UNIVERSITÉ PARIS-DAUPHINE U.F.R. MATHÉMATIQUES DE LA DÉCISION

 \mathbf{N}^{o} attribué par la bibliothèque

<u>THÈSE</u>

pour obtenir le grade de DOCTEUR ÈS-SCIENCES SPÉCIALITÉ MATHÉMATIQUES APPLIQUÉES

présentée et soutenue publiquement par Romuald ELIE

le 11 décembre 2006

sous le titre

CONTRÔLE STOCHASTIQUE ET MÉTHODES NUMÉRIQUES EN FINANCE MATHÉMATIQUE

Directeur de Thèse

M. Nizar TOUZI, Professeur à l'École Polytechnique

Jury

Rapporteurs :	M. Emmanuel GOBET, Professeur à l'INP Grenoble
	M. Arturo KOHATSU-HIGA, Professeur à l'Université d'Osaka
	Mme Thaleia ZARIPHOPOULOU, Professeur à l'Université du Texas
Examinateurs :	Mme Nicole EL KAROUI, Professeur à à l'École Polytechnique
	M. Bernard LAPEYRE, Professeur à l'ENPC
	M. Huyên PHAM, Professeur à l'Université Paris VII

L'université n'entend donner aucune approbation ou improbation aux opinions émises dans les thèses: ces opinions doivent être considérées comme propres à leurs auteurs.

Remerciements

Certains voient la thèse comme une course d'endurance, je préfère la comparer à l'escalade d'une falaise. Il y a trois ans, je me trouvais au bas de cette falaise, essayant d'entrevoir le sommet et tentant d'effectuer mes premiers mouvements sur cette roche inconnue. J'observais avec envie certain grimpeurs expérimentés qui alliaient technique, agilité et originalité dans leurs gestes.

C'est Nizar Touzi qui a pris le temps de me guider tout au long de cette aventure. Grimpant tout d'abord en tête afin de me montrer les pas, il a su me transmettre l'envie de me lancer seul sur certaines voies, parfois sans issues, et me donner le courage de recommencer à grimper lorsque mes forces m'abandonnaient. Toujours encourageant, il m'a donné des clefs pour déchiffrer les voies et m'a incité à prendre des risques, à choisir des parcours plus exposés. Au sens propre comme au sens figuré, mon second partenaire d'escalade a été Bruno Bouchard. Il a fait preuve d'une très grande disponibilité et a partagé avec moi son expérience sur certaines parois plus techniques ou surprenantes. Leur confiance à tous deux m'a permis de dépasser de nombreux obstacles imprévus.

Un grand merci à Emmanuel Gobet, Arturo Kohatsu-Higa et Thaleia Zariphopoulou pour avoir accepté d'examiner cette thèse. Leurs travaux sont pour moi une grande source d'inspiration et je suis honoré et flatté du temps qu'ils ont consacré à la relecture de ma thèse. Toute ma gratitude va également à Nicole El Karoui, Bernard Lapeyre et Huyen Pham qui ont accepté d'être membres de mon jury de thèse. Il y a trois ans, alors qu'il me restait un long chemin à parcourir, ils m'ont donné de précieux conseils sur la manière de mener à bien cette entreprise.

Mes remerciements vont également aux joyeuses équipes de l'entresol de l'ENSAE et du laboratoire de Finance-Assurance du CREST. De pauses café en bonnes humeurs, de déjeuners animés en discussions mathématiques, chacun d'entre eux a créé les conditions indispensables à l'équilibre détente-travail dans un environnement scientifiquement très stimulant. Je tiens particulièrement à saluer Arnaud, Arthur, Emmanuel, Fabian, Imen, Mathieu, Philippe, Xav' et Xavier. L'ENSAE m'a aussi donné l'opportunité d'enseigner dans mes domaines de recherche et de participer aux choix d'orientation des enseignements de l'Ecole. A ce titre, je remercie Sylviane Gastaldo et Christian Gourieroux pour leur accueil chaleureux et pour les responsabilités qu'ils ont su me confier. Je ne saurais oublier les chercheurs avec qui j'ai eu la chance de travailler ou d'échanger des idées. Je pense en particulier à Jean-David Fermanian, co-auteur d'un des articles présentés ici, Paul Doukhan ou Francois Delarue, et je les remercie pour leurs conseils éclairés. Je tiens enfin à exprimer ma sincère reconnaissance à ma famille et à mes amis qui m'ont aidé à avancer jusqu'à aujourd'hui. Si pour certains le langage mathématique est un monde mystérieux, ils ont su accepter mon rythme et être présents dans les moments de doute comme dans ceux de sérénité. Quant à ceux pour qui ce monde est plus familier, qui sait, peut-être serons-nous amenés un jour à progresser ensemble sur quelques sujets verticaux ? Merci à ma compagne pour sa présence riante à mes côtés comme pour sa longue absence outre-Atlantique, synonyme pour moi de période de travail intense. Nous recherchons ensemble l'excitant vertige du grimpeur face au vide, finalement peut-être identique aux émotions du chercheur face aux objets abstraits qu'il manipule ?

Si ces trois années de thèse ont été l'occasion d'échanges forts avec de nombreux compagnons de cordée, elles ont aussi été le moment de réflexions personnelles dans la solitude de la recherche. Quand nous parcourons de nouveaux domaines, nous nous mettons en jeu en explorant nos capacités et en cherchant notre équilibre. Plus qu'un aboutissement en soi, cette thèse est pour moi, je l'espère, le commencement du long et humble apprentissage des connaissances et compétences me permettant de participer pleinement à la recherche en mathématiques financières.

Résumé

Cette thèse présente trois sujets de recherche indépendants appartenant au domaine des méthodes numériques et du contrôle stochastique avec des applications en mathématiques financières.

Nous présentons dans la première partie une méthode non-paramétrique d'estimation des sensibilités des prix d'options. A l'aide d'une perturbation aléatoire du paramètre d'intérêt, nous représentons ces sensibilités sous forme d'espérance conditionnelle, que nous estimons à l'aide de simulations Monte Carlo et de régression par noyaux. Par des arguments d'intégration par parties, nous proposons plusieurs estimateurs à noyaux de ces sensibilités, qui ne nécessitent pas la connaissance de la densité du sous-jacent, et nous obtenons leurs propriétés asymptotiques. Lorsque la fonction payoff est irrégulière, ils convergent plus vite que les estimateurs par différences finies, ce que l'on vérifie numériquement.

La deuxième partie s'intéresse à la résolution numérique de systèmes découplés d'équations différentielles stochastiques progressives rétrogrades. Pour des coefficients Lipschitz, nous proposons un schéma de discrétisation qui converge plus vite que $n^{-1/2+\varepsilon}$, pour tout $\varepsilon > 0$, lorsque le pas de temps 1/n tends vers 0. Lorsque les coefficients sont C_b^1 à dérivées Lipschitz, ou que le terme de saut du processus tangent de la composante progressive de l'équation satisfait une condition de non-dégénérescence, nous obtenons la vitesse optimale en $n^{-1/2}$. L'utilisation pratique de ce schéma nécessite le calcul d'un grand nombre d'espérances conditionnelles, que nous approchons à l'aide de techniques d'estimation non-paramétrique. Nous contrôlons l'erreur globale commise par l'algorithme ce qui permet le choix simultané de ses paramètres, et nous présentons des exemples de résolution numérique de systèmes couplés d'EDP semi-linéaires.

Enfin, la dernière partie de cette thèse étudie le comportement d'un gestionnaire de fond, maximisant l'utilité intertemporelle de sa consommation, sous la contrainte que la valeur de son portefeuille ne descende pas en dessous d'une fraction fixée de son maximum courant. Nous considérons une classe générale de fonctions d'utilité, et un marché financier composé d'un actif risqué de dynamique Black-Scholes. Lorsque le gestionnaire se fixe un horizon de temps infini, nous obtenons sous forme explicite sa stratégie optimale d'investissement et de consommation, ainsi que la fonction valeur du problème. En horizon fini, nous caractérisons la fonction valeur comme unique solution de viscosité de l'équation d'Hamilton-Jacobi-Bellman correspondante.

Abstract

This PhD dissertation presents three independent research topics in the fields of numerical methods and stochastic control with applications to financial mathematics.

The first part of this thesis is dedicated to the estimation of the sensitivities of option prices, by means of non-parametric techniques. When the density of the underlying is unknown, we propose several non-parametric estimators of the so called *Greeks*, based on the randomization of the parameter of interest combined with Monte Carlo simulations and Kernel regression techniques. We provide an asymptotic analysis of the mean squared error of these estimators, as well as their asymptotic distributions. For a discontinuous payoff function, the kernel estimators outperforms the classical finite differences one in terms of the asymptotic rate of convergence. This result is confirmed by our numerical experiments.

The second part of this dissertation deals with the numerical resolution of systems of decoupled forward-backward stochastic differential equations with jumps. Assuming that the coefficients are Lipschitz-continuous, we propose a convergent discrete-time scheme whose rate of convergence is at least $n^{-1/2+\varepsilon}$, for any $\varepsilon > 0$, when the number of time steps n goes to infinity. Under the additional condition that, either all the coefficients are C_b^1 with Lipschitz derivatives, or the jump coefficient of the first variation process of the forward component satisfies a non-degeneracy condition which ensures its invertibility, we achieve the optimal convergence rate $n^{-1/2}$. The implementation of this scheme requires the computation of a large number of conditional expectations, that we approximate by means of non parametric regression techniques. We control the global error of the algorithm, allowing to calibrate all the parameters of estimation at the same time, and provide the numerical solution of systems of coupled semilinear parabolic PDE's.

The third part of this thesis is concerned with the resolution of the optimal consumptioninvestment problem under a drawdown constraint, i.e. the wealth process never falls below a fixed fraction of its running maximum. We assume that the risky asset is driven by the constant coefficients Black and Scholes model and we consider a general class of utility functions. On an infinite time horizon, we provide the value function in explicit form, and we derive closed-form expressions for the optimal consumption and investment strategy. On a finite time horizon, we interpret the value function as the unique viscosity solution of its corresponding Hamilton-Jacobi-Bellman equation.

Contents

Introduction Générale			3	
		Calcul	de sensibilité de prix d'options	3
		Résolu	tion numérique d'EDSPR découplées avec sauts	10
		Investi	is sement et consommation sous contrainte drawdown	21
Ι	Op	timal	Greek weight by Kernel estimation	31
	1	Introd	uction	34
	2 The Greek weights set		37	
		2.1	Definition	37
		2.2	Malliavin Greek weights	39
		2.3	Examples of Malliavin Greek weights	40
	3	Kernel	estimation and optimal Greek weight	42
		3.1	Randomization of the parameter	42
		3.2	A first kernel estimator of the Greek	43
		3.3	A simpler kernel estimator of the Greek	44
		3.4	Differentiating the kernel estimator of the price	45
	4 Asymptotic results		ptotic results	46
		4.1	Asymptotic results for the single kernel-based estimators	47
		4.2	Asymptotic properties of the double Kernel-based estimator	51
		4.3	Optimal choice of N and h	52
		4.4	The case of a uniform randomizing distribution	53
		4.5	The case of a truncated exponential randomizing distribution $\ .$.	55
		4.6	Comparison with the finite differences estimators $\ldots \ldots \ldots$	56
	5	Numerical results		58
		5.1	Computation of the optimal bandwidth $\ldots \ldots \ldots \ldots \ldots \ldots$	58
		5.2	Numerical comparison of the estimators $\ldots \ldots \ldots \ldots \ldots$	60
	6	Short :	maturity asymptotics	63

		6.1	Singularity of the Greek weights for short maturity	64
		6.2	Parameterized stochastic differential equation $\ldots \ldots \ldots \ldots$	65
		6.3	Asymptotic properties	66
	7	Asymp	ptotic properties of $\tilde{\beta}_N$	71
		7.1	Preliminaries	71
		7.2	A suitable decomposition $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	75
		7.3	Asymptotic bias and variance	81
		7.4	Central limit theorem	83
II	Nı	ımerio	cal approximation of BSDEs with jumps	85
1	Disc	crete ti	ime approximation	89
	1.1	Introd	uction	89
	1.2	Discre	te time approximation of decoupled FBSDE with jumps	93
		1.2.1	Decoupled forward backward SDE's	93
		1.2.2	Discrete time approximation	95
		1.2.3	Convergence of the approximation scheme	97
		1.2.4	Path-regularity and convergence rate under additional assumption	s101
		1.2.5	Possible Extensions	103
	1.3	Mallia	vin calculus for FBSDE	105
		1.3.1	Generalities	106
		1.3.2	Malliavin calculus on the Forward SDE	111
		1.3.3	Malliavin calculus on the Backward SDE	112
	1.4	Repres	sentation results and path regularity for the BSDE	116
		1.4.1	Representation	116
		1.4.2	Path regularity	119
	1.5	Appen	dix: A priori estimates	125
2	Algo	\mathbf{p}	and numerical results	131
	2.1	A fully	/ implementable algorithm	131
		2.1.1	A localization procedure	132
		2.1.2	Description of the algorithm	133
		2.1.3	Discussion on the global error of the algorithm $\ldots \ldots \ldots$	135
		2.1.4	Control of the statistical error	137
	2.2	Numer	rical examples	141
		2.2.1	Put option with default risk on the seller	141

2.6

2.2.2	Fully coupled system of PDE	143
2.2.3	A more complex example	. 146

III Consumption-investment strategy under drawdown constraint 149 Explicit solution in infinite time horizon 1 1531531.1 1.2155Consumption-portfolio strategies and the drawdown constraint . 1.2.11551.2.21561.2.3The optimal consumption-investment problem 1581.31601.3.1The corresponding dynamic programming equation 1601.3.21601.3.3161 1.3.4Explicit solution under drawdown constraint 1631.3.51651.3.61681.41711.5The verification argument 1741.5.11741.5.2178PDE characterization in finite time horizon 1852 2.11852.21862.2.1Consumption-portfolio strategies and the drawdown constraint . 1872.2.2The finite horizon consumption-investment problem 188 2.31892.3.11892.3.21922.4Numerical examples 1942.5Viscosity property 1972.5.11972.5.2198

A comparison result

200

Introduction Générale

Cette thèse est composée de trois sujets de recherche pouvant être lus indépendamment. Ces travaux ont été motivés par des exemples d'applications en mathématiques financières, mais certains résultats, en particulier ceux de la deuxième partie, s'inscrivent dans un cadre plus général. La première partie propose une nouvelle méthode numérique non paramétrique pour estimer les sensibilités de prix d'options. La deuxième s'intéresse à la résolution numérique d'équations différentielles stochastiques progressives rétrogrades (EDSPR) découplées avec sauts. Enfin la dernière traite de la résolution d'un problème de contrôle optimal stochastique de gestion de portefeuille, sous une contrainte de type drawdown, qui interdit à la valeur d'un portefeuille de descendre en dessous d'une fraction $\alpha \in [0, 1)$ de son maximum courant. Cette introduction suit la structure générale de cette thèse en présentant successivement ces trois parties, qui bénéficient chacune de notations qui leurs sont propres.

Calcul de sensibilité de prix d'options

Motivation

Une option Européenne sur un actif financier est un contrat par lequel son vendeur s'engage à délivrer à une date T un paiement aléatoire dépendant de la trajectoire de cet actif sous-jacent contre le versement d'une prime à la date 0. Ces produits sont fréquemment échangés sur les marchés financiers car ils bénéficient d'un fort effet de levier et permettent de se couvrir facilement contre les évolutions non souhaitées de l'actif financier sous-jacent. L'estimation du coût de couverture contre ces risques nécessite alors la valorisation de ces options, c'est à dire le calcul de la prime à verser à l'instant t = 0.

En 1973, Black et Scholes [17] définissent le prix d'une option comme la valeur à la date t = 0 d'une stratégie dynamique d'investissement dans l'actif risqué sous-jacent et dans un actif sans risque permettant de répliquer parfaitement le paiement aléatoire de l'option à l'instant T. En effet, si cette relation n'était pas vérifiée, il y aurait des possibilités d'arbitrage sur le marché. Sous certaines hypothèses (en particulier l'absence de coûts de transaction et la complétude du marché), les options sont réplicables et il est possible de créer artificiellement un univers dans lequel tous les intervenants du marché peuvent être considérés neutres au risque. Autrement dit, dans cet univers caractérisé par une probabilité risque neutre, la valeur donnée par tout agent à cette option est simplement l'espérance actualisée des flux futurs qu'elle engendre. Considérons alors une option de payoff terminal réactualisé $\phi[Z(\lambda)]$, avec ϕ une fonction déterministe et

 $Z(\lambda)$ une variable aléatoire traduisant l'évolution de l'actif financier sous-jacent jusqu'en T, dépendant d'un paramètre λ de dimension d dicté par la modélisation choisie. Sa valeur $V^{\phi}(\lambda)$ s'écrit ainsi

$$V^{\phi}(\lambda) = \mathbb{E}[\phi(Z(\lambda))], \qquad (1)$$

où l'espérance est prise sous la probabilité risque-neutre.

Etant donnée une dynamique d'évolution pour l'actif financier sous-jacent, $Z(\lambda)$ est directement relié à la solution d'une équation différentielle stochastique, et la valeur $V^{\phi}(\lambda)$ de l'option, donnée par (1), n'est que très rarement explicitement calculable. Les méthodes numériques généralement envisagées pour estimer le prix de l'option se séparent en deux grandes classes. D'une part, le prix de l'option donné par (1) s'interprète comme solution d'une équation aux dérivées partielles, caractérisation qui sera d'ailleurs discutée dans le deuxième chapitre de cette thèse. L'EDP alors obtenue peut être résolue à l'aide de schémas numériques d'approximation à base de différences finies ou d'éléments finis, dont le livre d'Achdou et Pironneau [1] présente les principaux résultats de convergence et celui de Tavella [101] donne de précieux conseils pour leur mise en oeuvre pratique. D'autre part, la solution de l'équation différentielle stochastique peut être approchée par un schéma de type Euler le long d'une discrétisation en temps, et l'espérance peut alors être estimée par une méthode de Monte Carlo.

Une fois une méthode adoptée pour la calcul du prix de l'option, se pose la question de sa sensibilité face aux variations des paramètres caractérisant le marché et l'évolution de l'actif financier sous-jacent. Ces sensibilités appelées *Grecques* sont données à la valeur λ^0 du paramètre d'intérêt λ par

$$\beta^0 := \nabla_{\lambda} V^{\phi}(\lambda^0) = \nabla_{\lambda} \mathbb{E}[\phi(Z(\lambda^0))].$$
(2)

Selon le choix de λ , ces *Grecques* prennent des significations bien sûr différentes mais ont souvent des interprétations très utiles en pratique. Par exemple, lorsque λ est la valeur actuelle de l'actif financier sous-jacent, cette sensibilité nommée *Delta* s'interprète comme la quantité d'actif risqué à détenir dans le portefeuille de duplication de l'option. De même, le *Vega*, sensibilité du prix par rapport à la volatilité du sous-jacent, permet, entre autres, de mesurer le risque de mauvaise calibration du modèle d'évolution de l'actif.

Etat de l'art

Nous présentons ici les principales méthodes numériques probabilistes utilisées pour le calcul des Grecques, dont, par exemple, Kohatsu-Higa et Montero [69] font une descrip-

tion très détaillée.

La méthode des différences finies repose sur l'approximation de la dérivée du prix par sa variation en réponse à une petite perturbation ϵ du paramètre λ comme suit

$$\beta^0 \sim \frac{\mathbb{E}[\phi(Z(\lambda^0 + \epsilon))] - \mathbb{E}[\phi(Z(\lambda^0))]}{\epsilon}.$$

Les deux espérances sont alors approchées à l'aide de simulations Monte Carlo pouvant être réalisées avec des jeux de trajectoires différentes ou identiques, modifiant ainsi la variance de l'estimateur. Ce dernier est biaisé et le choix de la perturbation ϵ est crucial car il repose sur un équilibre entre biais et variance. Comme étudié précisément par L'Ecuyer et Perron [42] puis Detemple, Garcia et Rindisbacher [36] ou Milstein et Tretyakov [81], cet estimateur converge avec la vitesse paramétrique $N^{-1/2}$ si la fonction ϕ est suffisamment régulière, mais n'atteint qu'une vitesse en $N^{-1/3}$ (ou $N^{-2/5}$ pour un estimateur centré symétrique) lorsque le nombre de points de discontinuités de ϕ est dénombrable.

La *pathwise method* proposée par Broadie et Glasserman [23] repose sur une interversion entre les opérateurs de dérivation et d'espérance

$$\beta^0 = \mathbb{E}[\phi'(Z(\lambda^0))\nabla_{\lambda}Z(\lambda^0)],$$

où $\nabla_{\lambda} Z(\lambda)$ représente le processus tangent associé à $Z(\lambda)$. L'espérance précédente est approchée à l'aide de simulations Monte Carlo et l'estimateur obtenu est non biaisé. Cependant, son calcul nécessite la simulation du processus $\nabla_{\lambda} Z(\lambda^0)$ et des conditions fortes de régularité sur la fonction payoff ϕ .

La méthode du rapport de vraisemblance également introduite par Broadie et Glasserman [23] repose cette fois sur l'interversion entre les opérateurs de dérivation et d'intégration, lorsque la variable aléatoire $Z(\lambda)$ admet une densité régulière $f(\lambda, .)$:

$$\beta^0 = \mathbb{E}[\phi(Z(\lambda^0))s(\lambda^0, Z(\lambda^0))], \text{ avec } s(\lambda, z) := \frac{\partial}{\partial \lambda}\ln(f(\lambda, z)).$$

A moins d'utiliser la densité artificielle du schéma d'Euler associé au sous-jacent, cette méthode nécessite l'existence et la connaissance de la densité f. Cette technique a été généralisée par Fournié et al [50, 51] qui utilisent le calcul de Malliavin pour caractériser l'ensemble des poids

$$\mathcal{W} := \left\{ \pi \in \mathcal{L}^2(\Omega, \mathbb{R}^d) : \beta^0 = E\left[\phi[Z(\lambda^0)]\pi\right] \text{ pour tout } \phi \in \mathcal{L}^\infty(\mathbb{R}^n, \mathbb{R}) \right\}.$$

Cette caractérisation, détaillée dans la Section 2, permet, dans certains cas, d'obtenir au prix de lourds calculs analytiques un panel de poids π utilisables. Parmi tous les poids $\pi \in \mathcal{W}$ possibles, $s(\lambda^0, Z(\lambda^0))$ est celui qui minimise $\mathbb{V}ar[\phi(Z(\lambda^0))\pi]$. Lorsque la densité de $Z(\lambda)$ est connue, il est donc optimal d'utiliser la méthode du rapport de vraisemblance.

Résultats nouveaux

La première partie de cette thèse est un travail réalisé en collaboration avec Jean-David Fermanian et Nizar Touzi qui propose de nouveaux estimateurs de β^0 reposant sur des techniques d'estimation non paramétrique. Nous nous plaçons dans un cadre de travail où la densité $f(\lambda, .)$ de $Z(\lambda)$ est inconnue et où la fonction payoff ϕ est peu régulière. Alors la méthode du rapport de vraisemblance et la *pathwise method* ne sont pas utilisables et les estimateurs à différences finies ont au mieux une vitesse de convergence en $N^{-2/5}$. Comme détaillé ci-après, les estimateurs non-paramétriques que nous proposons bénéficient d'une vitesse de convergence plus rapide.

Nous perturbons de manière aléatoire notre paramètre λ autour de λ^0 à l'aide d'une densité régulière $\ell(\lambda^0 - .)$. Le prix $V^{\phi}(\lambda^0)$ et sa sensibilité β^0 peuvent alors s'écrire

$$V^{\phi}(\lambda^{0}) = \mathbb{E}\left[\phi(Z)|\Lambda = \lambda^{0}\right] \text{ et } \beta^{0} = \mathbb{E}\left[\phi(Z)s(\Lambda, Z)|\Lambda = \lambda^{0}\right], \tag{3}$$

où (Λ, Z) est une variable aléatoire de densité $\varphi(\lambda, z) := \ell(\lambda^0 - \lambda)f(\lambda, z)$ et donc telle que Z sachant $\{\Lambda = \lambda\}$ ait pour densité $f(., \lambda)$. L'intérêt de cette perturbation est d'introduire artificiellement une densité régulière sur laquelle nous pourrons reporter l'opérateur de dérivation. Considérant ainsi N réalisations indépendantes (Λ_i, Z_i) de la variable aléatoire (Λ, Z) , ces espérances conditionnelles peuvent être approchées par les estimateurs à noyaux

$$V_N^{\phi}(\lambda^0) := \frac{1}{\ell(0)Nh^d} \sum_{i=1}^N \phi(Z_i) K\left(\frac{\Lambda_i - \lambda^0}{h}\right)$$
(4)

 et

$$\bar{\beta}_N := \frac{1}{\ell(0)Nh^d} \sum_{i=1}^N \phi(Z_i) s(\Lambda_i, Z_i) K\left(\frac{\Lambda_i - \lambda^0}{h}\right) , \qquad (5)$$

où h > 0 est la fenêtre de l'estimateur et K un noyau régulier. Rappelons brièvement que les techniques d'estimation par noyaux reposent simplement sur l'approximation de la masse de dirac $\delta_{\Lambda_i=\lambda^0}$ par $K[(\Lambda_i - \lambda^0)/h]/h$ pour une fenêtre h petite et K une fonction qui peut s'interpréter comme une densité. Cette fonction K est caractérisée par son ordre p, plus petit entier q tel que $\int K(x)x^q dx \neq 0$, qui indique ainsi le premier terme non nul dans les développements asymptotiques de l'erreur d'approximation. L'ordre du noyau est directement relié à la régularité de la fonction que l'on cherche à estimer et influence fortement la vitesse de convergence de l'estimateur. Remarquons que le processus Z sachant $\{\Lambda = \lambda\}$ étant caractérisé par une équation différentielle stochastique paramétrée par λ , la simulation de N réalisations indépendantes de (Λ, Z) , à l'aide par exemple d'un schéma d'Euler, ne nécessite pas la connaissance de la densité f. Malheureusement, la fonction score s étant elle aussi inconnue, on ne peut utiliser directement l'estimateur $\bar{\beta}_N$ de β_0 introduit dans (5).

Néanmoins, l'écriture du prix sous forme d'un estimateur à noyaux permet de reporter l'opération de dérivation par rapport au paramètre λ sur la densité ℓ et le noyau Kréguliers. En dérivant par rapport à λ l'estimateur $V_N^{\phi}(\lambda^0)$ du prix donné en (4), nous obtenons alors l'estimateur de β^0 suivant

$$\check{\beta}_N := \frac{1}{\ell(0)Nh^{d+1}} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) - hK \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \frac{\nabla \ell}{\ell}(0) \right).$$
(6)

Lorsque N tend vers l'infini à h fixé, l'estimateur $\bar{\beta}_N$ de β_0 introduit dans (5) converge et, par un argument d'intégration par partie détaillé dans la Section 3.3, sa limite se réécrit comme la limite lorsque N tend vers l'infini d'un nouvel estimateur

$$\hat{\beta}_N := \frac{1}{\ell(0)Nh^{d+1}} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + hK \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \frac{\nabla \ell}{\ell} (\lambda^0 - \Lambda_i) \right).$$
(7)

La densité f peut également être directement approchée à l'aide de techniques d'estimation par noyaux, dont l'on déduit un estimateur de s par une opération de dérivation. Reportant cette approximation dans (5), on construit alors un dernier estimateur $\tilde{\beta}_N$ de β^0 fondé sur deux fonctions noyaux et défini précisément en Section 3.2. Cependant, il s'avère que la vitesse de convergence de $\tilde{\beta}_N$ est identique à celle de $\check{\beta}_N$ et de $\hat{\beta}_N$ mais nécessite des hypothèses plus fortes, en particulier sur la régularité de ϕ . Comme il est, de surcroît, plus coûteux en temps de calcul, nous concentrons la suite de notre étude sur les deux estimateurs $\check{\beta}_N$ et $\hat{\beta}_N$.

Sous des hypothèses de régularité sur les densités ℓ et f liées à l'ordre p du noyau, les comportements asymptotiques de ces deux estimateurs sont identiques. Lorsque N tend vers l'infini et h tend vers 0, nous obtenons des équivalents du biais et de la variance asymptotique de la forme

$$\mathbb{E}\left[\hat{\beta}_{N}\right] - \beta^{0} \sim Ch^{p} \quad \text{et} \quad \mathbb{V}\mathrm{ar}\left[\hat{\beta}_{N}\right] \sim \frac{\Sigma}{Nh^{d+2}}, \qquad (8)$$

où d est la dimension du paramètre λ et p est l'ordre du noyau K. Lorsque de plus Nh^{d+2+2p} tend vers 0, on en déduit le théorème central limite

$$\sqrt{Nh^{d+2}} \left(\hat{\beta}_N - \beta^0 \right) \longrightarrow \mathcal{N}(0, \Sigma) .$$
(9)

Dans le cas particulier où l'on choisit pour ℓ une densité uniforme ou exponentielle tronquée de largeur h, nous améliorons le comportement asymptotique de nos estimateurs en enlevant la dimension d dans les équivalents (8) et (9). Le choix de la fenêtre est primordial lors de l'utilisation d'estimation par noyaux et repose sur un équilibre entre le biais et la variance de l'estimateur. La fenêtre optimale h^* vaut ici $C^* N^{-1/(2p+2)}$ et donne à nos estimateurs la vitesse de convergence $N^{-p/(2p+2)}$. Il est à noter que l'implémentation pratique de nos estimateurs nécessite l'estimation de cette constante C^* , pour laquelle nous proposons une méthode reposant sur un faible nombre de simulations Monte Carlo et une adaptation de "la règle du pouce" de Silvermann. L'avantage majeur de nos estimateurs est que leur vitesse de convergence ne nécessite aucune hypothèse sur la régularité de ϕ . En comparaison aux estimateurs à différences finies dont la vitesse de convergence est limitée à $N^{-2/5}$ lorsque ϕ a un nombre dénombrable de discontinuités, notre estimateur est donc plus rapide dès que l'ordre du noyau p est supérieur à 4.

Nous présentons des résultats numériques pour le calcul du delta d'une option digitale Européenne ou Asiatique dans le modèle de Black-Scholes. Les résultats obtenus confirment les résultats théoriques précédents mais notre méthode nécessite un grand nombre de trajectoires Monte Carlo pour être plus précise que les estimateurs à différences finies. Ce nombre de simulations peut toutefois être considérablement réduit à l'aide de techniques de réduction de variance sur la densité ℓ . En revanche, les estimateurs fondés sur le calcul de Malliavin, bien que de variance non optimale en comparaison à celui de likelihood ratio, sont tout de même plus précis. Ils bénéficient en effet d'une vitesse de convergence paramétrique en $N^{-1/2}$. Cependant, l'obtention de ces estimateurs dans des modèles plus complexes nécessite de lourds calculs analytiques, que l'on ne peut pas toujours mener à terme, et souffrent d'une variance trop importante au voisinage de la maturité, comme détaillé par Fournié, Lasry, Lebouchoux, Lions et Touzi [50]. Nous étudions donc plus précisément le cas où $Z(\lambda)$ est la solution d'une équation différentielle stochastique paramétrée par λ qui diffuse sur un intervalle de temps très court. Par une étude du comportement asymptotique de notre estimateur en temps petit, nous obtenons des équivalents plus précis sur sa variance et son biais asymptotiques, dont on déduit en particulier une méthode plus simple d'estimation de la fenêtre optimale h^* .

Perspectives

Au vu du grand nombre de simulations nécessaires à nos estimateurs, différentes pistes de recherche pourraient être envisagées. Tout d'abord, d'un point de vue simplement numérique, des tests pourraient être réalisés dans des modèles plus complexes, où les estimateurs de Malliavin ne sont pas disponibles. Différentes techniques de réduction de variance pourraient également être appliquées. De plus, une étude approfondie de l'influence la densité ℓ sur la précision de l'estimation pourrait permettre l'obtention de critères de choix, permettant d'adapter par exemple cette densité au modèle sous-jacent ou à la forme de la fonction payoff. Les recherches que nous avons effectuées dans cette direction restent encore infructueuses et les tests numériques réalisés avec différents choix de densités produisent des résultats comparables. Il est également possible que le choix de cette densité et du noyau K puisse en fait se restreindre au choix d'une unique fonction, ayant des propriétés particulières permettant de retranscrire la forme des estimateurs étudiés ici.

L'estimation par noyaux n'est pas le seul outil de statistique non paramétrique à notre disposition pour approcher des espérances conditionnelles. Nous pourrions également envisager des estimations à l'aide de splines ou de polynômes locaux du type de ceux utilisés dans le chapitre 2. Un outil très puissant pour estimer des espérances conditionnelles repose sur la projection sur des bases d'ondelettes. En comparaison à l'utilisation de base orthogonales classiques, elles permettent la localisation de l'information en fréquence mais aussi en temps. Pour une régularité donnée de la fonction à estimer, caractérisée par un espace de Besov auquel elle appartient, les estimateurs linéaires par ondelettes de la régression sont très souvent optimaux au sens minimax. De plus, la puissance des base d'ondelettes repose principalement sur l'utilisation de techniques de seuillage des coefficients qui leur permet d'assurer l'optimalité minimax sur une classe de fonctions plus importante, mais surtout de s'adapter à une régularité inconnue du signal. Un exposé détaillé de ces techniques est présenté par Donoho, Johnstone, Kerkyacharian et Picard [40]. On peut alors imaginer appliquer au calcul des grecques les techniques d'estimation par ondelettes de la dérivée d'une fonction de régression, en s'inspirant par exemple de la méthode de Cai [25]. Interprétant ce problème comme un cas particulier d'une théorie générale d'estimation fonctionnelle dans le cadre de problème inverse, il démontre, sur une large classe de fonctions Hölderiennes, l'optimalité minimax locale adaptative pour l'estimation ponctuelle de la dérivée. Ce résultat est obtenu à l'aide d'une technique propre aux ondelettes: le seuillage par bloc.

Résolution numérique d'EDSPR découplées avec sauts

Motivation

Il est désormais classique d'associer la solution de l'équation de la chaleur au comportement du mouvement Brownien. De manière plus générale, les solutions d'équations aux dérivée partielles (EDP) linéaires du second ordre s'interprètent à l'aide d'équations différentielles stochastiques (EDS). Cette représentation dîte de Feynman-Kac est une passerelle qui permet de transposer des résultats d'ordre analytique à la théorie des processus stochastiques, et inversement. D'un point de vue numérique, il est alors possible de résoudre un problème entièrement déterministe, s'interprétant à l'aide d'une équation aux dérivées partielles, par des techniques probabilistes de simulation.

Ce lien entre la théorie des processus stochastiques et l'univers des équations aux dérivées partielles fut étendue par Pardoux et Peng [84, 85] au cadre d'EDP semi-linéaires du second ordre, dont la solution de viscosité s'interprète à l'aide d'un processus, solution d'un système découplé de deux EDS, l'une progressive, l'autre rétrograde. On parle alors d'équation différentielle stochastique progressive rétrograde (EDSPR) découplée, au sens où la dynamique du processus progressif est indépendante de la solution de l'EDS rétrograde. Différents schémas numériques probabilistes ont été proposés ces dernières années pour résoudre les EDSPR découplées et concurrencent ainsi les méthodes numériques plus classiques de résolution d'EDP, particulièrement en grande dimension.

Tang et Li [100] ont étudié les conséquences de l'ajout de sauts à la dynamique du processus stochastique solution de l'EDSPR découplée et ont obtenu des résultats d'existence et d'unicité. Comme observé par Barles, Buckdahn et Pardoux [5] et Pardoux, Pradeilles et Rao [86], cette solution s'interprète à l'aide d'équations intégro-différentielles partielles (EIDP) semi-linéaires, voire dans certains cas plus particuliers, à l'aide de solutions de système couplé d'EDP semi-linéaire.

Le champ d'applications nécessitant la résolution d'équations aux dérivées partielles est très vaste et nous ne présentons ici que quelques exemples. Il couvre en particulier le domaine du contrôle optimal stochastique, où Bismut [15] a donné naissance aux EDS rétrogrades, et son pendant déterministe: les équations d'Hamilton-Jacobi-Bellman. Pham [87] présente en détail les liens entre EDSPR découplée et la résolution de problèmes de contrôle optimal stochastique, et Tang et Li [100] détaillent en particulier de nombreuses applications de la résolution d'EDSR avec sauts dans ce domaine. Ces techniques se lient ainsi aux opérations de maximisation de fonctions d'utilité ou de minimisation de risque, et démontrent leur intérêt dans les domaines de l'économie et de la finance. El Karoui, Peng et Quenez [47] présentent par exemple un large panorama des applications en mathématiques financières de la résolution d'EDSPR sans sauts, le lien avec la valorisation par indifférence d'utilité étant discuté plus en détail par Rouge et El Karoui [94]. L'ajout de sauts dans la dynamique des actifs financiers permet une représentation plus réaliste de leur évolution. Ainsi Becherer [9] ou Eyraud-Loisel [48], par exemple, se heurtent à la résolution d'EDSPR avec sauts lorsqu'ils traitent des problèmes de couverture d'actifs financiers avec sauts par indifférence d'utilité et en présence d'*insider* sur le marché. Notez également que la résolution de système couplé d'EDP semi-linéaires permet entre autres l'évaluation de produits financiers classiques soumis en sus à un risque de défaut, dont nous présentons un exemple en Section 2.2. L'utilisation de ces techniques pour la valorisation de produits plus complexes tels que les obligations convertibles est également en cours d'étude par Bielecky, Crépey, Jeanblanc et Rutkowsky [32].

Etat de l'art

Nous présentons ici plus en détail les notions d'EDSPR avec et sans sauts ainsi que les méthodes numériques à notre disposition pour les résoudre.

Détaillons tout d'abord la notion d'EDSPR découplée sans sauts. Soient $b : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,1] \times \mathbb{R}^d \to \mathbb{M}^d$, $g : \mathbb{R}^d \to \mathbb{R}$ et $h : [0,1] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ des fonctions Lipschitziennes. Considérons l'équation aux dérivées partielles semi-linéaire suivante:

$$\begin{cases} 0 = \mathcal{L}^X u(t,x) - h(t,x,u(t,x),\sigma(t,x) \nabla_x u(t,x)) & \text{sur } [0,1] \times \mathbb{R}^d, \\ g(x) = u(1,x) & \text{sur } \mathbb{R}^d, \end{cases}$$
(10)

où \mathcal{L}^X est l'opérateur linéaire de dérivation

$$\mathcal{L}^X u := \frac{\partial u}{\partial t} + \nabla_x u \, b + \frac{1}{2} \sum_{i,j=1}^d (\sigma \sigma^*)^{i,j} \frac{\partial^2 u}{\partial x^i \partial x^j} \, .$$

Etant donné un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, cet opérateur de dérivation s'interprète comme l'opérateur de Dynkin associé à la solution de l'EDS suivante

$$X_t = x + \int_0^t b(s, X_s) ds - \int_0^t \sigma(s, X_s) \cdot dW_s \quad t \le 1,$$
 (11)

où W est un mouvement Brownien sous \mathbb{P} et x la valeur initiale du processus X, solution dont l'existence et l'unicité sont assurées par le caractère Lipschitzien de b et σ . Heuristiquement, si u est une solution régulière de l'EDP (10), en appliquant la formule d'Itô au processus défini sur [0,1] par $Y_t := u(t, X_t)$ et en posant $Z_t := \sigma(t, X_t) \nabla_x u(t, X_t)$, on obtient la relation

$$Y_t = g(1, X_1) + \int_t^1 h(s, X_s, Y_s, Z_s) ds - \int_t^1 Z_s \cdot dW_s, \quad t \le 1.$$
(12)

L'EDP (10) est donc étroitement connectée à l'EDSPR découplée donnée par (11)-(12).

Inversement, partant directement d'une équation rétrograde de la forme (12), Pardoux et Peng [84, 85] ont démontré l'existence d'une unique solution progressivement mesurable $(Y, Z) \in \mathcal{S}^2_{[0,1]} \times \mathcal{L}^2_{[0,1]}$ satisfaisant les conditions d'intégrabilité

$$\|Y\|_{\mathcal{S}^{2}_{[0,1]}} + \|Z\|_{\mathcal{L}^{2}_{[0,1]}} := \mathbb{E}\left[\sup_{0 \le s \le 1} |Y_{s}|^{2}\right]^{\frac{1}{2}} + \mathbb{E}\left[\left(\int_{0}^{1} |Z_{s}|^{2} ds\right)\right]^{\frac{1}{2}} < \infty.$$
(13)

De plus, Y_t peut s'écrire sous la forme $u(t, X_t)$, où la fonction déterministe u est solution de viscosité de l'EDP semi-linéaire (10). La valeur à la date t = 0 de la fonction u que l'on cherche à estimer est donc donnée par

$$u(0,x) = \mathbb{E}\left[g(X_1) + \int_0^1 h(s, X_s, Y_s, Z_s)ds\right].$$
 (14)

Pour un entier n > 0 donné, considérons une grille régulière $\pi := (t_i)_{i \le n}$ de [0, 1] et introduisons X^{π} le schéma d'Euler associé au processus X défini récursivement par

$$X_0^{\pi} := x, \quad \text{et} \quad X_{t_i+1}^{\pi} := b(t_i, X_{t_i}^{\pi}) \Delta t_i + \sigma(t_i, X_{t_i}^{\pi}) \Delta W_{t_i}, \quad i < n,$$
(15)

où $\Delta t_i := t_{i+1} - t_i = 1/n$ et $\Delta W_{t_i} := W_{t_{i+1}} - W_{t_i}$. Dans le cas d'EDPs linéaires, on se trouve dans le cadre de la représentation de Feynman-Kac et le générateur hest une fonction qui ne dépend que de ses deux premières composantes. Il est possible d'approcher numériquement de manière classique u(0, x), donné par (14), à l'aide de simulations Monte Carlo de X^{π} . L'erreur d'approximation est la superposition de l'erreur statistique due à l'utilisation de simulations Monte Carlo pour approcher l'opérateur d'espérance et de l'erreur de discrétisation due à l'utilisation de X^{π} à la place de X, cette deuxième étant de l'ordre de $n^{-1/2}$ (voir [67] par exemple).

Dans le cas d'EDPs semi-linéaires, cette approche ne s'applique plus car elle nécessite la connaissance des processus (Y, Z) le long de chaque trajectoire. De nombreux algorithmes reposant sur l'approximation du mouvement Brownien par un processus ne prenant qu'un nombre fini de valeurs on été proposés, par exemple dans [3], [21], [26], [28] ou [76]. Zhang [104, 105] puis Bouchard et Touzi [19] ont proposé le schéma numérique naturel suivant. Ils approchent tout d'abord le processus progressif X par son schéma d'Euler X^{π} à l'aide de (15), et $\bar{Y}_1^{\pi} := g(X_1^{\pi})$ fournit une approximation du processus Y à maturité. Pout tout i < n, on déduit alors de manière rétrograde une approximation $(\bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi})$ de (Y_{t_i}, Z_{t_i}) à l'aide de la relation

$$\begin{cases} \bar{Z}_{t_{i}}^{\pi} = n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \Delta W_{i+1} \mid \mathcal{F}_{t_{i}} \right] \\ \bar{Y}_{t_{i}}^{\pi} = \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_{i}} \right] + \frac{1}{n} h \left(t_{i}, X_{t_{i}}^{\pi}, \bar{Y}_{t_{i}}^{\pi}, \bar{Z}_{t_{i}}^{\pi} \right) . \end{cases}$$
(16)

La dernière équation étant implicite, elle se résout numériquement à l'aide d'une procédure de point fixe. Comme $(Y, Z) \in \mathcal{S}^2_{[0,1]} \times \mathcal{L}^2_{[0,1]}$, l'erreur de discrétisation du schéma est définie par

$$\operatorname{Err}_{n}(Y,Z) := \left\{ \max_{i < n} \sup_{t \in [t_{i}, t_{i+1}]} \mathbb{E}\left[|Y_{t} - \bar{Y}_{t_{i}}^{\pi}|^{2} \right] + \sum_{i=0}^{n-1} \int_{t_{i}}^{t_{i+1}} \mathbb{E}\left[|Z_{t} - \bar{Z}_{t_{i}}^{\pi}|^{2} \right] dt \right\}^{\frac{1}{2}}$$

Cette erreur est directement liée à la régularité de (Y, Z) et est traduite ici par la quantité

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 \right] dt, \quad \text{où} \quad \bar{Z}_{t_i} := n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} Z_t dt \mid \mathcal{F}_{t_i} \right]$$

Lorsque b, σ, g et h sont Lipschitz, Zhang [78] a démontré que ce terme est de l'ordre de n^{-1} conduisant à un contrôle sur l'erreur globale de discrétisation $\operatorname{Err}_n(Y, Z)$ en $n^{-1/2}$. Gobet, Lemor et Warin [73] ont obtenu une vitesse de convergence similaire en considérant un schéma totalement explicite où la deuxième équation de (16) est remplacée par

$$\bar{Y}_{t_i}^{\pi} = \mathbb{E}\left[\bar{Y}_{t_{i+1}}^{\pi} + \frac{1}{n} h\left(t_i, X_{t_i}^{\pi}, \bar{Y}_{t_{i+1}}^{\pi}, \bar{Z}_{t_i}^{\pi}\right) \mid \mathcal{F}_{t_i}\right].$$

Pour être utilisables en pratique, ces deux schémas nécessitent le calcul de nombreuses espérances conditionelles. Trois principales méthodes ont été proposées pour combiner ces schémas à des techniques d'approximation des opérateurs d'espérance conditionnelle. Gobet, Lemor et Warin [73] étudient une adaptation de l'algorithme de Longstaff et Schwartz reposant sur des techniques de régression non paramétrique. Bally et Pages [8] utilisent des techniques de quantification dans le cas particulier d'équations rétrogrades réfléchies où h ne dépend pas de Z, techniques qui furent reprises par Delarue et Menozzi [38, 39] dans un cadre très général d'EDSPR couplée. Enfin, Bouchard et Touzi [19] utilisent une technique d'intégration par parties reposant sur le calcul de Malliavin.

Introduisons maintenant une mesure de Poisson μ , indépendante de W, d'espace de marque E et de compensateur $\bar{\mu}(de, ds) := \mu(de, ds) - \lambda(de)ds$ avec λ une mesure finie.

Ajoutant des sauts à la dynamique de X à l'aide de $\beta : \mathbb{R}^d \times E \to \mathbb{R}^d$, la représentation martingale de Y fait apparaître des sauts dans sa dynamique. On considère alors l'EDSPR découplée plus générale

$$\begin{cases} X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s + \int_0^t \int_E \beta(s, X_{s-}, e) \bar{\mu}(de, ds) , \\ Y_t = g(X_1) + \int_t^1 h(s, X_s, Y_s, Z_s, \Gamma_s) ds - \int_t^1 Z_s \cdot dW_s - \int_t^1 \int_E U_s(e) \bar{\mu}(de, ds) . \end{cases}$$
(17)

avec $\Gamma := \int_E \rho(e) U(e) \lambda(de)$ et ρ une fonction donnée. En supposant $\beta(0,.)$ et ρ bornés ainsi que $\beta(.,e)$ Lipschitz uniformément en $e \in E$, Tang et Li [100] ont obtenu l'existence d'une unique solution $(X, Y, Z, U) \in \mathcal{S}^2_{[0,1]} \times \mathcal{S}^2_{[0,1]} \times \mathcal{L}^2_{[0,1]} \times \mathcal{L}^2_{\lambda,[0,1]}$ à l'EDSPR (17) satisfaisant (13) et

$$\|U\|_{\mathbf{L}^{2}_{\lambda,[0,1]}} := \mathbb{E}\left[\int_{0}^{1} \int_{E} |U_{s}(e)|^{2} \lambda(de) ds\right]^{\frac{1}{2}} < \infty.$$
(18)

Barles, Buckdahn et Pardoux [5] remarquent que, pout tout t, Y_t s'écrit toujours $u(t, X_t)$, avec u solution de viscosité de l'équation Intégro-différentielle suivante

$$\begin{cases} 0 = \mathcal{L}^X u - \int_E \beta(., e) \lambda(de) + \mathcal{I}^1[u] - h(., u, \sigma \nabla_x u, \mathcal{I}^{\rho}[u]) & \text{sur } [0, 1] \times \mathbb{R}, \\ g = y(1, .) & \text{sur } \mathbb{R}, \end{cases}$$
(19)

où \mathcal{I}^{ϱ} est un opérateur Intégro-différentiel défini par

$$\mathcal{I}^{\varrho}[u](t,x) := \int_{E} \{u(t,x+\beta(x,e)) - u(t,x)\} \,\varrho(e) \,\lambda(de) \;. \tag{20}$$

Précisons pour finir que, dans le cas particulier où le générateur h ne dépend pas de Γ , c'est à dire de U, le schéma proposé par Gobet, Lemor et Warin [73] permet également la résolution de l'EDSPR (17) avec une erreur de l'ordre de $n^{-1/2}$.

Résultats nouveaux

La deuxième partie de cette thèse propose un algorithme numérique probabiliste de résolution de système d'EDSPR découplées de la forme (17). Nous présentons tout d'abord un travail réalisé en collaboration avec Bruno Bouchard qui généralise les schémas numériques présentés ci-dessus à la résolution de ce type d'équations. Puis, nous étudions l'erreur statistique due à l'approximation des espérances conditionnelles de ce schéma à l'aide de techniques de régression non-paramétrique, et nous présentons des résultats numériques.

Afin d'assurer l'existence d'une unique solution à (17) satisfaisant (13) et (18), nous supposons que les fonctions b, σ, g, h et $\beta(., e)$ sont Lipschitz uniformément en $e \in E$, et que $\beta(0, .)$ et ρ sont bornées.

L'approximation d'Euler X^{π} de X présentée en (15) prend désormais la forme suivante

$$\begin{cases} X_0^{\pi} := x \\ X_{t_{i+1}}^{\pi} := X_{t_i}^{\pi} + \frac{1}{n} b(X_{t_i}^{\pi}) + \sigma(X_{t_i}^{\pi}) \Delta W_{i+1} + \int_E \beta(X_{t_i}^{\pi}, e) \bar{\mu}(de, (t_i, t_{i+1}]) \,. \end{cases}$$
(21)

On en déduit l'approximation $\bar{Y}_1^{\pi} := g(1, X_1^{\pi})$ de Y_1 mais, afin d'adapter l'approximation rétrograde de Y présentée dans (16), il faut trouver un moyen d'approcher le processus (Z, Γ) . Etudions donc plus précisément le comportement de (Y, Z, U) sur chaque intervalle $[t_i, t_{i+1}]$. Etant donnée $\bar{Y}_{t_{i+1}}^{\pi}$ approximation de $Y_{t_{i+1}}$, le théorème de représentation des martingales assure l'existence d'un processus $(Z^{\pi}, U^{\pi}) \in \mathcal{L}^2_{[t_i, t_{i+1}]} \times \mathcal{L}^2_{\lambda, [t_i, t_{i+1}]}$ satisfaisant

$$\bar{Y}_{t_{i+1}}^{\pi} = \mathbb{E}\left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_i}\right] + \int_{t_i}^{t_{i+1}} Z_s^{\pi} \cdot dW_s + \int_{t_i}^{t_{i+1}} \int_E U_s^{\pi}(e)\bar{\mu}(de, ds) \,.$$

Remarquons que les meilleures approximations dans $\mathcal{L}^2_{[t_i,t_{i+1}]}$ des deux processus Z^{π} et $\Gamma^{\pi} := \int_E \rho(e) U^{\pi}(e) \lambda(de)$ par des variables aléatoires \mathcal{F}_{t_i} -mesurable sont données par

$$\bar{Z}_{t_i}^{\pi} := \mathbb{E}\left[\int_{t_i}^{t_{i+1}} Z_s^{\pi} ds \mid \mathcal{F}_{t_i}\right] \quad \text{et} \quad \bar{\Gamma}_{t_i}^{\pi} := \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \int_E \rho(e) U_s^{\pi}(e) \lambda(de) ds \mid \mathcal{F}_{t_i}\right],$$

qui sont donc candidats pour approcher Z et Γ . Gelant sur l'intervalle $[t_i, t_{i+1}]$, le processus (X, Y, Z, Γ) en la variable aléatoire \mathcal{F}_{t_i} -mesurable $(X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi})$, avec $\bar{Y}_{t_i}^{\pi}$ encore indéterminé, nous obtenons

$$\bar{Y}_{t_i}^{\pi} = \bar{Y}_{t_{i+1}}^{\pi} + h(t_i, X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi}) \Delta t_i - \int_{t_i}^{t_{i+1}} Z_s^{\pi} \cdot dW_s - \int_{t_i}^{t_{i+1}} \int_E U_s^{\pi}(e) \bar{\mu}(de, ds) \, .$$

Prenant alors l'espérance conditionnelle sachant \mathcal{F}_{t_i} de cette équation, multipliée respectivement par 1, ΔW_{t_i} et $\int_E \rho(e)\bar{\mu}(de, (t_i, t_{i+1}])$, nous proposons le schéma récursif suivant

$$\begin{cases} \bar{Z}_{t_{i}}^{\pi} := n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \Delta W_{i+1} \mid \mathcal{F}_{t_{i}} \right] \\ \bar{\Gamma}_{t_{i}}^{\pi} := n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \int_{E} \rho(e) \bar{\mu}(de, (t_{i}, t_{i+1}]) \mid \mathcal{F}_{t_{i}} \right] \\ \bar{Y}_{t_{i}}^{\pi} := \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_{i}} \right] + \frac{1}{n} h\left(t_{i}, X_{t_{i}}^{\pi}, \bar{Y}_{t_{i}}^{\pi}, \bar{Z}_{t_{i}}^{\pi}, \bar{\Gamma}_{t_{i}}^{\pi} \right) . \end{cases}$$
(22)

L'erreur de discrétisation de ce schéma doit prendre en compte l'erreur d'estimation de Γ et est donnée par

$$\operatorname{Err}_{n}(Y, Z, U) := \left\{ \max_{i < n} \sup_{t \in [t_{i}, t_{i+1}]} \mathbb{E}\left[|Y_{t} - \bar{Y}_{t_{i}}^{\pi}|^{2} \right] + \sum_{i=0}^{n-1} \int_{t_{i}}^{t_{i+1}} \mathbb{E}\left[|Z_{t} - \bar{Z}_{t_{i}}^{\pi}|^{2} + |\Gamma_{t} - \bar{\Gamma}_{t_{i}}^{\pi}|^{2} \right] dt \right\}^{\frac{1}{2}}.$$

Nous obtenons alors le contrôle suivant sur cette erreur

$$\operatorname{Err}_{n}(Y, Z, U) \leq C \left(n^{-1/2} + \|Z - \bar{Z}\|_{\mathcal{L}^{2}} + \|\Gamma - \bar{\Gamma}\|_{\mathcal{L}^{2}} \right) \underset{n \to \infty}{\longrightarrow} 0, \qquad (23)$$

où C est une constante générique et $(\overline{Z}, \overline{\Gamma})$ est, sur chaque intervalle $[t_i, t_{i+1}]$, un processus égal à la meilleure approximation dans $\mathcal{L}^2_{[t_i, t_{i+1}]}$ de (Z, Γ) par une variable aléatoire \mathcal{F}_{t_i} -mesurable. Ce processus est donné sur chaque intervalle $[t_i, t_{i+1}]$ par

$$\bar{Z}_t := n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} Z_s \, ds \mid \mathcal{F}_{t_i}\right] \quad \text{et} \quad \bar{\Gamma}_t := n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \Gamma_s \, ds \mid \mathcal{F}_{t_i}\right] \,,$$

et permet de traduire une fois de plus la régularité de la solution de l'EDSPR (17). Notons également qu'un schéma explicite adapté de [73], où la dernière équation de (22) est remplacée par

$$\bar{Y}_{t_i}^{\pi} := \mathbb{E}\left[\bar{Y}_{t_{i+1}}^{\pi} + \frac{1}{n} h\left(X_{t_i}^{\pi}, \bar{Y}_{t_{i+1}}^{\pi}, \bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi}\right) \mid \mathcal{F}_{t_i}\right],$$

bénéficie également d'un contrôle sur l'erreur de type (23).

