ (ce qui signifie que $(N_t - \int_0^t \lambda_s \mathrm{d}s)_t$ est une martingale locale).
Différents modèles pour l'intensité λ ont été abordés dans la littérature, avec la tentative de prendre en compte des phénomènes de contagion qui peuvent se manifester dans la réalité des marchés. Les principaux modèles qui expriment l'intensité λ_t comme fonction de N_t sont le modèle linéaire ($\lambda_t := c_0 + c_1 N_t$, cf. Frey et Backhaus [FB08]) et le modèle exponentiel ($\lambda_t := c_0 e^{c_1 N_t}$, cf. Davis et Lo [DL01]).
Nous optons pour un modèle intermédiaire entre les modèles linéaire et exponentiel, introduit par Frey et Backhaus [FB08]. Il est défini par

$$\lambda_t = \lambda(t, L_t, N_t) := c_0 + \frac{c_1}{c_2}\left(e^{c_2(N_t-\mu(t))^+} - 1\right),$$

où c_0, c_1 et c_2 sont des constantes positives, et μ représente le nombre moyen de défauts auquel on s'attend *a priori*. Ainsi, il n'y a pas de contagion si le nombre de défauts observé est inférieur à μ (si $N_t < \mu$, $\lambda_t = c_0$), sinon, la contagion prolifère exponentiellement en fonction du nombre de défauts (dans un scénario catastrophique, par exemple). On fait remarquer aussi la flexibilité de ce modèle, car on peut le rendre presque linéaire en choisissant c_2 "petit" (en effet, lorsque $c_2 \simeq 0$, $\lambda_t \simeq c_0 + c_1(N_t - \mu)^+$). De plus, on suppose que $\lambda(t, l, n) = 0$ dès que $n \geq n_0$ (puisque chacun des n_0 noms est supposé faire défaut une seule fois au plus).

Les taux de recouvrement $(R_q)_q$

Dans la littérature traitant de la valorisation et de la couverture des CDO, les taux de recouvrement $(R_q)_q$ sont souvent supposés déterministes, dans un but de simplification, et égaux à une constante (40%, par exemple). Cependant, cette hypothèse peut ne pas être assez réaliste, et, comme indiqué dans [Ron04], l'impact du taux de recouvrement sur la loi des défauts peut être très important. Il est montré aussi dans [DG01] que la loi des taux de recouvrement peut varier d'un type de dette à un autre.
Dans notre étude, les taux de recouvrement $(R_q)_q$ sont des variables aléatoires i.i.d., indépendantes aussi des temps de défauts $(\tau_q)_q$, et on suppose, comme dans [Ron04], qu'ils suivent la loi bêta sur $(0, 1)$, de paramètres $a > 0$ et $b > 0$. La densité de cette loi est définie par

$$\beta^{a,b}(z) = \frac{1}{B(a,b)} z^{a-1}(1-z)^{b-1} \mathbb{1}_{(0,1)}(z),$$

où $B(a, b) = \int_0^1 z^{a-1}(1-z)^{b-1}\mathrm{d}z = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, et Γ est la fonction gamma.
Ceci nous donne une famille variée de lois à support borné et prenant différentes formes. De plus, l'avantage principal de prendre une loi beta pour les taux de recouvrement est de pouvoir calibrer les paramètres a et b à partir de la connaissance de la moyenne $\mu := \mathbb{E}(R)$ et de la variance $\sigma^2 := \mathbb{V}ar(R)$ de ces taux.

2.2.3 Notre apport

Dans le cadre d'une approche markovienne de type *top down*, et autorisant les taux de recouvrement à être non déterministes (nous supposons seulement qu'ils sont des variables aléatoires i.i.d.), nous proposons une méthode de Monte-Carlo séquentielle afin de calculer efficacement le prix d'une tranche de CDO. Nous appliquons l'algorithme des *variables de contrôle séquentielles* (VCS en abrégé), introduit par Gobet et Maire [GM05a] pour le calcul des solutions d'EDP linéaires de second ordre ayant une représentation de type Feynman-Kac (voir Chapitre 7).

2.2.4 La fonction prix

Nous définissons le *processus de proportion de défauts* $(P_t)_{t\in[0,T]}$ (à valeurs dans $[0,1]$) par

$$P_t := \frac{N_t}{n_0} = \frac{1}{n_0}\sum_{q\geq 1} \mathbb{1}_{\tau_q \leq t}.$$

Le processus de proportion de défauts et le processus de perte partant resp. de p et l au temps t sont désignés resp. par $(P_s^{t,l,p})_{t\leq s\leq T}$ et $(L_s^{t,l,p})_{t\leq s\leq T}$.
Nous notons le processus de Markov (L,P) par X:

$$X_t := (L_t, P_t), \quad X_s^{t,l,p} := (L_s^{t,l,p}, P_s^{t,l,p}).$$

Soit $\bar{D} := [0,T]\times[0,1]\times[0,1]$. Nous supposons maintenant que $\lambda_t = \lambda(t, L_t, P_t)$, avec λ uniformément bornée sur $\bar{D}$, et que $\lambda(t,l,p) = 0$ si $p > 1 - \frac{1}{n_0}$ (c.-à-d. dès que $n > n_0 - 1$) ou si $l > p$ (ce qui est naturel, puisque $L_t \leq P_t$).
Pour des fonctions génériques uniformément bornées $\tilde{f}$ et $\tilde{g}$, definies sur $\bar{D}$, et pour $f(t,l,p) := 0$ et $g(l,p) := (l-K)^+$, nous posons (pour $(t,l,p)\in\bar{D}$)

$$\begin{aligned}
\Psi(\tilde{f},\tilde{g},X^{t,l,p}) &:= \tilde{g}(X_T^{t,l,p}) - \int_t^T \tilde{f}(s, X_s^{t,l,p})\mathrm{d}s,\\
v(t,l,p) &:= \mathbb{E}[\Psi(\tilde{f},\tilde{g},X^{t,l,p})],\\
u(t,l,p) &:= \mathbb{E}[\Psi(f,g,X^{t,l,p})] = \mathbb{E}[(L_T-K)^+ | L_t = l, P_t = p].
\end{aligned} \tag{2.10}$$

Nous montrons (cf. Théorème 8.2.1) que v est la *solution classique* (c.-à-d. uniformément bornée par rapport à (t,l,p) et dérivable par rapport à t) de

$$\begin{cases} \mathcal{A}v &= \tilde{f} \quad \text{when } t < T,\\ \mathcal{B}v &= \tilde{g} \quad \text{when } t = T, \end{cases} \tag{2.11}$$

où

$$\begin{aligned}
\mathcal{A}v(t,l,p) &:= \partial_t v(t,l,p) + \lambda(t,l,p)\int_0^1 \left(v(t, l+\frac{y}{n_0}, p+\frac{1}{n_0}) - v(t,l,p)\right)\beta^{b,a}(y)dy,\\
\mathcal{B}v(t,l,p) &:= v(t,l,p),
\end{aligned}$$

et ainsi l'est la fonction prix u qu'on cherche à calculer, avec $\tilde{f} = f$ et $\tilde{g} = g$. Ceci nous permettra d'utiliser la méthode VCS pour calculer u.

2.2.5 Calcul de la fonction prix en utilisant VCS

La méthode de Monte-Carlo standard convient particulièrement aux problèmes de grande dimension, mais présente une convergence lente. Deux techniques principales de réduction de variance sont souvent utilisées pour minimiser l'erreur de convergence: la technique des *variables de contrôle* et celle des *fonctions d'importance* (cf.

Halton [Hal70] et Glasserman [Gla04], par exemple). De plus, des méthodes de Monte-Carlo séquentielles, utilisant itérativement les deux techniques précédentes, ont été développées (cf. [Hal62] et [Hal70], par exemple).
Gobet et Maire ont introduit dans [GM04] l'algorithme de *variables de contrôle séquentielles* dans le but de résoudre l'équation de Poisson avec conditions aux bords de type Dirichlet sur des domaines carrés. Ensuite, ils ont étendu leur approche à des processus de Markov généraux (tels que les diffusions et les processus à sauts) liés à des EDP elliptiques et paraboliques linéaires de second ordre, avec conditions aux bords linéaires.
L'idée de l'algorithme VCS est d'approcher la solution u par une suite $(u_n)_n$ telle que, à la n^{ieme} itération,

$$u_n(x_i) = u_{n-1}(x_i) + \bar{y}_n(x_i),$$

où les $(x_i)_{1\leq i\leq N}$ sont des points de $\bar{D}$ (nœuds d'interpolation, par exemple), et $\bar{y}_n(x_i)$ est une approximation par Monte-Carlo du terme de correction $y_n(x_i) := u(x_i) - u_{n-1}(x_i)$, représenté comme une espérance: en effet, à partir de (2.10) et (2.11), et par linéarité de la fonctionnelle Ψ, on a

$$y_n(x) = \mathbb{E}[\Psi(f - \mathcal{A}u_{n-1}, g - \mathcal{B}u_{n-1}, X(x))].$$

Une approximation globale basée sur les valeurs aux points $(x_i)_i$ est obtenue par

$$u_n = u_{n-1} + \mathcal{P}\bar{y}_n,$$

où $\mathcal{P}$ est un opérateur d'approximation linéaire (un opérateur d'interpolation, par exemple), basé sur les points $(x_i)_i$, c.-à-d. , pour toute fonction φ et pour certaines fonctions de base $(\mathcal{C}_i)_i$ données,

$$\mathcal{P}\varphi(.) = \sum_{i=1}^{N} \varphi(x_i)\mathcal{C}_i(.).$$

L'opérateur $\mathcal{P}$ est supposé avoir la propriété de stabilité suivante

$$\mathcal{P}(\mathcal{P}\varphi) = \mathcal{P}\varphi.$$

Cette propriété est satisfaite par les opérateurs d'interpolation et de projection.
Voici l'algorithme:

Algorithme 2.2.1 (VCS)**.**

1. *Initialisation:* $u_0 := 0$.
2. *A l'itération n: pour chaque x_i, simuler M trajectoires indépendantes $(X^{n,m}(x_i))_{1\leq m\leq M}$ de $X(x_i)$ (indépendamment de tout le reste), et poser*

$$\bar{y}_n(x_i) := \frac{1}{M}\sum_{m=1}^{M} \Psi(f - \mathcal{A}u_{n-1}, g - \mathcal{B}u_{n-1}, X^{n,m}(x_i)).$$

3. *Toujours à l'itération n: construire la solution globale en utilisant l'approximation linéaire du terme de correction:*

$$u_n := u_{n-1} + \mathcal{P}\bar{y}_n = u_{n-1} + \sum_{i=1}^{N} \bar{y}_n(x_i)\mathcal{C}_i.$$

Signalons que, dans leur article, Gobet et Maire ont présenté le cas général où le processus $X(x)$ est obtenu par une procédure de discrétisation $X^{\Delta}(x)$, avec Δ le paramètre de discrétisation (par exemple, le pas de temps pour le schéma d'Euler). Dans notre cadre, la simulation de $X = (L, P)$ est exacte.
De plus, Gobet et Maire ont étudié le biais $m_n := \sup_{1\leq i\leq N} |\mathbb{E}(u_n(x_i) - u(x_i))|$ et la variance $v_n := \sup_{1\leq i\leq N} |\mathbb{V}ar(u_n(x_i))|$. Essentiellement, ils ont montré que ces deux erreurs convergent géométriquement avec le nombre d'itérations de l'algorithme, à un terme d'erreur près, qui est lié à la qualité d'approximation de u par l'opérateur $\mathcal{P}$ (nous ne détaillons pas les conditions requises pour les fonctions de base $(\mathcal{C}_i)_i$):

Théorème 2.2.1. *(Théorèmes 3.1 et 3.2 dans [GM05a])*

i)

$$m_n \leq \rho_m m_{n-1} + \mathcal{E}_m(\mathcal{P}u - u),$$

où $\rho_m < 1$ (et dépend de Δ, N et $\mathcal{P}$) quand Δ est assez petit, et $\mathcal{E}_m(\mathcal{P}u - u)$ est une norme (dependant de Δ et de N) liée à l'erreur d'approximation $\mathcal{P}u - u$. Et, par conséquent,

$$\limsup_n m_n \leq \frac{1}{1-\rho_m}\mathcal{E}_m(\mathcal{P}u - u).$$

ii)

$$v_n \leq \rho_v v_{n-1} + \frac{1}{M}\mathcal{E}_v(\mathcal{P}u - u),$$

où $\rho_v < 1$ (et dépend de Δ, N, M et $\mathcal{P}$) quand Δ est assez petit et M est assez grand, et $\mathcal{E}_v(\mathcal{P}u - u)$ est une norme (dépendant de Δ et de N) liée à l'erreur d'approximation $\mathcal{P}u - u$. Et, par conséquent,

$$\limsup_n v_n \leq \frac{1}{1-\rho_v}\frac{1}{M}\mathcal{E}_v(\mathcal{P}u - u).$$

Remarque 2.2.1. *Si u appartient à l'espace d'approximation choisi, c.-à-d. $\mathcal{P}u = u$, alors $\mathcal{E}_m(\mathcal{P}u - u) = \mathcal{E}_v(\mathcal{P}u - u) = 0$, et et le biais et la variance convergent géométriquement vers 0 (lorsque Δ est assez petit et M est assez grand).*

Les estimations ci-dessus montrent qu'on n'a pas besoin de prendre M grand pour avoir une valeur précise, puisque de petites valeurs de M peuvent être compensées par une faible erreur d'approximation $\mathcal{P}u - u$: c'est la propriété essentielle de la méthode VCS.

2.2.6 Résultats numériques

Nous présentons d'abord des méthodes de simulation exacte du processus $X = (L, P)$ et des taux de recouvrement (de loi bêta). Pour la procédure d'interpolation, nous utilisons l'interpolation de Gauss-Lobatto-Tchebychev par rapport à chacune des trois variables t, l et p, qui permet d'inclure les bornes des intervalles respectifs $[0, T]$, $[0, 1]$ et $[0, 1]$ (puisqu'on veut calculer la valeur de $u(0, 0, 0)$).
A travers des résultats numériques, nous montrons d'abord, avec $K = 0$ et λ constante (qui est un cas où on connaît explicitement la solution exacte), que l'algorithme VCS s'avère être environ 10 fois plus efficace que la méthode de Monte-Carlo brute. Ensuite, avec notre modèle général, nous montrons que VCS est plus rapide et précis que la méthode de Monte-Carlo brute quand le point d'attachement K n'est pas trop grand par rapport à la perte moyenne (voir Chapitre 8, section 8.4).

2.3 La couverture en Delta-Gamma

2.3.1 La couverture en Delta

Le contraste entre la nature continue en temps des modèles en finance et la nature évidemment discrète en temps des transactions entraîne une *erreur de couverture* (dite *tracking error*) inévitable dès qu'on essaie de couvrir un produit dérivé (avec une stratégie autofinançante).
Dans le modèle de Black-Scholes, la stratégie de couverture en Delta (SCD en abrégé) consiste à détenir un nombre fixe δ_{t_i} d'actifs S entre deux dates de rebalancement t_i et t_{i+1}, pour $0 \leq i \leq N-1$, où N est le nombre d'intervalles de temps. La quantité δ_{t_i} est égale à $\partial_S u(t_i, S_{t_i})$, où u est la fonction prix. A la date de maturité T, la valeur terminale $V_T^{\Delta,N}$ du portefeuille ainsi construit peut être différente du payoff à répliquer $g(S_T)$, entraînant la *tracking error* $\mathcal{E}_N^\Delta := V_T^{\Delta,N} - g(S_T)$. Sous la probabilité risque-neutre, la *tracking error* actualisée $\bar{\mathcal{E}}_N^\Delta := e^{-rT}\mathcal{E}_N^\Delta$ est une martingale qui s'écrit comme

$$\bar{\mathcal{E}}_N^\Delta = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (\delta_{t_i} - \delta_t) \mathrm{d}\bar{S}_t, \tag{2.12}$$

où r est le taux d'intérêt (constant) et $\bar{S}_t := e^{-rT} S_t$ (qui est aussi une martingale sous la probabilité risque-neutre).
Zhang [Zha99] a montré que, si g est lipschitzienne, alors l'erreur $\mathbf{L}_2$, $(\mathbb{E}|\mathcal{E}_N^\Delta|^2)^{\frac{1}{2}}$, converge avec la vitesse $N^{1/2}$ avec des dates de rebalancement uniformes $t_i = iT/N$. Gobet et Temam [GT01] ont montré que la vitesse de convergence dépend fortement, en fait, de la régularité de la fonction payoff g: par exemple, pour les options digitales, l'ordre de convergence vaut $1/4$ au lieu de $1/2$. Geiss *et al.* ont étudié cet aspect en profondeur: dans une série d'articles commencée par Geiss [Gei02], ils ont préparé le terrain pour relier la vitesse de convergence de la *tracking error* de SCD à la régularité fractionnaire du payoff. De plus, ils ont montré que, pour un payoff donné, on peut choisir une grille en temps non uniforme avec N dates de façon à atteindre la vitesse

$N^{1/2}$: plus le payoff est irrégulier, plus les points de la grille sont concentrés près de T. Ces idées ont été étendues dans cette thèse (Partie I) aux EDSR et aux modèles de diffusion multidimensionnels.

2.3.2 La couverture en Delta-Gamma

Dans tous les travaux ci-dessus (et aussi à travers l'expression (2.12)), on peut voir que l'estimation de $\bar{\mathcal{E}}_N^{\Delta}$ se ramène à l'estimation d'une certaine intégrale pondérée de ce qu'on appelle le Gamma $(\partial_S^2 u(t, S_t))_{0\leq t<T}$ de l'option. Ainsi, afin de réduire la *tracking error*, on peut penser à réduire le Gamma du portefeuille en incorporant un autre instrument, de prix $(C(t, S_t))_{0\leq t\leq T}$, dans le portefeuille de couverture: en pratique, c'est une option vanilla liquide (Call/Put dans la monnaie). Cette stratégie est appelée stratégie de couverture en Delta-Gamma (SCDG en abrégé), et est connue des praticiens (cf. Hull [Hul09], p.371). Elle est obtenue en égalisant la sensibilité de second ordre du portefeuille de couverture avec celle de la valeur de l'option. Il s'en suit que les quantités d'options C et d'actifs à détenir à la date t_i vallent respectivement (dans le cadre unidimensionnel)

$$\delta_{t_i}^C := \frac{\partial_S^2 u(t_i, S_{t_i})}{\partial_S^2 C(t_i, S_{t_i})}, \qquad \delta_{t_i} := \partial_S u(t_i, S_{t_i}) - \frac{\partial_S^2 u(t_i, S_{t_i})}{\partial_S^2 C(t_i, S_{t_i})} \partial_S C(t_i, S_{t_i}).$$

Cette spécification d'une stratégie constante par morceaux sur chaque intervalle $[t_i, t_{i+1}[$ mène à une valeur finale du portefeuille, V_T^N, qu'on espère plus proche de $g(S_T)$ comparée avec celle obtenue par SCD. Note objectif est de quantifier de combien la *tracking error* inhérente $\mathcal{E}_N^{\Delta\Gamma} := V_T^N - g(S_T)$ est réduite en fonction du nombre de dates de rebalancement. Plus précisément, nous cherchons à lier la vitesse de convergence (au sens $\mathbf{L}_2$) de $\mathcal{E}_N^{\Delta\Gamma}$ à la régularité du payoff, tout en examinant l'impact du choix des dates de rebalancement.

2.3.3 Cadre

Dans un cadre multidimensionnel, nous considérons seulement des mouvements browniens géométriques pour la modélisation de $S = (S^1, ..., S^d)$ (cependant, nous pensons que nos résultats sont valables pour une classe plus large de modèles d'EDS pour S). Notre modèle s'écrit alors comme $(1 \leq j \leq d)$

$$\begin{cases} S_0^j &= s_0^j, \\ \mathrm{d}S_t^j &= \mu_j S_t^j \mathrm{d}t + \sigma_j S_t^j \mathrm{d}\hat{W}_t^j, \end{cases}$$

où S_t^j, s_0^j et σ_j sont positifs, et $\hat{W} = (\hat{W}^1, ..., \hat{W}^d)$ est un mouvement brownien de dimension d, défini sur un espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$, où $\mathbb{P}$ est la probabilité historique, et $(\mathcal{F}_t)_{0\leq t\leq T}$ est la filtration naturelle augmentée de $\hat{W}$. Nous supposons que les composantes de $\hat{W}$ sont corrélées, avec $\langle \hat{W}^j, \hat{W}^k \rangle_t = \rho_{j,k} t$, et que la matrice $(\rho_{j,k})_{1\leq j,k\leq d}$ est de rang plein.
Concernant les instruments de couverture, nous proposons d'utiliser des Calls pour

gérer les gammas individuels, et des options d'échange (option *spread* avec *strike* nul) pour gérer les gammas croisés, avec une date de maturité commune $T_2 > T$. La fonction prix de chacun de ces instruments s'écrit:

$$C^{j,k}(t, S^j, S^k) := \mathbb{E}_{\mathbb{Q}}\left[e^{-r(T_2-t)}(S^k_{T_2} - K_{j,k}S^j_{T_2})_+ | S^j_t = S^j, S^k_t = S^k\right],$$

où $\mathbb{Q}$ est la probabilité risque-neutre. Puisque les modèles pour S sont des mouvements browniens géométriques corrélés, on peut tirer profit des formules fermées de Black-Scholes et de Margrabe pour de telles options (cf. Appendice 10.A.1). Nous désignons par $\delta^{j,k}_{t_i}$ le nombre de chacun de ces instruments qui doit être détenu entre les dates t_i et t_{i+1}.
Dans l'esprit de la section 2.1, nous cherchons à utiliser la régularité fractionnaire de $g \in \mathbf{L}_{2,\alpha}$ (avec $\mathbf{L}_{2,\alpha}$, $K^\alpha(g)$ et $V_{t,T}(g)$ definis par (2.6), (2.7) et (2.8)) pour estimer l'erreur $(\mathbb{E}|\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi)|^2)^{1/2}$ et trouver une grille en temps permettant d'atteindre la vitesse optimale.
De plus, nous supposons que $\mathbb{E}_{\mathbb{P}}|g(S_T)|^{2p_0} < \infty$ pour un certain $p_0 > 1$, et nous définissons

$$K^{\alpha,p_0}(g) := \left(\mathbb{E}_{\mathbb{P}}|g(S_T)|^{2p_0}\right)^{\frac{1}{p_0}} + \sup_{t\in[0,T)} \frac{V_{t,T}(g)}{(T-t)^\alpha}.$$

2.3.4 Principaux résultats et outils

Un outil clé de notre étude est l'utilisation de la propriété de martingale (sous la probabilité risque-neutre) satisfaite par le processus de prix actualisé (que nous notons par $\bar{u}(t)$) et par ses dérivées (que nous notons, à des constantes multiplicatives près, par $\bar{u}^{(1)}_l(t)$ pour la dérivée première par rapport à S^l, par $\bar{u}^{(2)}_{l,m}(t)$ pour la dérivée seconde par rapport à (S^l, S^m), et par $\bar{u}^{(3)}_{l,m,n}(t)$ pour la dérivée troisième par rapport à (S^l, S^m, S^n)).
Nous commençons par décomposer la *tracking error* (actualisée) comme somme d'intégrales stochastiques triples par rapport à W, où $(W^j_t := \hat{W}^j_t + \lambda_j t)_{1\le j\le d}$ est un $\mathbb{Q}$-mouvement brownien (les corrélations restant inchangées). Cette décomposition consiste en un terme principal, lié aux dérivées troisièmes de la fonction prix de l'option, et des termes résiduels.

Théorème 2.3.1. *Supposons que* $\mathbb{E}_{\mathbb{P}}|g(S_T)|^{2p_0} < \infty$ *pour un certain* $p_0 > 1$. *On a*

$$\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi) = -\sum_{i=0}^{N-1}\sum_{l,m,n=1}^{d}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\left(\bar{u}^{(3)}_{l,m,n}(r) + R^{i,(3)}_{l,m,n}(r)\right)\mathrm{d}W^n_r\mathrm{d}W^m_s\mathrm{d}W^l_t, \tag{2.13}$$

où

$$
\begin{aligned}
R^{i,(3)}_{l,m,n}(t) = &\left(\sigma_m \bar{u}^{(2)}_{l,m}(t)\mathbb{1}_{n=m} + \sigma_l \bar{u}^{(2)}_{l,m}(t)\mathbb{1}_{n=l} + \sigma_l \bar{u}^{(2)}_{l,n}(t)\mathbb{1}_{m=l}\right) \\
&+ \sigma_l^2 \left(\bar{u}^{(1)}_l(t) - \frac{\bar{S}^l_t}{\bar{S}^l_{t_i}}\bar{u}^{(1)}_l(t_i)\right)\mathbb{1}_{n=m=l} \\
&- \sum_{0\le j<k\le d} \delta^{j,k}_{t_i} \bar{C}^{j,k,(3)}_{l,m,n}(t) \\
&- \sum_{0\le j<k\le d} \delta^{j,k}_{t_i} \left(\sigma_m \bar{C}^{j,k,(2)}_{l,m}(t)\mathbb{1}_{n=m} + \sigma_l \bar{C}^{j,k,(2)}_{l,m}(t)\mathbb{1}_{n=l} + \sigma_l \bar{C}^{j,k,(2)}_{l,n}(t)\mathbb{1}_{m=l}\right) \\
&- \sigma_l^2 \sum_{0\le j<k\le d} \delta^{j,k}_{t_i} \left(\bar{C}^{j,k,(1)}_l(t) - \frac{\bar{S}^l_t}{\bar{S}^l_{t_i}}\bar{C}^{j,k,(1)}_l(t_i)\right)\mathbb{1}_{n=m=l}. \qquad (2.14)
\end{aligned}
$$

Commentons, de manière schématique, les conséquences de (2.13). Posons $T^{i,(3)}_{l,m,n}(r) := \bar{u}^{(3)}_{l,m,n}(r) + R^{i,(3)}_{l,m,n}(r)$. A première vue, si $T^{i,(3)}_{l,m,n}(r)$ avait un moment $\mathbf{L}_2(\mathbb{Q})$ uniformément borné en r, on déduirait que

$$
\begin{aligned}
\mathbb{E}_{\mathbb{Q}}|\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi)|^2 &\le \sum_{i=0}^{N-1}\sum_{l,m,n=1}^{d}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\sup_{0\le r\le T}|T^{i,(3)}_{l,m,n}(r)|^2_{\mathbf{L}_2}\mathrm{d}r\mathrm{d}s\mathrm{d}t \\
&\le C \sup_{i,l,m,n;\ 0\le r\le T}|T^{i,(3)}_{l,m,n}(r)|^2_{\mathbf{L}_2}|\pi|^2,
\end{aligned}
$$

c.-à-d. la *tracking error* aurait une norme $\mathbf{L}_2(\mathbb{Q})$ de l'ordre du pas de temps $|\pi|$. En particulier, N dates uniformes de rebalancement assureraient une vitesse de convergence optimale. Toutefois, cette argumentation n'est pas valable dans notre cadre, car les moments de $T^{i,(3)}_{l,m,n}(r)$ peuvent exploser quand r tend vers T (en effet, pour un payoff irrégulier, les grècques peuvent tendre vers $\pm\infty$ quand r tend vers T). Ainsi, un aspect important de notre analyse de l'erreur de SCDG consiste à comprendre le comportement de $|T^{i,(3)}_{l,m,n}(r)|^2_{\mathbf{L}_2}$ lorsque $r \to T$. Nous en établissons des estimations fines à travers la régularité fractionnaire de g (cf. Proposition 2.3.1 et Corollaire 2.3.1).

Proposition 2.3.1. *On a les estimations suivantes:*

- $$\mathbb{E}_{\mathbb{P}}\left|\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi)\right|^2 \le C\sum_{i=0}^{N-1}\sum_{l,m,n=1}^{d}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(3)}_{l,m,n}(r) + R^{i,(3)}_{l,m,n}(r)\right|^2\mathrm{d}r\mathrm{d}s\mathrm{d}t. \qquad (2.15)$$

- *Pour $l,m,n = 1...d$ et $0 \le t < T$,*
$$\mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(3)}_{l,m,n}(t)\right|^2 \le C\frac{V_{t,T}(g)}{(T-t)^3}. \qquad (2.16)$$

- *Il existe une constante positive* $\pi^{\text{threshold}}$ *(dependant de* p_0*,* T *et* T_2*) telle que, si* $|\pi| \leq \pi^{\text{threshold}}$*, alors, pour* $0 \leq t_i \leq t < t_{i+1} \leq T$*,*

$$\mathbb{E}_{\mathbb{P}} \left| R^{i,(3)}_{l,m,n}(t) \right|^2 \leq C \frac{V_{t,T}(g) + \left(\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0} \right)^{\frac{1}{p_0}}}{(T-t)^2}. \tag{2.17}$$

L'estimation suivante en est une conséquence directe.

Corollaire 2.3.1. *Soit* $\alpha \in (0,1]$*. Supposons que* $g \in \mathbf{L}_{2,\alpha}$ *et que* $\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0} < \infty$ *pour un certain* $p_0 > 1$*. Alors, si* $|\pi| \leq \pi^{\text{threshold}}$*, et pour* $0 \leq t < T$*,*

$$\mathbb{E}_{\mathbb{P}} \left| \bar{u}^{(3)}_{l,m,n}(t) + R^{i,(3)}_{l,m,n}(t) \right|^2 \leq C \frac{K^{\alpha,p_0}(g)}{(T-t)^{3-\alpha}}. \tag{2.18}$$

Afin d'obtenir l'estimation (2.16), nous utilisons la représentation usuelle des grècques en termes de poids issus du calcul de Malliavin. Il faut juste exprimer cette représentation sous la probabilité risque-neutre $\mathbb{P}$.
Cependant, l'estimation (2.17) est plus délicate. En effet, on peut voir à travers l'expression (2.14) qu'on a besoin d'estimer, dans $\mathbf{L}_p$, des quotients de grècques des instruments de couverture. Dans le cas où le numérateur et le dénominateur sont évalués à la même date t, leurs parties exponentielles gaussiennes se simplifient (en utilisant la formule de Margrabe). La difficulté apparaît dès que les deux dates sont différentes, comme pour $\frac{\bar{C}^{j,k,(2)}_{l,m}(t)}{\bar{C}^{j,k,(2)}_{l,m}(t_i)}$, par exemple (avec $t_i \leq t \leq t_{i+1}$). En effet, si $t - t_i$ est trop grand, la quantité $\mathbb{E}_{\mathbb{P}}[|\bar{C}^{j,k,(2)}_{l,m}(t)|^p | \mathcal{F}_{t_i}]$ devient à queue trop lourde par rapport au dénominateur $\bar{C}^{j,k,(2)}_{l,m}(t_i)$, et le quotient explose. Ainsi, nous devons faire attention à la condition que la taille $|\pi|$ de la grille en temps doit satisfaire pour éviter cette explosion. Ceci mène à un seuil maximal qu'on désigne par $\pi^{\text{threshold}}$.
Les estimations (2.15) et (2.18) nous permettent d'établir notre résultat principal. La grille en temps $\pi^{(\beta)}$ est toujours définie par (2.9).

Théorème 2.3.2. *Soit* $\alpha \in (0,1]$*. Supposons que* $g \in \mathbf{L}_{2,\alpha}$ *et que* $\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0} < \infty$ *pour un certain* $p_0 > 1$*.*

1. Grille régulière $\pi^{(1)}$*. Pour* N *assez grand pour assurer* $|\pi^{(1)}| = \frac{T}{N} \leq \pi^{\text{threshold}}$*, on a*

$$\left(\mathbb{E}_{\mathbb{P}} \left| \overline{\mathcal{E}}^{\Delta\Gamma}_N (g, \pi^{(1)}) \right|^2 \right)^{1/2} = \mathcal{O}(\frac{1}{N^{\alpha/2}}). \tag{2.19}$$

2. Grille non régulière $\pi^{(\beta)}$, $\beta \in (0,1)$*. En prenant* N *assez grand pour assurer* $|\pi^{(\beta)}| \leq \pi^{\text{threshold}}$*, on a*

$$\left(\mathbb{E}_{\mathbb{P}} \left| \overline{\mathcal{E}}^{\Delta\Gamma}_N (g, \pi^{(\beta)}) \right|^2 \right)^{1/2} = \begin{cases} \mathcal{O}(\frac{1}{N^{\frac{\alpha}{2\beta}}}) \text{ si } \beta \in (\frac{\alpha}{2}, 1), \\ \mathcal{O}(\frac{\sqrt{\log N}}{N}) \text{ si } \beta = \frac{\alpha}{2}, \\ \mathcal{O}(\frac{1}{N}) \text{ si } \beta \in (0, \frac{\alpha}{2}). \end{cases}$$

Nous signalons qu'avec la grille régulière, l'estimation (2.19) de la vitesse de convergence $N^{\alpha/2}$ est en générale optimale pour $\alpha < 1$ (car l'estimation (2.18) est optimale, cf. les discussions dans Geiss et Hujo [GH07] et dans Gobet et Makhlouf [GM08]). De plus, elle coïncide avec la vitesse de convergence de la stratégie de couverture en Delta: en d'autres termes, pour des payoffs non réguliers, la stratégie de couverture en Delta-Gamma, utilisée avec une grille régulière, n'améliore pas la vitesse de convergence de la *tracking error*. L'utilisation de grilles non régulières est cruciale pour l'augmentation de cette vitesse, pour toute régularité fractionnaire $\alpha \in (0, 1]$, pour atteindre la vitesse N. Ceci est confirmé par nos résultats numériques (voir Chapitre 10, section 10.7): à travers la figure 2.1, nous montrons, pour un Call digital (dont le degré de régularité α est égal à 0.5), qu'on peut améliorer la vitesse de convergence grâce à un bon choix de la grille en temps.

En conclusion, et comme application, nos résultats peuvent être utilisés dans le but

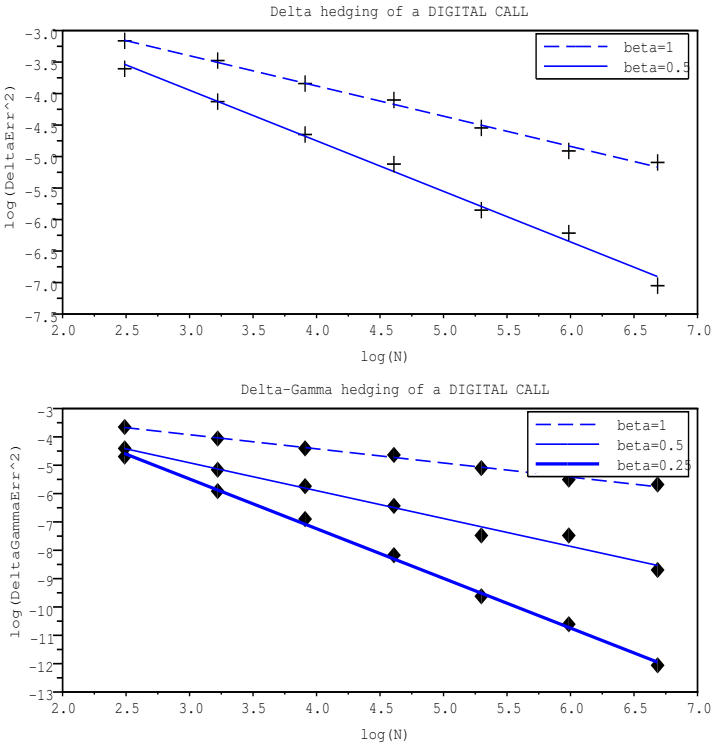

Figure 2.1: Pour un Call digital: en haut (stratégie de couverture en Delta), $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N^{\Delta}(g, \pi^{(\beta)})|^2)$ vs $\log(N)$. En bas (stratégie de couverture en Delta-Gamma), $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g, \pi^{(\beta)})|^2)$ vs $\log(N)$.

de réduire les risques inhérents à la couverture des options européennes.

Part I

$\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions

Chapter 3

Background results for BSDEs and motivations

Hereafter, W is a q-dimensional Brownian motion, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $(\mathcal{F}_t)_{0\leq t\leq T}$ (T is a fixed terminal time) is the natural filtration of W, augmented with $\mathbb{P}$-null sets.
We denote the conditional expectation $\mathbb{E}(X|\mathcal{F}_t)$ of a random variable X by $\mathbb{E}^{\mathcal{F}_t}(X)$.
We will mainly consider Markovian BSDEs with Lipschitz-continuous driver:

$$\left\{\begin{array}{ll} X_0 & = x_0, \\ \mathrm{d}X_t & = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \\ -\mathrm{d}Y_t & = f(t, X_t, Y_t, Z_t)\mathrm{d}t - Z_t\mathrm{d}W_t, \\ Y_T & = g(X_T), \end{array}\right. \tag{3.1}$$

with $X, x_0 \in \mathbb{R}^d$, $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times q}$ and $f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^{1\times d} \to \mathbb{R}$.

3.1 Existence and uniqueness results

The existence and uniqueness of square-integrable solutions to BSDE (3.1) were first investigated in [PP90] and in [KPQ97b]. Then, the assumptions have been relaxed, and the $\mathbf{L}_p$-integrability (for $p > 1$) has been studied. We give here a result stated by Briand *et al.* [BDH$^+$03], for the general BSDE

$$\left\{\begin{array}{ll} -dY_t & = f(t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T & = \xi. \end{array}\right. \tag{3.2}$$

Although this wider framework seems to be a bit too general to study (3.1), it will be useful in our analysis when exploding drivers will be involved (see Chapter 5, paragraph 5.8.2).

Assumption 3.1.1. **(i)** *For some $p > 1$,*

$$\mathbb{E}|\xi|^p + \mathbb{E}\left(\int_0^T |f(t,0,0)|\mathrm{d}t\right)^p < \infty.$$

(ii) *There exist constants $C_1 \geq 0$ and $C_2 \in \mathbb{R}$ such that, $\mathbb{P}$-a.s., for every $(t, y, y', z, z') \in [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{1\times q} \times \mathbb{R}^{1\times q}$,*

$$|f(t,y,z) - f(t,y,z')| \leq C_1|z - z'|,$$
$$\langle y - y', f(t,y,z) - f(t,y',z)\rangle \leq C_2|y - y'|^2.$$

(iii) *$\mathbb{P}$-a.s., for every $(t,z) \in [0,T]\times\mathbb{R}^{1\times q}$, the function $y \mapsto f(t,y,z)$ is continuous.*

(iv) *$\forall R > 0$, $\psi_R(t) := \sup_{|y|\leq R} |f(t,y,0) - f(t,0,0)| \in \mathbf{L}_1([0,T] \times \Omega, \mathrm{d}t \otimes \mathrm{d}\mathbb{P})$.*

The space $\mathcal{S}^p$ is defined to be the set of adapted and càdlàg processes Y such that

$$\mathbb{E}\big[\sup_{t\in[0,T]} |Y_t|^p\big] < +\infty,$$

and $\mathcal{M}^p$ the set of predictable processes Z such that

$$\mathbb{E}\big[\big(\int_0^T |Z_s|^2 ds\big)^{p/2}\big] < +\infty.$$

Theorem 3.1.1 (see Theorem 4.2 and Proposition 3.2 in [BDH$^+$03])**.** *Under Assumption 3.1.1, BSDE (3.2) has a unique solution in $\mathcal{S}^p \times \mathcal{M}^p$. Moreover, one has the following a priori estimate, for some constant c_p depending only on p, and for any $a \geq C_2 + \frac{C_1^2}{1\wedge(p-1)}$:*

$$\mathbb{E}(\sup_t e^{apt}|Y_t|^p) + \mathbb{E}\left(\int_0^T e^{2ar}|Z_r|^2\mathrm{d}r\right)^{\frac{p}{2}} \leq c_p[\mathbb{E}(e^{apT}|\xi|^p) + \mathbb{E}\left(\int_0^T e^{ar}|f(r,0,0)|\mathrm{d}r\right)^p].$$

Besides, the Y-component is clearly time continuous.

In particular, with a Lipschitz-continuous driver f, and if Assumption 3.1.1-(i) is fulfilled with $p = 2$ (and $\xi = g(X_T)$), BSDE (3.1) has a unique solution in $\mathcal{S}^2 \times \mathcal{M}^2$. Existence and uniqueness results have been stated for a non-Lipschitz driver f: see e.g. [BLM07] for f monotonic in y and non-Lipschitz in z, [Kob00] for f having quadratic growth in z and for bounded terminal conditions, and [BH06] for driver with quadratic growth and unbounded terminal condition (but with bounded exponential moment).

3.2 Applications

3.2.1 Finance

In finance, in a complete market model, and considering Itô dynamics for the risky assets prices $S_t^1, ..., S_t^d$ ($\mathrm{d}S_t^j = S_t^j(b_t^j \mathrm{d}t + \sum_{k=1}^d \sigma_t^{j,k} \mathrm{d}W_t^k)$), with predictable bounded processes $b = (b^1, ..., b^d)^*$ and $\sigma = (\sigma^{j,k})_{1 \leq j,k \leq d}$, and $W = (W^1, ..., W^d)^*$ is a standard Brownian motion), the price process of a positive contingent claim ξ, with maturity T, solves a linear BSDE (LBSDE for short).
Indeed, if one holds a *self-financing portfolio* (see definition in [KPQ97b]) whose value process is V, and invests an amount δ_t^j of the wealth V_t in the jth stock (the pair (V, δ), with $\delta = (\delta^1, ..., \delta^d)^*$, is called a *self-financing trading strategy*), then the aim is to hedge against ξ, i.e. to make a self-financing trading strategy (V, δ) such that $V_T = \xi$ (and (V, δ) is called a *hedging strategy*).
The *fair price* Y_t at time t of the hedgeable claim ξ is the smallest endowment needed at time t to hedge ξ:

$$Y_t := \inf\{y \geq 0 : \exists (V, \delta) \text{ a hedging strategy with initial value } V_t = y\}.$$

The following theorem shows that the proces Y satisfies a LBSDE with terminal condition ξ (under some assumptions on the coefficients b and σ). For the proof, see e.g. [KS91] and [KPQ97b].

Theorem 3.2.1. *Let ξ be a positive square-integrable contingent claim. There exists a hedging strategy (Y, δ) against ξ such that*

$$\begin{cases} \mathrm{d}Y_t &= r_t Y_t \mathrm{d}t + \delta_t^* \sigma_t \theta_t \mathrm{d}t + \delta_t^* \sigma_t \mathrm{d}W_t, \\ Y_T &= \xi \end{cases}$$

(r_t being the bond rate and θ is a risk premium defined by $b_t - r_t \mathbf{1} = \sigma_t \theta_t$, $\mathrm{d}\mathbb{P} \otimes \mathrm{d}t$-a.s., $\mathbf{1}$ is the vector whose every component is 1), and such that the portfolio value Y is the fair price of the claim.
Moreover, one has the following explicit expression a.s. for Y:

$$Y_t = \mathbb{E}^{\mathcal{F}_t}(H_T^t \xi), \tag{3.3}$$

where $(H_s^t)_{s \geq t}$ is the deflator started at time t, defined by

$$\begin{cases} \mathrm{d}H_s^t &= -H_s^t(r_s \mathrm{d}s + \theta_s^* \mathrm{d}W_s), \\ H_t^t &= 1. \end{cases}$$

Remark 3.2.1. *Such a process Y is not unique in general under the assumptions of the theorem. However, if one considers only square-integrable, self-financing strategies (with a square-integrable contingent claim), there exists a unique square-integrable hedging strategy.*
The representation (3.3) is equivalent to the usual risk-neutral rule (see Harrison and Pliska [HP81]).

Nonlinear BSDEs appear when the trading strategy (V, δ) is subject to additional constraints. Indeed, if the investor is allowed to borrow money at time t at an interest rate $R_t > r_t$, a nonlinear term is to be added to the driver, and a corresponding hedging strategy (Y, δ) against ξ satisfies

$$\begin{cases} \mathrm{d}Y_t &= r_t Y_t \mathrm{d}t + \delta_t^* \sigma_t \theta_t \mathrm{d}t - (R_t - r_t)(Y_t - \sum_{j=1}^d \delta_t^j)^- \mathrm{d}t + \delta_t^* \sigma_t \mathrm{d}W_t, \\ Y_T &= \xi. \end{cases}$$

In incomplete markets, where only some *primary securities* (say the first $\tilde{d}$ ones, $\tilde{d} < d$, instead of all the d securities) can be traded, the vector δ has to be constrainted to $(\delta^{\tilde{d}+1}, ..., \delta^d) = (0, ..., 0)$, which makes the related BSDE nonlinear also. Furthermore, always in incomplete markets, for the so-called Föllmer-Schweizer strategy, a tracking error $-\phi_t$ is introduced in order to measure the gap between ξ and V_T (this gap is equal to 0 in a complete market). The process ϕ is assumed to be a càdlàg semimartingale satisfying $\phi_0 = 0$. It turns out that the Föllmer-Schweizer strategy is the solution to a LBSDE (see Proposition 1.1 in [KPQ97b]).
In addition, in the continuous-time, non-deterministic case, recursive utilities are solutions of BSDEs. Duffie and Epstein [DE92] introduced two classes of recursive utilities Y:

$$-\mathrm{d}Y_t = f(t, c_t, Y_t)\mathrm{d}t - Z_t^* \mathrm{d}W_t$$

and

$$-\mathrm{d}Y_t = [f(c_t, Y_t) - A(Y_t)\frac{1}{2} Z_t^* Z_t]\mathrm{d}t - Z_t^* \mathrm{d}W_t,$$

where c_t is the consumption rate at time t and f is a concave *generator*. A more general class of recursive utilities, defined by BSDEs with concave generator f depending on t, c_t, Y_t and Z_t, has been studied in [KPQ97b].
BSDEs have other applications in stochastic zero-sum differential games (see [HL95]), and in the switching problem (called also the starting and stopping problem), which involves reflected BSDEs (RBSDEs for short), see [HJ07]. RBSDEs have been introduced by El Karoui *et al.* in [KKP$^+$97], and their development has been motivated by pricing American contingent claims (see [KPQ97a]).
More details on this section and on other applications of BSDEs can be found, for instance, in [KPQ97b] and in [KHM09].

3.2.2 Semilinear PDEs

In the Markovian case, it turns out that there is a link between BSDEs and semilinear parabolic partial differential equations, generalizing the well-known Feynman-Kac formula.
Let us illustrate this link. The framework is the following (it is similar to that of BSDE (3.1), with a SDE starting from x at time t):

$$\begin{cases} X_s^{t,x} &= x, \quad 0 \le s \le t, \\ \mathrm{d}X_s^{t,x} &= b(s, X_s^{t,x})\mathrm{d}s + \sigma(s, X_s^{t,x})\mathrm{d}W_s, \\ -\mathrm{d}Y_s^{t,x} &= f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\mathrm{d}s - Z_s^{t,x}\mathrm{d}W_s, \\ Y_T^{t,x} &= g(X_T^{t,x}). \end{cases} \tag{3.4}$$

The assumptions are the following:

Assumption 3.2.1. *i) The functions b and σ are uniformly Lipschitz continuous w.r.t. x.*

ii) The function f is uniformly Lipschitz continuous w.r.t. y and z.

iii) There exists a constant C such that, for any (t,x), $|b(t,x)|+|\sigma(t,x)| \leq C(1+|x|)$.

iv) There exists a constant C such that, for any (t,x,y,z), $|f(t,x,y,z)| + |g(x)| \leq C(1+|x|^p)$, for real $p \geq \frac{1}{2}$.

Let us consider the following system of parabolic PDEs:

$$\begin{cases} \partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),\nabla_x u(t,x)\sigma(t,x)) = 0 \text{ for } t < T, \\ u(T,x) = g(x), \end{cases} \tag{3.5}$$

where $\mathcal{L}$ is the second-order differential operator given by

$$\mathcal{L}u(t,x) := \sum_{i=1}^{d} b_i(t,x)\partial_{x_i} u(t,x) + \frac{1}{2}\sum_{i,j=1}^{d} [\sigma\sigma^*]_{i,j}(t,x)\partial^2_{x_i,x_j} u(t,x).$$

If one supposes that v is the unique solution of PDE (3.5) of class $\mathcal{C}^{1,2}$ and such that $|v(t,x)|+|\sigma(t,x)^*\nabla_x v(t,x)| \leq C(1+|x|)$, then $v(t,x) = Y_t^{t,x}$, where $(Y_s^{t,x}, Z_s^{t,x})_{t\leq s\leq T}$ is the unique solution of BSDE (3.4), and $(Y_s^{t,x}, Z_s^{t,x}) = (v(s,X_s^{t,x}), \nabla_x v(s,X_s^{t,x})\sigma(s,X_s^{t,x}))$, for $t \leq s \leq T$. This result was stated by Pardoux and Peng [PP92], and readily follows from an application of Itô's rule.
Conversely, if f and g are uniformly continuous w.r.t. x, then the function $(t,x) \mapsto Y_t^{t,x}$ is a viscosity solution of PDE (3.5), whose definition is as follows.

Definition 3.2.1. *Let $u \in \mathcal{C}([0,T]\times\mathbb{R}^d)$ such that $u(T,x) = g(x), \forall x \in \mathbb{R}^d$.*

i) u is said to be a viscosity subsolution (resp. supersolution) of PDE (3.5) if, for any $(t,x) \in [0,T]\times\mathbb{R}^d$ and $\phi \in \mathcal{C}^{1,2}([0,T]\times\mathbb{R}^d)$ such that $\phi(t,x) = u(t,x)$ and (t,x) is a maximum (resp. minimum) of $u-\phi$,

$$\begin{aligned} &-\partial_t\phi(t,x) - \mathcal{L}\phi(t,x) - f(t,x,\phi(t,x),\nabla_x\phi(t,x)\sigma(t,x)) \leq 0 \\ (\textit{resp. } &-\partial_t\phi(t,x) - \mathcal{L}\phi(t,x) - f(t,x,\phi(t,x),\nabla_x\phi(t,x)\sigma(t,x)) \geq 0). \end{aligned}$$

ii) u is a viscosity solution of PDE (3.5) if it is both a viscosity subsolution and a viscosity supersolution of PDE (3.5).

For the following result, we refer to [PP92] or to [KPQ97b]).

Theorem 3.2.2. *Suppose that Assumption 3.2.1 holds, and that f and g are uniformly continuous w.r.t. x. Then, the function u defined by $u(t,x) := Y_t^{t,x}$ is a viscosity solution of (3.5).*

3.3 $\mathbf{L}_2$-time regularity of Z

In the sequel, we focus on the markovian setting:

$$\begin{cases} X_0 &= x_0, \\ \mathrm{d}X_t &= b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \\ -\mathrm{d}Y_t &= f(t, X_t, Y_t, Z_t)\mathrm{d}t - Z_t\mathrm{d}W_t, \\ Y_T &= g(X_T). \end{cases} \tag{3.6}$$

For a time partition $\pi = \{0 = t_0 < \cdots < t_i < \cdots < t_N = T\}$ of $[0, T]$, the $\mathbf{L}_2$-time regularity of Z is defined by

$$\mathcal{E}(Z, \pi) := \sum_{i=0}^{N-1} \mathbb{E} \int_{t_i}^{t_{i+1}} |Z_t - \bar{Z}_{t_i}|^2 dt, \tag{3.7}$$

where

$$\bar{Z}_{t_i} := \frac{1}{t_{i+1} - t_i} \mathbb{E}^{\mathcal{F}_{t_i}} \int_{t_i}^{t_{i+1}} Z_s ds \tag{3.8}$$

is the projection of $(Z_s)_{t_i \le s \le t_{i+1}}$ on the space of $\mathcal{F}_{t_i}$-measurable random variables, according to the scalar product $< u, v >= \mathbb{E} \int_{t_i}^{t_{i+1}} u_s v_s ds$.
Besides, we will consider the following backward dynamic programming equation (for $0 \le i < N$) associated to (3.6):

$$\begin{cases} Y^\pi_{t_i} &= \mathbb{E}^{\mathcal{F}_{t_i}}(Y^\pi_{t_{i+1}} + (t_{i+1} - t_i) f(t_i, X^\pi_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})), \\ Z^\pi_{t_i} &= \frac{1}{(t_{i+1} - t_i)} \mathbb{E}^{\mathcal{F}_{t_i}}(Y^\pi_{t_{i+1}} (W_{t_{i+1}} - W_{t_i})^*), \end{cases} \tag{3.9}$$

where * denotes the transposition, $Y^\pi_{t_N} := g(X^\pi_T)$ and $(X^\pi_{t_i})_{0 \le i \le N}$ is the Euler approximation of the forward component $(X_t)_{0 \le t \le T}$.
It is known that when one wants to study the convergence of (3.9) for the purpose of the numerical simulation of BSDE (3.2), one has to examine the regularity of Z (see section 3.4).
In the framework of the studies devoted to the path regularity and discretization errors of the Z-component, Zhang[Zha04] studied not exactly $\mathcal{E}(Z, \pi)$, but the similar quantity

$$\tilde{\mathcal{E}}(Z, \pi) := \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \{|Z_s - Z_{t_k}|^2 + \left|Z_s - Z_{t_{k+1}}\right|^2\} ds.$$

Zhang showed that, when the terminal function g is Lipschitz continuous (and more generally if g is an L^∞-Lipschitz continuous functional, see [Zha04]), and under some regularity assumptions on the different coefficients, $\tilde{\mathcal{E}}(Z, \pi)$ is of order the time mesh size $|\pi|$. Here is the theorem.

Assumption 3.3.1. *For some constant K,*

(i) $f, b, \sigma \in \mathcal{C}^{1/2,1}$, and, for $h = f, b$ and σ, $|h(t, x) - h(s, y)| \le K(|t - s|^{\frac{1}{2}} + |x - y|)$.

(ii) $\sup_{t\in[0,T]}\{|b(t,0)|+|\sigma(t,0)|+|f(t,0,0,0)|\}+|g(0)|\leq K$.

(iii) g *is Lipschitz continuous and* $|g(x)-g(y)|\leq K|x-y|$.