Afin d'améliorer la borne obtenue sur notre erreur, nous avons étudié plus en détail la régularité de (Y, Z, U) à l'aide du calcul de Malliavin sur l'espace de Wiener. En effet, le processus (X, Y, Z, U) est différentiable au sens de Malliavin, et sa dérivée satisfait une EDSPR découplée linéaire. Ainsi, remarquant que Z s'interprète à l'aide de la dérivée de Malliavin de Y et que U traduit les sauts de Y, nous avons obtenu des propriétés de régularité trajectorielle sur les processus (X, Y, Z, U), qui impliquent en particulier

$$\|\Gamma - \bar{\Gamma}\|_{\mathcal{L}^2} \leq C n^{-1/2} \quad \text{et} \quad \|Z - \bar{Z}\|_{\mathcal{L}^2} \leq C_{\varepsilon} n^{-1/2 + \epsilon}, \quad \text{pour tout } \epsilon > 0.$$

On obtient ainsi une borne en $n^{-1/2+\epsilon}$ pour tout $\epsilon > 0$ sur la vitesse de convergence de l'algorithme. Dans le cas particulier où le terme de sauts du processus progressif X satisfait une condition de non-dégénérescence, nous obtenons la vitesse optimale en $n^{-1/2}$ en étudiant l'EDSPR dont le processus tangent de (X, Y, Z, U) est solution. Cette vitesse optimale est également obtenue lorsque les coefficients b, σ, g, h et $\beta(., e)$ sont C_b^1 à dérivées Lipschitz, uniformément en $e \in E$.

Afin d'être utilisable en pratique, ce schéma nécessite l'estimation d'un grand nombre d'espérances conditionnelles. Nous étendons les résultats de Gobet, Lemor et Warin [73] en étudiant la propagation de l'erreur statistique due à l'approximation des opérateurs d'espérance conditionnelle à l'aide de techniques de régression non-paramétrique. Nous obtenons un majorant de l'erreur globale de l'algorithme qui nous permet de choisir dans le même temps le nombre de simulations Monte Carlo, le pas de discrétisation en temps et le nombre de fonctions de base à utiliser.

Application aux systèmes couplés d'EDP semi-linéaires

Un autre résultat remarquable sur les EDSPR avec sauts est la manière dont elles peuvent se lier à des solutions de systèmes couplés d'EDP. Considérons en effet un système couplé de deux EDPs de la forme suivante

$$\begin{cases} \mathcal{L}_0^X u_0 + h_0(\cdot, (u_0, u_1), \sigma_0 \nabla_x u_0) = 0, & u_0(1, \cdot) = g_0, \\ \mathcal{L}_1^X u_1 + h_1(\cdot, (u_0, u_1), \sigma_1 \nabla_x u_1) = 0, & u_1(1, \cdot) = g_1, \end{cases}$$
(24)

où, pour i = 0 ou 1, b_i , σ_i , g_i et h_i sont des fonctions Lipschitz et \mathcal{L}_i^X est l'opérateur linéaire associé à b_i et σ_i . Les fonctions h_0 et h_1 sont des fonctions du couple solution (u_0, u_1) que nous modifions comme suit

$$\tilde{h}_0: (., u, ., \gamma) \mapsto h_0(., (u, u + \gamma), z) - \lambda \gamma \quad \text{et} \quad \tilde{h}_1: (., u, ., \gamma) \mapsto h_1(., (u + \gamma, u), z) - \lambda \gamma,$$

en se fixant λ quelconque dans \mathbb{R} . Oubliant le dernier terme technique de compensation de la forme $\lambda\gamma$, cette modification permet d'écrire respectivement $h_0(., (u_0, u_1), .)$ et $h_1(., (u_0, u_1), .)$ sous la forme de fonctions de $(u_0, u_1 - u_0)$ et $(u_1, u_0 - u_1)$.

Introduisons alors une mesure de poisson μ sur $E = \{1\}$ de compensateur égal à la mesure de comptage multipliée par λ et considérons l'EDSPR suivante

$$\begin{cases} M_t \equiv \int_0^t \int_E e\,\mu(de, ds) \pmod{2} ,\\ X_t = \int_0^t b_{M_r}(r, X_r) dr + \int_0^t \sigma_{M_r}(r, X_r) dW_r ,\\ Y_t = g_{M_1}(X_1) + \int_t^1 \tilde{h}_{M_r}(r, X_r, Y_r, Z_r, U_r(1)) dr - \int_t^1 Z_r \cdot dW_r - \int_t^1 \int_E U_r(e) \bar{\mu}(de, dr) . \end{cases}$$

Pardoux, Pradeilles et Rao [86] ont démontré que le couple (u_0, u_1) de fonctions déterministes, tel que la composante rétrograde de la solution de cette EDSPR satisfait $Y_t = u_{M_t}(t, X_t)$ sur [0, 1], est solution de viscosité du système couplé d'EDP (24). La première composante du processus progressif (M, X) est un processus de sauts pur basculant à chaque saut entre les valeurs 0 et 1. Sa valeur va s'interpréter comme le numéro de la composante de la solution de (24). En effet, plaçons nous entre deux sauts consécutifs et appliquons les résultats de liens entre EDSPR sans sauts et EDP semi-linéaire présentés préalablement. Lorsque M = 0 et si $U(1) = u_1(., X) - u_0(., X)$, l'utilisation du générateur \tilde{h}_0 modifié permet de lier l'EDSPR sans sauts considérée à la solution $u_M = u_0$ de la première équation du système. De même, si M = 1 et $U(1) = u_0(., X) - u_1(., X)$, la solution de l'EDSPR s'interprète à l'aide de $u_M = u_1$. Comme le processus U(1) traduit les sauts de Y, il est naturel qu'il prenne successivement les valeurs $u_1(., X) - u_0(., X)$ et $u_0(., X) - u_1(., X)$ dès que $Y = u_M(., X)$, ce qui justifie le raisonnement précédent. Notre algorithme s'adapte également à la résolution d'EDSPR de cette forme. En effet, nous simulons tout d'abord parfaitement le processus de saut pur M, puis le processus progressif X à l'aide de son schéma d'Euler X^{π} en ajoutant dans la grille régulière π les temps de sauts de M. Nous obtenons donc une approximation $\bar{Y}_1^{\pi} = g_{M_1}(1, X_1^{\pi})$ de Y_1 et n'ayant pas d'information sur la régularité du générateur \tilde{h} comme fonction de Mnous adaptons la version explicite du schéma (22) en le remplaçant par

$$\begin{cases} \bar{Z}_{t_{i}}^{\pi} := n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \Delta W_{i+1} \mid \mathcal{F}_{t_{i}} \right] \\ \bar{\Gamma}_{t_{i}}^{\pi} := n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \int_{E} \bar{\mu}(de, (t_{i}, t_{i+1}]) \mid \mathcal{F}_{t_{i}} \right] \\ \bar{Y}_{t_{i}}^{\pi} := \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} + \int_{t_{i}}^{t_{i+1}} \tilde{h}_{M_{s}} \left(t_{i}, X_{t_{i}}^{\pi}, \bar{Y}_{t_{i+1}}^{\pi}, \bar{Z}_{t_{i}}^{\pi}, \bar{\Gamma}_{t_{i}}^{\pi} \right) ds \mid \mathcal{F}_{t_{i}} \right].$$

$$(25)$$

Cet algorithme converge et nous obtenons le contrôle de l'erreur suivant

$$\operatorname{Err}_{n}(Y, Z, U) \leq C\left(n^{-1/2} + \|H - \bar{H}\|_{\mathcal{L}^{2}}\right) \underset{n \to \infty}{\longrightarrow} 0, \qquad (26)$$

où H et \bar{H} sont définis sur chaque intervalle $[t_i, t_{i+1}]$ par $H_t := h_{M_t}(t_i, X_{t_i}, Y_{t_i}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i})$ et $\bar{H}_t = \mathbb{E}\left[\int_{t_i}^{t_{i+1}} H_s ds \mid \mathcal{F}_{t_i}\right]$. Ainsi \bar{H} est la meilleure approximation de H dans tout $\mathcal{L}^2[t_i, t_{i+1}]$ par une variable aléatoire \mathcal{F}_{t_i} - mesurable, et le terme $||H - \bar{H}||_{\mathcal{L}^2}$ traduit la régularité de la solution de l'EDSPR par rapport à M, soit l'écart entre les deux solutions du système (24). Pour tout entier k, notre algorithme permet également la résolution de système couplé de k EDPs, le processus M faisant alors des sauts de k-1tailles différentes.

Nous présentons dans la Section 2.2 quelques exemples numériques de résolution de système couplé d'EDPs, dans lesquels nous approchons les opérateurs d'espérance conditionnelle à l'aide de projection sur des bases de polynômes. Nous considérons en particulier la valorisation d'un produit dérivé dont le vendeur peut faire défaut et les résultats numériques sont probants quant à la convergence de l'algorithme.

Perspectives

Dans un premier temps, nous pourrions étudier plus précisément la vitesse exacte de convergence de l'algorithme (25) en regardant en particulier l'influence du paramètre λ qui calibre la fréquence des sauts. Empiriquement, si λ est très petit, l'algorithme a des difficultés à capturer la dynamique de chacune des deux solutions du système d'EDPs. De même, si λ est très élevé, la précision des estimations souffre d'un nombre de saut trop élevé sur chaque intervalle $[t_i, t_{i+1}]$. Un choix arbitraire de λ conduit à des estimations précises mais il serait intéressant d'essayer de déterminer le choix du λ calibrant la fréquence de sauts optimale sur chaque intervalle $[t_i, t_{i+1}]$. La difficulté théorique pour l'obtention de cette fréquence de saut optimale réside dans la dépendance en λ du générateur \tilde{h} et donc de sa constante de Lipschitz.

De la même manière que l'on peut lier les EDPs semi-linéaires à les EDSPR découplées, les EDPS quasi-linéaire peuvent également s'interpréter à l'aide d'EDSPR couplées. Il s'agit donc d'EDSPR dont la dynamique du processus progressif dépend de la solution de l'équation rétrograde et en l'absence de sauts prennent la forme suivante

$$\begin{cases} X_t = X_0 + \int_0^t b(s, X_s, Y_s, Z_s) ds + \int_0^t \sigma(s, X_s, Y_s) dW_s \\ Y_t = g(X_1) + \int_t^1 h(s, X_s, Y_s, Z_s, \Gamma_s) ds - \int_t^1 Z_s \cdot dW_s , \end{cases}$$
(27)

L'existence et l'unicité du triplet (X, Y, Z) solution de ce système sont assurées pour des coefficients Lipschitz et une volatilité σ non dégénérée (voir par exemple les travaux de Delarue [37]). La difficulté numérique pour la résolution de tels systèmes réside dans la nécessité de simuler le processus progressif et d'estimer le processus rétrograde dans le même temps. Delarue et Menozzi [38, 39] proposent un algorithme reposant sur des techniques de quantification permettant la résolution de ce type d'EDSPR couplée. Citons également Bender et Zhang [11] qui, à l'aide d'un algorithme itératif, approchent numériquement la solution de ces équations, dans le cas particulier où *b* ne dépend pas de *Z*. Une piste de recherche serait l'étude du cadre dans lequel ces deux algorithmes peuvent être adaptés à la résolution d'EDSPR couplées avec sauts, équations pour lesquelles les résultats de Pardoux et Sow [98] peuvent assurer l'existence d'une unique solution. De même, de récents travaux de Bouchard et Chassagneux [18] ont amélioré les résultats de convergence obtenus par Zhang [104] pour la résolution numérique d'EDSPR réfléchies, et l'ont pourrait étudier l'influence sur leurs résultats de l'ajout des sauts à la dynamique des processus.

La convergence de notre algorithme nécessite actuellement la manipulation d'EDSPR à coefficients Lipschitz, hypothèses que l'on souhaiterait pouvoir réduire. Le générateur h peut par exemple se contenter d'être 1/2-Hölder en temps, mais diminuer les autres hypothèses semble malheureusement difficile. Il existe de nombreux résultats d'existence de solution aux EDSR sous des hypothèses plus faibles, lorsque le générateur est, par exemple, continue, monotone en Y ou quadratique en Z comme remarqué récemment par Briand et Hu [22], mais les résultats d'unicité sont plutôt rares. L'obtention nécessaire de régularité sur la solution que l'on cherche à approcher en est alors fortement compromise. Cependant, lorsque la fonction g est bornée, les EDSPR dont le générateur est simplement quadratique en Z admette une unique solution. Ces résultats ont été obtenus par Kobylanski [68], qui s'est inspirée de techniques issues de l'étude d'EDP, puis généralisés

par Rong [95] et Becherer [9] lorsque l'on ajoute des sauts aux processus. Leurs démonstrations reposent sur un changement de variable de type exponentiel rendu possible car le processus Y est borné dès que la fonction g l'est également. Notons cependant qu'une EDSR quadratique de condition terminale non bornée admet également une solution comme remarqué par Briand et Hu [22]. La difficulté pour faire converger notre algorithme réside dans l'obtention de la régularité trajectorielle du processus (Y, Z). Il est toujours possible d'approcher le générateur h à l'aide d'une suite de fonctions h_p de constante de Lipschitz K_p tendant vers l'infini. L'algorithme obtenu est convergent mais avec de vitesse très lente. En effet, l'utilisation du lemme de Gronwall entraîne l'apparition de termes en e^{K_p} dans la borne de l'erreur d'approximation.

Remarquons qu'une autre méthode est également possible en étudiant le cas particulier d'un générateur quadratique h qui se décompose en la somme d'une fonction Lipschitz h' et de $z \mapsto z^2$. L'EDSPR considérée a alors la forme suivante

$$\begin{cases} X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s \\ Y_t = g(X_1) + \int_t^1 \left[h'(s, X_s, Y_s, Z_s) + Z_s^2 \right] ds - \int_t^1 Z_s \cdot dW_s . \end{cases}$$
(28)

Comme détaillé par Ankirchner, Imkeller et Popier [2], le processus $\int_0^{\cdot} Z_s \cdot dW_s$ est une BMO martingale. Ainsi, le processus W^z défini sur [0, 1] par $W_t^z := W_t - \int_0^t Z_s ds$ est un mouvement Brownien sous une nouvelle probabilité. L'EDSPR (28) s'écrit alors sous la forme

$$\begin{cases} X_t = X_0 + \int_0^t [b(s, X_s) + Z_s] ds + \int_0^t \sigma(s, X_s) dW_s^z \\ Y_t = g(X_1) + \int_t^1 h'(s, X_s, Y_s, Z_s) ds - \int_t^1 Z_s \cdot dW_s^z. \end{cases}$$

qui est une EDSPR couplée, dont on peut approcher la solution à l'aide de l'algorithme de Delarue et Menozzi [38, 39]. Cependant cet algorithme présente le défaut de nécessiter une discrétisation de l'espace, au risque de perdre en grande dimension l'avantage possible des méthode probabilistes sur leurs équivalents déterministes. En cela, un algorithme reposant sur les simulations du processus progressif puis sur une approximation rétrograde de Y pourrait être plus performant en grande dimension. Signalons qu'une résolution numérique efficace de ce type d'EDSPR serait extrêmement utile au domaine de contrôle optimal stochastique, pour lequel, par exemple, la maximisation d'utilité de type exponentielle conduit à l'obtention d'EDSPR quadratiques. Citons, par exemple, les récents travaux de Porchet, Touzi et Warin [91] qui utilisent justement ce type de techniques.

Investissement et consommation sous contrainte drawdown

Motivation

Les marchés offrent de nombreuses opportunités d'investissement dans divers produits financiers. Chaque gestionnaire de fond doit alors choisir dans quels actifs investir, dans quelles proportions et sur quelle période. Etant donnée une fonction d'utilité Ucaractérisant ses préférences ou celles des investisseurs qu'il représente, le gestionnaire cherche donc une stratégie optimale d'investissement θ dans un panier d'actif S lui permettant de maximiser l'utilité de ses revenus futurs. En lui donnant, de plus, la possibilité de verser aux investisseurs une rente, s'interprétant économiquement comme une consommation C, la valeur $X^{x,C,\theta}$ de son portefeuille de capital initial x s'écrit

$$X_t^{x,C,\theta} = x + \int_0^t \theta_r \cdot dS_r - \int_0^t C_r dr, \quad t \ge 0.$$
⁽²⁹⁾

Etant donné un horizon de vie T à son porte feuille, le gestionnaire a le comportement d'un agent économique cherchant à résou dre

$$\max_{C,\theta} \int_0^T e^{-\beta s} U(C_s) \, ds \,, \tag{30}$$

le facteur β traduisant sa préférence pour le présent.

Merton [79, 80] propose en 1970 une solution à ces problèmes dans un cadre continu d'évolution des actifs financiers. Supposant une dynamique de type Black-Scholes sur ces actifs, il parvient à résoudre l'équation d'Hamilton-Jacobi-Bellman correspondante pour certaines fonction d'utilité dont la fonction d'utilité puissance

$$U_p(x) = \frac{x^p}{p}, \quad x \ge 0, \quad \text{avec } p \in (0,1).$$
 (31)

A l'aide d'un principe de dualité, Bismut [16] obtient une nouvelle démonstration de ces résultats, qui, adaptée par Cox et Huang [29] et Karatzas, Lehoczky et Shreeve [64], permet de traiter le cas d'actifs financiers de dynamique non Markovienne. Ils généralisèrent ainsi les conclusions de Pliska [89] qui portaient sur un agent maximisant l'utilité de sa richesse terminale. Une littérature très vaste traite de l'extension de ces résultats en présence de différents types d'imperfections sur le marché, dont voici quelques exemples. L'introduction de contraintes sur la stratégie d'investissement est ainsi traitée de manière probabiliste par Cvitanic et Karatzas [33], ou à l'aide de techniques déterministes dans un cadre Markovien par Zariphopoulou [103]. L'ajout de coûts de transaction proportionnels est, entre autres, discuté par Constantinides et Magill [27], Davis et Norman [35] ou Shreve et Soner [97]. Permettre à l'investisseur de toucher un revenu en plus de ses investissements a été étudié par He et Pagès [62] ainsi qu'El Karoui et Jeanblanc [44]. Citons également l'article de Ben Tahar, Soner et Touzi [10] qui étudie un marché financier comportant des taxes sur les plus-values en capitaux. Pour finir, El Karoui, Jeanblanc et Lacoste [45] imposent à la richesse de l'investisseur de dominer à tout instant un processus donné, problème proche de ce que nous étudions ici.

Nous considérons un gestionnaire de fond qui cherche à attirer de nouveaux investisseurs et à leur proposer certaines garanties. Afin de les convaincre, il a besoin d'indicateurs traduisant les performances de leurs portefeuilles. En particulier, le drawdown d'un portefeuille est, par définition, donné par la différence entre le maximum courant du portefeuille et sa valeur actuelle. Les gestionnaires de fond peuvent en effet être remerciés suite à un drawdown trop important en valeur ou simplement trop long en durée. Nous considérons alors un gestionnaire de fond qui s'engage auprès de ses investisseurs à ce que la valeur du portefeuille ne descende pas en dessous d'une fraction $\alpha \in [0, 1)$ de son maximum courant. Il cherche la stratégie d'investissement θ et de consommation C lui permettant de maximiser l'utilité intertemporelle de sa consommation, donnée par (30), sous la "contrainte drawdown"

$$X^{x,C,\theta} \geq \alpha \left(X^{x,C,\theta} \right)^*, \quad \text{avec} \quad \left(X^{x,C,\theta}_t \right)^*_t := \max_{s \leq t} X^{x,C,\theta}_s, \quad t \geq 0.$$
(32)

La valeur du portefeuille doit ainsi rester au dessus d'un certain palier, dénommé "barrière *drawdown*", dont la valeur dépend des performances passées de ses investissements.

Etat de l'art

Dans un marché contenant un actif sans risque à rendement constant et un actif risqué de type Black-Scholes, Grossman et Zhou [59] furent les premiers à analyser le comportement d'un investisseur soumis à une contrainte *drawdown*. Cet agent ne bénéficie pas de possibilité de consommation intermédiaire et cherche à maximiser le taux de croissance à long terme de l'utilité de la valeur terminale de son portefeuille X, c'est à dire

$$\limsup_{T \to \infty} \frac{1}{T} \ln E[U_p(X_T)].$$

La stratégie optimale d'investissement, obtenue par résolution de l'équation d'Hamilton-Jacobi-Bellman correspondante, est alors une fonction linéaire de la distance entre la valeur du porte feuille et la fraction α de son maximum courant.

Cvitanic et Karatzas [34] étendent ces résultats au cadre d'un marché financier composé de plusieurs actifs de dynamique très générale, en imposant cependant à la contrainte
drawdown de porter sur les valeurs actualisées du portefeuille. Ils observent que toute stratégie d'investissement en proportion aléatoire de $(X - \alpha X^*)$ produit un portefeuille vérifiant la contrainte drawdown. Leur approche probabiliste très fine repose sur les propriétés de la martingale exponentielle $(X - \alpha X^*)(X^*)^{\frac{\alpha}{1-\alpha}}$ dès lors que la stratégie d'investissement s'exprime en proportion aléatoire de $(X - \alpha X^*)$. Notons cependant que Klass et Nowicki [66] démontrent que la stratégie proposée n'est plus optimale dans le cadre d'un marché où les actifs évoluent à des dates de temps discrètes. Citons enfin les travaux récents d'El Karoui et Meziou [43] qui considèrent des contraintes de type drawdown non nécessairement linéaires, et dont nous discutons les résultats à la fin de cette section. La principale critique que l'on peut formuler sur le critère de maximisation du taux de croissance à long terme de l'utilité espérée est que l'investisseur peut employer n'importe quelle stratégie d'investissement, si elle coincide avec la stratégie optimale à partir d'une date donnée.

Considérant un marché financier identique à celui de Grossman et Zhou [59], Roche [93] étudie le comportement d'un gestionnaire de fond cherchant à maximiser, sous une contrainte *drawdown*, l'utilité intertemporelle de sa consommation en horizon infini. Dans la cas particulier d'une utilité puissance, il propose une stratégie optimale d'investissement et de consommation du gestionnaire. Malgré une interprétation économique de ses résultats, il ne justifie cependant pas que sa solution résout le problème posé. Nous avons étudié le comportement d'un gestionnaire sujet à des objectifs similaires. Pour une classe générale de fonctions d'utilité, nous obtenons la stratégie optimale explicite en horizon infini, et nous donnons une caractérisation par EDP de la solution du problème en horizon fini.

Résultats nouveaux

Considérons un marché financier composé d'un actif risqué de dynamique

$$dS_t = \sigma S_t \left(dW_t + \lambda dt \right) \,,$$

avec W un mouvement Brownien, et d'un actif sans risque de valeur 1. Cette normalisation à l'unité de l'actif sans risque signifie simplement que les actifs financiers sont déjà écrits sous leur forme actualisée. Etant donné un capital initial x, la stratégie d'un gestionnaire de fond consistant à investir θ dans l'actif risqué et à consommer C, produit un portefeuille dont la valeur $X^{C,\theta}$ est donc donnée par

$$X_t^{x,C,\theta} = x - \int_0^t C_r dr + \int_0^t \sigma \theta_r \left(dW_r + \lambda dr \right), \quad t \ge 0.$$
(33)

Selon que l'investisseur ait ou non la possibilité de retirer ses fonds à tout instant, nous étudions le comportement d'un gestionnaire maximisant l'utilité intertemporelle de sa consommation sur un horizon fini ou infini.

Horizon infini

Le gestionnaire, caractérisé par une fonction d'utilité U quelconque, cherche à résoudre

$$\sup_{(C,\theta)\in\mathcal{A}_{\alpha}(x)} \mathbb{E}\left[\int_{0}^{\infty} e^{-\beta t} U\left(C_{t}\right) dt\right], \qquad (34)$$

ou $\mathcal{A}_{\alpha}(x)$ représente l'ensemble des stratégies satisfaisant certaines conditions d'intégrabilité ainsi que la contrainte drawdown (32). Pour simplifier cette présentation, nous supposons sans perte de généralité que U(0) = 0. Nous introduisons une version dynamique de notre problème

$$u^{\alpha}(x,z) := \sup_{(C,\theta)\in\mathcal{A}_{\alpha}(x,z)} \mathbb{E}\left[\int_{0}^{\infty} e^{-\beta t} U(C_{t}) dt\right], \qquad (35)$$

où x et z correspondent aux valeurs initiales des processus $X^{x,C,\theta}$ donné par (33) et $Z^{x,z,C,\theta} := z \vee (X^{x,C,\theta})^*$, et $\mathcal{A}_{\alpha}(x,z)$ est l'ensemble des stratégies satisfaisant de bonnes conditions d'intégrabilité ainsi que

$$X_t^{x,C,\theta} \ge \alpha Z_t^{x,z,C,\theta} \quad \text{p.s.}, \qquad t \ge 0.$$
(36)

Ainsi le domaine de définition de u^{α} est l'adhérence de $\mathbf{D}_{\alpha} := \{(x, z) : 0 < \alpha z < x \leq z\}$ dans \mathbb{R}^2 , dont nous notons $\partial^{\alpha} \mathbf{D}_{\alpha}$ et $\partial^1 \mathbf{D}_{\alpha}$ les bords contenant respectivement les éléments de la forme ($\alpha z, z$) et (z, z) avec z > 0. L'équation de la programmation dynamique associée à (35) est reliée à l'opérateur différentiel

$$\mathcal{L}\varphi := \sup_{C \ge 0, \theta \in \mathbb{R}} \mathcal{L}^{C,\theta}\varphi, \quad \text{avec } \mathcal{L}^{C,\theta}\varphi := -\beta\varphi + U(C) + (\theta\sigma\lambda - C)\varphi_x + \frac{\theta^2\sigma^2}{2}\varphi_{xx}.$$

Comme pour Cvitanic et Karatzas [34], la contrainte drawdown est exprimée en terme de processus actualisé. Une fois que la valeur du portefeuille du gestionnaire a touché sa barrière drawdown, il ne lui reste aucune possibilité d'investissement ou de consommation. La fonction valeur u^{α} est donc soumise à la contrainte de Dirichlet $u^{\alpha} = 0$ sur $\partial^{\alpha} \mathbf{D}_{\alpha}$. L'autre bord $\partial^{1} \mathbf{D}_{\alpha}$ du domaine $\overline{\mathbf{D}}_{\alpha}$ joue le rôle d'une barrière réfléchissante, et u^{α} y est soumis à la contrainte de Neumann $u_{z}^{\alpha} = 0$. Nous nous attendons donc à ce que la fonction valeur soit solution de l'équation de la programmation dynamique

$$-\mathcal{L}u^{\alpha} = 0 \text{ sur } \mathbf{D}_{\alpha}; \quad -u_{z}^{\alpha} = 0 \text{ sur } \partial^{1}\mathbf{D}_{\alpha}; \quad u^{\alpha} = 0 \text{ sur } \partial^{\alpha}\mathbf{D}_{\alpha} \cup \{(0,0)\}.$$
(37)

Notre démarche fut alors de trouver une solution régulière à cette équation puis d'appliquer un théorème de vérification nous assurant que notre candidat était bien solution du problème posé.

Les arguments de Cvitanic et Karatzas [34] peuvent être adaptés à notre problème, et toute stratégie (C, θ) écrite en proportion (c, π) de la distance entre la valeur du portefeuille et sa barrière drawdown est admissible, sous réserve de bonnes conditions d'intégrabilité des processus c et π . Nous cherchons donc une stratégie optimale de cette forme. Afin d'utiliser un principe de dualité, nous supposons que la fonction d'utilité Uest croissante, concave, continûment dérivable et satisfait les conditions d'Inada. Nous étudions alors la formulation duale de notre problème en introduisant la transformée de Legendre-Fenchel associée

$$v^{\alpha}(y,z) := \sup_{x \ge 0} (u^{\alpha}(x,z) - xy) .$$
 (38)

Comme observé par Xu [102], la duale v^0 de la fonction valeur u^0 du problème non contraint satisfait une EDP linéaire. La clef de notre résolution repose sur l'observation que v^{α} est également solution d'une EDP linéaire dès que u^{α} vérifie (37). Introduisant les fonctions φ et ψ définies sur \mathbb{R}_+ par $\varphi(z) = u_x^{\alpha}(z, z)$ et $\psi(z) = u_x^{\alpha}(\alpha z, z)$, v^{α} est en effet solution d'une EDP linéaire sur $[\varphi(z), \psi(z)]$ et satisfait

$$v_z^{\alpha}(y,z) = \varphi(z) - y \text{ pour } y \le \varphi(z), \text{ et } v_z^{\alpha}(y,z) = -\alpha yz \text{ pour } y \ge \psi(z).$$

Comme aucune possibilité de gain n'est possible pour le gestionnaire dès que la valeur de son portefeuille touche la barrière *drawdown*, nous cherchons une solution satisfaisant de plus $\psi = \infty$. De lourds calculs analytiques nous permettent alors de déterminer explicitement l'inverse de la fonction φ et d'en déduire v^{α} sous la condition

$$\frac{\gamma}{1+\gamma} < 1-\alpha, \quad \text{avec} \quad \gamma := \frac{2\beta}{\lambda^2},$$
(39)

qui est toujours vérifiée dans le cas non contraint $\alpha = 0$. Une inversion de la fonction v_y^{α} nous donne alors notre candidat à la résolution de (35) ainsi que les stratégies optimales d'investissement.

Afin d'assurer à notre problème d'être bien posé, nous supposons que l'élasticité asymptotique AE(U) de la fonction d'utilité du gestionnaire satisfait

$$AE(U) := \limsup_{x \to \infty} \frac{xU'(x)}{U(x)} \le (1-\alpha)\frac{\gamma}{\gamma+1}.$$

Dans un cadre très général, Kramkov et Shachermayer [70] ont introduit ce type d'hypothèse qui assure l'existence d'une stratégie optimale. Remarquons également que cette



Figure 1: Stratégie optimale vs la proportion de richesse x/z, pour α entre 0 et 0.6

hypothèse coincide avec celle de Merton pour la maximisation sans contrainte d'une utilité puissance. Nous ajoutons également une hypothèse technique sous laquelle l'équation différentielle stochastique vérifiée par la valeur \hat{X} du portefeuille associée à la stratégie optimale d'investissement et de consommation, admet une unique solution. Comme dans 38 605 528 762 cet Karatzas [34], notre stratégie optimale s'écrit en proportion de la distance entre \hat{X} et sa barrière drawdown $\alpha \hat{Z}$, et le processus $(\hat{X} - \alpha \hat{Z})\hat{Z}^{\frac{\alpha}{1-\alpha}}$ est une martingale exponentielle. Cette observation nous permet d'obtenir la condition de transversalité nécessaire pour l'argument de vérification qui conclut que notre candidat est bien solution du problème (35).

L'écriture analytique précise de la solution est donnée en Section 1.3.4 et nous présentons ici un exemple numérique dans le cas particulier où l'utilité est une fonction puissance du type (31), le choix des paramètres étant $\{p, \sigma, \lambda, \beta\} = \{0.2, 1, 3, 3\}$. La stratégie optimale associée à un portefeuille de valeur x et de maximum courant z, s'écrit alors en proportion de z à l'aide de fonctions dépendant uniquement de x/z. Cette caractéristique, qui provient de la propriété d'homogénéité de la fonction d'utilité puissance, avait permis à Roche [93] de deviner la forme de la solution et d'observer des résultats similaires. La Figure 1 présente la stratégie optimale du gestionnaire (en proportion de z) pour différentes valeurs de α satisfaisant (39), courbes qui se différencient facilement puisqu'elles partent de 0 au point $x/z = \alpha$. Son comportement s'interprète de la manière suivante. Lorsqu'il est proche de sa barrière drawdown, son investissement dans l'actif risqué et sa consommation diminuent si α augmente. L'investisseur anticipe en effet la possibilité de toucher sa barrière drawdown dans le futur. En revanche, pour α suffisamment grand, il a tendance à réduire son investissement et à augmenter sa consommation lorsqu'il approche de son maximum. Il a alors peur d'atteindre son maximum qui aurait pour conséquence de rehausser sa barrière *drawdown*. Dans le cas limite où $\alpha = 1/(1 + \gamma) = 0.6$, le gestionnaire ne cherche plus à augmenter la valeur de son portefeuille et se contente de consommer.

Horizon fini

Nous étudions maintenant le comportement de notre gestionnaire de fond ayant en charge un portefeuille de durée de vie déterminée. Soumis à la contrainte drawdown, il cherche à maximiser l'utilité intertemporelle de sa consommation sur une période donnée [0, T]. La version dynamique du problème prend alors la forme

$$u(t,x,z) := \sup_{(C,\theta)\in\mathcal{A}_{\alpha}(t,x,z)} \mathbb{E}\left[\int_{t}^{T} e^{-\beta r} U(C_{r}) dr\right], \qquad (40)$$

où x et z sont les valeurs initiales des processus définis sur [t, T] par

$$X_s^{t,x,C,\theta} = x - \int_t^s C_r dr + \int_t^s \theta_r \frac{dS_r}{S_r} \quad \text{et} \quad Z_s^{t,x,z,C,\theta} := z \vee \left\{ X^{t,x,C,\theta} \right\}_s^*,$$

et $\mathcal{A}_{\alpha}(t, x, z)$ l'ensemble des stratégies, satisfaisant de bonnes conditions d'intégrabilité, et vérifiant la contrainte drawdown (36) sur la période [t, T]. Le domaine de définition de u est ainsi donné par l'adhérence dans \mathbb{R}^3 de $\mathcal{O}_{\alpha} := [0, T) \times \{(x, z) : 0 < \alpha z < x < z\}$. Nous divisons le bord de ce domaine en quatre ensembles disjoints :

$$\partial^{\alpha} \mathcal{O}_{\alpha} := [0, T] \times \partial^{\alpha} \mathbf{D}_{\alpha}, \qquad \partial^{0} \mathcal{O}_{\alpha} := [0, T] \times \{(0, 0)\}, \\ \partial^{1} \mathcal{O}_{\alpha} := [0, T) \times \partial^{\alpha} \mathbf{D}_{\alpha}, \qquad \partial^{T} \mathcal{O}_{\alpha} := \{T\} \times \overline{\mathbf{D}}_{\alpha}.$$

L'introduction d'une dépendance temporelle dans la fonction valeur u empêche l'utilisation de notre approche précédente, rendant inextricables les calculs analytiques précédents déjà complexes. Cependant la fonction valeur u peut s'interpréter comme solution de viscosité de l'équation de la programmation dynamique correspondante. Cette notion de solution faible d'EDP, introduite par Crandall et Lions [31] est en effet très bien adaptée à la forme des équations d'Hamilton-Jacobi-Bellman. Son utilisation ne requière aucune régularité de la fonction candidate car les propriétés qu'elle doit satisfaire ne portent que sur ses enveloppes semi-continues. Signalons de plus que les schémas numériques d'approximation de solutions de viscosité convergent sous de très faibles propriétés de stabilité, comme observé par Barles et Souganidis [7]. Le lecteur intéressé pourra se reporter à l'article de Crandall, Ishii et Lions [30] pour une présentation complète et pédagogique de cette notion, ainsi qu'aux travaux de Huyen Pham [88] pour leurs applications en contrôle optimal stochastique et en finance. Pour toute fonction d'utilité U croissante et concave, nous démontrons que la fonction valeur u, définie en (40), est solution de viscosité de l'équation

$$u_t + \mathcal{L} u = 0 \quad \text{sur } \mathcal{O}_\alpha \cup \partial^\alpha \mathcal{O}_\alpha \,, \quad -u_z = 0 \quad \text{sur } \partial^1 \mathcal{O}_\alpha \,, \quad u = 0 \quad \text{sur } \partial^0 \mathcal{O}_\alpha \cup \partial^T \mathcal{O}_\alpha \,, \tag{41}$$

avec des conditions aux bords relaxées pour la propriété de sur-solution. L'obtention d'un théorème de comparaison fût ensuite nécessaire pour caractériser u comme l'unique solution de cette équation dans une classe de fonctions satisfaisant trois propriétés vérifiées par u, que nous détaillons ici. Tout d'abord, nous avons considéré des fonctions d'utilité U d'élasticité asymptotique inférieure à $\gamma/(\gamma + 1)$ afin de contrôler la croissance de u. Ensuite, nous avons remarqué que u s'annulait sur $\partial^0 \mathcal{O}_{\alpha} \cup \partial^T \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha}$ puisqu'aucune possibilité d'investissement et de consommation n'est alors possible. Enfin, grâce à la continuité à droite de $u \operatorname{sur} \overline{\mathcal{O}}_{\alpha} \setminus \partial^{\alpha} \mathcal{O}_{\alpha}$ le long de la bissectrice x = z, nous avons contourné la difficulté considérable due à l'absence de borne sur l'ensemble des stratégies admissibles. Notons également que des hypothèses plus fortes sur la fonction d'utilité U, permettant d'utiliser la fonction valeur u^{α} du problème en horizon infini comme majorant régulier de u, étendent le théorème de comparaison à une classe de fonctions non nécessairement nulles sur $\partial^{\alpha} \mathcal{O}_{\alpha}$ mais bénéficiant de continuité à droite sur $\overline{\mathcal{O}}_{\alpha}$ le long de la bissectrice x = z. Ces résultats d'unicité permettent ainsi l'approximation numérique de la fonction valeur u et sa comparaison à la solution u^{α} du problème en horizon infini.

Perspectives

Remarquons tout d'abord que la caractérisation par EDP de la solution du problème en horizon fini devrait pouvoir se généraliser assez facilement à l'étude d'un marché contenant des actifs financiers Markoviens de dynamique donnée par une équation différentielle stochastique assez générale. L'obtention d'une solution explicite au problème en horizon infini est également envisageable mais passe par une bonne compréhension de la dépendance temporelle de la fonction valeur et nécessite des calculs analytiques conséquents. Une étude numérique précise de la convergence de la fonction valeur en horizon fini vers la solution en horizon infini pourrait nous apporter des éclaircissements sur le type de solutions recherchées. Cette étude gagnerait à être complétée par une comparaison entre le comportement de gestionnaires soumis à des fonctions d'utilité de formes différentes.

La solution en horizon infini bien qu'explicite n'est pas entièrement satisfaisante. En particulier, son obtention nécessite l'inversion successive de deux fonctions. La première permet d'obtenir la frontière libre de la solution de l'EDP duale et la deuxième de déduire

la fonction valeur à partir de la solution du problème dual associé. Il est ainsi peut être possible de déterminer directement la fonction valeur sous une forme entièrement explicite. D'autre part, il est tentant de rechercher une démonstration purement probabiliste des résultats obtenus. Le cas échéant, il serait envisageable de les généraliser à l'étude du comportement d'un gestionnaire de fond pouvant investir dans des actifs financiers de dynamique plus complexe, éventuellement non Markovienne.

Citons pour finir les récents travaux d'El Karoui et Meziou [43] qui considèrent des contraintes *drawdown* non nécessairement linéaires de la forme

$$X_t \ge w(X_t^*) \quad \text{p.s.}, \qquad t \ge 0, \tag{42}$$

avec w une fonction plus petite que l'identité. Pour un actif financier réactualisé S de dynamique générale, elles démontrent que la martingale d'Azéma Yor M associée à l'inverse de la solution de l'EDP $[x - w(x)]\phi'(x) = \phi(x)$ est un portefeuille autofinançant réactualisé de dynamique

$$dM_t = (M_t - w[(M)_t^*]) \frac{dS_t}{S_t},$$

satisfaisant la contrainte (42). Cette martingale coincide avec le portefeuille optimal satisfaisant la contrainte drawdown linéaire (32), dans le cadre de travail de Cvitanic et Karatzas [34]. Ces observations sont encourageantes quant à la meilleure compréhension de nos résultats par des arguments probabilistes et à l'éventuelle généralisation de ceux-ci sous des contraintes drawdown de forme plus générale. En particulier, il est possible que cette caractérisation permette d'obtenir un analogue de la martingale $(X - \alpha X^*)(X^*)^{\frac{\alpha}{1-\alpha}}$, décisive quant à l'obtention de la condition de transversalité utilisée pour l'argument de vérification.

Liste des travaux ayant contribué à la rédaction de la thèse

• R. Elie, J.D. Fermanian et N. Touzi, *Optimal greek weight by Kernel estimation*, en révision pour Annals of Applied probability;

• B. Bouchard et R. Elie, *Discrete time approximation of decoupled Forward-Backward* SDE with jumps, en révision pour Stochastic Processes and Applications;

• R. Elie, et N. Touzi, *Optimal lifetime consumption and investment under drawdown constraint*, soumis à *Finance and Stochastics*;

• R. Elie, *Optimal consumption and investment in finite horizon under drawdown constraint*, en préparation.

Part I

Optimal Greek weight by Kernel estimation

Abstract

A Greek weight associated to a parameterized random variable $Z(\lambda)$ is a random variable π such that $\nabla_{\lambda} E[\phi(Z(\lambda))] = E[\phi(Z(\lambda))\pi]$ for any function ϕ . The importance of the set of Greek weights for the purpose of Monte Carlo simulations has been highlighted in the recent literature. Our main concern in this chapter is to device methods which produce the optimal weight, which is well-known to be given by the score, in a general context where the density of $Z(\lambda)$ is not explicitly known. To do this, we randomize the parameter λ by introducing an a priori distribution, and we use classical kernel estimation techniques in order to estimate the score function. By an integration by parts argument on the limit of this first kernel estimator, we define an alternative simpler kernel-based estimator which turns out to be closely related to the partial gradient of the kernel-based estimator of $\mathbb{E}[\phi(Z(\lambda))]$. We provide an asymptotic analysis of the mean squared error of these estimators, as well as their asymptotic distributions. For a discontinuous payoff function, the kernel estimators outperforms the classical finite differences one in terms of the asymptotic rate of convergence. This result is confirmed by our numerical experiments. We finally investigate further the short maturity properties of these estimators.

Keywords: Greek weights, Monte Carlo simulation, Non-parametric regression.

Note

The content from Section 1 to Section 5 of this part is based on a paper, written in collaboration with Jean-David Fermanian and Nizar Touzi, in revision for *Annals of Applied Probability*. Since classical estimators of the Greeks suffer from a singularity for short maturity options, an additional careful study of the short time asymptotic properties of the Kernel estimators is reported in Section 6. The heavy asymptotic analysis of the double Kernel based estimator introduced in Section 3.2, is also provided in Section 7.

1 Introduction

Let λ be some given parameter in \mathbb{R}^d , and define the function

$$V^{\phi}(\lambda) := \mathbb{E}\left[\phi\left(Z(\lambda)\right)\right],$$

where Z(.) is a parameterized random variable with values in \mathbb{R}^n and $\phi : \mathbb{R}^n \to \mathbb{R}$ is a measurable function. In many applications, we are interested in the numerical computation of the function $V^{\phi}(\lambda)$ for some parameter λ^0 , together with the sensitivities of V^{ϕ} with respect to the parameter λ .

In particular, in the financial literature, V^{ϕ} represents the no-arbitrage price of a contingent claim, defined by the payoff $\phi(Z(\lambda))$, in the context of a complete market with prices measured in terms of the price of the non-risky asset (so that the model is reduced to the zero-interest rate situation). The sensitivities of V^{ϕ} with respect to the parameter λ are called *Greeks*, and are widely used by the practitioners in their *hedging strategies*. In the context of the Black-Scholes model, the derivative of the option price with respect to the current underlying asset price is the so-called *Delta*, and represents the number of shares of risky asset to be held at each time in order to realize a dynamic perfect hedge of the option. The *Gamma* is the second derivative of the option price, with respect to the underlying asset price. It is an indicator of the variation of the hedging portfolio. Another important *Greek* is the so-called *Vega* (although not a Greek letter !) which is the derivative of the option price with respect to the *volatility* coefficient (see e.g. Hull [63], for more details).

Given a numerical scheme for the computation of the function V^{ϕ} , the first natural idea for the numerical computation of the *Greeks* is the finite differences approximation of the corresponding derivative. In addition to the generic standard error on the numerical computation of the expectation, this approximation leads to a biased estimator at a finite distance and appears to be inefficient for discontinuous payoff functions ϕ . We refer to L'Ecuyer and Perron [42], Detemple, Garcia and Rindisbacher [36] or Milstein and Tretyakov [81] for a theoretical analysis of the rate of convergence of this estimator. Two direct methods for computing the *Greeks* have been presented by Broadie and Glasserman [23] : (i) the pathwise method, which consists in differentiating the random variable $\phi(Z(\lambda))$ inside the expectation operator, and (ii) the likelihood ratio method which reports the differentiation on the distribution of $Z(\lambda)$. The first method requires the computation of the gradient of the payoff function ϕ , which is a serious limitation in practice as ϕ is typically highly complicated or even not differentiable, see also Giles and Glasserman [53] for further developments in this direction. As for the second method (ii), it was (apparently) restricted to the very special cases where the distribution of $Z(\lambda)$ is known explicitly. This difficulty was overcome by Fournié, Lasry, Lebuchoux, Lions and Touzi [50] who exploited the Malliavin integration-by-parts formula to show that, for *smooth* random variables Z(.),

$$\nabla_{\lambda} \mathbb{E}[\phi(Z(\lambda))] = \mathbb{E}[\phi(Z(\lambda))\pi], \qquad (I.1)$$

where π , the so-called *Greek weight*, is a random variable independent of the pay-off function ϕ . A quick overview of the notion of Greek weights is reported in Section 2. Further developments of the results of [50] were obtained by Gobet and Kohatsu-Higa [55]. The comparison of the above different methods is available in the survey paper of Kohatsu-Higa and Montero [69].

An important observation is that the set of Greek weights which satisfy (I.1) is a convex set of random variables. By an easy variance reduction argument, it is easily seen that the score $\pi^* := \nabla_{\lambda} \ln f(\lambda^0, Z(\lambda^0))$ minimizes $\mathbb{V}ar[\phi(Z(\lambda))\pi]$, whenever the density $f(\lambda, z)$ of the random variable $Z(\lambda)$ exists and is sufficiently smooth. In general, the use of the Malliavin calculus does not lead to this optimal Greek weight, except in trivial cases where the density $f(\lambda, z)$ is explicitly known, which corresponds to the case covered by [23].

The main purpose of this chapter is to focus on the use of the optimal Greek weight in order to estimate the corresponding Greek by the Monte Carlo method. To do this, our main idea is to randomize the parameter λ and to re-write V^{ϕ} as a regression function :

$$V^{\phi}(\lambda) := \mathbb{E}\left[\phi(Z(\Lambda))|\Lambda = \lambda\right],$$

where $Z(\Lambda)$ is a random variable with density $\varphi(\lambda, z) := \ell(\lambda^0 - \lambda)f(\lambda, z)$, and $\ell(\lambda^0 - .)$ is some given randomizing distribution on the parameter λ around λ^0 . In other words, the random variable $Z(\Lambda)$ given $\{\Lambda = \lambda\}$ has the same distribution as the random variable $Z(\lambda)$ defined by the density $f(\lambda, z)$. We next assume that our observations consist of a family $\{(\Lambda_i, Z_i), 1 \le i \le N\}$ of independent pairs (Λ_i, Z_i) drawn in the density φ , and we define various kernel estimators of the Greek

$$\nabla_{\lambda} \mathbb{E}[\phi(Z(\lambda))]_{|\lambda=\lambda^0} = \mathbb{E}\left[\phi\left(Z(\lambda^0)\right)s\left(\lambda^0, Z(\lambda^0)\right)\right], \qquad (I.2)$$

where $s(\lambda, z) := \nabla_{\lambda} \ln f(\lambda, z)$ is the score function. The first natural idea is to notice that

$$\mathbb{E}\left[\phi\left(Z(\lambda^{0})\right)s\left(\lambda^{0},Z(\lambda^{0})\right)\right] = \mathbb{E}\left[\phi\left(Z(\Lambda)\right)s\left(\Lambda,Z(\Lambda)\right) \mid \Lambda = \lambda^{0}\right], \quad (I.3)$$

which is a usual regression function. Thus, a two-steps estimation method is proposed : we first perform a kernel-based estimator \hat{s} of the score function, and then we define a kernel regression estimator of the Greek by substituting \hat{s} to s. In the sequel, the resulting estimator is referred to as the double kernel-based estimator and is denoted by $\tilde{\beta}$.

Our next kernel estimator of the Greek is based on a convenient integration-by-parts in (I.2). This leads to a much simpler estimator $\hat{\beta}$ which turns out to be closely related to the estimator $\check{\beta}$, obtained by direct differentiation of the classical kernel regression estimator of $V^{\phi}(\lambda) = \mathbb{E}\left[\phi\left(Z(\Lambda)\right) \mid \Lambda = \lambda^{0}\right]$. These two estimators will be referred to as the single kernel-based estimators.

These three estimators are defined precisely in Section 3, and their asymptotic properties are discussed in Section 4. We show that $\hat{\beta}$ and $\check{\beta}$ are asymptotically equivalent. The asymptotic properties of $\tilde{\beta}$ are derived under stronger conditions on the pay-off function ϕ and the kernel functions. The simultaneous choice of the bandwidth, and the number of observations is also more restrictive in the latter case.

An important observation is that the two single kernel based estimators coincide if and only if the randomizing distribution ℓ is a truncated exponential distribution. In this case, by conveniently relating the support of the truncated exponential distribution to the kernel bandwidth, we observe that the rate of convergence is independent of the dimension of the parameter λ . We next solve the optimal choice of the randomizing distribution within this class by minimizing the corresponding mean square error.

Our asymptotic results imply the following main property of the single kernel based estimators: for a discontinuous payoff function ϕ , the asymptotic rate of convergence of our estimator is better than the classical finite differences one, whenever the order of the kernel function is larger than some explicit threshold. In the case of a truncated exponential randomizing distribution, with support related to the kernel bandwidth, the single kernel based estimator has a better asymptotic rate of convergence whenever the order of the kernel function is larger than four.

Some numerical results are reported in Section 5. We estimate the delta of an European and an Asian digital call option. Our experiments show that the Malliavin-based estimators defined in [50] or [23] are the most efficient, as documented by the previous literature. As predicted by our theoretical asymptotic results, the single-kernel based estimator outperforms the finite differences one, but this is only observed for a large number of simulations. We believe that this does not restrict the interest in our new suggested method as this is just a matter of computer power, and the required number of simulations can be significantly reduced by using variance reduction techniques. For instance, the technique of antithetic variables applied to the randomizing density appears to be very efficient. Finally, Section 6 compares the short time performance of the single-Kernel estimator $\hat{\beta}$ with the Malliavin-based estimator, whose Greek weight is well-known to suffer from a singularity for short maturity problems. We shall derive the asymptotic properties of $\hat{\beta}$ in the situation where the bandwidth of the Kernel and the maturity shrink to zero, and the number of observations goes to infinity. This allows to fix the theoretical relative orders for these three parameters in order to obtain the optimal rate of convergence.

2 The Greek weights set

Throughout this chapter, we consider a classical canonical filtered space of continuous functions equipped with the Wiener measure. The generic point $\omega = \omega(.) \in \Omega$ of this space is a continuous function on \mathbb{R}_+ with $\omega(0) = 0$. We denote by \mathcal{F}_t the σ -algebra generated by the family $\{\omega(s), s \leq t\}$ augmented by all *P*-null sets of Ω . This defines a probability space (Ω, \mathcal{F}, P) carrying a *m*-dimensional standard Brownian motion $\{W_t, t \leq T\}$, with \mathcal{F}_t the smallest filtration that contains the filtration generated by $\{W_s, s \leq t\}$ and satisfying the usual assumptions.

complete probability space (Ω, \mathcal{F}, P) . Let $Z(\lambda)$ be some random variable, valued in \mathbb{R}^n , depending on some finite dimensional parameter $\lambda \in \mathbb{R}^d$, and set

$$V^{\phi}(\lambda) := E[\phi(Z(\lambda))] \text{ for } \phi \in \mathcal{L}^{\infty}(\mathbb{R}^n, \mathbb{R}).$$

In order to simplify the presentation, we shall focus our attention on some fixed particular value λ^0 of λ , and we denote

$$Z^0 := Z(\lambda^0)$$

The chief goal of this chapter is to device efficient methods for the computation of the sensitivity parameter

$$\beta^0 := \nabla_{\lambda} V^{\phi}(\lambda^0),$$

for arbitrary functions ϕ chosen from a suitable large class.

2.1 Definition

We assume that the distribution of $Z(\lambda)$ is absolutely continuous with respect to the Lebesgue measure, and we denote by $f(\lambda, z)$ the associated density, i.e.

$$\mathbb{E}\left[\phi(Z(\lambda))\right] = \int \phi(z) f(\lambda, z) dz \quad \text{for all} \quad \phi \in \mathcal{L}^{\infty}(\mathbb{R}^n, \mathbb{R}) \,.$$

Under mild smoothness assumptions on the density f, we directly compute that

$$\nabla_{\lambda} V^{\phi}(\lambda^{0}) := \frac{\partial V^{\phi}}{\partial \lambda}(\lambda^{0}) = E\left[\phi(Z^{0})S^{0}\right], \qquad S^{0} := s(\lambda^{0}, Z^{0}),$$

where the function s is independent of ϕ and is explicitly given by

$$s(\lambda, z) := \nabla_{\lambda} \{ \ln f(\lambda, z) \}$$

This idea was introduced by Broadie and Glasserman [23] in the context of the Black-Scholes model where the density $f(\lambda, z)$ is explicitly known. We shall always assume that

ve shall always assume that

$$\mathbb{E}\left|S^{0}\right|^{2} < \infty. \tag{I.4}$$

Under this condition, the set

$$\mathcal{W} := \left\{ \pi \in \mathcal{L}^2(\Omega, \mathbb{R}^d) : \nabla_{\lambda} V^{\phi}(\lambda^0) = E\left[\phi(Z^0)\pi\right] \text{ for all } \phi \in \mathcal{L}^{\infty}(\mathbb{R}^n, \mathbb{R}) \right\}$$

is not empty. From the arbitrariness of $\phi \in \mathcal{L}^{\infty}(\mathbb{R}^n, \mathbb{R})$, it is immediately seen that

$$\mathcal{W} = \left\{ \pi \in \mathcal{L}^2(\Omega, \mathbb{R}^d) : E[\pi | Z^0] = S^0 \right\} \,,$$

and therefore

$$\begin{aligned} \operatorname{\mathbb{V}ar}\left[\phi(Z^{0})\pi\right] &= E\left[\phi(Z^{0})^{2}E[\pi\pi'|Z^{0}]\right] - \nabla V^{\phi}(\lambda^{0})\nabla V^{\phi}(\lambda^{0})'\\ &\geq E\left[\phi(Z^{0})^{2}E[\pi|Z^{0}]E[\pi|Z^{0}]'\right] - \nabla V^{\phi}(\lambda^{0})\nabla V^{\phi}(\lambda^{0})'\\ &= E\left[\phi(Z^{0})^{2}S^{0}S^{0'}\right] - \nabla V^{\phi}(\lambda^{0})\nabla V^{\phi}(\lambda^{0})' = \operatorname{\mathbb{V}ar}\left[\phi(Z^{0})S^{0}\right],\end{aligned}$$

where ' denotes the transposition operator. Hence

$$S^0 \in \mathcal{W}$$
 is a minimizer of $\operatorname{Var}\left[\phi(Z^0)\pi\right]$, $\pi \in \mathcal{W}$.

Throughout this chapter, we call S^0 the optimal Greek weight. As reported briefly in subsection 2.2, when the density function $f(\lambda, z)$ is not known, it was suggested in [50] to obtain (inefficient) Greek weights from the set \mathcal{W} by exploiting the integration by-parts-formula from Malliavin calculus. Our main objective here is to derive Monte Carlo estimators of the Greek value β^0 , which asymptotically achieve the minimum variance, by using methods from non-parametric statistics to approximate the above optimal Greek weight S^0 .

2.2 Malliavin Greek weights

We first recall the definition of the Malliavin gradient operator. Let S be the set

$$\left\{F = f\left(\int_{\mathbb{R}_+} h_t^1 \cdot dW_t, \dots, \int_{\mathbb{R}_+} h_t^n \cdot dW_t\right), n \in \mathbb{N}, f \in \mathbf{C}_p^{\infty}(\mathbb{R}^n), h^i \in L^2(\mathbb{R}_+, \mathbb{R}^m)\right\},\$$

where $\mathbf{C}_p^{\infty}(\mathbb{R}^n)$ is the set of all infinitely continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ such that f and all of its partial derivatives have polynomial growth. The Malliavin derivative of any random variable F in S is defined by :

$$D_t F := \sum_{i=1}^n \nabla_{x_i} f\left(\int_{\mathbb{R}_+} h_t^1 \cdot dW_t, \cdots, \int_{\mathbb{R}_+} h_t^n \cdot dW_t\right) h_t^i$$

This operator is then extended to $\mathcal{L}^2(\Omega, \mathbb{R}^d)$, by taking the closure \mathcal{S} with respect to the semi norm $||F|| := \left(\mathbb{E}|F|^2 + \mathbb{E}\int_{\mathbb{R}_+} |D_t F|^2 dt\right)^{1/2}$ (see e.g. Nualart [82]). This produces the domain $\mathbb{ID}^{1,2}$ of the Malliavin operator D, as a dense subset of $\mathcal{L}^2(\Omega, \mathbb{R}^d)$. The Malliavin derivative of functions valued in \mathbb{R}^d is defined componentwise.

We next show how the operator D allows to derive Greek weights in \mathcal{W} , without appealing to the explicit knowledge of the density $f(\lambda, z)$. Observe that, for every $\pi \in \mathcal{W}$, we have

$$\mathbb{E}[\pi] = \mathbb{E}[S^0] = \frac{\partial}{\partial \lambda} \int f(\lambda, t) dt = 0.$$

If, in addition, $\pi \in \mathcal{L}^2(\Omega, \mathbb{R}^d)$, then it follows from the representation theorem that

$$\pi = \int_0^\infty u_s \ dW_s$$

for some $u \in \mathcal{L}^2_a(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d, m))$ with $\mathbb{E}\left[\int_0^\infty |u_s|^2 ds\right] < \infty$. Here, $\mathbb{M}_{\mathbb{R}}(d, m)$ is the collection of all real matrices with d rows and m columns, and $\mathcal{L}^2_a(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d, m))$ is the set of all adapted processes with values in $\mathbb{M}_{\mathbb{R}}(d, m)$. Assume that

$$Z^0 \in \mathbb{ID}^{1,2}, \tag{I.5}$$

and let ϕ be a $\mathbf{C}_b^1(\mathbb{R}^n, \mathbb{R})$ function and $\pi \in \mathcal{W}$. Then, it follows from the Malliavin integration by parts formula that

$$\nabla V^{\phi}(\lambda^{0}) = \mathbb{E}\left[\phi(Z^{0})\int_{0}^{\infty} u_{s} \, dW_{s}\right] = \mathbb{E}\left[\int_{0}^{\infty} u_{s} D_{s} \phi(Z^{0}) \, ds\right]$$
$$= \mathbb{E}\left[\int_{0}^{\infty} u_{s} (D_{s} Z^{0})' ds \, \nabla \phi(Z^{0})\right], \quad (I.6)$$

where $(D_s Z^0)_{ij} = (D_s Z^0_i)_j$ and Z^0_i is the *i*-th entry of Z^0 , i = 1, ..., n, j = 1, ..., m. On the other hand,

$$\nabla V^{\phi}(\lambda^{0}) = \mathbb{E}\left[\nabla Z^{0} \nabla \phi(Z^{0})\right] \quad \text{where} \quad \nabla Z^{0} := \frac{\partial Z'}{\partial \lambda}(\lambda^{0}). \tag{I.7}$$

By arbitrariness of $\phi \in \mathbf{C}_b^1(\mathbb{R}^n, \mathbb{R})$, we deduce from (I.6) and (I.7) that

$$\mathbb{E}\left[\int_0^\infty u_s(D_s Z^0)' ds \,\middle|\, Z^0\right] = \mathbb{E}\left[\nabla Z^0 \,\middle|\, Z^0\right]. \tag{I.8}$$

Conversely, let u be a process in $\mathcal{L}^2(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d, m))$ and integrable in the Skorohod sense, i.e. in $Dom(\delta)$, satisfying (I.8). Observe that u does not need to be adapted. Then $\pi := \int_0^\infty u_s dW_s$ satisfies $\nabla V^{\phi}(\lambda^0) = \mathbb{E}[\phi(Z^0)\pi]$ for every $\phi \in \mathbf{C}_b^1(\mathbb{R}^n, \mathbb{R})$. By a density argument, this property is easily seen to hold for every $\phi \in L^\infty(\mathbb{R}^n, \mathbb{R})$. Hence $\pi \in \mathcal{W} \cap \mathcal{L}^2(\Omega, \mathbb{R}^d)$. We have then proved the following result :

Proposition 2.1 Assume that $Z^0 \in \mathbb{D}^{1,2}$. Then

$$\mathcal{W} = \left\{ \int_0^\infty u_s \, dW_s : u \in \mathcal{L}^2 \left(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d, m) \right) \text{ and (I.8) holds} \right\}$$

This result allows to obtain a family of Greek weights without any knowledge of the density distribution of the random variable Z^0 . However there is no guarantee for the weight defined by some process $u \in \mathcal{L}^2(\mathbb{R}_+ \times \Omega, \mathbb{M}_{\mathbb{R}}(d, m))$ satisfying (I.8) to produce the optimal Greek weight: see the last two examples of the subsequent Subsection 2.3. The chief goal of this chapter is to introduce kernel-based estimators which focus on the optimal weight S^0 . Of course, our estimators do not have the parametric rate of convergence, but we believe that this critic does not exclude our estimators in finite samples. The main advantage of our estimators remains their simplicity of computation in comparison to Malliavin-based estimators.

Note that the Malliavin Greek weights also lead to estimators of the Greeks which do not have the parametric rate of convergence. Indeed, except the trivial gaussian case, the Malliavin weight is a stochastic integral which needs to be approximated on some given time grid. This leads to a loss of the parametric rate.

2.3 Examples of Malliavin Greek weights

We now provide some examples in the context of the Black-Scholes model. In the first two examples, we derive the optimal Greek weight by the Malliavin integration by parts technique. The last examples show the limitation of this technique as the optimal Greek weight can not be derived. The reader interested in our statistical results can move straight away to the next section.

Let T > 0 be some given finite maturity, and define

$$S_T^{s,\mu,\sigma} := s \exp\left[\left(\mu - (\sigma^2/2)\right)T + \sigma W_T\right].$$

In this simple example, the Malliavin derivative process is given by

$$D_r S_T = \sigma S_T \mathbf{1}_{r \leq T}$$
 for all $r \geq 0$.

Example 2.1 (Delta of a European option, Black-Scholes model)

With $Z^0(s) := S_T^{s,\mu,\sigma}$, we directly compute that $\int_0^\infty D_r Z^0 u_r dr = \sigma S_T \int_0^T u_r dr$ for every $u \in L^2(\mathbb{R}_+ \times \Omega, \mathbb{R})$. Clearly the constant process $u_r^0 := (\sigma sT)^{-1} \mathbf{1}_{r \leq T}$ satisfies Condition (I.8), and the associated Greek weight is

$$\pi^0 = \int_0^T u_r^0 dW_r = (\sigma sT)^{-1} W_T + (\sigma$$

Since π^0 is a deterministic function of S_T , we see that π^0 is the optimal Greek weight.

Example 2.2 (Vega of a European option, Black-Scholes model)

We now consider the case $Z^0(\sigma) := S_T^{s,\mu,\sigma}$. It is easily checked that the constant process $u_r^0 := [(\sigma T)^{-1} W_T - 1] \mathbf{1}_{r \leq T}$ satisfies Condition (I.8), and the associated Greek weight is

$$\pi^{0} = \int_{0}^{T} u_{r}^{0} dW_{r} = (\sigma T)^{-1} \left[-\sigma T W_{T} + W_{T}^{2} - T \right] \,.$$

Since π^0 is a deterministic function of S_T , we see that π^0 is the optimal Greek weight.

Example 2.3 (Delta of an Asian option, Black-Scholes model)

We now set $Z^0(s) := \int_0^T S_t^{s,\mu,\sigma} dt$. We directly compute that $D_r Z^0 = \sigma \int_r^T S_t dt \mathbf{1}_{r \leq T}$ for all $r \geq 0$, so that Condition (I.8) reduces to

$$\sigma s \mathbb{E}\left[\int_0^T \int_r^T S_t u_r dt \, dr \, \middle| \, Z^0\right] = \int_0^T S_t dt$$

Direct computation shows that the process $u_r^0 := 2 \left(\sigma s \int_0^T S_t dt\right)^{-1} S_r$ satisfies Condition (I.8), and the associated Greek weight is

$$\pi^{0} = \int_{0}^{T} u_{r}^{0} dW_{r} = \frac{2}{\sigma^{2} s} \left[-\mu + \frac{\sigma^{2}}{2} + \frac{S_{T} - s}{\int_{0}^{T} S_{t} dt} \right].$$

Observe that π^0 is not $\sigma(Z^0)$ -measurable. Hence π^0 is not the optimal Greek weight.

Example 2.4 (Delta of an Euro-Asian option, Black-Scholes model) We now set $Z^0(s) := \left(S_T^{s,\mu,\sigma}, \int_0^T S_t^{s,\mu,\sigma} dt\right)$. We directly compute that, for all $r \ge 0$, $D_r Z^0 = \sigma \left(S_T, \int_r^T S_t dt\right) \mathbf{1}_{r \le T}$, so that Condition (I.8) reduces to

$$\sigma s \mathbb{E}\left[\int_0^T u_r dr \bigg| Z^0\right] = 1 \text{ and } \sigma s \mathbb{E}\left[\int_0^T \int_r^T S_t u_r dt dr \bigg| Z^0\right] = \int_0^T S_t dt.$$

By direct computation, we see that this condition is satisfied by the process

$$u_r^0 := \frac{2}{\sigma s} \left[\frac{-S_r}{\int_0^T S_t dt} + 3 \frac{S_r \int_r^T S_t dt}{\left(\int_0^T S_t dt\right)^2} \right]$$

and the associated Greek weight is given by

$$\pi^{0} = \int_{0}^{T} u_{r}^{0} dW_{r} = \frac{1}{\sigma^{2}s} \left[-\mu + 3\sigma^{2} - \frac{2S_{T} + 4s}{\int_{0}^{T} S_{t} dt} + 6 \frac{\int_{0}^{T} S_{t}^{2} dt}{\left(\int_{0}^{T} S_{t} dt\right)^{2}} \right]$$

Observe that π^0 is not $\sigma(Z^0)$ -measurable. Hence it is not the optimal Greek weight.

3 Kernel estimation and optimal Greek weight

3.1 Randomization of the parameter

The main idea of this chapter is to randomize the parameter λ in order to estimate the Greek by the classical kernel estimation technique. This randomization can be exploited from two viewpoints. First, one can use it in order to estimate the optimal Greek weight, i.e. the score function. An alternative viewpoint is to take advantage of the smoothness of the randomizing distribution in order to obtain an integration by parts formula similar to the Malliavin integration by parts technique. This technique is well known in the non-parametric statistics litterature, see eg [4].

Let $\ell : \mathbb{R}^d \longrightarrow \mathbb{R}$ be some given probability density function, with support containing the origin in its interior, and set

$$\varphi(\lambda, z) := \ell(\lambda^0 - \lambda) f(\lambda, z) \text{ for } \lambda \in \mathbb{R}^d \text{ and } z \in \mathbb{R}^n,$$

where λ^0 is the parameter of interest. We consider a sequence

$$(\Lambda_i, Z_i)_{1 \le i \le N}$$
 of N independent r.v. with distribution $\varphi(\lambda, z)$, (I.9)

so that, for any $i \leq N$, $\ell(\lambda^0 - .)$ is the density of Λ^i and $f(\Lambda^i, .)$ is the conditional density of Z^i given Λ^i .

Remark 3.1 Notice that the simulation of $(\Lambda_i, Z_i)_{i\geq 1}$ can be performed easily even in cases where the density φ can not be written explicitly. This applies typically to the case where $Z(\lambda) = X_T(\lambda)$, for some integer T, where $\{X_t(\lambda), t \in \mathbb{N}\}$ is a Markov chain with given transition density. Then, for a given value of λ , the simulation of Z is easily feasible by usual methods. However the marginal distribution of $Z(\lambda)$ is typically very complicated so that it is useless for the numerical computation of the score function $s(\lambda, z)$.

In this section, we provide various estimation methods of β^0 based on non-parametric kernel methods. We then introduce the kernel function

$$K : \mathbb{R}^d \longrightarrow \mathbb{R} \text{ with } \int K = 1,$$

whose precise properties will be detailed at the beginning of section 4.

3.2 A first kernel estimator of the Greek

The main idea is that the optimal weight S^0 requires a priori the knowledge of the probability density function $f(\lambda, z)$ and the associated *score* function $s(\lambda, z)$. Indeed, if these functions were explicitly known, then a natural non-parametric estimator of the Greek β^0 using the observations (I.9) is

$$\bar{\beta}_N := \frac{1}{\ell(0) N h^d} \sum_{i=1}^N \phi(Z_i) \ s(\Lambda_i, Z_i) \ K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \ . \tag{I.10}$$

Although s is not explicitly known in our applications of interest, one could approximate it by means of an additional kernel estimator based on another kernel function H defined on \mathbb{R}^n . We introduce our first kernel-based estimator of β^0

$$\tilde{\beta}_N := \frac{1}{\ell(0) N h^d} \sum_{i=1}^N \phi(Z_i) \, \hat{s}_N^{-i}(\Lambda_i, Z_i) \, K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \,, \tag{I.11}$$

where s_N^{-i} is an approximation of s given by

$$\hat{s}_N^{-i}(\lambda, z) := \frac{\hat{\varphi}_\lambda^{-i}}{\hat{\varphi}^{-i} + (\delta/3 - \varphi^{-i})\mathbf{1}_{|\varphi^{-i}| < \delta/3}}(\lambda, z) + \frac{\nabla \ell(\lambda^0 - \lambda)}{\ell(\lambda^0 - \lambda)}, \quad (I.12)$$

with δ some small fixed parameter, and

$$\hat{\varphi}^{-i}(\lambda, z) := \frac{h^{-d-n}}{N-1} \sum_{j=1, j \neq i}^{N} K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right), \qquad (I.13)$$

$$\hat{\varphi_{\lambda}}^{-i}(\lambda, z) := \nabla_{\lambda} \hat{\varphi}^{-i}(\lambda, z) = \frac{h^{-d-n-1}}{N-1} \sum_{j=1, j \neq i}^{N} \nabla K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right).$$
(I.14)

Remark 3.2 Observe that the denominator $\hat{\varphi}^{-i} + (\delta/3 - \varphi^{-i})\mathbf{1}_{|\varphi^{-i}| < \delta/3}$ in (I.12) is simply a truncation which avoids the small values of $\hat{\varphi}^{-i}$. This technical trick allows to avoid the explosion of the estimator and the error due to this truncation is controlled by imposing some constraints on the small values of φ , detailed in Assumption S below. In fact, $\hat{s}_N^{-i}(\lambda, z)$ behaves like

$$\frac{\hat{\varphi_{\lambda}}^{-i}}{\hat{\varphi}^{-i}}(\lambda, z) + \frac{\nabla \ell(\lambda^0 - \lambda)}{\ell(\lambda^0 - \lambda)},$$

$$= \frac{\partial}{\partial \lambda} \ln \left\{ \frac{1}{\ell(\lambda^0 - \lambda) (N - 1)h^{d+n}} \sum_{j=1, j \neq i}^N K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right) \right\}.$$

From a practical point of view, this estimator displays two drawbacks. First, its expression involves a product of two (possibly multidimensional) kernels K and H. Thus, it suffers from the so-called "curse of dimensionality". Moreover, its calculation is timeconsuming. In the subsequent subsections, we introduce two alternative kernel estimators of β^0 , which involve a single kernel function and a single summation.

From a theoretical point of view, we shall see that this estimator achieves the same rate of convergence as the two following ones but requires more stringent conditions, and involves heavy calculations.

3.3 A simpler kernel estimator of the Greek

For convenience, we continue our discussion under the condition that

the kernel function
$$K$$
 has compact support. (I.15)

The latter condition is essentially technical. It could be removed, but at the price of additional regularity assumptions, that would be related to the tails of the underlying distributions and K. Moreover, without (I.15), the relations between our estimators would be more involved and less nice. We still consider the natural estimator given by (I.10). For fixed h > 0, it follows from the law of large numbers that

$$\bar{\beta}_N \xrightarrow[N \to \infty]{} \frac{1}{\ell(0)h^d} \mathbb{E}\left[\phi(Z)s(\Lambda, Z) \ K\left(\frac{\lambda^0 - \Lambda}{h}\right)\right], \ P - \text{a.s.}$$
(I.16)

where (Λ, Z) is a random variable with distribution $\varphi(\lambda, z)$. Recalling the definition of s, and integrating by parts with respect to the variables $\lambda_1, \ldots, \lambda_d$, we see that for h > 0

sufficiently small, we have

$$\mathbb{E}\left[\bar{\beta}_{N}\right] = \frac{1}{\ell(0)h^{d}} \int \phi(z) K\left(\frac{\lambda^{0}-\lambda}{h}\right) \ell(\lambda^{0}-\lambda) \nabla_{\lambda} f(\lambda,z) \, d\lambda dz \\ = \frac{h^{-d-1}}{\ell(0)} \int \phi(z) \left(\nabla K\left(\frac{\lambda^{0}-\lambda}{h}\right) + hK\left(\frac{\lambda^{0}-\lambda}{h}\right) \frac{\nabla \ell}{\ell}(\lambda^{0}-\lambda)\right) \varphi(\lambda,z) \, d\lambda \, dz \\ = \frac{1}{\ell(0)h^{d+1}} \mathbb{E}\left[\phi(Z) \left(\nabla K\left(\frac{\lambda^{0}-\Lambda}{h}\right) + hK\left(\frac{\lambda^{0}-\Lambda}{h}\right) \frac{\nabla \ell}{\ell}(\lambda^{0}-\Lambda)\right)\right],$$

where we used (I.15). This suggests the following simpler kernel estimator β^0 :

$$\hat{\beta}_N := \frac{1}{\ell(0)Nh^{d+1}} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + hK \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \frac{\nabla \ell}{\ell} (\lambda^0 - \Lambda_i) \right).$$
(I.17)

The asymptotic properties of $\hat{\beta}_N$ will be provided in Section 4.

3.4 Differentiating the kernel estimator of the price

We next start out from the natural kernel estimator of the price $V^{\phi}(\lambda)$:

$$\hat{V}_N^{\phi}(\lambda) := \frac{1}{Nh^d \,\ell(\lambda^0 - \lambda)} \,\sum_{i=1}^N \phi(Z_i) K\left(\frac{\lambda - \Lambda_i}{h}\right) \,.$$

Differentiating $\hat{V}_N^{\phi}(\lambda)$ with respect to λ at the point λ^0 , we obtain our final kernel estimator of the Greek:

$$\check{\beta}_N := \frac{1}{\ell(0)Nh^{d+1}} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + hK \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \frac{\nabla \ell}{\ell}(0) \right).$$
(I.18)

Observe that our two estimators $\hat{\beta}_N$ and $\check{\beta}_N$ are closely related by :

$$\check{\beta}_N = \hat{\beta}_N + \frac{1}{\ell(0)Nh^d} \sum_{i=1}^N \phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \left(\frac{\nabla \ell}{\ell}(0) - \frac{\nabla \ell}{\ell}(\lambda^0 - \Lambda_i)\right) \,.$$

In particular,

 $\check{\beta}_N = \hat{\beta}_N$ whenever $\ell : l \mapsto e^{a_0 + a_1 \cdot \ell} \mathbf{1}_B(\ell)$ is a truncated exponential distribution, (I.19)

for some parameters $a_0 \in \mathbb{R}$, $a_1 \in \mathbb{R}^d$ and some subset B of \mathbb{R}^d containing the origin in its interior.

The asymptotic properties of this third estimator will also be provided in Section 4.

4 Asymptotic results

We now compare the estimators defined in the previous section from the viewpoint of their asymptotic distributions. The main result of this section is that there is no advantage from using the cumbersome *double Kernel-based* estimator. From a theoretical point of view, it is proved to achieve the same asymptotic rate of convergence as the single Kernel ones but under more stringent condition and, from a practical point of view, the use of this estimator is much more time consuming.

We shall first show that the two single kernel-based estimators have equal asymptotic rates of convergence. We then derive the same rate of convergence for the double Kernel based estimator but under stronger conditions so that we next focus on the study of the single Kernel based ones. We deduce the optimal choice of the number of simulations N and the bandwidth h of the kernel function K, by using the classical mean square error minimization criterion. We next specialize the discussion to the case of a uniform or truncated exponential randomizing distribution (I.19) with support defined by $B := [-\varepsilon, \varepsilon]^d$. In this setting, we observe that the rate of convergence of the kernel estimator is independent of the dimension of the parameter λ for some convenient choice of ε in terms of the bandwidth h. We then discuss the optimal choice of the randomizing density ℓ within the class of truncated exponential distribution, and we provide a quasi-explicit characterization of the optimal truncated exponential randomizing distribution in the sense of the mean square error criterion. Finally, we compare the rate of convergence of our estimators to the finite differences one.

Before stating our results, we recall that the order of the kernel function K is defined as the smallest non zero integer p such that there exist some integers (j_1, \ldots, j_p) , with $j_k \in \{1, \ldots, d\}$, such that

$$\int l_{\alpha_1} \dots l_{\alpha_r} K(l) dl = 0 \quad \text{for } 0 < r < p, \, \alpha_k \in \{1, \dots, d\}, \text{ and}$$
$$\int l_{j_1} \dots l_{j_p} K(l) dl \neq 0.$$

Typically, if K is the product of d even univariate kernels, then it is of order p = 2 (at least). The regularity hypothesis on the kernel function K will be the following.

Assumption K The kernel function $K : \mathbb{R}^d \to \mathbb{R}$ is C^1 , compactly supported, satisfies $\int K = 1$, and is of order $p \ge 2$.

In the subsequent subsections, we shall use the notation

$$\xi_K^p[\psi](\lambda, z) := \frac{(-1)^p}{p!} \sum_{j_1, \dots, j_p=1}^d \left(\int l_{j_1} \dots l_{j_p} K(l) dl \right) \nabla_{\lambda_{j_1} \dots \lambda_{j_p}}^p \psi(\lambda, z) , \quad (I.20)$$

for every smooth function ψ defined on $\mathbb{R}^d \times \mathbb{R}^n$. We shall also denote $A^{\otimes} := AA'$ for every matrix A, and C denotes a constant whose value may change from line to line.

4.1 Asymptotic results for the single kernel-based estimators

Our first result requires some regularity conditions on the density functions f and ℓ .

Assumption R1 For every z, the functions $f(\cdot, z)$ and ℓ are p+1 times differentiable, and for every integer $i \leq p$, the function $\lambda \mapsto \nabla_{\lambda}^{i} \{\ell(\lambda^{0} - \lambda)\nabla_{\lambda}f(\lambda, z)\}$ is continuous at λ^{0} uniformly with respect to $z \in S$, for some subset S s.t. Supp $(\phi) \subset int(S)$.

Proposition 4.1 Under Assumptions K and R1, as $N \to \infty$ and $h \to 0$, the bias and the variance of $\hat{\beta}_N$ satisfy

$$\mathbb{E}\left[\hat{\beta}_{N}\right] - \beta^{0} \sim C_{1}h^{p} \quad and \quad \mathbb{Var}\left[\hat{\beta}_{N}\right] \sim \frac{\Sigma}{Nh^{d+2}}, \qquad (I.21)$$

where

$$C_1 := \frac{1}{\ell(0)} \int \xi_K^p \left[\ell(\lambda^0 - .) f_\lambda \right] (\lambda^0, z) \phi(z) dz \quad and \quad \Sigma := \frac{\mathbb{E}[\phi^2(Z^0)]}{\ell(0)} \int \nabla K^{\otimes}.$$
 (I.22)

Proof. By definition of $\hat{\beta}_N$, we have $\mathbb{E}\left[\hat{\beta}_N\right] = \mathbb{E}\left[\bar{\beta}_N\right]$. By (I.16), this provides

$$\psi(h) := \mathbb{E}\left[\hat{\beta}_{N}\right] = \frac{1}{\ell(0)h^{d}} \int \phi(z)\ell(\lambda^{0}-\lambda)\nabla_{\lambda}f(\lambda,z)K\left(\frac{\lambda^{0}-\lambda}{h}\right)d\lambda dz$$
$$= \frac{1}{\ell(0)} \int \phi(z)\ell(hl) f_{\lambda}(\lambda^{0}-hl,z)K(l)dl dz.$$

Clearly, $\psi(0) = \int \phi(z) f_{\lambda}(\lambda^0, z) dz = \beta^0$. Moreover, since K has compact support, it follows from Assumption R1 that the function ψ is p times differentiable at zero, with derivatives obtained by differentiating inside the integral sign, so that its *i*-th iterated derivative denoted $\psi^{(i)}(0)$ are given by

$$\frac{(-1)^{i}}{\ell(0)} \sum_{j_{1},\dots,j_{i}=1}^{d} \left(\int l_{j_{1}}\dots l_{j_{i}}K(l) dl \right) \left(\int \phi(z) \left[\nabla^{i}_{\lambda_{j_{1}},\dots,\lambda_{j_{i}}} \{\ell(\lambda^{0}-.) f_{\lambda}\} \right] (\lambda^{0},z) dz \right)$$

for every $1 \leq i \leq p$. Since p is the order of K, observe that $\psi^{(i)}(0) = 0$ for every $1 \leq i < p$, so that a Taylor expansion of ψ provides the first part of the Proposition. As for the variance, we directly compute that

$$\mathbb{V}\mathrm{ar}\left[\hat{\beta}_{N}\right] = \frac{(v_{1}-v_{2}^{\otimes})}{Nh^{2d+2}\ell(0)^{2}} ,$$

where

$$v_1 := \mathbb{E}\left[\phi(Z)^2 \left(\nabla K\left(\frac{\lambda^0 - \Lambda}{h}\right) + hK\left(\frac{\lambda^0 - \Lambda}{h}\right)\frac{\nabla \ell}{\ell}(\lambda^0 - \Lambda)\right)^{\otimes}\right],$$

$$v_2 := \mathbb{E}\left[\phi(Z) \left(\nabla K\left(\frac{\lambda^0 - \Lambda}{h}\right) + hK\left(\frac{\lambda^0 - \Lambda}{h}\right)\frac{\nabla \ell}{\ell}(\lambda^0 - \Lambda)\right)\right].$$

By a similar argument as in the first part of this proof, we compute that

$$v_1 = h^d \int \phi^2(z) \left(\nabla K(l) + hK(l) \frac{\nabla \ell}{\ell}(hl) \right)^{\otimes} \ell(hl) f(\lambda^0 - hl, z) dl dz$$

 $\sim h^d \ell(0) \left(\int \nabla K(l)^{\otimes} dl \right) \mathbb{E} \left[\phi^2(Z^0) \right] .$

The required result follows by observing that $v_2 = O(h^{d+1})$.

We are now ready for our first main result.

Theorem 4.1 (i) Let the conditions of Proposition 4.1 hold, and assume that

$$h \longrightarrow 0 \quad and \quad N h^{d+2} \longrightarrow \infty \quad as \quad N \to \infty.$$
 (I.23)

Then, with Σ as in (I.22), we have $\sqrt{Nh^{d+2}} \left(\hat{\beta}_N - \mathbb{E}[\hat{\beta}_N]\right) \longrightarrow \mathcal{N}(0, \Sigma)$ in distribution. (ii) In addition to the above conditions, assume that

$$N h^{d+2+2p} \longrightarrow 0 \quad as \quad N \to \infty.$$
 (I.24)

Then the bias vanishes and $\sqrt{Nh^{d+2}}\left(\hat{\beta}_N-\beta^0\right)\longrightarrow \mathcal{N}\left(0,\Sigma\right)$ in distribution.

Proof. We shall prove this result by verifying the Lyapounov conditions (see e.g. Billingsley [14], p. 44). Let a be an arbitrary vector in \mathbb{R}^d , and define, for every $i = 1, \ldots, N$,

$$\begin{split} Y_i^N &:= \frac{1}{Nh^{d+1}\ell(0)} \phi(Z_i) \left(\nabla K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) + hK\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \frac{\nabla \ell}{\ell} (\lambda^0 - \Lambda_i) \right) \\ X_i^N &:= a' \left(Y_i^N - \mathbb{E}[Y_i^N] \right) \,. \end{split}$$

In view of Proposition 4.1, the only condition which remains to check in order to verify the Lyapounov conditions is the existence of $\delta > 2$ such that

$$\sup_{N} \frac{1}{\sigma_{N}^{\delta}} \sum_{i=1}^{N} \mathbb{E}[|X_{i}^{N}|^{\delta}] < +\infty \quad \text{where} \quad \sigma_{N}^{2} := \mathbb{V}\operatorname{ar}\left[\sum_{i=1}^{N} X_{i}^{N}\right].$$
(I.25)

In order to prove (I.25), we start by observing from (I.21) that

$$\sigma_N^2 \sim \frac{\Sigma_a}{Nh^{d+2}}$$
 with $\Sigma_a := \frac{1}{\ell(0)} \mathbb{E}[\phi^2(Z^0)] \int |a' \nabla K(l)|^2 dl$.

We next estimate by the Minkowski inequality and (I.21) that

$$\begin{aligned} \|X_i^N\|_{\delta} &\leq \|a'Y_i^N\|_{\delta} + |a'\mathbb{E}[Y_i^N]| \\ &= \|a'Y_i^N\|_{\delta} + \frac{1}{N} \left|a'\mathbb{E}[\hat{\beta}_N]\right| \\ &\leq \frac{\sum_{i=1}^d \left\|\phi(Z)a_i\left(\nabla_i K\left(\frac{\lambda^0 - \Lambda}{h}\right) + hK\left(\frac{\lambda^0 - \Lambda}{h}\right)\frac{\nabla_i \ell}{\ell}(\lambda^0 - \Lambda)\right)\right\|_{\delta}}{Nh^{d+1}\ell(0)} + O\left(\frac{1}{N}\right) \end{aligned}$$

By a Taylor expansion with respect to the h variable, in the neighborhood of the origin, following the method used in the proof of Proposition 4.1, we deduce

$$\left\|X_i^N\right\|_{\delta} \leq C \left(\frac{h^{d/\delta}}{Nh^{d+1}} + \frac{1}{N}\right).$$

Hence, we have

$$\frac{1}{\sigma_N^{\delta}} \sum_{i=1}^N \mathbb{E}\left[|X_i^N|^{\delta} \right] \leq C N \frac{h^d}{(Nh^{d+1})^{\delta}} (Nh^{d+2})^{\delta/2} \leq \frac{C}{(Nh^d)^{(\delta-2)/2}},$$

and condition (I.25) is satisfied when $Nh^d \to \infty$, as assumed in (I.23). Therefore, $\sqrt{Nh^{d+2}} \sum_{i=1}^{N} X_i^N$ is asymptotically gaussian, with a variance matrix given by Σ_a . By the arbitrariness of $a \in \mathbb{R}^d$, the required result follows from the Cramer-Wold device. \Box

We next turn to the estimator $\check{\beta}$ which was defined as the gradient, with respect to λ , of the kernel based estimator $\hat{V}_N^{\phi}(\lambda)$ of the function $V_N^{\phi}(\lambda)$. The asymptotic properties of this estimator are obtained by following the techniques of the previous proofs and require the following regularity condition on the densities f and ℓ .

Assumption R2 For every z, the functions $f(\cdot, z)$ and ℓ are p+1 times differentiable, and for every integer $i \leq p$, the function $\lambda \mapsto \nabla_{\lambda}^{i} \{\ell(\lambda^{0} - \lambda)f(\lambda, z)\}$ is continuous at λ^{0} uniformly with respects to $z \in S$, for some subset S s.t. Supp $(\phi) \subset int(S)$. **Proposition 4.2** Under Assumptions K and R2, as $N \to \infty$ and $h \to 0$, the bias and the variance of $\check{\beta}_N$ satisfy

$$\mathbb{E}[\check{\beta}_N] - \beta^0 \sim C_2 h^p \quad and \quad \mathbb{V}\mathrm{ar}[\check{\beta}_N] \sim \frac{\Sigma}{Nh^{d+2}},$$

where Σ is given by (I.22), and

$$C_2 := \frac{1}{\ell(0)} \int \left(\xi_K^p[\varphi_\lambda] + \frac{\nabla \ell}{\ell}(0)\,\xi_K^p[\varphi]\right) \left(\lambda^0, z\right) \phi(z)\,dz$$

Proof. The proof is essentially similar to the one of Proposition 4.1. Recall that the estimators $\check{\beta}_N$ and $\hat{\beta}_N$ are related by :

$$\check{\beta}_N = \hat{\beta}_N + \frac{1}{\ell(0)Nh^d} \sum_{i=1}^N \phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \left(\frac{\nabla\ell}{\ell}(0) - \frac{\nabla\ell}{\ell}(\lambda^0 - \Lambda_i)\right).$$
(I.26)

We start by analyzing the bias term. Recall from the proof of Proposition 4.1 that :

$$\mathbb{E}\left[\hat{\beta}_{N}\right] = \frac{1}{\ell(0)} \int \phi(z)\ell(hl)f_{\lambda}(\lambda^{0} - hl, z)K(l) \, dl \, dz$$

We then deduce from (I.26) that :

$$\mathbb{E}\left[\check{\beta}_{N}\right] = \frac{1}{\ell(0)} \int \phi(z) \left(\varphi_{\lambda}(\lambda^{0} - hl, z) + \frac{\nabla \ell}{\ell}(0)\varphi(\lambda^{0} - hl, z)\right) K(l) \, dl \, dz \, .$$

We now observe that Assumption R2 allows to derive an expansion in the h variable of the above expression, near the origin, up to the order p. The coefficients of the expansion are obtained by simple differentiation inside the integral sign. Finally, since p is the order of the kernel K, it is easily seen that the coefficients of h^i , in this expansion vanish for i < p, and the only non-zero coefficient is that of h^p .

The variance of β_N is also treated by the same argument as in the proof of Proposition 4.1, and the dominant term in the expansion of the variance is easily seen to be the same as in that proof.

Proposition 4.2 says that $\hat{\beta}_N$ and $\check{\beta}_N$ have the same asymptotic variance, and the orders of their asymptotic biases are the same. Our next result states that these two estimators have exactly the same asymptotic distribution.

Theorem 4.2 (i) Let the conditions of Proposition 4.2 hold, and assume further that (I.23) holds. Then, with Σ as in (I.22), we have $\sqrt{Nh^{d+2}} \left(\check{\beta}_N - \mathbb{E}[\check{\beta}_N]\right) \longrightarrow \mathcal{N}(0, \Sigma)$ in distribution.

(ii) Let (I.24) hold, in addition to the above conditions. Then the bias vanishes and

$$\sqrt{Nh^{d+2}} \left(\check{\beta}_N - \beta^0\right) \xrightarrow[N \to \infty]{} \mathcal{N}\left(0, \Sigma\right) \quad in \ distribution$$

Proof. Define the sequence

$$Y_i^N := \frac{1}{Nh^{d+1}\ell(0)} \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda^i}{h} \right) + hK \left(\frac{\lambda^0 - \Lambda^i}{h} \right) \frac{\nabla \ell}{\ell}(0) \right) ,$$

and follow the lines of the proof of Theorem 4.1.

4.2 Asymptotic properties of the double Kernel-based estimator

As in the previous section, we start by analyzing the asymptotics of the bias and the variance of $\tilde{\beta}_N$. We first introduce some additional conditions which will be needed in our subsequent analysis.

Assumption KH K and H are the product of some univariate compactly supported lipschitz Kernels with orders respectively p and q, and ∇K has bounded variation.

Assumption S ϕ is continuous and has compact support. Moreover, there exist $\delta > 0$ such that, for every $z \in \mathbb{R}^n$, $\inf \{\varphi(\lambda, z) : (\lambda, z) \in \mathcal{V}(\lambda^0) \times C_{\phi}\} > \delta$, for some neighborhood $\mathcal{V}(\lambda^0)$ of λ^0 , and some compact subset C_{ϕ} of \mathbb{R}^n with $\operatorname{Supp}(\phi) \subset \operatorname{int}(C_{\phi})$.

Assumption R3 For every λ , the function $\nabla_{\lambda} f(\lambda, \cdot)$ is q times differentiable, and for every integer $i \leq q$, the function $\lambda \mapsto \nabla_z^i \nabla_\lambda \varphi(\lambda, z)$ is continuous at $\lambda = \lambda^0$ uniformly with respect to $z \in S$, for some subset S s.t. Supp $(\phi) \subset int(S)$.

Notice that Assumption S restricts seriously the choice of ϕ .

Following the notation (I.20), we define for any function ψ on $\mathbb{R}^d \times \mathbb{R}^n$:

$$\xi_{H}^{q}[\psi](\lambda,z) := \frac{(-1)^{q}}{q!} \sum_{j_{1},\dots,j_{q}=1}^{d} \left(\int v_{j_{1}}\dots v_{j_{q}}H(v)dv \right) \nabla_{z_{j_{1}}\dots z_{j_{q}}}^{q}\psi(\lambda,z) . \quad (I.27)$$

Proposition 4.3 Under Assumptions KH, S, R1, R2 and R3, choose N and h so that

$$h \longrightarrow 0 \quad and \quad \frac{(\ln N)^4}{N \ h^{d+n+n\vee 2}} \longrightarrow 0 \quad as \ N \to \infty.$$
 (I.28)

Then, the bias and the variance of $\tilde{\beta}_N$ satisfy

$$\mathbb{E}\left[\tilde{\beta}_{N}\right] - \beta^{0} \sim C_{3}h^{p} + C_{4}h^{q} + \frac{C_{5}}{Nh^{d+n+1}} \quad and \quad \mathbb{V}\mathrm{ar}\left[\tilde{\beta}_{N}\right] \sim \frac{\tilde{\Sigma}}{Nh^{d+2}}, \quad (I.29)$$

where

$$C_{3} := \frac{1}{\ell(0)} \int \left[\xi_{K}^{p} \left[\ell(\lambda^{0} - .)f_{\lambda} + \varphi_{\lambda} \right] - \frac{\varphi_{\lambda}}{\varphi} \xi_{K}^{p} \left[\varphi \right] \right] (\lambda^{0}, z) \phi(z) dz$$

$$C_{4} := \frac{1}{\ell(0)} \int \left[\xi_{H}^{q} \left[\varphi_{\lambda} \right] - \frac{\varphi_{\lambda}}{\varphi} \xi_{H}^{q} \left[\varphi \right] \right] (\lambda^{0}, z) \phi(z) dz$$

$$C_{5} := \frac{1}{\ell(0)} \int \frac{\phi(z)}{\varphi(\lambda^{0}, z)} K(l_{2} - l_{1}) K(l_{1}) \nabla K(l_{1}) H^{2}(v) dl_{1} dl_{2} dv dz$$

$$\tilde{\Sigma} := \frac{\mathbb{E}[\phi^{2}(Z^{0})]}{\ell(0)} \int \left\{ \int K(l_{2} - l_{1}) \nabla K(l_{1}) dl_{1} \right\}^{\otimes} dl_{2}.$$

The proof of this result involves heavy calculations, and is reported in Section 7.

Theorem 4.3 (i) Under the conditions of Proposition 4.3 hold, we have

$$\sqrt{Nh^{d+2}} \left(\tilde{\beta}_N - \mathbb{E}[\tilde{\beta}_N] \right) \xrightarrow[N \to \infty]{law} \mathcal{N} \left(0, \tilde{\Sigma} \right)$$

(ii) If in addition $Nh^{d+2+2(p \wedge q)} \to 0$, then the bias vanishes and

$$\sqrt{Nh^{d+2}} \left(\tilde{\beta}_N - \beta^0 \right) \xrightarrow[N \to \infty]{law} \mathcal{N} \left(0, \tilde{\Sigma} \right)$$

The proof is also reported in Section 7. Note that it is necessary to have $n < (p \land q) + 1$, in order to satisfy (I.28) and the condition of (ii). Thus, for basket derivatives or bermudean options, it would be necessary to consider high-order kernels.

4.3 Optimal choice of N and h

The two single kernel-based estimators $\hat{\beta}_N$ and $\check{\beta}_N$ have similar asymptotic properties. They both have a bias of order h^p , a variance of order $1/(Nh^{d+2})$ and a convergence in distribution at the rate $\sqrt{Nh^{d+2}}$. Therefore, the determination methods of the optimal N and h will be similar for both of them, and we only detailed calculations for the estimator $\hat{\beta}_N$. Let the conditions of Proposition 4.1 hold. Then (I.21) holds, and we calculate an asymptotic equivalent for the mean square error between $\hat{\beta}_N$ and β^0

$$MSE(\hat{\beta}_N) := \mathbb{E}\left[|\hat{\beta}_N - \beta^0|^2\right] \sim \frac{\mathrm{Tr}(\Sigma)}{Nh^{d+2}} + h^{2p}|C_1|^2$$

Minimizing the MSE in h, we get the asymptotically optimal bandwidth selector :

$$\hat{h} = \left(\frac{(d+2)Tr(\Sigma)}{2p|C_1|^2N}\right)^{1/(d+2p+2)}.$$
(I.30)

Note that \hat{h} is of order $N^{-1/(d+2p+2)}$, leading to an MSE of order $N^{-2p/(d+2p+2)}$. Similarly, the asymptotically optimal bandwidth selector for $\check{\beta}_N$ is

$$\check{h} = \left(\frac{(d+2)Tr(\Sigma)}{2p|C_2|^2N}\right)^{1/(d+2p+2)}.$$
(I.31)

These results imply an asymptotic theoretical choices for h relative to N, but we may still encounter difficulties in the numerical calculation of h. Even if the optimal order of h were known, we still need to evaluate the associated constant coefficients. From our empirical experiments, we observed that the accuracy of our estimators depends heavily on the choice of the bandwidth h, as usual in kernel estimation.

4.4 The case of a uniform randomizing distribution

We first study further the case where the randomizing density is uniform on the sphere of \mathbb{R}^d centered at 0 with radius ϵ . This means we consider the function

$$\ell: l \mapsto \frac{1}{(2\epsilon)^d} \mathbf{1}_{[-\epsilon,\epsilon]}(l) .$$

Observe that this is a particular example from the truncated exponential distributions (I.19) for which the single kernel density estimators coincide :

$$\hat{\beta}_N = \check{\beta}_N = \frac{(2\epsilon)^d}{Nh^{d+1}} \sum_{i=1}^N \phi(Z_i) \nabla K\left(\frac{\lambda^0 - \Lambda_i}{h}\right)$$

Without loss of generality, we assume that the kernel K has support on $[-1, 1]^d$. We first rewrite Assumption R1 in the setting of this section.

Assumption R4 For every z, the function $f(\cdot, z)$ is p + 1 times differentiable, and for every integer $i \leq p + 1$, the function $\lambda \mapsto \nabla^i_{\lambda} f(\lambda, z)$ is continuous at λ^0 uniformly with respects to $z \in S$, for some subset S s.t. $\operatorname{Supp}(\phi) \subset \operatorname{int}(S)$.

Proposition 4.4 Let Assumptions K and R4 hold. Then, as $N \to \infty$, $h \to 0$ and $\epsilon \to 0$ with $\epsilon \ge h$, we have

$$\mathbb{E}\left[\hat{\beta}_{N}\right] - \beta^{0} \sim C_{u}h^{p} \quad and \quad \mathbb{V}\mathrm{ar}\left[\hat{\beta}_{N}\right] \sim N^{-1}h^{-d-2}\epsilon^{d} \Sigma_{u} , \qquad (I.32)$$

where

$$C_u := \int \xi_K^p [f_\lambda] (\lambda^0, z) \phi(z) dz \quad and \quad \Sigma_u := 2^d \mathbb{E}[\phi^2(Z^0)] \int \nabla K^{\otimes}. \quad (I.33)$$

Proof. The proof is similar to the one of Proposition 4.1. Denoting by $\mathbf{1}_{\mathbf{d}}$ the vector of \mathbb{R}^d with unit component, we rewrite

$$\mathbb{E}\left[\hat{\beta}_{N}\right] = \frac{1}{h^{d+1}} \int_{\mathbb{R}^{n}} \phi(z) \left(\int_{\lambda^{0} - \epsilon \mathbf{1}_{\mathbf{d}}}^{\lambda^{0} + \epsilon \mathbf{1}_{\mathbf{d}}} \nabla K\left(\frac{\lambda^{0} - \lambda}{h}\right) f(\lambda, z) \, d\lambda \right) dz$$
$$= \frac{1}{h} \int_{\mathbb{R}^{n}} \phi(z) \left(\int_{\left[-\frac{\epsilon}{h}, \frac{\epsilon}{h}\right]^{d}} \nabla K(u) f(\lambda^{0} - uh, z) \, du \right) dz.$$

Since $\epsilon \geq h$ and K is supported on $[-1,1]^d$, we may replace in our last term the integration on $[-\frac{\epsilon}{h}, \frac{\epsilon}{h}]^d$ by an integration on \mathbb{R}^d , which is necessary to get the convergence of our estimator to β^0 . Then, as in the proof of Proposition 4.1, an integration by parts followed by Taylor expansions lead to the expected equivalent of the bias. The same argument applies for the computation of the variance of $\hat{\beta}_N$. \Box Sending ϵ to zero, we obtain the same asymptotic properties as in Proposition 4.1, as

long as $\epsilon \geq h$. Therefore, the asymptotic optimal ϵ is simply the bandwidth h. The kernel-based estimator $\hat{\beta}_N^u$, associated with this optimal uniform density ℓ is then given by

$$\hat{\beta}_N^u := \frac{2^d}{Nh} \sum_{i=1}^N \phi(Z_i) \nabla K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) , \qquad (I.34)$$

and satisfies

$$\mathbb{E}\left[\hat{\beta}_{N}^{u}\right] - \beta^{0} \sim C_{u}h^{p} \quad \text{and} \quad \mathbb{V}\mathrm{ar}\left[\hat{\beta}_{N}^{u}\right] \sim N^{-1}h^{-2}\Sigma_{u} , \qquad (I.35)$$

with C_u and Σ_u defined in (I.33). Minimizing the corresponding mean square error, we obtain the optimal bandwidth

$$h^{u} := \left(\frac{Tr\Sigma_{u}}{p|C_{u}|^{2}N}\right)^{\frac{1}{2p+2}}.$$
 (I.36)

As in the study of the previous estimators, we also obtain a central limit theorem for the estimator $\hat{\beta}_N^u$.

Theorem 4.4 (i) Let the conditions of Proposition 4.4 hold in the particular case where $\epsilon = h$, and assume further that

$$h \longrightarrow 0 \quad and \quad N h^2 \longrightarrow \infty \quad as \quad N \to \infty.$$
 (I.37)

Then, with Σ_u as in (I.33), we have $\sqrt{Nh^2} \left(\hat{\beta}_N^u - \mathbb{E}[\hat{\beta}_N^u] \right) \longrightarrow \mathcal{N}(0, \Sigma_u)$ in distribution. (ii) If in addition $Nh^{2p+2} \to 0$, then the bias vanishes and :

$$\sqrt{Nh^2}\left(\hat{\beta}_N^u - \beta^0\right) \longrightarrow \mathcal{N}(0, \Sigma_u) \quad in \ distribution.$$

A remarkable feature of the above asymptotic result is that the rate of convergence is independent of the dimension d of the parameter λ^0 .

4.5 The case of a truncated exponential randomizing distribution

Actually, it is possible to improve the asymptotic properties by choosing other densities ℓ . In this subsection, we specialize the discussion to the one-dimensional case, and we consider a truncated exponential randomizing distribution :

$$\ell(l) := \theta \frac{e^{\theta l}}{e^{\theta \epsilon} - e^{-\theta \epsilon}} \mathbf{1}_{[-\epsilon,\epsilon]}(l) ,$$

with the parameter $\theta \in \mathbb{R}$, so that the two single kernel estimators associated to this density coincide:

$$\check{\beta}_N = \hat{\beta}_N = \frac{1}{\ell(0)Nh^{d+1}} \sum_{i=1}^N \phi(Z_i) \left(\nabla K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) + \theta h K \left(\frac{\lambda^0 - \Lambda_i}{h} \right) \right) \,.$$

Using the same line of arguments as in Proposition 4.4, we see that, under Assumptions K and R4, as $N \to \infty$, $h \to 0$ and $\epsilon \to 0$ with $\epsilon \ge h$, we have

$$\mathbb{E}\left[\hat{\beta}_{N}\right] - \beta^{0} \sim C_{e}h^{p} \quad \text{and} \quad \mathbb{V}\mathrm{ar}\left[\hat{\beta}_{N}\right] \sim N^{-1}h^{-3}\epsilon \Sigma_{e} , \qquad (I.38)$$

where $\Sigma_e := \Sigma_u$ defined in (I.33) and

$$C_e := \frac{(-1)^p}{p!} \left(\int u^p K(u) du \right) \sum_{k=1}^{p+1} {p \choose k-1} \left(\int \nabla^k_{\lambda} f(\lambda^0, z) \, \phi(z) \, dz \right) (-\theta)^{p-k+1}.$$
(I.39)

Again, the asymptotic optimal ϵ is simply the bandwidth h and the kernel-based estimator $\hat{\beta}_N^e$, associated with this optimal exponential density is given by

$$\hat{\beta}_{N}^{e} := \frac{e^{\theta h} - e^{-\theta h}}{\theta N h^{2}} \sum_{i=1}^{N} \phi(Z_{i}) \left(\nabla K \left(\frac{\lambda^{0} - \Lambda_{i}}{h} \right) + \theta h K \left(\frac{\lambda^{0} - \Lambda_{i}}{h} \right) \right).$$
(I.40)

The optimal bandwidth is obtained by minimizing the corresponding mean squared error:

$$h^e := \left(\frac{Tr\Sigma_e}{p|C_e|^2 N}\right)^{\frac{1}{2p+2}}, \qquad (I.41)$$

which leads to the following MSE :

$$MSE\left(\hat{\beta}_{N}^{e}\right) = 2(p+1)p^{-\frac{p}{p+1}}\left[|C_{e}|^{2}(Tr\Sigma_{e})^{p}\right]^{\frac{1}{p+1}} N^{-\frac{p}{p+1}}.$$
 (I.42)

As in Theorem 4.4, a central limit theorem for the estimator $\hat{\beta}_N^e$ can be derived.

Remark 4.1 From the asymptotic viewpoint, the estimators based on the truncated exponential randomizing density differ by their bias, as the constants C_e depends on θ while the variance $\Sigma_e = \Sigma_u$ is independent of θ . The optimal truncated exponential randomizing density is then obtained by minimizing the squared bias, defined by the polynomial function C_e^2 , with respect to θ . In our numerical experiments of Section 5, this minimization is performed by classical Newton-Raphson iterations. Unfortunately, it seems to be impossible to exhibit some "universal" ℓ families that would provide some "sharp" lower bounds in every case. Even finding explicitly the "most relevant" ℓ family for a given density f and given dimensions d, n seems to be inaccessible. So, in practice, we advise to introduce a one or two parameters ℓ family, and, as we have done with the truncated exponential family, to choose the parameter values that minimize the asymptotic MSE.

Remark 4.2 Notice that, in both cases, the choice of the radius ϵ of ℓ depends on the kernel function K only through its support. For instance, if $\operatorname{supp}(K) = [-M, M]^d$, then the optimal radius is $\epsilon = Mh$.

4.6 Comparison with the finite differences estimators

We first start by recalling the finite differences estimators. For ease of presentation, we let d = 1. The finite differences estimator of the parameter $\beta^0 := \nabla_{\lambda} \mathbb{E}[\phi(Z(\lambda^0))]$ is based on the finite differences approximation of the gradient

$$\nabla_{\lambda} \mathbb{E}[\phi(Z(\lambda^0))] \sim \frac{\mathbb{E}[\phi(Z(\lambda^0 + \alpha \varepsilon))] - \mathbb{E}[\phi(Z(\lambda^0 - (1 - \alpha)\varepsilon))]}{\varepsilon}$$

where $\varepsilon > 0$ is a "small" parameter, and $\alpha \in [0, 1]$. The values $\alpha = 0, 0.5$ and 1 correspond respectively to the backward, centered and forward finite difference. The above finite difference approximation suggests the following finite differences estimator of β^0 :

$$\hat{\beta}_N^{FD} = \frac{1}{N\varepsilon} \sum_{i=1}^N \left(\phi \left[Z^i (\lambda^0 + \alpha \varepsilon) \right] - \phi \left[Z^i (\lambda^0 - (1 - \alpha) \varepsilon) \right] \right) \ .$$

The asymptotic properties of these estimators were first studied by L'Ecuyer and Perron (1994). In the case where $\lambda \mapsto \phi[Z(\lambda)] \in C_b^3(\mathbb{R}^d)$, when $N \to \infty$ and $\varepsilon \to 0$ with $N^{1/4}\varepsilon \to 0$, they obtained a parametric rate of convergence :

$$\sqrt{N} \left(\hat{\beta}_N^{FD} - \beta^0\right) \xrightarrow[N \to \infty]{} \mathcal{N}\left(0, \Sigma_\alpha\right) \text{ in distribution,} \quad \text{for } \alpha = 0 \text{ , } \frac{1}{2} \text{ and } 1 \text{ .}$$

When the payoff function ϕ has a countable number of discontinuities, Detemple, Garcia and Rindisbacher (2005) obtained the following central limit theorems :

For
$$\alpha = \frac{1}{2}$$
, when $N^{1/5} \varepsilon \to 0$, $N^{2/5} \left(\hat{\beta}_N^{FD} - \beta^0 \right) \underset{N \to \infty}{\longrightarrow} \mathcal{N}(0, \Sigma_{\alpha})$ in distribution
For $\alpha = 0, 1$, when $N^{1/3} \varepsilon \to 0$, $N^{1/3} \left(\hat{\beta}_N^{FD} - \beta^0 \right) \underset{N \to \infty}{\longrightarrow} \mathcal{N}(0, \Sigma_{\alpha})$ in distribution

In the general case $d \ge 1$, the finite differences estimators are defined componentwise, and therefore, the rate of convergence is not affected by the dimension d of the parameter λ^0 .