Theorem 3.3.1. *Under assumption 3.3.1, and if Z is càdlàg, then, there exists a positive constant C such that, for any partition π of $[0,T]$, the following estimate holds:*

$$\tilde{\mathcal{E}}(Z,\pi)=\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\{|Z_s-Z_{t_k}|^2+\left|Z_s-Z_{t_{k+1}}\right|^2\}ds\leq C\left(1+|x_0|^2\right)|\pi|.$$

Two types of sufficient conditions for the path regularity of Z are given by Zhang: (i) If $d=q$ (i.e. $dim(X)=dim(W)$) and $\sigma\sigma^*\geq\delta I_d$ for some constant $\delta>0$ (ellipticity condition) and g is Lipschitz continuous, then Z admits a càdlàg version. (ii) If, in addition to the assumption 3.3.1, σ is uniformly Lipschitz continuous w.r.t. t, then Z is càdlàg.
Moreover, Zhang gives some other estimates for (X,Y,Z):

Theorem 3.3.2. *Assume Assumption 3.3.1. Then, for all $p\geq 2$, there exists a constant C_p, depending only on T, K and p, such that*

$$\|Z\|_{\mathcal{S}^p}:=\left(\mathbb{E}\big[\sup_{t\in[0,T]}|Z_t|^p\big]\right)^{\frac{1}{p}}\leq C_p(1+|x_0|).$$

Moreover, there exists a positive constant C (depending only on T, K, but not on the partition π) such that, for any time partition π,

$$\max_{1\leq i\leq N}\sup_{t\in(t_{i-1},t_i]}\mathbb{E}\{|X_t-X_{t_{i-1}}|^2+|Y_t-Y_{t_{i-1}}|^2\}\leq C(1+|x_0|^2)|\pi|.$$

3.4 Application to time discretization of BSDEs

Zhang has given an application of these different estimates to the study of the rate of convergence of a numerical approximation of BSDEs. The numerical (still explicit) scheme that he considered was a continuous version of (3.9): for $i=N-1,...,0$

$$\begin{cases}\tilde{Y}^\pi_{t_N} &= \xi^\pi,\\ \tilde{Y}^\pi_t &= \tilde{Y}^\pi_{t_{i+1}}+(t_{i+1}-t)f(t_{i+1},X^\pi_{t_{i+1}},\tilde{Y}^\pi_{t_{i+1}},\tilde{Z}^{\pi,1}_{t_{i+1}})-\int_t^{t_{i+1}}\tilde{Z}^\pi_r dW_r,\ t\in[t_i,t_{i+1}),\end{cases}\tag{3.10}$$

where $\xi^\pi\in\mathbf{L}_2(\mathcal{F}_T)$ is an approximation of $\xi=g(X_T)$, $\tilde{Z}^{\pi,1}_{t_i}:=\frac{1}{\Delta t_i}\mathbb{E}\left\{\int_{t_i}^{t_{i+1}}\tilde{Z}^\pi_r dr|\mathcal{F}_{t_i}\right\}$ and $\Delta t_i:=t_{i+1}-t_i$.
By Itô isometry, one has

$$\tilde{Z}^{\pi,1}_{t_i}=\frac{1}{\Delta t_i}\mathbb{E}((\tilde{Y}^\pi_{t_{i+1}}+\Delta t_i f(t_{i+1},X^\pi_{t_{i+1}},\tilde{Y}^\pi_{t_{i+1}},\tilde{Z}^{\pi,1}_{t_{i+1}}))(W_{t_{i+1}}-W_{t_i})^*|\mathcal{F}_{t_i}).$$

Then, one neglects the term "$\Delta t_i f(\dots)$" in the above equality to recognize $Z^\pi_{t_i}$ defined by (3.9). Besides, one writes

$$\tilde{Y}^\pi_{t_i} = \mathbb{E}(\tilde{Y}^\pi_{t_{i+1}} + \Delta t_i f(t_{i+1}, X^\pi_{t_{i+1}}, \tilde{Y}^\pi_{t_{i+1}}, \tilde{Z}^{\pi,1}_{t_{i+1}})|\mathcal{F}_{t_i}).$$

Since $\tilde{Z}^{\pi,1}_{t_{i+1}}$ is close to $Z^\pi_{t_i}$, one writes

$$\tilde{Y}^\pi_{t_i} \simeq \mathbb{E}(\tilde{Y}^\pi_{t_{i+1}} + \Delta t_i f(t_{i+1}, X^\pi_{t_{i+1}}, \tilde{Y}^\pi_{t_{i+1}}, Z^\pi_{t_i})|\mathcal{F}_{t_i})$$

to recognize $Y^\pi_{t_i}$ also defined by (3.9).

3.4.1 The key role of the $\mathbf{L}_2$-time regularity of Z

In the proof of the convergence of (3.10), Zhang proved the following intermediate estimate, that emphasizes the importance of the quantity $\tilde{\mathcal{E}}(Z,\pi)$ in the approximation of BSDEs. Note that, since we consider an Euler scheme for the forward component X, the related approximation error of X is of order $|\pi|^{\frac{1}{2}}$.

Proposition 3.4.1. *Assume (i) and (ii) of Assumption 3.3.1. Suppose moreover that there exists a constant $K > 0$ such that, $\forall i = 1...N, t_i - t_{i-1} \geq \frac{|\pi|}{K}$ (the partition π is then called K-uniform). Then*

$$\begin{aligned} e(\tilde{Y}^\pi - Y, \tilde{Z}^\pi - Z, \pi) &:= \sup_{0\leq i\leq N} \mathbb{E}(\tilde{Y}^\pi_{t_i} - Y_{t_i})^2 + \mathbb{E}\int_0^T |\tilde{Z}^\pi_r - Z_r|^2 dr \\ &\leq C(\tilde{\mathcal{E}}(Z,\pi) + \mathbb{E}|\xi - \xi^\pi|^2 + (1+|x_0|^2)|\pi|). \end{aligned}$$

However, the above result only holds for K-uniform time grids in the paper of Zhang. In fact, this condition is not needed to obtain the same kind of result. Either in [BT04] or in [GL06], the authors have showed it holds true for the approximation (Y^π, Z^π), with any partition π with size $|\pi| \leq \frac{C}{N}$.

Proposition 3.4.2. *Assume (i) and (ii) of Assumption 3.3.1, and that there exists a constant C such that $|\pi| \leq \frac{C}{N}$. Then*

$$\begin{aligned} e(Y^\pi - Y, Z^\pi - Z, \pi) &:= \sup_{0\leq i\leq N} \mathbb{E}(Y^\pi_{t_i} - Y_{t_i})^2 + \sum_{i=0}^{N-1} \mathbb{E}\int_{t_i}^{t_{i+1}} |Z^\pi_{t_i} - Z_t|^2 dt \\ &\leq C(\mathcal{E}(Z,\pi) + \mathbb{E}|\xi - \xi^\pi|^2 + (1+|x_0|^2)|\pi|). \end{aligned}$$

Consequently, if $\xi^\pi = g(X^\pi_T)$ and g is Lipschitz-continuous, then (from Theorem 3.3.1), one has

$$e(Y^\pi - Y, Z^\pi - Z, \pi) \leq C(1+|x_0|^2)|\pi|.$$

Here is the proof from [GL06]. For the sake of notational simplicity, we present it in a one-dimensional setting, but the proof is the same in the multidimensional case.

Proof. One has

$$Y_{t_i}-Y^\pi_{t_i}=\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)+\mathbb{E}^{\mathcal{F}_{t_i}}\left(\int_{t_i}^{t_{i+1}}\left\{f(s,X_s,Y_s,Z_s)-f(t_i,X^\pi_{t_i},Y^\pi_{t_{i+1}},Z^\pi_{t_i})\right\}ds\right).$$

Using Young's inequality (i.e. $(a+b)^2\leq(1+\gamma|\pi|)a^2+(1+\frac{1}{\gamma|\pi|})b^2$, with a positive parameter γ to be chosen later) together with Cauchy-Schwartz inequality and the Lipschitz property of f, one obtains

$$\begin{aligned}\mathbb{E}|Y_{t_i}-Y^\pi_{t_i}|^2 &\leq (1+\gamma|\pi|)\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)\right|^2\\ &\quad+(|\pi|+\frac{1}{\gamma})\mathbb{E}\int_{t_i}^{t_{i+1}}\left|f(s,X_s,Y_s,Z_s)-f(t_i,X^\pi_{t_i},Y^\pi_{t_{i+1}},Z^\pi_{t_i})\right|^2ds\\ &\leq (1+\gamma|\pi|)\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)\right|^2\\ &\quad+C(|\pi|+\frac{1}{\gamma})\left(|\pi|^2+\mathbb{E}\int_{t_i}^{t_{i+1}}\left|X_s-X^\pi_{t_i}\right|^2ds\right)\\ &\quad+C(|\pi|+\frac{1}{\gamma})\left(\mathbb{E}\int_{t_i}^{t_{i+1}}\left|Y_s-Y^\pi_{t_{i+1}}\right|^2ds+\mathbb{E}\int_{t_i}^{t_{i+1}}\left|Z_s-Z^\pi_{t_i}\right|^2ds\right).\end{aligned}\tag{3.11}$$

Now, since $\bar{Z}_{t_i}$ is the projection of $(Z_s)_{t_i\leq s\leq t_{i+1}}$ on the space of $\mathcal{F}_{t_i}$-measurable random variables, according to the scalar product $<u,v>=\mathbb{E}\int_{t_i}^{t_{i+1}}u_sv_sds$, Pythagore's theorem yield

$$\mathbb{E}\int_{t_i}^{t_{i+1}}\left|Z_s-Z^\pi_{t_i}\right|^2ds=\mathbb{E}\int_{t_i}^{t_{i+1}}\left|Z_s-\bar{Z}_{t_i}\right|^2ds+|\pi|\mathbb{E}\left|\bar{Z}_{t_i}-Z^\pi_{t_i}\right|^2.\tag{3.12}$$

From the definitions of $\bar{Z}_{t_i}$ (3.8) and $Z^\pi_{t_i}$ (3.9),

$$\bar{Z}_{t_i}-Z^\pi_{t_i}=\frac{1}{|\pi|}\mathbb{E}^{\mathcal{F}_{t_i}}\left\{\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)\Delta W_i\right\}+\frac{1}{|\pi|}\mathbb{E}^{\mathcal{F}_{t_i}}\left\{\left(\int_{t_i}^{t_{i+1}}f(s,X_s,Y_s,Z_s)ds\right)\Delta W_i\right\},$$

where we set $\Delta W_i:=W_{t_{i+1}}-W_{t_i}$.
One applies Cauchy-Schwartz inequality to both terms to have (denoting $\mathbb{E}^{\mathcal{F}_t}|X-\mathbb{E}^{\mathcal{F}_t}X|^2$ by $\mathbb{V}ar^{\mathcal{F}_t}$)

$$\begin{aligned}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left\{\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)\Delta W_i\right\}\right|^2 &\leq \mathbb{V}ar^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)\mathbb{V}ar^{\mathcal{F}_{t_i}}(\Delta W_i)\\ &\leq|\pi|\left\{\mathbb{E}^{\mathcal{F}_{t_i}}\left|Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right|^2-\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)\right|^2\right\},\end{aligned}$$

and

$$\begin{aligned}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left\{\left(\int_{t_i}^{t_{i+1}}f(s,X_s,Y_s,Z_s)ds\right)\Delta W_i\right\}\right|^2 &\leq|\pi|\mathbb{E}^{\mathcal{F}_{t_i}}\left(\int_{t_i}^{t_{i+1}}f(s,X_s,Y_s,Z_s)ds\right)^2\\ &\leq|\pi|^2\mathbb{E}^{\mathcal{F}_{t_i}}\int_{t_i}^{t_{i+1}}|f(s,X_s,Y_s,Z_s)|^2ds.\end{aligned}$$

Therefore

$$|\pi|\mathbb{E}\left|\bar{Z}_{t_i} - Z^\pi_{t_i}\right|^2 \leq 2\mathbb{E}\left\{\mathbb{E}^{\mathcal{F}_{t_i}}\left(\left|Y_{t_{i+1}} - Y^\pi_{t_{i+1}}\right|^2\right) - \left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}} - Y^\pi_{t_{i+1}}\right)\right|^2\right\}$$
$$+ 2|\pi|\mathbb{E}\int_{t_i}^{t_{i+1}} |f(s, X_s, Y_s, Z_s)|^2 ds. \tag{3.13}$$

Plugging (3.13) into (3.12), then into (3.11), one gets

$$\begin{aligned}
\mathbb{E}|Y_{t_i} - Y^\pi_{t_i}|^2 \leq{}& (1+\gamma|\pi|)\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}} - Y^\pi_{t_{i+1}}\right)\right|^2 \\
&+ C(|\pi| + \frac{1}{\gamma})\mathbb{E}\int_{t_i}^{t_{i+1}} \left|Z_s - \bar{Z}_{t_i}\right|^2 ds \\
&+ C(|\pi| + \frac{1}{\gamma})\left(|\pi|^2 + \int_{t_i}^{t_{i+1}} \mathbb{E}\left|Y_s - Y^\pi_{t_{i+1}}\right|^2 ds\right) \\
&+ C(|\pi| + \frac{1}{\gamma})\mathbb{E}\left\{\mathbb{E}^{\mathcal{F}_{t_i}}\left(\left|Y_{t_{i+1}} - Y^\pi_{t_{i+1}}\right|^2\right) - \left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}} - Y^\pi_{t_{i+1}}\right)\right|^2\right\} \\
&+ C(|\pi| + \frac{1}{\gamma})\left(|\pi|^2 + \mathbb{E}\int_{t_i}^{t_{i+1}} \left|X_s - X^\pi_{t_i}\right|^2 ds\right) \\
&+ C|\pi|(|\pi| + \frac{1}{\gamma})\mathbb{E}\int_{t_i}^{t_{i+1}} |f(s, X_s, Y_s, Z_s)|^2 ds.
\end{aligned}$$

Taking $\gamma = C$, the term $\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}} - Y^\pi_{t_{i+1}}\right)\right|^2$ above cancels.
Then, for $|\pi|$ small enough, and using $\sup_{t_i \leq s \leq t_{i+1}} \mathbb{E}|X_s - X^\pi_{t_i}|^2 \leq C|\pi|$,

$$\begin{aligned}
\mathbb{E}|Y_{t_i} - Y^\pi_{t_i}|^2 \leq{}& (1 + C|\pi|)\mathbb{E}\left|Y_{t_{i+1}} - Y^\pi_{t_{i+1}}\right|^2 + C|\pi|^2 \\
&+ C\mathbb{E}\int_{t_i}^{t_{i+1}} \left|Z_s - \bar{Z}_{t_i}\right|^2 ds + C\int_{t_i}^{t_{i+1}} \mathbb{E}\left|Y_s - Y_{t_{i+1}}\right|^2 ds \\
&+ C|\pi|\mathbb{E}\int_{t_i}^{t_{i+1}} |f(s, X_s, Y_s, Z_s)|^2 ds.
\end{aligned} \tag{3.14}$$

Discrete Gronwall's lemma gives (using $|\pi| \leq \frac{C}{N}$)

$$\begin{aligned}
\max_{0 \leq k \leq N} \mathbb{E}|Y_{t_i} - Y^\pi_{t_i}|^2 \leq{}& \mathbb{E}|\xi - \xi^\pi|^2 + C|\pi| \\
&+ C\sum_{i=0}^{N-1}\left(\mathbb{E}\int_{t_i}^{t_{i+1}} \left|Z_s - \bar{Z}_{t_i}\right|^2 ds + \int_{t_i}^{t_{i+1}} \mathbb{E}\left|Y_s - Y_{t_{i+1}}\right|^2 ds\right) \\
&+ C|\pi|\mathbb{E}\int_0^T |f(s, X_s, Y_s, Z_s)|^2 ds.
\end{aligned} \tag{3.15}$$

Now, from the BSDE satisfied by Y, one has

$$\begin{aligned}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\mathbb{E}\left|Y_s-Y_{t_{i+1}}\right|^2ds &\leq 2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\left((t_{i+1}-s)\int_s^{t_{i+1}}\mathbb{E}|f(u,X_u,Y_u,Z_u)|^2du\right)ds\\ &\quad+2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\left(\int_s^{t_{i+1}}\mathbb{E}\left|Z_u\right|^2du\right)ds\\ &\leq 2|\pi|\left(|\pi|\mathbb{E}\int_0^T|f(u,X_u,Y_u,Z_u)|^2du+\mathbb{E}\int_0^T\left|Z_u\right|^2du\right).\end{aligned}$$

Using the Lipschitz property of f, one has

$$|f(s,X_s,Y_s,Z_s)|^2\leq C(|f(s,0,0,0)|^2+|X_s|^2+|Y_s|^2+|Z_s|^2),$$

and since $(X,Y,Z)\in\mathcal{M}^2$, one obtains

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\mathbb{E}\left|Y_s-Y_{t_{i+1}}\right|^2ds\leq C|\pi| \tag{3.16}$$

and

$$\mathbb{E}\int_0^T|f(s,X_s,Y_s,Z_s)|^2ds\leq C. \tag{3.17}$$

Plugging (3.16) and (3.17) into (3.15) yields the estimate of $\max_{0\leq k\leq N}\mathbb{E}|Y_{t_i}-Y^\pi_{t_i}|^2$ in Proposition 3.4.2.
For the second estimate, one has, from (3.13),

$$\begin{aligned}|\pi|\sum_{i=0}^{N-1}\mathbb{E}\left|\bar{Z}_{t_i}-Z^\pi_{t_i}\right|^2 &\leq 2\sum_{i=0}^{N-1}\left\{\mathbb{E}\left|Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right|^2-\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)\right|^2\right\}\\ &\quad+2|\pi|\sum_{i=0}^{N-1}\mathbb{E}\int_{t_i}^{t_{i+1}}|f(s,X_s,Y_s,Z_s)|^2ds\\ &=2\sum_{i=1}^{N-1}\left\{\mathbb{E}\left|Y_{t_i}-Y^\pi_{t_i}\right|^2-\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_{i+1}}-Y^\pi_{t_{i+1}}\right)\right|^2\right\}\\ &\quad+2\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_N}-Y^\pi_{t_N}\right)\right|^2-2\mathbb{E}\left|\mathbb{E}^{\mathcal{F}_{t_i}}\left(Y_{t_1}-Y^\pi_{t_1}\right)\right|^2\\ &\quad+2|\pi|\mathbb{E}\int_0^T|f(s,X_s,Y_s,Z_s)|^2ds.\end{aligned}$$

From (3.14), we obtain

$$|\pi| \sum_{i=0}^{N-1} \mathbb{E} \left|\bar{Z}_{t_i} - Z^\pi_{t_i}\right|^2 \leq C|\pi| \sum_{i=0}^{N-1} \mathbb{E} \left|Y_{t_{i+1}} - Y^\pi_{t_{i+1}}\right|^2 + C|\pi| + C\mathbb{E}\left|\xi - \xi^\pi\right|^2$$
$$+ C \sum_{i=0}^{N-1} \mathbb{E} \int_{t_i}^{t_{i+1}} \left|Z_s - \bar{Z}_{t_i}\right|^2 ds + C \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left|Y_s - Y_{t_{i+1}}\right|^2 ds$$
$$+ 2|\pi|\mathbb{E} \int_0^T |f(s, X_s, Y_s, Z_s)|^2 ds.$$

The estimate that we have established for $\max_{0\leq k\leq N} \mathbb{E}|Y_{t_i} - Y^\pi_{t_i}|^2$, together with the estimates (3.16), (3.17), readily yield

$$|\pi| \sum_{i=0}^{N-1} \mathbb{E} \left|\bar{Z}_{t_i} - Z^\pi_{t_i}\right|^2 \leq C\mathcal{E}(Z, \pi) + C|\pi| + C\mathbb{E}\left|\xi - \xi^\pi\right|^2. \tag{3.18}$$

Recalling the equality (3.12), we have proved the estimate of $\sum_{i=0}^{N-1} \mathbb{E} \int_{t_i}^{t_{i+1}} |Z^\pi_{t_i} - Z_t|^2 dt$ in Proposition 3.4.2. □

3.4.2 Other related results

In [GLW06], it is shown that Proposition 3.4.2 can be extended to jump diffusion forward component X and Markovian BSDEs with a non Brownian filtration, of the form $-dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t - dL_t$, where L is a scalar càdlàg martingale orthogonal to W (with $L_0 = 0$). An additional contribution related to the approximation of X by X^π has to be put.
The previous results show the importance of controling the Z-component regularity. Similar studies have been done for BSDEs with reflection (see Chassagneux [Cha09], Bouchard and Chassagneux [BC08]), for BSDEs with random terminal time (see [BM08]) and for BSDEs with jump forward diffusion (see [BE08]). In all these works, the regularity of Z is crucial.
We mention another point of view regarding the weak convergence: in Gobet and Labart [GL07], the convergence in law of the process $\left(\sqrt{N}(Y^\pi_t - Y_t), \sqrt{N}(Z^\pi_t - Z_t)\right)_{0\leq t\leq T}$ is established ($(Y^\pi_t, Z^\pi_t)_{0\leq t\leq T}$ being a continuous version of $(Y^\pi_{t_i}, Z^\pi_{t_i})_{0\leq i\leq N}$).

3.5 Numerical computations of the dynamic programming equation (3.9)

One of the challenging issues is how to compute the conditional expectations in (3.9). In [BT04], the authors used a Malliavin calculus approach in order to compute the regression functions $\mathbb{E}^{\mathcal{F}_{t_i}}(Y^\pi_{t_{i+1}})$ and $\mathbb{E}^{\mathcal{F}_{t_i}}(Y^\pi_{t_{i+1}}(W_{t_{i+1}} - W_{t_i})^*)$.
For a real-valued mapping ρ and a vector random variable ξ independent of $\sigma(X^\pi_{t_i}, 1 \leq$

$i \leq N$), we set $R = \rho(X^\pi_{t_{i+1}}, \xi)a(W_{t_{i+1}} - W_{t_i})$ for some affine function a defined on $\mathbb{R}^q$ (so that R can be either $Y^\pi_{t_{i+1}}$ or $Y^\pi_{t_{i+1}}(W_{t_{i+1}} - W_{t_i})^*$). The main idea is to use a Malliavin integration by parts technique to rewrite the conditional expectation of R as a ratio of non-conditional expectations (see Corollary 3.1 in [BET04] or a slight extension of it, Theorem 6.1 in [BT04]):

$$\mathbb{E}\left(R|X^\pi_{t_i} = x\right) = \frac{\mathbb{E}\left(Q^R[h_i, \varphi](x)\right)}{\mathbb{E}\left(Q^1[h_i, \psi](x)\right)},$$

where

$$Q^R[h_i, \varphi](x) := H_x(X^\pi_{t_i})\rho(X^\pi_{t_{i+1}}, \xi)S^{h_i}[a(W_{t_{i+1}} - W_{t_i})\varphi(X^\pi_{t_i} - x)],$$

H_x is the Heaviside function ($H_x(y) = \prod_{j=1}^d \mathbb{1}_{x_j \leq y_j}$), φ and ψ are smooth localizing functions (bounded continuous functions mapping from $\mathbb{R}^d_+$ to $\mathbb{R}$, as well as their derivatives), $S^u[F] := \int_0^\infty F u_t . dW_t$ whenever it exists in the Skorohod sense, and the h_i's are suitable processes involving the derivative ∇X of X. Such a representation imposes an extra non degeneracy condition (ellipticity for instance).
Then, for each discretization time t_i, one computes the expectations using the usual Monte-Carlo method by simulating a set of M paths of $X^\pi_{t_i}$, all the sets of simulations (each relative to a discretization time) being independent.
Let $(\hat{Y}^\pi, \hat{Z}^\pi)$ be the approximation obtained by means of the above algorithm. Under some conditions on the localizing functions, the authors stated the following estimate of the regression error (Theorem 6.2 in [BT04]).

Theorem 3.5.1. *For $p > 1$,*

$$\limsup_{|\pi| \to 0} \max_{0 \leq i \leq N} |\pi|^{p + \frac{d}{4}} M^{\frac{1}{2}} \left\| \hat{Y}^\pi_{t_i} - Y^\pi_{t_i} \right\|^p_{\mathbf{L}_p} < \infty.$$

Taking $p = 2$ in the above theorem, the squared regression error is of order $N^{2+\frac{d}{4}} M^{-\frac{1}{2}}$.
However, as mentioned in [GL06], in the most favorable case where X is the geometric Brownian motion and the Malliavin weights are easy to compute, the complexity is $\mathcal{C} = \mathcal{O}(NM^2)$[1]. If one wants the squared regression error to be of order $|\pi|$ (which is the order of the squared discretization error $\left\| Y_{t_i} - Y^\pi_{t_i} \right\|^2_{\mathbf{L}_2}$ in general), the resulting order is $\mathcal{C}^{-\frac{1}{13+d}}$, which means that one has to multiply the computational time by 2^{13+d} if one wants to divide the squared error by 2.
In [GLW06], the authors propose an algorithm using an empirical regression method. It consists in approximating the conditional expectations in the dynamic programming equation (3.9), which are regression functions of $X^\pi_{t_i}$, by their projections on finite-dimensional bases. The projection coefficients are computed using the standard Monte-Carlo method with a fixed set of M simulations of X, and solving a

[1] Actually, using sorting techniques, coupled with a "divide to conquer" approach, one can reach a better complexity $\mathcal{C} = \mathcal{O}(NM(\ln M)^{(d-1)\vee 1})$.

least squares problem at each discretization time. The difference with [BT04] is that only one set of independent paths are needed, and no non degeneracy condition is required. The related squared error is of order $\tilde{\mathcal{C}}^{-\frac{1}{4+2d}}$, where $\tilde{\mathcal{C}}$ is the complexity of this algorithm (and, with a slightly modified algorithm, one can even reach the order $\tilde{\mathcal{C}}^{-\frac{1}{4+d}}$). In the case of geometric Brownian motion, this cost is cheaper than that obtained by [BT04] only if $d \leq 9$, but the approach of [GLW06] remains much more model-free, so is more suitable for the general models of the diffusion X.
In [BP03] and [Che97], quantization techniques have been explored, but the driver f does not depend on the z variable.

Remark 3.5.1. *Numerical schemes similar to (3.9) have been developed for forward-backward stochastic differential equations (FBSDEs, for short), that represent an SDE which is fully-coupled to a BSDE:*

$$\begin{cases} X_0 &= x_0, \\ dX_t &= b(t, X_t, Y_t, Z_t)dt + \sigma(t, X_t, Y_t, Z_t)dW_t, \\ -dY_t &= f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T &= g(X_T), \end{cases} \tag{3.19}$$

(see, for example, [DM06]). The corresponding convergence analysis relies on the use of the quasi-linear PDE associated to (3.19), and so requires strong regularity assumptions on the coefficients.

3.6 Approximation of the terminal condition

The second problem related to (3.9) concerns the numerical approximation of the terminal condition ξ, when $\xi = g(X_T)$. The natural approximation is $\xi^\pi = g(X_T^\pi)$, and one is interested in the resulting error $(\mathbb{E}|\xi - \xi^\pi|^2)^{\frac{1}{2}}$.
Let $X^{\pi,E}$ (resp. $X^{\pi,M}$) be the approximation of X by the Euler scheme (resp. Milstein scheme, if $d = 1$). If g is Lipschitz-continuous, the approximation error is of order $|\pi|^{\frac{1}{2}}$ with the Euler scheme, and of order $|\pi|$ with the Milstein scheme.
In [Avi09], the approximation error is studied in the one-dimensional framework for functions g of bounded variation. Let us recall the definition of this class of functions.

Definition 3.6.1. *For $x \in \mathbb{R}$, let $\Delta(x)$ be the set of all partitions $(x_j^n)_{j=0}^n$ such that $-\infty < x_0 < x_1 < ... < x_n = x$, for $n \geq 1$. The total variation function of a function g is defined by*

$$T_g(x) := \sup_{(x_j^n)\in\Delta(x)} \sum_{j=1}^n |g(x_j^n) - g(x_{j-1}^n)|.$$

The function g is said to be of bounded variation ($g \in BV$) if $V(g) := \lim_{x\to\infty} T_g(x)$ is finite ($V(g)$ is called the total variation of g).

Here is the result from [Avi09] (Corollary 4.2 and Theorem 5.4), for these two most common approximation schemes of X (under some (standard) assumptions on

the coefficients b and σ that we do not give here). However, this result holds only for one-dimensional diffusions.

Theorem 3.6.1. *Let $g \in BV$ and $1 \leq p < \infty$. Then, for $0 < \varepsilon < \frac{1}{2}$,*

$$\mathbb{E}|g(X_T) - g(X_T^{\pi,E})|^p \leq 3^p(\sup p_{X_T} \vee \sqrt{\sup p_{X_T}})V^p(g)C_\varepsilon|\pi|^{\frac{1}{2}-\varepsilon},$$

and for $0 < \varepsilon < 1$,

$$\mathbb{E}|g(X_T) - g(X_T^{\pi,M})|^p \leq 3^p(\sup p_{X_T} \vee \sqrt{\sup p_{X_T}})V^p(g)C'_\varepsilon|\pi|^{1-\varepsilon},$$

where p_{X_T} is the density of X_T, and C_ε and C'_ε depend on ε and the constants of the corresponding schemes.

This theorem obviously includes the case of indicator functions. The case of intermediate smoothness (between Lipschitz and step functions) is an open issue to our knowledge.

Chapter 4

Fractional regularity

4.1 First results on irregular terminal conditions

The results of Chapter 3, dealing with the rate of convergence of the dynamic programming equation (3.9), have been stated assuming always some regularity on the terminal function g, namely Lipschitz continuity. Our aim is to relax this assumption and study the related $\mathbf{L}_2$-regularity of Z, $\mathcal{E}(Z,\pi)$.
The first works in this direction were carried out, to our knowledge, by Gobet and Temam [GT01], but only in the specific framework of linear driver (w.r.t. y and z), which is technically equivalent to the null driver case ($f \equiv 0$). They considered general one-dimensional SDEs for X (under regularity and ellipticity assumptions on the coefficients):

$$\begin{cases} X_0 &= x_0, \\ dX_t &= \mu(X_t)X_t dt + \sigma(X_t)X_t dW_t, \end{cases}$$

but particular expressions of the terminal function g, in order to study the rate of convergence of the traking error when hedging, discretely in time, European options with standard payoffs: $g(x) = \mathbb{1}_{x \geq K}$ (digital option) or $g(x) = (x-K)_+^a$, for $a \in (0, \frac{1}{2})$. Their mathematical analysis is based upon explicit computations.
In the above paper, the authors considered only equidistant time nets, and showed that the optimal rates for $\mathcal{E}(z,\pi)$ associated to the above functions g are related to their Hölder regularity, and vary from $N^{\frac{1}{2}}$ to N. Here are their main results (Theorems 1 and 2 in [GT01]), when $\mu(x) = r$ (r is the interest rate and the error is computed under the risk-neutral measure).

Theorem 4.1.1. *Let $\pi = (T\frac{i}{N})_{i=0}^N$ be the equidistant time net.*

(i) For $g(x) = \mathbb{1}_{x \geq K}$, one has

$$\mathcal{E}(z,\pi) = \left(\frac{T}{N}\right)^{\frac{1}{2}} \frac{C_0}{4\sqrt{\pi}} K\sigma^3(K)e^{-2rT} p_T(x_0,K) + o\left(\frac{1}{N^{\frac{1}{2}}}\right),$$

where C_0 is a universal constant and $p_t(.,.)$ is the transition density of the process $(X_t)_t$.

(ii) For $g(x) = (x-K)_+^a$, $a \in (0, \frac{1}{2})$, *one has*

$$\mathcal{E}(z,\pi) = \left(\frac{T}{N}\right)^{\frac{1}{2}+a} \hat{C}_a K^{1+2a} \sigma^{3+2a}(K) e^{-2rT} p_T(x_0, K) + o\left(\frac{1}{N^{\frac{1}{2}+a}}\right),$$

where $\hat{C}_a$ *is a universal constant.*

4.2 Generalization using fractional regularity

The above results have been extended in a series of papers by Geiss *et al.*, and were essentially presented in [GH07]. In this paper, the authors studied the discrete $\mathbf{L}_2$-approximation of stochastic integrals, that arises for instance in finance when one replaces a continuously time-adjusted porfolio by a discretely time-adjusted one, and wants to estimate the resulting hedging error. They have characterized the rate of this error according to the number of the time net points by some weak regularity of the terminal function (related to the terminal value of the portfolio), expressed as a fractional smoothness by means of a real interpolation between the Sobolev space $\mathbf{D}_{1,2}(\gamma)$ and $\mathbf{L}_2(\gamma)$ (see definitions below), where γ is the standard Gaussian measure on the real line ($d\gamma(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$). In [GH07], the choice of the diffusion process is limited to either the standard Brownian motion or the geometric Brownian motion, so that all the calculations use explicit Hermite polynomials expansions in order to carry out a deterministic analysis of the expansion coefficients instead of using stochastic techniques. In [GG04], more general, but only one-dimensional SDEs are considered. Let us present the work of [GH07] in more details.
Following again the notations of [GH07], we denote by $\mathbf{L}_2(\gamma)$ the space of functions that are square-integrable with respect to the scalar Gaussian measure γ, and $(h_k)_{k=0}^{\infty}$ the orthonormal basis of Hermite polynomials of $\mathbf{L}_2(\gamma)$:

$$h_k(x) := \frac{(-1)^k}{\sqrt{k!}} e^{\frac{x^2}{2}} \frac{d^k}{dx^k}(e^{-\frac{x^2}{2}}).$$

The associated norm of a function $\phi = \sum_{k=0}^{\infty} \alpha_k h_k \in \mathbf{L}_2(\gamma)$ is then

$$\|\phi\|_{\mathbf{L}_2(\gamma)} := \left(\sum_{k=0}^{\infty} \alpha_k^2\right)^{\frac{1}{2}} < \infty.$$

The space $\mathbf{D}_{1,2}(\gamma)$ is defined by the subspace of $\mathbf{L}_2(\gamma)$ composed of all the functions $\phi = \sum_{k=0}^{\infty} \alpha_k h_k \in \mathbf{L}_2(\gamma)$ such that

$$\|\phi\|_{\mathbf{D}_{1,2}(\gamma)} := \left(\sum_{k=0}^{\infty} (k+1)\alpha_k^2\right)^{\frac{1}{2}} < \infty.$$

Definition 4.2.1. *1. Let* E_0 *and* E_1 *be compatible Banach spaces. For* $e \in E_0 + E_1$ *and* $\lambda > 0$, *the* K*-functional is defined by*

$$K(e, \lambda; E_0, E_1) := \inf\{\|e_0\|_{E_0} + \lambda \|e_1\|_{E_1} : e = e_0 + e_1\}.$$

2. *Given $\theta \in (0,1)$ and $q \in [1,\infty]$, the interpolation space $(E_0, E_1)_{\theta,q}$ of the spaces (E_0, E_1) is defined to be the subspace of $E_0 + E_1$ such that, for $e \in E_0 + E_1$,*

$$\|e\|_{(E_0,E_1)_{\theta,q}} := \left\| \lambda^{-\theta} K(e, \lambda; E_0, E_1) \right\|_{L_q((0,\infty),\frac{d\lambda}{\lambda})} < \infty.$$

3. *For $\theta \in (0,1)$ and $q \in [1,\infty]$, the Besov spaces are defined by*

$$\mathbf{B}^{\theta}_{2,q}(\gamma) := (\mathbf{L}_2(\gamma), \mathbf{D}_{1,2}(\gamma))_{\theta,q}.$$

The previous fractional smoothness will be shown to be in relation with the behaviour of an approximation $a(\xi;\pi)$, on a time net π, of $\xi = g(W_1)$ (that will be viewed as the sum of its expectation with a stochastic integral). Here is the definition.

Definition 4.2.2. *Let X be the Brownian motion or the geometric Brownian motion. Given a deterministic time net $\pi = (t_i)_{i=0}^N$ such that $0 = t_0 < ... < t_N = 1$ and a random variable $\xi \in \mathbf{L}_2(\Omega, \mathcal{F}_1, \mathbb{P})$,*

$$a_X(\xi;\pi) := \inf \left\| (\xi - \mathbb{E}\xi) - \sum_{i=1}^{N} v_{i-1}(X_{t_i} - X_{t_{i-1}}) \right\|_{\mathbf{L}_2}, \tag{4.1}$$

where the infimum is taken over all $\mathcal{F}_{t_{i-1}}$-measurable random variables v_{i-1} such that $\mathbb{E}|v_{i-1}(X_{t_i} - X_{t_{i-1}})|^2 < \infty$.

Let $\xi = g(X_1)$ and $(y_t, z_t)_{0\le t\le 1}$ be the solution of the linear BSDE:

$$y_t = g(X_1) - \int_t^T z_s dW_s.$$

Then $\xi - \mathbb{E}\xi = \xi - y_0 = \int_0^1 z_s dW_s$, and one has $v_i = \bar{z}_{t_i} := \frac{1}{t_{i+1}-t_i}\mathbb{E}^{\mathcal{F}_{t_i}} \int_{t_i}^{t_{i+1}} z_s ds$, which means that (see the definition (3.7) of $\mathcal{E}(Z,\pi)$)

$$|a_X(\xi;\pi)|^2 = \mathcal{E}(z,\pi).$$

Let $u(t,x) := \mathbb{E}(g(X_1)|X_t = x)$. Then

$$y_t = u(t, X_t), \quad z_t = \nabla_x u(t, X_t)\sigma(t, X_t).$$

The following theorem (Corollary 2.3 in [GH07]) characterizes the fractional regularity of the function g by the rate of decreasing of the expected conditional variance of $g(X_1)$ given X_t as t goes to T. The notation $A \sim_c B$ is an abbreviation for $\frac{A}{c} \le B \le cA$).

Theorem 4.2.1. *Assume X to be either the Brownian motion or the geometric Brownian motion. Let $q \in [1,\infty]$ and $\theta \in (0,1)$. Then*

$$\|g\|_{\mathbf{B}^{\theta}_{2,q}(\gamma)} \sim_c \|g\|_{\mathbf{L}_2(\gamma)} + \left\| (1-t)^{-\frac{\theta}{2}} \|g(X_T) - y_t\|_{\mathbf{L}_2} \right\|_{\mathbf{L}_q([0,1),\frac{dt}{1-t})}.$$

One of the main results of [GH07] shows the link between the fractional regularity of the function g and the behavior of $a_X(\xi;\pi)$ when π is an equidistant time net: it is stated in the following theorem. For a sequence $a = (a_k)_{k=0}^{\infty}$ of elements of a Banach space, $\|a\|_{l_q} := (\sum_{k=0}^{\infty} \|a_k\|^q)^{\frac{1}{q}}$.

Theorem 4.2.2. *Assume X to be either the Brownian motion or the geometric Brownian motion, and $T = 1$. Let $q \in [1,\infty]$ and $\theta \in (0,1)$. Then*

$$\|g\|_{\mathbf{B}^{\theta}_{2,q}(\gamma)} \sim_c \|g\|_{\mathbf{L}_2(\gamma)} + \left\| \left(N^{\frac{\theta}{2}-\frac{1}{q}} a_X(g(X_1);\pi_N) \right)_{N=1}^{\infty} \right\|_{l_q},$$

where $c = c(\theta) \geq 1$ and $\pi_N := (\frac{i}{N})_{i=0}^{N}$.

Taking $q = \infty$ in Theorem 4.2.2, the two following assertions are equivalent:

(i) $g \in \mathbf{B}^{\theta}_{2,\infty}(\gamma)$.

(ii) There exists $c > 0$ such that, for all $N \geq 1$ and $\pi_N = (\frac{i}{N})_{i=0}^{N}$,

$$a_X(g(X_1);\pi_N) \leq cN^{-\frac{\theta}{2}}.$$

The rate $N^{-\frac{\theta}{2}}$ in the assertion (ii) above can be improved by choosing non-equidistant time nets. Indeed, in [GG04] (Theorems 2.3, 2.5), it is shown that assertion (ii) implies

$$\sup_{N\geq 1} \sqrt{N} a_X(g(X_1);\pi_N^{\beta}) < \infty,$$

where, for some $\beta \in (0,\theta)$, $\pi_N^{\beta} = (t^{N,\beta})_{i=0}^{N}$ is defined by $t_i^{N,\beta} := 1 - (1 - \frac{i}{N})^{\frac{1}{\beta}}$, which makes the time knots more and more concentrated near the terminal time (equal to 1 here).

As stated in [GH07], it is not clear that one can obtain a natural characterization (and not just an implication as previously) of $\mathbf{B}^{\theta}_{2,q}(\gamma)$ when $q \neq 2$ by time nets realizing the optimal rate $\sqrt{N}$. For $q = 2$, the authors showed that the three following assertions are equivalent:

(i) $g \in \mathbf{B}^{\theta}_{2,2}(\gamma)$.

(ii) There exists $c_2 > 0$ such that, for all $N \geq 1$ and $\pi = (t_i)_{i=0}^{N}$,

$$a_X(g(X_1);\pi) \leq c_2 \sup_{i=1,\dots,N} \frac{(t_i - t_{i-1})^{\frac{1}{2}}}{(1 - t_{i-1})^{\frac{1-\theta}{2}}}.$$

(iii) There exists $c_3 > 0$ such that, for all $N \geq 1$,

$$a_X(g(X_1);\pi_N^{\theta}) \leq \frac{c_3}{\sqrt{N}}.$$

Either in the results of Gobet *et al.* [GT01] or Geiss *et al.* [GH07][GG04], the constants depend on some suitable integral of the second moment of the second derivative of the function u. Indeed, Lemma 3.10 in [GH07] states that, for $g \in \mathbf{L}_2(\gamma)$ and a time net $\pi = (t_i)_{i=0}^N$,

$$a_X(g(X_1);\pi) \sim_c \left(\sum_{i=1}^N \int_{t_{i-1}}^{t_i} (t_i - r)\mathbb{E} \left| \frac{\partial^2 u}{\partial x^2}(r, W_r) \right|^2 dr \right)^{\frac{1}{2}}.$$

That is why the study of the order of $\mathcal{E}(z,\pi)$ often reduces to that of the rate of inreasing in time of the $\mathbf{L}_2$-moment of the second derivative of u.

4.3 Other related works on sensitivities involving fractional regularity

This issue was investigated by Gobet and Munos [GM05b]. In this paper, they gave some sensitivity formulae for the expected cost $J(\alpha) := \mathbb{E}g(X_T^\alpha)$, where $(X_t^\alpha)_{0\le t\le T}$ is a multidimensional diffusion process with initial value $X_0 = x$, and whose dynamics depends on a parameter α:

$$\begin{cases} X_0 &= x, \\ dX_t &= b(t, X_t, \alpha)dt + \sum_{j=1}^q \sigma_j(t, X_t, \alpha)dW_t^j. \end{cases} \tag{4.2}$$

The terminal functions g that were considered were the bounded measurable functions satisfying the following fractional regularity for $p_0 > 1$:

$$\int_0^T \frac{\left\| g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T)) \right\|_{\mathbf{L}_{p_0}}}{T-t} dt < \infty, \tag{4.3}$$

which is satisfied for instance by uniformly Hölder continuous functions and by the indicator function of a domain (under some conditions on this domain). It is also satisfied, in the case of $X = W$ and $p_0 = 2$, by functions $g \in \cap_{0<\theta<1}\mathbf{B}_{2,2}^\theta(\gamma)$ given by Theorem 4.2.1.

Using Itô-Malliavin integration-by-parts techniques, the authors gave representation formulae for $\nabla_x u(t, X_t)$ and $D^2u(t, X_t)$ ($u(t,x)$ still stands for $\mathbb{E}(g(X_T)|X_t = x)$) in terms of g, assumed to be only bounded (see Theorem 2.11 in [GM05b] and its proof):

$$\nabla_x u(t, X_t) = \mathbb{E}^{\mathcal{F}_t}\left[g(X_T) H_{t,T}^{(1)} \right],$$
$$D^2 u(t, X_t) = \mathbb{E}^{\mathcal{F}_t}\left[g(X_T) H_{t,T}^{(2)} \right],$$

where $H_{t,T}^{(1)}$ and $H_{t,T}^{(2)}$ are suitable Malliavin weights that have explicit forms (see formulae (5.47) and (5.48) in Appendix 5.A, valid when $d = q$), and that can be then estimated using the condition (4.3). We will refer to these techniques in our work, using a bit different type of fractional regularity instead of that given by (4.3).

4.4 Conclusion

Our objective is threefold: to extend the results of [GT01] to any terminal function g having some fractional smoothness, to take a multidimensional diffusion X instead of the one-dimensional one in [GH07], and to study the associated $\mathbf{L}_2$-regularity of the Z-component of general BSDEs with non null generator, according to the fractional regularity of the function g.
We first establish, in the null driver case, the equivalence between the fractional regularity of g and some estimates on the $\mathbf{L}_2$-moments of the derivatives of the function u, from which we deduce the behaviour of $\mathcal{E}(z, \pi)$. Then, showing that the non null driver case is just a perturbation of the null driver one, we state the results for $\mathcal{E}(Z, \pi)$.

Chapter 5

$\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions

This chapter corresponds to an article submitted for publication in "Stochastic Processes and their Applications", currently under minor revision.

5.1 Abstract

We study the $\mathbf{L}_2$-time regularity of the Z-component of a Markovian BSDE, whose terminal condition is a function g of a forward SDE $(X_t)_{0\leq t\leq T}$. When g is Lipschitz continuous, Zhang [Zha04] proved that the related squared $\mathbf{L}_2$-time regularity is of order one with respect to the size of the time mesh. We extend this type of result to any function g, including irregular functions such as indicator functions for instance. We show that the order of convergence is explicitly connected to the rate of decreasing of the expected conditional variance of $g(X_T)$ given X_t as t goes to T. This holds true for any Lipschitz continuous generator. The results are optimal.

5.2 Framework

In the past decade, a lot of attention has been paid to the numerical resolution of Backward Stochastic Differential Equations (BSDEs in short). In this work, we focus on Markovian BSDEs, that is the case where the BSDE is coupled to a forward SDE. For fixed initial condition x_0 and terminal time $T > 0$, it writes

$$\begin{cases} X_0 &= x_0, \\ dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ -dY_t &= f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T &= g(X_T), \end{cases} \tag{5.1}$$

where $g(X_T) \in \mathbf{L}_2$, W is a standard Brownian motion. A solution to (5.1) is a triplet (X, Y, Z) adapted to the filtration of the Brownian motion, and in some

appropriate $\mathbf{L}_2$ spaces (defined later). When the generator f equals 0, Y is given by the conditional expectation $Y_t = \mathbb{E}^{\mathcal{F}_t}(g(X_T))$ and Z is the predictable process arising from the predictable representation theorem. This type of closed representation can be extended to f that are linear w.r.t. the variables y and z (called linear BSDEs). In the other cases (truly non linear), usually no closed representation is available and one needs to compute numerical solutions. As explained later, the corner stone to derive a rate of convergence for numerical schemes solving (5.1) is the $\mathbf{L}_2$-time regularity of Z. It is defined for a given time mesh $\pi = \{0 = t_0 < \cdots < t_i < \cdots < t_N = T\}$ by

$$\mathcal{E}(Z,\pi) = \sum_{i=0}^{N-1} \mathbb{E} \int_{t_i}^{t_{i+1}} |Z_t - \bar{Z}_{t_i}|^2 dt \tag{5.2}$$

where $\bar{Z}_{t_i} = \frac{1}{t_{i+1}-t_i}\mathbb{E}^{\mathcal{F}_{t_i}} \int_{t_i}^{t_{i+1}} Z_s ds$. Note that $\bar{Z}_{t_i}$ is the projection of $(Z_s)_{t_i \le s \le t_{i+1}}$ on the space of $\mathcal{F}_{t_i}$-measurable random variables, according to the scalar product $< u, v >= \mathbb{E} \int_{t_i}^{t_{i+1}} u_s v_s ds$. The objective of this work is to provide tight estimates of $\mathcal{E}(Z,\pi)$, according to the regularity of the function g and the mesh size $|\pi| = \sup_{0 \le i < N}(t_{i+1} - t_i)$. In all the sequel, we only consider time mesh with N deterministic points.

5.3 A brief account on numerical methods for BSDEs

There are three main approaches for the numerical solution of (Y, Z) (the simulation of forward component X is standard). Firstly, under appropriate conditions, $Y_t = v(t, X_t)$ where v solves a semi-linear PDE (and Z is analogously related to the gradient of v) (see [Par98] for instance): hence one may solve this PDE by deterministic methods and then, we get Y by simulating X. Secondly, one may approach the BSDE by a sequence of linear BSDEs (Picard iteration scheme): this is efficiently achieved by Gobet and Labart [GL09], by coupling the resolution with iterative control variates that drastically improves the accuracy. The third approach is strongly related to the motivation of this work: it uses a backward dynamic programming equation of the form ($i < N$)

$$\begin{cases} Y^\pi_{t_i} &= \mathbb{E}^{\mathcal{F}_{t_i}}(Y^\pi_{t_{i+1}} + (t_{i+1} - t_i) f(t_i, X_{t_i}, Y^\pi_{t_{i+1}}, Z^\pi_{t_i})), \\ Z^\pi_{t_i} &= \frac{1}{(t_{i+1}-t_i)}\mathbb{E}^{\mathcal{F}_{t_i}}(Y^\pi_{t_{i+1}}(W_{t_{i+1}} - W_{t_i})^*), \end{cases} \tag{5.3}$$

where * denotes the transposition and $Y^\pi_{t_N} = g(X_T)$. In addition, possibly X can be replaced by a process easier to simulate (Euler scheme for instance) and close to X. The equations (5.3) defines an explicit scheme but it could be implicit as well, replacing in f the quantity $Y^\pi_{t_{i+1}}$ by $Y^\pi_{t_i}$: this does not modify the convergence results. The next big issue would be how to compute the conditional expectations: we do not discuss these aspects here and we refer to [BP03] for quantization techniques, to [BT04] for Malliavin calculus tools, to [GLW06] for empirical regression methods. Let us focus on the error estimate between (Y^π, Z^π) and (Y, Z). Actually under

standard Lipschitz assumptions on f, it is now well known (even in the more general case where jumps are included in the equations, see [GL06]) that the error can be estimated as follows:

$$\begin{aligned} e(Y^\pi - Y, Z^\pi - Z, \pi) &:= \sup_{0\le i\le N} \mathbb{E}(Y^\pi_{t_i} - Y_{t_i})^2 + \sum_{i=0}^{N-1} \mathbb{E}\int_{t_i}^{t_{i+1}} |Z^\pi_{t_i} - Z_t|^2 dt \\ &\le C(|\pi| + \mathcal{E}(Z,\pi)). \end{aligned}$$

Thus, it is clear that the $\mathbf{L}_2$-time regularity of Z plays a crucial role in the rate of convergence of the dynamic programming equation (5.3).

5.4 Known results on the $\mathbf{L}_2$-regularity of Z

In the BSDE framework, the best result to our knowledge has been obtained by Zhang [Zha04]: $\mathcal{E}(Z,\pi)$ is of order $|\pi|$ when g is a Lipschitz continuous function. Consequently, $e(Y^\pi - Y, Z^\pi - Z, \pi)$ is also of order $|\pi|$ and uniform time grids ($t_i = iT/N$) are sufficient for the approximation scheme. However, in practice g may be an indicator function: in that case, one expects that it worsens the rate of convergence $\mathcal{E}(Z,\pi)$ to 0. Actually, this downgrade phenomenon is well known when the generator f is null. This problem is related to the approximation of stochastic integrals and of hedging strategy in finance. In [GT01], it is proved for instance that for indicator functions in dimension 1, one gets $\mathcal{E}(Z,\pi) = O(N^{-1/2})$ for an uniform time grid. In fact, any rate $N^{-\alpha}$ with $\alpha \in (0,1)$ can be obtained, by picking an appropriate non smooth function g_α. The larger α is, the smoother g_α is. The above results are extended by Geiss and his coauthors (see [GH07] and references therein) by considering functions g in a Besov space $\mathbf{B}^\alpha_{2,2}$ ($\alpha \in (0,1]$). For a uniform time grid, they prove that the regularity index α exactly gives the rate of convergence: $\mathcal{E}(Z,\pi) = O(N^{-\alpha})$. In addition, to get the rate N^{-1} using a grid with N points, one has to consider points appropriately concentrated near T. However, their results are restricted either to the Brownian motion case for X (as in [GH07]) or to one-dimensional time-homogeneous SDEs (as in [GG04]).

The purpose of this work is twofold: firstly, to extend this type of results to general SDEs; secondly, to deal with general BSDEs (i.e. with non null generator).

5.5 Summary of our results

For general SDE model, the characterization of the rate of convergence of $\mathcal{E}(Z,\pi)$ in terms of Besov space is no more relevant. It appears more natural to consider the following space ($\alpha \in (0,1]$)

$$\mathbf{L}_{2,\alpha} = \{g \text{ s.t. } \mathbb{E}(g(X_T)^2) + \sup_{0\le t<T} \frac{\mathbb{E}(g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T)))^2}{(T-t)^\alpha} < +\infty\}. \tag{5.4}$$

It describes the rate of decreasing of the expected conditional variance of $g(X_T)$ given $\mathcal{F}_t$ as t goes to T. If X is a Brownian motion and $T = 1$, $g \in \mathbf{L}_{2,\alpha}$ is equivalent to $g \in \mathbf{B}^{\alpha}_{2,\infty}$ provided $\alpha \neq 1$ (see Corollary 2.3 in [GH07]). However, the current characterization is more flexible because it is adapted to the process and the time horizon T (see [GG04] in dimension 1). In addition, we show that this quantity is intrinsic to the the time regularity of Z (even in the BSDE case). For uniform grids, the rate of convergence is of order $N^{-\alpha}$ (Theorem 5.9.2, (a)). Also, one can take non uniform grids to get the rate N^{-1} (Theorem 5.9.2, (b)). To achieve these results, we first estimate the error in the null generator case (thus extending the results by Geiss et al. in a non trivial way) (Theorem 5.7.1, Theorem 5.7.2 and Corollary 5.7.2). Then we prove that the non null generator case (involving $(Z_t)_t$) is a perturbation of the null case (with $(z_t)_t$), so that the former results still apply (Theorem 5.9.1):

$$\mathcal{E}(Z, \pi) \leq C(\mathcal{E}(z, \pi) + |\pi|).$$

More precisely, we establish that Z is the superposition of z plus a time smoother term (Theorem 5.8.1). This result seems to be original in our framework. It allows us to reduce the study of the $\mathbf{L}_2$ time regularity of Z to that of z (the former case) and that of the smoother term (which is easier). The decomposition may be also interesting to get tight estimates on the behavior of Z as t goes to T (Corollary 5.8.1). Our proof relies on stochastic analysis techniques combining PDEs, martingales, Itô calculus and BSDEs in $\mathbf{L}_p$ ($p \in (1, 2]$). We mention that usually with these tools, g is supposed to have a polynomial growth, ensuring that $g(X_T)$ is in any $\mathbf{L}_p$ for $p > 0$. Here we stress the fact that we only assume $g(X_T) \in \mathbf{L}_2$ which is the minimal condition to discuss the existence and uniqueness of the solution of (5.1) in $\mathbf{L}_2$ spaces.