The main objective of this paragraph is to provide an asymptotic comparison of the single-kernel based estimator with the finite differences one. The key point of our single-kernel based estimators is that the differentiation with respect to the parameter λ is reported on the density of $Z(\lambda)$ so that our asymptotic results do not involve the regularity of the pay-off function ϕ . For any pay-off function ϕ , and when $N h^{d+2p+2} \longrightarrow 0$, we derived in Theorems 4.1 and 4.2 that

$$\sqrt{Nh^{d+2}} \left(\hat{\beta}_N - \beta^0\right) \xrightarrow[N \to \infty]{} \mathcal{N}\left(0, \Sigma\right)$$
 in distribution,

where p is the order of the kernel function. Minimizing the corresponding MSE, we obtained in Section 4.3 an optimal h of order $N^{-1/(d+2p+2)}$ which, of course, almost satisfies the condition required for the convergence in distribution. Therefore, taking a bandwidth h of order $N^{-1/(d+2p+2)-2\delta/(d+2)}$ with $\delta > 0$ sufficiently small leads to a convergence in distribution at rate N^r with $r := p/(d+2p+2) - \delta > 0$. Therefore, the single-kernel based estimators, with kernel of order p > 2d + 4 and δ sufficiently small, achieve a convergence rate of order r > 2/5. Hence, they outperform all the finite differences estimators in the case of discontinuous payoffs.

Notice that, by taking kernel functions of order p sufficiently large, we can obtain a convergence rate in distribution as close as desired to the parametric rate \sqrt{N} .

Remark 4.3 Consider the optimized kernel estimators $\hat{\beta}_n^u$ and $\hat{\beta}_n^e$, based on uniform or exponential density ℓ on the sphere with radius h, as derived in section 4.4. Then, for $Nh^{2p+2} \to 0$, we obtain a rate of convergence of $\sqrt{Nh^2}$. Therefore, in order to outperform the finite differences estimators of a Greek associated to a discontinuous payoff function ϕ , one just needs to use a kernel function of order p > 4.

5 Numerical results

In this section, we present some numerical results obtained in the Black-Scholes model :

$$S_t^x := x \exp\left[\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t\right], \quad t \ge 0, \ x > 0,$$

where W is a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in \mathbb{R} , and $r \in \mathbb{R}$, $\sigma > 0$ are two given constants. We focus on the estimation of the so-called Delta :

$$\beta^0 := \nabla_x \mathbb{E}[\phi(Z^x)],$$

where $Z^x = S_T^x$ for an European option and $Z^x = \int_0^T S_t^x dt$ for an Asian option. As in the previous sections, we denote by f(x, .) the density of Z^x .

We simulate independent observations X_i distributed in the (optimal) exponential randomizing distribution ℓ on the sphere centered at $S_0 = x$ with radius h, as derived in section 4.5. The single-kernel based estimator $\hat{\beta}_N^e$ is therefore given by (I.40).

5.1 Computation of the optimal bandwidth

As the "bumping" parameter ϵ for the finite differences estimator, the bandwidth in kernel estimation needs to be chosen carefully. The asymptotic results of Section 4 provide the expression of the asymptotic optimal bandwidth. For the truncated exponential randomizing distribution, we obtain

$$h^e = \left(\frac{\Sigma_e}{pC_e^2 N}\right)^{\frac{1}{2p+2}},$$

where $\Sigma_e = 2 \mathbb{E}[\phi^2(Z^x)] \int (\nabla K)^2$ and

$$C_e := \frac{(-1)^p}{p!} \left(\int u^p K(u) du \right) \sum_{k=1}^{p+1} {p \choose k-1} \mathbb{E} \left[\phi(Z^x) \frac{\nabla_x^k f(x, Z^x)}{f(x, Z^x)} \right] (-\theta)^{p-k+1}$$

Given a kernel function K, the coefficient Σ_e can be estimated by a standard Monte Carlo procedure. We next focus on the estimation of the parameter

$$\mathcal{E}_k := \mathbb{E}\left[\phi(Z^x) \frac{\nabla_x^k f(x, Z^x)}{f(x, Z^x)}\right]$$

for a given $k \in \{1, ..., p+1\}$.

(i) Let $Z^x = S_T^x = x e^Y$, where Y has a normal distribution with mean $m := (r - \frac{\sigma^2}{2})T$ and variance $\Sigma := \sigma^2 T$. Then, it is easily checked that :

$$\nabla_x^k f(x,z) = \left[\sum_{i=0}^k a_i^k d(x,z)^i\right] \frac{f(x,z)}{x^k}$$
5. NUMERICAL RESULTS

where

$$d(x,z) := \frac{\ln z - \ln x - m}{\Sigma}, \qquad (I.43)$$

and the coefficients $(a_i^j)_{(i,j)\in\{0,\dots,k\}^2}$ are given by

$$a_i^0 = \mathbf{1}_{\{i=0\}}, \quad a_i^{j+1} = a_{i-1}^j - j a_i^j - \frac{i+1}{\Sigma} a_{i+1}^j,$$
 (I.44)

with the convention $a_i^j = 0$ for i < 0 and i > j. Hence :

$$\mathcal{E}_k = \frac{1}{x^k} \mathbb{E}\left[\phi(Z^x)\left(\sum_{i=0}^k a_i^k d(x, Z^x)^i\right)\right],$$

and this parameter can be estimated by a straightforward Monte Carlo procedure.

(ii) In practice, the distribution function is unknown, and the calculation of the previous paragraph can not be used to estimate \mathcal{E}_k . We suggest to mimic the same principle as the usual Silverman's rule-of-thumb in kernel estimation (see Scott [99] e.g.) : let \hat{m} and $\hat{\Sigma}$ be respectively two given estimates of the mean the variance $\ln(Z^x/x)$, and define $\hat{d}(x,z)$ and $(\hat{a}_i^j)_{(i,j)\in\{0,\ldots,k\}^2}$ by substituting $(\hat{m},\hat{\Sigma})$ to (m,Σ) in (I.43)-(I.44); then the coefficient \mathcal{E}_k is approximated by

$$\hat{\mathcal{E}}_k = \frac{1}{x^k} \mathbb{E}\left[\phi(Z^x) \left(\sum_{i=0}^k \hat{a}_i^k \hat{d}(x, Z^x)^i\right)\right].$$

Once the coefficients \mathcal{E}_k estimated for $1 \leq k \leq p+1$, the parameter θ is chosen through a numerical minimization, see Remark 4.1. In the particular case of an uniform randomizing distribution ($\theta = 0$), remark that only the estimation of \mathcal{E}_{p+1} is necessary.

Therefore, the numerical procedure is divided in three steps: first, we estimate the terms detailed in the previous subsection Σ_e , \mathcal{E}_k , \hat{m} and $\hat{\Sigma}_e$ through a Monte Carlo procedure with very few simulations. Then, we calibrate the parameter θ by minimization and we deduce the exponential optimal theoretical bandwidth . Finally we estimate the delta of the option by means of a single-kernel based estimator with the estimated bandwidth.

Remark 5.1 The numerical effort dedicated to the calculation of the optimal bandwidth parameter h is also encountered in the classical finite differences method, as the optimal bumping parameter ϵ involves some a priori numerical simulations.

5.2 Numerical comparison of the estimators

We present here numerical results obtained for the estimation of the delta of an European and an Asian at-the-money digital calls, i.e. with a payoff of the form $\phi(s) = \mathbf{1}_{s>K}$. Since this payoff function is discontinuous, the results of Section 4.6 show that the singlekernel based estimator achieves a better rate of convergence than the finite differences estimators, whenever the kernel has order p > 4. The main object of this section is to verify the empirical validity of these asymptotic results.

In order to compare their behavior, each estimator has been computed 200 times and their empirical distributions have been smoothed by a Gaussian kernel.

Our numerical experiments are performed with the following values of the parameters :

$$S_0 = 120, \quad r = 0, \quad \sigma = 0.2, \quad T = 1, \text{ and } K = 120.$$

We use the following polynomial kernel functions of order 2, 4 and 6, respectively, with support on [-1, 1]:

$$K_{2}(u) = \frac{3}{4}(1-u^{2}),$$

$$K_{4}(u) = \frac{15}{32}(1-u^{2})(3-7u^{2}),$$

$$K_{6}(u) = \frac{105}{256}(1-u^{2})(33u^{4}-30u^{2}+5)$$

From the viewpoint of computing time, kernel based or finite differences estimations with the same number of simulations are comparable. All the numerical tests have been realized in Visual C++ on a Pentium 4 xeon 3 GHz processor with 1 Gb of RAM.

European Digital Call Option In the context of the Black-Scholes model, it was observed by [50] that the optimal weight for European options can be obtained by means of the Malliavin integration by part formula, and coincides with the likelihood estimator introduced by [23]. Therefore, we are not hoping to compete with the Malliavin-based Monte Carlo estimator.

From our numerical experiments, we observed that the gain from using kernel estimators based on an exponential rather than a uniform randomizing distribution ℓ was very poor, especially when the order of the kernel function increases. From a numerical viewpoint, the gain obtained at most counter-balanced the numerical price of the minimization procedure. The examples presented here are therefore based on a uniform randomization distribution ℓ .



Figure 2: Delta of an European Digital Call, N = 1 Million



Figure 3: Delta of an European Digital Call, N = 1 Billion

The distributions of the different estimators based on $N = 10^6$ simulations are reported in Figure 2. The good performance of the Malliavin estimator is confirmed by our numerical experiments. However, we observe surprisingly that the three kernel based estimators are less accurate than the centered finite differences one, although their numerical computing times are comparable, of the order of 2 seconds. According to Section 4.6, the kernel of order 6 should perform better than the other ones, but this is not the case here. Actually, the terms C_e and Σ_e are such that the constant term of the mean square error increases very fast with the variability of K, which naturally increases with its order. For example, the MSE of the estimator based on the kernel of order 4 is ten times bigger than the one of the finite differences one, although they have the same rate of convergence. Furthermore, the optimal bandwidth h increases with the order of the kernel, so that the asymptotic approximations become less accurate.

In order to further investigate this effect, we increase the number of simulations. Figure 3 shows the distribution of the finite differences estimator and the kernel based estimator of order 6 based on $N = 10^9$ simulations where each simulation takes approximately 30 minutes on our computer. In this case, we observe that the kernel based estimator of order 6 truly outperforms the finite differences one: its bias and its variance are two times smaller. This confirms the theoretical asymptotic results obtained in section 4.6. We do not consider that the high number of simulations required is a serious restriction since it is just a matter of computer power or time given to the simulation. Furthermore, the good performance of the kernel based estimators of high order can be observed for a smaller number of simulation if we use in addition variance reduction technique. For example, by performing the simple antithetic variable technique with respect to the randomizing density ℓ , we observe that the kernel based estimator of order 6 outperforms the finite differences estimator as soon as the number of simulations exceeds $6 * 10^7$, corresponding to a computer time of about 2 minutes.

Asian Digital Call Option We next investigate the case of an Asian option, where the Malliavin integration by parts formula does not lead to the optimal weight, see [50]. The distribution of the different estimators based on $N = 10^6$ simulations are reported in Figure 4, where the "true value" of the Greek has been approximated by an unbiased Malliavin estimation with a very large number of simulations. Even if the Malliavin weight is not optimal, the Malliavin estimator still outperforms the other estimators. As for the European digital call, the finite differences estimator outperforms the kernel based estimators but one simply requires more simulations in order to make the kernel estimator of order 6 more efficient than the finite differences one.



Figure 4: Delta of an Asian Digital Call, N = 1 Million

Conclusion (numerical results) Other tests realized with different parameters, payoff functions or randomizing densities lead to rather similar results. Our kernel based estimator with order p > 4 of the delta of a digital option outperforms asymptotically the finite differences one, but one requires a large number of simulation to verify this fact empirically. Nevertheless, the high number of simulations required can be significantly reduced by means of variance reduction techniques. When the density of the underlying is unknown and the pay-off function is irregular, the Malliavin based estimator is still more efficient than the others. Nevertheless, in general, Malliavin weights are very difficult to derive analytically and this is precisely the advantage of the other estimators which are straightforward to implement.

6 Short maturity asymptotics

In this section, we study further the asymptotic properties of the single kernel based estimators when $Z(\lambda)$ is the time t realization of a Markov process defined by a stochastic differential equation parameterized by λ . We first justify the importance of this short time analysis for the purpose of financial applications, by presenting several examples pointing out the singularity of the Greek weights of the Malliavin-based estimators in this context. We then study the behavior of $\hat{\beta}_N$ when the bandwidth of the kernel and the maturity shrink to zero, as the number of observations goes to infinity. This allows to derive the (theoretical) relative orders for these three parameters and provides a simpler method for the estimation of the optimal bandwidth.

6.1 Singularity of the Greek weights for short maturity

Example 6.1 Vanilla options with short maturity.

Let $Z(x) := x + W_t$. Then, the density of Z(x) is Gaussian and the score function is given by $s(x,z) := \nabla_x \ln f(x,z) = (z-x)/t$. Hence the optimal Greek weight is the random variable $S^0 := W_t/t$.

This example shows the explosion of the Greek weight for short maturity. This feature is by no means specific to the gaussian case. It is shown in [50] that this is the rule for any continuous-time process defined as the solution of a (smooth) stochastic differential equation. the next examples show that the problem of short maturity singularity is encountered in a larger class of problem beyond the above case of European options.

Example 6.2 Path dependent options with fixed maturity. Let π : $0 = t_0 < t_1 < \ldots < t_s = 1$ be a partition of the interval [0, 1], and let

$$Z(\lambda) = \phi(X_{t_1}(\lambda), \dots, X_{t_s}(\lambda))$$

where $\{X_t(\lambda), t \in [0, 1]\}$ is some given continuous-time Markov process parameterized by λ . The partition π is typically a time-grid on which the continuous-time process is discretized. So one should think about the mesh $\max\{|t_i - t_{i-1}| : 1 \leq i \leq s\}$ to be small. By the Markov property, we have $\mathbb{E}[\phi(X_{t_1}(\lambda), \ldots, X_{t_s}(\lambda))] = \mathbb{E}[\bar{\phi}(X_{t_1}(\lambda))],$ where $\bar{\phi}(x) := \mathbb{E}[\phi(X_{t_1}(\lambda), \ldots, X_{t_s}(\lambda))|X_{t_1}(\lambda) = x]$. Therefore, the Malliavin Greek weights derived in [50] or [54] for this path dependent option are the same as those derived for the random variable

$$Z(\lambda) := X_{t_1}(\lambda),$$

so that we are reduced to the short maturity t_1 which induces singular Greek weights.

Example 6.3 American option / optimal stopping problems with fixed maturity.

Consider the Bermudean approximation V_0 of an American style option with fixed maturity, i.e. the optimal stopping problem with stopping possibilities restricted to the partition π defined in the previous example. Then, by the so-called dynamic programming principle, the value of the Bermudean option can be computed by the backward scheme :

$$V_s(\lambda) := \phi(X_{t_s}(\lambda)) \text{ and } V_{i-1}(\lambda) := \max\left\{\phi\left(X_{t_{i-1}}(\lambda)\right), \mathbb{E}\left[V_i(\lambda)|\mathcal{F}_{t_{i-1}}\right]\right\}$$

From the Markov property of the process $X(\lambda)$, $V_0(\lambda) = \mathbb{E} [\psi(X_{t_1}(\lambda)) | X_0 = x]$ in the *continuation region* $\{x : \phi(x) < V_0(\lambda)\}$, where $\psi(X_{t_1}(\lambda)) = V_1(\lambda)$, and we are reduced again to a short maturity context, implying the singularity of the Greek weights.

6.2 Parameterized stochastic differential equation

Let $\{X_u^{\lambda}, u \geq 0\}$ be a process with values in \mathbb{R}^n defined by the stochastic differential equation

$$X_0^{\lambda} = x(\lambda) , \qquad dX_u^{\lambda} = \mu(u,\lambda,X_u^{\lambda})du + \sigma(u,\lambda,X_u^{\lambda})dW_u , \qquad (I.45)$$

where W is a Brownian motion with values in \mathbb{R}^n , and the functions x, μ and σ satisfy the following assumption:

Assumption SDE The function x(.) belongs to $\mathbf{C}^{p+2}(\mathbb{R}^d, \mathbb{R}^n)$, and the coefficients μ, σ are continuous with $\mu(u,.,.) \in \mathbf{C}_b^{p+3}(\mathbb{R}^d \times \mathbb{R}^n, \mathbb{R}^n)$ and $\sigma(u,.,.) \in \mathbf{C}_b^{p+3}(\mathbb{R}^d \times \mathbb{R}^n, \mathcal{M}_{\mathbb{R}}^n)$ for every $u \in \mathbb{R}_+$.

In this section, we are interested in the behaviour of the estimator $\hat{\beta}_N$ when $Z(\lambda) = X_t^{\lambda}$ for a small t > 0. Since t is now an important variable, we shall emphasize more the dependence of V^{ϕ} on t by denoting

$$V^{\phi}(\lambda, t) := \mathbb{E}\left[\phi\left(X_t^{\lambda}\right)\right]$$

The main objective of our analysis is to device an optimal choice of the number of simulations N and the bandwidth h for $\hat{\beta}_N$ given a short maturity t, i.e. given $t \longrightarrow 0$. Since $\nabla_{\lambda} V^{\phi}(\lambda^0, t)$ converges to $\beta^0 := \nabla_{\lambda} \phi\left(X_0^{\lambda^0}\right)$, the present context requires further smoothness conditions on the function ϕ .

Lemma 6.1 Under Assumption SDE, the solution X^{λ} of the stochastic differential equation (I.45) is p + 2 times differentiable in λ and each of the derivatives $\nabla_{\lambda}^{i} X_{t}^{\lambda}$ is locally (α, β) -Hölder continuous in (t, λ) for any $\alpha < 1$ and $\beta < \frac{1}{2}$. Furthermore, for any compact sets $\mathcal{K} \subset \mathbb{R}^{k}$ and $\mathcal{L} \subset \mathbb{R}$, we can find $M \in \mathbb{R}$ such that for any $\lambda_{1}, \lambda_{2} \in \mathcal{K}$, $t_{1}, t_{2} \in \mathcal{L}$:

$$\mathbb{E}\left[\left|\nabla_{\lambda}^{i}X_{t_{1}}^{\lambda_{1}}-\nabla_{\lambda}^{i}X_{t_{2}}^{\lambda_{2}}\right|\right] \leq M\left(\left|\lambda_{1}-\lambda_{2}\right|+\left|t_{1}-t_{2}\right|^{\frac{1}{2}}\right), \quad (I.46)$$

and

$$\sup_{\lambda \in \mathcal{K}, t \in \mathcal{L}} \mathbb{E}\left[\left| \nabla_{\lambda}^{i} X_{t}^{\lambda} \right|^{k} \right] < \infty \quad for \ all \quad k \in \mathbb{N} \,. \tag{I.47}$$

Proof. We first introduce the functions $\bar{\mu}(u, x, \lambda) := (\mu(u, \lambda, x)', 0) \in \mathbb{R}^{n+d}$ and $\bar{\sigma}(u, x, \lambda) := (\sigma(u, \lambda, x)', 0) \in \mathcal{M}^{n+d}_{\mathbb{R}}$, and consider the process Y defined by the stochastic differential equation

$$Y_0 = y$$
 and $dY_u = \bar{\mu}(u, Y_u)du + \bar{\sigma}(u, Y_u)dW_u$

so that the parameterized process X^{λ} coincides with the first *n* components of the process *Y* with initial condition $y = (x(\lambda), \lambda)$. Under Assumption SDE, the coefficients of the stochastic differential equation defining *Y* are in $\mathbf{C}_{b}^{p+2}(\mathbb{R}^{d}, \mathbb{R}^{n})$. From Theorem 3.3 p. 223 in Kunita [71], we conclude that the flow $Y_{t}(y)$ is p + 2 times differentiable with respect to its initial value *y*, and every derivative $\nabla_{y}^{k}Y_{t}$ is locally (α, β) -Hölder continuous in (t, y) for any $\alpha < 1$ and $\beta < \frac{1}{2}$. Now, since the function $\lambda \mapsto x(\lambda)$ is smooth, this property is inherited by the process X^{λ} .

We next turn to the proof of (I.46). It is shown in the proof of Theorem 3.3 p. 223 in [71], that, given two solutions Y^1 and Y^2 starting respectively at y_1 and y_2 , there exists a constant C such that, for any p > 2, we have, for $s, t \ge 0$,

$$\mathbb{E}\left[|\nabla^{k}Y_{t}^{1} - \nabla^{k}Y_{s}^{2}|^{p}\right] \leq C\left(|y_{1} - y_{2}|^{p} + (1 + |y_{1}| + |y_{2}|)^{p}|s - t|^{\frac{p}{2}}\right).$$
(I.48)

Since the \mathcal{L}^1 norm is dominated by the \mathcal{L}^p norm, and the function x(.) is locally Lipschitz, this implies (I.46). Finally, (I.47) is a direct consequence of (I.48).

6.3 Asymptotic properties

The infinitesimal generator of the process X^{λ} is given by

$$\mathcal{L}_t^{\lambda}g(x) := \mu(t,\lambda,x) \cdot \nabla g(x) + \frac{1}{2} \operatorname{Tr} \left[\sigma \sigma'(t,\lambda,x) \nabla^2 g(x)\right], \qquad g \in C^2(\mathbb{R}^n,\mathbb{R})$$

As in the previous section, we consider a sequence (Λ_i, X_t^i) of independent pairs of random variables where the distribution density of Λ_i is $\ell(\lambda^0 - .)$, and X_t^i is the solution of the stochastic differential equation (I.45) with parameter λ fixed to Λ_i . In view of the results of the previous section, we shall only consider the estimator of the Greek defined by

$$\hat{\beta}_N^t = \frac{1}{\ell(0)Nh^{d+1}} \sum_{i=1}^N \phi(X_t^i) \left(\nabla K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) - hK\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \frac{\nabla \ell}{\ell} (\lambda^0 - \Lambda_i) \right) \,.$$

Theorem 6.1 Let the Kernel function K be of order p > 0, and let Assumption SDE hold. Assume further that the density function ℓ is in $C^{p+1}(\mathbb{R}^d, \mathbb{R})$ and ϕ is in $C^{p+3}(\mathbb{R}^d, \mathbb{R})$. If we have

$$h \longrightarrow 0, \quad t \longrightarrow 0 \quad and \quad N h^{d+2} \longrightarrow \infty \quad as \quad N \to \infty,$$
 (I.49)

then the bias and the variance of $\hat{\beta}_N$ satisfy

 $\mathbb{E}[\hat{\beta}_N^t] - \beta^0 \sim C_6 h^p + C_7 t, \quad \mathbb{E}[\hat{\beta}_N^t] - \nabla_\lambda V^\phi(\lambda^0, t) \sim C_6 h^p \text{ and } \mathbb{Var}[\hat{\beta}_N] \sim \frac{\Sigma_0}{Nh^{d+2}},$

where

$$C_{6} := \frac{(-1)^{p}}{p!\,\ell(0)} \sum_{j_{1},\dots,j_{p}=1}^{d} \nabla^{p}_{\lambda_{j_{1}},\dots,\lambda_{j_{p}}}[\ell(0)\nabla(\phi \circ x)(\lambda^{0})] \int l_{j_{1}}\dots l_{j_{p}}K(l) \, dl ,$$

$$C_{7} := \nabla_{\lambda} \left[\mathcal{L}_{0}^{\lambda^{0}}\phi\left(x(\lambda^{0})\right) \right] ,$$

$$\Sigma_{0} := \frac{1}{\ell(0)} \phi^{2}(x(\lambda^{0})) \int \nabla K^{\otimes}(l) dl ,$$

and the asymptotic distribution of $\hat{\beta}_N^t$ is given by

$$\sqrt{Nh^{d+2}} \left(\hat{\beta}_N^t - \mathbb{E} \left[\hat{\beta}_N^t \right] \right) \xrightarrow[N \to \infty]{law} \mathcal{N} \left(0, \Sigma_0 \right) .$$
 (I.50)

If, in addition, $Nh^{d+2+2p} \longrightarrow 0$ as $N \rightarrow \infty$, we get :

$$\sqrt{Nh^{d+2}} \left(\hat{\beta}_N^t - \nabla_\lambda V^\phi(\lambda^0, t) \right) \xrightarrow[N \to \infty]{law} \mathcal{N}(0, \Sigma_0) .$$
 (I.51)

And the addition of condition $Nh^{d+2}t^2 \longrightarrow 0$ as $N \rightarrow \infty$, leads to

$$\sqrt{Nh^{d+2}} \left(\hat{\beta}_N^t - \beta^0 \right) \xrightarrow[N \to \infty]{law} \mathcal{N} \left(0, \Sigma_0 \right) . \tag{I.52}$$

Before proceeding to the proof of this result, let us comment on the optimal choice of N and h given a short time t. Since, we are trying to estimate $\nabla_{\lambda} V^{\phi}(\lambda^{0}, t)$ using $\hat{\beta}_{N}^{t}$, we have to minimize

$$\mathbb{E}\left[\left|\hat{\beta}_N^t - \nabla_\lambda V^\phi(\lambda^0, t)\right|^2\right] \sim \frac{Tr(\Sigma_0)}{Nh^{d+2}} + |C_6|^2 h^{2p}.$$

Then, as in the fixed time study, the optimal bandwidth h^* is given by :

$$h^* = \left(\frac{(d+2)Tr(\Sigma^0)}{2p|C_6|^2N}\right)^{1/(d+2p+2)}.$$
 (I.53)

Indeed, Theorem 6.1 says that, considering a process X evaluated at a short time t, the asymptotic equivalents of the bias and of the variance are obtained by sending t to

zero in the expressions of the fixed maturity case. From a practical point of view, the interest is that the constants C_6 and Σ^0 do not depend on time t and are much easier to evaluate than the corresponding C_1 and Σ .

Proof of Theorem 6.1 We split the proof in three steps.

1. We first study the bias term. Using the same technique as in the proof of Proposition 4.1, we obtain

$$\mathbb{E}\left[\hat{\beta}_{N}^{t}\right] = \frac{1}{\ell(0)h^{d+1}} \mathbb{E}\left[\phi\left(X_{t}^{\Lambda}\right)\left(\nabla K\left(\frac{\lambda^{0}-\Lambda}{h}\right)-hK\left(\frac{\lambda^{0}-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\lambda^{0}-\Lambda)\right)\right] \\
= \frac{1}{\ell(0)h^{d+1}} \mathbb{E}\left[V^{\phi}(\Lambda,t)\left(\nabla K\left(\frac{\lambda^{0}-\Lambda}{h}\right)-hK\left(\frac{\lambda^{0}-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\lambda^{0}-\Lambda)\right)\right] \\
= \frac{(-1)}{\ell(0)h^{d}}\int V^{\phi}(\lambda,t)\nabla\left(K\left(\frac{\lambda^{0}-\lambda}{h}\right)\ell(\lambda^{0}-\lambda)\right)d\lambda \\
= \frac{1}{\ell(0)h^{d}}\int \nabla_{\lambda}V^{\phi}(\lambda,t)K\left(\frac{\lambda^{0}-\lambda}{h}\right)\ell(\lambda^{0}-\lambda)d\lambda \\
= \frac{1}{\ell(0)}\int \nabla_{\lambda}V^{\phi}(\lambda^{0}-hl,t)\ell(hl)K(l)\,dl\,.$$
(I.54)

We will use the latter expression in order to derive an expansion of the bias with respect to the pair (h, t) near the origin. Before this, let us derive a suitable representation of $\nabla_{\lambda} V^{\phi}(\lambda^0 - hl, t)$. Since $\nabla \phi$ and σ are bounded, it follows from Itô's lemma that :

$$V^{\phi}(\lambda^{0} - hl, t) = \phi\left(x(\lambda^{0} - hl)\right) + \mathbb{E}\left[\int_{0}^{t} \mathcal{L}_{s}^{\lambda^{0} - hl}\phi\left(X_{s}^{\lambda^{0} - hl}\right)ds\right].$$

By (I.47) of Lemma 6.1, the above expression is differentiable with respect to λ^0 , and :

$$\nabla_{\lambda} V^{\phi}(\lambda^{0} - hl, t) = \nabla_{\lambda}(\phi \circ x)(\lambda^{0} - hl) + \mathbb{E}\left[\int_{0}^{t} \nabla_{\lambda} [\mathcal{L}_{s}^{\lambda^{0} - hl}\phi\left(X_{s}^{\lambda^{0} - hl}\right)]ds\right]$$

Combining this equality with (I.54), we decompose $\mathbb{E}[\hat{\beta}_N] - \beta^0$ into three pieces

$$\mathbb{E}[\hat{\beta}_N^t] - \beta^0 = \mathbf{A} + \mathbf{B} + \mathbf{C},$$

where A, B and C are defined below.

(i) The first term is given by

$$\mathbf{A} := \frac{1}{\ell(0)} \int \nabla_{\lambda} (\phi \circ x) (\lambda^0 - hl) \ell(hl) K(l) dl - \beta^0 \sim C_6 h^p$$

where C_6 is defined in the statement of the theorem, and the latter equivalence follows from the fact that p is the order of K by the same argument as in Proposition 4.1. (ii) The second term is given by

$$\mathbf{B} := \mathbb{E}\left[\int_0^t \nabla_{\lambda} [\mathcal{L}_s^{\lambda^0} \phi\left(X_s^{\lambda^0}\right)] ds\right] \sim t \nabla_{\lambda} [\mathcal{L}_0^{\lambda^0} \phi\left(x(\lambda^0)\right)] = C_7 t,$$

as a consequence of the a.s. continuity at the origin of the map $s \mapsto \nabla_{\lambda^0} \mathcal{L}_s \phi(X_s^{\lambda^0})$, and the dominated convergence theorem together with (I.46) of Lemma 6.1.

(iii) We now show that the remaining term C which rewrites

$$C := \int \mathbb{E}\left[\int_0^t \left(\nabla_{\lambda} [\mathcal{L}_s^{hl}\phi] \left(X_s^{\lambda^0 - hl}\right) \frac{\ell(hl)}{\ell(0)} - \nabla_{\lambda} [\mathcal{L}_s^{\lambda^0}\phi \left(X_s^{\lambda^0}\right)]\right) ds\right] K(l) dl,$$

is dominated by A and B. To see this, observe that the first p-1 terms of the order p Taylor expansion of the integrand disappear, by the fact that p is the order of the Kernel K. Using (I.47) of Lemma 6.1 and the regularity of the derivatives of μ , σ and ϕ , the expectation of the remainder term in the expansion can be bounded uniformly in s and l. Therefore, $|\mathbf{C}| = O(th^p)$ and \mathbf{C} is negligible with respect to A and B. Thus $E\left[\hat{\beta}_N^t\right] - \beta^0 \sim C_6 h^p + C_7 t$ as announced in the statement of theorem. And, noticing simply that $\mathbb{E}[\hat{\beta}_1^t] - \nabla_{\lambda} V^{\phi}(\lambda^0 | t) = A + C$, we get the second announced result

noticing simply that $\mathbb{E}[\hat{\beta}_N^t] - \nabla_{\lambda} V^{\phi}(\lambda^0, t) = A + C$, we get the second announced result $\mathbb{E}[\hat{\beta}_N^t] - \nabla_{\lambda} V^{\phi}(\lambda^0, t) \sim C_6 h^p$.

2. We now compute the variance of $\hat{\beta}_N^t$. As in Proposition 4.1, we can rewrite :

$$\operatorname{Var}\left[\phi\left(X_{t}^{\Lambda}\right)\left(\nabla K\left(\frac{\lambda^{0}-\Lambda}{h}\right)-hK\left(\frac{\lambda^{0}-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\lambda^{0}-\Lambda)\right)\right]=v_{1}-v_{2}^{\otimes},$$

where

$$v_2 := \mathbb{E}\left[\phi(X_t^{\Lambda})\left(\nabla K\left(\frac{\lambda^0 - \Lambda}{h}\right) - hK\left(\frac{\lambda^0 - \Lambda}{h}\right)\frac{\nabla \ell}{\ell}(\lambda^0 - \Lambda)\right)\right] = O\left(h^{d+1}\right),$$

and,

$$v_{1} := \mathbb{E}\left[\phi^{2}\left(X_{t}^{\Lambda}\right)\left(\nabla K\left(\frac{\lambda^{0}-\Lambda}{h}\right)-hK\left(\frac{\lambda^{0}-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\lambda^{0}-\Lambda)\right)^{\otimes}\right]$$
$$= \mathbb{E}\left[V^{\phi^{2}}(\Lambda,t)\left(\nabla K\left(\frac{\lambda^{0}-\Lambda}{h}\right)-hK\left(\frac{\lambda^{0}-\Lambda}{h}\right)\frac{\nabla\ell}{\ell}(\lambda^{0}-\Lambda)\right)^{\otimes}\right]$$
$$= h^{d}\int V^{\phi^{2}}(\lambda^{0}-hl,t)\left(\nabla K(l)-hK(l)\frac{\nabla\ell}{\ell}(hl)\right)^{\otimes}\ell(hl)\,dl\,.$$

Now observe that the following equivalence

$$\left(\nabla K(l) - hK(l)\frac{\nabla \ell}{\ell}(hl)\right)^{\otimes} \ell(hl) \sim \nabla K(l)^{\otimes} \ell(0) + hK(l)\frac{\nabla \ell}{\ell}(0) + C lh,$$

holds uniformly in l for some constant C. Also, from the first step of this proof, we have

$$V^{\phi^2}(\lambda^0 - hl, t) = \phi^2\left(x(\lambda^0)\right) + O(t),$$

uniformly with respect to l in a compact subset. Then

$$v_1 \sim h^d \ell(0) \phi^2(x(\lambda^0)) \int \nabla K^{\otimes}.$$

Hence v_2^{\otimes} is dominated by v_1 , and we get the expression of the variance reported in the statement of the theorem follows from the last equivalence.

3. We now turn to the derivation of the asymptotic distribution of $\hat{\beta}_N^t$. The proof is again similar to that of Theorem 4.1, and consists in verifying the Lyapounov conditions (Billingsley [14], p.44). Let *a* be a *d*-dimensional vector and let us define, for every $i = 1, \ldots, N$,

$$\begin{split} U_i^N &:= \frac{1}{Nh^{d+1}\ell(0)} \phi\left(X_t^i\right) \left(\nabla K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) - hK\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \frac{\nabla \ell}{\ell} (\lambda^0 - \Lambda_i)\right) \,, \\ V_i^N &:= a' U_i^N - \mathbb{E}\left[a' U_i^N\right] \,. \end{split}$$

It is sufficient to show that, for some $\delta > 2$, we have

$$\sup_{N} \frac{1}{\sigma_{N}^{\delta}} \sum_{i=1}^{N} \mathbb{E}\left[|V_{i}^{N}|^{\delta} \right] < \infty \quad \text{where} \quad \sigma_{N}^{2} := \mathbb{V} \operatorname{ar}\left[\sum_{i=1}^{N} V_{i}^{N} \right].$$
(I.55)

To check this, we directly estimate by the Minkowski inequality that $\|V_i^N\|_{\delta}$ is bounded by

$$\begin{split} \|V_i^N\|_{\delta} &\leq \frac{\sum_{i=1}^d \left\|a_i \phi\left(X_t^\Lambda\right) \left(\nabla_i K\left(\frac{\lambda^0 - \Lambda}{h}\right) - hK\left(\frac{\lambda^0 - \Lambda}{h}\right) \frac{\nabla_i \ell}{\ell} (\lambda^0 - \Lambda)\right)\right\|_{\delta}}{Nh^{d+1} \ell(0)} + \frac{C}{N} \\ &\leq \frac{|a|_{\infty} \sum_{i=1}^d \left\|\{V^{\phi^\delta}(\Lambda, t)\}^{1/\delta} \left(\nabla_i K\left(\frac{\lambda^0 - \lambda}{h}\right) - hK\left(\frac{\lambda^0 - \lambda}{h}\right) \frac{\nabla_i \ell}{\ell} (\lambda^0 - \Lambda)\right)\right\|_{\delta}}{Nh^{d+1} \ell(0)} + \frac{C}{N} \\ &\leq C \left(\frac{h^{d/\delta}}{Nh^{d+1}} + \frac{1}{N}\right) \,, \end{split}$$

by the usual change of variable and Taylor expansion. On the other hand, it follows from the equivalent of the variance derived in the previous step of this proof that

$$\sigma_N^2 \sim \frac{\phi^2\left(x(\lambda^0)\right)}{Nh^{d+2}\ell(0)} \int \left|a'\nabla K(l)\right|^2 dl$$

The two last estimates imply that Condition (I.55) holds under Condition (I.49). Therefore, $\sum_{i=1}^{N} V_i^N$ is asymptotically gaussian for any $a \in \mathbb{R}^d$, and the Cramer-Wold device concludes the proof.

7 Asymptotic properties of $\tilde{\beta}_N$

This section is dedicated to the proof of Proposition 4.3 and Theorem 4.3, characterizing the asymptotic behavior of $\tilde{\beta}_N$. In this section, we shall always work under the Assumptions of Proposition 4.3.

7.1 Preliminaries

Recall that

$$\tilde{\beta}_N := \frac{1}{\ell(0) N h^d} \sum_{i=1}^N \phi(Z_i) \, \hat{s}_N^{-i}(\Lambda_i, Z_i) \, K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \,, \tag{I.56}$$

where

$$\hat{s}_N^{-i}(\lambda, z) := \frac{\hat{\varphi_\lambda}^{-i}}{\hat{\varphi}^{-i,\delta}}(\lambda, z) + \frac{\nabla \ell}{\ell}(\lambda^0 - \lambda),$$

with $\hat{\varphi}^{-i,\delta} := \hat{\varphi}^{-i} + (\delta/3 - \hat{\varphi}^{-i}) \mathbf{1}_{|\hat{\varphi}^{-i}| \le \delta/3}$ a truncated version of $\hat{\varphi}^{-i}(\lambda, z)$ defined by

$$\hat{\varphi}^{-i}(\lambda, z) := \frac{h^{-d-n}}{N-1} \sum_{j=1, j \neq i}^{N} K\left(\frac{\lambda - \Lambda_j}{h}\right) H\left(\frac{z - Z_j}{h}\right) \quad \text{and} \quad \hat{\varphi_{\lambda}}^{-i} = \nabla_{\lambda} \hat{\varphi}^{-i}.$$

For every λ, z , we set

$$\bar{\varphi}(\lambda, z) := E[\hat{\varphi}^{-1}(\lambda, z)] = \int K(l)H(v)\varphi(\lambda - hl, z - hv) \, dl \, dv \,,$$

and its derivative is given by

$$\bar{\varphi}_{\lambda}(\lambda, z) = h^{-1} \int \nabla K(l) H(v) \varphi(\lambda - hl, z - hv) \, dl \, dv$$

Arguing as in the proof of Proposition 4.1, we next compute that

$$\bar{\varphi}(\lambda, z) - \varphi(\lambda, z) = \xi_K^p[\varphi](\lambda, z) h^p + \xi_H^q[\varphi](\lambda, z) h^q + o(h^{p \wedge q}).$$
(I.57)

Similarly, we get

$$\bar{\varphi}_{\lambda}(\lambda, z) - \varphi_{\lambda}(\lambda, z) = \xi_{K}^{p}[\varphi_{\lambda}](\lambda, z) h^{p} + \xi_{H}^{q}[\varphi_{\lambda}](\lambda, z) h^{q} + o(h^{p \wedge q}).$$
(I.58)

Remark 7.1 Since ϕ and K have compact support by Assumption S, it follows that, for sufficiently small h, the sum in (I.56) is restricted to pairs (Λ_i, Z_i) with values in $C_K \times C_{\phi}$ where $C_K \subset \mathcal{V}(\lambda^0)$ is defined in Assumption S, and C_{ϕ} is a compact subset of \mathbb{R}^n such that $\operatorname{Supp} \phi \subset C_{\phi}$. For any function ψ defined on $C_K \times C_{\phi}$, we set

$$||\psi||_{\infty} := \sup_{(\lambda,z)\in C_K\times C_\phi} |\psi(\lambda,z)|,$$

and, in the following, $||.||_r$ refers to the $\mathcal{L}_r(\Omega)$ -norm.

Remark 7.2 By Assumptions R2 and R3, since (λ, z) vary in a compact subset of $\mathbb{R}^d \times \mathbb{R}^n$, the remainder terms in (I.57) and (I.58) are uniformly bounded in (λ, z) . By the same argument, we also see that $\xi_K^p[\varphi]$, $\xi_H^q[\varphi]$, $\xi_K^p[\varphi_\lambda]$ and $\xi_H^q[\varphi_\lambda]$ are uniformly bounded so that :

$$\|\bar{\varphi} - \varphi\|_{\infty} = O(h^{p \wedge q}) \quad and \quad \|\bar{\varphi}_{\lambda} - \varphi_{\lambda}\|_{\infty} = O(h^{p \wedge q}).$$
 (I.59)

We now study further the tails of the estimators $\hat{\varphi}^{-i}$ and we obtain the following estimates.

Lemma 7.1 There exists α_1 and α_2 such that

$$\mathbb{P}[|\hat{\varphi}^{-i} - \bar{\varphi}|(\lambda, z) > t] \leq 2e^{-\frac{t^2}{\alpha_1 + \alpha_2 t}Nh^{d+n}}, \quad (\lambda, z) \in C_K \times C_\phi.$$
(I.60)

Furthermore, for any t > 0, there exists $C_t > 0$ and $c_t > 0$ satisfying

$$\mathbb{P}[\sup_{i \le N} \|\hat{\varphi}^{-i} - \bar{\varphi}\|_{\infty} > t] \le C_t N^3 e^{-c_t N h^{d+n}}.$$
(I.61)

Finally, for any integer $r \ge 1$, we have

$$\left\| \sup_{1 \le i \le N} \left\| \hat{\varphi}^{-i} - \bar{\varphi} \right\|_{\infty} \right\|_{2r} = O\left(\frac{\ln(N)}{\sqrt{Nh^{d+n}}} \right).$$
(I.62)

Proof. Observe first that there exists α_1 and α_2 such that, for any $(\lambda, z) \in C_K \times C_{\phi}$, the random variables $K[(\lambda - \Lambda^i)/h]H[(z - z^i)/h]$ are bounded by $3\alpha_2/2$ and, by the usual change of variable, their variance are bounded from above by $\alpha_1 h^{d+n}/2$. Therefore (I.60) follows directly from the Bernstein inequality.

We now turn to the proof of the second estimate and first observe that

$$\mathbb{P}[\sup_{i \le N} \|\hat{\varphi}^{-i} - \bar{\varphi}\|_{\infty} > t] \le N \mathbb{P}[\|\hat{\varphi} - \bar{\varphi}\|_{\infty} > t], \qquad (I.63)$$

where, for ease of notation, we introduce $\hat{\varphi} := \hat{\varphi}^{-1}$. Applying the Liebscher's strategy, see [74], we recover the compact set $C_K \times C_{\phi}$ by $C_0(R_{N,h})^{-d-n}$ balls $B_j := B((\lambda_j, z_j), R_{N,h})$, with C_0 a constant chosen large enough. On each ball B_j , we have

$$\sup_{B_{j}} |\hat{\varphi} - \bar{\varphi}| \leq |\hat{\varphi} - \bar{\varphi}|(\lambda_{j}, z_{j}) + \sup_{(\lambda, z) \in B_{j}} |\hat{\varphi}(\lambda, z) - \hat{\varphi}(\lambda_{j}, z_{j})|$$

$$+ \sup_{(\lambda, z) \in B_{j}} |\bar{\varphi}(\lambda, z) - \bar{\varphi}(\lambda_{j}, z_{j})|$$
(I.64)

According to Assumption KH, the kernel functions K and H are lipschitz and compactly supported. Therefore, there exists M > 0 such that

$$\sup_{(\lambda,z)\in B_j} |\hat{\varphi}(\lambda,z) - \hat{\varphi}(\lambda_j,z_j)| \leq C \frac{R_{N,h}}{h} \hat{\psi}(\lambda_j,z_j),$$

where $\hat{\psi}$ is the classical histogram Kernel estimator of the density φ defined by

$$\hat{\psi}(\lambda,z) := \frac{1}{4M^2 N h^{d+n}} \sum_{i=1}^N \mathbf{1}_{|\Lambda_i - \lambda| \le M h} \mathbf{1}_{|Z_i - z| \le M h}.$$

Introducing the notation $\bar{\psi} := \mathbb{E}[\hat{\psi}]$ and choosing $R_{N,h}$ such that $R_{N,h} = o(h)$, we then deduce from (I.64) that

$$\sup_{B_j} |\hat{\varphi} - \bar{\varphi}| \leq |\hat{\varphi} - \bar{\varphi}| (\lambda_j, z_j) + |\hat{\psi} - \bar{\psi}| (\lambda_j, z_j) + 2C \frac{R_{N,h}}{h} \bar{\psi}(\lambda_j, z_j) + C \frac{R_{N,h}}{h} |\hat{\psi}(\lambda_j, z_j)| \leq C \frac{R_{$$

Summing up over all the balls B_j , we get

$$\mathbb{P}[\|\hat{\varphi} - \bar{\varphi}\|_{\infty} > t] \leq C_0 R_{N,h}^{-(d+n)} \left(\mathbb{P}[|\hat{\varphi} - \bar{\varphi}|(\lambda_j, z_j) > t/3] + \mathbb{P}[|\hat{\psi} - \bar{\psi}|(\lambda_j, z_j) > t/3] \right) \\
+ C_0 R_{N,h}^{-(d+n)} \mathbb{P}[2Ch^{-1}R_{N,h} |\bar{\psi}|(\lambda_j, z_j) > t/3].$$

Therefore, applying estimate (I.60) to both kernel estimators $\hat{\varphi}$ and $\hat{\psi}$, we deduce the existence of γ_1 and γ_2 satisfying

$$\mathbb{P}[\|\hat{\varphi} - \bar{\varphi}\|_{\infty} > t] \le CR_{N,h}^{-(d+n)} \left(e^{-\frac{t^2}{\gamma_1 + \gamma_2 t}Nh^{d+n}} + \mathbb{P}\left[2C\frac{R_{N,h}}{h} |\bar{\psi}|(\lambda_j, z_j) > t/3 \right] \right).$$
(I.65)

But $\bar{\psi}$ is bounded so that for any given t the last term on the right hand side equals 0 for h small enough. Since $Nh^{d+n} \to \infty$ according to (I.28), choosing $R_{N,h} = h^2$, we deduce (I.61) from (I.63).

We now turn to the moment inequalities and introduce the notation

$$Y_N := \frac{\sqrt{Nh^{d+n}}}{\ln(N)} \sup_{i \le N} \|\hat{\varphi} - \bar{\varphi}\|_{\infty},$$

so that we simply need to prove $||Y_N||_{2r} < \infty$ for all integer $r \ge 1$. Fix $r \in \mathbb{N}^*$ and observe that

$$\mathbb{E}\left[Y^{2r}\right] = \int_0^\infty 2rs^{2r-1} \mathbb{P}[Y_N > s] ds \le C_a + \int_a^\infty 2rs^{2r-1} \mathbb{P}[Y_N > s] ds, \quad (I.66)$$

for any a>0. We now fix s large enough and take $R_{N,h} = h \ln(N) / \sqrt{Nh^{d+n}}$ in (I.65) and (I.63), so that we get, for N large enough, the existence of δ_1 and δ_2 satisfying

$$\mathbb{P}[Y_N > s] \le CN\left(\frac{\sqrt{Nh^{d+n}}}{hln(N)}\right)^{d+n} e^{-\frac{s\ln(N)^2}{\delta_1 + \delta_2 s\ln(N)/\sqrt{Nh^{d+n}}}}$$

Since $ln(N)/\sqrt{Nh^{d+n}} \to 0$ and $h \to 0$, we deduce that for N large enough, we have

$$\mathbb{P}[Y_N > s] \le CN^{d+n} e^{-\frac{s \ln(N)^2}{\delta_1 + \delta_2 s \ln(N)/\sqrt{Nh^{d+n}}}} \le Ce^{(d+n)\ln(N) - s(\ln N)^{3/2}} \le Ce^{-s}.$$

Plugging this estimate into (I.66) completes the proof.

Since ∇K has bounded variation, the exact same reasoning can apply to the estimators $\hat{\varphi}_{\lambda}^{-i}$ and we similarly derive

$$\left\| \sup_{1 \le i \le N} \left\| \hat{\varphi_{\lambda}}^{-i} - \bar{\varphi_{\lambda}} \right\|_{\infty} \right\|_{2r} = O\left(\frac{\ln N}{h\sqrt{Nh^{d+n}}} \right), \quad r \in \mathbb{N}^*.$$
(I.67)

The estimates of the previous lemma also allow to control the error due to the truncation of $\hat{\varphi}^{-i}$. Indeed, since the function φ admits δ as a lower bound according to Assumption S, it follows from (I.59) that that $\bar{\varphi} > 2\delta/3$ for h small enough, and (I.60) leads to

$$\mathbb{P}[|\hat{\varphi}^{-1}(\lambda, z)| < \delta/3] \leq \mathbb{P}[|\hat{\varphi}^{-1} - \bar{\varphi}|(\lambda, z) > \delta/3] \leq 2 e^{-CNh^{d+n}}.$$
(I.68)

Introducing $\bar{\varphi}^{\delta} := \mathbb{E}\left[\hat{\varphi}^{-1,\delta}\right]$, we derive

$$\left\|\bar{\varphi}^{\delta} - \bar{\varphi}\right\|_{\infty} \leq \frac{\delta}{3} \sup_{C_K \times C_{\phi}} \mathbb{P}[|\hat{\varphi}^{-1}|(\lambda, z) < \delta/3] \leq \frac{2\delta}{3} e^{-CNh^{d+n}}, \quad (I.69)$$

and combining (I.28) and (I.59), we deduce

$$\left\|\bar{\varphi}^{\delta} - \varphi\right\|_{\infty} = O\left(h^{p\wedge q}\right) \,. \tag{I.70}$$

Similarly, applying (I.61), we get

$$\begin{aligned} \left\| \sup_{1 \le i \le N} \left\| \hat{\varphi}^{-i,\delta} - \hat{\varphi}^{-i} \right\|_{\infty} \right\|_{2r} &\le \delta \mathbb{P} \left[\sup_{i \le N} \left\| \hat{\varphi}^{-i} - \bar{\varphi} \right\|_{\infty} > \delta/3 \right] \\ &\le C \delta N^3 e^{-CNh^{d+n}}, \quad r \in \mathbb{N} \,. \end{aligned}$$
(I.71)

Observe also that (I.69) and (I.71) combined with (I.28) allows to derive

$$\left\| \sup_{1 \le i \le N} \left\| \hat{\varphi}^{-i,\delta} - \bar{\varphi}^{\delta} \right\|_{\infty} \right\|_{2r} = O\left(\frac{\ln N}{\sqrt{Nh^{d+n}}} \right), \quad \text{for any } r \in \mathbb{N}^*.$$
 (I.72)

Finally, since (λ, z) vary in a compact subset, Assumptions R2, R3 and S imply that

$$\|\varphi\|_{\infty} + \|\varphi_{\lambda}\|_{\infty} + \|1/\varphi\|_{\infty} < \infty.$$
(I.73)

It then follows from equation (I.59), (I.70) and the truncation procedure that

$$\|\bar{\varphi}\|_{\infty} + \left\|\bar{\varphi}^{\delta}\right\|_{\infty} + \|\bar{\varphi}_{\lambda}\|_{\infty} + \|1/\bar{\varphi}\|_{\infty} + \left\|1/\bar{\varphi}^{\delta}\right\|_{\infty} + \sup_{1 \le i \le N} \left\|1/\hat{\varphi}^{-i,\delta}\right\|_{\infty} < \infty.$$
(I.74)

7.2 A suitable decomposition

For any $N \in \mathbb{N}$ and $i \leq N$, we define the following functions $t_{i,N}^1, \ldots, t_{i,N}^9$ on $\mathbb{R}^d \times \mathbb{R}^n \times \Omega$:

$$\begin{split} t^{1}_{i,N} &:= s \,, \ t^{2}_{i,N} &:= \frac{\bar{\varphi_{\lambda}} - \varphi_{\lambda}}{\varphi} \,, \ t^{3}_{i,N} &:= \frac{(\varphi - \bar{\varphi^{\delta}}) \,\varphi_{\lambda}}{\varphi^{2}} \,, \ t^{4}_{i,N} &:= \frac{(\varphi - \bar{\varphi^{\delta}}) \,(\bar{\varphi_{\lambda}} \,\varphi - \bar{\varphi^{\delta}} \,\varphi_{\lambda})}{\varphi^{2} \,\bar{\varphi^{\delta}}} \,, \\ t^{5}_{i,N} &:= \frac{\hat{\varphi_{\lambda}}^{-i} - \bar{\varphi_{\lambda}}}{\varphi} \,, \ t^{6}_{i,N} &:= \frac{(\bar{\varphi^{\delta}} - \hat{\varphi^{-i,\delta}}) \,\bar{\varphi_{\lambda}}}{(\bar{\varphi^{\delta}})^{2}} \,, \ t^{7}_{i,N} &:= \frac{(\hat{\varphi_{\lambda}}^{-i} - \bar{\varphi_{\lambda}}) \,(\varphi^{\delta} - \bar{\varphi^{\delta}})}{\varphi^{\delta} \,\bar{\varphi^{\delta}}} \,, \\ t^{8}_{i,N} &:= \frac{(\bar{\varphi^{\delta}} - \hat{\varphi^{-i,\delta}}) (\hat{\varphi_{\lambda}}^{-i} - \bar{\varphi_{\lambda}})}{\hat{\varphi^{-i,\delta}} \,\bar{\varphi^{\delta}}} \, \text{ and } \ t^{9}_{i,N} &:= \frac{(\bar{\varphi^{\delta}} - \hat{\varphi^{-i,\delta}})^{2} \bar{\varphi_{\lambda}}}{\hat{\varphi^{-i,\delta}} (\bar{\varphi^{\delta}})^{2}} \,, \end{split}$$

so that
$$\hat{s}_N^{-i}(\Lambda_i, Z_i) = \sum_{j=1}^9 t_{i,N}^j(\Lambda_i, Z_i).$$

This implies the following decomposition of the estimator $\tilde{\beta}_N$:

$$\tilde{\beta}_{N} = \sum_{j=1}^{9} T_{N}^{j}, \text{ where } T_{N}^{j} := \frac{1}{\ell(0) N h^{d}} \sum_{i=1}^{N} \phi(Z_{i}) t_{i,N}^{j}(\Lambda_{i}, Z_{i}) K\left(\frac{\lambda^{0} - \Lambda_{i}}{h}\right), \quad (I.75)$$

for every $j = 1, \ldots, 9$. By (I.73) and (I.74), we observe that

$$\left\| t_{i,N}^{j} \right\|_{\infty} < \infty$$
, for all $j = 1, \dots, 4$.