We finally discuss the choice of time grids (uniform or alpha dependent) and the optimality of the results.

5.6 Preliminaries

Hereafter, W is a q-dimensional Brownian motion, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $(\mathcal{F}_t)_{0 \leq t \leq T}$ (T is a fixed terminal time) is the natural filtration of W, augmented with $\mathbb{P}$-null sets.
We denote the conditional expectation $\mathbb{E}(X|\mathcal{F}_t)$ of a random variable X by $\mathbb{E}^{\mathcal{F}_t}(X)$. $A \leq_c B$ means $A \leq cB$ with a constant c depending on T, b, σ, f and universal constants.
For a $r \times c$ matrix A ($r, c \geq 1$), that will be considered as an element of $\mathbb{R}^{r \times c}$, A^* stands for its transpose, A_j for its jth column, and $|A|$ for its Euclidean norm ($|A| := \sqrt{\mathrm{Tr}(AA^*)}$).
If $\varphi : \mathbb{R}^{p_1} \to \mathbb{R}^{p_2}$ is a differentiable function, its gradient $\nabla_x \varphi(x) := (\partial_{x_1} \varphi(x), ..., \partial_{x_{p_1}} \varphi(x))$ takes values in $\mathbb{R}^{p_2 \times p_1}$.
If $p_2 = 1$, $D^2 \varphi(x) := (\partial^2_{x_i, x_j} \varphi(x))_{i,j=1...d}$ stands for the Hessian matrix of φ and takes values in $\mathbb{R}^{p_1 \times p_1}$.

a) The forward component:

$$\begin{cases} X_0 &= x_0, \\ dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \end{cases} \tag{5.5}$$

where $X, x_0 \in \mathbb{R}^d$, $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times q}$. We will assume that the coefficients of this SDE satisfy the following assumption:

> **($\mathbf{A}_{b,\sigma}$)** The functions b and σ are bounded and twice continuously differentiable with respect to the space variable, with uniformly bounded and γ-Hölder continuous derivatives, for some $\gamma \in (0,1]$. In addition, b and σ are $\frac{1}{2}$-Hölder continuous in time. σ is also assumed to be uniformly elliptic: there exists $\delta > 0$ such that, $\forall (t,x) \in [0,T] \times \mathbb{R}^d$, $[\sigma\sigma^*](t,x) \geq \delta I_d$.

We denote by ∇X_s the gradient of X_s with respect to x_0, and by $(D_t X_s)_{0\leq t\leq s}$ its Malliavin derivative (see [Nua05]). It is known that $(\nabla X_s)_{s\geq 0}$ and $(D_t X_s)_{s\geq t}$ satisfy the following linear SDEs

$$\nabla X_s = I_d + \int_0^s \nabla_x b(r, X_r)\nabla X_r dr + \sum_{j=1}^q \int_0^s \nabla_x \sigma_j(r, X_r)\nabla X_r dW_r^j; \tag{5.6}$$

$$D_t X_s = \sigma(t, X_t) + \int_t^s \nabla_x b(r, X_r) D_t X_r dr + \sum_{j=1}^q \int_t^s \nabla_x \sigma_j(r, X_r) D_t X_r dW_r^j. \tag{5.7}$$

The following estimates are standard results in SDE literature.

Lemma 5.6.1. *Assume ($\mathbf{A}_{b,\sigma}$). For any $p \geq 2$, there exits a constant C_p such that*

$$\mathbb{E} \sup_{0\leq s\leq T} |X_s|^p \leq C_p(1 + |x_0|^p),$$
$$\mathbb{E}|X_s - X_t|^p \leq C_p|s-t|^{\frac{p}{2}}.$$

From equation (5.6), one gets the following estimate, that will be used in this work (it is a standard estimate if $t = 0$; one can deduce the estimate for $t \neq 0$ since $\nabla X_s[\nabla X_t]^{-1}$ is the derivative of X_s with respect to X_t):

$$\mathbb{E}^{\mathcal{F}_t} \sup_{t\leq s\leq T} |\nabla X_s[\nabla X_t]^{-1}|^p \leq C_p. \tag{5.8}$$

Since

$$D_t X_s = \nabla X_s(\nabla X_t)^{-1}\sigma(t, X_t)\mathbb{1}_{t\leq s} \tag{5.9}$$

and σ is bounded, the same estimate applies to $(D_t X_s)_{t\leq s\leq T}$.

b) The backward component:

$$\begin{cases} -dY_t &= f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T &= g(X_T). \end{cases} \tag{5.10}$$

We define the space $\mathcal{S}^p$ to be the set of continuous adapted processes Y such that

$$\mathbb{E}\big[\sup_{t\in[0,T]} |Y_t|^p\big] < +\infty,$$

and $\mathcal{M}^p$ the set of predictable processes Z such that

$$\mathbb{E}\big[\big(\int_0^T |Z_s|^2 ds\big)^{p/2}\big] < +\infty.$$

In the following, Y is always considered as a one dimensional process, but we think that all our study would remain valid if it were multidimensional (the analysis, however, would be more intricate since we have a system of coupled PDEs in this case).

A solution to (5.10) is a triplet (X, Y, Z), where X is a continuous adapted $\mathbb{R}^d$-valued process with $\mathbb{E}(\sup_{t\leq T} |X_t|^2) < +\infty$, solution to the SDE (5.5), and $(Y, Z) \in \mathcal{S}^2 \times \mathcal{M}^2$. We make use of the following assumption on the generator:

> **($\mathbf{A}_f$)** The function f is continuous with respect to its four arguments, and continuously differentiable with respect to (x, y, z) with uniformly bounded derivatives. Moreover, $\int_0^T |f(s, 0, 0, 0)|ds < +\infty$.

In Theorem 5.9.3, f is assumed to be only Lipschitz continuous in (x, y, z), but not necessarily continuously differentiable.

Under the assumptions **($\mathbf{A}_{b,\sigma}$)** and **($\mathbf{A}_f$)**, and when $\mathbb{E}|g(X_T)|^2 < +\infty$, the FB-SDE (5.1) has a unique solution $(X, Y, Z) \in \mathcal{S}^2 \times \mathcal{S}^2 \times \mathcal{M}^2$.

c) Linear PDE and linear BSDE. Some of our intermediate results require the following boundedness assumption on the terminal function g:

> **($\mathbf{A}_g$)** g is a bounded measurable function.

Under the assumptions **($\mathbf{A}_{b,\sigma}$)** and **($\mathbf{A}_g$)**, and setting $u(t, x) := \mathbb{E}\left[g(X_T^{t,x})\right]$, one has $u(t, x) = \int_{\mathbb{R}^d} p(t, x; T, y)g(y)dy$ where p is the probability transition density function of X. It is well known that p is a smooth function for $t < T$ (see Friedman [Fri64]) and this regularity transfers to u since g is bounded. Indeed, Gaussian type estimates on p and its derivatives enable us to apply the Lebesgue derivation theorem. Of course, boundedness assumptions are too strong for this statement, and sub-exponential growth would be enough. However, assuming only at this stage that g is such that $\mathbb{E}(g^2(X_T)) < +\infty$ leads to technicalities that we have not been able to overcome.
To sum up, under **($\mathbf{A}_g$)**, $\nabla_x u$, $D^2 u$, $\partial_x^3 u$, $\partial_t u$, $\partial_t \nabla_x u$ exist and are continuous for $t < T$, and u is the smooth solution (on $[0, T) \times \mathbb{R}^d$) of the partial differential equation (PDE)

$$\begin{aligned}
&\partial_t u(t, x) + \sum_{i=1}^d b_i(t, x)\partial_{x_i} u(t, x) + \frac{1}{2}\sum_{i,j=1}^d [\sigma\sigma^*]_{i,j}(t, x)\partial^2_{x_i,x_j} u(t, x) = 0 \text{ for } t < T,\\
&u(T, x) = g(x). \qquad (5.11)
\end{aligned}$$

Let $(y_t, z_t)_{0\le t\le T}$ be the solution of the linear BSDE:

$$y_t = g(X_T) - \int_t^T z_s dW_s. \tag{5.12}$$

Then

$$y_t = u(t, X_t), \quad z_t = \nabla_x u(t, X_t)\sigma(t, X_t). \tag{5.13}$$

d) The space $\mathbf{L}_{2,\alpha}$. For a measurable function g satisfying $\mathbb{E}|g(X_T)|^2 < +\infty$, we set

$$V_{t,T}(g) := \mathbb{E}\left|g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T))\right|^2, \tag{5.14}$$

and, when g belongs to $\mathbf{L}_{2,\alpha}$, we define $K^\alpha(g)$ as

$$K^\alpha(g) := \mathbb{E}|g(X_T)|^2 + \sup_{t\in[0,T)} \frac{V_{t,T}(g)}{(T-t)^\alpha}.$$

Notice that $\bigcup_{\alpha\in(0,1]} \mathbf{L}_{2,\alpha}$ obviously contains uniformly Hölder continuous functions, but also some non-smooth functions, such as the indicator function of a domain (under some conditions on the functions b and σ and on the domain: see Gobet and Munos [GM05b]).

Examples:

- If g is β-Hölder continuous, then $g \in \mathbf{L}_{2,\beta}$.
- If $d = q = 1$, $X = W$ and $g(x) = \mathbb{1}_{[0,+\infty)}(x)$, then $g \in \mathbf{L}_{2,\frac{1}{2}}$ (see paragraph 5.7.2).
- More generally, for an indicator function of a smooth domain (a domain whose boundary is compact and of class $\mathcal{C}^2$), $g \in \mathbf{L}_{2,\frac{1}{2}}$.

e) The time net. In all what follows, $\pi := (t_k)_{k=0\dots N}$ is a deterministic time net, such that $0 = t_0 < t_1 < \dots < t_N = T$, and $|\pi| := \sup_{k=0\dots N}(t_{k+1} - t_k)$. We shall use the following net ($\beta \in (0, 1]$)

$$\pi^{(\beta)} := \left\{t_k^{(N,\beta)} := T - T\big(1 - \frac{k}{N}\big)^{\frac{1}{\beta}}, 0 \le k \le N\right\}. \tag{5.15}$$

Note that $\pi^{(1)} = (t_k^{(N,1)})$ coincides with the equidistant net. For $\beta < 1$, the points in $\pi^{(\beta)}$ are more concentrated near T.

f) The constants. We emphasize the fact that, whenever a constant depends on the function g, the dependence will be expressed explicitly, so that all the constants such as C or c do not depend on g, but may depend on b, σ, f, α and other universal constants. They may also depend on T, but remain bounded when $T \to 0$.

5.7 The BSDE with null generator ($f = 0$) and bounded terminal condition g

5.7.1 Main results

In this section, we study the solution $(y_t, z_t)_{0\le t\le T}$ of the BSDE with null generator, and with terminal condition $g(X_T)$. We derive estimates that will be usefull for the

next sections, and in particular we study the $\mathbf{L}_2$-regularity of the integrand $(z_t)_{0\le t\le T}$. It is known for years that the $\mathbf{L}_2$-time regularity of z is strongly related to the rate of explosion of the derivatives of $u(t,x)$ as t goes to T (see [GT01], [GG04] and [GH07]). We give below standard and also new related estimates, that will be useful in the proofs.
The following estimate is standard:

$$\mathbb{E}\sup_{0\le t\le T}|y_t|^2+\mathbb{E}\int_0^T|z_s|^2ds\le_c\mathbb{E}|g(X_T)|^2, \tag{5.16}$$

and, it follows from (5.13) and (5.16), under the ellipticity assumption,

$$\mathbb{E}\sup_{0\le t\le T}|u(t,X_t)|^2+\mathbb{E}\int_0^T|\nabla_x u(s,X_s)|^2ds\le_c\mathbb{E}|g(X_T)|^2, \tag{5.17}$$

We now bring together different estimates on $\nabla_x u$ and D^2u in terms of the suitable integrability of $V_{t,T}(g)$ as t goes to T.

Lemma 5.7.1 ($\mathbf{L}_2$-estimates for u and its derivatives)**.** *Assume* $(\boldsymbol{A}_{b,\sigma})$ *and* $(\boldsymbol{A}_g)$*. Then, there exists a positive constant C, such that, for all $t\in[0,T)$,*

$$\begin{aligned}\mathbb{E}|u(t,X_t)|^2&\le\mathbb{E}|g(X_T)|^2,\\ \mathbb{E}|\nabla_x u(t,X_t)|^2&\le C\frac{V_{t,T}(g)}{T-t},\\ \mathbb{E}|D^2u(t,X_t)|^2&\le C\frac{V_{t,T}(g)}{(T-t)^2}.\end{aligned}$$

For the proof, see section 5.7.3. The powers of $(T-t)$ appearing in Lemma 5.7.1 are standard, but note that the $\mathbf{L}_2$-norms depend on $V_{t,T}(g)$ but not on the supremum norm of g.

The following estimate, which is a consequence of Lemma 5.7.1, will be usefull in our work:

Corollary 5.7.1. *Assume* $(\boldsymbol{A}_{b,\sigma})$*and* $(\boldsymbol{A}_g)$*. Assume moreover that $g\in\mathbf{L}_{2,\alpha}$, for some $\alpha\in(0,1]$. Then, there exists a positive constant C, such that*

$$\mathbb{E}\left(\int_0^T|\nabla_x u(t,X_t)|+|D^2u(t,X_t)|dt\right)^2\le CT^\alpha K^\alpha(g).$$

For the proof, see section 5.7.3. We will show (see the proof of Theorem 5.7.2), that

$$\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}|z_s-\bar z_{t_k}|^2\,ds\le_c\frac{1}{N}+\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}(t_{k+1}-r)\mathbb{E}|D^2u(r,X_r)|^2dr$$

(and we have equivalence if $(X_t)_t$ is the Brownian motion).
Now, Geiss and Hujo [GH07] (Lemma 3.8) showed that, if $\varphi : [0,T) \longrightarrow [0,\infty)$ is a non-decreasing continuous function, then

$$\sum_{k=0}^{N-1} \int_{t_k^{(N,\beta)}}^{t_{k+1}^{(N,\beta)}} (t_{k+1}^{(N,\beta)} - r)\varphi(r)dr \leq \frac{c}{N} \Longleftrightarrow \int_0^T (T-r)^{1-\beta}\varphi(r)dr < \infty. \tag{5.18}$$

If $(X_t)_t$ is a Brownian Motion, $(D^2u(r,X_r))_{r<T}$ is an $\mathbf{L}_2$-martingale, which easily implies that $\varphi(r) = \mathbb{E}|D^2u(r,X_r)|^2$ defines a non-decreasing continuous function.
In the following proposition, we give two (more explicit) new characterizations of the integrability of $(T-r)^{1-\beta}\varphi(r)$.

Proposition 5.7.1. *Let $\beta \in (0,1)$, and assume $(\boldsymbol{A}_{b,\sigma})$, $(\boldsymbol{A}_g)$. Then the following assertions are equivalent:*
(i) $\displaystyle\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2u(r,X_r)|^2dr < +\infty.$
(ii) $\displaystyle\int_0^T (T-r)^{-\beta}\mathbb{E}|\nabla_x u(r,X_r)|^2dr < +\infty.$
(iii) $\displaystyle\int_0^T (T-r)^{-1-\beta}V_{r,T}(g)dr < +\infty.$

For the proof, see section 5.7.3.

Remark 5.7.1. *Actually, in the above result, the boundedness assumptions of g can be relaxed into a sub-exponential growth condition.*
The characterizations given by Proposition 5.7.1 are no longer true when $\beta = 1$. A counterexample is given by $g(x) = x$ with $(X_t) \equiv (W_t)$, which gives $u(t,x) = x$: assertion(i) is satisfied, but neither (ii) nor (iii) are. In fact, it can easily be seen that if we take any infinitely smooth but non constant function g (always with $(X_t) \equiv (W_t)$), the assertion (iii) is never satisfied with $\beta = 1$.
That's why we did not define the space $\mathbf{L}_{2,\beta}$ as the space of functions satisfying the assertion (iii) : otherwise, a Lipschitz continuous function such that $g(x) = x$ would belong to $\mathbf{L}_{2,\beta}$ with $\beta < 1$ but not to $\mathbf{L}_{2,1}$. Thus, it would imply (see below) to work with the non equidistant time grid $\pi^{(\beta)}$ instead of the equidistant one as it has usually been done when one has a Lipschitz continuous terminal condition.
And it is clear that, if $g \in \mathbf{L}_{2,\alpha}$ $(\alpha \leq 1)$, then all the three assertions of Proposition 5.7.1 are satisfied with $\beta < \alpha$.

Now, we state tight estimates on $\nabla_x u$ and D^2u according to $g \in \mathbf{L}_{2,\alpha}$ for $\alpha \in (0,1]$ (note that $\alpha = 1$ is allowed).

Lemma 5.7.2. *Let $\alpha \in (0,1]$, and assume $(\boldsymbol{A}_{b,\sigma})$ and $(\boldsymbol{A}_g)$. Then the three following assertions are equivalent:*
(i) $g \in \mathbf{L}_{2,\alpha}$.
(ii) $\exists C^\alpha(g) > 0$, *such that,* $\forall t \in [0,T)$,

$$\int_0^t \mathbb{E}\left|D^2u(s,X_s)\right|^2 ds \leq \frac{C^\alpha(g)}{(T-t)^{1-\alpha}}.$$

(iii) $\exists C^\alpha(g) > 0$, *such that,* $\forall t \in [0, T)$,

$$\mathbb{E}\left|\nabla_x u(t, X_t)\right|^2 \leq \frac{C^\alpha(g)}{(T-t)^{1-\alpha}}.$$

And, if $g \in \mathbf{L}_{2,\alpha}$, *one can take* $C^\alpha(g) = CK^\alpha(g)$ *in (ii) and (iii).*
If $\alpha < 1$ *(resp.* $\alpha = 1$*), the previous three assertions are also equivalent to (resp. lead to) the following one:*
(iv) $\exists C^\alpha(g) > 0$, *such that,* $\forall t \in [0, T)$,

$$\mathbb{E}\left|D^2 u(t, X_t)\right|^2 \leq \frac{C^\alpha(g)}{(T-t)^{2-\alpha}}$$

(and one can take $C^\alpha(g) = CK^\alpha(g)$*).*

Remark 5.7.2. *The assumption "*$g \in \mathbf{L}_{2,\alpha}$*" is natural in our framework, if we want a rate of convergence for* $\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2 \, ds$ *which is polynomial with respect to the time-step* $|\pi|$. *In fact, Geiss and Hujo [GH07] (Theorem 1.3 and Theorem 2.1) showed that, if* B *is either the Brownian motion or the geometric Brownian motion, then the following assertions are equivalent (*$\alpha \in (0, 1)$*):*
(GH-i) $\exists C > 0, \forall t \in [0, T)$,

$$\mathbb{E}\left|\nabla_x u(t, B_t)\right|^2 \leq \frac{C}{(T-t)^{1-\alpha}}.$$

(GH-ii) $\exists C > 0$ *such that, for all time-nets* $\pi = (t_k)_{k=0\ldots N}$,

$$\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2 \, ds \leq \frac{C}{N^\alpha}.$$

Lemma 5.7.2 shows that the assertion (GH-i), written with a general diffusion X *instead of* B, *is equivalent to the assertion "*$g \in \mathbf{L}_{2,\alpha}$*" (even for* $\alpha = 1$*).*

The previous estimates are sufficient to assert that, if $g \in \mathbf{L}_{2,\alpha}$, then the equidistant time net provides an $\mathcal{E}(z, \pi^{(1)})$ of order $\frac{1}{N^\alpha}$.

Theorem 5.7.1. *Assume* $(\boldsymbol{A}_{b,\sigma})$ *and* $(\boldsymbol{A}_g)$. *Assume moreover that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0, 1]$. *Then, with the choice of the equidistant time net,*

$$\mathcal{E}(z, \pi^{(1)}) = \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k^{(N,1)}}^{t_{k+1}^{(N,1)}} \left|z_s - \bar{z}_{t_k^{(N,1)}}\right|^2 ds \leq CK^\alpha(g) \left(\frac{T}{N}\right)^\alpha$$

(where C *does not depend on* N*).*

Proof. One knows by (5.13) that $z_s = \nabla_x u(s, X_s)\sigma(s, X_s)$. Thus, by a projection argument, one has

$$\begin{aligned}\mathbb{E} \int_{t_k}^{t_{k+1}} |z_s - \bar{z}_{t_k}|^2 ds &\leq \mathbb{E} \int_{t_k}^{t_{k+1}} |z_s - z_{t_k}|^2 ds \\ &= \mathbb{E} \int_{t_k}^{t_{k+1}} |\nabla_x u(s, X_s)\sigma(s, X_s) - \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k})|^2 ds.\end{aligned}$$

Now, write $\nabla_x u(s,X_s)\sigma(s,X_s) - \nabla_x u(t_k,X_{t_k})\sigma(t_k,X_{t_k}) = \nabla_x u(t_k,X_{t_k})\left\{\sigma(s,X_s)-\sigma(t_k,X_{t_k})\right\} + \left\{\nabla_x u(s,X_s)-\nabla_x u(t_k,X_{t_k})\right\}\sigma(s,X_s)$. Then, using the assumptions $(\mathbf{A}_{b,\sigma})$, and for $s\in[t_k,t_{k+1}]$,

$$\begin{aligned}
&\mathbb{E}|\nabla_x u(s,X_s)\sigma(s,X_s)-\nabla_x u(t_k,X_{t_k})\sigma(t_k,X_{t_k})|^2\\
&\le_c \mathbb{E}\left\{\left(|s-t_k|^{\frac{1}{2}}+|X_s-X_{t_k}|\right)^2|\nabla_x u(t_k,X_{t_k})|^2\right\}+\mathbb{E}|\nabla_x u(s,X_s)-\nabla_x u(t_k,X_{t_k})|^2\\
&= E_1+E_2.
\end{aligned}$$

Clearly, and by means of Lemma 5.7.1,

$$E_1 \le_c |\pi|\mathbb{E}|\nabla_x u(t_k,X_{t_k})|^2 \le |\pi|\frac{K^\alpha(g)}{(T-t_k)^{1-\alpha}} \le |\pi|\frac{K^\alpha(g)}{(T-s)^{1-\alpha}}.$$

As in the proof of Proposition 5.7.1 (see (5.22)), one gets the general estimate (under $(\mathbf{A}_g)$)

$$\begin{aligned}
E_2 &\le_c \int_{t_k}^s \mathbb{E}|\nabla_x u(r,X_r)|^2dr+\int_{t_k}^s \mathbb{E}|D^2u(r,X_r)|^2dr\\
&\le_c |\pi|\frac{K^\alpha(g)}{(T-s)^{1-\alpha}}+\int_{t_k}^s \mathbb{E}|D^2u(r,X_r)|^2dr
\end{aligned}$$

using (iii) Lemma 5.7.2. Therefore

$$\begin{aligned}
\mathcal{E}(z,\pi) &\le_c |\pi|K^\alpha(g)\int_0^T \frac{1}{(T-s)^{1-\alpha}}ds+\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}\int_{t_k}^s \mathbb{E}|D^2u(r,X_r)|^2drds\\
&= |\pi|K^\alpha(g)T^\alpha+\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}(t_{k+1}-r)\mathbb{E}|D^2u(r,X_r)|^2dr, \qquad (5.19)
\end{aligned}$$

where we have used an integration by parts. Note that the above upper bound is available for any time net π.
Now, if π is the equidistant time net, $\mathcal{E}(z,\pi^{(1)})$ is bounded (up to a constant c) by

$$K^\alpha(g)\frac{T^{\alpha+1}}{N}+\frac{T}{N}\int_0^{T-\frac{T}{N}}\mathbb{E}|D^2u(r,X_r)|^2dr+\int_{T-\frac{T}{N}}^T(T-r)\mathbb{E}|D^2u(r,X_r)|^2dr.$$

Using Lemma 5.7.1, one gets

$$\begin{aligned}
\mathcal{E}(z,\pi^{(1)}) &\le_c K^\alpha(g)\frac{T^{\alpha+1}}{N}+\frac{T}{N}\int_0^{T-\frac{T}{N}}\frac{K^\alpha(g)}{(T-r)^{2-\alpha}}dr+\int_{T-\frac{T}{N}}^T\frac{K^\alpha(g)}{(T-r)^{1-\alpha}}dr\\
&\le_c K^\alpha(g)\frac{T^{\alpha+1}}{N}+K^\alpha(g)\frac{T}{N}\left(\left(\frac{T}{N}\right)^{-1+\alpha}-T^{-1+\alpha}\right)+K^\alpha(g)\left(\frac{T}{N}\right)^\alpha\\
&\le_c K^\alpha(g)\left(\frac{T}{N}\right)^\alpha.
\end{aligned}$$

□

To get the rate $\frac{1}{N}$ in the case $\alpha \in (0,1)$, one should consider time nets with a higher concentration of points near T to compensate the faster rate of explosion of (D^2u). For non equidistant time nets, we state the following universal bounds.

Theorem 5.7.2. *Assume* $(\boldsymbol{A}_{b,\sigma})$ *and* $(\boldsymbol{A}_g)$. *Assume moreover that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0,1]$.
Now, take $\beta = 1$, *if* $\alpha = 1$, *and* $\beta < \alpha$ *otherwise. Then,* $\exists C > 0$ *such that, for any time net* $\pi = \{t_k, k = 0...N\}$,

$$\mathcal{E}(z,\pi) \leq CK^\alpha(g)T^\alpha|\pi| + CK^\alpha(g)T^{\alpha-\beta} \sup_{k=0...N-1} \left(\frac{t_{k+1} - t_k}{(T-t_k)^{1-\beta}} \right).$$

Proof. Owing to inequality (5.19), $\mathcal{E}(z,\pi)$ is bounded by

$$C \left(|\pi| K^\alpha(g) T^\alpha + \left\{ \sup_{k=0...N-1} \sup_{r \in [t_k, t_{k+1})} \left(\frac{t_{k+1} - r}{(T-r)^{1-\beta}} \right) \right\} \int_0^T (T-r)^{1-\beta} \mathbb{E}|D^2u(r,X_r)|^2 dr \right).$$

Now, for $r \in [t_k, t_{k+1})$,

$$\begin{aligned} \frac{t_{k+1} - r}{(T-r)^{1-\beta}} &= \left(1 - \frac{T - t_{k+1}}{T-r}\right)(T-r)^\beta \\ &\leq \left(1 - \frac{T - t_{k+1}}{T-t_k}\right)(T-t_k)^\beta = \frac{t_{k+1} - t_k}{(T-t_k)^{1-\beta}}, \end{aligned}$$

which leads to $\sup_{r \in [t_k, t_{k+1})} \left(\frac{t_{k+1} - r}{(T-r)^{1-\beta}} \right) = \frac{t_{k+1} - t_k}{(T-t_k)^{1-\beta}}$. Then,

$$\mathcal{E}(z,\pi) \leq_c |\pi| K^\alpha(g) T^\alpha + \sup_{k=0...N-1} \left(\frac{t_{k+1} - t_k}{(T-t_k)^{1-\beta}} \right) \int_0^T (T-r)^{1-\beta} \mathbb{E}|D^2u(r,X_r)|^2 dr.$$

If $\beta = \alpha = 1$, then, from Lemma 5.7.2, one has $\int_0^T \mathbb{E}|D^2u(r,X_r)|^2 dr \leq_c K^\alpha(g)$.
And, if $\beta < \alpha < 1$, then, from Lemma 5.7.1,

$$\int_0^T (T-r)^{1-\beta} \mathbb{E}|D^2u(r,X_r)|^2 dr \leq_c \int_0^T (T-r)^{1-\beta} \frac{K^\alpha(g)}{(T-t)^{2-\alpha}} dr \leq K^\alpha(g) T^{\alpha-\beta}.$$

We conclude that, in both cases,

$$\int_0^T (T-r)^{1-\beta} \mathbb{E}|D^2u(r,X_r)|^2 dr \leq_c K^\alpha(g) T^{\alpha-\beta}.$$

The proof is complete. □

Corollary 5.7.2. *Assume* $(\boldsymbol{A}_{b,\sigma})$, $(\boldsymbol{A}_g)$, *and that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0,1]$. *Let* β *be as in Theorem 5.7.2. Then, with the choice* $\pi^{(\beta)}$ *(defined in (5.15))*

$$\mathcal{E}(z,\pi^{(\beta)}) = \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k^{(N,\beta)}}^{t_{k+1}^{(N,\beta)}} \left| z_s - \bar{z}_{t_k^{(N,\beta)}} \right|^2 ds \leq CK^\alpha(g) \frac{T^\alpha}{N}$$

(where C *does not depend on* N*).*

Proof. Recall that $t_k^{(N,\beta)} = T - T\left(1 - \frac{k}{N}\right)^{\frac{1}{\beta}}$. Since the function $r \mapsto T - T(1-r)^{\frac{1}{\beta}}$ is concave on $[0,1]$, one has

$$t_{k+1}^{(N,\beta)} - t_k^{(N,\beta)} \leq \frac{T}{\beta N}\left(1 - \frac{k}{N}\right)^{\frac{1}{\beta}-1}.$$

Therefore,

$$\frac{t_{k+1}^{(N,\beta)} - t_k^{(N,\beta)}}{(T - t_k^{(N,\beta)})^{1-\beta}} \leq \frac{T}{\beta N}\frac{\left(1 - \frac{k}{N}\right)^{\frac{1}{\beta}-1}}{T^{1-\beta}\left(1 - \frac{k}{N}\right)^{\frac{1}{\beta}(1-\beta)}} = \frac{T^\beta}{\beta N}.$$

This, combined with Theorem 5.7.2, proves Corollary 5.7.2. □

5.7.2 Optimality of the time net

One may raise the following question: if $\alpha < 1$ and $g \in \mathbf{L}_{2,\alpha}$, is the time net $\pi^{(\beta)} = \{t_k^{(N,\beta)} : 0 \leq k \leq N\}$, with $\beta < \alpha$, optimal? In other words, can we take $\beta = \alpha$ so as to have a rate of convergence of $\frac{1}{N}$? It follows from the previous results that the answer is no. Let us give a counterexample.
Let $g(x) = \mathbf{1}_{[0,\infty)}(x)$ and $(X_t) \equiv (W_t)$. Then

$$\begin{aligned}
u(t,x) &= \mathbb{P}(x + W_T - W_t \geq 0),\\
u_x'(t,x) &= \frac{1}{\sqrt{2\pi(T-t)}}\exp -\frac{x^2}{2(T-t)},\\
\mathbb{E}|\nabla_x u(t,X_t)|^2 &= \int_{\mathbb{R}} \frac{1}{2\pi(T-t)}\exp\{-\frac{x^2}{T-t}\}\frac{1}{\sqrt{2\pi t}}\exp\{-\frac{x^2}{2t}\}dx\\
&= \frac{1}{2\pi(T-t)}\frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}}\exp\{-\frac{T+t}{2t(T-t)}x^2\}dx\\
&= \frac{1}{2\pi\sqrt{T+t}\sqrt{T-t}},
\end{aligned}$$

which is equivalent to $(T-t)^{-1/2}$, up to a constant, when $t \to T$. Then, it follows from Lemma 5.7.2 that $g \in \mathbf{L}_{2,\alpha}$ with $\alpha = \frac{1}{2}$ (but not with $\alpha > \frac{1}{2}$).
However, assertion (ii) of Proposition 5.7.1 cannot be satisfied, for this example, with $\beta = \alpha = \frac{1}{2}$, so neither assertion (i) (which is necessary to have the rate $\frac{1}{N}$, in view of the equivalence (5.18)) is.

Remark 5.7.3. *If the assertion (iii) of Proposition 5.7.1 is satisfied for some $\beta < 1$, using analogous arguments we obtain that the rate of convergence is $\frac{1}{N^\beta}$ with the equidistant time net $\pi^{(1)}$, and $\frac{1}{N}$ with the non equidistant net $\pi^{(\beta)}$.*

□

5.7.3 Proofs

Proof of Lemma 5.7.1.
- *Estimate on u.* One has $u(t, X_t) = \mathbb{E}^{\mathcal{F}_t} g(X_T)$, so $\mathbb{E}|u(t, X_t)|^2 \leq \mathbb{E}(\mathbb{E}^{\mathcal{F}_t}|g(X_T)|)^2 \leq \mathbb{E}|g(X_T)|^2$.
- *First derivative of u.*
Suppose first that $d = q$. Then, under the ellipticity assumption, σ is invertible. It is known that, $\forall t \in [0, T)$, $\nabla u(t, .)$ can be represented as a conditional expectation (see Gobet and Munos [GM05b] among others):

$$\nabla_x u(t, X_t) = \mathbb{E}^{\mathcal{F}_t}\left[g(X_T) H^{(1)}_{t,T}\right],$$

where $H^{(1)}_{t,T}$ is the random variable (called Malliavin weight) given by

$$H^{(1)}_{t,T} = \frac{1}{T-t}\int_t^T \sigma^{-1}(s, X_s)\nabla X_s[\nabla X_t]^{-1} dW_s.$$

One uses the estimate (5.8) to get

$$\mathbb{E}^{\mathcal{F}_t}|H^{(1)}_{t,T}|^2 \leq \frac{1}{(T-t)^2}\int_t^T \mathbb{E}^{\mathcal{F}_t}|\nabla X_s[\nabla X_t]^{-1}|^2 ds \leq_c \frac{1}{T-t}.$$

Now, since $\mathbb{E}^{\mathcal{F}_t}(H^{(1)}_{t,T}) = 0$, one can write $\nabla_x u(t, X_t) = \mathbb{E}^{\mathcal{F}_t}[(g(X_T) - \mathbb{E}^{\mathcal{F}_t} g(X_T))H^{(1)}_{t,T}]$. The Cauchy-Schwartz inequality yields

$$|\nabla_x u(t, X_t)|^2 \leq_c \frac{\mathbb{E}^{\mathcal{F}_t}|g(X_T) - \mathbb{E}^{\mathcal{F}_t} g(X_T)|^2}{T-t}. \tag{5.20}$$

So, we obtains $\mathbb{E}|\nabla_x u(t, X_t)|^2 \leq_c \frac{V_{t,T}(g)}{T-t}$.
If $d \neq q$ (and always under the ellipticity assumption on σ), there exists a $d \times d$ symmetric invertible matrix Σ such that $\sigma\sigma^* = \Sigma^2$ (see Stroock and Varadhan [SV79], Lemma 5.2.1, to define a square root of $\sigma\sigma^*$). In addition, Σ satisfies the same regularity estimates as σ. Then, one can carry on the proof above, replacing σ by Σ, since the PDE (5.11) satisfied by u depends on σ only through $\sigma\sigma^*$.
Second derivative of u.
Suppose first that $d = q$, hence σ invertible. It is also known (see again Gobet and Munos [GM05b]) that $\forall t \in [0, T)$, there exists a random variable $H^{(2)}_{t,T}$ such that

$$D^2 u(t, X_t) = \mathbb{E}^{\mathcal{F}_t}\left[g(X_T) H^{(2)}_{t,T}\right],$$

and we can prove (as for the previous estimate) that $H^{(2)}_{t,T}$ satisfies,

$$\mathbb{E}^{\mathcal{F}_t}[H^{(2)}_{t,T}] = 0, \qquad \mathbb{E}^{\mathcal{F}_t}|H^{(2)}_{t,T}|^2 \leq_c \frac{1}{(T-t)^2}.$$

Then the proof of the estimate of $\mathbb{E}|D^2u(t,X_t)|^2$ is the same as for $\mathbb{E}|\nabla_x u(t,X_t)|^2$. Note that the existence of $H^{(2)}_{t,T}$ relies on the existence of $\nabla(\nabla X)$, which holds under $(\mathbf{A}_{b,\sigma})$ because b and σ are both of class $\mathcal{C}^{2+\gamma}$ $(\gamma > 0)$.
If $d \neq q$ and $\sigma\sigma^*$ is elliptic, we proceed as for $\nabla_x u$ using the matrix $\Sigma = (\sigma\sigma^*)^{-1/2}$. □

Proof of Corollary 5.7.1.
First, note that

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^T |\nabla_x u(t,X_t)| + |D^2u(t,X_t)|dt\Big)^2 \\
&\leq 2\Big\{\int_0^T \big(\mathbb{E}|\nabla_x u(t,X_t)|^2\big)^{\frac{1}{2}}dt\Big\}^2 + 2\Big\{\int_0^T \big(\mathbb{E}|D^2u(t,X_t)|^2\big)^{\frac{1}{2}}dt\Big\}^2,
\end{aligned}$$

using the generalized Minkowski inequality. Besides, from Lemma 5.7.1, and using that $g \in \mathbf{L}_{2,\alpha}$, one obtains

$$\mathbb{E}\,|\nabla_x u(t,X_t)|^2 \leq_c \frac{K^\alpha(g)}{(T-t)^{1-\alpha}},\ \mathbb{E}\,\big|D^2u(t,X_t)\big|^2 \leq_c \frac{K^\alpha(g)}{(T-t)^{2-\alpha}}.$$

Now, the required result easily follows. □

Proof of Proposition 5.7.1.
We prove that (i)⇒ (ii)⇒ (iii)⇒ (i).
<u>(i)⇒ (ii)</u>. By Itô's rule,

$$\begin{aligned}
\partial_{x_k}u(t,X_t) &= \partial_{x_k}u(0,X_0) + \int_0^t \{\nabla(\partial_{x_k}u)\sigma\}(s,X_s)dW_s \\
&\quad + \int_0^t \Big\{\partial_t\partial_{x_k}u + \nabla(\partial_{x_k}u)b + \frac{1}{2}\mathrm{Tr}\left[\sigma\sigma^* D^2(\partial_{x_k}u)\right]\Big\}(s,X_s)ds.
\end{aligned}$$

In order to get rid of the terms $\partial_t\partial_{x_k}u$ and $D^2(\partial_{x_k}u)$, differentiate the PDE (5.11) solved by u:

$$\begin{aligned}
0 &= \partial_{x_k}\big(\partial_t u + \nabla u b + \frac{1}{2}\mathrm{Tr}\big[\sigma\sigma^* D^2 u\big]\big) \\
&= \big(\partial_t\partial_{x_k}u + \nabla(\partial_{x_k}u)b + \frac{1}{2}\mathrm{Tr}\big[\sigma\sigma^* D^2(\partial_{x_k}u)\big]\big) + \big(\nabla u\partial_{x_k}b + \frac{1}{2}\mathrm{Tr}\big[\partial_{x_k}(\sigma\sigma^*)D^2u\big]\big).
\end{aligned}$$

Consequently

$$\begin{aligned}
\partial_{x_k}u(t,X_t) &= \partial_{x_k}u(0,X_0) - \int_0^t \Big\{\nabla u\partial_{x_k}b + \frac{1}{2}\mathrm{Tr}\left[\partial_{x_k}(\sigma\sigma^*)D^2u\right]\Big\}(s,X_s)ds \\
&\quad + \int_0^t \{\nabla(\partial_{x_k}u)\sigma\}(s,X_s)dW_s. \qquad (5.21)
\end{aligned}$$

Then,

$$\begin{aligned}
&\mathbb{E}\left|\nabla_x u(t,X_t)\right|^2 = \sum_{k=1}^{d}\mathbb{E}|\partial_{x_k}u(t,X_t)|^2\\
&\leq_c \mathbb{E}\left|\nabla_x u(0,X_0)\right|^2 + \mathbb{E}\left(\int_0^t \left\{|\nabla_x u| + |D^2u|\right\}(s,X_s)ds\right)^2 + \int_0^t \mathbb{E}|D^2u(s,X_s)|^2 ds\\
&\leq_c \mathbb{E}\left|\nabla_x u(0,X_0)\right|^2 + \int_0^t \mathbb{E}|\nabla_x u(s,X_s)|^2 ds + \int_0^t \mathbb{E}|D^2u(s,X_s)|^2 ds \qquad (5.22)\\
&\leq_c \frac{K^\alpha(g)}{T^{1-\alpha}} + \mathbb{E}|g(X_T)|^2 + \int_0^t \mathbb{E}|D^2u(s,X_s)|^2 ds := \phi(t). \qquad (5.23)
\end{aligned}$$

Then, by integrating by parts, one has

$$\begin{aligned}
&\int_0^T (T-r)^{-\beta}\mathbb{E}|\nabla_x u(r,X_r)|^2 dr \leq_c \varlimsup_{s\uparrow T}\int_0^s (T-r)^{-\beta}\phi(r)dr\\
&= \varlimsup_{s\uparrow T}\left(-\left[\frac{(T-r)^{1-\beta}}{1-\beta}\phi(r)\right]_0^s + \frac{1}{1-\beta}\int_0^s (T-r)^{1-\beta}\phi'(r)dr\right).
\end{aligned}$$

Note that the first term in the limit is bounded by $\frac{T^{1-\beta}}{1-\beta}\left(\frac{K^\alpha(g)}{T^{1-\alpha}} + \mathbb{E}|g(X_T)|^2\right)$. The second term is bounded by $\frac{1}{1-\beta}\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2u(r,X_r)|^2 dr$, which is finite because assertion (i) is in force.

<u>(ii)$\Rightarrow$ (iii)</u>: the proof is similar to the previous one. In view of (5.12-5.13), one has

$$\mathbb{E}\left|g(X_T) - \mathbb{E}^{\mathcal{F}_r}(g(X_T))\right|^2 \leq_c \int_r^T \mathbb{E}|\nabla_x u(s,X_s)|^2 ds := \Psi(r).$$

Then, using an integration by parts, one gets

$$\begin{aligned}
&\int_0^T (T-r)^{-1-\beta}\mathbb{E}\left|g(X_T) - \mathbb{E}^{\mathcal{F}_r}(g(X_T))\right|^2 dr\\
&\leq_c \varlimsup_{s\uparrow T}\int_0^s (T-r)^{-1-\beta}\Psi(r)dr\\
&\leq_c \varlimsup_{s\uparrow T}\left(\left[\frac{(T-r)^{-\beta}}{\beta}\Psi(r)\right]_0^s + \frac{1}{\beta}\int_0^s (T-r)^{-\beta}\mathbb{E}|\nabla_x u(r,X_r)|^2 dr\right).
\end{aligned}$$

The second term is finite according to assertion (ii). The first one is bounded by

$$\varlimsup_{s\uparrow T}\frac{(T-s)^{-\beta}}{\beta}\int_s^T \mathbb{E}|\nabla_x u(r,X_r)|^2 dr \leq \varlimsup_{s\uparrow T}\frac{1}{\beta}\int_s^T (T-r)^{-\beta}\mathbb{E}|\nabla_x u(r,X_r)|^2 dr,$$

because $(T-r)^{-\beta}$ is increasing with respect to r. The limit above equals then to 0 since the related integral is convergent.

(iii)$\Rightarrow$ (i). From Lemma 5.7.1, one has $\mathbb{E}|D^2u(r,X_r)|^2 \leq_c \frac{V_{r,T}(g)}{(T-r)^2}$, from which we deduce (using assertion (iii)) that

$$\int_0^T (T-r)^{1-\beta}\mathbb{E}|D^2u(r,X_r)|^2dr \leq_c \int_0^T (T-r)^{-1-\beta}V_{r,T}(g)dr < +\infty.$$

□

Proof of Lemma 5.7.2.
We prove that (i)$\Rightarrow$ (ii)$\Rightarrow$ (iii)$\Rightarrow$(i).

(i)$\Rightarrow$ (ii):

$$\begin{aligned}\int_0^t \mathbb{E}|D^2u(s,X_s)|^2ds &= \sum_{k=1}^d \int_0^t \mathbb{E}|\nabla(\partial_{x_k}u)|^2(s,X_s)ds \\ &\leq_c \sum_{k=1}^d \int_0^t \mathbb{E}|\nabla(\partial_{x_k}u)\sigma|^2(s,X_s)ds,\end{aligned}$$

by the ellipticity assumption. Then, by the isometry property and equality (5.21), one gets that $\int_0^t \mathbb{E}|D^2u(s,X_s)|^2ds$ is bounded (up to a constant) by

$$\begin{aligned}&\sum_{k=1}^d \mathbb{E}\big(\int_0^t \nabla(\partial_{x_k}u)\sigma(s,X_s)dW_s\big)^2 \\ &= \sum_{k=1}^d \mathbb{E}\big(\partial_{x_k}u(t,X_t) - \partial_{x_k}u(0,X_0) + \int_0^t \big\{\nabla u\partial_{x_k}b + \frac{1}{2}\mathrm{Tr}\big[\partial_{x_k}(\sigma\sigma^*)D^2u\big]\big\}(s,X_s)ds\big)^2 \\ &\leq_c \mathbb{E}\big|\nabla_x u(t,X_t)\big|^2 + \mathbb{E}\big|\nabla_x u(0,X_0)\big|^2 + \mathbb{E}\big(\int_0^t \big\{|\nabla_x u| + |D^2u|\big\}(s,X_s)ds\big)^2 \\ &\leq_c \frac{K^\alpha(g)}{(T-t)^{1-\alpha}} + \frac{K^\alpha(g)}{T^{1-\alpha}} + K^\alpha(g)T^\alpha,\end{aligned}$$

where we have used Lemma 5.7.1 and Corollary 5.7.1.

(ii)$\Rightarrow$ (iii): this is an immediate consequence of the inequality (5.23).

(iii)$\Rightarrow$ (i): from equation (5.12), one gets

$$V_{t,T}(g) \leq_c \int_t^T \mathbb{E}\left|\nabla_x u(s,X_s)\right|^2 ds \leq_c \int_t^T \frac{C^\alpha(g)}{(T-s)^{1-\alpha}}ds \leq_c C^\alpha(g)(T-t)^\alpha,$$

which means that $g \in \mathbf{L}_{2,\alpha}$.

The fact that (i) leads to (iv) follows from Lemma 5.7.1, and it is clear that, when $\alpha < 1$, (iv) leads to (ii). □

In the following section, we state some results on $Z_t - z_t$ that, put together with those of section 5.7, will be crucial to study the $\mathbf{L}_2$-regularity of the Z-component of BSDEs with non null generator.

5.8 A representation and an estimate of $Z_t - z_t$ when the terminal condition g is bounded and belongs to $\mathbf{L}_{2,\alpha}$

5.8.1 The main result

We define

$$Y_t^0 := Y_t - y_t, \qquad Z_t^0 := Z_t - z_t.$$

Then, the process (Y^0, Z^0) is the solution, in $\mathcal{S}^2 \times \mathcal{M}^2$ (because (Y, Z) and (y, z) are in such spaces), of the BSDE with null terminal condition and generator

$$f^0(t, x, y, z) := f\left(t, x, y + u(t, x), z + \nabla_x u(t, x)\sigma(t, x)\right),$$

i.e.

$$Y_t^0 = \int_t^T f^0(s, X_s, Y_s^0, Z_s^0)ds - \int_t^T Z_s^0 dW_s.$$

We set

$$\begin{aligned} a_r^0 &:= \nabla_x f^0(r, X_r, Y_r^0, Z_r^0); \\ b_r^0 &:= \nabla_y f^0(r, X_r, Y_r^0, Z_r^0); \\ c_r^0 &:= \nabla_z f^0(r, X_r, Y_r^0, Z_r^0). \end{aligned}$$

These quantities play a key role in the further estimates.
Note that

$$\begin{aligned} &|a_r^0| \leq C\left(1 + |\nabla_x u(r, X_r)| + |D^2 u(r, X_r)|\right); \\ &|b_r^0| \leq C; \\ &|c_r^0| \leq C. \end{aligned} \tag{5.24}$$

Hence, f^0 is Lipschitz continuous with respect to y and z, but not with respect to x because $\nabla_x u$ and $D^2 u$ may explode as t goes to T.
Our purpose is to estimate $Z - z = Z^0$, and it is known that usually the Z^0-component is related to the Malliavin derivative of the Y^0-component (see Proposition 5.3. in [KPQ97b]). But this is stated under strong integrability conditions: namely in Prop.5.3.[KPQ97b], it is required that $\mathbb{E}\int_0^T \int_0^T |D_\theta f^0(s, X_s, y, z)|^2 ds d\theta < \infty$, for any y and z. This is not satisfied in our case (since it essentially means that $\mathbb{E}\int_0^T |D^2 u(s, X_s)|^2 ds < +\infty$). However, we are going to prove that the expected result (relating Z^0 to Malliavin derivatives) holds in our setting ($g \in \mathbf{L}_{2,\alpha}$). We proceed by a localization of the generator (see paragraph 5.8.2).
Before giving our main result, we introduce (U, V) the solution of the linear BSDE

$$\begin{aligned} U_t = &\int_t^T \left\{ a_r^0 + U_r \left(b_r^0 I_d + b_r^{'} + \sum_{j=1}^q c_{j,r}^0 \sigma_{j,r}^{'} \right) + \sum_{j=1}^q V_r^j \left(c_{j,r}^0 I_d + \sigma_{j,r}^{'} \right) \right\} dr \\ &- \sum_{j=1}^q \int_t^T V_r^j dW_r^j, \end{aligned} \tag{5.25}$$

where $b_r^{'}$ and $\sigma_{j,r}^{'}$ denote respectively $\nabla_x b(r, X_r)$ and $\nabla_x \sigma_j(r, X_r)$, and $c_{j,r}^0$ is the j-th component of c_r^0. It is well defined in $\mathcal{S}^2 \times \mathcal{M}^2$ (see Lemma 5.A.1 in Appendix 5.A) because it follows from Lemma 5.7.1 and inequality (5.24) that

$$\mathbb{E}\left(\int_0^T |a_r^0| dr\right)^2 \leq_c T^\alpha K^\alpha(g) < +\infty. \tag{5.26}$$

Our main result is stated as follows:

Theorem 5.8.1. *Assume* $(\boldsymbol{A}_{b,\sigma})$, $(\boldsymbol{A}_g)$ *and* $(\boldsymbol{A}_f)$. *Assume moreover that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0, 1]$. *Then,* $\mathrm{d}\mathbb{P} \otimes \mathrm{d}t - a.s.$, *one has*

$$Z_t - z_t = U_t \sigma(t, X_t).$$

In particular, since $z_t + U_t\sigma(t, X_t)$ is continuous, Z has a continuous version : this extends the results by Ma and Zhang [MZ02], in the case when g is continuously differentiable with bounded derivative. We work with this version in the sequel.

Remark 5.8.1. $Z_t - z_t$ *has also a closed representation as a conditional expectation: see equation (5.38) in the proof.*

It is now easy to derive pointwise and $\mathbf{L}_2$-estimates of $Z_t - z_t$ as t goes to T. We will not use the following estimates in the sequel, but we guess that they are interesting for themselves.

Corollary 5.8.1. *Assume* $(\boldsymbol{A}_{b,\sigma})$, $(\boldsymbol{A}_g)$ *and* $(\boldsymbol{A}_f)$. *Assume moreover that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0, 1]$. *Then, for all* $t \in [0, T)$, *the following pointwise estimate and* $\mathbf{L}_2$-*estimate hold:*

$$|Z_t - z_t| \leq C \int_t^T \frac{\sqrt{\mathbb{E}^{\mathcal{F}_t}\left[\left(g(X_T) - \mathbb{E}^{\mathcal{F}_s} g(X_T)\right)^2\right]}}{T - s} ds + C(T - t);$$

$$\mathbb{E}\,|Z_t - z_t|^2 \leq CK^\alpha(g)(T - t)^\alpha + C(T - t)^2.$$

Remark 5.8.2. *When* g *is bounded and* α-*Hölder continuous (i.e.* $|g(x) - g(y)| \leq C(g)|x - y|^\alpha$*), the pointwise estimate in Corollary 5.8.1 leads to* $|Z_t - z_t| \leq_c C(g)(T - t)^{\frac{\alpha}{2}} + (T - t)$.
Since $(z_t)_{0 \leq t \leq T}$ *may explode at time* T, *this is a way to assert that* Z *and* z *are equivalent for times close to* T.