Lemma 7.2 For any j = 1, ..., 4, we have $\mathbb{E}\left[T_N^j\right] = O\left(\left\|t_{1,N}^j\right\|_{\infty}\right)$.

Proof. The result is derived from the following inequality:

$$\begin{aligned} \left| \mathbb{E}[T_N^j] \right| &\leq \frac{1}{\ell(0) h^d} \left| \mathbb{E}\left[\phi(Z_1) t_{1,N}^j(\Lambda_1, Z_1) K\left(\frac{\lambda^0 - \Lambda_1}{h}\right) \right] \right| \\ &\leq \frac{1}{\ell(0)} \left| \int \phi(z) t_{1,N}^j(\lambda^0 - hl, z) K(l) dl dv \right| \\ &\leq C ||t_{1,N}^j||_{\infty} . \end{aligned}$$

Lemma 7.3 For every j = 1, ..., 4, $\operatorname{Var}(T_N^j) = O\left(N^{-1}h^{-d} \| t_{1,N}^j \|_{\infty}^2\right)$.

Proof. For any j = 1..., 4, the N random variables $T_N^j(\Lambda_i, Z_i)$ are independent and

$$\begin{aligned} \mathbb{V}\mathrm{ar}[T_{N}^{j}] &= \frac{1}{\ell(0)^{2} N h^{2d}} \mathbb{V}\mathrm{ar}\left[\phi(Z_{1}) \ t_{1,N}^{j}(\Lambda_{1}, Z_{1}) \ K\left(\frac{\lambda^{0} - \Lambda_{1}}{h}\right)\right] \\ &\leq \frac{1}{\ell(0)^{2} N h^{2d}} \mathbb{E}\left[\phi^{2}(Z_{1}) \ t_{1,N}^{j}(\Lambda_{1}, Z_{1})^{2} \ K^{2}\left(\frac{\lambda^{0} - \Lambda_{1}}{h}\right)\right] \\ &\leq \frac{\|t_{1,N}^{j}\|_{\infty}^{2}}{\ell(0)^{2} N h^{d}} \int \phi^{2}(z) \ K^{2}(l) \ dl \ dv \,.\end{aligned}$$

The analysis of T_N^j , for j > 4, requires more effort because of the dependence between the random variables $t_{i,N}^j(\Lambda_i, Z_i)$.

Lemma 7.4 $\mathbb{E}[T_N^5] = 0$ and $\mathbb{Var}(T_N^5) \sim \tilde{\Sigma}/(Nh^{d+2})$ where $\tilde{\Sigma}$ is defined in Proposition 4.3.

Proof. We introduce for any i = 1, ..., N and j = 1, ..., N:

$$\mathcal{T}_{ij} := \frac{\phi(Z_i)}{\varphi(\Lambda_i, Z_i)} K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \left\{ \nabla_\lambda K\left(\frac{\Lambda_i - \Lambda_j}{h}\right) H\left(\frac{Z_i - Z_j}{h}\right) - h^{d+n+1} \bar{\varphi}_\lambda(\Lambda_i, Z_i) \right\} ,$$

so that T_N^5 can be re-written in

$$T_N^5 = \frac{h^{-2d-n-1}}{\ell(0) N(N-1)} \sum_{i < j} \left(\mathcal{T}_{ij} + \mathcal{T}_{ji} \right) \; .$$

By definition, for any i = 1, ..., N and j = 1, ..., N with $i \neq j$, we have

$$\bar{\varphi_{\lambda}}(\Lambda_i, Z_i) = \frac{1}{h^{d+n+1}} \mathbb{E}\left[\nabla_{\lambda} K\left(\frac{\Lambda_i - \Lambda_j}{h}\right) H\left(\frac{Z_i - Z_j}{h}\right) \mid \Lambda_i, Z_i\right]$$

Therefore, $\mathbb{E}[\mathcal{T}_{ij}] = 0$ whenever $i \neq j$, leading to $\mathbb{E}[T_N^5] = 0$. Since the \mathcal{T}_{ij} are not independent, the computation of the variance requires to decompose T_N^5 into

$$T_N^5 = T_N^{5,1} + T_N^{5,2}, (I.76)$$

where

$$T_N^{5,1} := \frac{h^{-2d-n-1}}{\ell(0) N(N-1)} \sum_{i < j} \left(\mathcal{T}_{ij} + \mathcal{T}_{ji} - b(\Lambda_i, Z_i) - b(\Lambda_j, Z_j) \right) ,$$

$$T_N^{5,2} := \frac{h^{-2d-n-1}}{\ell(0) N(N-1)} \sum_{i < j} \left(b(\Lambda_i, Z_i) + b(\Lambda_j, Z_j) \right) .$$

and $b(\lambda, z) := E[\mathcal{T}_{12}|\Lambda_2 = \lambda, Z_2 = z].$

1. Let first study the term $T_N^{5,1}$. Setting $\Upsilon_{ij} := \mathcal{T}_{ij} + \mathcal{T}_{ji} - b(\Lambda_i, Z_i) - b(\Lambda_j, Z_j)$, we derive the key property :

$$\mathbb{E}[\Upsilon_{ij}|\Lambda_i, Z_i] = \mathbb{E}[\Upsilon_{ij}|\Lambda_j, Z_j] = 0.$$
(I.77)

Therefore $T_N^{5,1}$ has zero mean and we derive :

$$\mathbb{V}\mathrm{ar}[T_N^{5,1}] = \frac{h^{-4d-2n-2}}{\ell(0)^2 N^2 (N-1)^2} \sum_{i < j} \mathbb{E}[\Upsilon_{ij} \Upsilon'_{ij}] = \frac{h^{-4d-2n-2}}{2\ell(0)^2 N(N-1)} \mathbb{E}[\Upsilon_{12} \Upsilon'_{12}].$$

By (I.77), we compute :

$$\mathbb{E}[\Upsilon_{12}\,\Upsilon'_{12}] = 2\,\mathbb{E}[\mathcal{T}_{12}\mathcal{T}'_{12}] + 2\,\mathbb{E}[\mathcal{T}_{12}\mathcal{T}'_{21}] - 2\mathbb{E}[b^2(\Lambda_1, Z_1)].$$

We next estimate that $|\mathbb{E}[\mathcal{T}_{12}\mathcal{T}'_{12}]|$ is dominated by

$$\mathbb{E}\left[\frac{\phi^2(Z_1)}{\varphi^2(\Lambda_1, Z_1)} K^2\left(\frac{\lambda^0 - \Lambda_1}{h}\right) |\nabla_\lambda K|^2 \left(\frac{\Lambda_1 - \Lambda_2}{h}\right) H^2\left(\frac{Z_1 - Z_2}{h}\right)\right] + h^{2d+n} \int \phi^2(z) K^2(l_1) |\nabla_\lambda K|^2(l_2) H^2(v) \frac{\varphi(\lambda^0 - hl_1 - hl_2, z - hv)}{\varphi(\lambda^0 - hl_1, z)} dl_1 dl_2 dz dv ,$$

by the usual change of variables. Clearly, the first term on the right hand-side is of order $O(h^{2d+n})$, while the second one is a $O(h^{3d+2n+2})$ by (I.74). Similarly, we have $\mathbb{E}[\mathcal{T}_{12}\mathcal{T}'_{21}] = O(h^{2d+n})$. Moreover, $\mathbb{E}[b^2(\Lambda_1, Z_1)] = O(N^{-2}h^{-d-2})$. We deduce that

$$\operatorname{Var}(T_N^{5,1}) = O\left(\frac{1}{N^2 h^{2d+n+2}}\right) = o\left(\frac{1}{N h^{2+d}}\right),$$
 (I.78)

using the relations between N and h given by (I.28).

2. We next rewrite $T_N^{5,2}$ as

$$T_N^{5,2} = \frac{h^{-2d-n-1}}{\ell(0)N} \sum_i b(\Lambda_i, Z_i).$$

By the usual change of variables,

$$b(\lambda, z) = h^{d+n} \int \phi(z + hv) K\left(\frac{\lambda^0 - \lambda}{h} - l\right) \nabla K(l) H(v) \, dl \, dv$$
$$-h^{n+1} \int \phi(z) \, \bar{\varphi_{\lambda}}(\lambda^0 - hl, z) K(l) \, dl.$$

By direct calculation, it is easily checked that the second term is negligible. Then, by the usual change of variables, it follows that

$$\mathbb{E}[b(\Lambda_i, Z_i)b(\Lambda_i, Z_i)'] \sim h^{3d+2n} \int \left\{ \int \phi(z+hv)K(l_2-l_1)\nabla K(l_1)H(v) \, dl_1 \, dv \right\}^{\otimes} \varphi(\lambda^0-hl_2, z) \, dl_2 \, dz \, .$$

By Assumptions S and R3, we deduce from the dominated convergence theorem together with the fact that $\mathbb{E}[b(\Lambda_i, Z_i)] = 0$ that

$$\mathbb{V}\mathrm{ar}[T_N^{5,2}] \sim \frac{1}{Nh^{d+2}} \int \phi^2(z) \left\{ \int K(l_2 - l_1) \nabla K(l_1) \, dl_1 \right\}^{\otimes} \varphi(\lambda^0, z) \, dl_2 \, dz \,.$$
(I.79)

The proof is completed by collecting the estimates (I.78) and (I.79) into (I.76). \Box

Lemma 7.5 $\mathbb{E}[T_N^6] = o(h^{p \wedge q})$ and $\mathbb{Var}(T_N^6) = o(N^{-1}h^{-d-2}).$

Proof. We decompose $t_{i,N}^6$ into the sum of

$$t_{i,N}^{6,1} := \frac{(\bar{\varphi} - \hat{\varphi}^{-i})\,\bar{\varphi_{\lambda}}}{(\bar{\varphi}^{\delta})^2}, \quad t_{i,N}^{6,2} := \frac{(\hat{\varphi}^{-i} - \hat{\varphi}^{-i,\delta})\,\bar{\varphi_{\lambda}}}{(\bar{\varphi}^{\delta})^2} \quad \text{and} \quad t_{i,N}^{6,3} := \frac{(\bar{\varphi}^{\delta} - \bar{\varphi})\,\bar{\varphi_{\lambda}}}{(\bar{\varphi}^{\delta})^2}$$

and we study the corresponding $T_N^{6,1}$, $T_N^{6,2}$ and $T_N^{6,3}$ separately.

1. It can be checked easily that $T_N^{6,1}$ can be dealt with as T_N^5 . By the same calculation, we get $\mathbb{E}[T_N^{6,1}] = 0$ and

$$\mathbb{V}\mathrm{ar}(T_N^{6,1}) \sim \frac{h^{-4d-2n}}{\ell(0)^2 N^2} \sum_i \mathbb{V}\mathrm{ar}(\tilde{b}(\Lambda_i, Z_i))$$

where $\tilde{b}(\lambda, z)$ is given by :

$$E\left[\frac{\phi(Z_i)\varphi_{\lambda}(\Lambda_i, Z_i)}{\varphi(\Lambda_i, Z_i)^2} K\left(\frac{\lambda^0 - \Lambda_i}{h}\right) \left\{K\left(\frac{\Lambda_i - \lambda}{h}\right) H\left(\frac{Z_i - z}{h}\right) - h^{d+n}\bar{\varphi}(\Lambda_i, Z_i)\right\}\right]$$

The variables $\tilde{b}(\Lambda_i, Z_i)$ have also zero mean and, as in the proof of Lemma 7.4, the usual change of variables implies that

$$h^{-3d-2n} \operatorname{\mathbb{V}ar}(\tilde{b}(\Lambda_i, Z_i)) \sim \int [G_6(l_2, z)]^{\otimes} \varphi(\lambda^0 - hl_2, z) \, dl_2 \, dz \,,$$

with $G_6(l_2, z) := \int \phi(z + hv) \frac{\varphi_\lambda}{\varphi} (\lambda^0 + hl_1 - hl_2, z + hv) K(l_2 - l_1) K(l_1) H(v) \, dl_1 \, dv$

By the continuity and the uniform boundedness of ϕ and $\varphi_{\lambda}/\varphi$ implied by Assumptions S and R3, we derive

$$\operatorname{Var}(T_n^{6,1}) = O\left(\frac{1}{Nh^d}\right) = o\left(\frac{1}{Nh^{d+2}}\right).$$

2. We now turn to $T_N^{6,2}$ and compute

$$|T_N^{6,2}| \leq C \sup_{i \leq N} \left\| \hat{\varphi}^{-i,\delta} - \hat{\varphi}^{-i} \right\|_{\infty} \left(\frac{1}{Nh^d} \sum_{i=1}^N \left| \phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h} \right) \right| \right) \,.$$

Therefore, we deduce from Cauchy-Schwarz inequality that

$$\left| \mathbb{E} \left[T_N^{6,2} \right] \right| \leq C \left\| \sup_{i \leq N} \left\| \hat{\varphi}^{-i,\delta} - \hat{\varphi}^{-i} \right\|_{\infty} \right\|_2 \mathbb{E} \left[\left(\frac{1}{Nh^d} \sum_{i=1}^N \left| \phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h} \right) \right| \right)^2 \right]^{1/2},$$

and (I.28) combined with (I.71) lead to $\mathbb{E}\left[T_N^{6,2}\right] = o(h^{p \wedge q})$. Similarly, we get

$$Var(T_N^{6,2}) \leq C \left\| \sup_{i \leq N} \left\| \hat{\varphi}^{-i,\delta} - \hat{\varphi}^{-i} \right\|_{\infty} \right\|_{4} \mathbb{E} \left[\left(\frac{1}{Nh^d} \sum_{i=1}^{N} \left| \phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h} \right) \right| \right)^4 \right]^{1/4},$$

which leads to $\mathbb{V}ar(T_n^{6,2}) = o(N^{-1}h^{-d-2}).$

3. We finally observe that $T_N^{6,3}$ is treated similarly thanks to (I.69).

Lemma 7.6 $\mathbb{E}[T_N^7] = 0$ and $\mathbb{Var}(T_N^7) = o(N^{-1}h^{-d-2}).$

Proof. Observe that

$$t_N^7(\lambda, z) = t_N^5(\lambda, z)\psi(\lambda, z)$$
 where $\psi := \frac{\varphi - \bar{\varphi}^{\delta}}{\bar{\varphi}^{\delta}}$

Following the lines of the proof of Lemma 7.4, we see that $\mathbb{E}[T_N^7] = 0$, and we estimate

$$Nh^{d+2} \mathbb{V}ar(T_N^7) \sim \int [G_7(u,z)]^{\otimes} \varphi(\lambda^0 - hu,z) \, du \, dz \,,$$

with $G_7(u,z) := \int \phi(z+hv) \psi(\lambda^0 + hl - hu, z+hv) K(u-l) \nabla K(l) H(v) \, dl \, dv$

By (I.70) and (I.74) it follows that $\|\psi\|_{\infty} = O(h^{p \wedge q})$ and, since φ and ϕ are uniformly bounded, we deduce that

$$\mathbb{V}\mathrm{ar}(T_N^7) = O\left(\frac{h^{p\wedge q}}{Nh^{d+2}}\right) = o\left(\frac{1}{Nh^{d+2}}\right).$$

$$\Box$$

$$T_N^8] \sim \frac{h^{-d-n-1}}{\ell(0)N} \left(\int \phi\right) \left(\int H^2\right) \int K(l_1 - l_2)K(l_2)\nabla K(l_2)dl_1dl_2$$

Lemma 7.7 $\mathbb{E}\left[T_N^8\right] \sim \frac{h^{-a-n-1}}{\ell(0)N} \left(\int \phi\right) \left(\int H^2\right) \int K(l_1 - l_2)K(l_2)\nabla K(l_2)dl_1dl_2$ and $\mathbb{Var}(T_N^8) = o(N^{-1}h^{-d-2}).$

Proof. We split the proof it two steps.

1. We first estimate $\mathbb{E}\left[T_N^8\right]$. We rewrite $t_N^8(\lambda, z)$ as $t_N^{8,1}(\lambda, z) + t_N^{8,2}(\lambda, z) + t_N^{8,3}(\lambda, z)$ with

$$\begin{split} t^{8,1}_{i,N} &= \frac{(\bar{\varphi} - \hat{\varphi}^{-i})(\hat{\varphi}_{\lambda}^{-i} - \bar{\varphi}_{\lambda})}{\varphi^2}, \\ t^{8,2}_{i,N} &= \frac{(\bar{\varphi}^{\delta} - \bar{\varphi})(\hat{\varphi}_{\lambda}^{-i} - \bar{\varphi}_{\lambda})}{\varphi^2} + \frac{(\hat{\varphi}^{-i} - \hat{\varphi}^{-i,\delta})(\hat{\varphi}_{\lambda}^{-i} - \bar{\varphi}_{\lambda})}{\varphi^2}, \\ t^{8,3}_{i,N} &= \frac{(\bar{\varphi}^{\delta} - \hat{\varphi}^{-i,\delta})^2(\hat{\varphi}_{\lambda}^{-i} - \bar{\varphi}_{\lambda})}{\hat{\varphi}^{-i,\delta}(\bar{\varphi}^{\delta})^2} + \frac{(\bar{\varphi}^{\delta} - \hat{\varphi}^{-i,\delta})(\hat{\varphi}_{\lambda}^{-i} - \bar{\varphi}_{\lambda})(\varphi^2 - (\bar{\varphi}^{\delta})^2)}{\varphi^2(\bar{\varphi}^{\delta})^2}. \end{split}$$

Then $T_N^8 = T_N^{8,1} + T_N^{8,2} + T_N^{8,3}$, where

$$T_N^{8,k} := \frac{1}{\ell(0) N h^d} \sum_{i=1}^N \phi(Z_i) t_{i,N}^{8,k}(\Lambda_i, Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h}\right), \quad \text{for} \quad k = 1, 2, 3.$$

We now introduce

$$U_{ij} := \nabla_{\lambda} K\left(\frac{\Lambda_{i} - \Lambda_{j}}{h}\right) H\left(\frac{Z_{i} - Z_{j}}{h}\right) - \mathbb{E}\left[\nabla_{\lambda} K\left(\frac{\Lambda_{i} - \Lambda_{j}}{h}\right) H\left(\frac{Z_{i} - Z_{j}}{h}\right) |\Lambda_{i}, Z_{i}\right],$$

$$V_{ij} := K\left(\frac{\Lambda_{i} - \Lambda_{j}}{h}\right) H\left(\frac{Z_{i} - Z_{j}}{h}\right) - \mathbb{E}\left[K\left(\frac{\Lambda_{i} - \Lambda_{j}}{h}\right) H\left(\frac{Z_{i} - Z_{j}}{h}\right) |\Lambda_{i}, Z_{i}\right],$$

so that

$$\mathbb{E}\left[U_{ij}V_{ik}|\Lambda_i, Z_i\right] = \mathbb{E}\left[U_{ij}|\Lambda_i, Z_i\right] \mathbb{E}\left[V_{ik}|\Lambda_i, Z_i\right] = 0 \quad \text{whenever} \quad j \neq k.$$

Using this property, we compute directly that

$$\mathbb{E}\left[t_{N}^{8,1}(\Lambda_{1},Z_{1})|\Lambda_{1},Z_{1}\right] = \frac{h^{-2d-2n-1}}{(N-1)^{2}\varphi^{2}(\Lambda_{1},Z_{1})}\mathbb{E}\left[\sum_{j\neq 1}\sum_{k\neq 1}U_{1j}V_{1k}|\Lambda_{1},Z_{1}\right]$$
$$= \frac{h^{-2d-2n-1}}{(N-1)\varphi^{2}(\Lambda_{1},Z_{1})}\mathbb{E}\left[U_{12}V_{12}|\Lambda_{1},Z_{1}\right].$$

Since the expectation of ${\cal T}_N^{8,1}$ is given by :

$$\mathbb{E}\left[T_N^{8,1}\right] = \frac{h^{-d}}{\ell(0)} \mathbb{E}\left[\phi(Z_1) K\left(\frac{\lambda^0 - \Lambda_1}{h}\right) \mathbb{E}\left[t_{1,N}^{8,1}(\Lambda_1, Z_1) | \Lambda_1, Z_1\right]\right],$$

we derive by the usual change of variables,

$$\ell(0)Nh^{d+n+1} \mathbb{E}\left[T_N^{8,1}\right] \sim \int G_8(l_2, z)\varphi(\lambda^0 - hl_2, z) \, dl_2 \, dz \,,$$

with $G_8(l_2, z) := \int \frac{\phi(z+hv)}{\varphi(\lambda^0 + hl_1 - hl_2, z+hv)} K(l_2 - l_1)K(l_1)\nabla K(l_1)H^2(v) \, dl_1 \, dv \,.$

Finally, by the continuity and the uniform boundedness of φ and ϕ , we derive :

$$\mathbb{E}\left[T_N^{8,1}\right] \sim \frac{h^{-d-n-1}}{\ell(0)N} \int \phi(z) K(l_2 - l_1) K(l_1) \nabla K(l_1) H^2(v) \, dl_1 \, dv \, dl_2 \, dz \,.$$
(I.80)

Furthermore, by Cauchy-Schwarz inequality and (I.28), we have

$$\begin{aligned} \left| \mathbb{E} \left[T_N^{8,k} \right] \right| &\leq \left\| \sup_{i \leq N} \left\| t_{i,N}^{8,k} \right\|_{\infty} \right\|_2 \mathbb{E} \left[\left(\frac{1}{Nh^d} \sum_{i=1}^N \left| \phi(Z_i) K\left(\frac{\lambda^0 - \Lambda_i}{h} \right) \right| \right)^2 \right]^{1/2} \quad (I.81) \\ &\leq C \left\| \sup_{i \leq N} \left\| t_{i,N}^{8,k} \right\|_{\infty} \right\|_2, \quad k = 2, 3. \end{aligned}$$

Finally, combining relations (I.59)-(I.74), Cauchy-Schwarz inequality and (I.28), we get

$$\left\|\sup_{i\leq N} \left\|t_{i,N}^{8,2}\right\|_{\infty}\right\|_{2} = o\left(\frac{1}{Nh^{d+n+1}}\right),$$

and

$$\left\|\sup_{i\leq N} \left\|t_{i,N}^{8,3}\right\|_{\infty}\right\|_{2} = O\left(\frac{(lnN)^{3}}{Nh^{d+n+1}\sqrt{Nh^{d+n}}}\right) = o\left(\frac{1}{Nh^{d+n+1}}\right)$$

Therefore (I.80) and (I.81) lead to the expected equivalent for $\mathbb{E}[T_N^8]$. 2. We now study the variance of T_N^8 . We first notice that the Cauchy-Schwarz inequality and (I.28) lead to

$$Var\left[T_{N}^{8}\right] \leq C \left\|\sup_{i \leq N} \left\|t_{i,N}^{8}\right\|_{\infty}^{4}\right\|_{4}^{2}$$

But, using again Cauchy-Schwarz inequality and relations (I.28), (I.59), (I.74) and (I.72), we deduce that

$$\operatorname{Var}\left(T_{N}^{8}\right) = O\left(\frac{\ln^{4}N}{N^{2}h^{2d+2n+2}}\right) = O\left(\frac{1}{Nh^{d+2}}\right).$$

Lemma 7.8 $\mathbb{E}[T_N^9] = O(N^{-1}h^{-d-n})$ and $\mathbb{Var}(T_N^9) = O(N^{-1}h^{-d-2})$

Proof. It can be easily checked that T_N^9 can be dealt as T_N^8 and, following the lines of the proof of Lemma 7.7, we obtain the announced result.

7.3 Asymptotic bias and variance

This section is devoted to the proof of Proposition 4.3 characterizing the asymptotic bias and variance of the double Kernel based estimator $\tilde{\beta}_N$.

Proof of Proposition 4.3. We split the proof in two steps.

1. We first derive the expectation of $\tilde{\beta}_N$.

Notice that $T_N^1 = \bar{\beta}_N$ as defined in (I.10) which satisfies

$$\mathbb{E}\left[\bar{\beta}_{N}\right] = \frac{1}{\ell(0)} \int \phi(z) K(l) s(\lambda^{0} - hl, z) \varphi(\lambda^{0} - hl, z) \, dt \, dz$$

The regularity of function $s\varphi$ given by assumption R1 enables us to derive

$$\mathbb{E}[T_N^1] - \beta \sim \frac{h^p}{\ell(0)} \int \xi_K^p \left[\ell f_\lambda \right] (\lambda^0, z) \phi(z) \, dz \,. \tag{I.83}$$

Using remark 7.2, we deduce from (I.58) that we have

$$\mathbb{E}[T_N^2] = \frac{h^p}{\ell(0)} \int \xi_K^p \left[\varphi_\lambda\right] (\lambda^0, z) \phi(z) \, dz \, + \, \frac{h^q}{\ell(0)} \int \xi_H^q \left[\varphi_\lambda\right] (\lambda^0, z) \phi(z) \, dz \, + \, \mathrm{o}(h^{p \wedge q})$$

We now rewrite $t_{i,N}^3$ as the sum of

$$t_{i,N}^{3,1} := \frac{(\varphi - \bar{\varphi}) \varphi_{\lambda}}{\varphi^2} \text{ and } t_{i,N}^{3,2} := \frac{(\bar{\varphi}^{\delta} - \bar{\varphi}) \varphi_{\lambda}}{\varphi^2},$$

and study separately the corresponding $T_N^{3,1}$ and $T_N^{3,2}$. From (I.57), we derive

$$\mathbb{E}[T_N^{3,1}] = -\frac{h^p}{\ell(0)} \int \frac{\varphi_\lambda \xi_K^p[\varphi]}{\varphi} (\lambda^0, z) \phi(z) \, dz - \frac{h^q}{\ell(0)} \int \frac{\varphi_\lambda \xi_H^q[\varphi]}{\varphi} (\lambda^0, z) \phi(z) \, dz + o(h^{p \wedge q}) \, dz,$$

and we directly deduce from (I.28) and (I.69) that $\mathbb{E}[T_N^{3,2}] = o(h^{p \wedge q})$. Note that

$$t_{i,N}^4 = \frac{(\varphi - \bar{\varphi}^{\delta})^2 \varphi_{\lambda}}{\varphi^2 \bar{\varphi}^{\delta}} + \frac{(\bar{\varphi}_{\lambda} - \varphi_{\lambda})(\varphi - \bar{\varphi}^{\delta})}{\varphi \bar{\varphi}^{\delta}}$$

Then, using (I.59), (I.70), (I.73) and (I.74), we derive $||t_{i,N}^4||_{\infty} = o(h^{p \wedge q})$ and Lemma 7.2 leads to $\mathbb{E}(T_N^4) = o(h^{p \wedge q})$.

From Lemmas 7.4, 7.5 and 7.6, we have $\mathbb{E}(T_N^j) = 0$ for $j = 5 \dots 7$ and Lemma 7.7 gives

$$\mathbb{E}\left[T_{N}^{8}\right] \sim \frac{h^{-d-n-1}}{\ell(0)N} \int \frac{\phi(z)}{\varphi(\lambda^{0},z)} K(l_{2}-l_{1})K(l_{1})\nabla K(l_{1})H^{2}(v) \, dl_{1} \, dv \, dl_{2} \, dz \, .$$

Finally, Lemma 7.8 tells us $\mathbb{E}[T_N^9] = o(N^{-1}h^{-d-n-1}).$ We then obtain $\mathbb{E}[\tilde{\beta}_N]$ by summing up the $\mathbb{E}[T_N^j]$ for $j = 1, \dots, 9$.

2. We then analyze the variance of $\tilde{\beta}_N$. For any j = 1, ..., 4, expressions (I.59), (I.70), (I.73) and (I.74) imply $||t_N^j||_{\infty} = O(1)$. Then, Lemma 7.3 leads to

$$\operatorname{Var}(T_N^j) = o(N^{-1}h^{-d-2}) \quad \text{for every } j = 1, \dots, 4$$

From Lemma 7.4, we get

$$\mathbb{V}ar(T_N^5) \sim \frac{1}{\ell(0)Nh^{d+2}} \int \phi^2(z) \left\{ \int K(l_2 - l_1) \nabla K(l_1) dl_1 \right\}^{\otimes} f(\lambda^0, z) \, dz \, dl_2 \,.$$
(I.84)

Indeed, Lemmas 7.5 to 7.8 imply also

$$\mathbb{V}\mathrm{ar}(T_N^j) = o(N^{-1}h^{-d-2}) \quad \text{for every } j = 5, \dots, 9$$

Hence, $\operatorname{Cov}(T_N^j, T_N^k) = o(N^{-1}h^{-d-2})$ unless j = k = 5 and $\operatorname{Var}(\tilde{\beta}_N)$ is given by expression (I.84).

7.4 Central limit theorem

This section is devoted to the proof of Theorem 4.3, which provides a central limit theorem for the double Kernel based estimator $\tilde{\beta}_N$.

Proof of Proposition 4.3. As we saw in the proof of Proposition 4.3, the variance of $\tilde{\beta}_N$ is given by the variance of

$$T_N^{5,2} = \frac{h^{-2d-n-1}}{\ell(0)N} \sum_i b(\Lambda_i, Z_i),$$

where $b(\lambda, z) := h^{d+n} \int \phi(z+hv) K\left(\frac{\lambda^0 - \lambda}{h} - l\right) \nabla K(l) H(v) \, dl \, dv$
 $- h^{n+1} \int \phi(z) \, \bar{\varphi_\lambda}(\lambda^0 - hl, z) K(l) \, dl.$

As in the proofs of theorems 4.1 and 4.2, using Kolmogorov's condition with the fourth moment of b and the Cramer-Wold device, we derive that $T_N^{5,2}$ is asymptotically normal. We then finally deduce that

$$\sqrt{Nh^{d+2}} \left(\tilde{\beta}_N - \mathbb{E}[\tilde{\beta}_N] \right) \quad \stackrel{law}{\underset{N \to \infty}{\longrightarrow}} \quad \mathcal{N} \left(0, \tilde{\Sigma} \right) \; .$$

Under the additional condition $Nh^{d+2+2(p\wedge q)} \to 0$, we conclude the proof denoting that the bias vanishes in the previous expression.

Part II

Numerical approximation of BSDEs with jumps

Abstract

We first study a discrete-time approximation for solutions of systems of decoupled forward-backward stochastic differential equations with jumps. Assuming that the coefficients are Lipschitz-continuous, we prove the convergence of the scheme when the number of time steps n goes to infinity. The rate of convergence is at least $n^{-1/2+\varepsilon}$, for any $\varepsilon > 0$. When the jump coefficient of the first variation process of the forward component satisfies a non-degeneracy condition which ensures its invertibility, we achieve the optimal convergence rate $n^{-1/2}$. The proof is based on a generalization of a remarkable result on the path-regularity of the solution of the backward equation derived by Zhang [104, 105] in the no-jump case. A similar result is obtained without the non-degeneracy assumption whenever the coefficients are C_b^1 with Lipschitz derivatives. Adapting the arguments of Gobet et al [73], we control the statistical error induced by a fully implementable algorithm, where the conditional expectations operators are approximated by means of non-parametric estimation. Several extensions of these results are discussed. In particular, we propose a convergent scheme for the resolution of systems of coupled semilinear parabolic PDE's and provide some numerical examples.

Keywords: Discrete-time approximation, Forward-Backward SDE's with jumps, Malliavin calculus.

Note

The first chapter of this part is based on a paper, written in collaboration with Bruno Bouchard, in revision for *Stochastic Processes and Applications*. The additional second chapter presents a fully implementable algorithm, studies its induced statistical error and provides some numerical results.

Chapter 1

Discrete time approximation

1.1 Introduction

In this chapter, we study a discrete time approximation scheme for the solution of a system of decoupled Forward-Backward Stochastic Differential Equations (FBSDE in short) with jumps of the form

$$\begin{cases} X_t = X_0 + \int_0^t b(X_r) dr + \int_0^t \sigma(X_r) dW_r + \int_0^t \int_E \beta(X_{r-}, e) \bar{\mu}(de, dr) , \\ Y_t = g(X_1) + \int_t^1 h(\Theta_r) dr - \int_t^1 Z_r \cdot dW_r - \int_t^1 \int_E U_r(e) \bar{\mu}(de, dr) \end{cases}$$
(II.1)

where $\Theta := (X, Y, Z, \Gamma)$ with $\Gamma := \int_E \rho(e)U(e)\lambda(de)$. Here, the process W denotes a d-dimensional Brownian motion and $\bar{\mu}$ is an independent compensated Poisson measure $\bar{\mu}(de, dr) = \mu(de, dr) - \lambda(de)dr$. Such equations naturally appear in hedging problems, see e.g. Eyraud-Loisel [48], or in stochastic control, see e.g. Tang and Li [100] and the recent paper Becherer [9] for an application to exponential utility maximization in finance. Under standard Lipschitz assumptions on the coefficients b, σ, β, g and h, existence and uniqueness of the solution have been proved by Tang and Li [100], thus generalizing the seminal paper of Pardoux and Peng [85].

The main motivation for studying discrete time approximations of systems of the above form is that they provide an alternative to classical numerical schemes for a large class of (deterministic) PDE's of the form

$$-\mathcal{L}u(t,x) + h(t,x,u(t,x),\sigma(t,x)\nabla_x u(t,x),\mathcal{I}[u](t,x)) = 0, \ u(1,x) = g(x), \quad (\text{II.2})$$

where

$$\begin{aligned} \mathcal{L}u(t,x) &:= \frac{\partial u}{\partial t}(t,x) + \nabla_x u(t,x)b(x) + \frac{1}{2}\sum_{i,j=1}^d (\sigma\sigma^*(x))^{ij}\frac{\partial^2 u}{\partial x^i \partial x^j}(t,x) \\ &+ \int_E \{u(t,x+\beta(x,e)) - u(t,x) - \nabla_x u(t,x)\beta(x,e)\}\,\lambda(de) \ , \\ \mathcal{I}[u](t,x) &:= \int_E \{u(t,x+\beta(x,e)) - u(t,x)\}\,\rho(e)\,\lambda(de) \ . \end{aligned}$$

Indeed, it is well known that, under mild assumptions on the coefficients, the component Y of the solution can be related to the (viscosity) solution u of (II.2) in the sense that $Y_t = u(t, X_t)$, see e.g. [5]. Thus solving (II.1) or (II.2) is essentially the same. In the socalled four-steps scheme, this relation allows to approximate the solution of (II.1) by first estimating numerically u, see [41] and [77]. Here, we follow the converse approach. Since classical numerical schemes for PDE's generally do not perform well in high dimension, we want to estimate directly the solution of (II.1) so as to provide an approximation of u.

In the no-jump case, i.e. $\beta = 0$, the numerical approximation of (II.1) has already been studied in the literature, see e.g. Zhang [105], Bally and Pages [8], Bouchard and Touzi [19] or Gobet et al. [73]. In [19], the authors suggest the following implicit scheme. Given a regular grid $\pi = \{t_i = i/n, i = 0, ..., n\}$, they approximate X by its Euler scheme X^{π} and (Y, Z) by the discrete-time process $(\bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi})_{i \leq n}$ defined backward by

$$\begin{cases} \bar{Z}_{t_i}^{\pi} = n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \Delta W_{i+1} \mid \mathcal{F}_{t_i} \right] \\ \bar{Y}_{t_i}^{\pi} = \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_i} \right] + \frac{1}{n} h \left(X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi} \right) \end{cases}$$

where $\bar{Y}_{t_n}^{\pi} := g(X_{t_n}^{\pi})$ and $\Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}$. In the no-jump case, it turns out that the discretization error

$$\operatorname{Err}_{n}(Y,Z) := \left\{ \max_{i < n} \sup_{t \in [t_{i}, t_{i+1}]} \mathbb{E}\left[|Y_{t} - \bar{Y}_{t_{i}}^{\pi}|^{2} \right] + \sum_{i=0}^{n-1} \int_{t_{i}}^{t_{i+1}} \mathbb{E}\left[|Z_{t} - \bar{Z}_{t_{i}}^{\pi}|^{2} \right] dt \right\}^{\frac{1}{2}}$$

is intimately related to the quantity

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2\right] dt \quad \text{where} \quad \bar{Z}_{t_i} := n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} Z_t dt \mid \mathcal{F}_{t_i}\right] \ .$$

Under Lipschitz continuity conditions on the coefficients, Zhang [78] was able to prove that the later is of order of n^{-1} . This remarkable result allows to derive the bound $\operatorname{Err}_n(Y,Z) \leq Cn^{-1/2}$. Observe that this rate of convergence can not be improved in

1.1. INTRODUCTION

general. Consider for example the case where X is equal to the Brownian motion W, g is the identity and h = 0. Then, Y = W and $\bar{Y}_{t_i}^{\pi} = W_{t_i}$. Nevertheless, we refer to Gobet and Labart [56] who obtained, at each time t_i , an expansion of the error $|Y_{t_i} - \bar{Y}_{t_i}^{\pi}|$ in terms of $|X_{t_i} - X_{t_i}^{\pi}| \wedge n^{-1}$, so that the error at time 0 is finally of order n^{-1} , thus generalizing the results of Chevance [26].

In this chapter, we extend the approach of Bouchard and Touzi [19] and approximate the solution of (II.1) by the backward scheme

$$\begin{cases} \bar{Z}_{t_{i}}^{\pi} = n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \Delta W_{i+1} \mid \mathcal{F}_{t_{i}} \right] \\ \bar{\Gamma}_{t_{i}}^{\pi} = n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \int_{E} \rho(e) \bar{\mu}(de, (t_{i}, t_{i+1}]) \mid \mathcal{F}_{t_{i}} \right] \\ \bar{Y}_{t_{i}}^{\pi} = \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_{i}} \right] + \frac{1}{n} h \left(X_{t_{i}}^{\pi}, \bar{Y}_{t_{i}}^{\pi}, \bar{Z}_{t_{i}}^{\pi}, \bar{\Gamma}_{t_{i}}^{\pi} \right) \end{cases}$$

where $\bar{Y}_{t_n}^{\pi} := g(X_{t_n}^{\pi})$. By adapting the arguments of Gobet et al. [73], we first prove that our discretization error

$$\operatorname{Err}_{n}(Y, Z, U) := \left\{ \max_{i < n} \sup_{t \in [t_{i}, t_{i+1}]} \mathbb{E}\left[|Y_{t} - \bar{Y}_{t_{i}}^{\pi}|^{2} \right] + \sum_{i=0}^{n-1} \int_{t_{i}}^{t_{i+1}} \mathbb{E}\left[|Z_{t} - \bar{Z}_{t_{i}}^{\pi}|^{2} + |\Gamma_{t} - \bar{\Gamma}_{t_{i}}^{\pi}|^{2} \right] dt \right\}^{\frac{1}{2}}$$

converges to 0 as the discretization step 1/n tends to 0. We then provide upper bounds on

$$\max_{i < n} \sup_{t \in [t_i, t_{i+1}]} \mathbb{E}\left[|Y_t - Y_{t_i}|^2 \right] + \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2 \right] dt ,$$

where $\bar{\Gamma}_{t_i} := n \mathbb{E} \left[\int_{t_i}^{t_{i+1}} \Gamma_t dt \mid \mathcal{F}_{t_i} \right]$. When the coefficients are Lipschitz continuous, we obtain

$$\max_{i < n} \sup_{t \in [t_i, t_{i+1}]} \mathbb{E}\left[|Y_t - Y_{t_i}|^2 \right] + \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|\Gamma_t - \bar{\Gamma}_{t_i}|^2 \right] dt \leq C n^{-1}$$

and

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 \right] dt \leq C_{\varepsilon} n^{-1+\varepsilon}, \quad \text{for any } \varepsilon > 0.$$

Under some additional conditions on the inversibility of $\nabla \beta + I_d$, see \mathbf{H}_1 , or on the regularity of the coefficient, see \mathbf{H}_2 , we then prove that the previous inequality holds true for $\varepsilon = 0$. This extends to our framework the remarkable result derived by Zhang [104, 105] in the no-jump case. It allows us to show that our discrete-time scheme achieves, under the standard Lipschitz conditions, a rate of convergence of at least

 $n^{-1/2+\varepsilon}$, for any $\varepsilon > 0$, and the optimal rate $n^{-1/2}$ under the additional assumptions \mathbf{H}_1 or \mathbf{H}_2 .

Observe that, in opposition to algorithms based on the approximation of the Brownian motion by discrete processes taking a finite number of possible values (see [3], [21], [26], [28] and [76]), our scheme requires an additional numerical procedure to estimate a large number of conditional expectations. This issue can be solved by approximating the conditional expectation operators numerically in an efficient way. In the no-jump case, Bouchard and Touzi [19] use the Malliavin calculus to rewrite conditional expectations as the ratio of two unconditional expectations which can be estimated by standard Monte-Carlo methods. In the reflected case where h does not depend on Z, Bally and Pages [8] use a quantization approach. Finally, Gobet, Lemor and Warin [73, 57] have suggested an adaptation of the so-called Longstaff and Schwartz algorithm based on non-parametric regressions, see [75], which also works in the case where $\beta \neq 0$ but the driver does not depend on U. We refer to the next chapter for an adaptation of their result to the numerical resolution of systems of FBSDEs with jumps of the general form (II.1), as well as the presentation of some numerical results.

The rest of the chapter is organized as follows. In Section 1.2, we describe the approximation scheme and state our main convergence result. We also discuss several possible extensions. In particular, we propose a convergent scheme for the resolution of systems of coupled semilinear parabolic PDE's. Section 1.3 contains some results on the Malliavin derivatives of Forward and Backward SDE's. Applying these results in Section 1.4, we derive some regularity properties for the solution of the backward equation under additional smoothness assumptions on the coefficients. We finally use an approximation argument to conclude the proof of our main theorem.

Notations : Any element $x \in \mathbb{R}^d$ will be identified to a column vector with *i*-th component x^i and Euclidian norm |x|. For $x_i \in \mathbb{R}^{d_i}$, $i \leq n$ and $d_i \in \mathbb{N}$, we define (x_1, \ldots, x_n) as the column vector associated to $(x_1^1, \ldots, x_1^{d_1}, \ldots, x_n^1, \ldots, x_n^{d_n})$. The scalar product on \mathbb{R}^d is denoted by $x \cdot y$. For a $(d' \times d)$ -dimensional matrix M, we note $|M| := \sup\{|Mx|; x \in \mathbb{R}^d, |x| = 1\}, M^*$ its transpose and we write $M \in \mathbb{M}^d$ if d' = d. Given $p \in \mathbb{N}$ and a measured space (A, \mathcal{A}, μ_A) , we denote by $L^p(A, \mathcal{A}, \mu_A; \mathbb{R}^d)$, or simply $L^p(A, \mathcal{A})$ or $L^p(A)$ if no confusion is possible, the set of p-integrable \mathbb{R}^d -valued measurable maps on (A, \mathcal{A}, μ_A) . For $p = \infty, L^\infty(A, \mathcal{A}, \mu_A; \mathbb{R}^d)$ is the set of essentially bounded \mathbb{R}^d -valued measurable maps. The set of k-times differentiable maps with bounded derivatives up to order k is denoted by C_h^k and $C_h^\infty := \bigcap_{k>1} C_h^k$. For a

map $b : \mathbb{R}^d \mapsto \mathbb{R}^k$, we denote by ∇b is Jacobian matrix whenever it exists. In the following, we shall use these notations without specifying the dimension when it is clearly given by the context.

1.2 Discrete time approximation of decoupled FBSDE with jumps

1.2.1 Decoupled forward backward SDE's

As in [12], we shall work on a suitable product space $\Omega := \Omega_W \times \Omega_\mu$ where Ω_W is the set of continuous functions w from [0, 1] into \mathbb{R}^d , and Ω_μ is the set of integer-valued measures on $[0,1] \times E$ with $E := \mathbb{R}^{d'}$ for some $d' \geq 1$. For $\omega = (w,\eta) \in \Omega$, we set $W(w,\eta) = w$ and $\mu(w,\eta) = \eta$ and define $\mathbb{F}^W = (\mathcal{F}^W_t)_{t\leq 1}$ (resp. $\mathbb{F}^\mu = (\mathcal{F}^\mu_t)_{t\leq 1}$) as the smallest right-continuous filtration on Ω_W (resp. Ω_μ) such that W (resp. μ) is optional. We let \mathbb{P}_W be the Wiener measure on $(\Omega_W, \mathcal{F}^W_1)$ and \mathbb{P}_μ be the measure on $(\Omega_\mu, \mathcal{F}^\mu_1)$ under which μ is a Poisson measure with intensity $\nu(dt, de) = \lambda(de)dt$, for some finite measure λ on E, endowed with its Borel tribe \mathcal{E} . We then define the probability measure $\mathbb{P} := \mathbb{P}_W \otimes \mathbb{P}_\mu$ on $(\Omega, \mathcal{F}^W_1 \otimes \mathcal{F}^\mu_1)$. With this construction, W and μ are independent under \mathbb{P} . Without loss of generality, we can assume that the natural filtration $\mathbb{F} = (\mathcal{F}_t)_{t\leq 1}$ induced by (W, μ) is complete. We denote by $\bar{\mu} := \mu - \nu$ the compensated measure associated to μ .

Given K > 0, two K-Lipschitz continuous functions $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{M}^d$, and a measurable map $\beta : \mathbb{R}^d \times E \to \mathbb{R}^d$ such that

$$\sup_{e \in E} |\beta(0, e)| \leq K \quad \text{and} \quad \sup_{e \in E} |\beta(x, e) - \beta(x', e)| \leq K |x - x'|, \quad \forall x, x' \in \mathbb{R}^d, \text{ (II.3)}$$

we define X as the solution on [0, 1] of

$$X_t = X_0 + \int_0^t b(X_r) dr + \int_0^t \sigma(X_r) dW_r + \int_0^t \int_E \beta(X_{r-}, e) \bar{\mu}(de, dr) , \quad (\text{II.4})$$

for some initial condition $X_0 \in \mathbb{R}^d$. The existence and uniqueness of such a solution is well known under the above assumptions, see e.g. [52] and the Appendix for standard estimates for solutions of such SDE.

Before introducing the backward SDE, we need to define some additional notations. Given $s \leq t$ and some real number $p \geq 2$, we denote by $S_{[s,t]}^p$ the set of real valued adapted cadlag processes Y such that

$$\|Y\|_{\mathcal{S}^p_{[s,t]}} := \mathbb{E}\left[\sup_{s \le r \le t} |Y_r|^p\right]^{\frac{1}{p}} < \infty,$$

 $\mathbf{H}^p_{[s,t]}$ is the set of progressively measurable \mathbb{R}^d -valued processes Z such that

$$||Z||_{\mathbf{H}^p_{[s,t]}} := \mathbb{E}\left[\left(\int_s^t |Z_r|^2 dr\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} < \infty$$

 $\mathbf{L}^p_{\lambda,[s,t]}$ is the set of $\mathcal{P} \otimes \mathcal{E}$ measurable maps $U : \Omega \times [0,1] \times E \to \mathbb{R}$ such that

$$\|U\|_{\mathbf{L}^p_{\lambda,[s,t]}} := \mathbb{E}\left[\int_s^t \int_E |U_s(e)|^p \lambda(de) ds\right]^{\frac{1}{p}} < \infty$$

with \mathcal{P} defined as the σ -algebra of \mathbb{F} -predictable subsets of $\Omega \times [0,1]$. The space

$$\mathcal{B}^p_{[s,t]} := \mathcal{S}^p_{[s,t]} imes \mathbf{H}^p_{[s,t]} imes \mathbf{L}^p_{\lambda,[s,t]}$$

is endowed with the norm

$$\|(Y,Z,U)\|_{\mathcal{B}^{p}_{[s,t]}} := \left(\|Y\|^{p}_{\mathcal{S}^{p}_{[s,t]}} + \|Z\|^{p}_{\mathbf{H}^{p}_{[s,t]}} + \|U\|^{p}_{\mathbf{L}^{p}_{\lambda,[s,t]}}\right)^{\frac{1}{p}}$$

In the sequel, we shall omit the subscript [s,t] in these notations when (s,t) = (0,1). For ease of notations, we shall sometimes write that an \mathbb{R}^n -valued process is in $\mathcal{S}^p_{[s,t]}$ or $\mathbf{L}^p_{\lambda,[s,t]}$ meaning that each component is in the corresponding space. Similarly an element of $\mathbb{M}^{d'}$ is said to belong to $\mathbf{H}^p_{[s,t]}$ if each column belongs to $\mathbf{H}^p_{[s,t]}$. The norms are then naturally extended to such processes.

The aim of this chapter is to study a discrete time approximation of the triplet (Y, Z, U) solution on [0, 1] of the backward stochastic differential equation

$$Y_t = g(X_1) + \int_t^1 h(\Theta_r) \, dr - \int_t^1 Z_r \cdot dW_r - \int_t^1 \int_E U_r(e) \bar{\mu}(de, dr) \,, \tag{II.5}$$

where $\Theta := (X, Y, Z, \Gamma)$ and Γ is defined by

$$\Gamma := \int_E \rho(e) U(e) \lambda(de) ,$$

for some measurable map $\rho: E \to \mathbb{R}^{d'}$ satisfying

$$\sup_{e \in E} |\rho(e)| \leq K.$$
 (II.6)
By a solution, we mean a triplet $(Y, Z, U) \in \mathcal{B}^2$ satisfying (II.5).

In order to ensure the existence and uniqueness of a solution to (II.5), we assume that the map $g : \mathbb{R}^d \mapsto \mathbb{R}$ and $h : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d'} \to \mathbb{R}$ are K-Lipschitz continuous (see Lemma 1.5.2 in the Appendix).

For ease of notations, we shall denote by C_p a generic constant depending only on p and the constants K, $\lambda(E)$, b(0), $\sigma(0)$, h(0) and g(0). We write C_p^0 if it also depends on X_0 . In this chapter, p will always denote a real number greater than 2.

Remark 1.2.1 For the convenience of the reader, we have collected in the Appendix standard estimates for the solutions of Forward and Backward SDE's. In particular, they imply

$$\|(X, Y, Z, U)\|_{\mathcal{S}^p \times \mathcal{B}^p}^p \leq C_p (1 + |X_0|^p) , \ p \ge 2.$$
 (II.7)

The estimate on X is standard, see (II.82) of Lemma 1.5.1 in the Appendix. Plugging this in (II.86) of Lemma 1.5.2 leads to the bound on $||(Y, Z, U)||_{\mathcal{B}^p}$. Using (II.83) of Lemma 1.5.1, we also deduce that

$$\mathbb{E}\left[\sup_{s\leq u\leq t}|X_u-X_s|^p\right] \leq C_p \left(1+|X_0|^p\right)|t-s|, \qquad (\text{II.8})$$

while the previous estimates on X combined with (II.87) of Lemma 1.5.2 implies

$$\mathbb{E}\left[\sup_{s\leq u\leq t}|Y_u - Y_s|^p\right] \leq C_p \left\{ \left(1 + |X_0|^p\right)|t - s|^p + ||Z||^p_{\mathbf{H}^p_{[s,t]}} + ||U||^p_{\mathbf{L}^p_{\lambda,[s,t]}} \right\} .$$
 (II.9)

1.2.2 Discrete time approximation

We first fix a regular grid $\pi := \{t_i := i/n, i = 0, ..., n\}$ on [0, 1] and approximate X by its Euler scheme X^{π} defined by

$$\begin{cases} X_0^{\pi} := X_0 \\ X_{t_{i+1}}^{\pi} := X_{t_i}^{\pi} + \frac{1}{n} b(X_{t_i}^{\pi}) + \sigma(X_{t_i}^{\pi}) \Delta W_{i+1} + \int_E \beta(X_{t_i}^{\pi}, e) \bar{\mu}(de, (t_i, t_{i+1}]) \end{cases}$$
(II.10)

where $\Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}$. It is well known, see for example [24], that

$$\max_{i < n} \mathbb{E} \left[\sup_{t \in [t_i, t_{i+1}]} |X_t - X_{t_i}^{\pi}|^2 \right] \leq C_2^0 n^{-1}.$$
(II.11)

We then approximate (Y, Z, Γ) by $(\bar{Y}^{\pi}, \bar{Z}^{\pi}, \bar{\Gamma}^{\pi})$ defined by the backward implicit scheme

$$\begin{cases} \bar{Z}_{t}^{\pi} := n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \Delta W_{i+1} \mid \mathcal{F}_{t_{i}} \right] \\ \bar{\Gamma}_{t}^{\pi} := n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \int_{E} \rho(e) \bar{\mu}(de, (t_{i}, t_{i+1}]) \mid \mathcal{F}_{t_{i}} \right] \\ \bar{Y}_{t}^{\pi} := \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_{i}} \right] + \frac{1}{n} h \left(X_{t_{i}}^{\pi}, \bar{Y}_{t_{i}}^{\pi}, \bar{Z}_{t_{i}}^{\pi}, \bar{\Gamma}_{t_{i}}^{\pi} \right) \end{cases}$$
(II.12)

on each interval $[t_i, t_{i+1})$, where $\bar{Y}_{t_n}^{\pi} := g(X_{t_n}^{\pi})$. Observe that the resolution of the last equation in (II.12) may involve the use of a fixed point procedure. However, h being Lipschitz and multiplied by 1/n, the approximation error can be neglected for large values of n.