Proof of Corollary 5.8.1. Theorem 5.8.1 and the conditional version of estimate (5.50) yield

$$|Z_t - z_t| \leq_c |U_t| \leq_c \left\{\mathbb{E}^{\mathcal{F}_t}\left(\int_t^T |a_r^0| dr\right)^2\right\}^{\frac{1}{2}} \leq_c \int_t^T \left(\mathbb{E}^{\mathcal{F}_t}|a_r^0|^2\right)^{\frac{1}{2}} dr$$

using the generalized Minkowski inequality for the last inequality. From (5.24), one has

$$\mathbb{E}^{\mathcal{F}_t}|a_r^0|^2 \leq_c 1 + \mathbb{E}^{\mathcal{F}_t}|\nabla_x u(r, X_r)|^2 + \mathbb{E}^{\mathcal{F}_t}|D^2 u(r, X_r)|^2.$$

Therefore, by means of the pointwise estimates obtained in the proof of Lemma 5.7.1 (see inequality (5.20) for $\nabla_x u(r, X_r)$), one gets (for $t \leq r < T$)

$$\mathbb{E}^{\mathcal{F}_t}|a_r^0|^2 \leq_c 1 + \frac{\mathbb{E}^{\mathcal{F}_t}[(g(X_T) - \mathbb{E}^{\mathcal{F}_r} g(X_T))^2]}{(T-r)^2},$$

and

$$|Z_t - z_t| \leq_c \int_t^T 1 + \frac{\sqrt{\mathbb{E}^{\mathcal{F}_t}[(g(X_T) - \mathbb{E}^{\mathcal{F}_r} g(X_T))^2]}}{T-r} dr,$$

proving the pointwise estimate. Consequently, using the generalized Minkowski inequality and $g \in \mathbf{L}_{2,\alpha}$, one has

$$\begin{aligned}
\mathbb{E}\,|Z_t - z_t|^2 &\leq_c \Big(\int_t^T \frac{\sqrt{\mathbb{E}[(g(X_T) - \mathbb{E}^{\mathcal{F}_r} g(X_T))^2]}}{T-r} dr \Big)^2 + (T-t)^2 \\
&\leq_c \big(\sqrt{K^\alpha(g)} \int_t^T \frac{1}{(T-r)^{1-\frac{\alpha}{2}}} dr \big)^2 + (T-t)^2 \\
&\leq_c K^\alpha(g)(T-t)^\alpha + (T-t)^2.
\end{aligned}$$

□

5.8.2 Proof of Theorem 5.8.1

Since $a_r^0 = \nabla_x f^0(r, X_r, Y_r^0, Z_r^0)$ may explode as t goes to T, we proceed by a time localization of f^0 as follows: for $\varepsilon > 0$, we define:

$$f^\varepsilon(t, x, y, z) = f^0(t, x, y, z) \mathbb{1}_{t \leq T - \varepsilon}$$

and $(Y_t^\varepsilon, Z_t^\varepsilon)$ the solution, in $\mathcal{S}^2 \times \mathcal{M}^2$, of the localized BSDE:

$$Y_t^\varepsilon = \int_t^T f^\varepsilon(s, X_s, Y_s^\varepsilon, Z_s^\varepsilon) ds - \int_t^T Z_s^\varepsilon dW_s.$$

As for a_s^0, b_s^0 and c_s^0, we define

$$\begin{aligned}
a_s^\varepsilon &:= \nabla_x f^\varepsilon(s, X_s, Y_s^\varepsilon, Z_s^\varepsilon), \\
b_s^\varepsilon &:= \nabla_y f^\varepsilon(s, X_s, Y_s^\varepsilon, Z_s^\varepsilon), \\
c_s^\varepsilon &:= \nabla_z f^\varepsilon(s, X_s, Y_s^\varepsilon, Z_s^\varepsilon).
\end{aligned}$$

We assume $(\mathbf{A}_{b,\sigma}), (\mathbf{A}_g), (\mathbf{A}_f)$, and that $g \in \mathbf{L}_{2,\alpha}$, for some $\alpha \in (0, 1]$.
The idea of our proof of Theorem 5.8.1 is the following: we prove that Z^ε converges to Z^0 as ε goes to 0 (Lemma 5.8.1) and that $D_t Y_t^\varepsilon$ converges to some $\mathcal{D}_t Y_t^0$ satisfying a linear BSDE (Lemma 5.8.4). Then, since $Z_t^\varepsilon = D_t Y_t^\varepsilon$ (Lemma 5.8.2), we conclude

that $Z_t^0 = \mathcal{D}_t Y_t^0$. Finally, we derive the BSDE (5.25) satisfied by $(U_t)_{0\le t\le T}$ from that satisfied by $(\mathcal{D}_t Y_t^0)_{0\le t\le T}$.

Step 1: Stability

Lemma 5.8.1.

$$\lim_{\varepsilon\to 0}\mathbb{E}\left[\sup_{s\in[0,T]}|Y_s^\varepsilon - Y_s^0|^2 + \int_0^T |Z_s^\varepsilon - Z_s^0|^2 ds\right] = 0.$$

Proof. We denote $Y_s^\varepsilon - Y_s^0$ by δY_s and $Z_s^\varepsilon - Z_s^0$ by δZ_s. Then $(\delta Y_s, \delta Z_s)_{0\le s\le T}$ is the solution in $\mathcal{S}^2\times\mathcal{M}^2$ to the BSDE with a null terminal condition and the (random) generator

$$\begin{aligned}\delta f(t,y,z) &:= f^\varepsilon(t,X_t,y+Y_t^0,z+Z_t^0) - f^0(t,X_t,Y_t^0,Z_t^0)\\ &= \left[f^0(t,X_t,y+Y_t^0,z+Z_t^0) - f^0(t,X_t,Y_t^0,Z_t^0)\right]\mathbb{1}_{t\le T-\varepsilon}\\ &\quad - f^0(t,X_t,Y_t^0,Z_t^0)\mathbb{1}_{t>T-\varepsilon}.\end{aligned}$$

Since the function $f^0(t,x,.,.)$ is Lipschitz continuous (and its Lipschitz constant is the same as that of f),

$$|\delta f(t,y,z)| \le C|y| + C|z| + \mathbb{1}_{t>T-\varepsilon}|f^0(t,X_t,Y_t^0,Z_t^0)|.$$

Then, thanks to a standard stability result (see Proposition 3.2 in [BDH$^+$03]), one obtains the following estimate:

$$\mathbb{E}\left[\sup_{s\in[0,T]}|\delta Y_s|^2 + \int_0^T |\delta Z_s|^2 ds\right] \le C\mathbb{E}\left(\int_0^T \mathbb{1}_{s>T-\varepsilon}|f^0(s,X_s,Y_s^0,Z_s^0)|ds\right)^2.$$

Now, one has $f^0(s,X_s,Y_s^0,Z_s^0) = f(s,X_s,Y_s,Z_s)$, which is square integrable, since (Y,Z) is the solution in $\mathcal{S}^2\times\mathcal{M}^2$ of BSDE (5.1). Then, by the monotone convergence theorem, the above upper bound converges to 0 as ε goes to 0. □

Step 2: Malliavin derivatives

Lemma 5.8.2. *For any fixed $\varepsilon > 0$, $(Y^\varepsilon, Z^\varepsilon)$ belongs to $\mathbf{L}^2\left((0,T),\mathbb{D}^{1,2}\times(\mathbb{D}^{1,2})^q\right)$. Denoting their Malliavin derivatives by $(D_tY_s^\varepsilon, D_tZ_s^\varepsilon)$, one has, a.s., for all $t\in[0,T)$ (see [KPQ97b]),*

$$Z_t^\varepsilon = D_tY_t^\varepsilon.$$

Proof. This is a direct application of Proposition 5.9 of [KPQ97b]. It remains to show that f^ε has bounded derivatives w.r.t (x,y,z). One has, for $w = x,y,z$,

$$\nabla_w f^\varepsilon(t,x,y,z) = \nabla_w f^0(t,x,y,z)\mathbb{1}_{t\le T-\varepsilon},$$

and, if one defines $\theta := (t, x, y + u(t,x), z + \nabla_x u(t,x)\sigma(t,x))$, then

$$\begin{aligned}
\nabla_x f^0(t,x,y,z) &= \nabla_x f(\theta) + \nabla_y f(\theta)\nabla_x u(t,x) + \nabla_z f(\theta)\nabla_x[(\nabla_x u\sigma)^*](t,x),\\
\nabla_y f^0(t,x,y,z) &= \nabla_y f(\theta),\\
\nabla_z f^0(t,x,y,z) &= \nabla_z f(\theta).
\end{aligned}$$

Hence, only the boundedness of $\nabla_x f^\varepsilon(t,x,y,z)$ needs to be justified. This readily follows from $|\nabla_x u(t,x)| + |D^2 u(t,x)| \le C \|g\|_\infty / (T-t) \le C \|g\|_\infty / \varepsilon$, for $t \le T - \varepsilon$, where the first inequality is proved as in Lemma 5.7.1 (using the boundedness of g). Then, $dt \otimes d\mathbb{P}$-a.s., one has $Z_t^\varepsilon = D_t Y_t^\varepsilon$. In addition, in Ma and Zhang[MZ02], it is proved that the above processes have a continuous version, which enables to pass to an a.s. equality for any t. □

Note that, always from [KPQ97b], $(D_t Y_s^\varepsilon)_{s \ge t}$ $(\varepsilon > 0)$ satisfies to the following linear BSDE:

$$D_t Y_s^\varepsilon = \int_s^T \{a_r^\varepsilon D_t X_r + b_r^\varepsilon D_t Y_r^\varepsilon + \sum_{j=1}^q c_{j,r}^\varepsilon D_t Z_{j,r}^\varepsilon\} dr - \sum_{j=1}^q \int_s^T D_t Z_{j,r}^\varepsilon dW_r^j. \quad (5.27)$$

and, for $s \in [0,t)$, $(D_t Y_s^\varepsilon, D_t Z_s^\varepsilon) = (0,0)$.
We introduce $(\mathcal{D}_t Y_s^0, \mathcal{D}_t Z_s^0)_{t \le s \le T}$, as the unique solution, in $\mathcal{S}^p \times \mathcal{M}^p$ (for $p \in (1,2)$) to the following BSDE:

$$\mathcal{D}_t Y_s^0 = \int_s^T \left\{ a_r^0 D_t X_r + b_r^0 \mathcal{D}_t Y_r^0 + \sum_{j=1}^q c_{j,r}^0 \mathcal{D}_t Z_{j,r}^0 \right\} dr - \sum_{j=1}^q \int_s^T \mathcal{D}_t Z_{j,r}^0 dW_r^j. \quad (5.28)$$

For $s \in [0,t)$, $(\mathcal{D}_t Y_s^0, \mathcal{D}_t Z_s^0) := (0,0)$.
Note that BSDE (5.28) is well defined, applying Lemma 5.A.1. In fact, b_r^0 et c_r^0 are uniformly bounded, and from (5.24),

$$\mathbb{E}\left(\int_t^T \left|a_r^0 D_t X_r\right| dr\right)^p \le_c \mathbb{E}\left[\sup_{0 \le r \le T} |D_t X_r|^p \left(\int_0^T 1 + |\nabla_x u(r, X_r)| + |D^2 u(r, X_r)| dr\right)^p\right]. \quad (5.29)$$

This upper bound is finite using Hölder's inequality. Indeed, $\sup_{0 \le r \le T} |D_t X_r|^p$ is in any $\mathbf{L}_q$ (see remark after inequality (5.9)) and the integral term is in $\mathbf{L}_2$ (Corollary 5.7.1).
Note that $\mathcal{D}_t Y_s^0$ is given by the following closed formula (which is standard for linear BSDEs, see e.g. [KPQ97b])

$$\mathcal{D}_t Y_s^0 = \mathbb{E}^{\mathcal{F}_s} \int_s^T \Gamma_r^s a_r^0 D_t X_r dr, \quad (5.30)$$

where $(\Gamma_r^s)_{r \ge s}$ is the adjoint process defined by the forward linear SDE

$$d\Gamma_r^s = \Gamma_r^s \left(b_r^0 dr + c_r^0 dW_r\right), \qquad \Gamma_s^s = 1.$$

Furthermore, define $(\nabla Y_s^0, \nabla Z_s^0)_{0\le s\le T}$, to be the unique solution, in $\mathcal{S}^p \times \mathcal{M}^p$ $(1 < p < 2)$, to the following BSDE:

$$\nabla Y_s^0 = \int_s^T \left\{ a_r^0 \nabla X_r + b_r^0 \nabla Y_r^0 + \sum_{j=1}^q c_{j,r}^0 \nabla Z_{j,r}^0 \right\} dr - \sum_{j=1}^q \int_s^T \nabla Z_{j,r}^0 dW_r^j. \quad (5.31)$$

This is a slight abuse of notation because $(\nabla Y^0, \nabla Z^0)$ solves the BSDE obtained by differentiating with respect to x_0 the BSDE solved by (Y^0, Z^0), but we do not prove that $(\nabla Y^0, \nabla Z^0)$ are the gradients of (Y^0, Z^0) with respect to x_0 (however, this is true, using extra computations as before).
Then, from (5.9) it follows that

$$\mathcal{D}_t Y_s^0 = \nabla Y_s^0 [\nabla X_t]^{-1} \sigma(t, X_t), \qquad \mathcal{D}_t Z_s^0 = \nabla Z_s^0 [\nabla X_t]^{-1} \sigma(t, X_t). \quad (5.32)$$

Step 3: Convergence of $D_t Y_s^\varepsilon$ to $\mathcal{D}_t Y_s^0$ as ε goes to 0
To justify this step, we repeatedly use the lemma below.

Lemma 5.8.3. *Let φ be a bounded continuous deterministic function. Let (Θ^ε), (Θ^0) and (γ) be predictable processes such that*

$$(i)\ \mathbb{E}\int_0^T |\Theta_s^0|^2 ds < +\infty;$$

$$(ii)\ \lim_{\varepsilon\to 0} \mathbb{E}\int_0^T |\Theta_s^\varepsilon - \Theta_s^0|^2 ds = 0;$$

$$(iii)\ \exists p \in (1,2) \text{ such that } \mathbb{E}\Big(\int_0^T |\gamma_s| ds\Big)^p < +\infty.$$

Then, $\lim_{\varepsilon\to 0} \mathbb{E}\Big(\int_0^T |\varphi(\Theta_s^\varepsilon) - \varphi(\Theta_s^0)||\gamma_s| ds\Big)^p = 0.$

Proof. Let K_1, K_2 and δ be positive constants. It is clear that

$$\begin{aligned}
&\mathbb{E}\Big(\int_0^T |\varphi(\Theta_s^\varepsilon) - \varphi(\Theta_s^0)||\gamma_s| ds\Big)^p \\
&\le \big(2\|\varphi\|_\infty\big)^p \mathbb{E}\Big(\int_0^T |\gamma_s| \mathbb{1}_{|\gamma_s|>K_1} ds\Big)^p + K_1^p \big(2\|\varphi\|_\infty\big)^p \mathbb{E}\Big(\int_0^T \mathbb{1}_{|\Theta_s^0|>K_2} ds\Big)^p \\
&+ K_1^p \big(2\|\varphi\|_\infty\big)^p \mathbb{E}\Big(\int_0^T \mathbb{1}_{|\Theta_s^\varepsilon - \Theta_s^0|>\delta} ds\Big)^p \\
&+ K_1^p \mathbb{E}\Big(\int_0^T |\varphi(\Theta_s^\varepsilon) - \varphi(\Theta_s^0)| \mathbb{1}_{|\Theta_s^0|\le K_2, |\Theta_s^\varepsilon - \Theta_s^0|\le\delta} ds\Big)^p
\end{aligned} \quad (5.33)$$

Let $\eta > 0$. Firstly, by assumption (iii), K_1 can be chosen large enough to ensure that the first term is bounded by $\frac{\eta}{4}$. Besides, from Chebyshev inequality, one has

$$\mathbb{E}\mathbb{1}_{|\Theta_s^0|>K_2} + \mathbb{E}\mathbb{1}_{|\Theta_s^\varepsilon - \Theta_s^0|>\delta} \le \frac{\mathbb{E}|\Theta_s^0|^{2/p}}{K_2^{2/p}} + \frac{\mathbb{E}|\Theta_s^\varepsilon - \Theta_s^0|^{2/p}}{\delta^{2/p}},$$

so that, owing to assumption (i), one can choose K_2 large enough to make the second term in (5.33) bounded by $\frac{\eta}{4}$. Thirdly, since φ is continuous on the compact $[-K_2-1, K_2+1]$, it is also uniformly continuous on the same compact, and one can choose δ small enough to ensure that the last term in (5.33) is bounded by $\frac{\eta}{4}$. Finally, owing to assumption (ii), one can choose ε small enough to make the third term in (5.33) bounded by $\frac{\eta}{4}$.
Therefore, we have proved that for any $\eta > 0$, $\mathbb{E}\big(\int_0^T |\varphi(\Theta_s^\varepsilon) - \varphi(\Theta_s^0)||\gamma_s|ds\big)^p \leq \eta$, provided that ε is small enough. □

Lemma 5.8.4.

$$\lim_{\varepsilon\to 0} \mathbb{E}[\sup_{s\in[t,T]} |D_tY_s^\varepsilon - \mathcal{D}_tY_s^0|^p] = 0.$$

Proof. $(D_tY_s^\varepsilon - \mathcal{D}_tY_s^0, D_tZ_s^\varepsilon - \mathcal{D}_tZ_s^0)_{t\leq s\leq T}$ satisfies the linear BSDE:

$$\begin{aligned} D_tY_s^\varepsilon - \mathcal{D}_tY_s^0 = &\int_s^T [(a_r^\varepsilon - a_r^0)D_tX_r + (b_r^\varepsilon - b_r^0)\mathcal{D}_tY_r^0 + \sum_{j=1}^q (c_{j,r}^\varepsilon - c_{j,r}^0)\mathcal{D}_tZ_{j,r}^0]dr \\ &+ \int_s^T [b_r^\varepsilon\big(D_tY_r^\varepsilon - \mathcal{D}_tY_r^0\big) + \sum_{j=1}^q c_{j,r}^\varepsilon\big(D_tZ_r^\varepsilon - \mathcal{D}_tZ_{j,r}^0\big)]dr \\ &- \sum_{j=1}^q \int_s^T \big(D_tZ_r^\varepsilon - \mathcal{D}_tZ_{j,r}^0\big)dW_r^j. \end{aligned}$$

Set $\eta_r^{\varepsilon,a} := (a_r^\varepsilon - a_r^0)D_tX_r$, $\eta_r^{\varepsilon,b} := (b_r^\varepsilon - b_r^0)\mathcal{D}_tY_r^0$ and $\eta_r^{\varepsilon,c} := \sum_{j=1}^q (c_{j,r}^\varepsilon - c_{j,r}^0)\mathcal{D}_tZ_{j,r}^0$. Using the a priori $\mathbf{L}_p$-estimate in Lemma 5.A.1, one has

$$\mathbb{E}[\sup_{s\in[t,T]} |D_tY_s^\varepsilon - \mathcal{D}_tY_s^0|^p] \leq_c \mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,a}|dr\Big)^p + \mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,b}|dr\Big)^p + \mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,c}|dr\Big)^p. \tag{5.34}$$

Let us prove that each contribution with $\eta_\cdot^{\varepsilon,a}$, $\eta_\cdot^{\varepsilon,b}$ and $\eta_\cdot^{\varepsilon,c}$ converges to 0.
Contribution with $\eta_\cdot^{\varepsilon,a}$. It is clear that

$$\begin{aligned} |\eta_r^{\varepsilon,a}| \leq &|\nabla_x f^0(r, X_r, Y_r^\varepsilon, Z_r^\varepsilon) - \nabla_x f^0(r, X_r, Y_r^0, Z_r^0)||D_tX_r| \\ &+ \mathbb{1}_{r>T-\varepsilon}|\nabla_x f^0(r, X_r, Y_r^0, Z_r^0)||D_tX_r|. \end{aligned}$$

First, note that $|\nabla_x f^0(t,x,y,z)| \leq_c 1 + |\nabla_x u(t,x)| + |D^2u(t,x)|$. This implies that, uniformly in ε, $\mathbb{1}_{r>T-\varepsilon}|\nabla_x f^0(r, X_r, Y_r^0, Z_r^0)||D_tX_r| \leq_c \big(1 + |\nabla_x u(r,X_r)| + |D^2u(r,X_r)|\big)|D_tX_r|$, whose integral w.r.t. r belongs to $\mathbf{L}_p$ (this has been established in the proof of the existence of BSDE (5.28)), and is consequently a.s. finite. It readily follows from the dominated convergence theorem that $\mathbb{E}\big(\int_0^T \mathbb{1}_{r>T-\varepsilon}|\nabla_x f^0(r, X_r, Y_r^0, Z_r^0)||D_tX_r|\big)^p$ converges to 0 as ε goes to 0.
Next, setting $\Theta_r^\varepsilon := \big(r, X_r, Y_r^\varepsilon + u(r,X_r), Z_r^\varepsilon + \nabla_x u(r,X_r)\sigma(r,X_r)\big)$, $\varphi_w(\Theta) := \nabla_w f(\Theta)$, for $w = x,y,z$, and $\gamma_r := \big(1 + |\nabla_x u(r,X_r)| + |D^2u(r,X_r)|\big)|D_tX_r|$, one has

$$|\nabla_x f^0(\Theta_r^\varepsilon) - \nabla_x f^0(\Theta_r^0)||D_tX_r| \leq_c \sum_{w=x,y,z} |\varphi_w(\Theta_r^\varepsilon) - \varphi_w(\Theta_r^0)|\gamma_r.$$

Owing to assumption $(\mathbf{A}_f)$, φ_w is continuous and bounded. Assumptions (i) and (ii) of Lemma 5.8.3 hold thanks to Lemma 5.8.1 and to the fact that $\Theta_r^0 = (r, X_r, Y_r, Z_r)$. Assumption (iii) is checked since $\mathbb{E}(\int_0^T \gamma_s ds)^p < +\infty$, for $p \in (1,2)$ (see inequality (5.29) previously proved). It follows from Lemma 5.8.3 that $\mathbb{E}(\int_0^T |\nabla_x f^0(\Theta_r^\varepsilon) - \nabla_x f^0(\Theta_r^0)||D_t X_r|dr)^p$ converges to 0 as ε goes to 0.
Finally,

$$\mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,a}|dr\Big)^p \to 0. \tag{5.35}$$

Contribution with $\eta_\cdot^{\varepsilon,b}$. One has

$$|\eta_r^{\varepsilon,b}| \leq_c |\nabla_y f(\Theta_r^\varepsilon) - \nabla_y f(\Theta_r^0)||\mathcal{D}_t Y_r^0| + \mathbb{1}_{r>T-\varepsilon}|\nabla_y f^0(r, X_r, Y_r^0, Z_r^0)||\mathcal{D}_t Y_r^0|.$$

Then, we follow exactly the same proof as that of the contribution of $\eta_\cdot^{\varepsilon,a}$. One has only to check that $\gamma_r := |\mathcal{D}_t Y_r^0|$ satisfies assumption (iii) of Lemma 5.8.3. This readily follows from $(\mathcal{D}_t Y_r^0)_{t\leq r\leq T} \in \mathcal{S}^p$, which ensures that

$$\mathbb{E}\Big(\int_0^T \gamma_r dr\Big)^p \leq_c \mathbb{E}\sup_{r\in[t,T]} |\mathcal{D}_t Y_r^0|^p < +\infty.$$

Thus,

$$\mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,b}|dr\Big)^p \to 0. \tag{5.36}$$

Contribution with $\eta_\cdot^{\varepsilon,c}$. One has

$$|\eta_r^{\varepsilon,c}| \leq_c |\nabla_z f(\Theta_r^\varepsilon) - \nabla_z f(\Theta_r^0)||\mathcal{D}_t Z_r^0| + \mathbb{1}_{r>T-\varepsilon}|\nabla_y f^0(r, X_r, Y_r^0, Z_r^0)||\mathcal{D}_t Z_r^0|.$$

Similarly, we check the integrability of $\gamma_r := |\mathcal{D}_t Z_r^0|$. Since $(\mathcal{D}_t Z_r^0)_{t\leq r\leq T} \in \mathcal{M}^p$,

$$\mathbb{E}\Big(\int_0^T \gamma_r dr\Big)^p \leq_c \mathbb{E}\Big(\int_0^T |\mathcal{D}_t Z_r^0|^2 dr\Big)^{\frac{p}{2}} < +\infty.$$

This gives

$$\mathbb{E}\Big(\int_0^T |\eta_r^{\varepsilon,c}|dr\Big)^p \to 0. \tag{5.37}$$

From (5.34), (5.35), (5.36) and (5.37), the proof is complete. □

Step 4: Proof of Theorem 5.8.1

From Lemma 5.8.2 and Lemma 5.8.4, we know that $\lim_{\varepsilon\to 0}\mathbb{E}|Z_t^\varepsilon - D_t Y_t^0|^p = 0$. Besides, from Lemma 5.8.1, there is a subsequence $(\varepsilon_n)_{n\geq 1}$ decreasing to 0 such that $\lim_{n\to+\infty} Z_t^{\varepsilon_n} = Z_t^0$, $(\mathrm{d}\mathbb{P}\otimes \mathrm{d}t) - a.s.$. Thus, we conclude that, $(\mathrm{d}\mathbb{P}\otimes \mathrm{d}t) - a.s.$,

$$Z_t^0 = D_t Y_t^0 = \mathbb{E}^{\mathcal{F}_t}\int_t^T \Gamma_r^t a_r^0 D_t X_r dr, \tag{5.38}$$

taking advantage of the explicit representation of $D_t Y_t^0$ in (5.30).

Then, from (5.32) and setting $U_t := \nabla Y_t^0 [\nabla X_t]^{-1}$, we have proved $Z_t^0 = U_t \sigma(t, X_t)$. It remains to show that U satisfies the BSDE (5.25). It is clear that

$$-dU_t = (-d\nabla Y_t^0)[\nabla X_t]^{-1} + \nabla Y_t^0(-d[\nabla X_t]^{-1}) - d < \nabla Y_t^0, [\nabla X_t]^{-1} >_t .$$

Besides, it is known (see e.g. [GM05b]) that

$$-d[\nabla X_t]^{-1} = [\nabla X_t]^{-1}\big\{\big(b_t^{'} - \sum_{j=1}^{q}(\sigma_{j,t}^{'})^2\big)dt + \sum_{j=1}^{q}\sigma_{j,t}^{'}dW_t^j\big\}.$$

Then, from the expression of $d\nabla Y_t^0$ in (5.31), it follows that

$$\begin{aligned}
-dU_t &= \big(a_t^0 + b_t^0 \nabla Y_t^0 [\nabla X_t]^{-1} + \sum_{j=1}^{q} c_{j,t}^0 \nabla Z_{j,t}^0 [\nabla X_t]^{-1}\big)dt - \sum_{j=1}^{q} \nabla Z_{j,t}^0 [\nabla X_t]^{-1} dW_t^j \\
&\quad + \nabla Y_t^0 [\nabla X_t]^{-1}\big\{\big(b_t^{'} - \sum_{j=1}^{q}(\sigma_{j,t}^{'})^2\big)dt + \sum_{j=1}^{q}\sigma_{j,t}^{'}dW_t^j\big\} \\
&\quad + \sum_{j=1}^{q} \nabla Z_{j,t}^0 [\nabla X_t]^{-1} \sigma_{j,t}^{'} dt \\
&= a_t^0 dt + U_t\big(b_t^0 I_d + b_t^{'} + \sum_{j=1}^{q} c_{j,t}^0 \sigma_{j,t}^{'}\big)dt \\
&\quad + \sum_{j=1}^{q}\big(\nabla Z_{j,t}^0 [\nabla X_t]^{-1} - U_t \sigma_{j,t}^{'}\big)\big(c_{j,t}^0 I_d + \sigma_{j,t}^{'}\big)dt \\
&\quad - \sum_{j=1}^{q}\big(\nabla Z_{j,t}^0 [\nabla X_t]^{-1} - U_t \sigma_{j,t}^{'}\big)dW_t^j.
\end{aligned}$$

By setting $V_t^j := \nabla Z_{j,t}^0 [\nabla X_t]^{-1} - U_t \sigma_{j,t}^{'}$, we obtain

$$-dU_t = \big\{a_t^0 + U_t\big(b_t^0 I_d + b_t^{'} + \sum_{j=1}^{q} c_{j,t}^0 \sigma_{j,t}^{'}\big) + \sum_{j=1}^{q} V_t^j \big(c_{j,t}^0 I_d + \sigma_{j,t}^{'}\big)\big\}dt - \sum_{j=1}^{q} V_t^j dW_t^j.$$

□

5.9 $\mathbf{L}_2$-regularity of Z_t when the terminal condition $g \in \mathbf{L}_{2,\alpha}$, but is not necessarily bounded

5.9.1 The main results

In this section, we aim to establish an $\mathbf{L}_2$-regularity of the process $(Z_t)_t$, more precisely, to have a good rate of convergence of $\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left|Z_s - \bar{Z}_{t_k}\right|^2 ds$, where

$$\bar{Z}_{t_k} := \frac{1}{t_{k+1} - t_k} \mathbb{E}^{\mathcal{F}_{t_k}} \int_{t_k}^{t_{k+1}} Z_u du.$$

In Zhang [Zha04], it was shown that, for BSDEs with Lipschitz continuous terminal conditions, this rate is linear with respect to the time step $|\pi|$.

Here, we extend this result to non-Lipschitz terminal functions g: we show that, if we suppose that $g \in \bigcup_{\alpha\in(0,1]} \mathbf{L}_{2,\alpha}$, we can obtain the same rate $|\pi|^\alpha$ for the equidistant time net $\pi^{(1)}$ or the rate $|\pi|$ for an appropriate choice of the time net.
In fact, we show that this $\mathbf{L}_2$-regularity of $(Z_t)_{0\le t\le T}$ can be deduced, under the assumption above on g, from that of the process $(z_t)_{0\le t\le T}$ (Theorem 5.9.1).

This is an interesting fact, since the study of the martingale integrand of the initial nonlinear BSDE can be reduced to that of the martingale integrand of the linear simpler BSDE with a null generator. We can then derive the desired rate (Theorem 5.9.2). We state below these two main results, that hold even if g is not bounded. Their proofs are postponed to the next paragraph.

Theorem 5.9.1. *Assume* $(\boldsymbol{A}_{b,\sigma})$, $(\boldsymbol{A}_f)$ *and that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0,1]$. *Then, there is a positive constant* C *such that, for any time net* $\pi = \{t_k : k = 0...N\}$

$$\mathcal{E}(Z,\pi) = \sum_{k=0}^{N-1} \mathbb{E}\int_{t_k}^{t_{k+1}} \left|Z_s - \bar{Z}_{t_k}\right|^2 ds \le C\mathcal{E}(z,\pi) + C\big(K^\alpha(g)T^\alpha + T^2\big)|\pi|.$$

Theorem 5.9.2. *Assume* $(\boldsymbol{A}_{b,\sigma})$, $(\boldsymbol{A}_f)$ *and that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0,1]$. *Then, there is a positive constant* C *(which does not depend on* N*) such that*

a) for the choice of the equidistant time net $\pi^{(1)}$,

$$\mathcal{E}(Z,\pi^{(1)}) \le C\frac{K^\alpha(g)T^\alpha + T^3N^{-1+\alpha}}{N^\alpha};$$

b) for the choice of $\pi^{(\beta)}$, *with* β *as in Corollary 5.7.2,*

$$\mathcal{E}(Z,\pi^{(\beta)}) \le C\frac{K^\alpha(g)T^\alpha + T^3}{N}.$$

Since all the bounds depend on the regularity of f only through $\|\nabla_x f\|_\infty$, $\|\nabla_y f\|_\infty$ and $\|\nabla_z f\|_\infty$, we can state the following theorem.

Theorem 5.9.3. *Assume* $(\boldsymbol{A}_{b,\sigma})$ *and that* $g \in \mathbf{L}_{2,\alpha}$, *for some* $\alpha \in (0,1]$. *Then, Theorem 5.9.2 still holds when the generator* f *is uniformly Lipschitz continuous in* x, y *and* z *(but not necessarily continuously differentiable).*

The above theorem is derived from Theorem 5.9.2 using standard stability results, we omit details.

5.9.2 Proof of Theorem 5.9.1

In order to be able to use the results of the previous sections, we begin by assuming that g is bounded.

Step 1: Proof when g is bounded

Suppose for the moment that g is bounded (and belongs to $\mathbf{L}_{2,\alpha}$).
Recalling that $Z_t^0 = Z_t - z_t = U_t\sigma(t, X_t)$, we are going to exploit the BSDE (5.25) satisfied by U in order to bound $\mathbb{E}\int_{t_k}^{t_{k+1}} |Z_s^0 - \bar{Z}_{t_k}^0|^2 ds$.
From the BSDE (5.25) satisfied by $(U_t, V_t)_t$ (Theorem 5.8.1), and using Lemma 5.A.1, one obtains the following estimate:

$$\mathbb{E}\big[\sup_{r\in[0,T]} |U_r|^2 + \int_0^T |V_r|^2 dr\big] \leq_c \mathbb{E}\big(\int_0^T |a_r^0| dr\big)^2.$$

Since $|a_r^0| \leq_c 1 + |\nabla_x u(r, X_r)| + |D^2 u(r, X_r)|$, we use Corollary 5.7.1 to obtain

$$\mathbb{E}\big[\sup_{r\in[0,T]} |U_r|^2 + \int_0^T |V_r|^2 dr\big] \leq_c T^\alpha K^\alpha(g) + T^2. \tag{5.39}$$

Let $0 \leq t \leq s \leq T$. Always from BSDE (5.25), and using the estimate (5.39), one has

$$\begin{aligned}\mathbb{E}|U_s - U_t|^2 &\leq_c \mathbb{E}\big(\int_t^s |a_r^0| dr\big)^2 + (s-t)^2 \mathbb{E}\sup_r |U_r|^2 + \mathbb{E}\int_t^s |V_r|^2 dr \\ &\leq_c \mathbb{E}\big(\int_t^s |a_r^0| dr\big)^2 + \big[K^\alpha(g)T^\alpha + T^2\big](s-t)^2 + \mathbb{E}\int_t^s |V_r|^2 dr. \end{aligned} \tag{5.40}$$

We go back to the regularity of Z. As we did in the proof of Theorem 5.7.1, we write, using Theorem 5.8.1 to have $Z_s^0 = U_s\sigma(s, X_s)$, $(\mathrm{d}\mathbb{P} \otimes \mathrm{d}t) - a.s.$,

$$\mathbb{E}\int_{t_k}^{t_{k+1}} |Z_s^0 - \bar{Z}_{t_k}^0|^2 ds \leq \mathbb{E}\int_{t_k}^{t_{k+1}} |U_s\sigma(s, X_s) - U_{t_k}\sigma(t_k, X_{t_k})|^2 ds.$$

Now, for $s \in [t_k, t_{k+1}]$,

$$\begin{aligned}\mathbb{E}|U_s\sigma(s, X_s) - U_{t_k}\sigma(t_k, X_{t_k})|^2 &= \mathbb{E}|\big(U_s - U_{t_k}\big)\sigma(s, X_s) - U_{t_k}\big(\sigma(t_k, X_{t_k}) - \sigma(s, X_s)\big)|^2 \\ &\leq_c \mathbb{E}|U_s - U_{t_k}|^2 + |\pi|\mathbb{E}|U_{t_k}|^2.\end{aligned}$$

Then, using (5.39) and (5.40),

$$\mathbb{E}|U_s\sigma(s, X_s) - U_{t_k}\sigma(t_k, X_{t_k})|^2 \leq_c \mathbb{E}\big(\int_{t_k}^{t_{k+1}} |a_r^0| dr\big)^2 + \mathbb{E}\int_{t_k}^{t_{k+1}} |V_r|^2 dr + \big[T^\alpha K^\alpha(g) + T^2\big]|\pi|.$$

Therefore,

$$\begin{aligned}
&\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}\mathbb{E}\big|Z_s^0-\bar{Z}_{t_k}^0\big|^2ds\\
&\leq_c|\pi|\sum_{k=0}^{N-1}\big\{\mathbb{E}\big(\int_{t_k}^{t_{k+1}}|a_r^0|dr\big)^2+\mathbb{E}\int_{t_k}^{t_{k+1}}|V_r|^2dr\big\}+\big[T^\alpha K^\alpha(g)+T^2\big]|\pi|\\
&\leq_c|\pi|\big\{\mathbb{E}\big(\int_0^T|a_r^0|dr\big)^2+\mathbb{E}\int_0^T|V_r|^2dr+(T^\alpha K^\alpha(g)+T^2)\big\}\\
&\leq_c|\pi|\big(T^\alpha K^\alpha(g)+T^2\big).
\end{aligned}$$

Now, since clearly $\big|Z_s-\bar{Z}_{t_k}\big|^2=\big|(z_s+Z_s^0)-(\bar{z}_{t_k}+\bar{Z}_{t_k}^0)\big|^2\leq 2\big|z_s-\bar{z}_{t_k}\big|^2+2\big|Z_s^0-\bar{Z}_{t_k}^0\big|^2$, we conclude that

$$\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\big|Z_s-\bar{Z}_{t_k}\big|^2ds\leq_c\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}|z_s-\bar{z}_{t_k}|^2ds+|\pi|\big(T^\alpha K^\alpha(g)+T^2\big). \tag{5.41}$$

□

Now, we suppose that g is not necessarily bounded (and g belongs to $\mathbf{L}_{2,\alpha}$). We use the following bounded approximation of g. For $M>0$ and $y\in\mathbb{R}$, we set

$$\phi_M(y):=-M\vee y\wedge M, \tag{5.42}$$

and

$$g_M:=\phi_M\circ g. \tag{5.43}$$

It is clear that, when $M\to+\infty$, $g_M(x)\to g(x)$ for all $x\in\mathbb{R}$ such that $|g(x)|<+\infty$, and $g_M(X_T)\to g(X_T)$ in $\mathbf{L}_2$.
We denote by (Y^M,Z^M) (resp. (y^M,z^M)) the solution to BSDE (5.1) (resp. BSDE (5.12)) with $g_M(X_T)$ as terminal condition instead of $g(X_T)$.

Step 2: Some stability results when M goes to $+\infty$

Lemma 5.9.1. *Assume* $(\boldsymbol{A}_{b,\sigma})$ *and that* $g\in\mathbf{L}_{2,\alpha}$, *for some* $\alpha\in(0,1]$. *Then,* $g_M\in\mathbf{L}_{2,\alpha}$ *and*

$$K^\alpha(g_M)\leq K^\alpha(g).$$

Proof. Recall that $y_t=\mathbb{E}^{\mathcal{F}_t}g(X_T)$, and set $y_t^M:=\mathbb{E}^{\mathcal{F}_t}g_M(X_T)=\mathbb{E}^{\mathcal{F}_t}\phi_M(y_T)$, where ϕ_M is the function already defined by (5.42). Note that ϕ_M is Lipschitz-continuous, with a Lipschitz constant equal to 1. One has

$$\begin{aligned}
V_{t,T}(g_M)&=\mathbb{E}\big|y_T^M-y_t^M\big|^2=\mathbb{E}\big|\phi_M(y_T)-\mathbb{E}^{\mathcal{F}_t}\phi_M(y_T)\big|^2\\
&\leq\mathbb{E}\big|\phi_M(y_T)-\phi_M(y_t)\big|^2\leq\mathbb{E}|y_T-y_t|^2=V_{t,T}(g),
\end{aligned}$$

where we used a projection property on $\mathbf{L}_2(\mathcal{F}_t)$ for the first inequality.
In addition, clearly $|g_M(X_T)|\leq|g(X_T)|$, which readily finishes the proof. □

Lemma 5.9.2.

$$\lim_{M\to+\infty}\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|z_s^M-\bar{z}_{t_k}^M\right|^2ds=\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|z_s-\bar{z}_{t_k}\right|^2ds; \tag{5.44}$$

$$\lim_{M\to+\infty}\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|Z_s^M-\bar{Z}_{t_k}^M\right|^2ds=\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|Z_s-\bar{Z}_{t_k}\right|^2ds. \tag{5.45}$$

Proof. We only prove (5.44), since for (5.45), the arguments are the same. Write

$$\begin{aligned}
&\Big|\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|z_s^M-\bar{z}_{t_k}^M\right|^2ds-\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|z_s-\bar{z}_{t_k}\right|^2ds\Big|\\
&=\Big|\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\big(\left|z_s-\bar{z}_{t_k}\right|-\left|z_s^M-\bar{z}_{t_k}^M\right|\big)\big(\left|z_s-\bar{z}_{t_k}\right|+\left|z_s^M-\bar{z}_{t_k}^M\right|\big)ds\Big|\\
&\le\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|(z_s-z_s^M)-(\bar{z}_{t_k}-\bar{z}_{t_k}^M)\right|\big(\left|z_s-\bar{z}_{t_k}\right|+\left|z_s^M-\bar{z}_{t_k}^M\right|\big)ds\\
&\le\sum_{k=0}^{N-1}\sqrt{\mathbb{E}\int_{t_k}^{t_{k+1}}\left|(z_s-z_s^M)-(\bar{z}_{t_k}-\bar{z}_{t_k}^M)\right|^2ds}\sqrt{\mathbb{E}\int_{t_k}^{t_{k+1}}\big(\left|z_s-\bar{z}_{t_k}\right|+\left|z_s^M-\bar{z}_{t_k}^M\right|\big)^2ds}.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}\int_{t_k}^{t_{k+1}}\left|(z_s-z_s^M)-(\bar{z}_{t_k}-\bar{z}_{t_k}^M)\right|^2ds&\le 2\mathbb{E}\int_0^T\left|z_s-z_s^M\right|^2ds+2\mathbb{E}\int_0^T\left|\bar{z}_{t_k}-\bar{z}_{t_k}^M\right|^2ds\\
&\le 4\mathbb{E}\int_0^T\left|z_s-z_s^M\right|^2ds,
\end{aligned}$$

where we used a projection argument. By classical stability results for BSDEs, the term above tends to 0 when $M\to+\infty$. Besides,

$$\begin{aligned}
\mathbb{E}\int_{t_k}^{t_{k+1}}\big(\left|z_s-\bar{z}_{t_k}\right|+\left|z_s^M-\bar{z}_{t_k}^M\right|\big)^2ds&\le 4\mathbb{E}\int_0^T\big(\left|z_s\right|^2+\left|\bar{z}_{t_k}\right|^2+\left|z_s^M\right|^2+\left|z_{t_k}^M\right|^2\big)ds\\
&\le C\mathbb{E}|g(X_T)|^2,
\end{aligned}$$

where C does not depend on M (still using the classical a priori estimate for BSDEs). Thus, we have proved (5.44). □

Step 3: Proof when g is not necessarily bounded

Applying (5.41) in Step 1 with g_M, one has

$$\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|Z_s^M-\bar{Z}_{t_k}^M\right|^2ds\le_c\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\left|z_s^M-\bar{z}_{t_k}^M\right|^2ds+|\pi|\big(T^\alpha K^\alpha(g_M)+T^2\big),$$

and, since $K^\alpha(g_M) \leq K^\alpha(g)$ (Lemma 5.9.1),

$$\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left|Z_s^M - \bar{Z}_{t_k}^M\right|^2 ds \leq_c \sum_{k=0}^{N-1} \mathbb{E} \int_{t_k}^{t_{k+1}} \left|z_s^M - \bar{z}_{t_k}^M\right|^2 ds + |\pi|\big(T^\alpha K^\alpha(g) + T^2\big). \tag{5.46}$$

Passing to the limit when $M \to +\infty$ and using Lemma 5.9.2, we prove Theorem 5.9.1. □

5.9.3 Proof of Theorem 5.9.2

a) Equidistant time net $\pi^{(1)}$. As a direct consequence of Theorem 5.9.1 and Theorem 5.7.1, one has

$$\begin{aligned}\sum_{k=0}^{N-1} \mathbb{E} \int_{t_k^{(N,1)}}^{t_{k+1}^{(N,1)}} \left|Z_s^M - \bar{Z}_{t_k^{(N,1)}}^M\right|^2 ds &\leq_c K^\alpha(g_M)\big(\frac{T^\alpha}{N^\alpha} + \frac{T^{1+\alpha}}{N}\big) + \frac{T^3}{N} \\ &\leq_c K^\alpha(g)\frac{T^\alpha}{N^\alpha} + \frac{T^3}{N}\end{aligned}$$

using Lemma 5.9.1. Passing to the limit when $M \to +\infty$ (Lemma 5.9.2), we prove estimate (a) of Theorem 5.9.2.
b) Time net $\pi^{(\beta)}$. The proof is the same, using Theorem 5.9.1, Corollary 5.7.2 and Lemma 5.9.1. □

5.A Appendix

5.A.1 Malliavin weights expressions

These expressions are given by Gobet and Munos [GM05b].

(i) $\nabla_x u(t, X_t) = \mathbb{E}^{\mathcal{F}_t}\left[g(X_T)H_{t,T}^{(1)}\right]$,where

$$H_{t,T}^{(1)} = \frac{1}{T-t}\left[\int_t^T [\sigma^{-1}(s, X_s)\nabla X_s[\nabla X_t]^{-1}]^* dW_s\right]^*. \tag{5.47}$$

(ii) $\nabla_x(\partial_{x_k} u)(t, X_t) = \mathbb{E}^{\mathcal{F}_t}\left[g(X_T)H_{t,T}^{(2),k}\right]$, where

$$H_{t,T}^{(2),k} = \left[H_{\frac{T+t}{2},T}^{(1)}\right] e^k.\left[H_{t,\frac{T+t}{2}}^{(1)}\right]^* + e^k.\nabla_{x_0}\left\{\left[H_{t,\frac{T+t}{2}}^{(1)}\right]^*\right\}[\nabla X_t]^{-1}, \tag{5.48}$$

and $e^k = (0...010...0)^*$, 1 being in the kth coordinate.

5.A.2 An estimate for linear BSDEs

The following lemma gives an a priori estimate for linear BSDEs, and is a direct consequence of Proposition 3.2 in [BDH$^+$03] (applied with $f(\omega, t, y, z)$ of the form $a_t + b_t y + c_t z$).

Lemma 5.A.1. *Consider the linear BSDE*

$$U_t = \xi + \int_t^T \big(a_r + U_r b_r + \sum_{j=1}^q V_{j,r} c_{j,r}\big) dr - \sum_{j=1}^q \int_t^T V_{j,r} dW_r^j \tag{5.49}$$

where $\xi \in \mathbb{R}^{1\times d}, a_r \in \mathbb{R}^{1\times d}, b_r \in \mathbb{R}^{d\times d}, c_{j,r} \in \mathbb{R}^{d\times d}, U_r \in \mathbb{R}^{1\times d}, V_{j,r} \in \mathbb{R}^{1\times d}, W_r \in \mathbb{R}^q$, *for some progressively measurable coefficients* $(a_r)_r$, $(b_r)_r$, $(c_r)_r$ *and a* $\mathcal{F}_T$*-measurable terminal condition* ξ.
If $|b_r|$, $|c_{j,r}|$ *are uniformly bounded, and if* $\mathbb{E}|\xi|^p + \mathbb{E}\big(\int_0^T |a_r| dr\big)^p < +\infty$, *then there exists an unique solution* (U, V) *in* $\mathcal{S}^p \times \mathcal{M}^p$ *to BSDE (5.49), and the following estimate holds:*

$$\mathbb{E} \sup_{r\in[0,T]} |U_r|^p + \mathbb{E}\Big(\int_0^T |V_r|^2 dr\Big)^{\frac{p}{2}} \leq C\big\{\mathbb{E}|\xi|^p + \mathbb{E}\Big(\int_0^T |a_r| dr\Big)^p\big\}. \tag{5.50}$$

Part II

Efficient evaluation of CDOs with stochastic recoveries

Chapter 6

A brief account on CDO pricing

6.1 Collateralized Debt Obligations (CDOs)

A *CDO tranche*, based on an underlying multi-name porfolio of assets (corporate bonds, bank loans, credit swaps, ...), is a bilateral contract consisting of a protection against the default losses in the portfolio, within an interval $[K_1, K_2]$, over some time period $[0, T]$. The (fixed) values K_1 and K_2 are respectively called the *lower attachment point* (or *attachment point*) and the *upper attachment point* (or *detachment point*) of the tranche. One can roughly distinguish the *Equity* tranche (that absorbs the first proportion, say 7%, of the losses in the underlying pool, and is then the first to be exposed), the *Mezzanine* tranche (that absorbs the next proportion, say 18%=25%-7%) and the *Senior* tranches (that absorb the last 75%=100%-25%, and are then the less exposed to risk).
When a name q of the underlying pool defaults, only a part of its notional ν_q can be recovered, and we denote by R_q the corresponding recovery rate (which may be non deterministic). Thus, the loss caused by the default of the name q, called the *loss given default*, is equal to $l_q := \nu_q(1 - R_q)$, and the total portfolio loss over a time period $[0, t]$ is equal to

$$L_t := \sum_q l_q \mathbb{1}_{\tau_q \leq t} = \sum_q \nu_q (1 - R_q) \mathbb{1}_{\tau_q \leq t},$$

where τ_q denotes the default time of the name q.
At each time $T_i \in [0, T]$ of the ith coupon paid by the protection buyer, the protection seller covers (up to the coupon rate) the accumulated loss within the tranche $[K_1, K_2]$, denoted by $L_{T_i}^{[K_1,K_2]}$: it is equal to 0 if $L_{T_i} < K_1$, to $L_{T_i} - K_1$ if $K_1 \leq L_{T_i} \leq K_2$, and to $K_2 - K_1$ if $L_{T_i} > K_2$. Thus,

$$L_{T_i}^{[K_1,K_2]} = (L_{T_i} - K_1)^+ - (L_{T_i} - K_2)^+,$$

and the valuation of the CDO tranche amounts to evaluating the *expected tranche notionals* $\mathbb{E}((L_T - K)^+)$, for different values of T (namely, the CDO maturities) and of K (namely, the attachment and detachment points). For more details, see e.g. [Das05].

6.2 The main loss modeling approaches

There are two main approaches that have been developed for the valuation of credit derivatives: the so-called *bottom-up* and *top-down* approaches (see [Gie08]).
In the bottom-up models, one starts from the description of each individual name dynamics (default time, default intensity, loss given default,...), as well as the possible dependencies between them; then, the accumulated loss dynamics implicitely follows. On the contrary, within the top-down approach, one first explicitely models the portfolio loss process (without reference to the underlying constituents), and then tries to find the dynamics of the individual names. This is usually done with the help of the *random thinning* technique, introduced (in the credit derivatives framework) by Giesecke and Goldberg [GG09]. Note that, within both models, some *auxiliary factor processes* may be introduced in order to take into account some additional information.
Each of the two approaches has its advantages and drawbacks, depending on the purpose for which it is developed. In practice, the bottom-up models are the best suited for hedging issues (since they enable one to compute single name sensitivities, see [HT09]), whereas the top-down models are the best suited for fast valuation and calibration (since the accumulated loss is the only underlying process, possibly in addititon to few factor processes). We refer to [Gie08] and [BCJ08] for more detailed comparison and critic of both approaches.

6.3 Numerical valuation of CDOs

Depending on the adopted model, there exist a variety of numerical methods for pricing CDO tranches. Hereafter, we give a brief (and non exhaustive) enumeration of such methods.

6.3.1 In bottom-up models

Within bottom-up models, it is often assumed that the time defaults (τ_q) are conditionally independent given a common known market factor U, and that the losses given default (l_q) are mutually independent (and independent of the default times). Under such assumptions, a FFT (Fast Fourier Transform) algorithm has been used by Gregory and Laurent [GL05], relying on the inversion of the characteristic function of the loss L_t.
A recursive algorithm has been suggested by Andersen *et al.* [ASB03] and [HW04]. At the first step, it computes the loss distribution of a single-name portfolio, and then, it iteratively computes the loss distribution related to $q+1$ names starting from the one related to q names.
Recently, an approximation method, based on Stein's method, has been proposed by El Karoui *et al.* in [KJK09] and [KJ09]. It consists in approximating the loss conditional distribution by a known distribution, namely the Gaussian or the Poisson distribution.

In addition, within a dependent-obligor model (namely, a multifactor Normal Copula model), but under the assumption that the recoveries are deterministic, Glasserman and Suchintabandid [GS07] have suggested a method based on a power series expansion of the CDO tranche price into the sum of independent-obligor tranche prices that are easy to compute.
A more detailed account on the previous methods and some others can be found in [Cou08](Section 1.1.2).

6.3.2 In top-down models

In the contexts where the loss process, together with the *default intensity*, are specified (possibly in addition to some factor processes, the whole defining a Markovian setting), either deterministic or Monte-Carlo methods have been developed in order to compute the CDO tranche price.
The deterministic methods rely on the discretization of the partial integro-differential equation (PIDE, for short) satisfied by the price function $u(t,l) = \mathbb{E}[(L_T - K)^+ | L_t = l]$ (this PIDE arises from the Itô formula). Finite-difference schemes for the derivatives terms, and quadrature approximations for the integral term(s) can then be performed[1]. Such methods have been used for pricing derivatives whose driving process has jumps, possibly added to a diffusion component. We refer for instance to [Duf] , [Duf06], [Vol05] and [SS08], among others.
In the particular case of *homogeneous portfolio*, where the losses given defaut (l_q) are assumed to be constant and all equal to c (so that the loss process is given by $L_t = cN_t$, where $N_t := \sum_q \mathbb{1}_{\tau_q \leq t}$ is the *default process*), Cont and Savescu [CS08] have derived the forward Dupire-type equation satisfied by the price function, that can be efficiently solved using high-order time stepping schemes. Indeed, benefitting from the discrete nature of the losses, this equation is just a bi-diagonal system of ordinary differential equations (ODEs, for short). This approach has also been used for the calibration of the default portfolio intensities (see [CS08] and [CM08]).
Always in the case of constant and equal losses to default, one can compute the CDO tranche price by inverting the Fourier transform of the default process N (expressed in terms of a Laplace transform of its *compensator* process): see Giesecke [Gie07].
In the general multifactor Markovian framework, the deterministic methods (relying on PDEs or PIDEs) are inadequate because of the curse of dimensionality (as soon as the model dimension is greater than 3 or 4 !). Then, Monte-Carlo methods, accompanied with variance reduction techniques (in order to remedy the slowness of the crude Monte-Carlo method, see section 7.1), seem to be imperative.
The main Monte-Carlo-type method proposed for the purpose of computing CDO prices has been the *interacting particle systems* (IPS for short) approach, developed by Carmona and Crepey [CC09] and Carmona *et al.* [CFV09]. It consists in a more sophisticated variant of the *importance sampling* (IS for short) approach (see section 7.1), that is essentially suited for the computation of rare events (here, defaults)

[1] Note that, in our context, there is no partial derivatives w.r.t. l, since $(L_t)_t$ is a jump process, which is a simplification for the PIDE.

probabilities. The IPS based method has been analyzed in full generality by Del Moral and Garnier [MG05]. However, the IPS approach becomes inappropiate in a strong contagion situation, where the defaults probabilities are no more "small".

6.3.3 Our contribution

Within a top-down Markovian approach, and without assuming homogeneity on the underlying portfolio (we will only suppose that the losses to default (l_q) are i.i.d random variables), we propose a sequential Monte-Carlo method in order to efficiently compute the CDO tranche price. We apply the *sequential control variates* algorithm, introduced by Gobet and Maire [GM05a] for the computation of solutions of second order linear PDEs that have a Feynman-Kac-type representation (see chapter 7). The case of stochastic, non identically distributed losses to default, is left to a future research.

Chapter 7

Sequential control variates for linear Markov processes

7.1 Introduction

Our aim is to evaluate efficiently the price of CDO tranches, through the computation of quantities like $\mathbb{E}[(L_T - K)_+]$, where $(L_t)_t$ is the loss process. In such a situation, and in many others (for example, in finance, in the pricing of European options, or in physics, in diffusion equations), one has to numerically compute expectations of the form $\mathbb{E}(\Psi(X_s : s \geq t)|X_t = x)$, where $(X_t)_t$ is a Markov process, that generally arise as the Feynman-Kac representations of the solutions of linear PDEs.
The well-known Monte-Carlo method consists in approximating the expectation of a random variable Y by its empirical mean $\bar{Y}_M := \frac{1}{M}\sum_{i=1}^{M} Y_i$, where $Y_1, ..., Y_M$ are M independent simulated copies of Y. This method is especially convenient for high-dimensional problems, but the $\mathbf{L}_2$-norm of the related convergence error is equal to $\frac{\sigma}{\sqrt{M}}$, where σ is the standard deviation of Y, which makes the convergence slow.
Two major variance-reduction techniques have been used in order to minimize the convergence error. The first is the so-called *control variates* approach, and uses simulations of $Y - Z$ instead of Y, where Z is a random variable (called the control variate) such that $\mathbb{E}(Z) = 0$. It is more efficient if Z is such that $\mathbb{V}ar(Y - Z) << \mathbb{V}ar(Y)$. The second is the so-called *importance sampling* technique, and is based on a change of probability measure: one simulates $Y\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$ under a new probability measure $\mathbb{Q}$ instead of simulating Y under the initial probability $\mathbb{P}$ (since $\mathbb{E}_{\mathbb{Q}}(Y\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}) = \mathbb{E}_{\mathbb{P}}(Y)$). The efficiency is achieved if $\mathbb{V}ar_{\mathbb{Q}}(Y\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}) << \mathbb{V}ar_{\mathbb{P}}(Y)$. We refer to Halton [Hal70], Glasserman [Gla04] and to references therein. Moreover, sequential Monte-Carlo algorithms, using iteratively the two previous techniques, have been developed (see, for instance, [Hal62] and [Hal70]).
Gobet and Maire have introduced in [GM04] a sequential Monte-Carlo algorithm to solve the Poisson equation with Dirichlet boundary conditions over square domains. Then, in [GM05a], they have extended their approach to general Markov processes (including diffusion and jump processes) related to linear elliptic and parabolic sec-

ond order PDEs, with linear boundary conditions. Recently, Gobet and Labart [GL09] have applied this technique, together with Picard's type iterations, to the numerical approximation of (nonlinear) BSDEs.
Since we will rely on the sequential control variate approach in order to compute the price of CDO tranches, let us describe this approach in detail (we use the same notations as in [GM05a]).

7.2 Framework

Let $X(x) = (X_t(x))_{t\geq 0}$ be a Markov process with initial value x belonging to a domain $D \subset \mathbb{R}^d$, and let f and g be two continuous functions, respectively defined on $\bar{D}$ and ∂D. Consider the following quantity to approximate:

$$u(x) := \mathbb{E}(\Psi(f, g, X(x))), \tag{7.1}$$

where Ψ is a functional that is linear w.r.t. the data (f, g).
Suppose moreover that u is a classic solution of the following PDE

$$\begin{cases} \mathcal{A}u &= f \quad \text{in } D, \\ \mathcal{B}u &= g \quad \text{on } \partial D, \end{cases} \tag{7.2}$$

where $\mathcal{A}$ and $\mathcal{B}$ are second order linear operators. This is the elliptic problem, but one can as well consider parabolic problems, where the operator $\mathcal{A}$ has to be replaced by $\partial_t + \mathcal{A}$, with a time-space process $(t, X_t)_t$, in the domain $[0, T) \times D$ instead of D. For instance, if X is a Brownian SDE: $X_t = x + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s$, and $u(t, x) = \mathbb{E}\left(g(X_T) - \int_t^T f(s, X_s)\mathrm{d}s | X_t = x\right)$, then $\partial_t u + \mathcal{L}u = f$ and $u(T, .) = g(.)$, where $\mathcal{L}$ is the infinitesimal generator of the diffusion X ($\mathcal{L}u(t, x) := \sum_{i=1}^d b_i(t, x)\partial_{x_i} u(t, x) + \frac{1}{2}\sum_{i,j=1}^d [\sigma\sigma^*]_{i,j}(t, x)\partial^2_{x_i,x_j} u(t, x)$).

7.3 Sequential control variates

The idea of the sequential control variates (SCV, for short) algorithm is to approximate the solution u by a sequence $(u_n)_n$ such that, at the nth step,

$$u_n(x_i) = u_{n-1}(x_i) + \bar{y}_n(x_i),$$

where $(x_i)_{1\leq i\leq N}$ are some points of $\bar{D}$ (interpolation knots, for example), and $\bar{y}_n(x_i)$ is a Monte-Carlo evaluation of the correction term $y_n(x_i) := u(x_i) - u_{n-1}(x_i)$, represented as an expectation: indeed, from (7.1) and (7.2), and from the linearity of the functional Ψ, one has

$$y_n(x) = \mathbb{E}[\Psi(f - \mathcal{A}u_{n-1}, g - \mathcal{B}u_{n-1}, X(x))].$$

A global approximation based on the values on the points $(x_i)_i$ is then obtained by

$$u_n = u_{n-1} + \mathcal{P}\bar{y}_n,$$

where $\mathcal{P}$ is a linear approximation operator (an interpolation operator, for example) based on the points $(x_i)_i$, i.e., for any function φ and for some basis functions $(\mathcal{C}_i)_i$ (assumed to be in $\mathcal{C}^2(\bar{D})$),

$$\mathcal{P}\varphi(.) = \sum_{i=1}^{N} \varphi(x_i)\mathcal{C}_i(.).$$

Here is the algorithm. $X^{\Delta}(x)$ denotes a discretization procedure for the process $X(x)$, with Δ being the discretization parameter (for example, the time step for the Euler procedure).

Algorithm 7.3.1 (SCV)**.**

1. *Initialization:* $u_0 := 0$.

2. *At iteration n: for each x_i, simulate M independent paths $(X^{\Delta,n,m}(x_i))_{1\leq m\leq M}$ of $X^{\Delta}(x_i)$ (independently of anything else also), and set*

$$\bar{y}_n(x_i) := \frac{1}{M}\sum_{m=1}^{M} \Psi(f - \mathcal{A}u_{n-1}, g - \mathcal{B}u_{n-1}, X^{\Delta,n,m}(x_i)).$$

3. *Always at iteration n: build the global solution using the linear approximation of the correction term:*

$$u_n := u_{n-1} + \mathcal{P}\bar{y}_n = u_{n-1} + \sum_{i=1}^{N} \bar{y}_n(x_i)\mathcal{C}_i.$$

Remark 7.3.1. *The operator $\mathcal{P}$ is assumed to have the stability property:*

$$\mathcal{P}(\mathcal{P}\varphi) = \mathcal{P}\varphi.$$

This property is satisfied by the interpolation and the projection operators.
Under this assumption, it is easy to see that, for all n, $u_n = \mathcal{P}u_n = \sum_{i=1}^{N} u_n(x_i)\mathcal{C}_i$.
The numerical computation of $\mathcal{A}u_{n-1}$ in the step 2 of Algorithm 7.3.1 is then performed through that of $(\mathcal{A}\mathcal{C}_i)_i$: indeed, by linearity of the operator $\mathcal{A}$, one has $\mathcal{A}u_{n-1} = \sum_{i=1}^{N} u_{n-1}(x_i)\mathcal{A}\mathcal{C}_i$. The computation of $\mathcal{B}u_{n-1}$ is similar.

In their paper, Gobet and Maire have studied the bias $m_n := \sup_{1\leq i\leq N} |\mathbb{E}(u_n(x_i) - u(x_i))|$ and the variance $v_n := \sup_{1\leq i\leq N} |\mathbb{V}ar(u_n(x_i))|$. They have mainly established that these two errors converge geometrically with the number of iterations of the algorithm, up to an error term related to the quality of approximation of u by the operator $\mathcal{P}$ (we do not detail the required assumptions on the basis functions $(\mathcal{C}_i)_i$):

Theorem 7.3.1. *(Theorems 3.1 and 3.2 in [GM05a])*

i)

$$m_n \leq \rho_m m_{n-1} + \mathcal{E}_m(\mathcal{P}u - u),$$

where $\rho_m < 1$ (and depends on Δ, N and $\mathcal{P}$) when Δ is small enough, and $\mathcal{E}_m(\mathcal{P}u - u)$ is some norm (depending on Δ and on N) related to the approximation error $\mathcal{P}u - u$. And, consequently,

$$\limsup_n m_n \leq \frac{1}{1-\rho_m}\mathcal{E}_m(\mathcal{P}u - u).$$

ii)

$$v_n \leq \rho_v v_{n-1} + \frac{1}{M}\mathcal{E}_v(\mathcal{P}u - u),$$

where $\rho_v < 1$ (and depends on Δ, N, M and $\mathcal{P}$) when Δ is small enough and M large enough, and $\mathcal{E}_v(\mathcal{P}u - u)$ is some norm (depending on Δ and on N) related to the approximation error $\mathcal{P}u - u$. And, consequently,

$$\limsup_n v_n \leq \frac{1}{1-\rho_v}\frac{1}{M}\mathcal{E}_v(\mathcal{P}u - u).$$

Remark 7.3.2. *If u belongs to the chosen approximation space, i.e. $\mathcal{P}u = u$, then $\mathcal{E}_m(\mathcal{P}u - u) = \mathcal{E}_v(\mathcal{P}u - u) = 0$, and the bias and the variance converge geometrically to 0 (when Δ is small enough and M is large enough).*

The above estimates show that we do not need to take M large to have an accurate evaluation, since small values of M may be compensated by small approximation error $\mathcal{P}u - u$. This is the key feature of SCV method.

Chapter 8

Valuation of CDOs with stochastic recoveries using sequential control variates

Hereafter, $(\Omega, \mathcal{F}, \mathbb{P})$ is a filtered probability space , where $\mathcal{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ (T is a fixed terminal time) is the filtration (satisfying the usual conditions) that represents the flow of information we handle for our pricing purpose.
For a given Feller process $X = (X_t)_t$, we denote by $\mathcal{F}^X$ its natural filtration, augmented with the $\mathbb{P}$-null sets (which also satisfies the usual conditions, see [RY94]).
We denote the conditional expectation $\mathbb{E}(Y|\mathcal{F}_t)$ of a random variable Y by $\mathbb{E}^{\mathcal{F}_t}(Y)$.

8.1 The model

8.1.1 The portfolio loss and default processes

We consider a credit portfolio composed of n_0 names (that range from 50 to 150 in practice), and we assume that each name (firm) q may default at most once in the time interval $[0, T]$, at a random time τ_q.
We define the (normalized) *portfolio loss process* $(L_t)_{t \in [0,T]}$ by

$$L_t := \frac{1}{n_0} \sum_{q=1}^{n_0} (1 - R_q) \mathbb{1}_{\tau_q \leq t},$$

where the (stochastic) *recoveries* $(R_q)_{1 \leq q \leq n_0}$ are i.i.d. random variables (also independent of the default times $(\tau_q)_q$) taking values in [0,1] (which implies that $L_t \in [0, 1]$). The names notionals (ν_q) are supposed to be equal.
As pointed out in [BCJ08](Section 2.3), the loss process L may not be a sufficient statistic for pricing credit derivatives in practice. As an additional factor on which our model will depend (in order to take into account a possible default contagion, see the paragraph 8.1.2), we introduce the *default process* $(N_t)_{t \in [0,T]}$, where N_t represents

the number of defaults occuring before time t:

$$N_t := \sum_{q=1}^{n_0} \mathbb{1}_{\tau_q \leq t}.$$

In this way, we have a one-factor model, and we adopt a *top* approach taking as a filtration $\mathcal{F}$ the one generated by both the loss process L and the default process N: $\mathcal{F} := \mathcal{F}^L \vee \mathcal{F}^N$.

8.1.2 The choice of the parameters

The intensity process λ

We will assume that the compensator of the (non-decreasing and adapted) process N is time-differentiable. We denote its time derivative, which represents the *default intensity process*, by $(\lambda_t)_{t\in[0,T]}$ (this means that $(N_t - \int_0^t \lambda_s \mathrm{d}s)_t$ is a local martingale). Different models for the intensity process λ have been suggested in the literature, attempting to take into account some contagion phenomena that may appear in the markets reality.
The so-called *affine* model was proposed by Duffie and Gârleanu in [DG01] (following a bottom-up approach). They supposed that λ is a *basic affine process with parameters* $(\kappa, \theta, \sigma, \mu, l)$, defined by the following SDE

$$\mathrm{d}\lambda_t = \kappa(\theta - \lambda_t)\mathrm{d}t + \sigma\sqrt{\lambda_t}\mathrm{d}W_t + \mathrm{d}J_t,$$

where J is a compound Poisson process, independent of the Brownian motion W, with intensity l and exponentially distributed jumps with mean μ. The class of such processes has the advantages of being stable by summation (under some restrictions on the parameters) and of providing a computationally tractable model for portfolio default and loss. The default intensity process of each name q is supposed to be equal to $\lambda_q := X + X_q$, where X and X_q are independent basic affine processes. Within this model (and other variants of it, where some dependence on the loss process L is introduced: we refer, for instance, to [LM07], [EGG09] and [GG09]), the contagion is expressed only through correlation in the changes of the obligors' default intensities (see section 3 in [DG01]), but not according to the number of defaults.
A proposal of a model in this direction, motivated by that of Duffie and Gârleanu, has been made by Longstaff and Rajan in [LR08]. In this paper, the authors assume that

$$L_t = 1 - e^{-\gamma_1 N_{1,t}} e^{-\gamma_2 N_{2,t}} e^{-\gamma_3 N_{3,t}},$$

where γ_1, γ_2 and γ_3 are positive constants defining jump sizes, and where the three factors $(N_{1,t})$, $(N_{2,t})$ and $(N_{3,t})$ are independent Poisson processes with respective intensities $(\lambda_{1,t})$, $(\lambda_{2,t})$ and $(\lambda_{3,t})$ that are independent Cox-Ingersoll-Ross (CIR) processes (i.e. basic affine processes with no jumps). These three factors can be seen as three levels of defaults spread: suppose that the recoveries are all equal to 0, then, when a jump of the ith Poisson process $N_{i,t}$ occurs ($i = 1$, 2 or 3), a proportion γ_i

of the available firms in the portfolio defaults.
In addition, a variety of models that express the intensity process λ_t as a function of N_t (although sometimes introduced in a bottom-up framework) have been considered, especially linear and exponential models.
A linear model has been mentioned by Frey and Backhaus in [FB08] (corresponding to the *linear counterparty risk model* of Jarrow and Yu [JY01]), where $\lambda_t := c_0 + c_1 N_t$ for some positive constants c_0 and c_1. A variant of this model, using an additional time change, has been suggested by Ding *et al.* in [DGT09], and another variant has been presented by Arnsdorff and Halperin [AH08] and by Lopatin and Misirpashayev [LM07], where an additional stochastic factor is substituted for the constants c_0 and c_1.
An exponential model has been suggested by Davis and Lo [DL01] (also used by Carmona and Crepey in [CC09]), where $\lambda_t := c_0 e^{c_1 N_t}$, expressing a fast contagion growth with respect to the number of defaults.
A kind of intermediate model between the linear description and the exponential one has been introduced by Frey and Backhaus in [FB08] (the *convex counterparty risk model*). It is defined by

$$\lambda_t := c_0 + \frac{c_1}{c_2}\left(e^{c_2(N_t-\mu(t))^+} - 1\right), \tag{8.1}$$

where c_0, c_1, c_2 are positive constants, and μ represents the number of defaults one can *a priori* expect. This means that there is no contagion if the observed number of defaults is less than the expected one (if $N_t < \mu$, $\lambda_t = c_0$), otherwise, the contagion grows exponentially with respect to the number of defaults (in a catastrophic scenario, for example). Note also that one can make this model almost linear by choosing c_2 "small" (indeed, when $c_2 \simeq 0$, $\lambda_t \simeq c_0 + c_1(N_t - \mu)^+$).
Our choice. Thanks to the Markovian projection theorem (or the "mimicking theorem", see Proposition 1 of [CM08]), we can assume (w.l.o.g. concerning the law of (L, N), and so the CDO tranche spreads: see Corollary 1 and Remark 2 of the same reference [CM08]) that λ_t is a (deterministic) function of (t, L_{t^-}, N_{t^-}):

$$\lambda_t := \lambda(t, L_{t^-}, N_{t^-})$$

(and $\lambda(t, l, n)$ vanishes when $n \geq n_0$). Indeed, for the purpose of computation of $\mathbb{E}((L_T - K)^+)$, only the one-dimensional marginal laws of (L_t, N_t) have to be specified. These marginal laws coincide with those of a Markovian process $(L_t, N_t)_{t\in[0,T]}$ with an intensity of the above type.
Besides, we adopt the Frey-Backhaus model (8.1) for its flexibility.

The recoveries $(R_q)_q$

In the literature devoted to the valuation and hedging of CDOs, the recoveries $(R_q)_q$ are often supposed to be deterministic, for the sake of simplicity, and equal to a constant (40%, for example). However, this assumption may not be realistic, and, as pointed out in [Ron04], the impact of the recovery rate on the defaults distribution

may be very important. It is also showed in [DG01] that the recoveries distribution may differ from one type of debt to another (but then, for simplicity, the authors assumed a uniform distribution).
In our study, the recoveries $(R_q)_q$ are i.i.d. random variables, also independent of the default times $(\tau_q)_q$, and we assume, as in [Ron04], that they follow a beta distribution on $(0,1)$, with parameters $a > 0$ and $b > 0$. The density of this distribution is defined by

$$\beta^{a,b}(z) = \frac{1}{B(a,b)} z^{a-1}(1-z)^{b-1}\mathbb{1}_{(0,1)}(z), \tag{8.2}$$

where $B(a,b) = \int_0^1 z^{a-1}(1-z)^{b-1}\mathrm{d}z = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, and Γ is the gamma function.
This provides a versatile family of distributions with bounded support. Indeed, if one varies the parameters a and b, one obtains different shapes of distributions (Figure 8.1):

- the case $a = b = 1$ corresponds to the uniform distribution,
- the case $a = b = \frac{1}{2}$ corresponds to the arcsine distribution,
- if $a = b$, it is a symmetric distribution, and it is asymmetric otherwise.

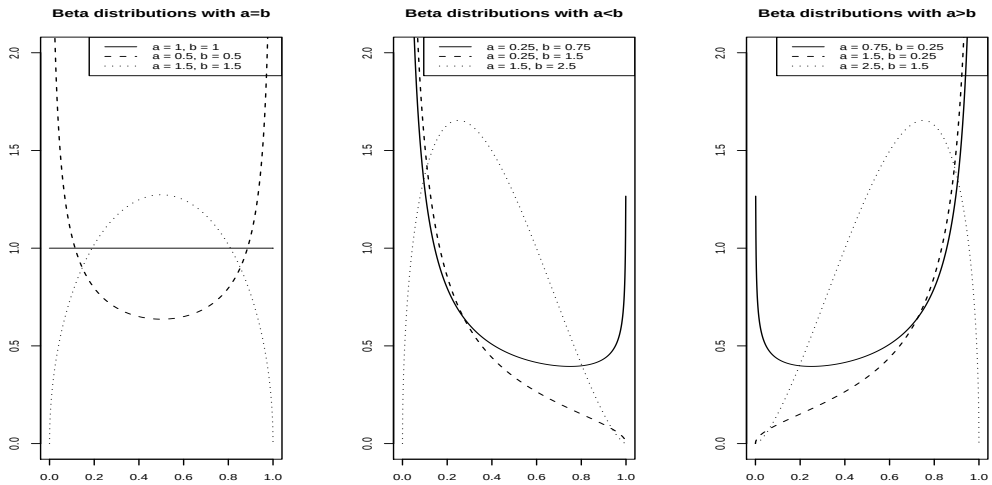

Figure 8.1: The beta distribution.

For the numerical simulation of the beta distribution, see paragraph 8.3.4.
The main advantage of taking a beta distribution for the recoveries is to be able to calibrate the parameters a and b from the knowledge of the mean $\mu := \mathbb{E}(R)$ and the

variance $\sigma^2 := \mathbb{V}ar(R)$ of the recoveries. Indeed (see [Ron04]),

$$a = \frac{\mu^2(1-\mu)}{\sigma^2} - \mu,$$
$$b = \frac{\mu(1-\mu)^2}{\sigma^2} - (1-\mu).$$

Remark 8.1.1. *It is easy to see that $1 - R_q$ follows the beta distribution (on $(0,1)$) with density function $\beta^{b,a}$.*

8.2 The price function

We recall (see section 6.1) that pricing CDO tranches amounts to evaluating expectations of the form $\mathbb{E}((L_T - K)^+)$, for different dates T (that may be equal to 1,2,...,10 years, for example) and different attachment points K (that typically may be equal to 0%, 3%, 6%, 9%, 12%, 22% and 100%, for example).
For fixed T and K, we introduce the price function u, defined on $[0,T] \times [0,\infty) \times \mathbb{N}$ by

$$u(t,l,n) := \mathbb{E}[g(L_T, N_T)|L_t = l, N_t = n], \tag{8.3}$$

where $g(l,n) := (l-K)^+$. We state the following result, showing that, under our model, the definition (8.3) of the function u may be interpreted as a representation of Feynman-Kac type of the solution of a linear partial integro-differential equation (PIDE for short).

Proposition 8.2.1 (Feynman-Kac representation)**.** *Let λ, $\tilde{f}$ and $\tilde{g}$ be uniformly bounded functions such that the following PIDE*

$$\begin{cases} \partial_t v(t,l,n) + \lambda(t,l,n) \int_0^1 \left(v(t, l + \frac{y}{n_0}, n+1) - v(t,l,n) \right) \beta^{b,a}(y)dy = \tilde{f}(t,l,n), \ t < T, \\ v(T,l,n) = \tilde{g}(l,n). \end{cases} \tag{8.4}$$

has a uniformly bounded solution $v(t,l,n) : [0,T] \times [0,\infty) \times \mathbb{N} \to \mathbb{R}$, which is differentiable w.r.t. to t (we will call this solution a classic solution). Then, this solution is unique, and is given by

$$v(t,l,n) = \mathbb{E}[\tilde{g}(L_T, N_T) - \int_t^T \tilde{f}(s, L_s, N_s)\mathrm{d}s | L_t = l, N_t = n], \tag{8.5}$$

for $(t,l,n) \in [0,T] \times [0,\infty) \times \mathbb{N}$.

Proof. We denote by $(N_s^{t,l,n})_{t \le s \le T}$ and $(L_s^{t,l,n})_{t \le s \le T}$ (the default process and the loss process, started, resp., from n and l at time t), the jump (Cox) processes such that $N_t^{t,l,n} = n$ and $L_t^{t,l,n} = l$, with simultaneous jumps, and with default intensity process $(\lambda(s, L_{s^-}^{t,l,n}, N_{s^-}^{t,l,n}))_{t<s\le T}$. The jumps of $(N_s^{t,l,n})_s$, denoted by $(\Delta N_s^{t,l,n})_s := (N_s^{t,l,n} - N_{s^-}^{t,l,n})_s$, follow the distribution of δ_1 (unit mass at 1), and those of $(L_s^{t,l,n})_s$,

denoted by $(\Delta L_s^{t,l,n})_s := (L_s^{t,l,n} - L_{s^-}^{t,l,n})_s$, follow the distribution of $\frac{1-R}{n_0}$, where R follows the beta distribution $\beta^{a,b}$.
One can decompose each of the processes $(L_s^{t,l,n})_s$ and $(N_s^{t,l,n})_s$ into the sum of its jumps. Then, noting that both processes always jump simultaneously, one writes the "Itô formula" between times t and T to obtain

$$\begin{aligned} v(t, L_t^{t,l,n}, N_t^{t,l,n}) &= v(T, L_T^{t,l,n}, L_T^{t,l,n}) - \int_t^T \partial_s v(s, L_s^{t,l,n}, N_s^{t,l,n}) \mathrm{d}s \\ &\quad - \sum_{t<s\leq T} (v(s, L_{s^-}^{t,l,n} + \Delta L_s^{t,l,n}, N_{s^-}^{t,l,n} + \Delta N_s^{t,l,n}) - v(s, L_{s^-}^{t,l,n}, N_{s^-}^{t,l,n})). \end{aligned}$$

Thus, taking the expectation of both sides,

$$\begin{aligned} v(t,l,n) &= \mathbb{E}(\tilde{g}(L_T^{t,l,n}, N_T^{t,l,n})) - \mathbb{E}(\int_t^T \partial_s v(s, L_s^{t,l,n}, N_s^{t,l,n}) \mathrm{d}s) \\ &\quad - \mathbb{E}(\sum_{t<s\leq T} (v(s, L_{s^-}^{t,l,n} + \Delta L_s^{t,l,n}, N_{s^-}^{t,l,n} + \Delta N_s^{t,l,n}) - v(s, L_{s^-}^{t,l,n}, N_{s^-}^{t,l,n}))) \\ &= \mathbb{E}(\tilde{g}(L_T^{t,l,n}, N_T^{t,l,n})) - \mathbb{E}(\int_t^T \partial_s v(s, L_s^{t,l,n}, N_s^{t,l,n}) \mathrm{d}s) \\ &\quad - \mathbb{E}(\int_t^T \mathrm{d}s \ \{\lambda(s, L_s^{t,l,n}, N_s^{t,l,n}) \int_0^1 \mathrm{d}y \\ &\quad \beta^{b,a}(y)[v(t, L_s^{t,l,n} + \frac{y}{n_0}, N_s^{t,l,n} + 1) - v(t, L_s^{t,l,n}, N_s^{t,l,n})]\}), \end{aligned}$$

where we have used, in the last equality, the independence between the jumps of L and the other variables, Remark 8.1.1 and the fact that $\{s : \Delta N_s^{t,l,n} \neq 0\}$ has a null Lebesgue-measure.
Now, using the equation (8.4) satisfied by v, we obtain

$$v(t,l,n) = \mathbb{E}[\tilde{g}(L_T^{t,l,n}, N_T^{t,l,n}) - \int_t^T \tilde{f}(s, L_s^{t,l,n}, N_s^{t,l,n}) \mathrm{d}s],$$

which proves the proposition (using the Markov property of the process (L, N)). □

Now, in our model, $\lambda(t,l,n)$ vanishes as soon as $n \geq n_0$. In this case, it turns out that the PIDE (8.4) has a classic solution (and it is the unique solution), assuming that the equation parameters are bounded.

Theorem 8.2.1. *Suppose that the functions λ, $\tilde{f}$ and $\tilde{g}$ are uniformly bounded. Moreover, assume that, when $n \geq n_0$, $\lambda(t,l,n) = 0$. Then, the PIDE (8.4) has a unique classic solution, and (from Proposition 8.2.1) the Feynman-Kac representation (8.5) holds.*

Proof. The PIDE (8.4) can be viewed as a system of linear ODEs (ordinary differential equations). Indeed, denoting $v(t,l,n)$ by $v_n(t,l)$, and $[\tilde{f}(t,l,n) - \lambda(t,l,n) \int_0^1 v(t, l+$

$\frac{y}{n_0}, n+1)\beta^{b,a}(y)dy]$ by $\tilde{I}(v_{n+1})(t,l,n)$, (8.4) writes

$$\begin{cases} \partial_t v_n(t,l) - \lambda(t,l,n)v_n(t,l) = \tilde{I}(v_{n+1})(t,l,n), \ t < T, \\ v_n(T,l,n) = \tilde{g}(l,n). \end{cases} \tag{8.6}$$

When $n \geq n_0$, $\lambda(t,l,n) = 0$, and it is clear that $v_n(t,l) = \tilde{g}(l,n) - \int_t^T \tilde{f}(s,l,n)\mathrm{d}s$ is the unique solution. When $n < n_0$, the system (8.6) can be solved backwardly in n: starting from the knowledge of v_{n_0}, one can express v_n in terms of v_{n+1}, for $n = n_0 - 1, ..., 0$, by solving explicitly (and the solution is unique) the recursive linear equation (8.6):

$$v_n(t,l) = \left(\tilde{g}(l,n) - \int_t^T \tilde{I}(v_{n+1})(s,l,n)\exp(\int_s^T \lambda(r,l,n)\mathrm{d}r)\mathrm{d}s\right)\exp(-\int_t^T \lambda(s,l,n)\mathrm{d}s).$$

Now, it is easy to check that v_n is time-differentiable and uniformly bounded (independently of n). □

8.3 Computation of the price function using SCV

Actually, the proof of Theorem 8.2.1 gives an iterative method to compute the price function. However, at each iteration n, one has to evaluate the integral term $\tilde{I}(v_{n+1})(s,l,n)$, which requires the knowledge of the value of $v_{n+1}(t,l)$ at many points l of $[0,1]$. In addition, one has errors due to the discretization of the space integrals and of the time integrals that spread from an iteration to the next. We will prefer to consider the system (8.4) as a single PIDE equation, with three continuous variables (t,l,p), where $p := \frac{n}{n_0}$, that we will solve in $\bar{D} := [0,T] \times [0,1] \times [0,1]$. The domain $\bar{D}$ being bounded, it is well adapted for interpolation, and we will apply the SCV algorithm to compute the solution.

8.3.1 Notations

From now on, u, v, λ, $\tilde{f}$ and $\tilde{g}$ are bounded functions of $(t,l,p) \in \bar{D}$ ($p := \frac{n}{n_0}$).
We define the *defaults proportion process* $(P_t)_{t\in[0,T]}$ by

$$P_t := \frac{N_t}{n_0} = \frac{1}{n_0}\sum_{q\geq 1} \mathbb{1}_{\tau_q \leq t}.$$

The defaults proportion process and the loss process started resp. from p and l at time t are denoted resp. by $(P_s^{t,l,p})_{t\leq s\leq T}$ and $(L_s^{t,l,p})_{t\leq s\leq T}$, and are defined in the same way as in the proof of Proposition 8.2.1.
We denote by X the Markov process (L,P):

$$X_t := (L_t, P_t), \quad X_s^{t,l,p} := (L_s^{t,l,p}, P_s^{t,l,p}).$$

We will assume that λ is uniformly bounded on $\bar{D}$, and that $\lambda(t,l,p) = 0$ when $p > 1 - \frac{1}{n_0}$ (i.e. when $n > n_0 - 1$) or when $l > p$ (which is natural, since $L_t \leq P_t$).

For generic uniformly bounded (on $\bar{D}$) functions $\tilde{f}$ and $\tilde{g}$, and for $f(t,l,p) := 0$ and $g(l,p) := (l-K)^+$, we set

$$\begin{aligned}\Psi(\tilde{f},\tilde{g},X^{t,l,p}) &:= \tilde{g}(X_T^{t,l,p}) - \int_t^T \tilde{f}(s,X_s^{t,l,p})\mathrm{d}s,\\ v(t,l,p) &:= \mathbb{E}[\Psi(\tilde{f},\tilde{g},X^{t,l,p})],\\ u(t,l,p) &:= \mathbb{E}[\Psi(f,g,X^{t,l,p})] = \mathbb{E}[(L_T-K)^+|L_t=l,P_t=p].\end{aligned}$$

From Theorem 8.2.1, v is the classic solution of

$$\begin{cases} \mathcal{A}v &= \tilde{f} \quad \text{when } t<T,\\ \mathcal{B}v &= \tilde{g} \quad \text{when } t=T,\end{cases}$$

where

$$\mathcal{A}v(t,l,p) := \partial_t v(t,l,p) + \lambda(t,l,p)\int_0^1 \left(v(t,l+\frac{y}{n_0},p+\frac{1}{n_0}) - v(t,l,p)\right)\beta^{b,a}(y)dy, \tag{8.7}$$

$$\mathcal{B}v(t,l,p) := v(t,l,p),$$

and so is the price function u we aim to compute, with $\tilde{f}=f$ and $\tilde{g}=g$.

8.3.2 Interpolation operator and points

Since one wants to evaluate $u(0,0,0)$, we will choose the Gauss-Lobatto-Tchebychev interpolation points: $(t_i)_{0\le i\le N^T}$, $(l_j)_{0\le j\le N^L}$ and $(p_k)_{0\le k\le N^P}$ (where N^T, N^L and N^P are fixed integer numbers) in order to include the boundaries of the respective intervals $[0,T]$, $[0,1]$ and $[0,1]$. These points are defined by

$$t_i := \frac{T}{2}(1+x_i^{N^T}),\ l_j := \frac{1}{2}(1+x_j^{N^L}),\ p_k := \frac{1}{2}(1+x_k^{N^P}),$$

where $(x_q^{N^0})_{0\le q\le N^0}$ (for some fixed integer N^0) denote the Gauss-Lobatto points on $[-1,1]$, defined by

$$x_q^{N^0} := \cos\left(\frac{N^0-q}{N^0}\pi\right),\ 0\le q\le N^0.$$

The related Gauss-Tchebychev interpolation basis functions on $[-1,1]$ are defined by

$$\mathcal{C}_q^{N^0}(.) := \sum_{q'=0}^{N^0} \alpha_{q,q'}^{N^0}\mathcal{T}_{q'}(.),\ 0\le q\le N^0,$$

where $(\mathcal{T}_q)_{0\le q\le N^0}$ are the Tchebychev polynomials ($\mathcal{T}_q(x) = \cos(q\arccos(x))$), and $(\alpha_{q,q'}^{N^0})_{0\le q,q'\le N^0}$ are constant coefficients (depending only on the points $(x_q^{N^0})$ and on

the $(\mathcal{T}_q)_q$), given by (see e.g. [GM05a])

$$\alpha^{N^0}_{q_1,q_2} = \begin{cases} \frac{1}{2N^0} & \text{if } q_1 = 0 \text{ and } q_2 = 0, \\ \frac{(-1)^{q_2}}{N^0} & \text{if } q_1 = 0 \text{ and } 0 < q_2 < N^0, \\ -\sum_{q_3=0}^{N^0-1} \alpha^{N^0}_{0,q_3} & \text{if } q_1 = 0 \text{ and } q_2 = N^0, \\ \frac{1}{N^0} & \text{if } 0 < q_1 < N^0 \text{ and } q_2 = 0, \\ \frac{\mathcal{T}_{q_2}(x^{N^0}_{q_1})}{2N^0} & \text{if } 0 < q_1 < N^0 \text{ and } 0 < q_2 < N^0, \\ -\sum_{q_3=0}^{N^0-1} \alpha^{N^0}_{q_1,q_3} & \text{if } 0 < q_1 < N^0 \text{ and } q_2 = N^0, \\ \frac{1}{2N^0} & \text{if } q_1 = N^0 \text{ and } q_2 = 0, \\ \frac{1}{N^0} & \text{if } q_1 = N^0 \text{ and } 0 < q_2 < N^0, \\ -\sum_{q_3=0}^{N^0-1} \alpha^{N^0}_{N^0,q_3} + 1 & \text{if } q_1 = N^0 \text{ and } q_2 = N^0. \end{cases}$$

The related interpolation operator is then given by

$$(\mathcal{P}^{N^0})\varphi(.) := \sum_q \varphi(x^{N^0}_q)\mathcal{C}^{N^0}_q(.) = \sum_{q_1,q_2} \alpha^{N^0}_{q_1,q_2}\varphi(x^{N^0}_{q_1})\mathcal{T}_{q_2}(.).$$

For $(t,l,p) \in [0,T] \times [0,1] \times [0,1]$, and for $(i,j,k) \in \{0,...,N^T\} \times \{0,...,N^L\} \times \{0,...,N^P\}$, we set

$$\mathcal{T}^T_i(t) := \mathcal{T}_i(\frac{2}{T}t - 1),\ \mathcal{T}^L_j(l) := \mathcal{T}_j(2l-1),\ \mathcal{T}^P_k(p) := \mathcal{T}_k(2p-1),$$

$$\tilde{\mathcal{C}}^T_i(t) := \mathcal{C}^{N^T}_i(\frac{2}{T}t - 1),\ \tilde{\mathcal{C}}^L_j(l) := \mathcal{C}^{N^L}_j(2l-1),\ \tilde{\mathcal{C}}^P_k(p) := \mathcal{C}^{N^P}_k(2p-1).$$

Our interpolation operator $\mathcal{P}$ then writes

$$\begin{aligned}(\mathcal{P}\varphi)(t,l,p) &:= \sum_{i,j,k} \varphi(t_i,l_j,p_k)\tilde{\mathcal{C}}_{i,j,k}(t,l,p) \\ &= \sum_{i_1,j_1,k_1,i_2,j_2,k_2} \alpha^{N^T}_{i_1,i_2}\alpha^{N^L}_{j_1,j_2}\alpha^{N^P}_{k_1,k_2}\varphi(t_{i_1},l_{j_1},p_{k_1})\mathcal{T}_{i_2,j_2,k_2}(t,l,p),\end{aligned}$$

where

$$\begin{aligned}\tilde{\mathcal{C}}_{i,j,k}(t,l,p) &:= \tilde{\mathcal{C}}^T_i(t)\tilde{\mathcal{C}}^L_j(l)\tilde{\mathcal{C}}^P_k(p), \\ \mathcal{T}_{i,j,k}(t,l,p) &:= \mathcal{T}^T_i(t)\mathcal{T}^L_j(l)\mathcal{T}^P_k(p).\end{aligned} \tag{8.8}$$

8.3.3 Algorithm

Here is the sequential control variates (SCV) algorithm, applied to the computation of the price function u:

Algorithm 8.3.1. *1. Initialization:* $u_0 := 0$.

2. *At iteration n: for each (t_i, l_j, p_k), simulate M independent paths $(X^{n,m}(t_i, l_j, p_k))_{1\leq m\leq M}$ $= (L^{n,m}(t_i, l_j, p_k), P^{n,m}(t_i, l_j, p_k))$ of $X^{t_i,l_j,p_k} = (L^{t_i,l_j,p_k}, P^{t_i,l_j,p_k})$ (independently of anything else), and set*

$$\bar{y}_n(t_i, l_j, p_k) := \frac{1}{M}\sum_{m=1}^{M} \Psi(f - \mathcal{A}u_{n-1}, g - \mathcal{B}u_{n-1}, X^{n,m}(t_i, l_j, p_k)).$$

3. *Always at iteration n: build the global solution using the interpolation of the correction term:*

$$u_n := u_{n-1} + \mathcal{P}\bar{y}_n.$$

Remark 8.3.1. *When approximating the correction term y_n in step 2 of the algorithm above, one needs a Monte-Carlo simulation of quantities of the form $e_n(t, l, p) := \mathbb{E}[\int_t^T \tilde{f}_n(s, X_s(t, l, p))\mathrm{d}s]$ (that come from Ψ). Now, it is easy to check that*

$$e_n(t, l, p) = (T - t)\mathbb{E}[\tilde{f}_n\left(T - (T - t)U, X_{T-(T-t)U}(t, l, p)\right)],$$

where U is a $[0, 1]$-uniformly distributed random variable, independent of X. Then, one can approximate $e_n(t, l, p)$ by

$$\bar{e}_n(t, l, p) := \frac{1}{M}\sum_{m=1}^{M}(T - t)\tilde{f}_n\left(T - (T - t)U^{n,m}, X^{n,m}_{T-(T-t)U^{n,m}}(t, l, p)\right),$$

where $(U^{n,m})_{n,m}$ are independent copies of U (and independent of anything else).

Now, all what we need to know is how to simulate the process $X = (L, P)$, and how to compute the functions $\mathcal{T}_{i,j,k}$ and (as mentioned in Remark 7.3.1) the functions $\mathcal{A}\mathcal{T}_{i,j,k}$. There are no extra computations to perform for $\mathcal{B}\mathcal{T}_{i,j,k} = \mathcal{T}_{i,j,k}$.

8.3.4 Simulation of the loss and the defaults proportion processes

We denote the jump times by $0 \leq \tau_1^* < ... < \tau_q^* < ... \leq T$.
We first give a simulation algorithm for the beta distribution, since the jump sizes $(L_{\tau_{q+1}^*} - L_{\tau_q^*})_q$ of the loss process L are i.i.d. random variables distributed as $\frac{Y}{n_0}$, where Y is a random variable with beta distribution density $\beta^{b,a}$. Then, we give an algorithm to generate the jump times and the process $X = (L, P)$.

Simulation of the beta distribution

Different algorithms for generating a random variable Y of beta distribution (on $(0, 1)$) with positive parameters (a, b) (defined by its density $\beta^{a,b}$, see (8.2)) have been proposed in the literature.
When $a, b \geq 1$, one can think about the pure acceptance-rejection method (see Example 2.2.6 in [Gla04]). Indeed, in this case, one can show that $\sup_z \beta^{a,b}(z) = \beta^{a,b}(\frac{a-1}{a+b-2}) := c$, so $\beta^{a,b} \leq c\beta^{1,1}$, where $\beta^{1,1}$ is the density of the uniform distribution on $[0, 1]$. Then, the acceptance-rejection algorithm is the following (for $a, b > 1$):

1. Generate Z and U of uniform distribution on $[0,1]$ until $\beta^{a,b}(Z) > cU$.

2. Return Z.

More sophisticated algorithms, available for $a, b > 0$, and based on change of variables with a more careful use of the acceptance-rejection technique, have been designed (see [Dev86] or [JKB95]). We present the following (and it is the one we will adopt), which, as mentioned in [JKB95], has the advantage of being reasonably fast (faster than the acceptance-rejection algorithm above) for values of a and b down to about 0.5.

Algorithm 8.3.2. *(Simulation of the beta distribution with parameters $a, b > 0$, Chen(1978)) Initialization: set $\alpha := a + b$. If $\min(a,b) \leq 1$ then set $\beta := \max(\frac{1}{a}, \frac{1}{b})$, else set $\beta := \sqrt{\frac{\alpha-2}{2ab-\alpha}}$. Set $\gamma := a + \frac{1}{\beta}$.*

1. *Generate uniform (on (0,1)) random variables U_1 and U_2, and set $V := \beta \log(\frac{U_1}{1-U_1})$, $W := ae^V$.*

2. *If $\alpha \log(\frac{\alpha}{b+W}) + \gamma V - 1.3862944 < \log(U_1^2 U_2)$ then return to step 1.*

3. *Return $Z := \frac{W}{b+W}$.*

We propose, in addition, an algorithm consisting in generating Z as $\tilde{G}^{a,b}(U)$. U denotes a uniform (on $[0,1]$) random variable, and $\tilde{G}^{a,b}$ is an approximation (whenever one has not the exact expression) of $G^{a,b} := [F^{a,b}]^{-1}$, where $F^{a,b}$ is the cumulative distribution function of Z ($F^{a,b}(z) := \int_{-\infty}^{z} \beta^{a,b}(z)\mathrm{d}z$), and $[F^{a,b}]^{-1}$ denotes its inverse function. This approximation is based on a series expansion of $G^{a,b}$: for $w \in (0,1)$,

$$G^{a,b}(w) = [F^{a,b}]^{-1}(w) \approx \tilde{G}^{a,b}(w) := z_0 + \sum_{i=1}^{I} \frac{c_i(w_0)}{i!}(w - w_0)^i, \tag{8.9}$$

where $z_0 \in (0,1)$ is a fixed initial approximation, $w_0 := F^{a,b}(z_0)$, and $c_i(w_0) := \frac{\mathrm{d}G^{a,b}}{\mathrm{d}w}(w_0)$ are constant coefficients (depending only on a, b and w_0). Actually, this series expansion method for the inverse of the beta distribution has been proposed by Abernathy and Smith [AS93] with the aim of finding percentiles of the F-distribution. They have given an iterative method for the exact computation of the series coefficients $(c_i(w_0))$ (see [AS93]).
This algorithm turns to be fast for values of a and b between about 0.5 and 1 (with these values, it is even three to four times faster than Algorithm 8.3.2), but it may either slow down or diverge for other values of a and b. In fact, as mentioned also in [AS93], the accuracy of the algorithm depends on the initial approximation z_0: if z_0 is too far from $G^{a,b}(w)$, then the partial sum in (8.9) may diverge.

Simulation of $X = (L, P)$

Conditioned on $(L_{\tau^*_{q-1}}, N_{\tau^*_{q-1}})$ (we recall that $N_t = n_0 P_t$ is the default process), the process $(N_t)_{\tau^*_{q-1} \le t < \tau^*_q}$ is a non-homogeneous Poisson process, with intensity $\lambda(t, L_{\tau^*_{q-1}}, P_{\tau^*_{q-1}})$. We then apply the simulation algorithm mentioned by Burnecki *et al.* in [BHW04] (also known as the *fictitious chock* technique, used in neutronics, see [LPS98], page 84), based on the acceptance-rejection method.
This algorithm assumes the intensity $\tilde{\lambda}(t)$ of the non-homogeneous Poisson process to simulate to be uniformly bounded, which is the case in our framework. We set

$$\bar{\lambda} := \sup_{(t,l,p)} \lambda(t, l, p).$$

Algorithm 8.3.3 (Simulation of (L, P)). *1. Set $\tau^*_0 := 0$, $\tau^* := 0$, $N_{\tau^*_0} := 0$ and $L_{\tau^*_0} := 0$.*

2. For $q = 1, 2, ...$ do

2a. Generate an exponential random variable E with intensity $\bar{\lambda}$.

2b. Set $\tau^* := \tau^* + E$.

2c. Generate a random variable U distributed uniformly on (0,1).

2d. If $\bar{\lambda} U > \lambda(\tau^*, L_{\tau^*_{q-1}}, P_{\tau^*_{q-1}})$ then return to step 2a (i.e. reject the jump time), else generate a random variable Y with beta distribution (on $(0, 1)$) of parameters (b,a) (using Algorithm 8.3.2), and set $\tau^*_q := \tau^*$ (i.e. accept the jump time), $P_{\tau^*_q} := P_{\tau^*_{q-1}} + \frac{1}{n_0}$ and $L_{\tau^*_q} := L_{\tau^*_{q-1}} + \frac{Y}{n_0}$.

8.3.5 Computation of $\mathcal{T}_{i,j,k}$ and $\mathcal{A}\mathcal{T}_{i,j,k}$

The $(\mathcal{T}_{i,j,k})$'s

From the definition of $\mathcal{T}_{i,j,k}$ (8.8), it is clear that all one needs for its computation is to calculate the Tchebychev polynomials $\mathcal{T}_q$. This can be performed iteratively on $q = 0, 1, ...$ in an exact way (instead of using the expression $\mathcal{T}_q(x) = \cos(q \arccos(x))$, where one is faced with the numerical approximation errors of the functions cos and arccos), using the following recursive relation (see e.g. [AS64]):

$$\begin{cases} \mathcal{T}_0(x) = 1, \ \mathcal{T}_1(x) = x, \\ \mathcal{T}_q(x) = 2x\mathcal{T}_{q-1}(x) - \mathcal{T}_{q-2}(x), \ q \ge 2. \end{cases} \tag{8.10}$$

The $(\mathcal{A}\mathcal{T}_{i,j,k})$'s

From the definitions of $\mathcal{A}$ (8.7) and $\mathcal{T}_{i,j,k}$ (8.8), we have

$$\begin{aligned} \mathcal{A}\mathcal{T}_{i,j,k}(t, l, p) = {} & \frac{\mathrm{d}\mathcal{T}_i^T}{\mathrm{d}t}(t)\mathcal{T}_j^L(l)\mathcal{T}_k^P(p) + \lambda(t, l, p)\mathcal{T}_i^T(t)\mathcal{T}_k^P(p + \frac{1}{n_0}) \int_0^1 \mathcal{T}_j^L(l + \frac{y}{n_0})\beta^{b,a}(y)dy \\ & - \lambda(t, l, p)\mathcal{T}_i^T(t)\mathcal{T}_j^L(l)\mathcal{T}_k^P(p). \end{aligned}$$

Thus, one only needs to compute $\frac{\mathrm{d}\mathcal{T}_i^T}{\mathrm{d}t}(t)$ and $I_j(l) := \int_0^1 \mathcal{T}_j^L(l + \frac{y}{n_0})\beta^{b,a}(y)dy$. Now, $\frac{\mathrm{d}\mathcal{T}_i^T}{\mathrm{d}t}(t) = \frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{T}_i(\frac{2}{T}t - 1)) = \frac{2}{T}\mathcal{D}_i(\frac{2}{T}t - 1)$, where $\mathcal{D}_i(x) := \frac{\mathrm{d}\mathcal{T}_i}{\mathrm{d}x}(x)$. Then, one computes the derivatives $(\mathcal{D}_q)$ of the Tchebychev polynomials, using the following relation (see e.g. [AS64]), combined with (8.10):

$$\begin{cases} \mathcal{D}_0(x) = 0, \\ (1-x^2)\mathcal{D}_q(x) = -qx\mathcal{T}_q(x) + q\mathcal{T}_{q-1}(x), \ q \geq 1. \end{cases}$$

Concerning the integral $I_j(l)$, one has

$$\begin{aligned} I_j(l) &= \int_0^1 \mathcal{T}_j(2l - 1 + 2\frac{y}{n_0})\beta^{b,a}(y)dy \\ &= \frac{1}{2^{a+b-2}B(b,a)} \int_{-1}^1 \mathcal{T}_j(2l - 1 + \frac{x+1}{n_0})(1+x)^{b-1}(1-x)^{a-1}dx, \end{aligned}$$

which represents (up to the constant $\frac{1}{2^{a+b-2}B(b,a)}$) the integral on $[-1,1]$ of the function $\psi_j^l : x \mapsto \mathcal{T}_j(2l - 1 + \frac{x+1}{n_0})$ (l is fixed) with the Jacobi weight $w(x) := (1+x)^{b-1}(1-x)^{a-1}$. Since $\psi_j^l(x)$ is a polynomial of degree j, one can compute the exact value of $I_j(l)$ using the following Gauss-Jacobi quadrature formula (that is actually exact for any polynomial of degree less or equal to $2N^L - 1$, instead of ψ_j^l: see [Fun92]).

$$I_j(l) = \frac{1}{2^{a+b-2}B(b,a)} \sum_{q=1}^{N^L} w_{q,N^L}^{b,a} \psi_j^l(x_{q,N^L}^{b,a}),$$

where $(w_{q,N^L}^{b,a})$ (the *Gauss-Jacobi weights*) and $(x_{q,N^L}^{b,a})$ (the *Gauss-Jacobi abscissae*) are constant coefficients (related to the zeros of the Jacobi polynomials) that do not depend on the function ψ_j^l. Numerical routines for the exact computation of these coefficients can be found, for instance, in [PTVF92].

Remark 8.3.2. *At first sight, one can think that we should have chosen the $(\mathcal{T}_j)$'s to be the Jacobi interpolation polynomials (with parameters b, a) instead of the Tchebychev ones, in order to be able to interprete the integral $I_j(l)$ as the inner product (with Jacobi weight $w(.)$) of a Jacobi polynomial $\mathcal{T}_j$ and 1, and then use the orthogonal property... However, this interpretation cannot hold, because, inside the integral $I_j(l)$, the polynomial $\mathcal{T}_j(.)$ and the weight $w(.)$ are not evaluated at the same point x.*

8.4 Numerical results

8.4.1 Preliminary tests

We want here to test the efficiency of the SCV algorithm in our framework. Whenever one explicitly knows the exact expression of $u(t,l,p) = \mathbb{E}[(L_T - K)^+ | L_t = l, P_t = p]$, one can compare the solution computed by SCV with u, and examine the convergence speed.

This is the case when $K = 0$ and $\lambda(t, l, p) = \lambda_0$ is constant for $p \in [0, 1]$ (and vanishes for $p > 1$): the default process $(N_s^{t,n})_{t \le s \le T}$ (starting from n at time t, and stopped at n_0) is given by

$$N_s^{t,n} = (n + \tilde{N}_{s-t}) \wedge n_0,$$

where $\tilde{N}$ is an homogeneous Poisson process with intensity λ_0. Then, one has (with $p = \frac{n}{n_0}$, and with the convention $\sum_{q=q_1}^{q_2} a_q := 0$ if $q_1 > q_2$)

$$L_T^{t,l,p} = l + \frac{1}{n_0} \sum_{q=n+1}^{N_T^{t,n}} (1 - R_q), \qquad P_T^{t,l,p} = \frac{N_T^{t,n}}{n_0},$$

and

$$\begin{aligned} u^0(t, l, p) &:= \mathbb{E}[L_T | L_t = l, P_t = p] = \mathbb{E}(L_T^{t,l,p}) \\ &= l + \frac{1}{n_0} \mathbb{E}(1 - R_1) \mathbb{E}(N_T^{t,n} - n) \\ &= l + \frac{1}{n_0} \mathbb{E}(1 - R_1) \mathbb{E}(\tilde{N}_{T-t} \wedge (n_0 - n)) \\ &= l + \frac{1}{n_0} \frac{b}{a+b} e^{-\lambda_0 (T-t)} \sum_{k=0}^{\infty} [k \wedge (n_0 - n)] \frac{(\lambda_0 (T-t))^k}{k!}. \end{aligned}$$

Therefore, the exact expression of u^0 is given by

$$u^0(t, l, p) = l + \frac{b}{n_0(a+b)} \left(n_0 - n - e^{-\lambda_0 (T-t)} \sum_{k=0}^{n_0 - n} (n_0 - n - k) \frac{(\lambda_0 (T-t))^k}{k!} \right).$$

For the SCV algorithm that we use to compute u^0, we use a Gauss-Lobatto-Tchebychev (GLT for short) interpolation of degree $N^T = 4$ for the variable t (since u^0 is infinitely smooth w.r.t. t, this is satisfying enough), a GLT interpolation of degree $N^L = 1$ for the variable l (this is an exact interpolation since u^0 is linear w.r.t. l). For the variable p, we will compare the GLT interpolation of degree 3 with the piecewise-constant (PC for short) interpolation with a $(N^P + 1)$-points equidistant grid (if $N^P = n_0$, the latter is an exact interpolation, since n takes the discrete values $0, 1, ..., n_0$).
We take $T = 1$ (and $K = 0$), and we compute a sequence $(u_I^0)_{I \ge 0}$ by SCV. We only focus on the value of $u^0(0, 0, 0)$: we stop the sequence when $|u_I^0(0, 0, 0) - u^0(0, 0, 0)| < 10^{-4}$, and we denote by I_{max} the corresponding number of performed iterations.
Concerning the number M_I of Monte-Carlo simulations at each iteration I, we start from $M_1 = 100$, and, if within 10 successive iterations the convergence is not reached yet, we multiply M by 2. We compare $M_{I_{max}}$ with the observed minimal number M_{min}^{MC} of crude Monte-Carlo simulations required for an error less than 10^{-4}, as well as the CPU times spent by both methods (as with SCV, the crude Monte-Carlo computations are made for all the grid points).