Remark 1.2.2 The above backward scheme, which is a natural extension of the one considered in [19] in the case $\beta = 0$, can be understood as follows. On each interval $[t_i, t_{i+1})$, we want to replace the arguments (X, Y, Z, Γ) of h in (II.5) by \mathcal{F}_{t_i} -measurable random variables $(\tilde{X}_{t_i}, \tilde{Y}_{t_i}, \tilde{Z}_{t_i}, \tilde{\Gamma}_{t_i})$. It is natural to take $\tilde{X}_{t_i} = X_{t_i}^{\pi}$. Taking conditional expectation, we obtain the approximation

$$Y_{t_i} \cong \mathbb{E}\left[Y_{t_{i+1}} \mid \mathcal{F}_{t_i}\right] + \frac{1}{n}h\left(X_{t_i}^{\pi}, \tilde{Y}_{t_i}, \tilde{Z}_{t_i}, \tilde{\Gamma}_{t_i}\right) \,.$$

This leads to a backward implicit scheme for Y of the form

$$\bar{Y}_{t_i}^{\pi} = \mathbb{E}\left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_i}\right] + \frac{1}{n}h\left(X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \tilde{Z}_{t_i}, \tilde{\Gamma}_{t_i}\right).$$
(II.13)

It remains to choose \tilde{Z}_{t_i} and $\tilde{\Gamma}_{t_i}$ in terms of $\bar{Y}^{\pi}_{t_{i+1}}$. By the representation theorem, there exist two processes $Z^{\pi} \in \mathbf{H}^2$ and $U^{\pi} \in \mathbf{L}^2_{\lambda}$ satisfying

$$\bar{Y}_{t_{i+1}}^{\pi} - \mathbb{E}\left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_i}\right] = \int_{t_i}^{t_{i+1}} Z_s^{\pi} \cdot dW_s + \int_{t_i}^{t_{i+1}} \int_E U_s^{\pi}(e)\bar{\mu}(ds, de)$$

Observe that they do not depend on the way $\bar{Y}_{t_i}^{\pi}$ is defined and that \bar{Z}^{π} and $\bar{\Gamma}^{\pi}$ defined in (II.12) satisfy

$$\bar{Z}_{t_i}^{\pi} = n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} Z_s^{\pi} ds \mid \mathcal{F}_{t_i}\right] \quad \text{and} \quad \bar{\Gamma}_{t_i}^{\pi} = n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \Gamma_s^{\pi} ds \mid \mathcal{F}_{t_i}\right] \quad (\text{II.14})$$

and thus coincide with the best $\mathbf{H}^2_{[t_i,t_{i+1}]}$ -approximations of the processes $(Z^{\pi}_t)_{t_i \leq t < t_{i+1}}$ and $(\Gamma^{\pi}_t)_{t_i \leq t < t_{i+1}} := (\int_E \rho(e) U^{\pi}_t(e) \lambda(de))_{t_i \leq t < t_{i+1}}$ by \mathcal{F}_{t_i} -measurable random variables (viewed as constant processes on $[t_i, t_{i+1}]$), i.e.

$$\mathbb{E}\left[\int_{t_{i}}^{t_{i+1}} |Z_{t}^{\pi} - \bar{Z}_{t_{i}}^{\pi}|^{2} dt\right] = \inf_{Z_{i} \in L^{2}(\Omega, \mathcal{F}_{t_{i}})} \mathbb{E}\left[\int_{t_{i}}^{t_{i+1}} |Z_{t}^{\pi} - Z_{i}|^{2} dt\right] \\
\mathbb{E}\left[\int_{t_{i}}^{t_{i+1}} |\Gamma_{t}^{\pi} - \bar{\Gamma}_{t_{i}}^{\pi}|^{2} dt\right] = \inf_{\Gamma_{i} \in L^{2}(\Omega, \mathcal{F}_{t_{i}})} \mathbb{E}\left[\int_{t_{i}}^{t_{i+1}} |\Gamma_{t}^{\pi} - \Gamma_{i}|^{2} dt\right]$$

Thus, it is natural to take $(\tilde{Z}_{t_i}, \tilde{\Gamma}_{t_i}) = (\bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi})$ in (II.13), so that

$$\bar{Y}_{t_i}^{\pi} = \bar{Y}_{t_{i+1}}^{\pi} + \frac{1}{n} h\left(X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi}\right) - \int_{t_i}^{t_{i+1}} Z_s^{\pi} \cdot dW_s - \int_{t_i}^{t_{i+1}} \int_E U_s^{\pi}(e) \bar{\mu}(ds, de) \,.$$

Finally, observe that, if we define Y^{π} on $[t_i, t_{i+1})$ by setting

$$Y_t^{\pi} := \bar{Y}_{t_i}^{\pi} - (t - t_i) h(X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi}) + \int_{t_i}^t Z_s^{\pi} dW_s + \int_{t_i}^t \int_E U_s^{\pi}(e) \bar{\mu}(ds, de) ,$$

we obtain

$$n\mathbb{E}\left[\int_{t_{i}}^{t_{i+1}} Y_{t}^{\pi} dt \mid \mathcal{F}_{t_{i}}\right] = \mathbb{E}\left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_{i}}\right] + \frac{1}{n}h\left(X_{t_{i}}^{\pi}, \bar{Y}_{t_{i}}^{\pi}, \bar{Z}_{t_{i}}^{\pi}, \bar{\Gamma}_{t_{i}}^{\pi}\right) = Y_{t_{i}}^{\pi} = \bar{Y}_{t_{i}}^{\pi}$$

Thus, in this scheme, $\bar{Y}_{t_i}^{\pi}$ is the best $\mathbf{H}^2_{[t_i,t_{i+1}]}$ -approximation of Y^{π} on $[t_i,t_{i+1})$ by an \mathcal{F}_{t_i} -measurable random variables (viewed as constant processes on $[t_i,t_{i+1})$). This explains the notation \bar{Y}^{π} which is consistent with the definition of \bar{Z}^{π} and $\bar{\Gamma}^{\pi}$.

Remark 1.2.3 One could also use an explicit scheme as in e.g. [8] or [73]. In this case, (II.12) has to be replaced by

$$\begin{cases} \tilde{Z}_{t_i}^{\pi} := n \mathbb{E} \left[\tilde{Y}_{t_{i+1}}^{\pi} \Delta W_{i+1} \mid \mathcal{F}_{t_i} \right] \\ \tilde{\Gamma}_{t_i}^{\pi} := n \mathbb{E} \left[\tilde{Y}_{t_{i+1}}^{\pi} \int_E \rho(e) \bar{\mu}(de, (t_i, t_{i+1}]) \mid \mathcal{F}_{t_i} \right] \\ \tilde{Y}_{t_i}^{\pi} := \mathbb{E} \left[\tilde{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_i} \right] + \frac{1}{n} \mathbb{E} \left[h \left(X_{t_i}^{\pi}, \tilde{Y}_{t_{i+1}}^{\pi}, \tilde{Z}_{t_i}^{\pi}, \tilde{\Gamma}_{t_i}^{\pi} \right) \mid \mathcal{F}_{t_i} \right] \end{cases}$$
(II.15)

with the terminal condition $\tilde{Y}_{t_n}^{\pi} = g(X_{t_n}^{\pi})$. The advantage of this scheme is that it does not require a fixed point procedure. However, from a numerical point of view, adding a term in the conditional expectation defining $\tilde{Y}_{t_i}^{\pi}$ makes it more difficult to estimate. We therefore think that the implicit scheme may be more tractable in practice. The convergence of the explicit scheme will be discussed in Remarks 1.2.6 and 1.2.8 below.

1.2.3 Convergence of the approximation scheme

In this subsection, we show that the approximation error

$$\operatorname{Err}_{\mathbf{n}}(Y, Z, U) := \left\{ \sup_{t \leq 1} \mathbb{E}\left[|Y_t - \bar{Y}_t^{\pi}|^2 \right] + \|Z - \bar{Z}^{\pi}\|_{\mathbf{H}^2}^2 + \|\Gamma - \bar{\Gamma}^{\pi}\|_{\mathbf{H}^2}^2 \right\}^{\frac{1}{2}}$$

converges to 0. Let us first introduce the processes $(\bar{Z}, \bar{\Gamma})$ defined on each interval $[t_i, t_{i+1})$ by

$$\bar{Z}_t := n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} Z_s \, ds \mid \mathcal{F}_{t_i}\right] \quad \text{and} \quad \bar{\Gamma}_t := n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \Gamma_s \, ds \mid \mathcal{F}_{t_i}\right] \,.$$

Remark 1.2.4 Observe that \bar{Z}_{t_i} and $\bar{\Gamma}_{t_i}$ are the counterparts of $\bar{Z}_{t_i}^{\pi}$ and $\bar{\Gamma}_{t_i}^{\pi}$ for the original backward SDE. They can also be interpreted as the best $\mathbf{H}^2_{[t_i,t_{i+1}]}$ -approximations of $(Z_t)_{t_i \leq t < t_{i+1}}$ and $(\Gamma_t)_{t_i \leq t < t_{i+1}}$ by \mathcal{F}_{t_i} -measurable random variables (viewed as constant processes on $[t_i, t_{i+1}]$), i.e.

$$\mathbb{E}\left[\int_{t_i}^{t_{i+1}} |Z_t - \bar{Z}_{t_i}|^2 dt\right] = \inf_{\substack{Z_i \in L^2(\Omega, \mathcal{F}_{t_i})}} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} |Z_t - Z_i|^2 dt\right]$$
$$\mathbb{E}\left[\int_{t_i}^{t_{i+1}} |\Gamma_t - \bar{\Gamma}_{t_i}|^2 dt\right] = \inf_{\Gamma_i \in L^2(\Omega, \mathcal{F}_{t_i})} \mathbb{E}\left[\int_{t_i}^{t_{i+1}} |\Gamma_t - \Gamma_i|^2 dt\right]$$

Proposition 1.2.1 We have

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Y_t - Y_{t_i}|^2 \right] dt \le C_2^0 \ n^{-1} \quad and \quad \|Z - \bar{Z}\|_{\mathbf{H}^2} + \|\Gamma - \bar{\Gamma}\|_{\mathbf{H}^2} \le \epsilon(n) \,, \quad (\text{II.16})$$

where $\epsilon(n) \to 0$ as $n \to \infty$. Moreover,

noreover,

$$\operatorname{Err}_{n}(Y, Z, U) \leq C_{2}^{0} \left(n^{-1/2} + \|Z - \bar{Z}\|_{\mathbf{H}^{2}} + \|\Gamma - \bar{\Gamma}\|_{\mathbf{H}^{2}} \right) , \qquad (\text{II.17})$$

so that

$$\operatorname{Err}_n(Y,Z,U) \xrightarrow[n \to \infty]{} 0$$

Proof. The proof follows from the same arguments as in [19]. We therefore only sketch it and refer to the above paper for more details. Recall from Remark 1.2.2 that

$$Y_t^{\pi} = \bar{Y}_{t_i}^{\pi} - (t - t_i) h(X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi}) + \int_{t_i}^t Z_s^{\pi} \cdot dW_s + \int_{t_i}^t \int_E U_s^{\pi}(e) \bar{\mu}(ds, de)$$

on $[t_i, t_{i+1})$ and that $\bar{Y}_{t_i}^{\pi} = Y_{t_i}^{\pi}$. For L = Y, Z or U, we set $\delta L := L - L^{\pi}$. It follows from the definition of \bar{Z}^{π} and \bar{U}^{π} in (II.14), Jensen's inequality and the bound on ρ that

$$\mathbb{E}\left[|\bar{Z}_{t_i} - \bar{Z}_{t_i}^{\pi}|^2\right] + \mathbb{E}\left[|\bar{\Gamma}_{t_i} - \bar{\Gamma}_{t_i}^{\pi}|^2\right] \le C_2 n \left(\|\delta Z\|^2_{\mathbf{H}^2_{[t_i, t_{i+1}]}} + \|\delta U\|^2_{\mathbf{L}^2_{\lambda, [t_i, t_{i+1}]}}\right).$$
(II.18)

For $t \in [t_i, t_{i+1})$, we deduce from Itô's Lemma, the Lipschitz property of h, (II.11) and (II.18) that

$$\mathbb{E}[|\delta Y_t|^2] + \|\delta Z\|^2_{\mathbf{H}^2_{[t,t_{i+1}]}} + \|\delta U\|^2_{\mathbf{L}^2_{\lambda,[t,t_{i+1}]}} \leq \mathbb{E}[|\delta Y_{t_{i+1}}|^2] + \alpha \int_t^{t_{i+1}} \mathbb{E}[|\delta Y_s|^2] ds + \frac{C_2^0}{\alpha} \left(n^{-2} + \bar{B}_i + B_i^{\pi}\right), \quad (\text{II.19})$$

where α is some positive constant to be chosen later, and (\bar{B}_i, B_t^{π}) is defined as

$$\bar{B}_{i} := \int_{t_{i}}^{t_{i+1}} \left(\mathbb{E}\left[|Y_{s} - Y_{t_{i}}|^{2} \right] + \mathbb{E}\left[|Z_{s} - \bar{Z}_{s}|^{2} \right] + \mathbb{E}\left[|\Gamma_{s} - \bar{\Gamma}_{s}|^{2} \right] \right) ds B_{i}^{\pi} := n^{-1} \mathbb{E}[|\delta Y_{t_{i}}|^{2}] + \|\delta Z\|_{\mathbf{H}^{2}_{[t_{i}, t_{i+1}]}}^{2} + \|\delta U\|_{\mathbf{L}^{2}_{\lambda, [t_{i}, t_{i+1}]}}^{2} .$$

Using Gronwall's Lemma, it follows that

$$\mathbb{E}[|\delta Y_t|^2] \leq \left(\mathbb{E}[|\delta Y_{t_{i+1}}|^2] + \frac{C_2^0}{\alpha} \left(n^{-2} + \bar{B}_i + B_i^{\pi} \right) \right) e^{\alpha/n} .$$
(II.20)

Let C denote an upper bound for the generic constants C_2^0 appearing in (II.19) and (II.20). Plugging (II.20) in (II.19) and taking $\alpha := 4C$ and n greater than $4Ce^1$ leads to

$$\mathbb{E}[|\delta Y_{t_i}|^2] + \frac{1}{2} \left(\|\delta Z\|^2_{\mathbf{H}^2_{[t_i, t_{i+1}]}} + \|\delta U\|^2_{\mathbf{L}^2_{\lambda, [t_i, t_{i+1}]}} \right) \leq (1 + \frac{C_2^0}{n}) \mathbb{E}[|\delta Y_{t_{i+1}}|^2] \quad (\text{II.21})$$

+ $C_2^0 \left(n^{-2} + \bar{B}_i + n^{-1} \mathbb{E}[|\delta Y_{t_i}|^2] \right) .$

For $n \ge 4Ce^1$, combining the last inequality with the identity $\delta Y_{t_n} = g(X_1) - g(X_1^{\pi})$ and the estimate (II.11) leads to

$$\mathbb{E}[|\delta Y_{t_i}|^2] \leq C_2^0 \left(n^{-1} + \bar{B} \right) \quad \text{where} \quad \bar{B} := \sum_{j=0}^{n-1} \bar{B}_j, \qquad (\text{II}.22)$$

which plugged into (II.21) implies

$$\mathbb{E}[|\delta Y_{t_i}|^2] + \eta \left(\|\delta Z\|^2_{\mathbf{H}^2_{[t_i, t_{i+1}]}} + \|\delta U\|^2_{\mathbf{L}^2_{\lambda, [t_i, t_{i+1}]}} \right) \leq \mathbb{E}[|\delta Y_{t_{i+1}}|^2] + C_2^0 \left(n^{-2} + \frac{\bar{B}}{n} + \bar{B}_i \right)$$

Summing up over i and using (II.20) and (II.22), we finally obtain

$$\operatorname{Err}_{n}(Y, Z, U)^{2} \leq C_{2}^{0}(n^{-1} + \bar{B})$$
 (II.23)

Since Y solves (II.5),

$$\mathbb{E}\left[|Y_t - Y_{t_i}|^2\right] \leq C_2^0 \int_{t_i}^t \mathbb{E}\left[|h(X_r, Y_r, Z_r, \Gamma_r)|^2 + |Z_r|^2 + \int_E |U_r(e)|^2 \lambda(de)\right] dr .$$

Combining the Lipschitz property of h with (II.7), it follows that

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Y_t - Y_{t_i}|^2\right] dt \leq \frac{C_2^0}{n}.$$

This is exactly the first part of (II.16) which combined with (II.23) leads to (II.17). It remains to prove the second part of (II.16). Since Z is \mathbb{F} -adapted, there is a sequence

of adapted processes $(Z^n)_n$ such that $Z_t^n = Z_{t_i}^n$ on each $[t_i, t_{i+1})$ and Z^n converges to Z in \mathbf{H}^2 . By Remark 1.2.4, we observe that

$$||Z - \bar{Z}||^2_{\mathbf{H}^2} \leq ||Z - Z^n||^2_{\mathbf{H}^2},$$

and applying the same reasoning to Γ concludes the proof.

Remark 1.2.5 If $\sigma = 0$, which implies $Z = \overline{Z}^{\pi} = 0$, or *h* does not depend on *Z*, the term \overline{B}_i in the above proof reduces to

$$\bar{B}_i = \int_{t_i}^{t_{i+1}} \left(\mathbb{E}\left[|Y_s - Y_{t_i}|^2 \right] + \mathbb{E}\left[|\Gamma_s - \bar{\Gamma}_s|^2 \right] \right) ds \,.$$

In this case, the assertion (II.17) of Proposition 1.2.1 can be replaced by

$$\operatorname{Err}_{n}(Y,U) := \left\{ \sup_{t \leq 1} \mathbb{E}\left[|Y_{t} - \bar{Y}_{t}^{\pi}|^{2} \right] + \|\Gamma - \bar{\Gamma}^{\pi}\|_{\mathbf{H}^{2}}^{2} \right\}^{\frac{1}{2}} \leq C_{2}^{0} \left(n^{-1/2} + \|\Gamma - \bar{\Gamma}\|_{\mathbf{H}^{2}} \right).$$
(II.24)

Remark 1.2.6 In this Remark, we explain how to adapt the proof of Proposition 1.2.1 to the explicit scheme defined in (II.15). First, we can find some $\hat{Z}^{\pi} \in \mathbf{H}^2$ and $\hat{U}^{\pi} \in \mathbf{L}^2_{\lambda}$ such that

$$\tilde{Y}_{t_{i+1}}^{\pi} = \mathbb{E}\left[\tilde{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_i}\right] + \int_{t_i}^{t_{i+1}} \hat{Z}_s^{\pi} \cdot dW_s + \int_{t_i}^{t_{i+1}} \int_E \hat{U}_s^{\pi}(e)\bar{\mu}(de, ds) + \int_{t_i}^{t_i} \hat{U}_s^{\pi}(e)\bar{\mu}(de, ds) + \int_{$$

We then define \hat{Y}^{π} on $[t_i, t_{i+1}]$ by

$$\hat{Y}_{t}^{\pi} = \tilde{Y}_{t_{i}}^{\pi} - (t - t_{i}) \mathbb{E} \left[h \left(X_{t_{i}}^{\pi}, \tilde{Y}_{t_{i+1}}^{\pi}, \tilde{Z}_{t_{i}}^{\pi}, \tilde{\Gamma}_{t_{i}}^{\pi} \right) \mid \mathcal{F}_{t_{i}} \right] + \int_{t_{i}}^{t} \hat{Z}_{s}^{\pi} \cdot dW_{s}$$

$$+ \int_{t_{i}}^{t} \int_{E} \hat{U}_{s}^{\pi}(e) \bar{\mu}(de, ds) \, .$$

Observe that $\hat{Y}_{t_{i+1}}^{\pi} = \tilde{Y}_{t_{i+1}}^{\pi}$ and

$$\tilde{Z}_{t_i}^{\pi} = n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \hat{Z}_s^{\pi} ds \mid \mathcal{F}_{t_i}\right] \quad , \quad \tilde{\Gamma}_{t_i}^{\pi} = n \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \hat{\Gamma}_s^{\pi} ds \mid \mathcal{F}_{t_i}\right] \; ,$$

for all i < n. Moreover

$$\begin{split} h(X_s, Y_s, Z_s, \Gamma_s) &= \mathbb{E} \left[h(X_{t_i}, Y_{t_{i+1}}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}) \mid \mathcal{F}_{t_i} \right] \\ &+ \mathbb{E} \left[h(X_{t_i}, Y_{t_i}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}) - h(X_{t_i}, Y_{t_{i+1}}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}) \mid \mathcal{F}_{t_i} \right] \\ &+ \left(h(X_s, Y_s, Z_s, \Gamma_s) - h(X_{t_i}, Y_{t_i}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}) \right) \,, \end{split}$$

where by the Lipschitz continuity of h and (i) of Theorem 1.2.1 below

$$\mathbb{E}\left[\left(\mathbb{E}\left[h(X_{t_i}, Y_{t_i}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}) - h(X_{t_i}, Y_{t_{i+1}}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}) \mid \mathcal{F}_{t_i}\right]\right)^2\right] \leq C_2^0/n ,$$

and

$$\mathbb{E}\left[\int_{t}^{t_{i+1}} \left(h(X_s, Y_s, Z_s, \Gamma_s) - h(X_{t_i}, Y_{t_i}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i})\right)^2 ds\right]$$

$$\leq C_2^0 \left(n^{-2} + \int_{t}^{t_{i+1}} \mathbb{E}\left[|Z_s - \bar{Z}_{t_i}|^2\right] + \mathbb{E}\left[|\Gamma_s - \bar{\Gamma}_{t_i}|^2\right] ds\right)$$

by (i) of Theorem 1.2.1 and (II.8). Using these remarks, the proof of Proposition 1.2.1 can be adapted in a straightforward way. This implies that the approximation error due to the explicit scheme is also upper-bounded by $C_2^0 \left(n^{-1/2} + \|Z - \bar{Z}\|_{\mathbf{H}^2} + \|\Gamma - \bar{\Gamma}\|_{\mathbf{H}^2} \right)$.

1.2.4 Path-regularity and convergence rate under additional assumptions

In view of Proposition 1.2.1, the discretization error converges to zero. In order to control its speed of convergence, it remains to study $||Z - \bar{Z}||^2_{\mathbf{H}^2} + ||\Gamma - \bar{\Gamma}||^2_{\mathbf{H}^2}$. Before to state our main result, let us introduce the following assumptions:

 \mathbf{H}_1 : For each $e \in E$, the map $x \in \mathbb{R}^d \mapsto \beta(x, e)$ admits a Jacobian matrix $\nabla \beta(x, e)$ such that the function

$$(x,\xi) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto a(x,\xi;e) := \xi' (\nabla \beta(x,e) + I_d) \xi$$

satisfies one of the following condition uniformly in $(x,\xi) \in \mathbb{R}^d \times \mathbb{R}^d$

$$a(x,\xi;e) \ge |\xi|^2 K^{-1}$$
 or $a(x,\xi;e) \le -|\xi|^2 K^{-1}$

 \mathbf{H}_2 : σ , b, $\beta(\cdot, e)$, h and g are C_b^1 functions with K-Lipschitz continuous derivatives, uniformly in $e \in E$.

Remark 1.2.7 Observe for later use that the condition \mathbf{H}_1 implies that, for each $(x, e) \in \mathbb{R}^d \times E$, the matrix $\nabla \beta(x, e) + I_d$ is invertible with inverse bounded by K. This ensures the inversibility of the first variation process ∇X of X, see Remark 1.3.5. Moreover, if q is a smooth density on \mathbb{R}^d with compact support, then the approximating functions β^k , $k \in \mathbb{N}$, defined by

$$\beta^k(x,e) := \int_{\mathbb{R}^d} k^d \beta(\bar{x},e) q(k[x-\bar{x}]) d\bar{x}$$

are smooth and also satisfy \mathbf{H}_1 .

Our main theorem is stated for a suitable version of (Z, U, Γ) . Observe that it does not change the quantity $\operatorname{Err}_n(Y, Z, U)$. **Theorem 1.2.1** The following holds.

(i) For all
$$i < n$$

$$\mathbb{E}\left[\sup_{t \in [t_i, t_{i+1}]} |Y_t - Y_{t_i}|^2\right] \leq C_2^0 n^{-1} \quad and \quad \mathbb{E}\left[\sup_{t \in [t_i, t_{i+1}]} |\Gamma_t - \Gamma_{t_i}|^2\right] \leq C_2^0 n^{-1} \quad (\text{II.25})$$
so that $\|\Gamma - \bar{\Gamma}\|_{-\infty}^2 \leq C_2^0 n^{-1} \quad and \quad \|\Gamma - \bar{\Gamma}\|_{-\infty}^2 \leq C_2^0 n^{-1} \quad Moreover, for any \epsilon > 0$

so that $\|\Gamma - \Gamma\|_{\mathcal{S}^2}^2 \le C_2^0 n^{-1}$ and $\|\Gamma - \Gamma\|_{\mathbf{H}^2}^2 \le C_2^0 n^{-1}$. Moreover, for any $\varepsilon > 0$, $\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - Z_{t_i}|^2\right] dt \le C_{\varepsilon}^0 n^{-1+\varepsilon}, \qquad (\text{II.26})$

so that $||Z - \overline{Z}||_{\mathbf{H}^2}^2 \leq C_{\varepsilon}^0 n^{-1+\varepsilon}$.

(ii) Assume that \mathbf{H}_1 holds. Then

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - Z_{t_i}|^2 \right] dt \leq C_2^0 n^{-1} , \qquad (\text{II.27})$$

so that $||Z - \bar{Z}||_{\mathbf{H}^2}^2 \le C_2^0 n^{-1}$.

(iii) Assume that \mathbf{H}_2 holds. Then, for all i < n and $t \in [t_i, t_{i+1}]$,

$$\mathbb{E}\left[|Z_t - Z_{t_i}|^2\right] \leq C_2^0 n^{-1}, \qquad (\text{II.28})$$

so that $||Z - \bar{Z}||_{\mathbf{H}^2}^2 \le C_2^0 n^{-1}$.

This regularity property will be proved in the subsequent sections. Combined with Proposition 1.2.1 and Remark 1.2.5, it provides an upper bound for the convergence rate of our backward implicit scheme.

Corollary 1.2.1 For any $\varepsilon > 0$

$$\operatorname{Err}_n(Y, Z, U) \leq C_{\varepsilon}^0 n^{-1/2+\varepsilon}$$

If either \mathbf{H}_1 or \mathbf{H}_2 holds, then

$$\operatorname{Err}_{n}(Y, Z, U) \leq C_{2}^{0} n^{-1/2}$$

If $\sigma = 0$ or h is independent of Z, then

$$\operatorname{Err}_n(Y,U) \leq C_2^0 n^{-1/2}$$

Remark 1.2.8 In view of Remark 1.2.6, the result of Corollary 1.2.1 can be extended to the explicit scheme defined in (II.15).

Remark 1.2.9 In comparison with the results of Zhang [105] in the no jump case, we obtain a speed of order $n^{-1/2+\varepsilon}$ for any $\varepsilon > 0$ under his assumptions and we require additional assumptions \mathbf{H}_1 or \mathbf{H}_1 to derive its optimal speed in $n^{-1/2}$.

1.2.5 Possible Extensions

(i) It will be clear from the proofs that all the results of this chapter hold if we let the maps b, σ, β , and h depend on t whenever these functions are 1/2-Hölder in t and the other assumptions are satisfied uniformly in t. In this case, the backward scheme (II.12) is modified by setting

$$\bar{Y}_{t_i}^{\pi} = \mathbb{E}\left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_i}\right] + \frac{1}{n} h(t_i, X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi}) .$$

(ii) The Euler approximation X^{π} of X could be replaced by any other adapted approximation satisfying (II.11).

(iii) Let M be the solution of the SDE

$$M_t = M_0 + \int_0^t b_M(M_r) dr + \int_0^t \int_E \beta_M(M_{r-}, e) \bar{\mu}(de, dr)$$

where $b_M : \mathbb{R}^k \mapsto \mathbb{R}^k$ and $\beta_M(\cdot, e) : \mathbb{R}^k \mapsto \mathbb{R}^k$, $k \ge 1$, are Lipschitz continuous uniformly in $e \in E$ with $|\beta_M(0, \cdot)|$ bounded, and consider the system

$$\begin{cases} X_t = X_0 + \int_0^t b(M_r, X_r) dr + \int_0^t \sigma(M_r, X_r) dW_r + \int_0^t \int_E \beta(M_{r-}, X_{r-}, e) \bar{\mu}(de, dr) \\ Y_t = g(M_1, X_1) + \int_t^1 h(M_r, \Theta_r) dr - \int_t^1 Z_r \cdot dW_r - \int_t^1 \int_E U_r(e) \bar{\mu}(de, dr) \end{cases}$$
(II.29)

where $b, \sigma, \beta(\cdot, e)$ and h are K-Lispchitz, uniformly in $e \in E$ and $|\beta(0, \cdot)|$ is bounded. Here, the discrete-time approximation of Y is given by

$$\bar{Y}_{t_n}^{\pi} = g(M_{t_n}^{\pi}, X_{t_n}^{\pi}) \quad , \quad \bar{Y}_{t_i}^{\pi} = \mathbb{E}\left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_i}\right] + \frac{1}{n} h\left(M_{t_i}^{\pi}, X_{t_i}^{\pi}, \bar{Y}_{t_i}^{\pi}, \bar{Z}_{t_i}^{\pi}, \bar{\Gamma}_{t_i}^{\pi}\right) \, ,$$

where (M^{π}, X^{π}) is the Euler scheme of (M, X). Considering (M, X) as an \mathbb{R}^{k+d} dimensional forward process, we can clearly apply the results of Proposition 1.2.1. Moreover, Theorem 1.2.1 holds when $b(m, \cdot)$, $\sigma(m, \cdot)$, $\beta(m, \cdot)$, $g(m, \cdot)$ and $h(m, \cdot)$ satisfies the conditions of this theorem as functions of (x, y, z, γ) uniformly in $m \in \mathbb{R}^k$. This comes from the fact that the dynamics of M are independent of X and that the Malliavin derivative of M with respect to the Brownian motion equals zero. This particular feature implies that the proofs of Section 1.3.3 and Section 1.4 work without any modification in this context.

(iv) In [86], see also [98], the authors consider a system of the form

$$\begin{cases} X_t = X_0 + \int_0^t b(M_r, X_r) dr + \int_0^t \sigma(M_r, X_r) dW_r \\ Y_t = g(M_1, X_1) + \int_t^1 h(M_r, \Theta_r) dr - \int_t^1 Z_r \cdot dW_r - \int_t^1 \int_E U_r(e) \bar{\mu}(de, dr) \end{cases}$$
(II.30)

where M is an \mathbb{F}^{μ} -adapted purely discontinuous jump process. In [86], it is shown that a large class of systems of (coupled) semilinear parabolic partial differential equations can be rewritten in terms of systems of BSDE of the form (II.30), where the backward components are decoupled. However, their particular construction implies that b, σ, h and g are not Lipschitz in their first variable m. In this remark, we explain how to consider this particular framework.

Hereafter, we assume that the path of M can be simulated exactly, which is the case in [86]. Then, recalling that $\lambda(E) < \infty$ so that μ has a.s. only a finite number of jumps on [0, 1], we can include the jump times of M in the Euler scheme X^{π} of X. Thus, even if b and σ are not Lipschitz in their first variable m, we can still define an approximating scheme X^{π} of X such that

$$\mathbb{E}\left[\sup_{t\in[t_{i},t_{i+1}]}|X_{t}-X_{t_{i}}^{\pi}|^{2}\right] \leq C_{2}^{0}|t_{i+1}-t_{i}|$$

whenever $b(m, \cdot)$ and $\sigma(m, \cdot)$ are Lipschitz in x and $|b(m, 0)| + |\sigma(m, 0)|$ is bounded, uniformly in m. We now explain how to construct a convergent scheme for the backward component even when g and h are not Lipschitz in m. We assume that $h(m, \cdot)$ is Lipschitz and h(m, 0) is bounded, uniformly in m. We make the same assumption on $g(m, \cdot)$. The approximation is defined as follows:

$$\begin{cases}
\bar{Z}_{t}^{\pi} := n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \Delta W_{i+1} \mid \mathcal{F}_{t_{i}} \right] \\
\bar{\Gamma}_{t}^{\pi} := n \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \int_{E} \rho(e) \bar{\mu}(de, (t_{i}, t_{i+1}]) \mid \mathcal{F}_{t_{i}} \right] \\
\bar{Y}_{t}^{\pi} := \mathbb{E} \left[\bar{Y}_{t_{i+1}}^{\pi} \mid \mathcal{F}_{t_{i}} \right] + \mathbb{E} \left[\int_{t_{i}}^{t_{i+1}} h\left(M_{s}, X_{t_{i}}^{\pi}, \bar{Y}_{t_{i+1}}^{\pi}, \bar{Z}_{t_{i}}^{\pi}, \bar{\Gamma}_{t_{i}}^{\pi} \right) ds \mid \mathcal{F}_{t_{i}} \right]$$
(II.31)

for $t \in [t_i, t_{i+1})$, with the terminal condition $\overline{Y}_{t_n}^{\pi} = g(M_{t_n}, X_{t_n}^{\pi})$. With this scheme, the proof of Proposition 1.2.1 can be modified as follows. We keep the same definition for Z^{π} and U^{π} but we now define Y^{π} as

$$Y_{t}^{\pi} = \bar{Y}_{t_{i}}^{\pi} - (t - t_{i}) \mathbb{E} \left[n \int_{t_{i}}^{t_{i+1}} h\left(M_{s}, X_{t_{i}}^{\pi}, \bar{Y}_{t_{i+1}}^{\pi}, \bar{Z}_{t_{i}}^{\pi}, \bar{\Gamma}_{t_{i}}^{\pi} \right) ds \mid \mathcal{F}_{t_{i}} \right]$$

+ $\int_{t_{i}}^{t} Z_{s}^{\pi} \cdot dW_{s} + \int_{t_{i}}^{t} \int_{E} U_{s}^{\pi}(e) \bar{\mu}(ds, de) .$

Let us introduce the processes $(H_t)_{t\leq 1}$ and $(\bar{H}_t)_{t\leq 1}$ defined, for $t \in [t_i, t_{i+1}]$, by

$$\begin{split} H_t &:= h(M_t, X_{t_i}, Y_{t_i}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}), \quad \bar{H}_t := \mathbb{E}\left[n \int_{t_i}^{t_{i+1}} h\left(M_s, X_{t_i}, Y_{t_i}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}\right) ds \mid \mathcal{F}_{t_i}\right]. \\ \text{Observe that } h(M_t, \Theta_t) - \mathbb{E}\left[n \int_{t_i}^{t_{i+1}} h\left(M_s, X_{t_i}, Y_{t_{i+1}}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}\right) ds \mid \mathcal{F}_{t_i}\right] \text{ can be written } ds \in \mathcal{F}_{t_i}. \end{split}$$

$$\mathbf{s}$$

$$h(M_t, \Theta_t) - H_t + H_t - \bar{H}_{t_i} + \bar{H}_{t_i} - \mathbb{E}\left[n \int_{t_i}^{t_{i+1}} h\left(M_s, X_{t_i}, Y_{t_{i+1}}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}\right) ds \mid \mathcal{F}_{t_i}\right] .$$

Recall from (iii) of this section that (i) of Theorem 1.2.1 holds for (II.30). Following the arguments of Remark 1.2.6, we get

$$\mathbb{E}\left[\left|\bar{H}_{t_i} - \mathbb{E}\left[n \int_{t_i}^{t_{i+1}} h\left(M_s, X_{t_i}, Y_{t_{i+1}}, \bar{Z}_{t_i}, \bar{\Gamma}_{t_i}\right) ds \mid \mathcal{F}_{t_i}\right]\right|^2\right] \leq \frac{C_2^0}{n}$$

By (i) of Theorem 1.2.1 and (II.8),

$$\int_{t_i}^{t_{i+1}} \mathbb{E}\left[|h(M_t,\Theta_t) - H_t|^2\right] dt \leq C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{\Gamma}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{T}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |\Gamma_t - \bar{T}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |T_t - \bar{T}_{t_i}|^2\right] dt\right) + C_2^0 \left(n^{-2} + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - \bar{Z}_{t_i}|^2 + |T_t - \bar{T}_{t_i}|^2\right] dt$$

We then deduce from the same arguments as in the proof of Proposition 1.2.1 that

$$\operatorname{Err}_{n}(Y, Z, U) \leq C_{2}^{0} \left(n^{-1/2} + \|Z - \bar{Z}\|_{\mathbf{H}^{2}} + \|\Gamma - \bar{\Gamma}\|_{\mathbf{H}^{2}} + \|H - \bar{H}\|_{\mathbf{H}^{2}} \right) ,$$

where

$$||Z - \bar{Z}||_{\mathbf{H}^2} + ||\Gamma - \bar{\Gamma}||_{\mathbf{H}^2} + ||H - \bar{H}||_{\mathbf{H}^2} \leq \epsilon(n)$$

for some map ϵ such that $\epsilon(n) \to 0$ when $n \to \infty$. This shows that the approximation scheme is convergent. Recall from (iii) of this section that the results of Theorem 1.2.1 hold for this system. Since here $\beta = 0$, it follows that $\|Z - \bar{Z}\|_{\mathbf{H}^2} + \|\Gamma - \bar{\Gamma}\|_{\mathbf{H}^2} \leq C_2^0 n^{-\frac{1}{2}}$, without any further assumption. We leave the study of $\|H - \bar{H}\|_{\mathbf{H}^2}$ to further research.

1.3 Malliavin calculus for FBSDE

In this section, we prove that the solution (Y, Z, U) of (II.5) is smooth in the Malliavin sense under the additional assumptions

$$\mathbf{C}_1^X : b, \sigma \text{ and } \beta(\cdot, e) \text{ are } C_b^1 \text{ uniformly in } e \in E$$

 $\mathbf{C}_1^Y : g \text{ and } h \text{ are } C_b^1.$

We shall also show that their derivatives are smooth under the stronger assumptions

 $\mathbf{C}_2^X : b, \sigma \text{ and } \beta(\cdot, e) \text{ are } C_b^2 \text{ with second derivatives bounded by K, uniformly in } e \in E$ $\mathbf{C}_2^Y : g \text{ and } h \text{ are } C_b^2 \text{ with second derivatives bounded by K.}$

This will allow us to provide representation and regularity results for Y, Z and U in Section 1.4. Under $\mathbf{C}_1^X \cdot \mathbf{C}_1^Y$, these results will immediately imply the first assertion of (i) of Theorem 1.2.1, while the second one (resp. (ii)) will be obtained by adapting the arguments of [18] (resp. [105] under the additional assumption \mathbf{H}_1). Under $\mathbf{C}_2^X \cdot \mathbf{C}_2^Y$, these results will also directly imply (iii). The proof of Theorem 1.2.1 will then be completed by appealing to an approximation argument.

This section is organized as follows. First we derive some properties for the Malliavin derivatives of stochastic integrals with respect to $\bar{\mu}$. Next, we recall some well known results on the Malliavin derivatives of the forward process X. Finally, we discuss the Malliavin differentiability of the solution of (II.5).

1.3.1 Generalities

The construction of the Malliavin derivatives on the Wiener space is standard, see e.g. [82], and can be easily extended to our setting by observing that there is an isometry between $L^2(\Omega_W \times \Omega_\mu)$ and $L^2(\Omega_W, L^2(\Omega_\mu))$, with obvious notations.

Let ${\bf S}$ denote the set of random variables of the form

$$F = \phi \left(\int_0^1 f^1(t) \cdot dW_t, \dots, \int_0^1 f^{\kappa}(t) \cdot dW_t, \mu \right) ,$$

where $\kappa \geq 1$, $f^i : [0,1] \mapsto \mathbb{R}^d$ is a bounded measurable map for each $i \leq \kappa$, ϕ is a real-valued measurable map on $\mathbb{R}^{\kappa} \times \Omega_{\mu}$ and $\phi(\cdot, \eta) \in C_b^{\infty}$, $\mathbb{P}_{\mu}(d\eta)$ -a.e.

We denote by D the Malliavin derivative operator with respect to the Brownian motion. For $F \in \mathbf{S}$ as above and $s \leq 1$, it is defined as

$$D_s F := \sum_{i \le \kappa} \nabla_i \phi \left(\int_0^1 f^1(t) \cdot dW_t, \dots, \int_0^1 f^\kappa(t) \cdot dW_t, \mu \right) f^i(s) ,$$

where $\nabla_i \phi$ is the derivative of ϕ with respect to its *i*-th argument. We then denote by $\mathrm{ID}^{1,2}$ the closure of **S** with respect to the norm

$$\|F\|_{\mathbb{D}^{1,2}} := \left\{ \mathbb{E}\left[F^2\right] + \mathbb{E}\left[\int_0^1 |D_s F|^2 ds\right] \right\}^{\frac{1}{2}}$$

and define $\mathbf{H}^2(\mathbb{D}^{1,2})$ as the set of elements $\xi \in \mathbf{H}^2$ such that $\xi_t \in \mathbb{D}^{1,2}$ for almost all $t \leq 1$ and such that, after possibly passing to a measurable version,

$$\|\xi\|_{\mathbf{H}^2(\mathbb{D}^{1,2})}^2 := \|\xi\|_{\mathbf{H}^2}^2 + \int_0^1 \|D_s\xi\|_{\mathbf{H}^2}^2 ds < \infty$$

Observe that for ψ in $\mathbf{L}^2_{\lambda}(\mathbb{F}^{\mu})$, the set of elements of \mathbf{L}^2_{λ} which are independent of W, we have $D\psi = 0$. We finally define $\mathbf{L}^2_{\lambda}(\mathbb{ID}^{1,2})$ as the closure of the set

$$\mathbf{L}_{\lambda}^{'2}(\mathbb{D}^{1,2}) \quad := \quad \operatorname{Vect}\left\{\psi = \xi\vartheta \ : \ \xi \in \mathbf{H}^{2}(\mathbb{D}^{1,2},\mathbb{F}^{W}), \ \vartheta \in \mathbf{L}_{\lambda}^{2}(\mathbb{F}^{\mu}), \ \|\psi\|_{\mathbf{L}_{\lambda}^{2}(\mathbb{D}^{1,2})} < \infty\right\}$$

for the norm

$$\|\psi\|^2_{\mathbf{L}^2_{\lambda}(\mathbb{D}^{1,2})} := \|\psi\|^2_{\mathbf{L}^2_{\lambda}} + \int_0^1 \|D_s\psi\|^2_{\mathbf{L}^2_{\lambda}} ds.$$

Here, $\mathbf{H}^2(\mathbb{ID}^{1,2}, \mathbb{F}^W)$ denotes the set of \mathbb{F}^W -adapted elements of $\mathbf{H}^2(\mathbb{ID}^{1,2})$ and $D_s(\xi\vartheta)$ equals $(D_s\xi)\vartheta$ for $\xi \in \mathbf{H}^2(\mathbb{ID}^{1,2}, \mathbb{F}^W)$, $\vartheta \in \mathbf{L}^2_\lambda(\mathbb{F}^\mu)$. Here again, we extend the definition of $\|\cdot\|_{\mathbf{H}^2(\mathbb{ID}^{1,2})}$ and $\|\cdot\|_{\mathbf{L}^2_\lambda(\mathbb{ID}^{1,2})}$ to processes with values in \mathbb{M}^d and \mathbb{R}^d in a natural way.

From now on, given a matrix A, we shall denote by A^i its *i*-th column. For $k \leq d$, we denote by D^k the Malliavin derivative with respect to W^k , meaning that $D^k F = (DF)^k$ for $F \in \mathbb{D}^{1,2}$.

Remark 1.3.1 With this construction, the operator D enjoys the usual properties of the Malliavin derivative operator on Wiener spaces. In particular, if $\xi \in \mathbf{H}^2(\mathbb{ID}^{1,2})$ and $f \in C_b^1(\mathbb{R}^d)$, then

$$D_s\left(\int_0^1 f(\xi_t)dt + \int_0^1 \xi_t \cdot dW_t\right) = \int_s^1 \nabla f(\xi_t) D_s \xi_t dt + \xi_s^* + \sum_{j=1}^d \int_s^1 D_s \xi_t^j \cdot dW_t^j$$

for all $s \leq 1$. Here * denotes transposition. It follows from the same argument as in [82], which we refer to for more details.

Remark 1.3.2 Fix $\xi \in \mathbf{H}^2(\mathbb{D}^{1,2}, \mathbb{F}^W)$. By Lemma 1.3.1 in [82], there exists a family of deterministic measurable kernels $f_m(t_1, \ldots, t_m, t)$ in $L^2([0, 1]^{m+1}), m \ge 0$, such that

$$\xi_t = \sum_{m \ge 0} I_m(f_m(\cdot, t)) \quad \text{and} \quad D_s \xi_t = \sum_{m \ge 1} m I_{m-1}(f_m(\cdot, s, t))$$

where I_m denotes the *m*-iterated Wiener integral, see Proposition 1.2.1 in [82]. Therefore, if τ is a random time bounded by 1 and independent of W, we have

$$\xi_{\tau} = \sum_{m \ge 0} I_m(f_m(\cdot, \tau))$$

and, by the same argument as in the proof of Proposition 1.2.1 in [82], $\xi_{\tau} \in \mathbb{D}^{1,2}$ whenever τ has a bounded density and

$$D_s(\xi_{\tau}) = \sum_{m \ge 1} m I_{m-1}(f_m(\cdot, s, \tau)) = (D_s \xi)_{\tau} .$$

The two following Lemmas are generalizations of Lemma 3.3 and Lemma 3.4 in [86] which correspond to the case where E is finite, see also Lemma 2.3 in [85] for the case of Itô integrals.

Lemma 1.3.1 Assume that $\psi \in \mathbf{L}^2_{\lambda}(\mathbb{D}^{1,2})$. Then,

$$H := \int_0^1 \int_E \psi_t(e) \bar{\mu}(de, dt) \in \mathbb{D}^{1,2}$$

and

$$D_s H := \int_0^1 \int_E D_s \psi_t(e) \overline{\mu}(de, dt) \quad \text{for all } s \le 1 \; .$$

Proof. First notice that it suffices to prove the required result when $\psi \in \mathbf{L}_{\lambda}^{'2}(\mathrm{ID}^{1,2})$. Indeed, we can retrieve the general case by considering a sequence $(\psi^n)_n$ in $\mathbf{L}_{\lambda}^{\prime 2}(\mathbb{ID}^{1,2})$ which converges to ψ in $\mathbf{L}^2_{\lambda}(\mathbb{ID}^{1,2})$, so that $H^n := \int_0^1 \int_E \psi_t^n(e) \bar{\mu}(de, dt)$ is a Cauchy sequence in $\mathbb{ID}^{1,2}$ which converges to H and $(D_s H^n)_{s\leq 1}$ converges to $(\int_0^1 \int_E D_s \psi_t(e) \bar{\mu}(de, dt)))_{s\leq 1}$ in \mathbf{H}^2 .

Therefore, we now assume that $\psi = \xi \vartheta$ where $\xi \in \mathbf{H}^2(\mathbb{D}^{1,2},\mathbb{F}^W), \ \vartheta \in \mathbf{L}^2_{\lambda}(\mathbb{F}^{\mu})$ and $\|\psi\|_{\mathbf{L}^2_{\lambda}(\mathbb{D}^{1,2})} < \infty$. Then,

$$\int_0^1 \int_E \psi_t(e) \bar{\mu}(de, dt) = \int_0^1 \int_E \xi_t \vartheta_t(e) \mu(de, dt) - \int_0^1 \xi_t \int_E \vartheta_t(e) \lambda(de) dt ,$$

where, by Remark 1.3.1 and the fact that $\int_E \vartheta_t(e)\lambda(de) \in L^2_\lambda(\mathbb{F}^\mu)$,

$$D_s \int_0^1 \xi_t \left(\int_E \vartheta_t(e)\lambda(de) \right) dt = \int_0^1 D_s \xi_t \int_E \vartheta_t(e)\lambda(de) dt = \int_0^1 \int_E (D_s \xi_t)\vartheta_t(e)\lambda(de) dt.$$

It remains to prove that

It remains to prove that

$$D_s \int_0^1 \int_E \xi_t \vartheta_t(e) \mu(de, dt) = \int_0^1 \int_E (D_s \xi_t) \vartheta_t(e) \mu(de, dt)$$

To see this, we define N by $N_t := \int_0^t \mu(E, ds)$ for $t \leq 1, (\tau_i)_{i \geq 1}$ as the sequence of jump times on [0,1] of N and $(\mathcal{E}_i)_{i\geq 1}$ by $\mathcal{E}_i := N_{\tau_i} - N_{\tau_i}$. With these notations, we have to show that

$$D_s \sum_{i \ge 1} \xi_{\tau_i} \vartheta_{\tau_i}(\mathcal{E}_i) = \sum_{i \ge 1} (D_s \xi)_{\tau_i} \vartheta_{\tau_i}(\mathcal{E}_i) .$$
(II.32)

Using Remark 1.3.2, we first observe that, for each $n \ge 1$,

$$D_s \sum_{i=1}^n \xi_{\tau_i} \vartheta_{\tau_i}(\mathcal{E}_i) = \sum_{i=1}^n (D_s \xi)_{\tau_i} \vartheta_{\tau_i}(\mathcal{E}_i) .$$

Passing to the limit in $L^2(\Omega \times [0,1])$ leads to (II.32) and concludes the proof.

Remark 1.3.3 Similar arguments as in the above proof shows that for $\psi \in \mathbf{L}^2_{\lambda}(\mathbb{D}^{1,2})$ and $f \in L^{\infty}(E)$, we have, for almost every $s \leq 1$,

$$\int_E \psi_s(e) f(e) \lambda(de) \in \mathbb{D}^{1,2}$$

and

$$D_t\left(\int_E \psi_s(e)f(e)\lambda(de)\right) := \int_E D_t\psi_s(e)f(e)\lambda(de)$$

Lemma 1.3.2 Let $\mathbf{S}(W)$ denote the set of random variables of the form

$$H^W = \phi\left(\int_0^1 f^1(t) \cdot dW_t, \dots, \int_0^1 f^\kappa(t) \cdot dW_t\right)$$

where $\kappa \geq 1$, $\phi \in C_b^{\infty}$ and $f^i : [0,1] \mapsto \mathbb{R}^d$ is a bounded measurable map for each $i \leq \kappa$. Then, $\operatorname{Vect}\{\mathbf{S}(W) \times L^{\infty}(\Omega_{\mu}, \mathcal{F}_1^{\mu})\}$ is dense in $\mathbb{D}^{1,2}$ for the norm $\|\cdot\|_{\mathbb{D}^{1,2}}$.

Proof. It suffices to prove that $\operatorname{Vect}\{\mathbf{S}(W) \times L^{\infty}(\Omega_{\mu}, \mathcal{F}_{1}^{\mu})\}$ is dense in **S**. Fix $H \in \mathbf{S}$ of the form

$$H = \phi\left(\int_0^1 f^1(t) \cdot dW_t, \dots, \int_0^1 f^{\kappa}(t) \cdot dW_t, \mu\right) .$$

Observe that Ω_{μ} can be identified to the space of finite (possibly empty) sequences $(t_i, e_i)_{i\geq 1}$ of $[0, 1] \times E$ such that $(t_i)_{i\geq 1}$ is increasing. Given $\eta \in \Omega_u$, we denote by $(t_i^{\eta}, e_i^{\eta})_{i\geq 1}$ the associated sequence, and we identify ϕ with a measurable map defined on $\mathbb{R}^{\kappa} \times ([0, 1] \times E)^{\mathbb{N}}$. We denote by ϕ_n its restriction to $\mathbb{R}^{\kappa} \times ([0, 1] \times E)^n$, $n \geq 0$. Let ψ_n denote the gradient of ϕ_n with respect to its first κ components and set $f := (f^1, \ldots, f^{\kappa})$ and $G := \left(\int_0^1 f^1(t) \cdot dW_t, \ldots, \int_0^1 f^{\kappa}(t) \cdot dW_t\right)$. Since

$$(H, D_s H) = \sum_{n \ge 0} (\phi_n (G, (t_i^{\mu}, e_i^{\mu})_{1 \le i \le n}), \psi_n (G, (t_i^{\mu}, e_i^{\mu})_{1 \le i \le n}) \cdot f(s)) \mathbf{1}_{\mu(E, [0, 1]) = n},$$

it suffices to prove that each $H_n := \phi_n (G, (t_i^{\mu}, e_i^{\mu})_{1 \le i \le n})$ can be approximated by linear combinations of elements of $\mathbf{S}(W) \times L^{\infty}(\Omega_{\mu}, \mathcal{F}_1^{\mu})$. Moreover, we can always assume that ϕ_n is C_b^{∞} on $\mathbb{R}^{\kappa} \times ([0, 1] \times E)^n$. Indeed, ϕ is already C_b^{∞} in its first κ components, a.e., and we can replace ϕ_n by its convolution with a sequence of smooth kernels acting only its last n components. Since both functions are continuous, we can then approximate (ϕ_n, ψ_n) pointwise by linear combinations of functions of the form $(\phi_n, \psi_n)(\cdot, (t_i, e_i)_{1 \le i \le n})\mathbf{1}_A$ where A is a Borel set of $([0,1] \times E)^n$ and $(t_i, e_i)_{1 \le i \le n} \in ([0,1] \times E)^n$. The required result then follows from the fact that

$$D_s \phi_n \left(G, (t_i, e_i)_{1 \le i \le n} \right) \mathbf{1}_A \left((t_i^{\mu}, e_i^{\mu})_{1 \le i \le n} \right) = \left(\psi_n \left(G, (t_i, e_i)_{1 \le i \le n} \right) \cdot f(s) \right) \mathbf{1}_A \left((t_i^{\mu}, e_i^{\mu})_{1 \le i \le n} \right).$$

Lemma 1.3.3 Fix $(\xi, \psi) \in \mathbf{H}^2 \times \mathbf{L}^2_{\lambda}$ and assume that

$$H := \int_0^1 \xi_t \cdot dW_t + \int_0^1 \int_E \psi_t(e) \bar{\mu}(de, dt) \in \mathbb{D}^{1,2} .$$

Then, $(\xi, \psi) \in \mathbf{H}^2(\mathrm{ID}^{1,2}) \times \mathbf{L}^2_{\lambda}(\mathrm{ID}^{1,2})$ and

$$D_s H := \xi_s^* + \int_0^1 \sum_{i=1}^d D_s \xi_t^i \, dW_t^i + \int_0^1 \int_E D_s \psi_t(e) \bar{\mu}(de, dt) \,,$$

where ξ^* denotes the transpose of ξ .

Proof. One easily deduces from Lemma 1.3.2 that

$$\mathcal{H} := \operatorname{Vect} \left\{ H^W H^{\bar{\mu}} : H^W \in \mathbf{S}(W) , H^{\bar{\mu}} \in L^{\infty}(\Omega_{\mu}, \mathcal{F}_1^{\mu}) , \mathbb{E} \left[H^W H^{\bar{\mu}} \right] = 0 \right\}$$

is dense in $\mathbb{D}^{1,2} \cap \{H \in L^2(\Omega, \mathcal{F}, \mathbb{P}) : \mathbb{E}[H] = 0\}$ for $\|\cdot\|_{\mathbb{D}^{1,2}}$. Thus, it suffices to prove the result for H of the form $H^W H^{\bar{\mu}}$ where $H^W \in \mathbf{S}(W), H^{\bar{\mu}} \in L^{\infty}(\Omega_{\mu}, \mathcal{F}_1^{\mu})$ and $\mathbb{E}[H^W H^{\bar{\mu}}] = 0$. By the representation theorem, there exists $\psi \in \mathbf{L}^2_{\lambda}$ such that

$$H^{\bar{\mu}} = \mathbb{E}\left[H^{\bar{\mu}}\right] + \int_0^1 \int_E \psi_t(e)\bar{\mu}(de, dt)$$

and by Ocone's formula, see e.g. Proposition 1.3.5 in [82],

$$H^{W} = \mathbb{E}\left[H^{W}\right] + \int_{0}^{1} \mathbb{E}\left[D_{t}H^{W} \mid \mathcal{F}_{t}^{W}\right] dW_{t} .$$

Thus it follows from Itô's Lemma that

$$H = \int_0^1 H_t^{\bar{\mu}} \mathbb{E}\left[D_t H^W \mid \mathcal{F}_t^W\right] dW_t + \int_0^1 \int_E H_t^W \psi_t(e) \bar{\mu}(de, dt)$$

where $H_t^{\bar{\mu}} = \mathbb{E}[H^{\bar{\mu}} | \mathcal{F}_t]$ and $H_t^W = \mathbb{E}[H^W | \mathcal{F}_t]$. Furthermore, easy computations show that the two integrands belong respectively to $\mathbf{H}^2(\mathbb{D}^{1,2})$ and $\mathbf{L}^2_{\lambda}(\mathbb{D}^{1,2})$. Thus, Remark 1.3.1 and Lemma 1.3.1 conclude the proof.

1.3.2 Malliavin calculus on the Forward SDE

In this section, we recall well-known properties concerning the differentiability in the Malliavin sense of the solution of a Forward SDE. In the case where $\beta = 0$ the following result is stated in e.g. [82]. The extension to the case $\beta \neq 0$ is easily obtained by conditioning by μ , see e.g. [49] for explanations in the case where E is finite, or by combining Remark 1.3.1, Lemma 1.3.1 with a fixed point procedure as in the proof of Theorem 2.2.1. in [82], see also Proposition 1.3.2 below.