	SCV			Crude MC	
Interpolation (t,l,p)	I_{max}	$M_{I_{max}}$	CPU time	M_{min}^{MC}	CPU time
(GLT,GLT,PC) with (4,2,4) points	21	400	between 0 and 10 sec	5.10^6	90 sec
(GLT,GLT,GLT) with (4,2,4) points	> 40	> 2000	> 1 mn	5.10^6	130 sec

Table 8.1: $n_0 = 3$ and $\lambda_0 = 1.5$.

	SCV			Crude MC	
Interpolation (t,l,p)	I_{max}	$M_{I_{max}}$	CPU time	M_{min}^{MC}	CPU time
(GLT,GLT,PC) with (4,2,101) points	∞	∞	∞	∞	∞
(GLT,GLT,PC) with (4,2,10) points	∞	∞	∞	∞	∞
(GLT,GLT,PC) with (4,2,3) points	31	800	between 0 and 10 sec	5.10^5	110 sec
(GLT,GLT,PC) with (4,2,2) points	23	400	between 0 and 4 sec	5.10^5	60 sec

Table 8.2: $n_0 = 100$ and $\lambda_0 = 50$.

In the case of $n_0 = 3$ and $\lambda_0 = 1.5$, Table 8.1 shows that, in this particular example (but not in general; see Remark 8.4.1), the PC interpolation w.r.t. p is more suitable than the GLT interpolation: this is obvious, since we have made an exact interpolation ($N^P = 4 = n_0 + 1$).
In the case of $n_0 = 100$ and $\lambda_0 = 50$, Table 8.2 shows that it is not necessary (and it is even not recommended when n_0 is greater than 10) to take N^P equal to $n_0 + 1$ in this example: 2 or 3 knots are sufficient. After taking into account these remarks, the SCV algorithm applied to the present example turns to be about 10 times more efficient than the crude Monte-Carlo method.

Remark 8.4.1. *From the study of the above particular example, one may think that we should, in general, adopt a PC interpolation w.r.t. the variable p. However, we have just been "lucky" because, for fixed t and l, the function $u^0(t,l,p)$ is almost constant when p is in some "large" interval $[0,p_0]$, so that a PC interpolation w.r.t. p induces a very good approximation of $u^0(0,0,0)$. Moreover, globally on the interval $[0,1]$, the GLT interpolation performs a much better approximation of $u^0(0,0,p)$ than the PC interpolation (see Figure 8.2, where we have taken $T = 1$, $n_0 = 100$, $\lambda_0 = 50$, and 3 knots for each of the two interpolations).*

8.4.2 Numerical results for CDO tranches pricing

We fix $T = 1$, $n_0 = 100$, and we adopt the Frey-Backhaus model (8.1) for the intensity function λ, with $c_0 = 20$, $c_1 = 1$, $c_2 = 0.0001$ and $\mu = 10$. These values mean that

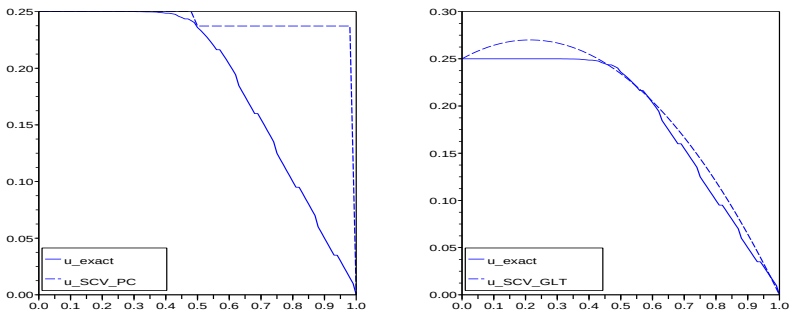

Figure 8.2: PC interpolation vs GLT interpolation, w.r.t. p.

the initial average number of defaults is 20 per year, but contagion appears as soon as 10 defaults occur, and then the default intensity grows almost linearly (see Figure 8.3). We fix also $a = 0.5$ and $b = 0.5$. One obtains (by Monte-Carlo simulations)

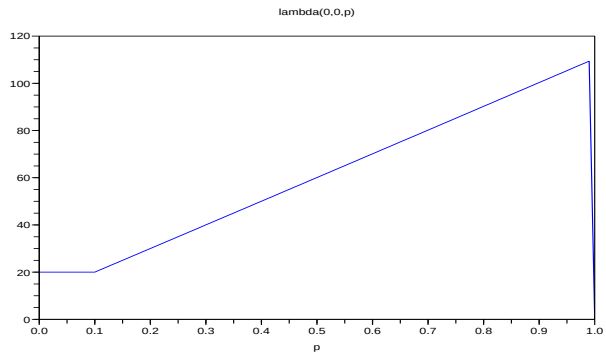

Figure 8.3: $\lambda(0, 0, p)$.

the distributions of the loss L_T and the defaults proportion P_T given by Figure 8.4. One can see that the loss is concentrated around about 0.1, between 0 and about 0.2. This will obviously have an impact on the computation of $\mathbb{E}(L_T - K)^+$ when K becomes large. Indeed, when we perform our SCV algorithm with a (GLT,GLT,GLT) interpolation with (3,3,3) knots, and we compare the result the the "quasi-exact" solution (computed by quite long Monte-Carlo simulations at 101 equidistant points of each of the intervals $[0, T]$, $[0, 1]$ and $[0, 1]$), we obtain a better global approximation with $K = 0$ (Figure 8.5) than with $K = 0.2$ (Figure 8.6).
Our goal is to use SCV to perform an accurate and fast computation of $u(0, 0, 0) = \mathbb{E}(L_T - K)^+$. We take a (GLT,GLT,GLT) interpolation grid with (2,2,2) knots, and we fix the number of simulations at each iteration to $M = 100$. Concerning the stopping criterion, we fix an absolute precision threshold $\varepsilon = 10^{-3}$, and we decide to stop at iteration I when $|u_I(0, 0, 0) - \bar{u}_I(0, 0, 0)| < \varepsilon$, where $\bar{u}_I(0, 0, 0) :=$

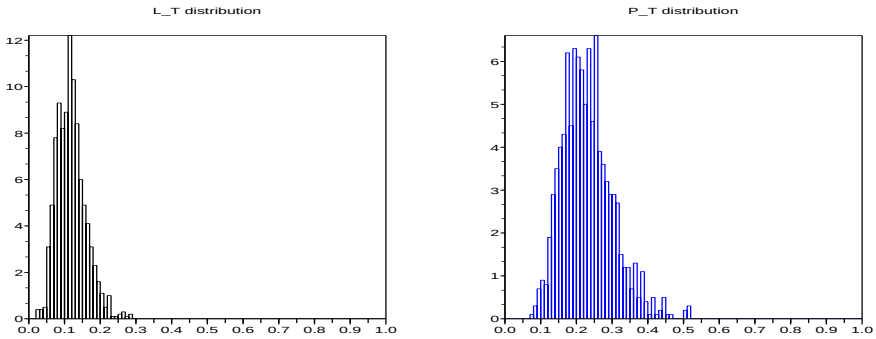

Figure 8.4: Distributions of L_T and P_T, based on 10000 samples.

$\frac{1}{I}\sum_{i=1}^{I} u_i(0,0,0)$ is the average value of the set of computed values between iteration 1 and iteration I. Then, we stop also as soon as the number of iterations reaches 10 (and, whenever this occurs, we take the final $\bar{u}_I(0,0,0)$). We have chosen this convergence criterion because we think it can express the ergodic property of the sequence of the $(u_I)'s$, viewed as a discrete first-order autoregressive (AR(1)) process. The study of this property, as well as the possible related optimal choice of the interpolation grid and of M, are left to a future research. From an empirical point of view, our choice seems to be quite efficient. Indeed, we have compared the final value $u_I(0,0,0)$ we obtain by SCV (using our criteria) with that obtained by 10^8 crude Monte-Carlo simulations (we assume that they lead to an almost exact value, but note that they take some hours!), and the difference between the two values is of order 10^{-3} most of the time.
In Table 8.3, we have reported the numerical results and the CPU times taken by the SCV and the crude Monte-Carlo methods. For the latter, we perform a number M of simulations that is approximately equal to $(\frac{2\sigma[(L_T-K)^+]}{\varepsilon})^2$, in order to approximately reach (with about 95% chance) the precision $\varepsilon = 10^{-3}$. The quantity $\sigma[(L_T-K)^+]$ denotes the standard deviation of $(L_T-K)^+$, and is computed by standard Monte-Carlo simulations.

			Crude MC		SCV	
K	$\mathbb{E}[(L_T-K)^+]$	$\sigma[(L_T-K)^+]$	M	CPU	M	CPU
0%	0.1169	5.10^{-2}	10000	1 sec	100	0.3 sec
3%	0.0869	4.10^{-2}	6400	0.6 sec	100	0.3 sec
6%	0.0573	4.10^{-2}	6400	0.6 sec	100	0.3 sec
9%	0.0316	3.10^{-2}	3600	0.4 sec	100	0.3 sec
12%	0.0142	3.10^{-2}	3600	0.4 sec	100	0.3 sec
20%	0.00068	5.10^{-3}	100	0 sec	100	0.3 sec
22%	0.00029	4.10^{-3}	64	0 sec	100	0.3 sec

Table 8.3: SCV, compared with crude Monte-Carlo, for different values of K.

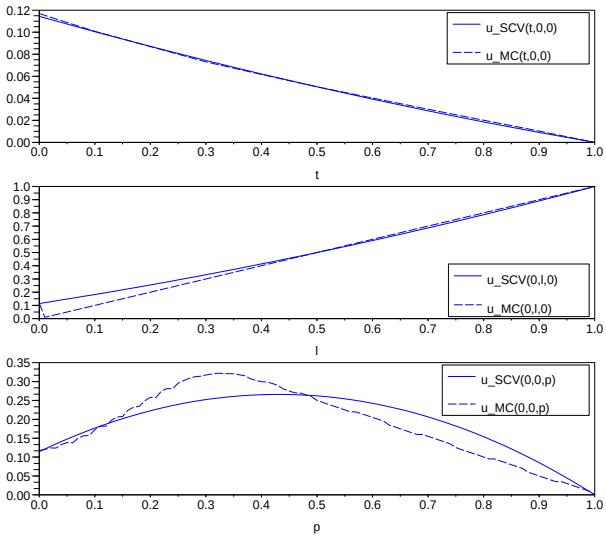

Figure 8.5: $K = 0$: u, obtained by MC (with 10^8 samples) and by SCV with (3,3,3) points of a (GLT,GLT,GLT) grid.

We see again that SCV is more suitable for values of K less than 0.2, otherwise crude Monte-Carlo simulations are faster and more accurate. This is due the fact that the standard deviation $\sigma[(L_T - K)^+]$ decreases when we increase K, and becomes of order 10^{-3} when K reaches 20%.

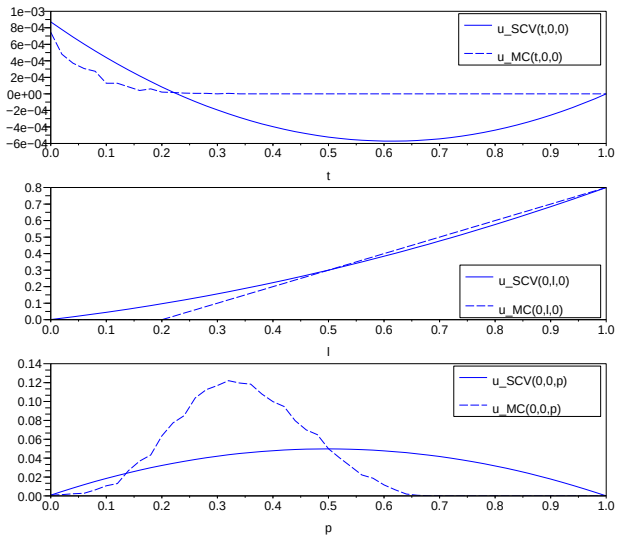

Figure 8.6: $K = 0.2$: u, obtained by MC (with 10^8 samples) and by SCV with (3,3,3) points of a (GLT,GLT,GLT) grid.

Part III

Delta-Gamma hedging

Chapter 9

Delta and Delta-Gamma hedging errors

Most of the models in finance are continuous in time, in contrast with the real practice of the markets where the transactions and the computations are carried out discretely in time. In particular, this means that one cannot make a perfect replication of a financial derivative security by a self-financing strategy, and might care about the related hedging error. Let us make a review of the literature devoted to the study of two common hedging strategies: the Delta hedging strategy (DHS for short) and the Delta-Gamma hedging strategy (DGHS for short).

9.1 Delta hedging

9.1.1 The L_2-tracking error

In the Black-Scholes model, a continuous strategy to perfectly replicate a payoff $g(S_T)$ (S is a risky asset) is well known: it consists in holding at time t a number δ_t of assets S given by $\delta_t = \partial_S u(t, S_t)$, where u is the price function (see Black and Scholes [BS73]). This is the so-called Delta hedging strategy. In the corresponding discrete strategy, one holds a fixed number δ_{t_i} of assets between two rebalancing times t_i and t_{i+1}, for $0 \leq i \leq N-1$, where N is the number of the time intervals. At the maturity time T, the terminal value of such a portfolio, denoted by V_T^N, may differ from the payoff to hedge $g(S_T)$, inducing a *tracking error* $\mathcal{E}_N^\Delta := V_T^N - g(S_T)$. In [Zha99], Zhang studied the order of convergence of the quadratic residual risk $\bar{R}_N^\Delta := \mathbb{E}|\bar{\mathcal{E}}_N^\Delta|^2$ (under both the historical probability and the risk-neutral one, we omit here the dependence of the expectation $\mathbb{E}$ on the measure), where

$$\bar{\mathcal{E}}_N^\Delta := e^{-rT}\mathcal{E}_N^\Delta = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (\delta_{t_i} - \delta_t)\mathrm{d}\bar{S}_t$$

is the discounted tracking error, r is the (constant) interest rate and $\bar{S}_t := e^{-rT}S_t$ (which is a martingale under the risk-neutral probability). We only present the

results in the one-dimensional framework (those of the multidimensional case are similar but stated under more restrictive assumptions, see chapter 3 in [Zha99]). The model that was considered for the non-risky asset X^0 and the risky asset X was the following, under the risk-neutral probability $\mathbb{Q}$:

$$\begin{cases} \frac{dX_t^0}{X_t^0} & = r\mathrm{d}t, \\ \frac{dX_t}{X_t} & = r\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t. \end{cases}$$

We do not detail the regularity assumptions assumed for the volatility coefficient $\sigma(.,.)$. The payoff function g is supposed to be differentiable with polynomial growth, as well as its derivative:

$$|g(x)| + |g'(x)| \leq L(1 + |x|^\gamma),$$

where L and γ are positive constants.
The time grid considered is the equidistant one: $t_i = \frac{T}{N}i, 0 \leq i \leq N$.
The following result (in [Zha99], it corresponds to Theorem 2.1.2 under the risk-neutral probability and to Theorem 2.4.1 under the historical one) states that the residual risk $\bar{R}_N^\Delta$ is of order of the time step $\frac{T}{N}$, when the number N of the rebalancing dates goes to infinity.

Theorem 9.1.1. *Under the above assumptions, one has*

$$\lim_{N\to\infty} N\bar{R}_N^\Delta = \frac{T}{2}\mathbb{E}\left(\int_0^T e^{-2rt}X_t^4\sigma^4(t, X_t)\left(\frac{\partial^2 u}{\partial x^2}(t, X_t)\right)^2 \mathrm{d}t\right).$$

Remark 9.1.1.

(i) *Under extra assumptions, the author moreover shows that the difference between $N\bar{R}_N^\Delta$ and the limit above is of order of $\mathcal{O}(N^{-\frac{1}{2}})$.*

(ii) *The author compared the quadratic residual risk $\bar{R}_N^\Delta$ of the Delta hedging (or what he called "the elementary approximative hedging") with that of what he called "the optimal approximative hedging", which consists in following the best time piecewise-constant strategy (δ_t^{opt}) ($\delta_t^{opt} = \delta_{\frac{T}{N}i}^{opt}$, for $t \in [\frac{T}{N}i, \frac{T}{N}(i+1))$) that minimizes the corresponding quadratic residual risk $\bar{R}_N^{\Delta,opt} := \mathbb{E}|\bar{\mathcal{E}}_N^{\Delta,opt}|^2$, the number N of rebalancing times being fixed (see Proposition 1.4.5 in [Zha99], or Martini and Patry [MP99]). It is stated (see Theorem 2.1.4 in [Zha99]) that $|\bar{R}_N^\Delta - \bar{R}_N^{\Delta,opt}| \leq \frac{C}{N^2}$, which means (taking into account the above theorem) that both strategies are asymptotically equally efficient.*

Then, Gobet and Temam [GT01] focused on some examples of more irregular payoffs in order to show that the order of convergence strongly depends on the regularity properties of the payoff function. Considering always uniform rebalancing dates $(\frac{T}{N}i)_{i=0}^N$, they stated (see Theorems 1 and 2 in [GT01])

Theorem 9.1.2.

(i) *For $g(x) = \mathbb{1}_{x \geq K}$, one has*

$$\bar{R}_N^{\Delta} = \left(\frac{T}{N}\right)^{\frac{1}{2}} \frac{C_0}{4\sqrt{\pi}} K\sigma^3(K)e^{-2rT} p_T(x_0, K) + o\left(\frac{1}{N^{\frac{1}{2}}}\right),$$

where C_0 is an universal constant and $p_t(.,.)$ is the transition density function of the process $(X_t)_t$.

(ii) *For $g(x) = (x-K)_+^a$, $a \in (0, \frac{1}{2})$, one has*

$$\bar{R}_N^{\Delta} = \left(\frac{T}{N}\right)^{\frac{1}{2}+a} \hat{C}_a K^{1+2a}\sigma^{3+2a}(K)e^{-2rT} p_T(x_0, K) + o\left(\frac{1}{N^{\frac{1}{2}+a}}\right),$$

where $\hat{C}_a$ is an universal constant.

Remark 9.1.2. *The above theorem was stated in [GT01] for a one-dimensional time-homogeneous diffusion model. It has been extended to the multidimensional model by Temam[Tem03] using Malliavin calculus techniques.*

The role played by the (fractional) regularity of the payoff function g has been then more systematically investigated by Geiss *et al.* , but mainly within the Brownian motion and the geometric Brownian motion model for the diffusion X (see e.g. [Gei02], [GG04] and [GH07]). The case of one-dimensional SDE has been studied in [GG04]. The authors have showed moreover that one can adapt the concentration of the time grid knots according to the fractional regularity of g in order to obtain a convergence rate for $\bar{R}_N^{\Delta}$ of N. For the details of Geiss *et al.* results, see Part I, Chapter 4.
This study has been recently extended by Gobet and Makhlouf [GM08] to the general multidimensional diffusion model (and also to BSDEs with non null generator). In this paper, the fractional regularity of the payoff function g is measured by the rate of decreasing of the expected conditional variance of $g(X_T)$ given $\mathcal{F}_t$ (denoted by $V_{t,T}(g) := \mathbb{E}\left|g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T))\right|^2$) as t goes to T. The corresponding space is defined by (for some $\alpha \in (0,1]$)

$$\mathbf{L}_{2,\alpha} := \{g \text{ s.t. } \mathbb{E}|g(X_T)|^2 + \sup_{0 \leq t < T} \frac{V_{t,T}(g)}{(T-t)^{\alpha}} < +\infty\},$$

and a related squared-norm is defined by

$$K^{\alpha}(g) := \mathbb{E}|g(X_T)|^2 + \sup_{t \in [0,T)} \frac{V_{t,T}(g)}{(T-t)^{\alpha}}.$$

The authors stated the following result (see Theorem 8 and Corollary 10 in [GM08]), under ellipticity and smoothness assumptions on the SDE coefficients of X that we do not detail. Notice the effect of the time net choice on the convergence rate.

Here, z_s corresponds to $\delta_s\sigma_s$. The error related to the discrete time approximation of z is defined by

$$\mathcal{E}(z,\pi) := \sum_{i=0}^{N-1} \mathbb{E}\int_{t_i}^{t_{i+1}} |z_t - \bar{z}_{t_i}|^2 dt, \tag{9.1}$$

where

$$\bar{z}_{t_i} := \frac{1}{t_{i+1}-t_i}\mathbb{E}^{\mathcal{F}_{t_i}}\int_{t_i}^{t_{i+1}} z_s ds \tag{9.2}$$

is the projection of $(z_s)_{t_i\le s\le t_{i+1}}$ on the space of $\mathcal{F}_{t_i}$-measurable random variables, according to the scalar product $< u,v >= \mathbb{E}\int_{t_i}^{t_{i+1}} u_s v_s ds$. The result still holds for the error $\mathbb{E}|\bar{\mathcal{E}}_N^{\Delta}|^2$ instead of $\mathcal{E}(z,\pi)$, under mild assumptions on σ.

Theorem 9.1.3. *Assume that $g \in \mathbf{L}_{2,\alpha}$, for some $\alpha \in (0,1]$. Then,*

(i) *with the choice of the equidistant time net $\pi^{(1)} := (\frac{T}{N}i)_{i=0}^N$, one has*

$$\mathcal{E}(z,\pi) \le CK^{\alpha}(g)\left(\frac{T}{N}\right)^{\alpha}$$

(where C does not depend on N).

(ii) *with the choice of the non-equidistant time net $\pi^{(\beta)} := (t_i^{(N,\beta)})_{i=0}^N$ defined by $t_i^{(N,\beta)} = T - T\big(1-\frac{k}{N}\big)^{\frac{1}{\beta}}, 0 \le i \le N$, and taking $\beta = 1$ if $\alpha = 1$ and $\beta < \alpha$ otherwise, one has*

$$\mathcal{E}(z,\pi) \le CK^{\alpha}(g)\frac{T^{\alpha}}{N}$$

(where C does not depend on N).

9.1.2 The weak convergence

In all the above results, the convergence criterion that has been considered is in the $\mathbf{L}_2$ sense. Instead, one may investigate the weak convergence rate of the (discounted) tracking error $\bar{\mathcal{E}}_N^{\Delta}$.
One already knows that $\bar{\mathcal{E}}_N^{\Delta}$ converges in probability to 0 as N goes to infinity (see [RY94], Proposition IV-2.13). Surprisingly, as pointed out in [GT01], this weak convergence occurs at rate $N^{\frac{1}{2}}$ even with irregular payoffs (the rebalancing dates are assumed to be equidistant):

Theorem 9.1.4 (Theorem 3 in [GT01])**.** *Assume that X solves the driftless and time-homogeneous Black-Scholes equation $dX_t = \sigma X_t \mathrm{d}W_t$. Then*

$$\sqrt{N}\bar{\mathcal{E}}_N^{\Delta} \to_d \tilde{W}_{\tau}, \quad N \to \infty,$$

where $\tau := \frac{1}{2}\int_0^T \left(\frac{\partial^2 u}{\partial x^2}(t,X_t)\right)^2 \sigma^4 X_t^4 \mathrm{d}t$, and $\tilde{W}$ is an extra Brownian motion independent of τ.

Remark 9.1.3.

(i) *The above theorem is derived from a general convergence result for the approximation of stochastic integrals, already given by Rootzen[Roo80]. Actually, it is related to the convergence of conditional Gaussian martingales, as definitely studied by Jacod [Jac97].*

(ii) *The random variable $\tilde{W}_\tau$ may not be square-integrable even with irregular payoffs. As stated in [GT01], it is possible to check that $\mathbb{E}(\tau) = \infty$ in the cases $g(x) = \mathbb{1}_{x \geq K}$ and $g(x) = (x-K)_+^a$ ($a \in (0, \frac{1}{2})$). Thus, these are nontrivial situations where convergence in law does not necessarily lead to convergence in $\mathbf{L}_2$ sense. The choice of the convergence criterion is then significant.*

(iii) *The above result has been extended by Hayashi and Mykland [HM05] to more general diffusions (namely positive Itô processes $\mathrm{d}X_t = b_t \mathrm{d}t + \sigma_t \mathrm{d}W_t$, where b_t and σ_t are adapted, suitably integrable processes, and $\sigma_t > 0$ for all t) and to more general payoffs $C_T \geq 0$ (and $\mathrm{d}C_t = \theta_t \mathrm{d}X_t$, where (C_t) is the price process of the (European) derivative to hedge, and (θ_t) is a self-financing strategy which is an adapted and suitably integrable process).*

(iv) *The asymptotic distribution of the tracking error when the rebalancing dates are non-equidistant is studied by Liang [Lia97] in the Black-Scholes framework.*

(v) *Geiss and Toivola [GT08] investigated the connection between the fractional smoothness of the payoff function and the weak convergence, for X being either the Brownian motion or the geometric Brownian motion. They also studied the $\mathbf{L}_p$-integrability of the limit ($p \geq 2$), and showed how a proper choice of the time net can ensure this integrability (contrary to the choice of the uniform grid as mentioned in (ii) above).*

9.1.3 Transaction costs

The transaction costs generally have an impact on the tracking error, and also urge the hedger to trade as least frequently as possible. Then, the influence of these costs have been investigated in some studies dealing with discrete-time hedging.
In the Black-Scholes context of the replication of a European Call, Leland [Lel85] was the first to suggest a hedging portfolio based on modified Black-Scholes Greeks (we omit the details). The cost of a single transaction is assumed to be proportional to its trading volume, with coefficient k_N (N is still the number of the rebalancing intervals). Denoting by $\hat{\delta}_t$ the number of shares of the stock (whose price process is S) held at time t, the portfolio value process is then given by

$$\hat{V}_t^N = \hat{V}_0^N + \int_0^t \hat{\delta}_u \mathrm{d}S_u - k_N \sum_{t_i \leq t} S_{t_i} |\hat{\delta}_{t_i} - \hat{\delta}_{t_{i-1}}|.$$

The new initial wealth $\hat{V}_0^N$ and the new deltas are usually computed using a Black-Scholes model with larger volatility (depending on k_N).

Kabanov and Safarian [KS97] showed that, depending on the transaction level k_N, Leland's strategy may lead to a hedging error that is asymptotically (when the rebalancing frequency goes to infinity) not equal to zero. Indeed, one can make a perfect asymptotic replication when $k_N = k_0 N^{-\gamma}$ (with $\gamma \in (0, \frac{1}{2}]$ and k_0 is a positive constant), but when $k_N = k_0 > 0$ is a constant, the option is always underpriced (contrary to Leland's conjecture in [Lel85]). Here is the result from [KS97] (Theorems 1 and 2). H denotes the option payoff with maturity $T = 1$ ($H = (S_1 - K)_+$).

Theorem 9.1.5.

(i) *If $k_N = k_0 N^{-\gamma}$, with $\gamma \in (0, \frac{1}{2}]$ and $k_0 > 0$,*

$$\hat{V}_1^N \to_d H, \quad N \to \infty.$$

(ii) *If $k_N = k_0 > 0$ is a constant,*

$$\hat{V}_1^N \to_d H - J, \quad N \to \infty,$$

with $0 \leq J \leq C k_0$ and C a constant depending on S_1 and K.

Remark 9.1.4.

(i) *In [Per03], Pergamenshchikov showed that the rate of convergence in Kabanov-Safarian theorem is $N^{\frac{1}{4}}$, and that the asymptotic distribution of $\hat{V}_1^N$ is a mixture of Gaussian distributions.*

(ii) *In [Den08], Denis has studied Leland's startegy with more general payoffs.*

In [Zha99], the author studied the expected transactions cost, C_N, according to the number N of rebalancing intervals. Considering more general European options with regular payoff functions, and assuming that the transaction level k_N is a constant, it is stated that $C_N \sim \sqrt{N}$ (see Proposition 2.5.1 in [Zha99]).

9.2 Delta-Gamma hedging

As one can notice through the previous results on the asymptotic Delta hedging error, the limit (either in the $\mathbf{L}_2$ sense or in distribution) involves a suitable integral of the second spatial derivative $\frac{\partial^2 u}{\partial x^2}(t, X_t)$ (called *Gamma*) of the option (see Theorems 9.1.1 and 9.1.4). Thus, while the DHS reduces the *portfolio Delta* (in such a way that, in the Black-Scholes model, the tracking error $\bar{\mathcal{E}}_N^{\Delta}$ can be decomposed as $\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} T^{i,(1)}(t) \mathrm{d}W_t$ with $T^{i,(1)}(t_i) = 0$ for $0 \leq i \leq N$, see e.g. the proof of Proposition 1.4.2), one may think of reducing in addition the *portfolio Gamma* (so that the related tracking error $\bar{\mathcal{E}}_N^{\Delta\Gamma}$ can hopefully be decomposed as $\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \int_{t_i}^{t} T^{i,(2)}(t) \mathrm{d}W_s \mathrm{d}W_t$ with $T^{i,(2)}(t_i) = 0$ for $0 \leq i \leq N$). This additional condition obviously requires to introduce another hedging instrument in the hedging

portfolio in addition to the underlying asset S. This strategy is the so-called Delta-Gamma hedging strategy (see [Hul09], p.371).
In practice, the extra instrument used is either a Call or a Put, with a time maturity T_2 higher than that of the option to hedge (T). The practioners have used DGHS and noticed that it was more efficient in general than DHS (if trading in the same number N of rebalancing times), but no rigorous mathematical proof in this direction has been provided, so far as we are aware.
However, we point out that Brodén and Wiktorsson [BW08] have investigated the DGHS of a European Call whose price process is $C_1(t, X_t)$, with a payoff equal to $(X_{T_1} - K_1)_+$, using a one-dimensional asset X with local volatility ($\mathrm{d}X_t = rX_t\mathrm{d}t + \sigma(X_t)X_t\mathrm{d}W_t$, under the risk-neutral probability $\mathbb{Q}$) and a European Call whose price process is $C_2(t, X_t)$, with a payoff equal to $(X_{T_2} - K_2)_+$, where $T_2 > T$. Following equidistant rebalancing dates ($t_i = \frac{T}{N}i, 0 \leq i \leq N$), the hedging porfolio consists of δ_{t_i} assets X and $\delta_{t_i}^{C_2}$ Calls C_2, to hold on each time interval $[t_i, t_{i+1})$, with

$$\delta_{t_i}^{C_2} := \frac{\partial_X^2 C_1(t_i, X_{t_i})}{\partial_X^2 C_2(t_i, X_{t_i})}, \qquad \delta_{t_i} := \partial_X C_1(t_i, X_{t_i}) - \frac{\partial_X^2 C_1(t_i, X_{t_i})}{\partial_X^2 C_2(t_i, X_{t_i})}\partial_X C_2(t_i, X_{t_i}).$$

Denoting the induced (discounted) tracking error by $\bar{\mathcal{E}}_N^{\Delta\Gamma}$, the main result of Brodén and Wiktorsson states that the convergence of the quadratic residual risk $\mathbb{E}|\bar{\mathcal{E}}_N^{\Delta\Gamma}|^2$ occurs at rate $N^{\frac{3}{2}}$. Notice however that their proofs present gaps when deriving some crucial estimates, namely those of the ratios $\frac{\partial_X^2 C_2(t,X_t)}{\partial_X^2 C_2(t_i,X_{t_i})}$ and $\frac{\partial_X^3 C_2(t,X_t)}{\partial_X^2 C_2(t_i,X_{t_i})}$. Here is the result (we do not give details of the required regularity and ellipticity assumptions on the diffusion coefficient $\sigma(.)$)

Theorem 9.2.1 (Proposition 2.1 in [BW08])**.**

$$\mathbb{E}|\bar{\mathcal{E}}_N^{\Delta\Gamma}|^2 = C\left(\frac{T_1}{N}\right)^{\frac{3}{2}} \lim_{t\to T_1} \varphi(t) + o\left(\frac{1}{N}\right)^{\frac{3}{2}},$$

where C is a universal constant (≈ 0.65), and

$$\varphi(t) := (T_1 - t)^{\frac{3}{2}}\mathbb{E}\left(e^{-2rt}\left(\partial_X^3 C_1(t, X_t)\right)^2 X_t^6\sigma^6(X_t)\right).$$

Remark 9.2.1.

(i) *One notices that, in order to investigate the tracking error order of DGHS, one needs estimates for the first three spatial derivatives of the price function to hedge, while only estimates of the first two spatial derivatives are required with DHS.*

(ii) *Compared with the (optimal) rate N obtained by Zhang [Zha99] for DHS, the above result shows the asymptotic efficiency of DGHS in comparison to DHS.*

9.3 Our contribution

We consider a mutidimensional Black-Scholes model for $S = (S^1, \ldots, S^d)$, with constant volatilities (log-normal model) and constant correlations. The hedging instruments that we use are Calls and exchange options (whose price functions are, within our model, explicitely given by the Black-Scholes and Margrabe formulas).
We write the Delta-Gamma tracking error $\bar{\mathcal{E}}_N^{\Delta\Gamma}$ as a (multiple) stochastic integral, involving the derivatives of the option price function up to order 3. Then, establishing estimates on the option Greeks and on ratios of the hedging instruments Greeks, we study the $\mathbf{L}_2$-convergence order of $\bar{\mathcal{E}}_N^{\Delta\Gamma}$, according to the number of rebalancing dates and to the regularity degree α of the payoff $g(S_T)$. Besides, we show that one can choose the trading dates so as to achieve an optimal convergence rate (of order the number of dates N).

Chapter 10

The tracking error rate of the Delta-Gamma hedging strategy

This chapter corresponds to an article submitted for publication in "Mathematical Finance".

10.1 Abstract

We analyse the convergence rate of the quadratic tracking error, when a Delta-Gamma hedging strategy is used at N discrete times. The fractional regularity of the payoff function plays a crucial role in the choice of the trading dates, in order to achieve optimal rates of convergence.

10.2 Statement of the problem

An investor who sells an option should hedge against the variations of the underlying asset price, either by making a static replication, or by making a dynamic hedging strategy, i.e. trading continuously in time in the hedging instruments. In practice, this is often done through the so-called delta hedging strategy (DHS in short), which ensures that the investor's portfolio remains delta-neutral (i.e. equalizing the first sensitivity δ of the hedging portfolio and that of the option value). When the market is complete (we assume this in all what follows), in a Markovian setting where the asset is modeled through a stochastic differential equation and the payoff at time T is of the form $g(S_T)$, it is known (cf Karatzas and Shreve [KS98]) that the number of assets to hold at time t is given by $\delta_t = \partial_S u(t, S_t)$ where u is the price function (solution of a valuation PDE). Theoretically, it means that the investor must trade, continuously at each time t between 0 and the maturity, in order to obtain zero residual risk.
However, due to practical considerations and to the transaction costs, one can only use discrete-time hedging strategies. Here we do not consider the impact of transaction costs (we refer to Kabanov and Safarian [KS97], Pergamenshchikov [Per03]

and references therein) and we focus only on the impact of discrete-time rebalancing. Thus, the number of assets between two successive rebalancing dates t_i and t_{i+1} is fixed to δ_{t_i}, and the associate hedging portfolio has a terminal value V_T^N which may differ from the required payoff $g(S_T)$. In other words, this induces a tracking error $\mathcal{E}_N^\Delta := V_T^N - g(S_T)$, which has been intensively studied in the literature. For Call/Put option, in Zhang [Zha99] it is shown that the $\mathbf{L}_2$ error has a convergence rate $N^{1/2}$ for uniform rebalancing dates $t_i = iT/N$. In Gobet and Temam [GT01], it is shown that the rate of convergence actually strongly depends on the smoothness of the payoff function g: for instance for digital options, the order of convergence becomes 1/4 instead of 1/2. Geiss and coauthors deeply investigated these features: in a series of papers initiated by Geiss [Gei02], they paved the way to connect the fractional regularity of the payoff and the rate of convergence of the related discrete time delta-hedging strategy. Moreover, they showed that for a given payoff, a suitable non-uniform grid with N dates can be chosen to achieve the rate $N^{1/2}$: the more irregular the payoff, the more concentrated near T the points. Recently in Gobet and Makhlouf [GM08], these ideas have been extended to BSDEs and to multidimensional diffusion models. One could alternatively measure the error using weak convergence techniques, instead of $\mathbf{L}_2$ norm. Surprisingly, the convergence rate may be $N^{1/2}$ even for non-smooth payoffs, i.e. $N^{1/2}\mathcal{E}_N^\Delta$ weakly converges as N goes to infinity to a non trivial random variable, which may not be square-integrable[1] for non-smooth payoffs. This convergence phenomenon has been noticed in Gobet and Temam [GT01] for digital options, then more systematically studied in Hayashi and Mykland [HM05]. The integrability of the weak limit for non-smooth payoffs is investigated in details in Geiss and Toivola [GT08].

In all these works, it is pointed out that the weak limit is related to a suitably weighted integral of the so-called Gamma $(\partial_S^2 u(t, S_t))_{0\le t<T}$ of the option[2]. Thus, in order to reduce the tracking error, one can think of reducing the *portfolio Gamma* by incorporating in the hedging portfolio another instrument, the price of which is $(C(t, S_t))_{0\le t\le T}$: in practice, it is a liquid vanilla option (At-The-Money Call/Put). The resulting strategy is called Delta-Gamma hedging strategy (DGHS in short) and is well-known from practitioners (see Hull [Hul09] p.371). It is obtained by equalizing the second sensitivity of the hedging portfolio and that of the option value. It follows that the numbers of options C and assets to hold at time t_i are equal to

$$\delta_{t_i}^C := \frac{\partial_S^2 u(t_i, S_{t_i})}{\partial_S^2 C(t_i, S_{t_i})}, \qquad \delta_{t_i} := \partial_S u(t_i, S_{t_i}) - \frac{\partial_S^2 u(t_i, S_{t_i})}{\partial_S^2 C(t_i, S_{t_i})}\partial_S C(t_i, S_{t_i}). \tag{10.1}$$

This specification of piecewise-constant strategy on each interval $[t_i, t_{i+1}[$ leads to a final wealth process V_T^N hopefully closer to $g(S_T)$ compared with that of the simple DHS.

[1]This gives a non trivial situation where the rate of convergence depends on the convergence criterion ($\mathbf{L}_2$ convergence or weak convergence).

[2]Similar integrals of the Gamma also appear in the measure of robustness of the Black and Scholes formula, cf El Karoui, Jeanblanc-Picqué and Shreve [KJS98].

10.3 Objectives

The purpose of this work is to quantify how much the resulting tracking error $\mathcal{E}_N^{\Delta\Gamma} := V_T^N - g(S_T)$ is reduced with respect to the number of rebalancing dates. More precisely, we aim at connecting the convergence rate of $\mathcal{E}_N^{\Delta\Gamma}$ to the payoff regularity, regarding the choice of rebalancing dates. These issues can be investigated under several points of view.

- Choice of the convergence criterion. We conjecture that in the usual cases, the rate of weak convergence of $\mathcal{E}_N^{\Delta}$ is N (instead of $N^{1/2}$), extending the results in Hayashi and Mykland [HM05], with a possible non square-integrable limit (our numerical experiments support this phenomenon). We leave the proof of this conjecture to further research and in this work, we prefer to investigate the $\mathbf{L}_2$ convergence because it leads to a more intuitive dispersion measure of the tracking error. Our results below state that equidistant trading dates with DGHS do not systematically yield a better convergence order w.r.t. N compared with DHS (see Theorem 10.6.2). But an appropriate choice of trading dates leads to a quadratic error equal to a $O(N^{-1})$. The choice of the trading dates is explicit and depends on the fractional regularity of the payoff.

- Specification of the model for S. In this work, we emphasize the tuning of the trading dates according to the payoff regularity. Analogously to the results in Geiss [Gei02] or Gobet and Makhlouf [GM08], we believe that the results on the rate of convergence (and thus the choice of trading dates) hold for a wide class of SDE model on S. To simplify the analysis, we only consider geometric Brownian motions for the modeling of S (see (10.2)). Nevertheless, the analysis remains rather intricate, but we hope that this model simplification will help the reader to focus on the measure of the payoff regularity, which is driven by the behavior of the expected conditional variance $V_{t,T}(g) = \mathbb{E}|g(S_T) - \mathbb{E}^{\mathcal{F}_t}(g(S_T))|^2$ as t goes to T. On the other hand, we allow the payoff to depend on several assets (basket options for instance). For this, we extend the single-asset formula (10.1) for DGHS to the multi-asset case, involving the *cross-Gammas* of the options.

- Choice of hedging instruments. Obviously, the instruments C used for DGHS have to yield non-zero gammas, in order to well define the strategy (see (10.1) for the single-asset case and Definition 10.6.1 for the multi-asset case). To manage the individual gammas, the most natural choice is Call/Put options whose price are convex in the spot variable (see Martini [Mar99] for general results on the propagation of convexity,). To handle cross gammas, we propose to use exchange options (spread option with strike zero). Since the models for S are correlated geometric Brownian motions, whenever needed we can take advantage of the closed Margrabe formula for such an exchange option (see Appendix 10.A.1).

The closest related work to ours is the one by Brodén and Wiktorsson [BW08]. In this reference, S is a one-dimensional asset, following a local volatility model. The payoff function g is the Call payoff and uniform trading dates are studied. They state that the $\mathbf{L}_2$ convergence holds at rate $N^{3/4}$. However, there are some gaps in their proofs, in particular regarding some key estimates related to lower/upper bounds for the second derivatives of the price function. The differences between their work and ours are the following: we restrict to log-normal modelling; we extend the study of DGHS to a multidimensional framework; we study the connection between the payoff regularity and the choice of the trading dates in order to achieve optimal convergence rates. Furthermore, our proofs leverage the intrinsic martingale properties of price processes (see Lemma 10.5.2), which leads to simplified computations.

10.4 Organization of the paper

In the next section, we define the stochastic model for the underlying assets, the payoff option to hegde and the other instruments used in the hedging portfolio. We state some preliminary results related to the Greeks' processes as martingales. We also introduce notations used in all the paper, in particular for the fractional regularity of the payoff function. The section 10.6 contains our main results: we define the DGHS strategy, show that the $\mathbf{L}_2$ norm of the tracking error is essentially measured by a suitable integral of the third derivatives of the price function, which can be accurately estimated through the payoff regularity. Then, we give sufficient conditions on the N trading dates to yield a tracking error of order $1/N$ as the number N of dates goes to infinity. In section 10.7, we present numerical results which corroborate these features. Section 10.8 is devoted to the proofs of the main results.

10.5 Preliminaries

10.5.1 Asset stochastic model, payoff, additional hedging instruments

Hereafter, $\hat{W} = (\hat{W}^1, ..., \hat{W}^d)$ is a d-dimensional Brownian motion, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is the historical probability and $(\mathcal{F}_t)_{0\leq t\leq T}$ (T is a fixed terminal time) is the natural filtration of $\hat{W}$, augmented with $\mathbb{P}$-null sets. We assume that the components of $\hat{W}$ are correlated, with $\langle \hat{W}^j, \hat{W}^k\rangle_t = \rho_{j,k}t$, and that the matrix $(\rho_{j,k})_{1\leq j,k\leq d}$ has a full rank.

Assets. We denote the non-risky asset by S^0 (bank account) and the risky assets by $(S^j)_{1\leq j\leq d}$. Although not essential for our results, we assume that the risky assets do not pay dividends and that the interest rates are constant and equal to r. Morever, we assume that the risky assets $(S^j)_{1\leq j\leq d}$ are lognormal processes with constant volatilities $(\sigma_j)_{1\leq j\leq d}$, constant historical drifts $(\mu_j)_{1\leq j\leq d}$ and correlations equal to

$(\rho_{j,k})_{j,k}$. It writes for $1 \leq j \leq d$:

$$\begin{cases} S_0^j &= s_0^j, \\ \mathrm{d}S_t^j &= \mu_j S_t^j \mathrm{d}t + \sigma_j S_t^j \mathrm{d}\hat{W}_t^j, \end{cases} \tag{10.2}$$

where S_t^j, s_0^j and σ_j are positive. Set $\lambda_j = \frac{\mu_j - r}{\sigma_j}$: then, we can define the so-called risk-neutral measure $\mathbb{Q}$ such that $(W_t^j := \hat{W}_t^j + \lambda_j t)_{1 \leq j \leq d}$ is a $\mathbb{Q}$-Brownian motion (the correlations remaining unchanged). Up to the correlation factor, λ is the market price of risk. Under $\mathbb{Q}$, the model writes

$$\begin{cases} S_0^j &= s_0^j, \\ \mathrm{d}S_t^j &= r S_t^j \mathrm{d}t + \sigma_j S_t^j \mathrm{d}W_t^j. \end{cases}$$

By a slight abuse of notation by setting $\sigma_0 := 0$ and $S_0^0 := 1$, we make the above equation also valid for $j = 0$, which might be useful in the following when short notations are needed.

At least, note that $\bar{S}_t^j := e^{-rt} S_t^j$ is a $\mathbb{Q}$-martingale (for any $0 \leq j \leq d$), and $\mathrm{d}\bar{S}_t^j = \sigma_j \bar{S}_t^j \mathrm{d}W_t^j$. More generally, if U is a stochastic process, $\bar{U}$ denotes its discounted value:

$$\bar{U}_t = e^{-rt} U_t.$$

When an expectation is computed under $\mathbb{P}$, we write $\mathbb{E}_{\mathbb{P}}(.)$, while under $\mathbb{Q}$ we write $\mathbb{E}_{\mathbb{Q}}(.)$. We add the superscript $\mathcal{F}_t$ to indicate the conditional expectations given $\mathcal{F}_t$, i.e. $\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_t}(.)$ and $\mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t}(.)$.

The payoff g. In what follows, g denotes the payoff function that defines the option to hedge. The price function of this option is then defined by $u(t, S) := \mathbb{E}_{\mathbb{Q}}\left[e^{-r(T-t)} g(S_T) | S_t = S\right]$, for $S = (S^1, ..., S^d) \in \mathbb{R}_+^d$. We assume that $\mathbb{E}_{\mathbb{P}} |g(S_T)|^{2p_0} < \infty$ for some $p_0 > 1$ (for instance, $|g(S)| \leq C(1 + |S|^n + |S|^{-n})$ for some $n \in \mathbb{N}$). Without additional regularity assumption on g, it is easy to check that u is a C^∞ function for $t < T$. In particular, for $l, m, n = 1...d$, $\partial_t u$, $\partial_{S^l} u$, $\partial^2_{S^l, S^m} u$ and $\partial^3_{S^l, S^m, S^n} u$ exist and are continuous for $t < T$.

As in Geiss [Gei02] and Gobet and Makhlouf [GM08], the fractional regularity of the payoff function g is measured through

$$V_{t,T}(g) := \mathbb{E}\left|g(X_T) - \mathbb{E}^{\mathcal{F}_t}(g(X_T))\right|^2,$$

and we consider the following space ($\alpha \in (0, 1]$)

$$\mathbf{L}_{2,\alpha} = \{g \text{ s.t. } \mathbb{E}|g(X_T)|^2 + \sup_{0 \leq t < T} \frac{V_{t,T}(g)}{(T-t)^\alpha} < +\infty\}.$$

It describes the rate of decreasing of the expected conditional variance of $g(S_T)$ given $\mathcal{F}_t$ as t goes to T. When g belongs to $\mathbf{L}_{2,\alpha}$ and $g(S_T) \in \mathbf{L}_{2p_0}$ (for some $p_0 > 1$), we

can define $K^\alpha(g)$ and $K^{\alpha,p_0}(g)$ as

$$K^\alpha(g) := \mathbb{E}_{\mathbb{P}}|g(S_T)|^2 + \sup_{t\in[0,T)} \frac{V_{t,T}(g)}{(T-t)^\alpha},$$
$$K^{\alpha,p_0}(g) := \left(\mathbb{E}_{\mathbb{P}}\,|g(S_T)|^{2p_0}\right)^{\frac{1}{p_0}} + \sup_{t\in[0,T)} \frac{V_{t,T}(g)}{(T-t)^\alpha}. \tag{10.3}$$

Notice that $\bigcup_{\alpha\in(0,1]} \mathbf{L}_{2,\alpha}$ obviously contains uniformly Hölder continuous functions, but also some non-smooth functions. Some examples are given in the lemma below, whose proof is done in Appendix 10.A.2.

Lemma 10.5.1.

- Call/Put options. *If $g(S) = (S^j - K)_+$ or $g(S) = (K - S^j)_+$ (for $1 \le j \le d$), then $g \in \mathbf{L}_{2,\alpha}$ with $\alpha = 1$.*
- Digital options. *If $g(S) = \mathbb{1}_{S^j > K}$ or $g(S) = \mathbb{1}_{S^j < K}$ (for $1 \le j \le d$), then $g \in \mathbf{L}_{2,\alpha}$ with $\alpha = 1/2$.*
- Stability by summation and product. *Let g_1 and g_2 belong resp. to $\mathbf{L}_{2,\alpha_1}$ and $\mathbf{L}_{2,\alpha_2}$. Then*
  - *$g_1 + g_2 \in \mathbf{L}_{2,\alpha}$ with $\alpha = \alpha_1 \wedge \alpha_2$;*
  - *If in addition g_1 and g_2 are bounded, then $g_1 g_2 \in \mathbf{L}_{2,\alpha}$ with $\alpha = \alpha_1 \wedge \alpha_2$.*

A straightforward application of the above rules shows for instance that Put payoffs with digital triggers (of the form $g(S) = \mathbb{1}_{S^1<K_1,\cdots,S^{d-1}<K_{d-1}}(K_d - S^d)_+$) belong to $\mathbf{L}_{2,\alpha}$ with $\alpha = 1/2$.

The additional instruments. To perform the DGHS, we need $d(d+1)/2$ extra financial instruments that have non vanishing gammas. To simplify the exposure, we assume that these instruments are options with the same time maturity $T_2 > T$. Their payoffs are given by (for $0 \le j < k \le d$)

$$(S^k_{T_2} - K_{j,k} S^j_{T_2})_+ \tag{10.4}$$

($K_{j,k} > 0$). If $j = 0$ and $1 \le k \le d$, it corresponds to a Call on the asset S^k with strike $K_{0,k}e^{rT_2}$. If $1 \le j < k \le d$, it is an exchange option between S^j and S^k. The price function of each of these instruments writes

$$C^{j,k}(t, S^j, S^k) := \mathbb{E}_{\mathbb{Q}}\left[e^{-r(T_2-t)}(S^k_{T_2} - K_{j,k}S^j_{T_2})_+ | S^j_t = S^j, S^k_t = S^k\right].$$

Since the model is log-normal with constant correlation, the price functions are explicit and are given by the Black-Scholes and Margrabe formulas (see Appendix 10.A.1).

10.5.2 Greeks and martingales

In this paragraph, we introduce some specific martingales which are related to the option Greeks. These properties are quite standard, but to our knowledge their importance in the analysis of the tracking error has not been emphasized so far: in particular, leveraging this property, we will deduce that the tracking error is an iterated stochastic integral w.r.t. the $\mathbb{Q}$-Brownian motion W (see Theorem 10.6.1).
Notation. For any function $\psi(t, S^1, ..., S^d)$ and a multi-index $(l_1, ..., l_p) \in \{1, ..., d\}^p$, we will denote $\partial^p_{S^{l_1},...,S^{l_p}}\psi(t, S^1_t, ..., S^d_t)$ by $\partial^p_{l_1,...,l_p}\psi(t)$.
For $l, m, n = 1...d$, we define

$$\bar{u}(t) := e^{-rt}u(t); \tag{10.5}$$

$$\bar{u}^{(1)}_l(t) := e^{-rt}\sigma_l S^l_t \partial_l u(t); \tag{10.6}$$

$$\bar{u}^{(2)}_{l,m}(t) := e^{-rt}\sigma_l\sigma_m S^l_t S^m_t \partial^2_{l,m}u(t); \tag{10.7}$$

$$\bar{u}^{(3)}_{l,m,n}(t) := e^{-rt}\sigma_l\sigma_m\sigma_n S^l_t S^m_t S^n_t \partial^3_{l,m,n}u(t). \tag{10.8}$$

In the same way, we define, for $0 \le j < k \le d$ and $l, m, n = 1...d$,

$$\begin{aligned}
\bar{C}^{j,k}(t) &:= e^{-rt}C^{j,k}(t);\\
\bar{C}^{j,k,(1)}_l(t) &:= e^{-rt}\sigma_l S^l_t \partial_l C^{j,k}(t);\\
\bar{C}^{j,k,(2)}_{l,m}(t) &:= e^{-rt}\sigma_l\sigma_m S^l_t S^m_t \partial^2_{l,m}C^{j,k}(t);\\
\bar{C}^{j,k,(3)}_{l,m,n}(t) &:= e^{-rt}\sigma_l\sigma_m\sigma_n S^l_t S^m_t S^n_t \partial^3_{l,m,n}C^{j,k}(t).
\end{aligned}$$

The lemma below will be useful.