Proposition 1.3.1 Assume that \mathbf{C}_1^X holds, then $X_t \in \mathbb{D}^{1,2}$ for all $t \leq 1$. For all $s \leq 1$ and $k \leq d$, $D_s^k X$ admits a version $\chi^{s,k}$ which solves on [s, 1]

$$\chi_t^{s,k} = \sigma^k(X_{s-}) + \int_s^t \nabla b(X_r) \chi_r^{s,k} dr + \int_s^t \sum_{j=1}^d \nabla \sigma^j(X_r) \chi_r^{s,k} dW_r^j$$
$$+ \int_s^t \int_E \nabla \beta(X_{r-}, e) \chi_{r-}^{s,k} \bar{\mu}(dr, de) .$$

If moreover \mathbf{C}_2^X holds, then $D_s^k X_t \in \mathbb{D}^{1,2}$ for all $s, t \leq 1$ and $k \leq d$. For all $u \leq 1$ and $\ell \leq d$, $D_u^\ell D_s^k X$ admits a version $\chi^{u,\ell,s,k}$ which solves on $[u \lor s, 1]$

$$\begin{aligned} \chi_{t}^{u,\ell,s,k} &= \nabla \sigma^{k}(X_{s-})\chi_{s-}^{u,\ell} + \nabla \sigma^{\ell}(X_{u-})\chi_{u-}^{s,k} \\ &+ \int_{s}^{t} \left(\nabla b(X_{r})\chi_{r}^{u,\ell,s,k} + \sum_{i=1}^{d} \nabla (\nabla b(X_{r}))^{i}\chi_{r}^{u,\ell}(\chi_{r}^{s,k})^{i} \right) dr \\ &+ \int_{s}^{t} \sum_{j=1}^{d} \left(\nabla \sigma^{j}(X_{r})\chi_{r}^{u,\ell,s,k} + \sum_{i=1}^{d} \nabla (\nabla \sigma^{j}(X_{r}))^{i}\chi_{r}^{u,\ell}(\chi_{r}^{s,k})^{i} \right) dW_{r}^{j} \end{aligned}$$
(II.33)
$$+ \int_{s}^{t} \int_{E} \left(\nabla \beta(X_{r-},e)\chi_{r-}^{u,\ell,s,k} + \sum_{i=1}^{d} \nabla (\nabla \beta(X_{r-},e))^{i}\chi_{r-}^{u,\ell}(\chi_{r-}^{s,k})^{i} \right) \bar{\mu}(dr,de) . \end{aligned}$$

Remark 1.3.4 Fix $p \ge 2$ and $r \le s \le t \le u \le 1$. Under \mathbf{C}_1^X , it follows from Lemma 1.5.1 applied to X and χ^s that

$$\|\chi^{s}\|_{\mathcal{S}^{p}}^{p} \leq C_{p} (1+|X_{0}|^{p})$$
(II.34)

$$\mathbb{E}\left[|\chi_{u}^{s} - \chi_{t}^{s}|^{p}\right] \leq C_{p} |u - t| (1 + |X_{0}|^{p})$$
(II.35)

$$\|\chi^s - \chi^r\|_{\mathcal{S}^p}^p \leq C_p |s - r| (1 + |X_0|^p) .$$
 (II.36)

If moreover \mathbf{C}_2^X holds then similar arguments show that

$$\|\chi^{r,s}\|_{\mathcal{S}^p}^p \leq C_p \left(1 + |X_0|^{2p}\right),$$
 (II.37)

where $\chi^{r,s} = (\chi^{r,\ell,s,k})_{\ell,k \leq d}$.

Remark 1.3.5 Under \mathbf{C}_1^X , we can define the first variation process ∇X of X which solves on [0, 1]

$$\nabla X_t = I_d + \int_0^t \nabla b(X_r) \nabla X_r dr + \int_0^t \sum_{j=1}^d \nabla \sigma^j(X_r) \nabla X_r dW_r^j + \int_0^t \int_E \nabla \beta(X_{r-}, e) \nabla X_{r-} \bar{\mu}(dr, de) .$$
(II.38)

Moreover, under \mathbf{H}_1 , see Remark 1.2.7, $(\nabla X)^{-1}$ is well defined and solves on [0, 1]

$$(\nabla X)_{t}^{-1} = I_{d} - \int_{0}^{t} (\nabla X)_{r}^{-1} \left[\nabla b(X_{r}) - \sum_{j=1}^{d} \nabla \sigma^{j}(X_{r}) \nabla \sigma^{j}(X_{r}) \right] dr + \int_{0}^{t} (\nabla X)_{r}^{-1} \int_{E} \nabla \beta(X_{r}, e) \lambda(de) dr - \int_{0}^{t} \sum_{j=1}^{d} (\nabla X)_{r}^{-1} \nabla \sigma^{j}(X_{r}) dW_{r}^{j} - \int_{0}^{t} \int_{E} (\nabla X)_{r-}^{-1} (\nabla \beta(X_{r-}, e) + I_{d})^{-1} \nabla \beta(X_{r-}, e) \mu(de, dr) .$$
 (II.39)

This can be checked by simply applying Itô's Lemma to the product $\nabla X(\nabla X)^{-1}$, see [82] p. 109 for the case where $\beta = 0$.

Remark 1.3.6 Fix $p \ge 2$. Under \mathbf{H}_1 - \mathbf{C}_1^X , it follows from Remark 1.2.7 and Lemma 1.5.1 applied to ∇X and $(\nabla X)^{-1}$ that

$$\|\nabla X\|_{\mathcal{S}^p} + \|(\nabla X)^{-1}\|_{\mathcal{S}^p} \le C_p$$
 (II.40)

Remark 1.3.7 Assume that \mathbf{H}_1 - \mathbf{C}_1^X holds and observe that $\chi^s = (\chi^{s,k})_{k \leq d}$ and ∇X solve the same equation up to the condition at time s. By uniqueness of the solution on [t, 1], it follows that

$$\chi_r^s = \nabla X_r (\nabla X_{s-})^{-1} \sigma(X_{s-}) \mathbf{1}_{s \le r} \quad \text{for all } s, r \le 1 .$$
 (II.41)

1.3.3 Malliavin calculus on the Backward SDE

In this section, we generalize the result of Proposition 3.1 in [86]. Let us denote by $\mathcal{B}^2(\mathbb{D}^{1,2})$ the set of triples $(Y, Z, U) \in \mathcal{B}^2$ such that $Y_t \in \mathbb{D}^{1,2}$, for any $t \leq 1$, and the process $(Z, U) \in \mathbf{H}^2(\mathbb{D}^{1,2}) \times \mathbf{L}^2_{\lambda}(\mathbb{D}^{1,2})$.

Proposition 1.3.2 Assume that $\mathbf{C}_1^X \cdot \mathbf{C}_1^Y$ holds.

(i) The triples (Y, Z, U) belongs to $\mathcal{B}^2(\mathbb{D}^{1,2})$. For each $s \leq 1$ and $k \leq d$, the equation

$$\Upsilon_{t}^{s,k} = \nabla g(X_{1})\chi_{1}^{s,k} + \int_{t}^{1} \nabla h(\Theta_{r})\Phi_{r}^{s,k}dr - \int_{t}^{1} \zeta_{r}^{s,k} \cdot dW_{r} - \int_{t}^{1} \int_{E} V_{r}^{s,k}(e)\bar{\mu}(de,dr),$$
(II.42)

with $\Phi^{s,k} := (\chi^{s,k}, \Upsilon^{s,k}, \zeta^{s,k}, \Gamma^{s,k})$ and $\Gamma^{s,k} := \int_E \rho(e) V^{s,k}(e) \lambda(de)$, admits a unique solution. Moreover, $(\Upsilon^{s,k}_t, \zeta^{s,k}_t, V^{s,k}_t)_{s,t \leq 1}$ is a version of $(D_s^k Y_t, D_s^k Z_t, D_s^k U_t)_{s,t \leq 1}$.

(ii) Assume further that $\mathbf{C}_2^X \cdot \mathbf{C}_2^Y$ holds. Then, for each $s \leq 1$ and $k \leq d$, $(D_s^k Y, D_s^k Z, D_s^k U)$ belongs to $\mathcal{B}^2(\mathbb{D}^{1,2})$. For each $u \leq 1$ and $\ell \leq d$, the equation

$$\Upsilon_t^{u,\ell,s,k} = \left(\chi_1^{u,\ell}\right)' [Hg](X_1)\chi_1^{s,k} + \nabla g(X_1)\chi_1^{u,\ell,s,k} + \int_t^1 \left[\nabla h(\Theta_r)\Phi^{u,\ell,s,k} + \left(D_u^{\ell}\Theta_r\right)' [Hh](\Theta_r)D_s^k\Theta_r\right] dr - \int_t^1 \zeta^{u,\ell,s,k} \cdot dW_r - \int_t^1 V_r^{u,\ell,s,k}(e)\bar{\mu}(de,dr), \qquad (\text{II.43})$$

where $\Phi^{u,\ell,s,k} := (\chi^{u,\ell,s,k}, \Upsilon^{u,\ell,s,k}, \zeta^{u,\ell,s,k}, \Gamma^{u,\ell,s,k})$ with $\Gamma^{u,\ell,s,k} := \int_E \rho(e) V^{u,\ell,s,k}(e) \lambda(de)$, and [Hg] (resp. [Hh]) denotes the Hessian matrix of g (resp. h), admits a unique solution. Moreover, $(\Upsilon^{u,\ell,s,k}_t, \zeta^{u,\ell,s,k}_t, V^{u,\ell,s,k}_t)_{u,s,t\leq 1}$ is a version of $(D^\ell_u D^k_s(Y_t, Z_t, U_t))_{u,s,t\leq 1}$.

Proof. For ease of notations, we only consider the case d = 1 and omit the indexes k and ℓ in the above notations.

(i) We proceed as in Proposition 5.3 in [47]. Combined with \mathbf{C}_X^1 - \mathbf{C}_Y^1 and (II.34), Lemma 1.5.2 implies that $(\Upsilon^s, \zeta^s, V^s)$ is well defined for each $s \leq 1$ and that we have

$$\sup_{s \le 1} \|(\Upsilon^s, \zeta^s, V^s)\|_{\mathcal{B}^p}^p \le C_p \ (1 + |X_0|^p) \quad \text{for all} \ p \ge 2.$$
(II.44)

We now define recursively the sequence $\Theta^n := (X, Y^n, Z^n, \Gamma^n)$ as follows. First, we set $\Theta^0 := (0, 0, 0)$. Then, given Θ^{n-1} , we define (Y^n, Z^n, U^n) as the unique solution in \mathcal{B}^2 of

$$Y_t^n = g(X_1) + \int_t^1 h(\Theta_r^{n-1}) dr - \int_t^1 Z_r^n dW_r - \int_t^1 \int_E U_r^n(e) \bar{\mu}(de, dr)$$

and set $\Gamma^n = \int_E \rho(e) U^n(e) \lambda(de)$. From the proof of Lemma 2.4 in [100], $(Y^n, Z^n, U^n)_n$ is a Cauchy sequence in \mathcal{B}^2 which converges to (Y, Z, U).

Moreover, using Proposition 1.3.1, Remark 1.3.1, Remark 1.3.3, Lemma 1.3.3 and an inductive argument, one obtains that $(Y^n, Z^n, U^n) \in \mathcal{B}^2(\mathbb{ID}^{1,2})$. For $s \leq 1$, set

$$\begin{split} & (\Upsilon^{s,n},\zeta^{s,n},V^{s,n}) := (D_sY^n,D_sZ^n,D_sU^n) \ , \ \Phi^{s,n} := (\chi^s,\Upsilon^{s,n},\zeta^{s,n},\Gamma^{s,n}) \ , \\ & \Xi^{s,n} := (\chi^s,\Upsilon^{s,n},\zeta^{s,n},U^{s,n}) \ \text{and} \ \Xi^s := (\chi^s,\Upsilon^s,\zeta^s,U^s) \ , \end{split}$$

where $\Gamma^{s,n} := \int_E \rho(e) V^{s,n}(e) \lambda(de)$. By Proposition 1.3.1, Remark 1.3.1, Lemma 1.3.3 and Remark 1.3.3, we have

$$\Upsilon_t^{s,n} = \nabla g(X_1)\chi_1^s + \int_t^1 \nabla h(\Theta_r^{n-1})\Phi_r^{s,n-1}dr - \int_t^1 \zeta_r^{s,n}dW_r - \int_t^1 V_r^{s,n}(e)\bar{\mu}(de,dr) .$$
(II.45)

Fix $I \in \mathbb{N}$ to be chosen later, set $\delta := 1/I$ and $\tau_i := i\delta$ for $0 \le i \le I$. By (II.88) of Lemma 1.5.2, we have

$$G_{i}^{s,n} := \|\Xi^{s} - \Xi^{s,n}\|_{\mathcal{S}^{4} \times \mathcal{B}_{[\tau_{i},\tau_{i+1}]}^{4}} \leq C_{4} \left(\mathbb{E} \left[|\Upsilon_{\tau_{i+1}}^{s} - \Upsilon_{\tau_{i+1}}^{s,n}|^{4} \right] + A_{i}^{s,n-1} + B_{i}^{s,n-1} \right), (\text{II.46})$$

where

$$\begin{split} A_i^{s,n-1} &:= & \left\| \{ \nabla h(\Theta^{n-1}) - \nabla h(\Theta) \} \Phi^s \right\|_{\mathbf{H}^4_{[\tau_i,\tau_{i+1}]}}^4 \\ B_i^{s,n-1} &:= & \mathbb{E} \left[\left(\int_{\tau_i}^{\tau_{i+1}} \nabla h(\Theta^{n-1}_r) \{ \Phi^s_r - \Phi^{s,n-1}_r \} dr \right)^4 \right] \,. \end{split}$$

Recalling that ρ and the derivatives of h are bounded, we deduce from Cauchy-Schwarz and Jensen's inequality that

$$B_i^{s,n-1} \leq C_4 \delta^2 G_i^{s,n-1},$$
 (II.47)

which combined with an inductive argument and (II.44)-(II.46) leads to

$$\sup_{s \le 1} G_i^{s,n} < \infty \quad \text{for all } n \ge 0.$$
 (II.48)

Since the derivatives of h are also continuous and Θ^{n-1} converges to Θ in $\mathcal{S}^2 \times \mathcal{B}^2$, we deduce from (II.34)-(II.44) that, after possibly passing to a subsequence,

$$\lim_{n \to \infty} \sup_{s \le 1} A_i^{s, n-1} = 0.$$
 (II.49)

It follows from (II.46)-(II.47)-(II.49) that for I large enough there is some $\alpha < 1$ such that for any $\varepsilon > 0$ we can find $N' \ge 0$, independent of s, such that

$$G_i^{s,n} \leq C_4 \mathbb{E}\left[|\Upsilon_{\tau_{i+1}}^s - \Upsilon_{\tau_{i+1}}^{s,n-1}|^4\right] + \varepsilon + \alpha G_i^{s,n-1} \quad \text{for } n \geq N' .$$
(II.50)

Since $\Upsilon_1^s = \Upsilon_1^{s,n-1}$, we deduce that for i = I - 1 and $n \ge N'$

$$\sup_{s \le 1} G_{I-1}^{s,n} \le \varepsilon + \alpha^{n-N'} \sup_{s \le 1} G_{I-1}^{s,N'} .$$

By (II.48), it follows that $\sup_{s\leq 1} G_{I-1}^{s,n} \to 0$ as $n \to \infty$. In view of (II.50), a straightforward induction argument shows that, for all $i \leq I-1$, $\sup_{s\leq 1} G_i^{s,n} \to 0$ as $n \to \infty$ so that, summing up over i, we get

$$\sup_{s \le 1} \| (\Xi^s - \Xi^{s,n}) \|_{\mathcal{S}^4 \times \mathcal{B}^4} \xrightarrow[n \to \infty]{} 0.$$
 (II.51)

Since (Y^n, Z^n, U^n) converges to (Y, Z, U) in \mathcal{B}^2 , this shows that $(Y, Z, U) \in \mathcal{B}^2(\mathbb{ID}^{1,2})$ and that there is a version of (DY, DZ, DU) given by (Υ, ζ, V) .

(ii) In view of (II.34)-(II.37)-(II.44) and $\mathbf{C}_2^X \cdot \mathbf{C}_2^Y$, it follows from Lemma 1.5.2 that $(\Upsilon^{u,s}, \zeta^{u,s}, V^{u,s})$ is well defined for $u, s \leq 1$ and that we have

$$\sup_{u,s\leq 1} \|(\Upsilon^{u,s},\zeta^{u,s},V^{u,s})\|_{\mathcal{B}^p}^p \leq C_p (1+|X_0|^{2p}) \quad \text{for all } p\geq 2.$$
(II.52)

Using Lemma 1.3.3, (II.45) and an inductive argument, we then deduce that we have $(DY^n, DZ^n, DU^n) \in \mathcal{B}^2(\mathbb{ID}^{1,2})$ and

$$\begin{split} \Upsilon_t^{u,s,n} &= \chi_1^u[Hg](X_1)\chi_1^s + \nabla g(X_1)\chi_1^{u,s} + \int_t^1 \nabla h(\Theta_r^{n-1})\Phi_r^{u,s,n-1}dr \\ &+ \int_t^1 \Phi_r^{u,n-1}[Hh](\Theta_r^{n-1})\Phi_r^{s,n-1}dr - \int_t^1 \zeta_r^{u,s,n}dW_r - \int_t^1 V_r^{u,s,n}(e)\bar{\mu}(de,dr) \end{split},$$

where $(\Upsilon^{u,s,n}, \zeta^{u,s,n}, V^{u,s,n}, \Phi^{u,s,n}) := D_u(\Upsilon^{s,n}, \zeta^{s,n}, V^{s,n}, \Phi^{s,n})$. By (i), (Υ^n, Z^n, U^n) goes to (Y, Z, U) in \mathcal{B}^2 and $(\Upsilon^{s,n}, \zeta^{s,n}, V^{s,n})$ converges to $(\Upsilon^s, \zeta^s, V^s)$ in \mathcal{B}^4 . Moreover, (II.51) implies

$$\sup_{n \ge 1} \sup_{s \le 1} \|(\Upsilon^{s,n}, \zeta^{s,n}, V^{s,n})\|_{\mathcal{B}^4}^4 < \infty , \qquad (\text{II.53})$$

so that, by dominated convergence, \mathbf{C}_{2}^{Y} and (II.52),

$$\|\Phi^{u,n}[Hh](\Theta^n)\Phi^{s,n} - \Phi^u[Hh](\Theta)\Phi^s\|_{\mathbf{H}^2} + \|(\nabla h(\Theta^n) - \nabla h(\Theta))\Phi^{u,s}\|_{\mathbf{H}^2} \xrightarrow[n \to \infty]{} 0,$$

after possibly passing to a subsequence. The rest of the proof follows step by step the arguments of (i) except that we now work on $S^2 \times B^2$ instead of $S^4 \times B^4$. \Box

Proposition 1.3.3 Assume that $\mathbf{C}_1^X \cdot \mathbf{C}_1^Y$ holds. For each $k \leq d$, the equation

$$\nabla Y_t^k = \nabla g(X_1) \nabla X_1^k + \int_t^1 \nabla h(\Theta_r) \nabla \Phi_r^k dr - \int_t^1 \nabla Z_r^k \cdot dW_r$$

-
$$\int_t^1 \int_E \nabla U_r^k(e) \bar{\mu}(de, dr), \qquad (\text{II.54})$$

with $\nabla \Phi^k = (\nabla X^k, \nabla Y^k, \nabla Z^k, \nabla \Gamma^k)$ and $\nabla \Gamma^k := \int_E \rho(e) \nabla U^k(e) \lambda(de)$, admits a unique solution $(\nabla Y^k, \nabla Z^k, \nabla U^k)$. Moreover, there is a version of $(\zeta_t^{s,k}, \Upsilon_t^{s,k}, V_t^{s,k})_{s,t\leq 1}$ given by $\{(\nabla Y_t, \nabla Z_t, \nabla U_t)(\nabla X_{s-})^{-1}\sigma^k(X_{s-})\mathbf{1}_{s\leq t}\}_{s,t\leq 1}$ where ∇Y_t denotes the matrix whose k-column is given by ∇Y_t^k and $\nabla Z_t, \nabla U_t$ are defined similarly.

Proof. In view of Proposition 1.3.2 and (II.41), this follows immediately from the uniqueness of the solution of (II.42). \Box

Remark 1.3.8 It follows from Lemma 1.5.2 and (II.40) that

$$\|(\nabla Y, \nabla Z, \nabla U)\|_{\mathcal{B}^p} \leq C_p \quad \text{for all } p \geq 2.$$
 (II.55)

1.4 Representation results and path regularity for the BSDE

In this section, we use the above results to obtain some regularity for the solution of the BSDE (II.5) under $\mathbf{C}_1^X \cdot \mathbf{C}_1^Y$, $\mathbf{C}_1^X \cdot \mathbf{C}_1^Y \cdot \mathbf{H}_1$ or $\mathbf{C}_2^X \cdot \mathbf{C}_2^Y$. Similar results without $\mathbf{C}_1^X \cdot \mathbf{C}_1^Y$ or with \mathbf{H}_2 instead of $\mathbf{C}_2^X \cdot \mathbf{C}_2^Y$ will then be obtained by using an approximation argument. Fix $(u, s, t, x) \in [0, 1]^3 \times \mathbb{R}^d$ and $k, \ell \leq d$. In the sequel, we shall denote by X(t, x) the solution of (II.4) on [t, 1] with initial condition $X(t, x)_t = x$, and by (Y(t, x), Z(t, x), U(t, x)) the solution of (II.5) with X(t, x) in place of X. We define similarly $(\Upsilon^{s,k}(t, x), \zeta^{s,k}(t, x), \nabla^{s,k}(t, x))$, $(\nabla Y(t, x), \nabla Z(t, x), \nabla U(t, x))$ and $(\Upsilon^{u,\ell,s,k}(t, x), \zeta^{u,\ell,s,k}(t, x), V^{u,\ell,s,k}(t, x))$. Observe finally that, with these notations, we have

 $(X(0, X_0), Y(0, X_0), Z(0, X_0), U(0, X_0)) = (X, Y, Z, U).$

1.4.1 Representation

We start this section by proving useful bounds for the (deterministic) maps defined on $[0,1] \times \mathbb{R}^d$ by

$$u(t,x) := Y(t,x)_t , \ \nabla u(t,x) := \nabla Y(t,x)_t , \ v^{s,k}(t,x) := \Upsilon^{s,k}(t,x)_t$$

and $w^{u,\ell,s,k}(t,x) := \Upsilon^{u,\ell,s,k}(t,x)_t$,

where $(u, s) \in [0, 1]^2$ and $k, \ell \leq d$.

Proposition 1.4.1 (i) Assume that \mathbf{C}_1^X and \mathbf{C}_1^Y hold, then,

$$|u(t,x)| + |v^{s,k}(t,x)| \le C_2 (1+|x|) \quad and \quad |\nabla u(t,x)| \le C_2$$
 (II.56)

for all $s, t \leq 1, k \leq d$ and $x \in \mathbb{R}^d$.

(ii) Assume that \mathbf{C}_2^X and \mathbf{C}_2^Y hold, then,

$$|w^{u,\ell,s,k}(t,x)| \le C_2 \ (1+|x|^2) \ , \tag{II.57}$$

for all $u, s, t \leq 1, \ell, k \leq d$ and $x \in \mathbb{R}^d$.

Proof. When $(t, x) = (0, X_0)$, the result follows from (II.7) in Remark 1.2.1, (II.44), (II.52) and (II.55). The general case is obtained similarly by changing the initial condition on X.

Proposition 1.4.2 Assume that C₁^X and C₁^Y hold.
(i) There is a version of Z given by (Υ^t_t)_{t<1} which satisfies

$$||Z||_{\mathcal{S}^p}^p \leq C_p \left(1 + |X_0|^p\right).$$
 (II.58)

(ii) Assume further that \mathbf{C}_2^X and \mathbf{C}_2^Y hold, then, for each $k \leq d$, there is a version of $(\zeta^{s,k})_t)_{s,t\leq 1}$ given by $((\Upsilon_t^{t,\ell,s,k})_{\ell\leq d})_{s,t\leq 1}$ which satisfies

$$\|\sup_{s\leq 1} |\zeta^{s,k}| \|_{\mathcal{S}^p}^p \leq C_p \left(1 + |X_0|^{2p}\right).$$
 (II.59)

Proof. Here again we only consider the case d = 1 and omit the indexes k, ℓ . By Proposition 1.3.2, (Y, Z, U) belongs to $\mathcal{B}^2(\mathbb{ID}^{1,2})$ and it follows from Lemma 1.3.3 that

$$D_s Y_t = Z_s - \int_s^t \nabla h(\Theta_r) D_s \Theta_r dr + \int_s^t D_s Z_r \, dW_r + \int_s^t D_s U_r(e) \bar{\mu}(de, dr) , \quad (\text{II.60})$$

for $0 < s \leq t \leq 1$. Taking s = t leads to the representation of Z. Thus, after possibly passing to a suitable version, we have $Z_t = D_t Y_t = \Upsilon_t^t$. By uniqueness of the solution of (II.4)-(II.5)-(II.42) for any initial condition in $L^2(\Omega, \mathcal{F}_t)$ at t, we have $\Upsilon_t^t = v^t(t, X_t)$. The bound on Z then follows from Proposition 1.4.1 combined with (II.7) of Remark 1.2.1. Under \mathbb{C}_2^X and \mathbb{C}_2^Y , the same arguments applied to $(\Upsilon^s, \zeta^s, V^s)$ instead of (Y, Z, U)lead to the second claim, see (ii) of Proposition 1.3.2, (ii) of Proposition 1.4.1 and recall (II.7).

Proposition 1.4.3 (i) Define \tilde{U} by

$$\tilde{U}_t(e) := u(t, X_{t-} + \beta(X_{t-}, e)) - \lim_{r \uparrow t} u(r, X_r) = Y_t - Y_{t-}.$$

Then \tilde{U} is a version of U and it satisfies

$$\|\sup_{e \in E} |\tilde{U}(e)| \|_{\mathcal{S}^p}^p \leq C_p (1 + |X_0|^p) .$$
 (II.61)

(ii) Assume that \mathbf{C}_1^X and \mathbf{C}_1^Y hold. Define $\nabla \tilde{U}$ by

$$\nabla \tilde{U}_t(e) := \nabla u\left(t, X_{t-} + \beta(X_{t-}, e)\right) - \lim_{r \uparrow t} \nabla u\left(r, X_r\right) \ .$$

Then $\nabla \tilde{U}$ is a version of ∇U and it satisfies

$$\|\sup_{e \in E} |\nabla \tilde{U}(e)| \|_{\mathcal{S}^p}^p \leq C_p .$$
(II.62)

(iii) Assume that \mathbf{C}_1^X and \mathbf{C}_1^Y hold, then, for each $k \leq d$, there is a version of $(V_t^{s,k})_{s,t\leq 1}$ given by $(\tilde{V}_t^{s,k})_{s,t\leq 1}$ defined as

$$\tilde{V}_t^{s,k}(e) := v^{s,k} (t, X_{t-} + \beta(X_{t-}, e)) - \lim_{r \uparrow t} v^{s,k} (r, X_r) .$$

It satisfies

$$\|\sup_{e \in E} \sup_{s \le 1} |\tilde{V}^{s,k}(e)| \|_{\mathcal{S}^p}^p \le C_p (1+|X_0|^p).$$
(II.63)

Remark 1.4.1 We will see in Proposition 1.4.4 below that u is continuous under \mathbf{C}_1^X and \mathbf{C}_1^Y so that

$$\tilde{U}_t(e) := u(t, X_{t-} + \beta(X_{t-}, e)) - u(t, X_{t-}) .$$

A similar representation is derived in [86] in a case where E is finite. One could similarly show that $v^{s,k}$ and ∇u are continuous under \mathbf{C}_2^X and \mathbf{C}_2^Y so that

$$\tilde{V}_{t}^{s,k}(e) := v^{s,k} (t, X_{t-} + \beta(X_{t-}, e)) - v^{s,k} (t, X_{t-})
\nabla \tilde{U}_{t}(e) := \nabla u (t, X_{t-} + \beta(X_{t-}, e)) - \nabla u (t, X_{t-}) .$$

However, since this result is not required for our main theorem, we do not provide its proof.

Proof of Proposition 1.4.3. We only provide the proof of (i), the other assertions are proved similarly.

1. By uniqueness of the solution of (II.4)-(II.5) for any initial condition in $L^2(\Omega, \mathcal{F}_t)$ at time t, one has $Y_t = u(t, X_t)$ a.s. for each $t \leq 1$. We shall prove in step 2. below that u is jointly continuous in x and right-continuous in t. This implies that $(u(t, X_t))_{t \leq 1}$ is right-continuous so that $Y_t = u(t, X_t)$ and $Y_{t-} = \lim_{r \uparrow t} u(r, X_r)$ for each $t \leq 1$ a.s., see Theorem I.2 in [92] and recall that X and Y are càdlàg. Thus

$$\int_E U_t(e)\mu(de, \{t\}) = Y_t - Y_{t-} = u(t, X_t) - \lim_{r \uparrow t} u(r, X_r) = \int_E \tilde{U}_t(e)\mu(de, \{t\}) ,$$

for each $t \leq 1$ a.s. and

$$\int_0^1 \int_E \left| \tilde{U}_t(e) - U_t(e) \right|^2 \mu(de, dt) = 0 ,$$

which, by taking expectation, implies

$$\mathbb{E}\left[\int_0^1 \int_E \left|\tilde{U}_t(e) - U_t(e)\right|^2 \lambda(de) dt\right] = 0.$$

2. We now prove that u is continuous in x and right-continuous on t. Fix $0 \le t_1 \le t_2 \le 1$ and $(x_1, x_2) \in \mathbb{R}^{2d}$. For A denoting X, Y, Z or U, we set $A^i := A(t_i, x_i)$ for i = 1, 2 and $\delta A := A^1 - A^2$. By (II.84) of Lemma 1.5.1, we derive

$$\|\delta X\|_{\mathcal{S}^{2}_{[t_{2},1]}}^{2} \leq C_{2} \{|x_{1} - x_{2}|^{2} + (1 + |x_{1}|^{2})|t_{2} - t_{1}|\} .$$
(II.64)

Plugging this estimate in (II.88) of Lemma 1.5.2 leads to

$$\|(\delta Y, \delta Z, \delta U)\|_{\mathcal{B}^{2}_{[t_{2}, 1]}}^{2} \leq C_{2} \{|x_{1} - x_{2}|^{2} + (1 + |x_{1}|^{2})|t_{2} - t_{1}|\} .$$
(II.65)

Now, observe that

$$|u(t_1, x_1) - u(t_2, x_2)|^2 = |Y_{t_1}^1 - Y_{t_2}^2|^2 \le C_2 \mathbb{E}\left[|Y_{t_2}^1 - Y_{t_1}^1|^2 + |Y_{t_2}^1 - Y_{t_2}^2|^2\right].$$
(II.66)

Since Y^1 is right-continuous and bounded in S^2 , the first term on the right-hand side goes to 0 as $t_2 \to t_1$, while the second is controlled by (II.65).

1.4.2 Path regularity

Proposition 1.4.4 Assume that \mathbf{C}_1^X and \mathbf{C}_1^Y hold. Then,

$$|u(t_1, x_1) - u(t_2, x_2)|^2 \leq C_2 \left\{ (1 + |x_1|^2) |t_2 - t_1| + |x_1 - x_2|^2 \right\}$$

for all $0 \le t_1 \le t_2 \le 1$ and $(x_1, x_2) \in \mathbb{R}^{2d}$.

Proof. It suffices to plug (II.58) and (II.61) in (II.9), which is possible since the norms in (II.9) do not change after passing to suitable versions, and appeal to (II.65) and (II.66). \Box

Remark 1.4.2 A similar result is obtained in [86] when λ has a finite support. The continuity of u is proved in [5] in a case where h is bounded.

Corollary 1.4.1 Assume that \mathbf{C}_1^X and \mathbf{C}_1^Y hold.

(i) There is a version of U such that

$$\mathbb{E}\left[\sup_{r\in[s,t]}|Y_r-Y_s|^2\right] + \mathbb{E}\left[\sup_{e\in E}\sup_{r\in[s,t]}|U_r(e)-U_s(e)|^2\right] \leq C_2 (1+|X_0|^2)|t-s|,$$

for all $s \leq t \leq 1$.

(ii) If moreover \mathbf{C}_2^X and \mathbf{C}_2^Y hold, then there is a version of Z such that

$$\mathbb{E}\left[|Z_t - Z_s|^2\right] \leq C_2 \left(1 + |X_0|^4\right) |t - s| ,$$

for all $s \leq t \leq 1$.

Proof. (i) Recall from the proof of Proposition 1.4.3 that $Y = u(\cdot, X_{\cdot})$ on [0, 1]. Thus, plugging (II.7) and (II.8) in the estimate of Proposition 1.4.4 gives the upper-bound on $\mathbb{E}\left[\sup_{r\in[s,t]}|Y_r-Y_s|^2\right]$. The upper-bound on $\mathbb{E}\left[\sup_{e\in E}\sup_{r\in[s,t]}|U_r(e)-U_s(e)|^2\right]$ is obtained similarly by passing to the version of U given in Remark 1.4.1.

(ii) By Proposition 1.4.2, a version of (Z_t) is given by (Υ_t^t) so that

$$\mathbb{E}\left[|Z_t - Z_s|^2\right] \leq C_2 \left(\mathbb{E}\left[|\Upsilon_t^t - \Upsilon_t^s|^2\right] + \mathbb{E}\left[|\Upsilon_t^s - \Upsilon_s^s|^2\right]\right).$$

By (II.87) of Lemma 1.5.2, (II.34), (II.59) and (II.63), we have

$$\mathbb{E}\left[|\Upsilon^s_t - \Upsilon^s_s|^2\right] \leq C_2 \left(1 + |X_0|^4\right)|t - s|.$$

By plugging (II.36) in (II.88) of Lemma 1.5.2, we then deduce that

$$\mathbb{E}\left[|\Upsilon_t^t - \Upsilon_t^s|^2\right] \leq C_2(1+|X_0|^2)|t-s|.$$

Proposition 1.4.5 Assume that \mathbf{H}_1 - \mathbf{C}_1^X - \mathbf{C}_1^Y holds. Then there is a version of Z such that for all $n \geq 1$

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - Z_{t_i}|^2 \right] dt \leq C_2^0 n^{-1} .$$

Proof. 1. We denote by $\nabla_x h$ (resp. $\nabla_y h$, $\nabla_z h$, $\nabla_\gamma h$) the gradient of h with respect to its x variable (resp. y, z, γ). We first introduce the processes Λ and M defined by

$$\Lambda_t := \exp\left(\int_0^t \nabla_y h(\Theta_r) \, dr\right) \quad , \quad M_t := 1 + \int_0^t M_r \, \nabla_z h(\Theta_r) \cdot dW_r \, d$$

Since h has bounded derivatives, it follows from Itô's Lemma and Proposition 1.4.2 that

$$\Lambda_t M_t Z_t = \mathbb{E} \left[M_1 \left(\Lambda_1 \nabla g(X_1) \chi_1^t + \int_t^1 \left(\nabla_x h(\Theta_r) \chi_r^t + \nabla_\gamma h(\Theta_r) \Gamma_r^t \right) \Lambda_r \, dr \right) \mid \mathcal{F}_t \right] \,.$$

By Remark 1.3.7 and Proposition 1.3.3, we deduce that

$$\Lambda_t M_t Z_t = \mathbb{E} \left[M_1 \left(\Lambda_1 \nabla g(X_1) \nabla X_1 + \int_t^1 F_r \Lambda_r \, dr \right) \mid \mathcal{F}_t \right] (\nabla X_{t-})^{-1} \sigma(X_{t-})$$

where the process F is defined by

$$F_r = \nabla_x h(\Theta_r) \nabla X_r + \nabla_y h(\Theta_r) \nabla \Gamma_r \quad \text{for } r \le 1$$

It follows that

$$\Lambda_t M_t Z_t = \left\{ \mathbb{E} \left[G \mid \mathcal{F}_t \right] - \int_0^t F_r \Lambda_r \, dr \right\} (\nabla X_{t-})^{-1} \sigma(X_{t-})$$
(II.67)

where

$$G := M_1 \left(\Lambda_1 \nabla g(X_1) \nabla X_1 + \int_0^1 F_r \Lambda_r dr \right) .$$

By (II.40) and (II.62), we deduce that

$$\mathbb{E}\left[|G|^p\right] \leq C_p^0 \quad \text{for all } p \geq 2.$$
 (II.68)

Set $m_s := \mathbb{E}[G \mid \mathcal{F}_s]$ and let $(\tilde{\zeta}, \tilde{V}) \in \mathbf{H}^2 \times \mathbf{L}^2_{\lambda}$ (with values in $\mathbb{M}^d \times \mathbb{R}^d$) be defined such that

$$m_s = G - \int_s^1 \tilde{\zeta}_r dW_r - \int_s^1 \int_E \tilde{V}_r(e) \bar{\mu}(de, dr) .$$

Applying (II.68) and Lemma 1.5.2 to $(m, \tilde{\zeta}, \tilde{V})$ implies that

$$\|(m, \tilde{\zeta}, \tilde{V})\|_{\mathcal{B}^p} \leq C_p^0 \quad \text{for all } p \geq 2.$$
 (II.69)

Using \mathbf{C}_1^X , (II.40), (II.62), (II.69), applying Lemma 1.5.1 to M^{-1} and using Itô's Lemma, we deduce from the last assertion that

$$\tilde{Z} := (\Lambda M)^{-1} \left(m - \int_0^{\cdot} F_r \Lambda_r \, dr \right) (\nabla X)^{-1}$$

can be written as

$$\tilde{Z}_t = \tilde{Z}_0 + \int_0^t \tilde{\mu}_r dr + \int_0^t \tilde{\sigma}_r dW_r + \int_0^t \int_E \tilde{\beta}_r(e) \bar{\mu}(de, dr) ,$$

where

$$\|\tilde{Z}\|_{\mathcal{S}^p}^p \leq C_p^0 \quad \text{for all} \quad p \ge 2 , \qquad (\text{II.70})$$

and $\tilde{\mu}, \, \tilde{\sigma}$ and $\tilde{\beta}$ are adapted processes satisfying

$$A^p_{[0,1]} \leq C^0_p \quad \text{for all} \quad p \geq 2 \tag{II.71}$$

where

$$A^{p}_{[s,t]} := \|\tilde{\mu}\|^{p}_{\mathbf{H}^{p}_{[s,t]}} + \|\tilde{\sigma}\|^{p}_{\mathbf{H}^{p}_{[s,t]}} + \|\tilde{\beta}\|^{p}_{\mathbf{L}^{p}_{\lambda,[s,t]}} , \ s \leq t \leq 1 .$$

2. Observe that

$$Z_t = \tilde{Z}_t \sigma(X_t) \mathbb{P} - \text{a.s.}$$

since the probability of having a jump at time t is equal to zero. It follows that, for all $i \leq n$ and $t \in [t_i, t_{i+1}]$,

$$\mathbb{E}\left[|Z_t - Z_{t_i}|^2\right] \leq C_2 \left(I_{t_i,t}^1 + I_{t_i,t}^2\right)$$
(II.72)

where

$$I_{t_i,t}^1 := \mathbb{E}\left[|\tilde{Z}_t - \tilde{Z}_{t_i}|^2 |\sigma(X_{t_i})|^2 \right] \text{ and } I_{t_i,t}^2 := \mathbb{E}\left[|\sigma(X_t) - \sigma(X_{t_i})|^2 |\tilde{Z}_t|^2 \right]$$

Observing that

$$I_{t_i,t}^1 = \mathbb{E}\left[\mathbb{E}\left[|\tilde{Z}_t - \tilde{Z}_{t_i}|^2 \mid \mathcal{F}_{t_i}\right] |\sigma(X_{t_i})|^2\right]$$

$$\leq C_2 \mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} \left[|\tilde{\mu}_r|^2 + |\tilde{\sigma}_r|^2 + \int_E |\tilde{\beta}_r(e)|^2 \lambda(de)\right] dr\right) |\sigma(X_{t_i})|^2\right]$$

we deduce from Hölder inequality, (II.7) and the linear growth assumption on σ that

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} I_{t_i,t}^1 dt \leq C_2 n^{-1} \mathbb{E} \left[\left(\int_0^1 \left[|\tilde{\mu}_r|^2 + |\tilde{\sigma}_r|^2 + \int_E |\tilde{\beta}_r(e)|^2 \lambda(de) \right] dr \right) \sup_{t \leq 1} |\sigma(X_t)|^2 \right] \\ \leq C_2^0 (A_{[0,1]}^4)^{\frac{1}{2}} n^{-1} .$$
(II.73)

Using the Lipschitz continuity of σ , we obtain

$$I_{t_i,t}^2 \leq C_2 \mathbb{E}\left[|X_t - X_{t_i}|^2 |\tilde{Z}_t|^2\right]$$
 (II.74)

Now observe that for each $k,l \leq d$

$$\mathbb{E}\left[(X_t^k - X_{t_i}^k)^2 (\tilde{Z}_t^l)^2 \right] \le C_2 \left(\mathbb{E}\left[(\tilde{Z}_t^l - \tilde{Z}_{t_i}^l)^2 (X_{t_i}^k)^2 \right] + \mathbb{E}\left[(X_t^k \tilde{Z}_t^l - X_{t_i}^k \tilde{Z}_{t_i}^l)^2 \right] \right).$$
(II.75)

Arguing as above, we obtain

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[(\tilde{Z}_t^l - \tilde{Z}_{t_i}^l)^2 (X_{t_i}^k)^2 \right] dt \le C_2^0 \left(1 + (A_{[0,1]}^4)^{\frac{1}{2}} \right) n^{-1} .$$
(II.76)

Moreover, we deduce from the linear growth condition on b, σ , β and (II.7), (II.70) and (II.71) that $X^k \tilde{Z}^l$ can be written as

$$X_{t}^{k}\tilde{Z}_{t}^{l} = X_{0}^{k}\tilde{Z}_{0}^{l} + \int_{0}^{t}\hat{\mu}_{r}^{kl}dr + \int_{0}^{t}\hat{\sigma}_{r}^{kl}dW_{r} + \int_{0}^{t}\int_{E}\hat{\beta}_{r}^{kl}(e)\bar{\mu}(de,dr) + \int_{0}^{t}\hat{\sigma}_{r}^{kl}dW_{r} + \int_{$$

with $\hat{\mu}^{kl}$, $\hat{\sigma}^{kl}$ and $\hat{\beta}^{kl}$ adapted processes satisfying $\|\hat{\mu}^{kl}\|_{\mathbf{H}^2} + \|\hat{\sigma}^{kl}\|_{\mathbf{H}^2} + \|\hat{\beta}^{kl}\|_{\mathbf{L}^2_{\lambda}} \leq C_2^0$. It follows that

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[(X_t^k \tilde{Z}_t^l - X_{t_i}^k \tilde{Z}_{t_i}^l)^2 \right] dt \leq C_2 n^{-1} \left(\|\hat{\mu}^{kl}\|_{\mathbf{H}^2}^2 + \|\hat{\sigma}^{kl}\|_{\mathbf{H}^2}^2 + \|\hat{\beta}^{kl}\|_{\mathbf{L}^2_\lambda}^2 \right)$$

which combined with (II.74), (II.75) and (II.76) leads to

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} I_{t_i,t}^2 dt \leq C_2^0 (1 + (A_{[0,1]}^4)^{\frac{1}{2}}) n^{-1} .$$
(II.77)

The proof is concluded by plugging (II.73)-(II.77) in (II.72) and recalling (II.71). \Box

Proposition 1.4.6 Assume that $\mathbf{C}_1^X \cdot \mathbf{C}_1^Y$ holds. Then there is a version of Z such that, for all $\varepsilon > 0$,

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|Z_t - Z_{t_i}|^2 \right] dt \leq C_{\varepsilon}^0 n^{-1+\varepsilon} ,$$

for all $n \geq 1$.

Proof. We adapt the arguments of [18]. Let Λ and M be defined as in the proof of Proposition 1.4.5 and recall that, after possibly passing to a suitable version, $Z_t = I_t^t$ where, for $s, t \leq 1$,

$$I_s^t := \mathbb{E}\left[M_1(\Lambda_t M_t)^{-1} \left(\Lambda_1 \nabla g(X_1)\chi_1^t + \int_t^1 \left(\nabla_x h(\Theta_r)\chi_r^t + \nabla_\gamma h(\Theta_r)\Gamma_r^t\right)\Lambda_r dr\right) \mid \mathcal{F}_s\right].$$

For $t \in [t_i, t_{i+1}], i \leq n-1$, we therefore have

$$|Z_t - Z_{t_i}|^2 \leq C_2 \left(|I_t^{t_i} - I_{t_i}^{t_i}|^2 + |I_t^t - I_t^{t_i}|^2 \right)$$

where, by (II.36), (II.88) below applied to (II.42), recall that ρ is bounded, and standard estimations on ΛM ,

$$\sup_{i \le n-1, t \in [t_i, t_{i+1}]} \mathbb{E}\left[|I_t^t - I_t^{t_i}|^2 \right] \le C_2^0 n^{-1}$$

Thus it suffices to prove that

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E} \left[|I_t^{t_i} - I_{t_i}^{t_i}|^2 \right] dt \leq C_{\varepsilon}^0 n^{-1+\varepsilon} ,$$

where $\varepsilon > 0$ is now fixed. To this purpose, we first observe that I^{t_i} is a martingale on $[t_i, t_{i+1}]$, which implies that

$$\mathbb{E}\left[|I_t^{t_i} - I_{t_i}^{t_i}|^2\right] \leq \mathbb{E}\left[|I_{t_{i+1}}^{t_i}|^2 - |I_{t_i}^{t_i}|^2\right].$$
(II.78)

Remark now that we have

$$\sum_{i=0}^{n-1} \mathbb{E}\left[|I_{t_{i+1}}^{t_i}|^2 - |I_{t_i}^{t_i}|^2 \right] = \mathbb{E}\left[|Z_1|^2 - |Z_0|^2 \right] + \sum_{i=1}^n \mathbb{E}\left[|I_{t_i}^{t_{i-1}}|^2 - |I_{t_i}^{t_i}|^2 \right] ,$$

which, combined with (II.58) and (II.78), leads to

$$\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|I_t^{t_i} - I_{t_i}^{t_i}|^2 \right] dt = C_2^0 n^{-1} \left(1 + \sum_{i=1}^n \mathbb{E}\left[|I_{t_i}^{t_{i-1}}|^2 - |I_{t_i}^{t_i}|^2 \right] \right) .$$

To conclude the proof, it remains to show that

$$\mathbb{E}\left[|I_{t_i}^{t_{i-1}}|^2 - |I_{t_i}^{t_i}|^2\right] \leq \mathbb{E}\left[|I_{t_i}^{t_{i-1}} - I_{t_i}^{t_i}| |I_{t_i}^{t_{i-1}} + I_{t_i}^{t_i}|\right] \leq C_{\varepsilon}^0 n^{-1+\varepsilon}$$

which follows from Hölder inequality, Remark 1.3.4 and Lemma 1.5.2 as above. $\hfill \Box$

We now complete the proof of Theorem 1.2.1.

Proof of Theorem 1.2.1. 1. We first prove (ii). Observe that the second assertion is a direct consequence of (II.27) and Remark 1.2.4.

We first show that (II.27) holds under \mathbf{H}_1 and \mathbf{C}_1^Y . We consider a C_b^{∞} density q on \mathbb{R}^d with compact support and set

$$(b^k, \sigma^k, \beta^k(\cdot, e))(x) = k^d \int_{\mathbb{R}^d} (b, \sigma, \beta(\cdot, e))(\bar{x}) q \left(k[x - \bar{x}]\right) d\bar{x} .$$

For large $k \in \mathbb{N}$, these functions are bounded by 2K at 0. Moreover, they are K-Lipschitz and C_b^1 . Using the continuity of σ , one also easily checks that σ^k is still invertible. By \mathbf{H}_1 and Remark 1.2.7, for each $e \in E$ and $x \in \mathbb{R}^d$, $I_d + \nabla \beta^k(x, e)$ is invertible with uniformly bounded inverse. We denote by (X^k, Y^k, Z^k, U^k) the solution of (II.4)-(II.5) with (b, σ, β) replaced by (b^k, σ^k, β^k) . Since (b^k, σ^k, β^k) converges pointwise to (b, σ, β) , one easily deduces from Lemma 1.5.1 and Lemma 1.5.2 that (X^k, Y^k, Z^k, U^k) converges to (X, Y, Z, U) in $S^2 \times B^2$. Since the result of Proposition 1.4.5 holds for (X^k, Y^k, Z^k, U^k) uniformly in k, this shows that (ii) holds under \mathbf{H}_1 and \mathbf{C}_1^Y .

We now prove that (II.27) holds under \mathbf{H}_1 . Let (X, Y^k, Z^k, U^k) be the solution of (II.4)-(II.5) with h^k instead of h, where h^k is constructed by considering a sequence of molifiers as above. For large k, $h^k(0)$ is bounded by 2K. By Lemma 1.5.2, (Y^k, Z^k, U^k) converges to (Y, Z, U) in $S^2 \times \mathcal{B}^2$ which implies (ii) by arguing as above.

2. The same approximation argument shows that (i) of Corollary 1.4.1 and Proposition 1.4.6 hold true without \mathbf{C}_1^X - \mathbf{C}_1^Y . Since ρ is bounded and $\lambda(E) < \infty$, this leads to (II.25). Now observe that

$$\mathbb{E}\left[\sup_{t\in[t_i,t_{i+1}]}|\Gamma_t-\bar{\Gamma}_{t_i}|^2\right] \leq 2\mathbb{E}\left[\sup_{t\in[t_i,t_{i+1}]}|\Gamma_t-\Gamma_{t_i}|^2\right] + 2\mathbb{E}\left[|\Gamma_{t_i}-\bar{\Gamma}_{t_i}|^2\right]$$

where, by Jensen's inequality and the fact that Γ_{t_i} is \mathcal{F}_{t_i} -measurable,

$$\mathbb{E}\left[|\Gamma_{t_i} - \bar{\Gamma}_{t_i}|^2\right] \leq \mathbb{E}\left[\left|n\int_{t_i}^{t_{i+1}}(\Gamma_{t_i} - \Gamma_s)ds\right|^2\right] \leq n\int_{t_i}^{t_{i+1}}\mathbb{E}\left[|\Gamma_{t_i} - \Gamma_s|^2\right]ds.$$

Thus, (II.25) implies $\|\Gamma - \overline{\Gamma}\|_{S^2}^2 \le C_2^0 n^{-1}$ and $\|\Gamma - \overline{\Gamma}\|_{\mathbf{H}^2}^2 \le C_2^0 n^{-1}$.

3. Item (iii) is proved similarly by using (ii) of Corollary 1.4.1.

1.5 Appendix: A priori estimates

For sake of completeness, we provide in this section some a priori estimates on solutions of forward and backward SDE's with jumps. The proofs being standard, we do not provide all the details.

Proposition 1.5.1 Given $\psi \in \mathbf{L}^2_{\lambda}$, let M be defined by $M_t = \int_0^t \int_E \psi_s(e) \bar{\mu}(ds, de)$ on [0, 1]. Then, for all $p \geq 2$,

$$k_p \|\psi\|_{\mathbf{L}^p_{\lambda,[0,1]}}^p \leq \|M\|_{\mathcal{S}^p_{[0,1]}}^p \leq K_p \|\psi\|_{\mathbf{L}^p_{\lambda,[0,1]}}^p.$$
(II.79)

where k_p , K_p are positive numbers that depend only on p and $\lambda(E)$.

Proof. 1. We first prove the left hand-side. Observe that for a sequence $(a_i)_{i \in I}$ of non-negative numbers we have

$$\sum_{i \in I} a_i^{\alpha} \leq \left(\max_{i \in I} a_i \right)^{\alpha - 1} \sum_{i \in I} a_i \leq \left(\sum_{i \in I} a_i \right)^{\alpha} \quad \text{for all } \alpha \geq 1 .$$
 (II.80)

It follows that

$$\|\psi\|_{\mathbf{L}^{p}_{\lambda,[0,1]}}^{p} = \mathbb{E}\left[\int_{0}^{1}\int_{E}|\psi_{s}(e)|^{p}\mu(de,ds)\right] \leq \mathbb{E}\left[\left|\int_{0}^{1}\int_{E}|\psi_{s}(e)|^{2}\mu(de,ds)\right|^{\frac{p}{2}}\right],$$

since $p/2 \ge 1$, and the result follows from Burkholder-Davis-Gundy inequality (see e.g. [92] p. 175).

2. We now prove the right hand-side inequality for $p \ge 1$, and denote K_p a generic positive number that depends only on p. We follow the inductive argument of [13]. For $p \in [1, 2]$, we deduce from Burkholder-Davis-Gundy inequality and (II.80) that

$$\mathbb{E}\left[\sup_{s\leq 1}|M_s|^p\right] \leq K_p \mathbb{E}\left[\left(\int_0^1 \int_E |\psi_s(e)|^2 \mu(de, ds)\right)^{\frac{p}{2}}\right]$$
$$\leq K_p \mathbb{E}\left[\int_0^1 \int_E |\psi_s(e)|^p \mu(de, ds)\right],$$

since $2/p \ge 1$. This implies the required result.

We now assume that the inequality is valid from some p > 1 and prove that it is also true for 2p. We define $\tilde{M}_t = \int_0^t \int_E \psi_s(e)^2 \bar{\mu}(de, ds)$, for $t \in [0, 1]$. Then, we have $[M, M]_1 = \tilde{M}_1 + \int_0^1 \int_E \psi_s(e)^2 \lambda(de) ds$. Applying Burkholder-Davis-Gundy inequality, we obtain $\mathbb{E}\left[\sup_{s \leq 1} |M_s|^{2p}\right] \leq \mathbb{E}\left[[M, M]_1^p\right]$ where

$$\mathbb{E}\left[\left[M,M\right]_{1}^{p}\right] \leq K_{p} \mathbb{E}\left[\left|\tilde{M}_{1}\right|^{p} + \left(\int_{0}^{1}\int_{E}\psi_{s}(e)^{2}\lambda(de)ds\right)^{p}\right].$$

Applying (II.79) to \tilde{M} , we obtain

$$\mathbb{E}\left[|\tilde{M}_1|^p\right] \leq K_p \mathbb{E}\left[\int_0^1 \int_E |\psi_s(e)|^{2p} \lambda(de) ds\right] \,.$$

On the other hand, it follows from Hölder inequality that

$$\int_0^1 \int_E \psi_s(e)^2 \lambda(de) ds \leq \left(\int_0^1 \int_E |\psi_s(e)|^{2p} \lambda(de) ds \right)^{\frac{1}{p}} \lambda(E)^{\frac{1}{q}}$$

where q = p/(p-1), recall that p > 1. Combining the two last inequalities leads to the required result.