Lemma 10.5.2. *The processes above are $\mathbb{Q}$-martingales on [0,T), and their Itô decompositions w.r.t. the $\mathbb{Q}$-Brownian motions W write*

$$\mathrm{d}\bar{u}(t) = \sum_{l=1}^{d} \bar{u}^{(1)}_l(t)\mathrm{d}W^l_t; \tag{10.9}$$

$$\mathrm{d}\bar{u}^{(1)}_l(t) = \sum_{m=1}^{d} \left(\bar{u}^{(2)}_{l,m}(t) + \sigma_l \bar{u}^{(1)}_l(t)\mathbb{1}_{m=l}\right)\mathrm{d}W^m_t; \tag{10.10}$$

$$\mathrm{d}\bar{u}^{(2)}_{l,m}(t) = \sum_{n=1}^{d} \left(\bar{u}^{(3)}_{l,m,n}(t) + \sigma_m \bar{u}^{(2)}_{l,m}(t)\mathbb{1}_{n=m} + \sigma_l \bar{u}^{(2)}_{l,m}(t)\mathbb{1}_{n=l}\right)\mathrm{d}W^n_t. \tag{10.11}$$

Similarly, one has

$$\begin{aligned}
\mathrm{d}\bar{C}^{j,k}(t) &= \sum_{l=1}^{d} \bar{C}_l^{j,k,(1)}(t)\mathrm{d}W_t^l;\\
\mathrm{d}\bar{C}_l^{j,k,(1)}(t) &= \sum_{m=1}^{d} \Big(\bar{C}_{l,m}^{j,k,(2)}(t) + \sigma_l \bar{C}_l^{j,k,(1)}(t)\mathbb{1}_{m=l}\Big)\,\mathrm{d}W_t^m; \qquad (10.12)\\
\mathrm{d}\bar{C}_{l,m}^{j,k,(2)}(t) &= \sum_{n=1}^{d} \Big(\bar{C}_{l,m,n}^{j,k,(3)}(t) + \sigma_m \bar{C}_{l,m}^{j,k,(2)}(t)\mathbb{1}_{n=m} + \sigma_l \bar{C}_{l,m}^{j,k,(2)}(t)\mathbb{1}_{n=l}\Big)\,\mathrm{d}W_t^n.
\end{aligned}$$

Proof. We show the result for the martingales involving u. The proof is the same for those involving $C^{j,k}$.
Let us define $\bar{u}(t,S) := e^{-rt}u(t,S)$, for $S = (S^1,...,S^d) \in \mathbb{R}_+^d$. Since $\bar{u}(t,S_t) := \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t}\left[e^{-rT}g(S_T)\right]$, clearly $(\bar{u}(t,S_t))_{t\le T}$ is a $\mathbb{Q}$-martingale for any initial values $(s_0^j)_{1\le j\le d}$. Then (see e.g Gobet and Munos [GM05b]), for a multi-index $l = (l_1,...,l_d)$, the process $\partial^{|l|}_{(s_0^1)^{l_1},...,(s_0^d)^{l_d}}\bar{u}(t,S_t^1,...,S_t^d)$, which is equal to $(\frac{S_t^1}{s_0^1})^{l_1}...(\frac{S_t^d}{s_0^d})^{l_d}\partial^{|l|}_{l_1,\cdots,l_d}\bar{u}(t)$ (since S_t^j is linear w.r.t. s_0^j), is a martingale (the estimates on the regularity for $t<T$ are given later in Lemma 10.8.1). Taking $|l|$ equal to 1 (resp. 2, resp. 3), we obtain that the process in (10.6) (resp. (10.7), resp. (10.8)) is a $\mathbb{Q}$-martingale on $[0,T[$.
Concerning the differentials, it is clear that, if $(\psi(t))_t = \big(\psi(t,S_t^1,...,S_t^d)\big)_t$ is a $\mathbb{Q}$-martingale where ψ is a smooth function, then its Itô decomposition contains only terms w.r.t. $\mathrm{d}W$:

$$\mathrm{d}\psi(t) = \sum_{l=1}^{d} \sigma_l S_t^l \partial_l \psi(t)\mathrm{d}W_t^l.$$

By taking $\psi(t) = \bar{u}(t)$ (resp. $\bar{u}_l^{(1)}(t)$, resp. $\bar{u}_{l,m}^{(2)}(t)$), the relation (10.9) (resp. (10.10), resp. (10.11)) follows. □

10.5.3 Other notations

The time net. In all what follows, $\pi := (t_k)_{k=0...N}$ is a deterministic time net, such that $0 = t_0 < t_1 < ... < t_N = T$, and $|\pi| := \sup_{k=0...N-1}(t_{k+1} - t_k)$. This defines the N trading dates of the DGHS. We shall use the following net ($\beta \in (0,1]$)

$$\pi^{(\beta)} := \big\{t_k^{(N,\beta)} := T - T\big(1 - \frac{k}{N}\big)^{\frac{1}{\beta}}, 0 \le k \le N\big\}.$$

Note that $\pi^{(1)} = (t_k^{(N,1)})$ coincides with the equidistant net. For $\beta < 1$, the points in $\pi^{(\beta)}$ are more concentrated near T.

The constants. In our estimates, we use numerous constants that will be generically denoted by C or c. We emphasize the fact that, whenever a constant depends on the function g, the dependence will be expressed explicitly, so that all the constants

such as C or c do not depend on g, but may depend on $r, (\mu_j)_j, (\sigma_j)_j, (\rho_{j,k})_{j,k}, (K_{j,k})_{j,k}, (s_0^j)_j, \alpha$ and other universal constants. They may also depend on T and on T_2, but remain bounded when $T \to 0$.
$A \leq_c B$ means $A \leq cB$ with a generic constant c.

10.6 The Delta-Gamma hedging strategy

10.6.1 Decomposition of the tracking error as multiple Itô integrals

In order to illustrate the ideas and results of the multidimensional case, which will be studied in paragraph 10.6.1, let us show what happens in the one-dimensional case ($d = 1$), with a single asset S^1.

The one-dimensional case

In that case, the additional hedging instrument is simply a Call on S^1 with strike $K_{0,1}S^0_{T_2}$ at maturity T_2 (see (10.4)). The related price function $C^{0,1}(t, S^0, S^1)$ is the Black-Scholes formula, where the usual variables are t and S^1.

Now let us decompose the tracking error. Taking advantage of the self-financing condition, the discounted terminal value of the hedging portfolio $\bar{V}_T^N = e^{-rT}V_T^N$ is

$$\begin{aligned}\bar{V}_T^N :=& \bar{u}(0, S_0) + \sum_{i=0}^{N-1} \delta^1_{t_i}(\bar{S}^1_{t_{i+1}} - \bar{S}^1_{t_i}) \\ &+ \sum_{i=0}^{N-1} \delta^{0,1}_{t_i}(\bar{C}^{0,1}(t_{i+1}, S^0_{t_{i+1}}, S^1_{t_{i+1}}) - \bar{C}^{0,1}(t_i, S^0_{t_i}, S^1_{t_i})),\end{aligned} \tag{10.13}$$

where

$$\delta^{0,1}_{t_i} = \frac{\bar{u}^{(2)}_{1,1}(t_i)}{\bar{C}^{0,1,(2)}_{1,1}(t_i)}, \tag{10.14}$$

$$\delta^1_{t_i} = \frac{1}{\sigma_1 \bar{S}^1_{t_i}}\bar{u}^{(1)}_1(t_i) - \frac{1}{\sigma_1 \bar{S}^1_{t_i}}\delta^{0,1}_{t_i}\bar{C}^{0,1,(1)}_1(t_i). \tag{10.15}$$

Note that $\delta^{0,1}_{t_i} = \frac{\partial^2_{S^1,S^1}u(t,S^1_t)}{\partial^2_{S^1,S^1}C^{0,1}(t,S^0_t,S^1_t)}$ and $\delta^1_{t_i} = \partial_{S^1}u(t, S^1_t) - \delta^{0,1}_{t_i}\partial_{S^1}C^{0,1}(t, S^0_t, S^1_t)$, which shows that the above representation is equivalent to that mentioned in the introduction (see identities (10.1)). However, representations (10.14) and (10.15) are more convenient for the mathematical analysis below.

The discounted tracking error is defined by

$$\begin{aligned}\overline{\mathcal{E}}^{\Delta\Gamma}_N(g, \pi) :=& \bar{V}_T^N - e^{-rT}g(S_T) \\ =& \bar{V}_T^N - \bar{u}(T, S_T) \\ =& \bar{V}_T^N - (\bar{u}(0, S_0) + \int_0^T \bar{u}^{(1)}_1(t)\mathrm{d}W_t),\end{aligned}$$

where we have used the Itô decomposition (10.9). Combining this with (10.13), one has

$$\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi) = -\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} T^{i,(1)}(t)\mathrm{d}W_t, \tag{10.16}$$

where

$$T^{i,(1)}(t) := \bar{u}_1^{(1)}(t) - \delta_{t_i}^1\sigma_1\bar{S}_t^1 - \delta_{t_i}^{0,1}\bar{C}_1^{0,1,(1)}(t).$$

Then, using (10.15), we observe that

$$T^{i,(1)}(t_i) = \bar{u}_1^{(1)}(t_i) - \left(\frac{\bar{u}_1^{(1)}(t_i)}{\sigma_1\bar{S}_{t_i}^1} - \frac{\delta_{t_i}^{0,1}\bar{C}_1^{0,1,(1)}(t_i)}{\sigma_1\bar{S}_{t_i}^1}\right)\sigma_1\bar{S}_{t_i}^1 - \delta_{t_i}^{0,1}\bar{C}_1^{0,1,(1)}(t_i) = 0.$$

This is the usual Delta-hedging condition. Applying Itô's rule and using Lemma 10.5.2, one obtains

$$T^{i,(1)}(t) = \int_{t_i}^{t} T^{i,(2)}(s)\mathrm{d}W_s \tag{10.17}$$

for $t \in [t_i, t_{i+1})$, with

$$T^{i,(2)}(s) = \bar{u}_{1,1}^{(2)}(s) + \sigma_1\bar{u}_1^{(1)}(s) - \delta_{t_i}^1[\sigma_1]^2\bar{S}_s^1 - \delta_{t_i}^{0,1}\left(\bar{C}_{1,1}^{0,1,(2)}(s) + \sigma_1\bar{C}_1^{0,1,(1)}(s)\right).$$

Then, using (10.15), we obtain

$$\begin{aligned}
T^{i,(2)}(t_i) =& \bar{u}_{1,1}^{(2)}(t_i) + \sigma_1\bar{u}_1^{(1)}(t_i) - \left(\frac{\bar{u}_1^{(1)}(t_i)}{\sigma_1\bar{S}_{t_i}^1} - \frac{\delta_{t_i}^{0,1}\bar{C}_1^{0,1,(1)}(t_i)}{\sigma_1\bar{S}_{t_i}^1}\right)[\sigma_1]^2\bar{S}_{t_i}^1 \\
&- \delta_{t_i}^{0,1}\left(\bar{C}_{1,1}^{0,1,(2)}(t_i) + \sigma_1\bar{C}_1^{0,1,(1)}(t_i)\right) \\
=& \bar{u}_{1,1}^{(2)}(t_i) - \delta_{t_i}^{0,1}\bar{C}_{1,1}^{0,1,(2)}(t_i).
\end{aligned}$$

Using (10.14), it leads to $T^{i,(2)}(t_i) = 0$: this is the Gamma-hedging condition. We apply once again Itô's rule and Lemma 10.5.2 to obtain

$$T^{i,(2)}(s) = \int_{t_i}^{s} T^{i,(3)}(r)\mathrm{d}W_r, \tag{10.18}$$

where $T^{i,(3)}(r)$ can be explicitly written using $\bar{u}_{1,1,1}^{(3)}$, $\bar{u}_{1,1}^{(2)}$, $\bar{u}_1^{(1)}$ and the similar Greeks for $C^{0,1}$. For the purpose of the current discussion, the exact espression of $T^{i,(3)}(t)$ is not important (it is given in the general case in Theorem 10.6.1). More interesting is to identify the form of the tracking error as a triple iterated stochastic integral w.r.t. W. Indeed, from (10.16), (10.17) and (10.18), one obtains

$$\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi) = -\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s} T^{i,(3)}(r)\mathrm{d}W_r\mathrm{d}W_s\mathrm{d}W_t. \tag{10.19}$$

The generalization of such a decomposition to the multidimensional asset model will be given in Theorem 10.6.1.

Now, let us formally comment the consequences of (10.19). At first sight, if $T^{i,(3)}(r)$ had a $\mathbb{Q}$-$\mathbf{L}_2$ moment uniformly bounded in r, we would deduce that

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi)|^2 &\leq \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\sup_{0\leq r\leq T}|T^{i,(3)}(r)|^2_{\mathbf{L}_2}\mathrm{d}r\mathrm{d}s\mathrm{d}t \\ &\leq \sup_{0\leq i\leq N-1,\ 0\leq r\leq T}|T^{i,(3)}(r)|^2_{\mathbf{L}_2}|\pi|^2\frac{T}{6},\end{aligned} \tag{10.20}$$

i.e. the tracking error has a $\mathbb{Q}$-$\mathbf{L}_2$ norm of the order of the time step $|\pi|$. In particular, N regular trading dates would lead to an optimal rate of convergence. Actually, this argumentation is not complete because the moments of $T^{i,(3)}(r)$ may explode as r goes to T (because for non smooth payoff, the Greeks may go to $\pm\infty$ as the time to maturity shrinks to 0). It illustrates that a significant piece of work in the Delta-Gamma error analysis lies in the understanding of the behavior of $|T^{i,(3)}(r)|^2_{\mathbf{L}_2}$ as $r\to T$. This feature is tightly estimated through the fractional regularity of g (see Proposition 10.6.1 and Corollary 10.6.1).

The multidimensional case

Following the previous decomposition for a single asset, we can handle the general case as well. The discounted terminal value of the hedging portfolio is

$$\begin{aligned}\bar{V}_T^N &:= \bar{u}(0,S_0)+\sum_{i=0}^{N-1}\sum_{j=1}^{d}\delta_{t_i}^j(\bar{S}_{t_{i+1}}^j-\bar{S}_{t_i}^j) \\ &+\sum_{i=0}^{N-1}\sum_{0\leq j<k\leq d}\delta_{t_i}^{j,k}(\bar{C}^{j,k}(t_{i+1},S_{t_{i+1}}^j,S_{t_{i+1}}^k)-\bar{C}^{j,k}(t_i,S_{t_i}^j,S_{t_i}^k)).\end{aligned} \tag{10.21}$$

Definition 10.6.1 (Delta-Gamma hedging strategy)**.** *We define the DGHS by*

$$\delta_{t_i}^{j,k} := \frac{\bar{u}_{j,k}^{(2)}(t_i)}{\bar{C}_{j,k}^{j,k,(2)}(t_i)} \quad (1\leq j<k\leq d,\ \textit{Exchange options}), \tag{10.22}$$

$$\delta_{t_i}^{0,l} := \frac{\bar{u}_{l,l}^{(2)}(t_i)}{\bar{C}_{l,l}^{0,l,(2)}(t_i)}-\frac{1}{\bar{C}_{l,l}^{0,l,(2)}(t_i)}\sum_{1\leq j<k\leq d}\delta_{t_i}^{j,k}\bar{C}_{l,l}^{j,k,(2)}(t_i) \quad (1\leq l\leq d,\ \textit{Call options}), \tag{10.23}$$

$$\delta_{t_i}^{l} := \frac{1}{\sigma_l\bar{S}_{t_i}^l}\bar{u}_l^{(1)}(t_i)-\frac{1}{\sigma_l\bar{S}_{t_i}^l}\sum_{0\leq j<k\leq d}\delta_{t_i}^{j,k}\bar{C}_l^{j,k,(1)}(t_i) \quad (1\leq l\leq d,\ \textit{assets}). \tag{10.24}$$

These are natural extensions of (10.14) and (10.15). Then, similarly, the discounted tracking error is defined by

$$\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi) := \bar{V}_T^N - e^{-rT}g(S_T).$$

Using expression (10.21) and Lemma 10.5.2, one has

$$\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi) = -\sum_{i=0}^{N-1}\sum_{l=1}^{d}\int_{t_i}^{t_{i+1}} T_l^{i,(1)}(t)\mathrm{d}W_t^l, \tag{10.25}$$

where

$$T_l^{i,(1)}(t) = \bar{u}_l^{(1)}(t) - \delta_{t_i}^l\sigma_l\bar{S}_t^l - \sum_{0\le j<k\le d}\delta_{t_i}^{j,k}\bar{C}_l^{j,k,(1)}(t) \tag{10.26}$$

for $t \in [t_i, t_{i+1})$.

10.6.2 Main results

In the general case, the (discounted) tracking error can be decomposed as a triple stochastic integral w.r.t. W, analogously to (10.19) in dimension 1. In the decomposition below, there are a principal part (related to the third derivative of the option price) and some residual terms.

Theorem 10.6.1. *Assume that $\mathbb{E}_{\mathbb{P}}|g(S_T)|^{2p_0} < \infty$ for some $p_0 > 1$. We have*

$$\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi) = -\sum_{i=0}^{N-1}\sum_{l,m,n=1}^{d}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\left(\bar{u}_{l,m,n}^{(3)}(r) + R_{l,m,n}^{i,(3)}(r)\right)\mathrm{d}W_r^n\mathrm{d}W_s^m\mathrm{d}W_t^l, \tag{10.27}$$

where

$$\begin{aligned}
R_{l,m,n}^{i,(3)}(t) = &\left(\sigma_m\bar{u}_{l,m}^{(2)}(t)\mathbb{1}_{n=m} + \sigma_l\bar{u}_{l,m}^{(2)}(t)\mathbb{1}_{n=l} + \sigma_l\bar{u}_{l,n}^{(2)}(t)\mathbb{1}_{m=l}\right)\\
&+\sigma_l^2\left(\bar{u}_l^{(1)}(t) - \frac{\bar{S}_t^l}{\bar{S}_{t_i}^l}\bar{u}_l^{(1)}(t_i)\right)\mathbb{1}_{n=m=l}\\
&-\sum_{0\le j<k\le d}\delta_{t_i}^{j,k}\bar{C}_{l,m,n}^{j,k,(3)}(t)\\
&-\sum_{0\le j<k\le d}\delta_{t_i}^{j,k}\left(\sigma_m\bar{C}_{l,m}^{j,k,(2)}(t)\mathbb{1}_{n=m} + \sigma_l\bar{C}_{l,m}^{j,k,(2)}(t)\mathbb{1}_{n=l} + \sigma_l\bar{C}_{l,n}^{j,k,(2)}(t)\mathbb{1}_{m=l}\right)\\
&-\sigma_l^2\sum_{0\le j<k\le d}\delta_{t_i}^{j,k}\left(\bar{C}_l^{j,k,(1)}(t) - \frac{\bar{S}_t^l}{\bar{S}_{t_i}^l}\bar{C}_l^{j,k,(1)}(t_i)\right)\mathbb{1}_{n=m=l}.
\end{aligned} \tag{10.28}$$

In addition, the following estimate holds:

$$\mathbb{E}_{\mathbb{P}}\left|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi)\right|^2 \le C\sum_{i=0}^{N-1}\sum_{l,m,n=1}^{d}\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s}\mathbb{E}_{\mathbb{P}}\left|\bar{u}_{l,m,n}^{(3)}(r) + R_{l,m,n}^{i,(3)}(r)\right|^2\mathrm{d}r\mathrm{d}s\mathrm{d}t. \tag{10.29}$$

The proof is postponed to Section 10.8. We now bring in the next proposition several estimates, as key ingredients of our main results. These estimates are proved in Section 10.8.

Proposition 10.6.1. *Assume that* $\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0} < \infty$ *for some* $p_0 > 1$.

- *For* $l, m, n = 1...d$ *and* $0 \leq t < T$,

$$\mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(3)}_{l,m,n}(t)\right|^2 \leq C\frac{V_{t,T}(g)}{(T-t)^3}. \tag{10.30}$$

- *There exists a positive* $\pi^{\text{threshold}}$ *(depending on* p_0, T *and* T_2*) such that, if* $|\pi| \leq \pi^{\text{threshold}}$, *then, for* $0 \leq t_i \leq t < t_{i+1} \leq T$,

$$\mathbb{E}_{\mathbb{P}}\left|R^{i,(3)}_{l,m,n}(t)\right|^2 \leq C\frac{V_{t,T}(g) + \left(\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0}\right)^{\frac{1}{p_0}}}{(T-t)^2}. \tag{10.31}$$

Thus, an easy and direct consequence of Proposition 10.6.1 and the definition (10.3) of $K^{\alpha,p_0}(g)$ is the following corollary.

Corollary 10.6.1. *Let* $\alpha \in (0,1]$. *Suppose that* $g \in \mathbf{L}_{2,\alpha}$ *and that* $\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0} < \infty$ *for some* $p_0 > 1$. *Then, when* $|\pi| \leq \pi^{\text{threshold}}$, *and for* $0 \leq t < T$,

$$\mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(3)}_{l,m,n}(t) + R^{i,(3)}_{l,m,n}(t)\right|^2 \leq C\frac{K^{\alpha,p_0}(g)}{(T-t)^{3-\alpha}}. \tag{10.32}$$

Now we are in a position to expose our main results.

Theorem 10.6.2. *Let* $\alpha \in (0,1]$. *Assume* $g \in \mathbf{L}_{2,\alpha}$ *and* $\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0} < \infty$ *for some* $p_0 > 1$.

1. Regular grid $\pi^{(1)}$. *For* N *sufficiently large to ensure* $|\pi^{(1)}| = \frac{T}{N} \leq \pi^{\text{threshold}}$, *one has*

$$\left(\mathbb{E}_{\mathbb{P}}\left|\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi^{(1)})\right|^2\right)^{1/2} \leq C\frac{(K^{\alpha,p_0}(g))^{1/2}\ T^{\alpha/2}}{N^{\alpha/2}}. \tag{10.33}$$

2. Non regular grid $\pi^{(\beta)}$, $\beta \in (0,1)$. *Taking* N *sufficiently large to ensure* $|\pi^{(\beta)}| \leq \pi^{\text{threshold}}$, *one has*

$$\left(\mathbb{E}_{\mathbb{P}}\left|\overline{\mathcal{E}}^{\Delta\Gamma}_N(g,\pi^{(\beta)})\right|^2\right)^{1/2} \leq \begin{cases} C\dfrac{(K^{\alpha,p_0}(g))^{1/2}\ T^{\alpha/2}}{N^{\frac{\alpha}{2\beta}}} & \text{if } \beta \in (\frac{\alpha}{2},1), \\ C\dfrac{(K^{\alpha,p_0}(g))^{1/2}\ T^{\alpha/2}}{N}\sqrt{\log(1+N)} & \text{if } \beta = \frac{\alpha}{2}, \\ C\dfrac{(K^{\alpha,p_0}(g))^{1/2}\ T^{\alpha/2}}{N} & \text{if } \beta \in (0,\frac{\alpha}{2}). \end{cases}$$

Before proving these results, we give few comments.

1. The use of irregular grids crucially helps to increase the convergence rate of the tracking error, for any fractional regularity $\alpha \in (0,1]$, up to the rate N.

2. For the regular grid, in general the estimate (10.33) of the rate of convergence $N^{\alpha/2}$ is tight for $\alpha < 1$ (because the estimate (10.32) is tight, see the discussions in Geiss and Hujo [GH07] and Gobet and Makhlouf [GM08]). In addition, it coincides with the rate of convergence of the Delta hedging strategy: in other words, for non-smooth payoffs, the DGHS used with regular grids does not improve the rate of convergence of the tracking error. This is confirmed by our numerical results in Section 10.7.

3. In the case $\alpha = 1$, our estimate (10.33) is not optimal. Indeed, for Call options, from Brodén and Wiktorsson [BW08] we expect the rate to be equal to $N^{3/4}$. For smoother payoffs (say C_b^3, leading also to $\alpha = 1$), the first three spatial derivatives of u are bounded and our computations lead to a convergence rate equal to N (see the estimates (10.20)).
   Hence, the case $\alpha = 1$ is a singular point, for which the convergence rate may go from $N^{1/2}$ to N. Tight estimates on $(\mathbb{E}_{\mathbb{P}} \left|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g, \pi^{(1)})\right|^2)^{1/2}$ can not be given only by assuming $g \in \mathbf{L}_{2,1}$: presumably, an extra condition on ∇g would be necessary. We leave this issue for further investigations.

Proof of Theorem 10.6.2. Put $T_{l,m,n}^{i,(3)}(r) = \bar{u}_{l,m,n}^{(3)}(r) + R_{l,m,n}^{i,(3)}(r)$ and define $f : \tau \mapsto \int_{t_i}^{\tau}\int_{t_i}^{t}\int_{t_i}^{s} \mathbb{E}_{\mathbb{P}} \left|T_{l,m,n}^{i,(3)}(r)\right|^2 \mathrm{d}r\mathrm{d}s\mathrm{d}t$. The Taylor expansion of order 2, with integral remainder term, of the function $f(.)$ between $\tau = t_i$ and $\tau = t_{i+1}$ gives

$$\int_{t_i}^{t_{i+1}}\int_{t_i}^{t}\int_{t_i}^{s} \mathbb{E}_{\mathbb{P}} \left|T_{l,m,n}^{i,(3)}(r)\right|^2 \mathrm{d}r\mathrm{d}s\mathrm{d}t = \frac{1}{2}\int_{t_i}^{t_{i+1}} (t_{i+1}-t)^2\, \mathbb{E}_{\mathbb{P}} \left|T_{l,m,n}^{i,(3)}(t)\right|^2 \mathrm{d}t.$$

Then, from Theorem 10.6.1 and Corollary 10.6.1, one obtains

$$\mathbb{E}_{\mathbb{P}} \left|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g, \pi^{(\beta)})\right|^2 \leq C \sum_{i=0}^{N-1} \int_{t_i^{(\beta)}}^{t_{i+1}^{(\beta)}} \left(t_{i+1}^{(\beta)} - t\right)^2 \frac{K^{\alpha,p_0}(g)}{(T-t)^{3-\alpha}} \mathrm{d}t. \tag{10.34}$$

The term $i = N-1$ in the above summation is equal to

$$C \int_{T-t_{N-1}^{(N,\beta)}}^{T} \frac{K^{\alpha,p_0}(g)}{(T-t)^{1-\alpha}} \mathrm{d}t = \frac{CK^{\alpha,p_0}(g)}{\alpha} \frac{T^\alpha}{N^{\alpha/\beta}}, \tag{10.35}$$

where we have used $T - t_{N-1}^{(N,\beta)} = \frac{T}{N^{1/\beta}}$. Furthermore, it is known (see Gobet and Makhlouf [GM08]) that

$$\sup_{i=0\ldots N-1} \sup_{t\in[t_i^{(\beta)}, t_{i+1}^{(\beta)})} \left(\frac{t_{i+1}^{(\beta)} - t}{(T-t)^{1-\beta}}\right) \leq \frac{T^\beta}{\beta N}.$$

From this, (10.34) and (10.35), it follows that

$$\mathbb{E}_{\mathbb{P}}\left|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi^{(\beta)})\right|^2 \leq \frac{CK^{\alpha,p_0}(g)}{\alpha}\frac{T^\alpha}{N^{\alpha/\beta}} + CK^{\alpha,p_0}(g)\big(\frac{T^\beta}{\beta N}\big)^2\int_0^{t_{N-1}^{(N,\beta)}}(T-t)^{2-2\beta}\frac{\mathrm{d}t}{(T-t)^{3-\alpha}}. \tag{10.36}$$

To complete the proof, note that it remains to upper bound the above integral in the three cases $\beta \in (\alpha/2, 1]$, $\beta = \alpha/2$ and $\beta \in (0, \alpha/2)$. Denote by I the second term in the r.h.s. of (10.36).

1. If $\beta > \alpha/2$, the function $f : t \mapsto \frac{(T-t)^{2-2\beta}}{(T-t)^{3-\alpha}}$ is not integrable at $t = T$ and we easily derive

$$I \leq CK^{\alpha,p_0}(g)\big(\frac{T^\beta}{\beta N}\big)^2\frac{T^{\alpha-2\beta}}{(2\beta-\alpha)N^{(\alpha-2\beta)/\beta}} = CK^{\alpha,p_0}(g)\frac{T^\alpha}{\beta^2(2\beta-\alpha)}\frac{1}{N^{\frac{\alpha}{\beta}}}.$$

2. If $\beta = \alpha/2$, the function f is still not integrable at $t = T$ and we obtain

$$I \leq CK^{\alpha,p_0}(g)\big(\frac{T^\beta}{\beta N}\big)^2\log(N^{1/\beta}).$$

3. If $\beta < \alpha/2$, the function f is integrable and it gives

$$I \leq CK^{\alpha,p_0}(g)\big(\frac{T^\beta}{\beta N}\big)^2\frac{T^{\alpha-2\beta}}{(\alpha-2\beta)} = CK^{\alpha,p_0}(g)\frac{T^\alpha}{\beta^2(\alpha-2\beta)}\frac{1}{N^2}.$$

□

10.7 Numerical Results

In this section, we present some experiments of Delta and Delta-Gamma hedging strategies, of both a European Call option ($g(S) = (S-K)_+$) and a European digital Call option ($g(S) = \mathbb{1}_{S\geq K}$) on a single asset ($d = 1$). The two payoff functions belong respectively to $\mathbf{L}_{2,1}$ and $\mathbf{L}_{2,\frac{1}{2}}$. In our experiments, we rely on the known explicit formulas for the prices of these two options. In the following tests, we take $r = 2\%$, $\mu = 1\%$, $\sigma = 25\%$, $s_0^1 = 100$, $T = 1$ year and $K = 100$. The additional hedging instrument is a Call with strike $K_{0,1} = 100$ and maturity $T_2 = 1.25$ year. All the experiments are carried out with 10000 simulations under the historical probability.

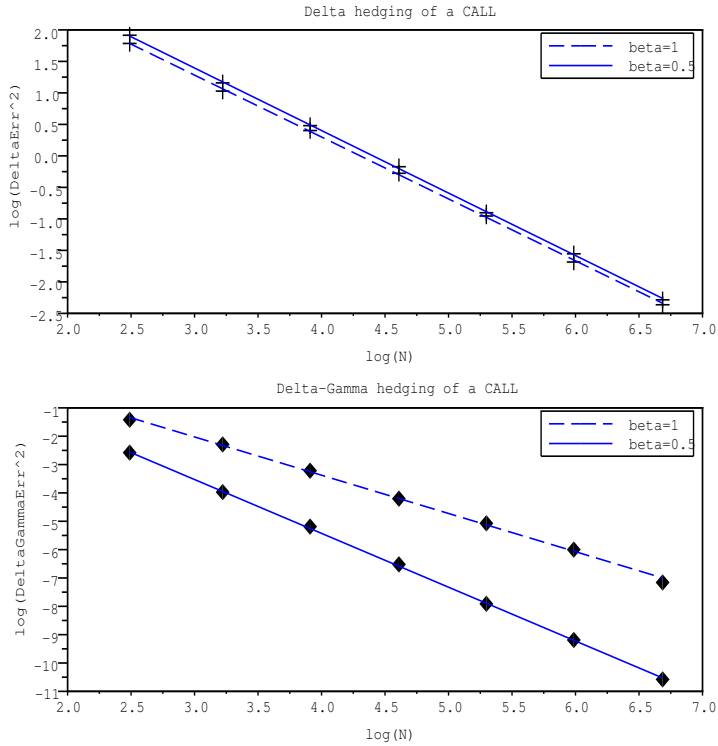

Figure 10.1: For a Call: at the top (Delta hedging strategy), $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N^{\Delta}(g,\pi^{(\beta)})|^2)$ vs $\log(N)$. At the bottom (Delta-Gamma hedging strategy), $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi^{(\beta)})|^2)$ vs $\log(N)$.

Comparison of the order of $\mathbf{L_2}$-convergence. For the figures 10.1 and 10.2, the simulations are made with $N = 12$, 25, 50, 100, 200, 400, 800 rebalancing dates, following different β's, i.e. different time nets. Note that, in practice, $N = 12$, $N = 25$ and $N = 50$ correspond approximately to a monthly, fortnightly and weekly rebalancing respectively.

Log-log plots. These figures are the log-log plots of the second moment of the tracking error versus the number of rebalancing dates (i.e. $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N(g,\pi^{(\beta)})|^2)$ versus $\log(N)$). We observe that they give straight lines, suggesting that not only upper bounds are available but presumably expansion results are also valid in these cases. The resulting slopes of the log-log plots represent twice the convergence order of $(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N(g,\pi^{(\beta)})|^2)^{1/2}$: these empirical convergence orders are reported in Table 10.1 for the Call and in Table 10.2 for the digital Call.

Theoretical order of convergence. In these tabulars, we also indicate the convergence order that one can expect from theoretical estimates, together with the reference.

Empirical order of convergence. The high convergence orders are not accurately estimated, possibly because of the relatively large statistical error due to simulations. Generally speaking, we notice that the rate of the Delta-Gamma tracking error is

	Delta hedging		Delta-Gamma hedging	
Beta	1.00	0.50	1.00	0.50
Expected order	0.50 see Zhang [Zha99]	0.50 see Geiss [Gei02]	0.75 see Brodén and Wiktorsson [BW08]	1.00 (up to a log factor) see our Theorem 10.6.2
Empirical order	0.49	0.49	0.68	0.95

Table 10.1: For a Call: order of convergence of the $\mathbf{L}_2$-norm of tracking errors $(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N(g,\pi^{(\beta)})|^2)^{1/2}$.

	Delta hedging		Delta-Gamma hedging		
Beta	1.00	0.50	1.00	0.50	0.25
Expected order	0.25 see Gobet and Temam [GT01]	0.50 (up to a log factor) see Geiss [Gei02]	0.25 see our Theorem 10.6.2	0.50 see our Theorem 10.6.2	1.00 (up to a log factor) see our Theorem 10.6.2
Empirical order	0.24	0.40	0.25	0.49	0.88

Table 10.2: For a Digital Call: order of convergence of the $\mathbf{L}_2$-norm of tracking errors $(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N(g,\pi^{(\beta)})|^2)^{1/2}$.

better than that of the Delta tracking error. However, when $\beta = 1$ (equidistant time net), there may be no significative difference between the two rates: consider for instance the digital Call (see Table 10.2). This shows the advantage of hedging at non equidistant rebalancing dates when the fractional regularity index α of the payoff function is smaller than 1.
Furthermore, the smaller β, the better the rate of convergence of the tracking error, either for the hedging of the Call or for that of the digital Call. However, one should not take β too small since the time net points become too close to each other near the maturity, which might cause numerical instabilities in the simulations or might lead to unrealistic trading dates. From Tables 10.1 and 10.2, observe that one obtains approximately the right convergence order for the $\mathbf{L}_2$-norm of tracking errors (order 1/2 with Delta hedging and order 1 with Delta-Gamma hedging) by taking the critical threshold for β, i.e. α for Delta hedging and $\frac{\alpha}{2}$ for Delta-Gamma hedging.

Comparison of convergence in distribution. Figures 10.3 and 10.4 show, for each option and for different β's, the histogram of the tracking error, superposed to the Gaussian density with mean zero and variance equal to the observed empirical variance of the error. It is obtained for $N = 1000$ trading dates. This comparison is aimed at checking two features:

1. is the asymptotic distribution Gaussian? the answer is generally no, as it can be easily observed in the figures. Additionally, it is known (cf Gobet and Temam

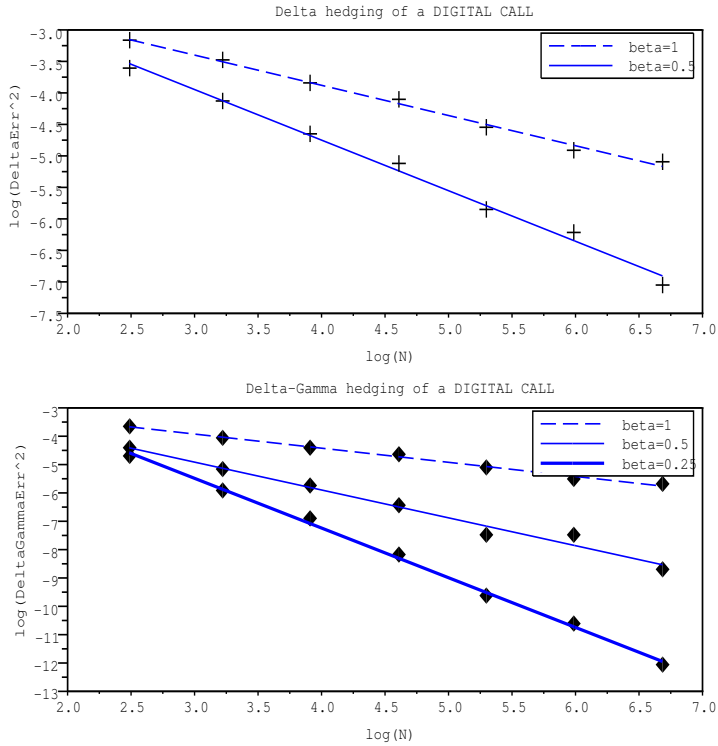

Figure 10.2: For a digital Call: at the top (Delta hedging strategy), $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N^{\Delta}(g,\pi^{(\beta)})|^2)$ vs $\log(N)$. At the bottom (Delta-Gamma hedging strategy), $\log(\mathbb{E}_{\mathbb{P}}|\overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi^{(\beta)})|^2)$ vs $\log(N)$.

[GT01], Hayashi and Mykland [HM05]) that the limiting distribution is not Gaussian but mixed Gaussian. Hence, we should consider the Gaussian distribution with the right variance as a benchmark and not as the true distribution: this representation helps to answer the next item.

2. do the convergence rates in $\mathbf{L}_2$ and in distribution differ? In introduction, we mention that it can happen.

When $\beta > \frac{\alpha}{2}$ with Delta-Gamma hedging and $\beta > \alpha$ with Delta hedging, one notices from Figures 10.3 and 10.4 that the empirical distribution of the tracking error and the related Gaussian distribution seem to be not in the same scale, which corroborates the fact that the convergence in $\mathbf{L}_2$ and in distribution hold at different rates.

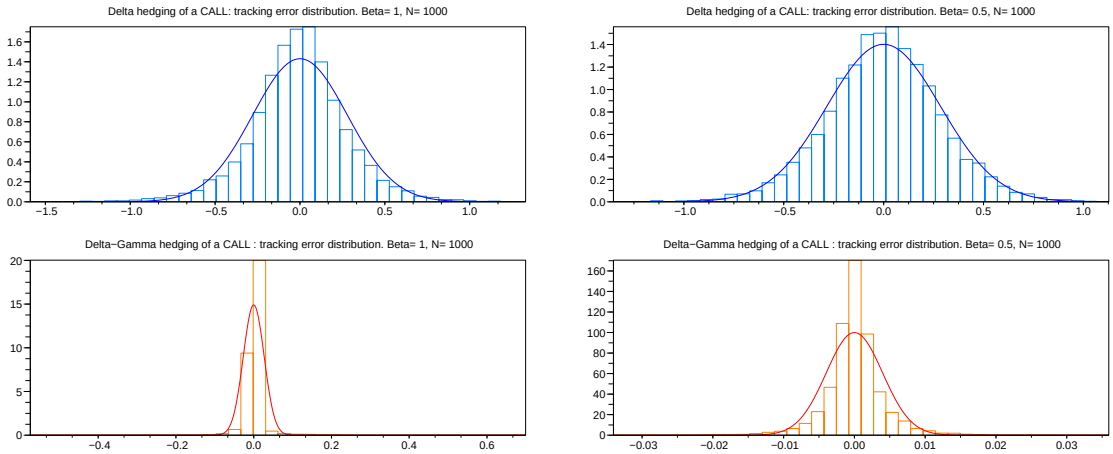

Figure 10.3: Distributions of the tracking errors for a Call

10.8 Proofs

10.8.1 Proof of Theorem 10.6.1.

Decomposition (10.27). It is clear that the definition of $\delta^l_{t_i}$ given by (10.24) ensures that $T^{i,(1)}_l(t_i) = 0$ for any $l = 1...d$, where $T^{i,(1)}_l(t)$ is defined by (10.26). Thus $T^{i,(1)}_l(t) = \int_{t_i}^t \mathrm{d}T^{i,(1)}_l(s)$, $t \in [t_i, t_{i+1})$. Applying Itô's rule and using Lemma 10.5.2, one obtains

$$\mathrm{d}T^{i,(1)}_l(t) = \sum_{m=1}^d T^{i,(2)}_{l,m}(t)\mathrm{d}W^m_t,$$

with

$$\begin{aligned} T^{i,(2)}_{l,m}(t) &= \bar{u}^{(2)}_{l,m}(t) + \sigma_l \bar{u}^{(1)}_l(t)\mathbb{1}_{m=l} - \delta^l_{t_i}\sigma_l^2 \bar{S}^l_t \mathbb{1}_{m=l} \\ &\quad - \sum_{0\le j<k\le d} \delta^{j,k}_{t_i}\left(\bar{C}^{j,k,(2)}_{l,m}(t) + \sigma_l \bar{C}^{j,k,(1)}_l(t)\mathbb{1}_{m=l}\right). \end{aligned}$$

Now, as in the one-dimensional case (see paragraph 10.6.1), we can check that the definitions of $\delta^{j,k}_{t_i}$ given by (10.22) (resp. (10.23)) ensure that $T^{i,(2)}_{l,m}(t_i) = 0$ for $1 \le l \ne m \le d$ (resp. $1 \le l = m \le d$) (in the previous cancellations, we strongly use that the price function $C^{j,k}$ depends on S only through the variables S^j and S^k). We apply once again Itô's rule and Lemma 10.5.2 to obtain

$$\mathrm{d}T^{i,(2)}_{l,m}(t) = \sum_{n=1}^d T^{i,(3)}_{l,m,n}(t)\mathrm{d}W^n_t$$

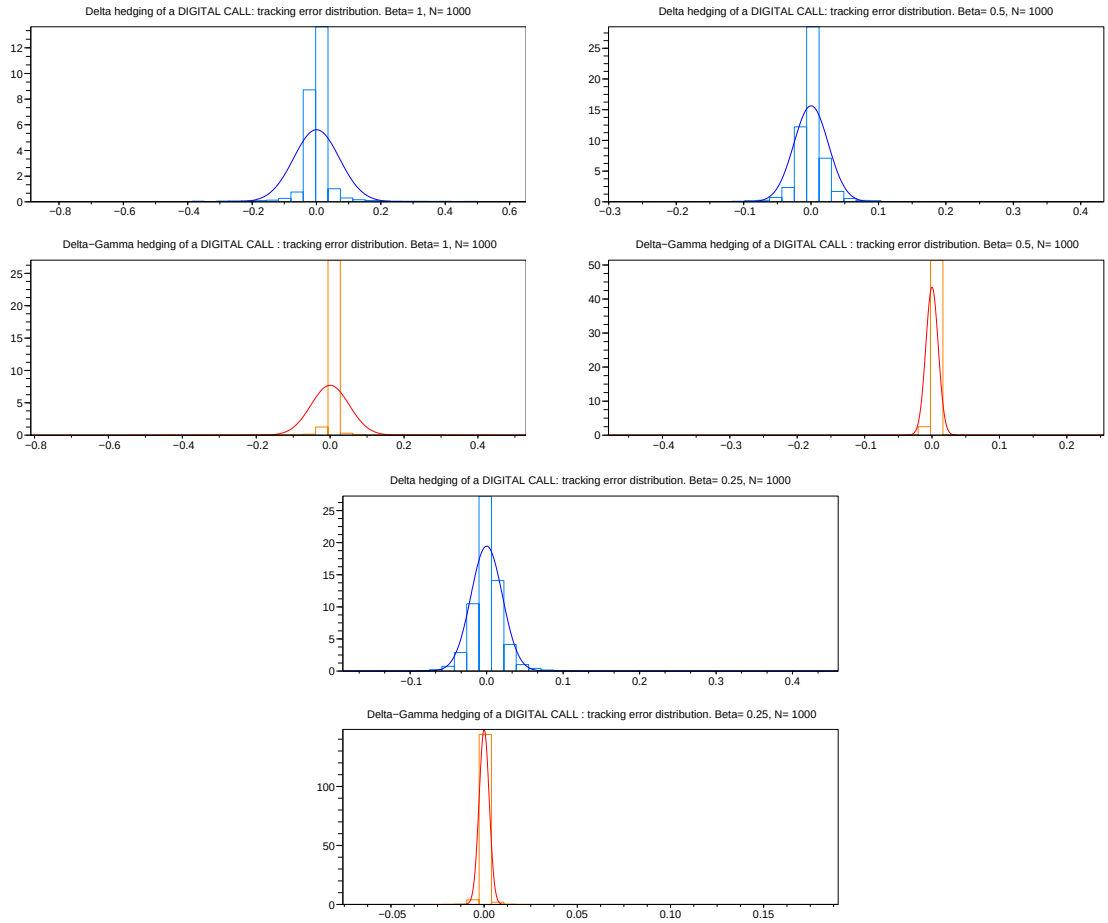

Figure 10.4: Distributions of the tracking errors for a Digital Call

for $t \in [t_i, t_{i+1})$, with

$$
\begin{aligned}
T^{i,(3)}_{l,m,n}(t) = {} & \bar{u}^{(3)}_{l,m,n}(t) + \sigma_m \bar{u}^{(2)}_{l,m}(t)\mathbb{1}_{n=m} + \sigma_l \bar{u}^{(2)}_{l,m}(t)\mathbb{1}_{n=l} \\
& + \left(\sigma_l \bar{u}^{(2)}_{l,n}(t) + \sigma_l^2 \bar{u}^{(1)}_{l}(t)\mathbb{1}_{n=l}\right)\mathbb{1}_{m=l} \\
& - \delta^l_{t_i}\sigma_l^3 \bar{S}^l_t \mathbb{1}_{n=m=l} \qquad (10.37) \\
& - \sum_{0\le j<k\le d} \delta^{j,k}_{t_i}\left(\bar{C}^{j,k,(3)}_{l,m,n}(t) + \sigma_m \bar{C}^{j,k,(2)}_{l,m}(t)\mathbb{1}_{n=m} + \sigma_l \bar{C}^{j,k,(2)}_{l,m}(t)\mathbb{1}_{n=l}\right) \\
& - \sum_{0\le j<k\le d} \delta^{j,k}_{t_i}\left(\sigma_l \bar{C}^{j,k,(2)}_{l,n}(t) + \sigma_l^2 \bar{C}^{j,k,(1)}_{l}(t)\mathbb{1}_{n=l}\right)\mathbb{1}_{m=l}.
\end{aligned}
$$

Substituting the expression (10.24) for $\delta^l_{t_i}$ in line (10.37) above, and rearranging the different terms, one readily obtains (10.28).

Estimate (10.29). The required estimate is computed under the historical probability $\mathbb{P}$. Note that an estimation of the $\mathbf{L}_2$-norm under the risk-neutral measure $\mathbb{Q}$ is straightforward using the Itô isometry.
Writing $\mathrm{d}W_t^l = \mathrm{d}\hat{W}_t^l + \lambda_l \mathrm{d}t$ in (10.25) yields (using the inequality $(a+b)^2 \leq 2(a^2+b^2)$)

$$\mathbb{E}_{\mathbb{P}} \left| \overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi) \right|^2 \leq 2\mathbb{E}_{\mathbb{P}} \left| \sum_{i=0}^{N-1} \sum_{l=1}^{d} \int_{t_i}^{t_{i+1}} T_l^{i,(1)}(t) \mathrm{d}\hat{W}_t^l \right|^2 + 2(\sup_{l=1\ldots d} \lambda_l^2) \mathbb{E}_{\mathbb{P}} \left| \sum_{i=0}^{N-1} \sum_{l=1}^{d} \int_{t_i}^{t_{i+1}} \left| T_l^{i,(1)}(t) \right| \mathrm{d}t \right|^2 .$$

One can apply Itô's isometry to the first term ($(\hat{W}_t)_t$ being a $\mathbb{P}$-Brownian motion) and Cauchy-Schwartz inequality to the second to obtain

$$\mathbb{E}_{\mathbb{P}} \left| \overline{\mathcal{E}}_N^{\Delta\Gamma}(g,\pi) \right|^2 \leq \left(2 + 2Td(\sup_{l=1\ldots d} \lambda_l^2)\right) \sum_{i=0}^{N-1} \sum_{l=1}^{d} \int_{t_i}^{t_{i+1}} \mathbb{E}_{\mathbb{P}} \left| T_l^{i,(1)}(t) \right|^2 \mathrm{d}t. \tag{10.38}$$

Now, since $T_l^{i,(1)}(t) = \sum_{m=1}^{d} \int_{t_i}^{t} T_{l,m}^{i,(2)}(s) \mathrm{d}W_s^m$ and $T_{l,m}^{i,(2)}(s) = \sum_{n=1}^{d} \int_{t_i}^{s} T_{l,m,n}^{i,(3)}(r) \mathrm{d}W_r^n$ with $T_{l,m,n}^{i,(3)}(r) = \bar{u}_{l,m,n}^{(3)}(r) + R_{l,m,n}^{i,(3)}(r)$, using the same arguments as above, we show that

$$\mathbb{E}_{\mathbb{P}} \left| T_l^{i,(1)}(t) \right|^2 \leq_c \sum_{m=1}^{d} \int_{t_i}^{t} \mathbb{E}_{\mathbb{P}} \left| T_{l,m}^{i,(2)}(s) \right|^2 \mathrm{d}s,$$

$$\mathbb{E}_{\mathbb{P}} \left| T_{l,m}^{i,(2)}(s) \right|^2 \leq_c \sum_{n=1}^{d} \int_{t_i}^{s} \mathbb{E}_{\mathbb{P}} \left| T_{l,m,n}^{i,(3)}(r) \right|^2 \mathrm{d}r.$$

Plugging these inequalities into (10.38) leads to (10.29). □

10.8.2 Proof of Proposition 10.6.1.

Estimate (10.30). It is part of the statement of the following lemma.

Lemma 10.8.1. *For $1 \leq l, m, n \leq d$ and $0 \leq t < T$,*

$$\mathbb{E}_{\mathbb{P}} \left| \bar{u}_l^{(1)}(t) \right|^2 \leq C \frac{V_{t,T}(g)}{(T-t)}, \tag{10.39}$$

$$\mathbb{E}_{\mathbb{P}} \left| \bar{u}_{l,m}^{(2)}(t) \right|^2 \leq C \frac{V_{t,T}(g)}{(T-t)^2}, \tag{10.40}$$

$$\mathbb{E}_{\mathbb{P}} \left| \bar{u}_{l,m,n}^{(3)}(t) \right|^2 \leq C \frac{V_{t,T}(g)}{(T-t)^3}. \tag{10.41}$$

Proof of Lemma 10.8.1. For $1 \leq l \leq d$, set $X_t^l := \ln S_t^l$ and $X_t := (X_t^1, ..., X_t^d)$: this defines the log-price process. We define the function $v(t, X)$ for $t \geq 0$ and

$X = (X^1, ..., X^d) \in \mathbb{R}^d$ by $v(t, X) = u(t, e^{X^1}, ..., e^{X^d})$. Then $v(t, X_t) = u(t, S_t)$. For notation simplicity, $(e^{X^1}, ..., e^{X^d})$ will be denoted by e^X. For $1 \leq l, m, n \leq d$,

$$\begin{aligned}
\partial_l^1 v(t, X) &= e^{X^l} \partial_l^1 u(t, e^X), \\
\partial_{lm}^2 v(t, X) &= e^{X^l} e^{X^m} \partial_{lm}^2 u(t, e^X) + e^{X^l} \partial_l^1 u(t, e^X) \mathbb{1}_{m=l}, \\
\partial_{lmn}^3 v(t, X) &= e^{X^l} e^{X^m} e^{X^n} \partial_{lmn}^3 u(t, e^X) + e^{X^l} e^{X^m} \partial_{lm}^2 u(t, e^X)(\mathbb{1}_{n=l} + \mathbb{1}_{n=m}) \\
&\quad + e^{X^l} e^{X^n} \partial_{ln}^2 u(t, e^X) \mathbb{1}_{m=l} + e^{X^l} \partial_l^1 u(t, e^X) \mathbb{1}_{n=m=l}.
\end{aligned}$$

Previous relations yield

$$\begin{aligned}
e^{X^l} \partial_l^1 u(t, e^X) &= \partial_l^1 v(t, X), \\
e^{X^l} e^{X^m} \partial_{lm}^2 u(t, e^X) &= \partial_{lm}^2 v(t, X) - \partial_l^1 v(t, X) \mathbb{1}_{m=l}, \\
e^{X^l} e^{X^m} e^{X^n} \partial_{lmn}^3 u(t, e^X) &= \partial_{lmn}^3 v(t, X) - \partial_{lm}^2 v(t, X)(\mathbb{1}_{n=l} + \mathbb{1}_{n=m}) \\
&\quad - \partial_{ln}^2 v(t, X) \mathbb{1}_{m=l} + 2\partial_l^1 v(t, X) \mathbb{1}_{n=m=l}.
\end{aligned} \tag{10.42}$$

We only prove (10.41), the proof is exactly the same for (10.39) and (10.40). From (10.42), we readily get

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}} \left| \bar{u}_{l,m,n}^{(3)}(t) \right|^2 \leq 16 \sup_{1 \leq j \leq d} |\sigma_j|^6 e^{-2rt} \big(&\mathbb{E}_{\mathbb{P}} \left| \partial_{lmn}^3 v(t, X_t) \right|^2 + \mathbb{E}_{\mathbb{P}} \left| \partial_{lm}^2 v(t, X_t) \right|^2 \\
&+ \mathbb{E}_{\mathbb{P}} \left| \partial_{ln}^2 v(t, X_t) \right|^2 + \mathbb{E}_{\mathbb{P}} \left| \partial_l^1 v(t, X_t) \right|^2 \big).
\end{aligned} \tag{10.43}$$

Now, write $v(t, x) = e^{-r(T-t)} \int_{\mathbb{R}^d} g\big(\cdots, e^{x^l + (r - \frac{1}{2}\sigma_l^2)(T-t) + \sigma_l y^l}, \cdots \big) p(t, x; T, y) \mathrm{d}y$, where $p(t, x; T, y)$ denotes the transition density function of the d-dimensional $\mathbb{Q}$-Brownian motion W. Thus we can explicitly differentiate this Gaussian density with respect to any component of x, and show that

$$\begin{aligned}
\partial_l^1 v(t, X_t) &= e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left(g(S_T) H_{t,T}^{(1)} \right), \\
\partial_{lm}^2 v(t, X_t) &= e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left(g(S_T) H_{t,T}^{(2)} \right), \\
\partial_{lmn}^3 v(t, X_t) &= e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}^{\mathcal{F}_t} \left(g(S_T) H_{t,T}^{(3)} \right)
\end{aligned}$$

(this is the usual representation of Greeks using Malliavin calculus weights, see Fournié *et al.* [FLL+99]). In the one dimensional case ($d = 1$), one has:

$$\begin{aligned}
H_{t,T}^{(1)} &= \frac{W_T - W_t}{T - t}, \\
H_{t,T}^{(2)} &= \frac{(W_T - W_t)^2}{(T - t)^2} - \frac{1}{T - t}, \\
H_{t,T}^{(3)} &= \frac{(W_T - W_t)^3}{(T - t)^3} - 3\frac{W_T - W_t}{T - t}.
\end{aligned}$$

In our more general setting, the random variables $H^{(1)}_{t,T}$, $H^{(2)}_{t,T}$ and $H^{(3)}_{t,T}$ are independent of $\mathcal{F}_t$, have zero mean and satisfy to the following estimates (for $q > 1$):

$$\mathbb{E}^{\mathcal{F}_t}_{\mathbb{Q}} \left| H^{(i)}_{t,T} \right|^q \leq \frac{C_q}{(T-t)^{i\frac{q}{2}}}, \quad 1 \leq i \leq 3. \tag{10.44}$$

To pass to $\mathbb{P}$-conditional expectations, we introduce $Z_t = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t}$ the Radon-Nikodym density of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ on $\mathcal{F}_t$. Therefore, using the zero-mean property of the weights $H^{(i)}_{t,T}$, one can write

$$\begin{aligned}
\partial^1_l v(t, X_t) &= e^{-r(T-t)} \mathbb{E}^{\mathcal{F}_t}_{\mathbb{Q}} \left[\left(g(S_T) - \mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} g(S_T) \right) H^{(1)}_{t,T} \right] \\
&= e^{-r(T-t)} \mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} \left[\left(g(S_T) - \mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} g(S_T) \right) H^{(1)}_{t,T} \frac{Z_T}{Z_t} \right], \\
\partial^2_{lm} v(t, X_t) &= e^{-r(T-t)} \mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} \left[\left(g(S_T) - \mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} g(S_T) \right) H^{(2)}_{t,T} \frac{Z_T}{Z_t} \right], \\
\partial^3_{lmn} v(t, X_t) &= e^{-r(T-t)} \mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} \left[\left(g(S_T) - \mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} g(S_T) \right) H^{(3)}_{t,T} \frac{Z_T}{Z_t} \right].
\end{aligned}$$

Thus, using (10.44) (together with the standard inequality $\mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} \left| \frac{Z_T}{Z_t} \right|^q \leq C_q$, for $q > 1$) and applying Cauchy-Schwartz inequality, one readily obtains

$$\begin{aligned}
\left| \partial^1_l v(t, X_t) \right|^2 &\leq C \frac{\mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} \left| g(S_T) - \mathbb{E}^{\mathcal{F}_t}_{\mathbb{P}} g(S_T) \right|^2}{T-t}, \\
\mathbb{E}_{\mathbb{P}} \left| \partial^1_l v(t, X_t) \right|^2 &\leq C \frac{V_{t,T}(g)}{(T-t)}.
\end{aligned} \tag{10.45}$$

Similarly,

$$\mathbb{E}_{\mathbb{P}} \left| \partial^2_{lm} v(t, X_t) \right|^2 \leq C \frac{V_{t,T}(g)}{(T-t)^2}, \quad \mathbb{E}_{\mathbb{P}} \left| \partial^3_{lmn} v(t, X_t) \right|^2 \leq C \frac{V_{t,T}(g)}{(T-t)^3}. \tag{10.46}$$

Then, (10.45), (10.46), together with (10.43), yield (10.41). □

Estimate (10.31). According to (10.28), one can write

$$R^{i,(3)}_{l,m,n}(t) = \bar{U}^{(2)}(t) + \bar{U}^{(1)}(t) - \sum_{0 \leq j < k \leq d} \left(\bar{U}^{C,j,k,(3)}(t) + \bar{U}^{C,j,k,(2)}(t) + \bar{U}^{C,j,k,(1)}(t) \right),$$

with

$$\bar{U}^{(2)}(t) := \big(\sigma_m \bar{u}^{(2)}_{l,m}(t)\mathbb{1}_{n=m} + \sigma_l \bar{u}^{(2)}_{l,m}(t)\mathbb{1}_{n=l} + \sigma_l \bar{u}^{(2)}_{l,n}(t)\mathbb{1}_{m=l}\big),$$
$$\bar{U}^{(1)}(t) := \sigma_l^2 \big(\bar{u}^{(1)}_l(t) - \frac{\bar{S}^l_t}{\bar{S}^l_{t_i}} \bar{u}^{(1)}_l(t_i)\big)\mathbb{1}_{n=m=l},$$
$$\bar{U}^{C,j,k,(3)}(t) := \delta^{j,k}_{t_i} \bar{C}^{j,k,(3)}_{l,m,n}(t),$$
$$\bar{U}^{C,j,k,(2)}(t) := \delta^{j,k}_{t_i} \big(\sigma_m \bar{C}^{j,k,(2)}_{l,m}(t)\mathbb{1}_{n=m} + \sigma_l \bar{C}^{j,k,(2)}_{l,m}(t)\mathbb{1}_{n=l} + \sigma_l \bar{C}^{j,k,(2)}_{l,n}(t)\mathbb{1}_{m=l}\big),$$
$$\bar{U}^{C,j,k,(1)}(t) := \sigma_l^2 \delta^{j,k}_{t_i} \big(\bar{C}^{j,k,(1)}_l(t) - \frac{\bar{S}^l_t}{\bar{S}^l_{t_i}} \bar{C}^{j,k,(1)}_l(t_i)\big)\mathbb{1}_{n=m=l}.$$

Then,

$$\mathbb{E}_{\mathbb{P}} \left| R^{i,(3)}_{l,m,n}(t) \right|^2 \leq_c \left(\mathbb{E}_{\mathbb{P}} \left| \bar{U}^{(2)}(t) \right|^2 + \mathbb{E}_{\mathbb{P}} \left| \bar{U}^{(1)}(t) \right|^2 \right)$$
$$+ \sup_{0 \leq j < k \leq d} \left(\mathbb{E}_{\mathbb{P}} \left| \bar{U}^{C,j,k,(3)}(t) \right|^2 + \mathbb{E}_{\mathbb{P}} \left| \bar{U}^{C,j,k,(2)}(t) \right|^2 + \mathbb{E}_{\mathbb{P}} \left| \bar{U}^{C,j,k,(1)}(t) \right|^2 \right). \quad (10.47)$$

Let us upper bound each term appearing in (10.47). From Lemma 10.8.1, one has

$$\mathbb{E}_{\mathbb{P}} \left| \bar{U}^{(2)}(t) \right|^2 \leq 3C(\sigma_l^2 + \sigma_m^2) \frac{V_{t,T}(g)}{(T-t)^2}, \quad (10.48)$$

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left| \bar{U}^{(1)}(t) \right|^2 &\leq 2\sigma_l^4 \mathbb{E}_{\mathbb{P}} \left| \bar{u}^{(1)}_l(t) \right|^2 + 2\sigma_l^4 \mathbb{E}_{\mathbb{P}} \big(\left| \bar{u}^{(1)}_l(t_i) \right|^2 \mathbb{E}^{\mathcal{F}_{t_i}}_{\mathbb{P}} \left| \frac{\bar{S}^l_t}{\bar{S}^l_{t_i}} \right|^2 \big) \\ &\leq_c \mathbb{E}_{\mathbb{P}} \left| \bar{u}^{(1)}_n(t) \right|^2 + \mathbb{E}_{\mathbb{P}} \left| \bar{u}^{(1)}_n(t_i) \right|^2 \\ &\leq_c \frac{V_{t,T}(g)}{(T-t)}, \end{aligned} \quad (10.49)$$

where we have used the fact that $\mathbb{E}^{\mathcal{F}_{t_i}}_{\mathbb{P}} \left| \frac{\bar{S}^n_t}{\bar{S}^n_{t_i}} \right|^2 = \mathbb{E}_{\mathbb{P}} \left| \frac{\bar{S}^n_t}{\bar{S}^n_{t_i}} \right|^2 \leq C$.

To handle the terms $\bar{U}^{C,j,k,(1)}(t)$, $\bar{U}^{C,j,k,(2)}(t)$, $\bar{U}^{C,j,k,(3)}(t)$ in (10.47), we need extra intermediate results, that we present as lemmas.

Lemma 10.8.2. *For $1 \leq l, m, n \leq d$ and $0 \leq t < T$, one has*

$$\mathbb{E}_{\mathbb{P}} \left| \bar{u}^{(1)}_l(t) \right|^{2p_0} \leq C \frac{\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0}}{(T-t)^{p_0}}, \qquad \mathbb{E}_{\mathbb{P}} \left| \bar{u}^{(2)}_{l,m}(t) \right|^{2p_0} \leq C \frac{\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0}}{(T-t)^{2p_0}}.$$

The proof of this lemma is very similar to that of Lemma 10.8.1. One has just to substitute the $\mathbf{L}_{2p_0}$-norm for the $\mathbf{L}_2$-norm and $\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0}$ for $V_{t,T}(g)$. We skip details.

Lemma 10.8.3. *For $0 \le j < k \le d$, $(l,m,n) \in \{(j,j,j),(j,j,k),(j,k,k),(k,k,k)\}$, $p > 1$ and $0 \le t_i \le t < t_{i+1} \le T$, one has*

$$\mathbb{E}_{\mathbb{P}} \left| \frac{\bar{C}^{j,k,(2)}_{l,m}(t)}{\bar{C}^{j,k,(2)}_{l,n}(t)} \right|^p \le C. \tag{10.50}$$

Proof of Lemma 10.8.3. Estimate (10.50) is a direct consequence from (10.63), (10.64), (10.65), together with $\frac{1}{\sqrt{T_2-t}} \le \frac{1}{\sqrt{T_2-T}}$ and the fact that

$$\mathbb{E}_{\mathbb{P}} \left| S^j_t \right|^q + \mathbb{E}_{\mathbb{P}} \left| S^j_t \right|^{-q} \le C_q \tag{10.51}$$

for $0 \le j \le d$, $t \in [0,T]$ and $q > 0$. □

Lemma 10.8.4. *For $0 \le j < k \le d$, $(l,m,n) \in \{(j,j,j),(j,j,k),(j,k,k),(k,k,k)\}$ and $p > 1$, there exists a positive constant $\pi^{\text{threshold}}$ (which depends on p, T and T_2) such that, for any time net whose size $|\pi| \le \pi^{\text{threshold}}$ and for any $0 \le t_i \le t < t_{i+1} \le T$,*

$$\mathbb{E}_{\mathbb{P}} \left| \frac{\bar{C}^{j,k,(2)}_{l,m}(t)}{\bar{C}^{j,k,(2)}_{l,m}(t_i)} \right|^p \le C, \tag{10.52}$$

$$\mathbb{E}_{\mathbb{P}} \left| \frac{\bar{C}^{j,k,(3)}_{l,m,n}(t)}{\bar{C}^{j,k,(2)}_{l,m}(t_i)} \right|^p \le C. \tag{10.53}$$

Proof of Lemma 10.8.4. We only detail the proof of (10.52), the proof of (10.53) is similar. The proof is divided into two steps. We first establish intermediate controls, then we complete the proof.

Step 1. In the following, q is a real number greater than p. One has

$$\begin{aligned}
d^{j,k}_1(t,S_t) &= \frac{1}{\sigma_{j,k}\sqrt{T_2-t}} \ln\left(\frac{S^k_t}{K_{j,k}S^j_t}\right) + \frac{1}{2}\sigma_{j,k}\sqrt{T_2-t} \\
&= \frac{1}{\sigma_{j,k}\sqrt{T_2-t}} \left(\ln s^k_0 + (\mu_k - \frac{1}{2}\sigma^2_k)t + \sigma_k \hat{W}^k_t\right) \\
&- \frac{1}{\sigma_{j,k}\sqrt{T_2-t}} \left(\ln K_{j,k} + \ln s^j_0 + (\mu_j - \frac{1}{2}\sigma^2_j)t + \sigma_j \hat{W}^j_t\right) + \frac{1}{2}\sigma_{j,k}\sqrt{T_2-t} \\
&= \frac{1}{\sigma_{j,k}\sqrt{T_2-t}} \left(\sigma_{j,k}\hat{W}^{j,k}_t - m^{j,k}(t)\right),
\end{aligned}$$

where $\sigma_{j,k} = \sqrt{\sigma^2_j + \sigma^2_k - 2\rho_{j,k}\sigma_j\sigma_k}$ (see 10.62), and

$$\begin{aligned}
\hat{W}^{j,k}_t &:= (\sigma_k \hat{W}^k_t - \sigma_j \hat{W}^j_t)/\sigma_{j,k}, \\
m^{j,k}(t) &:= -\left(\ln \frac{s^k_0}{s^j_0 K_{j,k}} + \left(\mu_k - \mu_j + \frac{1}{2}\sigma^2_j - \frac{1}{2}\sigma^2_k\right)t\right) + \frac{1}{2}\sigma^2_{j,k}(T_2 - t).
\end{aligned}$$

Note that $(\hat{W}_t^{j,k})_t$ is a $\mathbb{P}$-Brownian motion. We will denote $e^{-\frac{q}{2}\left|d_1^{j,k}(t,S_t)\right|^2}$ by $\phi^{j,k,q}(t)$. Then,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_{t_i}}\left(\phi^{j,k,q}(t)\right) &= \frac{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}}{\sqrt{q}}\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_{t_i}}\left(\frac{\sqrt{q}}{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}}\exp-\frac{\left(\sigma_{j,k}\hat{W}_t^{j,k}-m^{j,k}(t)\right)^2}{2\left(\sigma_{j,k}\sqrt{T_2-t}\right)^2/q}\right)\\ &= \frac{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}}{\sqrt{q}}(f_X * f_Y)(m^{j,k}(t)-\sigma_{j,k}\hat{W}_{t_i}^{j,k}),\end{aligned}$$

where f_X and f_Y are the respective probability density functions of two independent random variables X and Y having normal distributions $\mathcal{N}\big(0;\frac{\left(\sigma_{j,k}\sqrt{T_2-t}\right)^2}{q}\big)$ and $\mathcal{N}\big(0;\sigma_{j,k}^2(t-t_i)\big)$. Therefore,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_{t_i}}\left(\phi^{j,k,q}(t)\right) &= \frac{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}}{\sqrt{q}}f_{X+Y}(m^{j,k}(t)-\sigma_{j,k}\hat{W}_{t_i}^{j,k})\\ &= c(t_i,t)\exp\frac{-q\left(\sigma_{j,k}\hat{W}_{t_i}^{j,k}-m^{j,k}(t)\right)^2}{2\left(\sigma_{j,k}^2(T_2-t)+q\sigma_{j,k}^2(t-t_i)\right)}\end{aligned}$$

with

$$c(t_i,t) := \frac{\sigma_{j,k}\sqrt{T_2-t}}{\sqrt{\sigma_{j,k}^2(T_2-t)+q\sigma_{j,k}^2(t-t_i)}},$$

which is uniformly bounded by 1. Then

$$\frac{\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_{t_i}}\left(\phi^{j,k,q}(t)\right)}{\phi^{j,k,q}(t_i)} = c(t_i,t)e^{\varphi(t_i,t,\hat{W}_{t_i}^{j,k})},$$

where, for $0\le t_i\le t\le t_{i+1}\le T$ and $y\in\mathbb{R}$,

$$\begin{aligned}\varphi(t_i,t,y) &:= \frac{-q\left(\sigma_{j,k}y-m^{j,k}(t)\right)^2}{2\left(\sigma_{j,k}^2(T_2-t)+q\sigma_{j,k}^2(t-t_i)\right)}+\frac{q\left(\sigma_{j,k}y-m^{j,k}(t_i)\right)^2}{2\sigma_{j,k}^2(T_2-t_i)}\\ &= c_2(t_i,t)y^2+c_1(t_i,t)y+c_0(t_i,t),\end{aligned}$$

and $c_2(t_i,t)$, $c_1(t_i,t)$, $c_0(t_i,t)$ are uniformly bounded and deterministic, with

$$c_2(t_i,t) := \frac{q}{2(T_2-t_i)}-\frac{q}{2\left(T_2-t+q(t-t_i)\right)}.$$

Then, the quantity $\mathbb{E}_{\mathbb{P}}\left[\frac{\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_{t_i}}\left(\phi^{j,k,q}(t)\right)}{\phi^{j,k,q}(t_i)}\right]$ is finite (and uniformly bounded also) for all $t_i\le t\le t_{i+1}\le T$ if and only if

$$c_2(t_i,t)-\frac{1}{2t_i}<0, \forall t_i\le t\le t_{i+1}\le T. \tag{10.54}$$

From the expression of $c_2(t_i,t)$ taken at $t = t_{i+1}$, we see that, with a mesh size $|\pi| \geq (t_{i+1} - t_i)$ large enough, the condition (10.54) may fail. A sufficient condition on $|\pi|$ is $|\pi| \leq \frac{(T_2-T)^2}{Tq(q-1)} := \pi^{\text{threshold}}$. Indeed, under this condition, we have

$$c_2(t_i,t) = \frac{q(q-1)(t-t_i)}{2(T_2-t_i)\left(T_2-t+q(t-t_i)\right)} < \frac{q(q-1)|\pi|}{2(T_2-T)^2} \leq \frac{1}{2T} \leq \frac{1}{2t_i}.$$

Step 2. Now, it is clear from (10.63), (10.64) and (10.65), together with (10.51) that, for $(l,m) \in \{(j,j),(j,k),(k,k)\}$, one can apply Hölder's inequality to obtain

$$\mathbb{E}_{\mathbb{P}}\left|\frac{\bar{C}^{j,k,(2)}_{l,m}(t)}{\bar{C}^{j,k,(2)}_{l,m}(t_i)}\right|^p \leq C\left(\mathbb{E}_{\mathbb{P}}\left(\frac{\phi^{j,k,q}(t)}{\phi^{j,k,q}(t_i)}\right)\right)^{\frac{p}{q}} = C\left(\mathbb{E}_{\mathbb{P}}\left[\frac{\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_{t_i}}\left(\phi^{j,k,q}(t)\right)}{\phi^{j,k,q}(t_i)}\right]\right)^{\frac{p}{q}}$$

(where $1 < p < q$), which has just been shown to be finite under the condition above on the time net size. The lemma is proved. □

Completion of the upper bounds of (10.47). Let $0 \leq j < k \leq d$. If $j \neq 0$, use (10.22) to obtain

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left|\bar{U}^{C,j,k,(3)}(t)\right|^2 &= \mathbb{E}_{\mathbb{P}}\left|\delta^{j,k}_{t_i}\bar{C}^{j,k,(3)}_{l,m,n}(t)\right|^2 = \mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(2)}_{j,k}(t_i)\frac{\bar{C}^{j,k,(3)}_{l,m,n}(t)}{\bar{C}^{j,k,(2)}_{j,k}(t_i)}\right|^2 \\ &\leq \left(\mathbb{E}_{\mathbb{P}}\left|\bar{u}^{(2)}_{j,k}(t_i)\right|^{2p_0}\right)^{\frac{1}{p_0}}\left(\mathbb{E}_{\mathbb{P}}\left|\frac{\bar{C}^{j,k,(3)}_{l,m,n}(t)}{\bar{C}^{j,k,(2)}_{j,k}(t_i)}\right|^{2q}\right)^{\frac{1}{q}},\end{aligned}$$

where p_0 and q are conjugate real numbers. Thus, from Lemma 10.8.2 and Lemma 10.8.4, and for a time net π whose size is sufficiently small, we readily obtain

$$\mathbb{E}_{\mathbb{P}}\left|\bar{U}^{C,j,k,(3)}(t)\right|^2 \leq C\frac{\left(\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0}\right)^{\frac{1}{p_0}}}{(T-t)^2}.$$

If $j=0$, by using (10.23) we derive

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\left|\bar{U}^{C,0,k,(3)}(t)\right|^2 = \mathbb{E}_{\mathbb{P}}\left|\delta_{t_i}^{0,k}\bar{C}_{l,m,n}^{0,k,(3)}(t)\right|^2 \\
&= \mathbb{E}_{\mathbb{P}}\left|\frac{\bar{C}_{l,m,n}^{0,k,(3)}(t)}{\bar{C}_{k,k}^{0,k,(2)}(t_i)}\bar{u}_{k,k}^{(2)}(t_i) - \frac{\bar{C}_{l,m,n}^{0,k,(3)}(t)}{\bar{C}_{k,k}^{0,k,(2)}(t_i)}\sum_{1\le j_1<j_2\le d}\frac{\bar{C}_{k,k}^{j_1,j_2,(2)}(t_i)}{\bar{C}_{j_1,j_2}^{j_1,j_2,(2)}(t_i)}\bar{u}_{j_1,j_2}^{(2)}(t_i)\right|^2 \\
&\le 2\left(\mathbb{E}_{\mathbb{P}}\left|\frac{\bar{C}_{l,m,n}^{0,k,(3)}(t)}{\bar{C}_{k,k}^{0,k,(2)}(t_i)}\right|^{2q}\right)^{\frac{1}{q}}\left(\mathbb{E}_{\mathbb{P}}\left|\bar{u}_{k,k}^{(2)}(t_i)\right|^{2p_0}\right)^{\frac{1}{p_0}} \qquad (10.55) \\
&+ 2d^4\left(\mathbb{E}_{\mathbb{P}}\left|\frac{\bar{C}_{l,m,n}^{0,k,(3)}(t)}{\bar{C}_{k,k}^{0,k,(2)}(t_i)}\right|^{2q_1}\right)^{\frac{1}{q_1}} \qquad (10.56) \\
&\times \sup_{1\le j_1<j_2\le d}\left\{\left(\mathbb{E}_{\mathbb{P}}\left|\frac{\bar{C}_{k,k}^{j_1,j_2,(2)}(t_i)}{\bar{C}_{j_1,j_2}^{j_1,j_2,(2)}(t_i)}\right|^{2q_2}\right)^{\frac{1}{q_2}}\left(\mathbb{E}_{\mathbb{P}}\left|\bar{u}_{j_1,j_2}^{(2)}(t_i)\right|^{2p_0}\right)^{\frac{1}{p_0}}\right\}, \qquad (10.57)
\end{aligned}$$

where p_0, q, q_1 and q_2 are positive real numbers such that $\frac{1}{p_0}+\frac{1}{q}=\frac{1}{p_0}+\frac{1}{q_1}+\frac{1}{q_2}=1$. Now, we apply Lemma 10.8.4 and Lemma 10.8.2 for term (10.55), Lemma 10.8.4 for term (10.56) and Lemmas 10.8.3 and 10.8.2 for term (10.57) to conclude that, for $|\pi|$ small enough, $\mathbb{E}_{\mathbb{P}}\left|\bar{U}^{C,0,k,(3)}(t)\right|^2$ has the same bound as $\mathbb{E}_{\mathbb{P}}\left|\bar{U}^{C,j,k,(3)}(t)\right|^2$ with $j\neq 0$. Then, for every $0\le j<k\le d$,

$$\mathbb{E}_{\mathbb{P}}\left|\bar{U}^{C,j,k,(3)}(t)\right|^2 \le C\frac{\left(\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0}\right)^{\frac{1}{p_0}}}{(T-t)^2}. \qquad (10.58)$$

The same arguments yield, for every $0\le j<k\le d$,

$$\mathbb{E}_{\mathbb{P}}\left|\bar{U}^{C,j,k,(2)}(t)\right|^2 \le C\frac{\left(\mathbb{E}_{\mathbb{P}}\left|g(S_T)\right|^{2p_0}\right)^{\frac{1}{p_0}}}{(T-t)^2}. \qquad (10.59)$$

To handle $\bar{U}^{C,j,k,(1)}(t)$, we need to proceed a bit differently. We take advantage of the linear SDE satisfied by $\bar{U}^{C,j,k,(1)}$. Indeed, using (10.12), one has

$$\begin{aligned}
\mathrm{d}\bar{U}^{C,j,k,(1)}(t) &= \sigma_l^2\delta_{t_i}^{j,k}\mathbb{1}_{n=m=l}\left(\mathrm{d}\bar{C}_l^{j,k,(1)}(t) - \frac{\bar{C}_l^{j,k,(1)}(t_i)}{\bar{S}_{t_i}^l}\mathrm{d}\bar{S}_t^l\right) \\
&= \sigma_l^2\delta_{t_i}^{j,k}\mathbb{1}_{n=m=l}\sum_{m\prime=1}^{d}\left(\bar{C}_{l,m\prime}^{j,k,(2)}(t) + \sigma_l\bar{C}_l^{j,k,(1)}(t)\mathbb{1}_{m\prime=l}\right)\mathrm{d}W_t^{m\prime} \\
&- \sigma_l^2\delta_{t_i}^{j,k}\mathbb{1}_{n=m=l}\frac{\bar{C}_l^{j,k,(1)}(t_i)}{\bar{S}_{t_i}^l}\sigma_l\bar{S}_t^l\mathrm{d}W_t^l \\
&= \mathrm{d}\bar{V}^{C,j,k,(2)}(t) + \sigma_l\bar{U}^{C,j,k,(1)}(t)\mathrm{d}W_t^l,
\end{aligned}$$

where

$$\bar{V}^{C,j,k,(2)}(t) := \sigma_l^2 \delta_{t_i}^{j,k} \mathbb{1}_{n=m=l} \sum_{m\prime=1}^{d} \int_{t_i}^{t} \bar{C}_{l,m\prime}^{j,k,(2)}(s) \mathrm{d}W_s^{m\prime}.$$

Since $\bar{U}^{C,j,k,(1)}(t_i) = 0$, we obtain

$$\mathbb{E}_{\mathbb{P}} \left| \bar{U}^{C,j,k,(1)}(t) \right|^2 \leq 2\mathbb{E}_{\mathbb{P}} \left| \bar{V}^{C,j,k,(2)}(t) \right|^2 + 2\sigma_l^2 \int_{t_i}^{t} \mathbb{E}_{\mathbb{P}} \left| \bar{U}^{C,j,k,(1)}(s) \right|^2 \mathrm{d}s. \tag{10.60}$$

Following the ideas used to establish (10.29), we prove that

$$\mathbb{E}_{\mathbb{P}} \left| \bar{V}^{C,j,k,(2)}(t) \right|^2 \leq_c \sum_{m\prime=1}^{d} \int_{t_i}^{t} \mathbb{E}_{\mathbb{P}} \left| \delta_{t_i}^{j,k} \bar{C}_{l,m\prime}^{j,k,(2)}(s) \right|^2 \mathrm{d}s.$$

Then, the same arguments used for the proof of (10.58) and (10.59) yield

$$\mathbb{E}_{\mathbb{P}} \left| \bar{V}^{C,j,k,(2)}(t) \right|^2 \leq_c \int_{t_i}^{t} \frac{\left(\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0} \right)^{\frac{1}{p_0}}}{(T-s)^2} \mathrm{d}s \leq_c \frac{\left(\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0} \right)^{\frac{1}{p_0}}}{(T-t)^2} |\pi|.$$

Therefore, by invoking Gronwall's lemma in (10.60), we get

$$\mathbb{E}_{\mathbb{P}} \left| \bar{U}^{C,j,k,(1)}(t) \right|^2 \leq_c \frac{\left(\mathbb{E}_{\mathbb{P}} \left| g(S_T) \right|^{2p_0} \right)^{\frac{1}{p_0}}}{(T-t)^2} |\pi|.$$

Plugging this estimate, the inequalities (10.48), (10.49), (10.58) and (10.59) into (10.47), we complete the proof of (10.31) and that of Proposition 10.6.1. □

10.A Appendix

10.A.1 Margrabe formula

The following result is the so-called Margrabe formula (cf Margrabe (1978)): it is a generalization of the Black-Scholes formula, and gives an explicit expression for the Call and Exchange options.

Proposition 10.A.1 (The Margrabe Formula)**.** *For $t \in [0, T_2)$ and positive initial spot values S^j, S^k, one has*

$$C^{j,k}(t, S^j, S^k) = S^k \mathcal{N}(d_1^{j,k}(t,S)) - K_{j,k} S^j \mathcal{N}(d_2^{j,k}(t,S)), \tag{10.61}$$

where $\mathcal{N}(.)$ is the cumulative distribution function of the standard Gaussian distribution, and

$$\begin{aligned} d_1^{j,k}(t,S) &:= \frac{1}{\sigma_{j,k}\sqrt{T_2 - t}} \ln\left(\frac{S^k}{K_{j,k} S^j} \right) + \frac{1}{2}\sigma_{j,k}\sqrt{T_2 - t}, \\ d_2^{j,k}(t,S) &:= \frac{1}{\sigma_{j,k}\sqrt{T_2 - t}} \ln\left(\frac{S^k}{K_{j,k} S^j} \right) - \frac{1}{2}\sigma_{j,k}\sqrt{T_2 - t}, \\ \sigma_{j,k} &:= \sqrt{\sigma_j^2 + \sigma_k^2 - 2\rho_{j,k}\sigma_j\sigma_k}. \end{aligned} \tag{10.62}$$

We note that $|d_2^{j,k}(t,S)|^2 = |d_1^{j,k}(t,S)|^2 - 2\ln\left(\frac{S^k}{K_{j,k}S^j}\right)$, or equivalently, $e^{-\frac{\left|d_2^{j,k}(t,S)\right|^2}{2}} = \frac{S^k}{K_{j,k}S^j} e^{-\frac{\left|d_1^{j,k}(t,S)\right|^2}{2}}$. Then, straightforward computations give the following derivatives expressions for $C^{j,k}$.

a) For $0 \le j < k \le d$,

$$\partial_k^1 C^{j,k}(t,S^j,S^k) = \mathcal{N}(d_1^{j,k}(t,S));$$

$$\partial_{kk}^2 C^{j,k}(t,S^j,S^k) = \frac{1}{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}} \frac{1}{S^k} e^{-\frac{\left|d_1^{j,k}(t,S)\right|^2}{2}}; \tag{10.63}$$

$$\partial_{kkk}^3 C^{j,k}(t,S^j,S^k) = \frac{1}{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}} \frac{1}{|S^k|^2} \left(-1 - \frac{d_1^{j,k}(t,S)}{\sigma_{j,k}\sqrt{T_2-t}}\right) e^{-\frac{\left|d_1^{j,k}(t,S)\right|^2}{2}}.$$

b) For $1 \le j < k \le d$,

$$\partial_j^1 C^{j,k}(t,S^j,S^k) = -K_{j,k}\mathcal{N}(d_2^{j,k}(t,S));$$

$$\partial_{jj}^2 C^{j,k}(t,S^j,S^k) = \frac{1}{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}} \frac{S^k}{|S^j|^2} e^{-\frac{\left|d_1^{j,k}(t,S)\right|^2}{2}}; \tag{10.64}$$

$$\partial_{jk}^2 C^{j,k}(t,S^j,S^k) = -\frac{1}{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}} \frac{1}{S^j} e^{-\frac{\left|d_1^{j,k}(t,S)\right|^2}{2}}; \tag{10.65}$$

$$\partial_{jjk}^3 C^{j,k}(t,S^j,S^k) = -\frac{1}{\sqrt{2\pi}\sigma_{j,k}^2(T_2-t)} \frac{d_2^{j,k}(t,S)}{|S^j|^2} e^{-\frac{\left|d_1^{j,k}(t,S)\right|^2}{2}};$$

$$\partial_{jkk}^3 C^{j,k}(t,S^j,S^k) = \frac{1}{\sqrt{2\pi}\sigma_{j,k}^2(T_2-t)} \frac{d_1^{j,k}(t,S)}{S^j S^k} e^{-\frac{\left|d_1^{j,k}(t,S)\right|^2}{2}};$$

$$\partial_{jjj}^3 C^{j,k}(t,S^j,S^k) = \frac{1}{\sqrt{2\pi}\sigma_{j,k}\sqrt{T_2-t}} \frac{S^k}{|S^j|^3} \left(-1 + \frac{d_2^{j,k}(t,S)}{\sigma_{j,k}\sqrt{T_2-t}}\right) e^{-\frac{\left|d_1^{j,k}(t,S)\right|^2}{2}}.$$

10.A.2 Proof of Lemma 10.5.1

• Suppose that g is Lipschitz (including the case of Call/Put payoffs). Then observe that

$$\begin{aligned} V_{t,T}(g) = \mathbb{E}_{\mathbb{P}}|g(S_T) - \mathbb{E}_{\mathbb{P}}^{\mathcal{F}_t}(g(S_T))|^2 &\le \mathbb{E}_{\mathbb{P}}|g(S_T) - g(S_t)|^2 \\ &\le C\mathbb{E}_{\mathbb{P}}|S_T - S_t|^2 \le C(T-t) \end{aligned}$$

proving that $g \in \mathbf{L}_{2,\alpha}$ with $\alpha = 1$. The first item of Lemma 10.5.1 is proved.
• Now, consider $g(S) = \mathbb{1}_{S \ge K}$ and a single log-normal asset S. Then, passing to the log variables, we set

$$\begin{aligned} v(t,x) &= \mathbb{P}(\log(S_T) \ge \log(K) | \log(S_t) = x) \\ &= \mathcal{N}\left(\frac{1}{\sigma\sqrt{T-t}} \log\left(\frac{e^x}{Ke^{-\mu(T-t)}}\right) - \frac{1}{2}\sigma\sqrt{T-t}\right). \end{aligned}$$

An application of Itô formula to the $\mathbb{P}$-martingale $[v(s, \log(S_s)) = \mathbb{P}^{\mathcal{F}_s}(S_T \geq K)]_{0 \leq s \leq T}$ leads to

$$V_{t,T}(g) = \int_t^T \mathbb{E}_{\mathbb{P}}\Big(\frac{1}{2\pi(T-s)} e^{-(\frac{1}{\sigma\sqrt{T-s}} \log(\frac{S_s}{Ke^{-\mu(T-s)}}) - \frac{1}{2}\sigma\sqrt{T-s})^2}\Big) \mathrm{d}s.$$

Then, standard computations give $V_{t,T}(g) \leq \int_t^T \frac{C}{\sqrt{T-s}} \mathrm{d}s \leq C\sqrt{T-t}$, which shows that $g \in \mathbf{L}_{2,\alpha}$ with $\alpha = 1/2$.

• The stability by summation is obvious. Regarding the stability by product, write

$$\begin{aligned} V_{t,T}(g) &= \mathbb{E}_{\mathbb{P}}|g_1(S_T)g_2(S_T) - \mathbb{E}_{\mathbb{P}}^{\mathcal{F}_t}(g_1(S_T)g_2(S_T))|^2 \\ &\leq \mathbb{E}_{\mathbb{P}}|g_1(S_T)g_2(S_T) \pm g_1(S_T)\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_t}(g_2(S_T)) - \mathbb{E}_{\mathbb{P}}^{\mathcal{F}_t}(g_1(S_T))\mathbb{E}_{\mathbb{P}}^{\mathcal{F}_t}(g_2(S_T))|^2 \\ &\leq 2|g_1|_\infty^2 V_{t,T}(g_2) + 2|g_2|_\infty^2 V_{t,T}(g_1). \end{aligned}$$

Our statement readily follows from this. $\square$.

Bibliography

[AH08] M. Arnsdorf and I. Halperin. BSLP: Markovian Bivariate Spread-Loss Model for Portfolio Credit Derivatives. *J. Comput. Finance*, 12(2):77–107, 2008.

[AS64] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, 1964.

[AS93] R.W. Abernathy and R.P. Smith. Applying series expansion to the inverse beta distribution to find percentiles of the F-distribution. *ACM Trans. Math. Softw.*, 19(4):474–480, 1993.

[ASB03] L. Andersen, J. Sidenius, and S. Basu. All your hedges in one basket. *Risk November*, 16(11):67–72, 2003.

[Avi09] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13(3):381–401, 2009.

[BC08] B. Bouchard and J-F. Chassagneux. Discrete-time approximation for continuously and discretely reflected BSDEs. *Stochastic Processes Appl.*, 118(12):2269–2293, 2008.

[BCJ08] T.R. Bielecki, S. Crépey, and M. Jeanblanc. Up and down credit risk. 2008. Working paper. Available at `http://www.maths.univ-evry.fr/pages_perso/crepey/papers/up_and_down_credit_risk.pdf`.

[BDH+03] Ph. Briand, B. Delyon, Y. Hu, E. Pardoux, and L. Stoica. L^p solutions of backward stochastic differential equations. *Stochastic Processes and their Applications*, 108(1):109–129, 2003.

[BE08] B. Bouchard and R. Elie. Discrete-time approximation of decoupled forward-backward SDE with jumps. *Stochastic Processes and their Applications*, 118(1):53–75, 2008.

[BET04] B. Bouchard, I. Ekeland, and N. Touzi. On the Malliavin approach to Monte Carlo approximation of conditional expectations. *Finance Stoch.*, 8(1):45–71, 2004.

[BH06] Ph. Briand and Y. Hu. BSDE with quadratic growth and unbounded terminal value. *Probab. Theory Relat. Fields*, 136(4):604–618, 2006.

[BHW04] K. Burnecki, W. Härdle, and R. Weron. An introduction to simulation of risk processes. Research note, 2004. Available at `http://www.im.pwr.wroc.pl/~hugo/publ/BurneckiHaerdleRWeron_HSC_03_1.pdf`.

[Bis73] J.M. Bismut. Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl.*, 44:383–404, 1973.

[BLM07] Ph. Briand, J-P. Lepeltier, and J. San Martin. One-dimensional backward stochastic differential equations whose coefficient is monotonic in y and non-Lipschitz in z. *Bernoulli*, 13(1):80–91, 2007.

[BM08] B. Bouchard and S. Menozzi. Strong approximations of BSDEs in a domain. 2008. To appear in Bernoulli. Available on HAL preprint server.

[BP03] V. Bally and G. Pagès. Error analysis of the optimal quantization algorithm for obstacle problems. *Stochastic Processes and their Applications*, 106(1):1–40, 2003.

[BS73] F. Black and M.S. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[BT04] B. Bouchard and N. Touzi. Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications*, 111:175–206, 2004.

[BW08] M. Brodén and M. Wiktorsson. On the convergence of higher order hedging schemes. *Preprint*, 2008.

[CC09] R. Carmona and S. Crépey. Importance sampling and interacting particle systems for the estimation of Markovian credit portfolios loss distributions. 2009. Available at `http://grozny.maths.univ-evry.fr/pages_perso/crepey/papers/V-3.pdf`.

[CFV09] R. Carmona, J.-P. Fouque, and D. Vestal. Interacting particle systems for the computation of rare credit portfolio losses. *Finance and Stochastics*, 13(4):613–633, 2009.

[Cha09] J-F. Chassagneux. A discrete-time approximation for doubly reflected BSDEs. *Adv. Appl. Probab.*, 41(1):101–130, 2009.

[Che97] D. Chevance. Numerical methods for backward stochastic differential equations. In Numerical Methods in Finance (Rogers, L. C. G. (ed.) *et al.*). Cambridge Univ. Press. 232-244., 1997.

[CM08] R. Cont and A. Minca. Recovering portfolio default intensities implied by CDO quotes. 2008. Available on SSRN preprint server.

[Cou08] A. Cousin. *Analyse du risque et couverture des tranches de CDO synthétiques.* PhD thesis, Université Claude Bernard Lyon I, 2008.

[CS08] R. Cont and I. Savescu. Forward equations for portfolio credit derivatives. 2008. Available on SSRN preprint server.

[Das05] S. Das. *Credit Derivatives: CDOs and Structured Credit Products (Wiley Finance).* Wiley, 2005.

[DE92] D. Duffie and L.G. Epstein. Stochastic differential utility. *Econometrica*, 60(2):353–394, 1992.

[Den08] E. Denis. *Financial Markets with Transaction Costs; Leland's Strategies and Arbitrage.* PhD thesis, University of Besançon, 2008.

[Dev86] L. Devroye. *Non-Uniform Random Variate Generation.* Springer-Verlag, New York (original publisher), 1986.

[DG01] D. Duffie and N. Gârleanu. Risk and valuation of Collateralized Debt Obligations. *Financial Analysts Journal*, 57(1):41–59, 2001.

[DGT09] X. Ding, K. Giesecke, and P.I. Tomecek. Time-Changed Birth Processes and Multi-Name Credit Derivatives. *Operations Research*, 2009.

[DL01] M. Davis and V. Lo. Infectious Defaults. *Quantitative Finance*, 1(4):382–387, 2001.

[DM06] F. Delarue and S. Menozzi. A forward-backward stochastic algorithm for quasi-linear PDES. *Ann. Appl. Probab.*, 16(1):140–184, 2006.

[Duf] D.J. Duffy. Numerical analysis of jump diffusion models: A partial differential equation approach. Research report. Available at `www.datasimfinancial.com/UserFiles/articles/PIDE.pdf`.

[Duf06] D.J. Duffy. *Finite difference methods in financial engineering. A partial differential equation approach.* Wiley Finance Series. 423 p., 2006.

[EGG09] E. Errais, K. Giesecke, and L.R. Goldberg. Affine point processes and portfolio credit risk. 2009. Working paper. Available on SSRN preprint server.

[FB08] R. Frey and J. Backhaus. Pricing and hedging of portfolio credit derivatives with interacting default intensities. *International Journal of Theoretical and Applied Finance (IJTAF)*, 11(06):611–634, 2008.

[FLL+99] 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*, 3(4):391–412, 1999.

[Fri64] A. Friedman. *Partial differential equations of parabolic type.* Prentice-Hall, 1964.

[Fun92] D. Funaro. *Polynomial approximation of differential equations.* Lecture Notes in Physics. New Series m: Monographs. 8. Berlin: Springer- Verlag. x, 305 p., 1992.

[Gei02] S. Geiss. Quantitative approximation of certain stochastic integrals. *Stoch. Stoch. Rep.*, 73(3-4):241–270, 2002.

[GG04] C. Geiss and S. Geiss. On approximation of a class of stochastic integrals and interpolation. *Stochastics Stochastics Rep.*, 76(4):339–362, 2004.

[GG09] K. Giesecke and L.R. Goldberg. A top-down approach to multi-name credit. 2009. Working paper. Available on SSRN preprint server.

[GH07] S. Geiss and M. Hujo. Interpolation and approximation in $L_2(\gamma)$. *Journal of Approximation Theory*, 144:213–232, 2007.

[Gie07] K. Giesecke. The correlation-neutral measure for portfolio credit, 2007. Working paper. Available on SSRN preprint server.

[Gie08] K. Giesecke. Portfolio credit risk: Top down versus bottom up approaches. *Frontiers in Quantitative Finance: Credit Risk and Volatility Modeling, R. Cont (Ed.), Wiley.*, 2008.

[GL05] J. Gregory and J.-P. Laurent. Basket default swaps, CDOs and factor copulas. *Journal of Risk*, 7(4):103–122, 2005.

[GL06] E. Gobet and J. P. Lemor. Numerical simulation of BSDEs using empirical regression methods: theory and practice. In *Proceedings of the Fifth Colloquium on BSDEs (29th May - 1st June 2005, Shangai). Available on HAL preprint server*, 2006.

[GL07] E. Gobet and C. Labart. Error expansion for the discretization of backward stochastic differential equations. *Stochastic Processes and their Applications*, 117(7):803–829, 2007.

[GL09] E. Gobet and C. Labart. Solving BSDE with adaptive control variate. 2009. Submitted. Available on HAL preprint server.

[Gla04] P. Glasserman. *Monte Carlo methods in financial Engineering.* Applications of Mathematics 53. New York, NY: Springer. xiii, 596 p., 2004.

[GLW06] E. Gobet, J. P. Lemor, and X. Warin. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5):889–916, 2006.

[GM04] E. Gobet and S. Maire. A Spectral Monte Carlo Method for the Poisson Equation. *Monte Carlo Methods Appl.*, 10(3-4):275–285, 2004.

[GM05a] E. Gobet and S. Maire. Sequential control variates for functionals of markov processes. *SIAM Journal on Numerical Analysis*, 43(3):1256–1275, 2005.

[GM05b] E. Gobet and R. Munos. Sensitivity analysis using Itô-Malliavin calculus and martingales, and application to stochastic optimal control. *SIAM Journal on control and optimization*, 43(5):1676–1713, 2005.

[GM08] E. Gobet and A. Makhlouf. L_2-time regularity of BSDEs with irregular terminal functions. *In revision for Stochastic Processes and their Applications. Available on HAL preprint server*, 2008.

[GS07] P. Glasserman and S. Suchintabandid. Correlation expansions for CDO pricing. *Journal of Banking & Finance*, 31(5):1375–1398, 2007.

[GT01] E. Gobet and E. Temam. Discrete time hedging errors for options with irregular pay-offs. *Finance and Stochastics*, 5(3):357–367, 2001.

[GT08] S. Geiss and A. Toivola. Weak convergence of error processes in discretizations of stochastic integrals and Besov spaces. *Preprint, available at arXiv:0711.1439v2*, 2008.

[Hal62] J.H. Halton. Sequential Monte Carlo. *Proc. Camb. Philos. Soc.*, 58:57–78, 1962.

[Hal70] J.H. Halton. A retrospective and prospective survey of the Monte Carlo method. *SIAM Rev.*, 12:1–63, 1970.

[HJ07] S. Hamadène and M. Jeanblanc. On the starting and stopping problem: application in reversible investments. *Math. Oper. Res.*, 32(1):182–192, 2007.

[HL95] S. Hamadene and J.P. Lepeltier. Zero-sum stochastic differential games and backward equations. *Syst. Control Lett.*, 24(4):259–263, 1995.

[HM05] T. Hayashi and P. Mykland. Evaluating hedging errors: an asymptotic approach. *Mathematical Finance*, 15(2):309–343, 2005.

[HP81] J.M. Harrison and S.R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215 – 260, 1981.

[HT09] I. Halperin and P. Tomecek. Climbing down from the top: Single name dynamics in credit top down models, 2009. Working paper. Available on arXiv preprint server.

[Hul09] J.C. Hull. *Options, futures, and other derivatives. 7th ed.* Prentice-Hall International Editions. Upper Saddle River, NJ: Prentice Hall., 2009.

[HW04] J. Hull and A. White. Valuation of a CDO and nth to default CDS without Monte-Carlo simulation. *The Journal of Derivatives*, 12(2):8–23, 2004.

[Jac97] J. Jacod. On continuous conditional Gaussian martingales and stable convergence in law. 1997. Azéma, J. (ed.) et al., Séminaire de probabilités XXXI. Berlin: Springer. Lect. Notes Math. 1655, 232-246 (1997).

[JKB95] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, vol. 2, 2nd edition.* John Wiley & Sons, New York, 1995.

[JY01] R. Jarrow and F. Yu. Counterparty Risk and the Pricing of Defaultable Securities. *Journal of Finance*, 56:1765–1799, 2001.

[KHM09] N. El Karoui, S. Hamadène, and A. Matoussi. BSDEs and applications. Carmona, René (ed.) et al., Indifference pricing: Theory and applications. Princeton, NJ: Princeton University Press. Princeton Series in Financial Engineering, 267-320 (2009)., 2009.

[KJ09] N. El Karoui and Y. Jiao. Stein's method and zero bias transformation for CDO tranche pricing. *Finance and Stochastics*, 2009.

[KJK09] N. El Karoui, Y. Jiao, and D. Kurtz. Valuation and VaR computation for CDOs using Stein's method. Härdle, Wolfgang K. (ed.) et al., Applied quantitative finance. Berlin: Springer. 161-189., 2009.

[KJS98] N. El Karoui, M. Jeanblanc, and S.E. Shreve. Robustness of the Black and Scholes formula. *Math. Finance*, 8(2):93–126, 1998.

[KKP+97] N. El Karoui, C. Kapoudjian, E. Pardoux, S. Peng, and M.C. Quenez. Reflected solutions of backward SDE's, and related obstacle problems for PDE's. *Ann. Probab.*, 25(2):702–737, 1997.

[Kob00] M. Kobylanski. Backward stochastic differential equations and partial differential equations with quadratic growth. *Ann. Probab.*, 28(2):558–602, 2000.

[KPQ97a] N. El Karoui, E. Pardoux, and M.C. Quenez. Reflected backward SDEs and American options. Rogers, L. C. G. (ed.) et al., Numerical methods in finance. Session at the Isaac Newton Institute, Cambridge, GB, 1995. Cambridge: Cambridge Univ. Press. 215-231 (1997)., 1997.

[KPQ97b] N. El Karoui, S. Peng, and M.C. Quenez. Backward stochastic differential equations in finance. *Mathematical Finance*, 7(1):1–71, 1997.

[KS91] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus. 2nd ed.* Graduate Texts in Mathematics, 113. New York etc.: Springer-Verlag. xxiii, 470 p. DM 68.00/pbk , 1991.

[KS97] Y.M. Kabanov and M. Safarian. On Leland's strategy of option pricing with transaction costs. *Finance and Stochastics*, 1(3):239–250, 1997.

[KS98] I. Karatzas and S.E. Shreve. *Methods of mathematical finance.* Springer-Verlag, New York, 1998.

[Lel85] H.E. Leland. Option pricing and replication with transactions costs. *Journal of Finance*, 40(5):1283–1301, 1985.

[Lia97] B. Liang. *Options pricing with transaction costs: an asymptotic approach.* PhD thesis, Department of statistics, University of Chicago, 1997.

[LM07] A.V. Lopatin and T. Misirpashaev. Two-dimensional Markovian model for dynamics of aggregate credit loss, working paper, NumeriX. In *vol.22 of Advances in Econometrics*, 2007.

[LPS98] B. Lapeyre, E. Pardoux, and R. Sentis. *Monte-Carlo Methods for the Transport and Diffusion Equations (Méthodes de Monte-Carlo pour les Equations de Transport et de Diffusion.).* Mathématiques et Applications (Paris). 29. Berlin: Springer. x,178 p., 1998.

[LR08] F.A. Longstaff and A. Rajan. An empirical analysis of the pricing of collateralized debt obligations. *Journal of Finance*, 63(2):529–563, 2008.

[Mar78] W. Margrabe. The value of an option to exchange one asset for another. *Journal of Finance*, 33(1):177–186, 1978.

[Mar99] C. Martini. Propagation of convexity by Markovian and martingalian semigroups. *Potential Anal.*, 10(2):133–175, 1999.

[MG05] P. Del Moral and J. Garnier. Genealogical Particle Analysis of Rare Events. *Annals of Applied Probability*, 15:2496–2534, 2005.

[MP99] C. Martini and C. Patry. Variance optimal hedging in the Black-Scholes model for a given number of transactions. Research Report RR-3767, INRIA, 1999.

[MZ02] J. Ma and J. Zhang. Representation theorems for backward stochastic differential equations. *The Annals of Applied Probability*, 12(4):1390–1418, 2002.

[Nua05] D. Nualart. *Malliavin calculus and related topics, 2nd edition.* Springer-Verlag, New York, 2005.

[Par98] E. Pardoux. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order. In *Stochastic analysis and related topics, VI (Geilo, 1996)*, volume 42 of *Progr. Probab.*, pages 79–127. Birkhäuser Boston, Boston, MA, 1998.

[Per03] S. Pergamenshchikov. Limit theorem for Leland's strategy. *The Annals of Applied Probability*, 13:1099–1118, 2003.

[PP90] E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Systems Control Lett.*, 14(1):55–61, 1990.

[PP92] E. Pardoux and S. Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. Stochastic partial differential equations and their applications, Proc. IFIP Int. Conf., Charlotte/NC (USA) 1991, Lect. Notes Control Inf. Sci. 176, 200-217 (1992)., 1992.

[PTVF92] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C.* Cambridge University Press, 2nd edition, 1992.

[Ron04] T. Roncalli. *La Gestion des Risques Financiers.* Economica, 2004.

[Roo80] H. Rootzen. Limit distributions for the error in approximations of stochastic integrals. *Ann. Probab.*, 8:241–251, 1980.

[RY94] D. Revuz and M. Yor. *Continuous martingales and Brownian motion. 2nd ed.* Grundlehren der Mathematischen Wissenschaften. 293. Berlin: Springer-Verlag. xi, 560 p., 1994.

[SS08] E.W. Sachs and A.K. Strauss. Efficient solution of a partial integro-differential equation in finance. *Appl. Numer. Math.*, 58(11):1687–1703, 2008.

[SV79] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes.* Springer-Verlag, Berlin, 1979.

[Tem03] E. Temam. Analysis of error with Malliavin calculus: application to hedging. *Math. Finance*, 13(1):201–214, 2003.

[Vol05] E. Voltchkova. *Equations intégro-différentielles d'évolution : méthodes numériques et applications en finance.* PhD thesis, Ecole Polytechnique, 2005.

[Zha99] R. Zhang. *Couverture approchée des options Européennes.* PhD thesis, Ecole Nationale des Ponts et Chaussées, http://cermics.enpc.fr/theses/99/zhang-ruotao.ps.gz, 1999.

[Zha04] J. Zhang. A numerical scheme for BSDEs. *The Annals of Applied Probability*, 14(1):459–488, 2004.