We now consider some measurable maps

$$\begin{split} \tilde{b}^{i} &: \quad \Omega \times [0,1] \times \mathbb{R}^{d} \mapsto \mathbb{R}^{d} \\ \tilde{\sigma}^{i} &: \quad \Omega \times [0,1] \times \mathbb{R}^{d} \mapsto \mathbb{M}^{d} \\ \tilde{\beta}^{i} &: \quad \Omega \times [0,1] \times \mathbb{R}^{d} \times E \mapsto \mathbb{R}^{d} \\ \tilde{f}^{i} &: \quad \Omega \times [0,1] \times \mathbb{R} \times \mathbb{R}^{d} \times L^{2}(E,\mathcal{E},\lambda;\mathbb{R}) , \ i = 1,2 . \end{split}$$

Here $L^2(E, \mathcal{E}, \lambda; \mathbb{R})$ is endowed with the natural norm $(\int_E |a(e)|^2 \lambda(de))^{\frac{1}{2}}$. Omitting the dependence of these maps with respect to $\omega \in \Omega$, we assume that for each $t \leq 1$

 $\tilde{b}^i(t,\cdot)$, $\tilde{\sigma}^i(t,\cdot)$, $\tilde{\beta}^i(t,\cdot,e)$ and $\tilde{f}^i(t,\cdot)$ are a.s. K-Lipschitz continuous

uniformly in $e \in E$ for $\tilde{\beta}^i$. We also assume that $t \mapsto (\tilde{f}^i(t, \cdot), \tilde{b}^i(t, \cdot))$ is \mathbb{F} -progressively measurable, and $t \mapsto (\tilde{\sigma}^i(t, \cdot), \tilde{\beta}^i(t, \cdot))$ is \mathbb{F} -predictable, i = 1, 2.

Given some real number $p \geq 2$, we assume that $|\tilde{b}^i(\cdot, 0)|$, $|\tilde{\sigma}^i(\cdot, 0)|$ and $|\tilde{f}^i(\cdot, 0)|$ are in \mathbf{H}^p , and that $|\tilde{\beta}^i(\cdot, 0, \cdot)|$ is in \mathbf{L}^p_{λ} .

For $t_1 \leq t_2 \leq 1$, $\tilde{X}^i \in L^2(\Omega, \mathcal{F}_{t_i}, \mathbb{P}; \mathbb{R}^d)$ for i = 1, 2, we now denote by X^i the solution on $[t_i, 1]$ of

$$X_{t}^{i} = \tilde{X}^{i} + \int_{t_{i}}^{t} \tilde{b}^{i}(s, X_{s}^{i})ds + \int_{t_{i}}^{t} \tilde{\sigma}^{i}(s, X_{s}^{i})dW_{s} + \int_{t_{i}}^{t} \int_{E} \tilde{\beta}^{i}(s, e, X_{s-}^{i})\bar{\mu}(de, ds) .$$
(II.81)

Lemma 1.5.1

$$\|X^{1}\|_{\mathcal{S}^{p}_{[t_{1},1]}}^{p} \leq C_{p} \left\{ \mathbb{E}[|\tilde{X}^{1}|^{p}] + \|\tilde{b}^{1}(\cdot,0)\|_{\mathbf{H}^{p}_{[t_{1},1]}}^{p} + \|\tilde{\sigma}^{1}(\cdot,0)\|_{\mathbf{H}^{p}_{[t_{1},1]}}^{p} + \|\tilde{\beta}^{1}(\cdot,0,\cdot)\|_{\mathbf{L}^{p}_{\lambda,[t_{1},1]}}^{p} \right\}.$$
(II.82)

Moreover, for all $t_1 \leq s \leq t \leq 1$,

$$\mathbb{E}\left[\sup_{s\leq u\leq t}|X_u^1-X_s^1|^p\right] \leq C_p A_p^1|t-s|, \qquad (II.83)$$

where A_p^1 is defined as

$$\mathbb{E}[|\tilde{X}^{1}|^{p}] + \mathbb{E}\left[\sup_{t_{1} \le s \le 1} |\tilde{b}^{1}(s,0)|^{p} + \sup_{t_{1} \le s \le 1} |\tilde{\sigma}^{1}(s,0)|^{p} + \sup_{t_{1} \le s \le 1} \left\{ \int_{E} |\tilde{\beta}^{1}(s,0,e)|^{p} \lambda(de) \right\} \right],$$

and, for $t_{2} \le t \le 1$,

$$\begin{aligned} \|\delta X\|_{\mathcal{S}^{p}_{[t_{2},1]}}^{p} &\leq C_{p} \left(E|\tilde{X}^{1} - \tilde{X}^{2}|^{p} + A_{p}^{1}|t_{2} - t_{1}| \right) \\ &+ C_{p} \left(E\left(\int_{t_{2}}^{1} |\delta \tilde{b}_{t}| dt \right)^{p} + \|\delta \tilde{\sigma}\|_{\mathbf{H}^{p}_{[t_{2},1]}}^{p} + \|\delta \tilde{\beta}\|_{\mathbf{L}^{p}_{\lambda,[t_{2},1]}}^{p} \right), \quad (\mathrm{II.84}) \end{aligned}$$

where $\delta X := X^1 - X^2$, $\delta \tilde{b} = (\tilde{b}^1 - \tilde{b}^2)(\cdot, X^1)$ and $\delta \tilde{\sigma}$, $\delta \tilde{\beta}$ are defined similarly.

Lemma 1.5.2 (i) Let \tilde{f} be equal to \tilde{f}^1 or \tilde{f}^2 . Given $\tilde{Y} \in L^p(\Omega, \mathcal{F}_1, \mathbb{P}; \mathbb{R})$, the backward SDE

$$Y_t = \tilde{Y} + \int_t^1 \tilde{f}(s, Y_s, Z_s, U_s) ds + \int_t^1 Z_s \cdot dW_s + \int_t^1 \int_E U_s(e) \bar{\mu}(de, ds) \quad (\text{II.85})$$

has a unique solution (Y, Z, U) in \mathcal{B}^2 . It satisfies

$$\|(Y,Z,U)\|_{\mathcal{B}^{p}}^{p} \leq C_{p} \mathbb{E}\left[|\tilde{Y}|^{p} + \left(\int_{0}^{1} |\tilde{f}(t,0)|dt\right)^{p}\right].$$
 (II.86)

Moreover, if $A_p := \mathbb{E}\left[|\tilde{Y}|^p + \sup_{t \leq 1} |\tilde{f}(t,0)|^p\right] < \infty$, then

$$\mathbb{E}\left[\sup_{s\leq u\leq t}|Y_{u}-Y_{s}|^{p}\right]\leq C_{p}\left\{A_{p}|t-s|^{p}+\|Z\|_{\mathbf{H}^{p}_{[s,t]}}^{p}+\|U\|_{\mathbf{L}^{p}_{\lambda,[s,t]}}^{p}\right\}.$$
 (II.87)

(ii) Fix \tilde{Y}^1 and \tilde{Y}^2 in $L^p(\Omega, \mathcal{F}_1, \mathbb{P}; \mathbb{R})$ and let (Y^i, Z^i, U^i) be the solution of (II.86) with $(\tilde{Y}^i, \tilde{f}^i)$ in place of (\tilde{Y}, \tilde{f}) , i = 1, 2. Then, for all $t \leq 1$,

$$\|(\delta Y, \delta Z, \delta U)\|_{\mathcal{B}^{p}_{[t,1]}}^{p} \leq C_{p} \mathbb{E}\left[|\delta \tilde{Y}|^{p} + \left(\int_{t}^{1} |\delta \tilde{f}_{r}| dr\right)^{p}\right]$$
(II.88)

where $\delta \tilde{Y} := \tilde{Y}^1 - \tilde{Y}^2$, $\delta Y := Y^1 - Y^2$, $\delta Z := Z^1 - Z^2$, $\delta U := U^1 - U^2$ and

$$\delta \tilde{f}_{\cdot} := (\tilde{f}^1 - \tilde{f}^2)(\cdot, Y^1_{\cdot}, Z^1_{\cdot}, U^1_{\cdot})$$

Proof of Lemma 1.5.1. Applying Burkholder-Davis-Gundy inequality (see e.g. [92] p 175) and using Proposition 1.5.1, we get

$$\mathbb{E}\left[\sup_{s\in[t_{1},1]}|X_{s}^{1}|^{p}\right] \leq C_{p} \mathbb{E}\left[|\tilde{X}^{1}|^{p} + \left(\int_{t_{1}}^{1}|\tilde{b}^{1}(s,X_{s}^{1})|ds\right)^{p}\right] \\ + C_{p} \left(\|\tilde{\sigma}^{1}(\cdot,X_{\cdot}^{1})\|_{\mathbf{H}^{p}_{[t_{1},1]}}^{p} + \|\tilde{\beta}^{1}(\cdot,X_{\cdot}^{1},\cdot)\|_{\mathbf{L}^{p}_{\lambda,[t_{1},1]}}^{p}\right)$$

The estimate (II.82) is then deduced by using the Lipschitz properties of \tilde{b}^1 , $\tilde{\sigma}^1$, $\tilde{\beta}^1$ and Gronwall's Lemma. The estimate (II.83) is obtained by applying the same arguments to the process $|X_{\cdot}^1 - X_s^1|^p$ on [s, t]. To obtain the last assertion (II.84), we first apply the above argument to $\delta X = X^1 - X^2$ on $[t_2, 1]$. Then, decomposing $\tilde{b}^1(\cdot, X^1) - \tilde{b}^2(\cdot, X^2)$ as $\delta \tilde{b} + \tilde{b}^2(\cdot, X^1) - \tilde{b}^2(\cdot, X^2)$ and doing the same for $\tilde{\sigma}$ and $\tilde{\beta}^i$, the Lipschitz properties of \tilde{b}^2 , $\tilde{\sigma}^2$, $\tilde{\beta}^2$ combined with Gronwall's lemma leads to

$$\mathbb{E}\left[\sup_{s\in[t_2,1]}|\delta X_s|^p\right] \le C_p\left(\mathbb{E}|X_{t_2}^1 - \tilde{X}^2|^p + \mathbb{E}\left(\int_{t_2}^1 |\delta \tilde{b}_t|dt\right)^p + \|\delta \tilde{\sigma}\|_{\mathbf{H}^p_{[t_2,1]}}^p + \|\delta \tilde{\beta}\|_{\mathbf{L}^p_{\lambda,[t_2,1]}}^p\right)$$

We then conclude by using the (II.83).

Proof of Lemma 1.5.2. See [100] and [5] for existence and uniqueness.

(i) We divide [0, 1] in N intervals $[\tau_i, \tau_{i+1}]$ of equal length $\delta := 1/N$. For $\tau_i \leq t \leq s \leq \tau_{i+1}$

$$|Y_s| \leq \mathbb{E}\left[|Y_{\tau_{i+1}}| + \int_t^{\tau_{i+1}} |\tilde{f}(r, Y_r, Z_r, U_r)| dr \,|\,\mathcal{F}_s\right]$$

which, by Doob and Jensen's inequalities, implies

$$\mathbb{E}\left[\sup_{t\leq s\leq \tau_{i+1}}|Y_s|^p\right] \leq C_p \mathbb{E}\left[|Y_{\tau_{i+1}}|^p + \left(\int_t^{\tau_{i+1}}|\tilde{f}(r,Y_r,Z_r,U_r)|dr\right)^p\right].$$

Moreover, it follows from Burkholder-Davis-Gundy inequality (see e.g. [92] p. 175) and Proposition 1.5.1 that

$$\begin{aligned} \|Z\|_{\mathbf{H}^{p}_{[t,\tau_{i+1}]}}^{p} + \|U\|_{\mathbf{L}^{p}_{\lambda,[t,\tau_{i+1}]}}^{p} &\leq C_{p} \mathbb{E}\left[|Y_{\tau_{i+1}}|^{p} + \sup_{t \leq s \leq \tau_{i+1}} |Y_{s}|^{p}\right] \\ &+ C_{p} \mathbb{E}\left[+\left(\int_{t}^{\tau_{i+1}} |\tilde{f}(r,Y_{r},Z_{r},U_{r})|dr\right)^{p}\right]. \end{aligned}$$

Thus, using Hölder and Jensen's inequalities, we obtain

$$\begin{split} \|(Y,Z,U)\|_{\mathcal{B}^{p}_{[t,\tau_{i+1}]}}^{p} &\leq C_{p} \mathbb{E}\left[|Y_{\tau_{i+1}}|^{p} + \left(\int_{t}^{\tau_{i+1}} |\tilde{f}(r,Y_{r},Z_{r},U_{r})|dr\right)^{p}\right] \\ &\leq C_{p} \mathbb{E}\left[|Y_{\tau_{i+1}}|^{p} + \left(\int_{0}^{1} |\tilde{f}(t,0)|dt\right)^{p}\right] \\ &+ C_{p} \left\{\int_{t}^{\tau_{i+1}} \|Y\|_{\mathcal{S}^{p}_{[u,\tau_{i+1}]}}^{p} du\delta^{p/2} \left(\|Z\|_{\mathbf{H}^{p}_{[t,\tau_{i+1}]}}^{p} + \|U\|_{\mathbf{L}^{p}_{\lambda,[t,\tau_{i+1}]}}^{p}\right)\right\}, \end{split}$$

by the Lipschitz continuity assumption on \tilde{f} . For δ smaller than $(2C_p)^{-2/p}$, we then get

$$\|(Y,Z,U)\|_{\mathcal{B}^{p}_{[t,\tau_{i+1}]}}^{p} \leq C_{p} \left\{ \mathbb{E}\left[|Y_{\tau_{i+1}}|^{p}\right] + \left(\int_{0}^{1} |\tilde{f}(t,0)| dt\right)^{p} + \int_{t}^{\tau_{i+1}} \|Y\|_{\mathcal{S}^{p}_{[u,\tau_{i+1}]}}^{p} du \right\} .$$

Using Gronwall's Lemma, we deduce that

$$\|Y\|_{\mathcal{S}^{p}_{[\tau_{i},\tau_{i+1}]}}^{p} \leq C_{p} \left\{ \mathbb{E}\left[|Y_{\tau_{i+1}}|^{p}\right] + \left(\int_{0}^{1} |\tilde{f}(t,0)| dt\right)^{p} \right\} .$$

Plugging this estimate into the previous upper bound, we finally get

$$\|(Y,Z,U)\|_{\mathcal{B}^{p}_{[\tau_{i},\tau_{i+1}]}}^{p} \leq C_{p} \mathbb{E}\left[|Y_{\tau_{i+1}}|^{p} + \left(\int_{0}^{1} |\tilde{f}(t,0)| dt\right)^{p}\right].$$

This leads to (II.86).

By Burkholder-Davis-Gundy inequality and Proposition 1.5.1, we have

$$\mathbb{E}\left[\sup_{s\leq u\leq t}|Y_u-Y_s|^p\right] \leq C_p \mathbb{E}\left[\left(\int_s^t |\tilde{f}(r,Y_r,Z_r,U_r)|dr\right)^p + C_p \left\{\|Z\|_{\mathbf{H}^p_{[s,t]}}^p + \|U\|_{\mathbf{L}^p_{\lambda,[s,t]}}^p\right\}.$$

Using the Lipschitz continuity assumption on \tilde{f} together with (II.86) leads to (II.87).

(ii) The estimate (II.88) is obtained by applying similar arguments to $(\delta Y, \delta Z, \delta U)$. \Box
Chapter 2

Algorithm and numerical results

2.1 A fully implementable algorithm

This section presents a fully implementable convergent algorithm for the resolution of systems of decoupled FBSDEs with jumps. We studied in the previous chapter the error of a discrete time scheme which requires the computation of a large number of conditional expectations. We analyse here the propagation of the statistical error coming from the approximation of the conditional expectation operators by means of non parametric estimation techniques. This algorithm is a direct adaptation of the one proposed by Lemor, Gobet and Warin [73] and presented in detail in the PhD dissertation of Lemor [72]. They consider the case where the driver h does not depend on Γ and consequently on U, so that they do not require the estimations of the process Γ by $\overline{\Gamma}^{\pi}$. Our generalization mainly relies on handling the estimation of $\overline{\Gamma}^{\pi}$ by similar techniques used to estimate \overline{Z}^{π} . Our main result is that the additional dependence of the driver hin the jumps part of the BSDE does not modify the speed of the algorithm. We should refer to [73] for the obtention of some technical results and try to follow their notations. In particular, from now on, C denotes a generic constant which may depend on X^0 . We work under the assumptions of the previous chapter.

The section is organized as follows. We first modify the coefficients h and g in order to localize the solution of the BSDE with jumps. We then present the fully implementable algorithm and provide its statistical error, allowing to choose at the same time the different parameters of the algorithm. The technical proof of the control of the statistical error is reported in Section 2.1.4.

2.1.1 A localization procedure

For a given $R \in \mathbb{R}^d$, we localize functions h and g by:

$$h^R$$
 : $(x, y, z, \gamma) \mapsto h[-R \lor (x \land R), y, z, \gamma]$ and g^R : $x \mapsto g[-R \lor (x \land R)]$

where $-R \vee (x \wedge R)$ is computed componentwise. We denote $(\tilde{Y}^{\pi,R}, \tilde{Z}^{\pi,R}, \tilde{\Gamma}^{\pi,R})$ the solution of the localized version of the explicit discretization scheme studied in the previous chapter, where the coefficients h and g are respectively replaced by h^R and g^R . Therefore, we have $\tilde{Y}_1^{\pi,R} = g^R(X_1^{\pi})$ and ,on each interval $[t_i, t_{i+1}]$, we get

$$\begin{cases} \tilde{Z}_{t}^{\pi,R} := n \mathbb{E} \left[\tilde{Y}_{t_{i+1}}^{\pi,R} \Delta W_{i+1} \mid \mathcal{F}_{t_{i}} \right] \\ \tilde{\Gamma}_{t}^{\pi,R} := n \mathbb{E} \left[\tilde{Y}_{t_{i+1}}^{\pi,R} \int_{E} \rho(e) \bar{\mu}(de, (t_{i}, t_{i+1}]) \mid \mathcal{F}_{t_{i}} \right] \\ \tilde{Y}_{t}^{\pi,R} := \mathbb{E} \left[\tilde{Y}_{t_{i+1}}^{\pi,R} + \frac{1}{n} h^{R} \left(X_{t_{i}}^{\pi}, \tilde{Y}_{t_{i+1}}^{\pi,R}, \tilde{\Gamma}_{t_{i}}^{\pi,R}, \tilde{\Gamma}_{t_{i}}^{\pi,R} \right) \mid \mathcal{F}_{t_{i}} \right] . \end{cases}$$
(II.2.1)

Before going any further, notice that, since ρ is bounded by K, the application of the Cauchy-Schwarz inequality to the first and the second equations of (II.2.1), leads to the useful estimates

$$|\tilde{Z}_{t}^{\pi,R}|^{2} \leq n \left(\mathbb{E}[|\tilde{Y}_{t_{i+1}}^{\pi,R}|^{2} | \mathcal{F}_{t_{i}}] - \mathbb{E}[\tilde{Y}_{t_{i+1}}^{\pi,R} | \mathcal{F}_{t_{i}}]^{2} \right), \qquad (\text{II.2.2})$$

$$|\tilde{\Gamma}_{t}^{\pi,R}|^{2} \leq K^{2} n \left(\mathbb{E}[|\tilde{Y}_{t_{i+1}}^{\pi,R}|^{2} | \mathcal{F}_{t_{i}}] - \mathbb{E}[\tilde{Y}_{t_{i+1}}^{\pi,R} | \mathcal{F}_{t_{i}}]^{2} \right), \qquad (\text{II.2.3})$$

for any $t \in [t_i, t_{i+1})$. We emphasize that estimate (II.2.3) is crucial since it allows to control the error on Γ with the same procedure as the one used to handle the error on Z in [72], as detailed in the rest of the section.

The main purpose of the localization procedure is to obtain bounds on the approximation process $(\tilde{Y}^{\pi,R}, \tilde{Z}^{\pi,R}, \tilde{\Gamma}^{\pi,R})$, as stated in the next Proposition.

Proposition 2.1.1 There exists a constant C such that, denoting

$$C_y(R) := C\{||g^R||_{\infty} + ||h^R||_{\infty}\}, \quad C_z(R) := C_y(R)\sqrt{n} \quad and \quad C_\gamma(R) := K C_y(R)\sqrt{n},$$

we have, for n sufficiently large,

$$|\tilde{Y}^{\pi,R}| \leq C_y(R), \quad |\tilde{Z}^{\pi,R}| \leq C_z(R) \quad and \quad |\tilde{\Gamma}^{\pi,R}| \leq C_\gamma(R).$$

Proof. For any a > 0, combining the Lipschitz property of h with Youngs inequality applied to the last equation of (II.2.1), we derive

$$\begin{split} |\tilde{Y}_{t_{i}}^{\pi,R}|^{2} &\leq \left(1+\frac{a}{n}\right) |E[\tilde{Y}_{t_{i+1}}^{\pi,R} \mid \mathcal{F}_{t_{i}}]|^{2} \\ &+ \frac{C}{n^{2}} \left(1+\frac{n}{a}\right) \left\{ |h^{R}(X_{t_{i}}^{\pi},0,0,0)|^{2} + \mathbb{E}_{i}[|\tilde{Y}_{t_{i+1}}^{\pi,R}|^{2} \mid \mathcal{F}_{t_{i}}] + |\tilde{Z}_{t_{i}}^{\pi,R}|^{2} + |\tilde{\Gamma}_{t_{i}}^{\pi,R}|^{2} \right\} \,, \end{split}$$

for any $i \leq n$. Thanks to estimates (II.2.2) and (II.2.3), choosing a conveniently, we deduce

$$|\tilde{Y}_{t_i}^{\pi,R}|^2 \leq \left(1 + \frac{C}{n}\right) \mathbb{E}[|\tilde{Y}_{t_{i+1}}^{\pi,R}|^2 \mid \mathcal{F}_{t_i}] + \frac{C}{n} |h^R(X_{t_i}^{\pi}, 0, 0, 0, 0)|^2$$

Applying the discrete Gronwall lemma, we obtain the announced upper bound on $\tilde{Y}^{\pi,R}$. Plugging this estimate in (II.2.2) and (II.2.3) concludes the proof.

The error induced by the localization procedure is denoted

$$\operatorname{Err}^{loc}(Y, Z, U)^{2} := \max_{0 \le i \le n-1} \mathbb{E}\left[|\tilde{Y}_{t_{i}}^{\pi} - \tilde{Y}_{t_{i}}^{\pi, R}|^{2} \right] + \|\tilde{Z}^{\pi} - \tilde{Z}^{\pi, R}\|_{\mathbf{H}^{2}}^{2} + \|\tilde{\Gamma}^{\pi} - \tilde{\Gamma}^{\pi, R}\|_{\mathbf{H}^{2}}^{2} + \|\tilde{Y}_{t_{i}}^{\pi} - \tilde{Y}_{t_{i}}^{\pi, R}\|_{\mathbf{H}^{2}}^{2} + \|\tilde{Y}_{t_{i}}^{\pi} - \tilde{Y}_{t_{i}}^{\pi}\|_{\mathbf{H}^{2}}^{2} + \|\tilde{Y}_{t_{i}}^{\pi} - \tilde{Y}_{t_{i}}^{\pi}\|_{\mathbf{H}^{2}}^{2} + \|\tilde{Y}_{t_{i}}^{\pi} - \tilde{Y}_{t_{i}}^{\pi}\|_{\mathbf{H}^{2}}^{2} + \|\tilde{Y}_{t_{i}}^{\pi}\|_{\mathbf{H}^{2}}^{2} + \|\tilde{Y}_{t_{i}}^{\pi}\|_{\mathbf$$

The next Proposition provides a control on this error.

Proposition 2.1.2 Denoting $\Delta^R \varphi := \varphi - \varphi^R$ for $\varphi = g$ and h, we have

$$\operatorname{Err}^{loc}(Y, Z, U)^{2} \leq C \mathbb{E}[\Delta^{R}g(X_{1}^{\pi})] + \frac{C}{n} \mathbb{E}\sum_{i=0}^{n-1} \left| \Delta^{R}h\left(X_{t_{i}}^{\pi}, \tilde{Y}_{t_{i+1}}^{\pi}, \tilde{Z}_{t_{i}}^{\pi}, \tilde{\Gamma}_{t_{i}}^{\pi}\right) \right|^{2},$$

for n sufficiently large.

Proof. We omit the proof, which is a direct adaptation of the proof of Proposition 2 in [73], where we control the error on Γ by replacing estimates of the form (II.2.2) used to control Z, by estimates of (II.2.3) in the spirit of the previous proof.

Since the coefficients f and g are Lipschitz, the previous Proposition allows to control the error of localisation in terms of the tails of distributions of the process X^{π} . But, for any p > 0, we have $||X^{\pi}||_{\mathcal{S}^p} < C_p$, and we deduce

$$\operatorname{Err}^{loc}(Y, Z, U) \leq C_p n^{1/2} (1+R)^{1-p/2}$$

Thus, for any p > 0, this error is dominated by the error of discretization whenever R is of the order $n^{2/(p-2)+\epsilon}$, with $\varepsilon > 0$. As observed in [73], it suffices therefore to choose a fixed R large enough in order to obtain a very good approximation in practice, and we do so from now on.

2.1.2 Description of the algorithm

This section presents the fully implementable algorithm, direct adaptation of the one detailed in [72]. At each time t_i , the algorithm relies on a non parametric estimation of the deterministic functions $y_i^{\pi,R}$, $z_i^{\pi,R}$ and $\gamma_i^{\pi,R}$ characterized by

$$\tilde{Y}_{t_i}^{\pi,R} = y_i^{\pi,R}(X_{t_i}^{\pi}), \quad \tilde{Z}_{t_i}^{\pi,R} = z_i^{\pi,R}(X_{t_i}^{\pi}) \text{ and } \tilde{\Gamma}_{t_i}^{\pi,R} = \gamma_i^{\pi,R}(X_{t_i}^{\pi}).$$

In order to do this, we introduce d + d' + 1 deterministic function basis $(p_i^y), (p_{l,i}^z)_{0 \le l \le d}$ and $(p_{l',i}^{\gamma})_{0 \leq l' \leq d'}$. Let B be a parameter such that each basis is a vector composed by at most B functions and we denote by \mathcal{P}_i^y , $\mathcal{P}_{l,i}^z$ and $\mathcal{P}_{l,i}^\gamma$ the vector spaces respectively spanned by p_i^y , $p_{l,i}^z$ and $p_{l',i}^\gamma$.

For any function φ , we denote

$$[\varphi]_a(.) := -C_a(R) \wedge [\varphi(.) \vee C_a(R)], \quad \text{for } a = y, z \text{ and } \gamma.$$

The proposed algorithm is the following:

Time discretization π

Fix a regular discretization grid on [0, 1] with time step of order $\pi := 1/n$.

Monte Carlo simulation of the forward process X^{π}

At each time t_i , simulate M independent realizations of the increments of the Brownian Motion ΔW_{i+1} and the martingale $\int_E \bar{\mu}(de, (t_i, t_{i+1}])$. Compute for any path $m \leq M$, the approximation of X by its Euler scheme

$$\begin{cases} X_0^{\pi,m} &:= X_0 \\ X_{t_{i+1}}^{\pi,m} &:= X_{t_i}^{\pi,m} + \frac{1}{n} b(X_{t_i}^{\pi,m}) + \sigma(X_{t_i}^{\pi,m}) \Delta W_{i+1}^m + \int_E \beta(X_{t_i}^{\pi,m}, e) \bar{\mu}^m (de, (t_i, t_{i+1}]) \,. \end{cases}$$

Initialization of $y_n^{\pi,R,M}$ For each path $m \leq M$, we approximate the function $y_n^{\pi,R}$ by $y_n^{\pi,R,M} := g^R$.

Backward iteration at time t_i : from $y_{i+1}^{\pi,R,M}$ to $y_i^{\pi,R,M}$

• Simulation of an extra process \hat{X}^{π}

For each path $m \leq M$, simulate one realization $(\Delta \hat{W}_{i+1}^m, \int_E \hat{\mu}^m(de, (t_i, t_{i+1}]))$ of $(\Delta W_i, \int_E \bar{\mu}(de, (t_i, t_{i+1}]))$, independent of the previous simulations, and compute the process

$$\hat{X}_{t_{i+1}}^{\pi,m} := X_{t_i}^{\pi,m} + \frac{1}{n} b(X_{t_i}^{\pi,m}) + \sigma(X_{t_i}^{\pi,m}) \Delta \hat{W}_{i+1}^m + \int_E \beta(X_{t_i}^{\pi,m}, e) \hat{\mu}^m (de, (t_i, t_{i+1}]).$$

• Approximation of $z_i^{\pi,R}$ For $0 \leq l \leq d$, compute $\alpha_{l,i}^{z,M}$ solution of the ordinary least squares (OLS) problem

$$\inf_{\alpha_l} \frac{1}{M} \sum_{m=1}^{M} \left| n \, y_{i+1}^{\pi,R,M}(\hat{X}_{t_{i+1}}^{\pi,m}) \, \Delta \hat{W}_{l,i}^m - \alpha_l . p_{l,i}^z(X_{t_i}^{\pi,m}) \right|^2 \,,$$

and define the function $z_{l,i}^{\pi,R,M} := [\alpha_{l,i}^{z,M}.p_{l,i}^z]_z$.

• Approximation of $\gamma_i^{\pi,R}$ For $0 \leq l' \leq d'$, compute $\alpha_{l',i}^{\gamma,M}$ solution of the OLS problem

$$\inf_{\alpha_{l'}} \frac{1}{M} \sum_{m=1}^{M} \left| n \, y_{i+1}^{\pi,R,M}(\hat{X}_{t_{i+1}}^{\pi,m}) \, \int_{E} \rho(e) \hat{\mu}^{m}(de,(t_{i},t_{i+1}]) - \alpha_{l} \cdot p_{l',i}^{\gamma}(X_{t_{i}}^{\pi,m}) \right|^{2} \,,$$

and define the function $\gamma_{l,i}^{\pi,R,M} := [\alpha_{l',i}^{\gamma,M} \cdot p_{l',i}^{\gamma}]_z$.

• Approximation of $y_i^{\pi,R}$ Compute $\alpha_i^{y,M}$ solution of the OLS problem

$$\inf_{\alpha} \frac{1}{M} \sum_{m=1}^{M} \left| \frac{1}{n} h^{R} [X_{t_{i}}^{\pi,m}, y_{i+1}^{\pi,R,M}(\hat{X}_{t_{i+1}}^{\pi,m}), z_{i}^{\pi,R,M}(X_{t_{i}}^{\pi,m}), \gamma_{i}^{\pi,R,M}(X_{t_{i}}^{\pi,m})] + y_{i+1}^{\pi,R,M}(\hat{X}_{t_{i+1}}^{\pi,m}) - \alpha p_{i}^{y}(X_{t_{i}}^{\pi,m}) \right|^{2},$$

and define the function $y_i^{\pi,R,M} := [\alpha_i^{y,M}.p_i^y]_y$.

As explained in [73], we could avoid the simulation at each time t_i of M extra realizations of $(\Delta W_{i+1}, \int_E \bar{\mu}(de, (t_i, t_{i+1}]))$ and replace, for any $m \leq M$, $(\Delta \hat{W}_{i+1}^m, \int_E \hat{\mu}^m(de, (t_i, t_{i+1}])))$ by $(\Delta W_{i+1}^m, \int_E \bar{\mu}^m(de, (t_i, t_{i+1}])))$ in the previous expressions. To obtain a convergent algorithm, they require an additional truncation of the increments of the Brownian motion on each interval $[t_i, t_{i+1}]$ multiplied by $n^{-1/2}$. By similarly truncating the sum of the jumps on each interval $[t_i, t_{i+1}]$, we could also apply the same trick. The derived upper bound on the theoretical statistical error of the second algorithm is higher, but, according to [73], this modification does not seem to be relevant in practice.

2.1.3 Discussion on the global error of the algorithm

In this subsection, we control the statistical error of the algorithm and discuss briefly the relative orders of the parameters n, N and B.

For any function ψ , we denote

$$||\psi||_{i,M}^2 \ := \ \frac{1}{M} \sum_{m=1}^M |\psi(X_{t_i}^{\pi,m})|^2 \,.$$

The integrated empirical statistical error due to the approximations of the functions $y_i^{\pi,R}$, $z_i^{\pi,R}$ and $\gamma_i^{\pi,R}$ is the following

$$\begin{aligned} \operatorname{Err}^{stat\,emp}\left(Y,Z,U\right)^{2} &:= & \max_{0 \le i \le n} \mathbb{E}||y_{i}^{\pi,R} - y_{i}^{\pi,R,M}||_{i,M}^{2} + & \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}||z_{i}^{\pi,R} - z_{i}^{\pi,R,M}||_{i,M}^{2} \\ &+ & \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}||\gamma_{i}^{\pi,R} - \gamma_{i}^{\pi,R,M}||_{i,M}^{2} \end{aligned}$$

Theorem 2.1.1 For any $\beta \in (1, 2]$, the empirical statistical error satisfies

$$\operatorname{Err}^{stat\ emp}\left(Y, Z, U\right)^{2} \leq C(C_{y}(R)^{2} + ||h^{R}||_{\infty}^{2}) nBM^{-1} + C n^{1-\beta} + C C_{y}(R)^{2} n^{2} e^{CB \log(C_{y}(R)n^{\beta+1}) - \frac{M n^{-\beta-1}}{144C_{y}(R)^{2}}} + C \sum_{i=0}^{n-1} E(i),$$

for n sufficiently large, where, at time t_i , E(i) is defined by

$$E(i) := \inf_{\alpha} \mathbb{E}|y_{i}^{\pi,R}(X_{t_{i}}^{\pi}) - \alpha.p_{i}^{y}(X_{t_{i}}^{\pi})|^{2} + \sum_{l=1}^{d} \inf_{\alpha_{l}} \mathbb{E}|n^{-1/2} z_{l,i}^{\pi,R}(X_{t_{i}}^{\pi}) - \alpha_{l}.p_{l,i}^{z}(X_{t_{i}}^{\pi})|^{2} + \sum_{l'=1}^{d'} \inf_{\alpha_{l'}} \mathbb{E}|n^{-1/2} \gamma_{l',i}^{\pi,R}(X_{t_{i}}^{\pi}) - \alpha_{l'}.p_{l',i}^{\gamma}(X_{t_{i}}^{\pi})|^{2}.$$

The proof of this theorem is reported in section 2.1.4.

The previous statistical error is written in terms of the empirical law of $(X^{\pi,m})_{m \leq M}$, but we can also control the true statistical error written in terms of the law of X^{π} , which is defined by

$$\operatorname{Err}^{stat}(Y, Z, U)^{2} := \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}\left[|\gamma_{i}^{\pi, R, M}(X_{t_{i}}^{\pi}) - \gamma_{i}^{\pi, R}(X_{t_{i}}^{\pi})|^{2} + |z_{i}^{\pi, R, M}(X_{t_{i}}^{\pi}) - z_{i}^{\pi, R}(X_{t_{i}}^{\pi})|^{2} \right] \\ + \max_{0 \le i \le n} \mathbb{E}\left[|y_{i}^{\pi, R}(X_{t_{i}}^{\pi}) - y_{i}^{\pi, R, M}(X_{t_{i}}^{\pi})|^{2} \right].$$

Indeed, as presented in Remark 2 of [73] and more in details in Theorem II.3 and Theorem II.4 p.100-106 in [72], we deduce that

$$\operatorname{Err}^{stat}{(Y, Z, U)^2} \le C \operatorname{Err}^{stat\,emp}{(Y, Z, U)^2} + CC_y(R)^2 B \, n \, M^{-1} \log(M)$$

This result is obtained using techniques of covering numbers and refer to [60] for the control of the required quantity of numbers, see in particular Theorem 11.3 in [60]. It implies that the computation of the true statistical error instead of the empirical one does not affect the rate of convergence (up to the log(M) term).

Hence, the additional parameter γ in the driver h does not change the controls on the error of the algorithm derived by Lemor [72], and the optimal calibration of the number of basis function B, Monte Carlo simulations M and time steps n is similar. We refer to [72] or [73] for their very interesting discussion on the subject, whose results depends on the choice of basis functions. For example, considering a basis of hypercubes functions, the terms of the form E(i) are of order $B^{-2/d}$. Therefore, in order to get a statistical squared error of order $n^{1-\beta}$ with $\beta \in (1,2]$ where n is the time step, one should use a localization constant R large enough, a number of Monte Carlo simulation M of order $n^{\beta+1+d\beta/2} \ln(n)$ and a number of basis functions B of order $n^{d\beta/2}$. As detailed in [57], in terms of its complexity C, the squared error of the algorithm is of order $C^{-\frac{1}{4+d}}$, and of order $C^{-\frac{1}{4+2d}}$ for the algorithm without extra simulations. This result is independent of the model of the underlying and, as a benchmark, the algorithm of Bouchard and Touzi [19] is of order $C^{-\frac{1}{13+d}}$ in the particular Geometric Brownian setting.

Finally, the global error of the algorithm is bounded from above by

$$\operatorname{Err}_{n}(Y, Z, U) + \operatorname{Err}^{loc}(Y, Z, U) + \operatorname{Err}^{stat}(Y, Z, U)$$
,

up to a multiplicative constant C. We recall from Section 2.1.1 that we can neglect the localization error $\operatorname{Err}^{loc}(Y, Z, U)$ whenever R is chosen large enough. Since the discretization error $\operatorname{Err}_n(Y, Z, U)$ is of order $n^{-1/2}$ (or eventually $n^{-1/2+\varepsilon}$) as derived in Corollary 1.2.1, one should pick $\beta = 2$ (or $2(1 - \varepsilon)$) to obtain a statistical error $\operatorname{Err}^{stat}(Y, Z, U)$ of the same order. Therefore, the global error of the algorithm is of order $n^{-1/2+\varepsilon}$ for any $\varepsilon > 0$ and attains the optimal error of order $n^{-1/2}$, under Assumption \mathbf{H}_1 or \mathbf{H}_2 .

2.1.4 Control of the statistical error

This section is devoted to the proof of Theorem 2.1.1, which is adapted from [73] without the use of extra simulations and detailed in [72]. As already mentioned, the additional argument Γ in the driver function h is handled by similar procedures used to manage Z, and the key observation relies on the existence of estimates of the form (II.2.3). For sake of completeness, we present here the main steps of the demonstration. We try to follow the notations of [73] and to emphasize the required modifications of the proof in our context.

We first introduce some notations. We fix $i \leq n$ and denote $\alpha_i^{y,1,M}$ the solution of the

OLS problem

$$\inf_{\alpha} \frac{1}{M} \sum_{m=1}^{M} \left| y_{i+1}^{\pi,R,M}(\hat{X}_{t_{i+1}}^{\pi,m}) - \alpha . p_i^y(X_{t_i}^{\pi,m}) \right|^2 \, .$$

so that $\alpha_i^{y,M} = \alpha_i^{y,1,M} + \alpha_i^{y,2,M}$, where $\alpha_i^{y,2,M}$ is the solution of the OLS problem

$$\inf_{\alpha} \frac{1}{M} \sum_{m=1}^{M} \left| \frac{1}{n} h^{R}[X_{t_{i}}^{\pi,m}, y_{i+1}^{\pi,R,M}(\hat{X}_{t_{i+1}}^{\pi,m}), z_{i}^{\pi,R,M}(X_{t_{i}}^{\pi,m}), \gamma_{i}^{\pi,R,M}(X_{t_{i}}^{\pi,m})] - \alpha p_{i}^{y}(\bar{X}_{t_{i}}^{m}) \right|^{2}.$$

Following the notations of [72], we denote $\beta_i^{y,M}$ the solution of the OLS problem

$$\inf_{\beta} \frac{1}{M} \sum_{m=1}^{M} \left| \frac{1}{n} h^{R} [X_{t_{i}}^{\pi,m}, y_{i+1}^{\pi,R}(\hat{X}_{t_{i+1}}^{\pi,m}), z_{i}^{\pi,R}(X_{t_{i}}^{\pi,m}), \gamma_{i}^{\pi,R}(X_{t_{i}}^{\pi,m})] + y_{i+1}^{\pi,R}(\hat{X}_{t_{i+1}}^{\pi,m}) - \beta \cdot p_{i}^{y}(X_{t_{i}}^{\pi,m}) \right|^{2} .$$

The only difference between the definitions of $\beta_i^{y,M}$ and $\alpha_i^{y,M}$ relies on the use of the true unknown function $(y_i^{\pi,R}, z_i^{\pi,R} \gamma_i^{\pi,R})$ instead of its approximation $(y_i^{\pi,R,M}, z_i^{\pi,R,M} \gamma_i^{\pi,R,M})$. We then define $\beta_i^{y,1,M}$, $\beta_i^{y,2,M}$, $\beta_i^{z,M}$ and $\beta_i^{\gamma,M}$ using the same transformation. We now introduce the tribe $\hat{\mathcal{F}}_{i+1}$ induced by

$$\left\{ \left(\Delta W_{j+1}^m, \, \int_E \bar{\mu}^m(de, (t_j, t_{j+1}]) \right)_{0 \le j < n}, \, \left(\Delta \hat{W}_{k+1}^m, \int_E \hat{\mu}^m(de, (t_k, t_{k+1}]) \right)_{i < k < n} \right\}_{1 \le m \le M}$$

and denote $\hat{\mathbb{E}}_{i+1} := \mathbb{E}[.|\hat{\mathcal{F}}_{i+1}]$. For any projection coefficient of the forme α_i or β_i , we use the notation $\bar{\alpha}_i := \hat{\mathbb{E}}_{i+1}(\alpha_i)$ and $\bar{\beta}_i := \hat{\mathbb{E}}_{i+1}(\beta_i)$. For any function ψ , we finally introduce

$$||\psi||_{i,\hat{M}}^2 := \frac{1}{M} \sum_{m=1}^M |\psi(\hat{X}_{t_i}^{\pi,m})|^2$$

Proof of Theorem 2.1.1. We decompose the proof in two steps.

1. Propagation of the error.

We fix $i \leq n$ and look at the dependence of the approximation error at time t_i in terms of the approximation error at time t_{i+1} , in order to control its propagation. We first remark that, for any a > 0, Young's inequality leads to

$$\begin{aligned} ||\{\bar{\beta}_{i}^{y,M} - \alpha_{i}^{y,M}\} \cdot p_{i}^{y}||_{i,M}^{2} &\leq \left(1 + \frac{a}{n}\right) ||\{\bar{\beta}_{i}^{y,1,M} - \alpha_{i}^{y,1,M}\} \cdot p_{i}^{y}||_{i,M}^{2} \qquad (\text{II.2.4}) \\ &+ \left(1 + \frac{n}{a}\right) ||\{\bar{\beta}_{i}^{y,2,M} - \alpha_{i}^{y,2,M}\} \cdot p_{i}^{y}||_{i,M}^{2} .\end{aligned}$$

But the contraction property of the projection on $(p_i^y[X_{t_i}^{\pi,m}])_{m \leq M}$ and the $\hat{\mathcal{F}}_{i+1}$ measurability of $\bar{\beta}_i^{y,1,M}.p_i^y$ leads to

$$||\{\bar{\beta}_{i}^{y,1,M} - \alpha_{i}^{y,1,M}\} \cdot p_{i}^{y}||_{i,M}^{2} \leq ||\{\alpha_{i}^{y,1,M} - \bar{\alpha}_{i}^{y,1,M}\} \cdot p_{i}^{y}||_{i,M}^{2} + ||\hat{\mathbb{E}}_{i+1}[y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}]||_{i+1,\hat{M}}^{2} \cdot$$
 (II.2.5)

Combining (II.2.4) and (II.2.5) with the 1-lipschitz property of $[.]_y$, we compute that the error of interest satisfies

$$\begin{split} \mathbb{E}||y_{i}^{\pi,R} - y_{i}^{\pi,R,M}||_{i,M}^{2} &\leq \mathbb{E}||y_{i}^{\pi,R} - \bar{\beta}^{y,M}.p_{i}^{y}||_{i,M}^{2} \qquad (\text{II.2.6}) \\ &+ \left(1 + \frac{a}{n}\right) \mathbb{E}\left[||\hat{\mathbb{E}}_{i+1}[y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}]||_{i+1,\hat{M}}^{2}\right] \\ &+ \left(1 + \frac{a}{n}\right) \mathbb{E}\left[||\{\alpha_{i}^{y,1,M} - \bar{\alpha}_{i}^{y,1,M}\}.p_{i}^{y}||_{i,M}^{2}\right] \\ &+ C\left(1 + \frac{a}{n}\right) \mathbb{E}\left[||\{\beta_{i}^{y,2,M} - \bar{\beta}_{i}^{y,2,M}\}.p_{i}^{y}||_{i,M}^{2}\right] \\ &+ C\left(1 + \frac{n}{a}\right) \mathbb{E}\left[||\{\beta_{i}^{y,2,M} - \alpha_{i}^{y,2,M}\}.p_{i}^{y}||_{i,M}^{2}\right] \end{split}$$

From the Lipschitz property of h^R , the last term on the right hand side of the previous expression satisfies

$$\begin{aligned} ||\{\beta_{i}^{y,2,M} - \alpha_{i}^{y,2,M}\} \cdot p_{i}^{y}||_{i,M}^{2} &\leq \frac{C}{n^{2}} \sum_{l=1}^{d} ||z_{l,i}^{\pi,R} - z_{l,i}^{\pi,R,M}||_{i,M}^{2} \\ &+ \frac{C}{n^{2}} \left\{ ||y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}||_{i+1,\hat{M}}^{2} + \sum_{l'=1}^{d'} ||\gamma_{l',i}^{\pi,R} - \gamma_{l',i}^{\pi,R,M}||_{i,M}^{2} \right\}. \end{aligned}$$

Since the function $[.]_{\gamma}$ is 1-Lipschitz and $\gamma_i^{\pi,R} \leq C_{\gamma}(R)$, we have, for any $l' \leq d'$,

$$\begin{aligned} ||\gamma_{l',i}^{\pi,R} - \gamma_{l',i}^{\pi,R,M}||_{i,M}^2 &\leq C \, ||\gamma_{l',i}^{\pi,R} - \bar{\beta}_{l',i}^{\gamma,M} . p_{l',i}^{\gamma}||_{i,M}^2 \\ &+ C ||\{\bar{\alpha}_{l',i}^{\gamma,M} - \alpha_{l',i}^{\gamma,M}\} . p_{l',i}^{\gamma}||_{i,M}^2 + C ||\{\bar{\beta}_{l',i}^{\gamma,M} - \bar{\alpha}_{l',i}^{\gamma,M}\} . p_{l',i}^{\gamma}||_{i,M}^2 . \end{aligned}$$
(II.2.8)

For any $l' \leq d'$, we now deduce from the definition of $\bar{\alpha}_{l',i}^{\gamma,M}$ and $\bar{\beta}_{l',i}^{\gamma,M}$, that the contraction property of the projection on $(p_{l',i}^{\gamma}[X_{t_i}^{\pi,m}])_{m\leq M}$ combined with the Cauchy Schwarz inequality, leads to

$$\begin{split} & ||\{\bar{\beta}_{l',i}^{\gamma,M} - \bar{\alpha}_{l',i}^{\gamma,M}\}p_{l',i}^{\gamma}||_{i,M}^{2} \\ & \leq \frac{n}{M}\sum_{m=1}^{M} \left|\hat{\mathbb{E}}_{i+1}\left[\{y_{i+1}^{\pi,R}(\hat{X}_{t_{i+1}}^{\pi,m}) - y_{i+1}^{\pi,R,M}(\hat{X}_{t_{i+1}}^{\pi,m})\}\int_{E}\rho(e)\hat{\mu}^{m}(de,(t_{i},t_{i+1}])\right]\right|^{2} \\ & \leq K^{2}\left\{\hat{\mathbb{E}}_{i+1}\left[||y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}||_{i+1,\hat{M}}^{2}\right] - ||\hat{\mathbb{E}}_{i+1}[y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}]||_{i+1,\hat{M}}^{2}\right\}. \end{split}$$

Combining this inequality with (II.2.8) leads to a control on the term $||\gamma_{l,i}^{\pi,R} - \gamma_{l,i}^{\pi,R,M}||_{i,M}^2$, and the exact same reasoning provides an equivalent control on $||z_{l,i}^{\pi,R} - z_{l,i}^{\pi,R,M}||_{i,M}^2$, see [72] p. 87. Reporting those estimates and (II.2.7) in (II.2.6), a particular choice of *a* allows to get rid of the terms of the form $||\hat{\mathbb{E}}_{i+1}[y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}]||_{i+1,\hat{M}}^2$, and we derive

$$\mathbb{E}||y_{i}^{\pi,R} - y_{i}^{\pi,R,M}||_{i,M}^{2} \leq \left(1 + \frac{C}{n}\right)\mathbb{E}||y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}||_{i+1,\hat{M}}^{2} \qquad (\text{II.2.9}) \\
+ C\left(T_{i,M}^{y} + T_{i,M}^{z} + T_{i,M}^{\gamma}\right),$$

where $T_{i,M}^y$, $T_{i,M}^z$ and $T_{i,M}^\gamma$ are defined by

$$\begin{split} T_{i,M}^{y} &:= & \mathbb{E}||y_{i}^{\pi,R} - \bar{\beta}^{y,M}.p_{i}^{y}]||_{i,M}^{2} + & \mathbb{E}||\{\alpha_{i}^{y,1,M} - \bar{\alpha}_{i}^{y,1,M}\}.p_{i}^{y}||_{i,M}^{2} \\ &+ & n \ \mathbb{E}||\{\beta_{i}^{y,2,M} - \bar{\beta}_{i}^{y,2,M}\}.p_{i}^{y}||_{i,M}^{2}, \\ T_{i,M}^{z} &:= & \frac{1}{n} \sum_{l=1}^{d} \mathbb{E}\left[||z_{l,i}^{\pi,R} - \bar{\beta}_{l,i}^{z,M}.p_{l,i}^{z}||_{i,M}^{2} + ||\{\bar{\alpha}_{l,i}^{z,M} - \alpha_{l,i}^{z,M}\}.p_{l,i}^{z}||_{i,M}^{2}\right], \\ T_{i,M}^{\gamma} &:= & \frac{1}{n} \sum_{l'=1}^{d'} \mathbb{E}\left[||\gamma_{l',i}^{\pi,R} - \bar{\beta}_{l',i}^{\gamma,M}.p_{l',i}^{\gamma}||_{i,M}^{2} + ||\{\bar{\alpha}_{l',i}^{\gamma,M} - \alpha_{l',i}^{\gamma,M}\}.p_{l',i}^{\gamma}||_{i,M}^{2}\right]. \end{split}$$

From Proposition 4 in [73] (or Lemma II.1, Lemma II.2 and Lemma II.3 p. 90-92 in [72]), we have

$$T_{i,M}^{y} \leq C(C_{y}(R)^{2} + ||h^{R}||_{\infty}^{2}) B M^{-1} + \inf_{\alpha} \mathbb{E}|\alpha p_{i}^{y}(X_{t_{i}}^{\pi}) - y_{i}^{\pi,R}(X_{t_{i}}^{\pi})|^{2}.$$
(II.2.10)

From Proposition 4 in [73] (or Lemma II.4 and Lemma II.5 p. 94 in [72]), we derive

$$T_{i,M}^{z} \leq C C_{y}(R)^{2} B M^{-1} + C \sum_{l=1}^{d} \inf_{\alpha} \mathbb{E} |\alpha . p_{l,i}^{z}(X_{t_{i}}^{\pi}) - n^{-1/2} z_{l,i}^{\pi,R}(X_{t_{i}}^{\pi})|^{2} .$$
(II.2.11)

We remark that changing $C_z(R)$, ΔW_i and $\Delta \hat{W}_i$ to $C_\gamma(R)$, $\int_E \rho(e)\mu(e, (t_i, t_{i+1}])$ and $\int_E \rho(e)\hat{\mu}(e, (t_i, t_{i+1}])$ in the proofs of the previous estimate, the same argument leads to

$$T_{i,M}^{\gamma} \leq C K^2 C_y(R)^2 B M^{-1} + C K^2 \sum_{l'=1}^{d'} \inf_{\alpha} \mathbb{E} |\alpha. p_{l',i}^{\gamma}(X_{t_i}^{\pi}) - n^{-1/2} \gamma_{l',i}^{\pi,R}(X_{t_i}^{\pi})|^2 .$$
(II.2.12)

Reporting (II.2.10), (II.2.11) and (II.2.12) in (II.2.9), we finally deduce

$$\mathbb{E}||y_{i}^{\pi,R} - y_{i}^{\pi,R,M}||_{i,M}^{2} \leq \left(1 + \frac{C}{n}\right) \mathbb{E}||y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}||_{i+1,\hat{M}}^{2} \qquad (\text{II.2.13}) \\
+ C(C_{y}(R)^{2} + ||h^{R}||_{\infty}^{2})BM^{-1} + E(i).$$

2. Control of $||.||_{i+1,\hat{M}}^2 - ||.||_{i+1,M}^2$.

We now fix $\beta \in (1, 2]$ and introduce the following measurable set

$$\mathcal{A}_{i}^{M} := \left\{ \forall \psi \in \mathcal{P}_{i+1}^{y}, \ ||[\psi]_{y} - y_{i+1}^{\pi,R}||_{i+1,\hat{M}} - ||[\psi]_{y} - y_{i+1}^{\pi,R}||_{i+1,M} \le n^{-\frac{\beta+1}{2}} \right\}.$$

As detailed in Theorem II.1 p. 89 in [72], the introduction of this set allows to rewrite (II.2.13) as

$$\mathbb{E}||y_{i}^{\pi,R} - y_{i}^{\pi,R,M}||_{i,M}^{2} \leq \left(1 + \frac{C}{n}\right) \mathbb{E}||y_{i+1}^{\pi,R} - y_{i+1}^{\pi,R,M}||_{i+1,M}^{2} + Cn^{-\beta}$$
(II.2.14)
+ $CC_{y}(R)^{2} n \mathbb{P}([\mathcal{A}_{i}^{M}]^{c}) + C(C_{y}(R)^{2} + ||h^{R}||_{\infty}^{2})BM^{-1} + E(i).$

By arguments based on the use of covering numbers, Lemor [72] adapts the results of Gyorgi, Kohler, Krzyzak and Walk [60], and derives an upper bound on $\mathbb{P}([\mathcal{A}_i^M]^c)$. Therefore, referring to Proposition 4 in [73], we deduce

$$\mathbb{P}([\mathcal{A}_i^M]^c) \leq C e^{CB \log(C_y(R)n^{\beta+1}) - \frac{M n^{-\beta-1}}{144C_y(R)^2}}$$

Combining this estimate with (II.2.14), we conclude the proof by applying the discrete Gronwall's lemma. $\hfill \Box$

2.2 Numerical examples

As observed in Section 1.2.5 of the previous chapter, our algorithm can be adapted to the numerical resolution of systems of coupled PDE's. Since this algorithm is to our knowledge the only probabilistic method available to solve this type of systems of PDE, we present some numerical examples in this set-up. In this section, we therefore use a discrete time scheme of the form (II.31).

2.2.1 Put option with default risk on the seller

We first present a financial application by considering the pricing of a classical put option, when the seller of this option is in addition subject to a risk of default. This exemple belongs to the class of financial derivatives mixing credit risk and equity instruments, and the pricing via BSDEs with jumps of more complex products of this type, such as convertible bonds, are currently being studied by Bielecky, Crépey, Jeanblanc and Rutkowsky [32].

Consider a market composed by a non risky asset normalized to unity and a risky asset X with Black-Scholes dynamics. We denote \mathcal{L} its associated Dynkin operator and we

have

$$\mathcal{L} : u \mapsto u_t + \frac{1}{2} \sigma^2 x^2 u_{xx} \quad \text{with} \quad dX_t = \sigma X_t dW_t,$$

with W a Brownian Motion under a well chosen probability. Let u_1 defined on $[0, 1] \times \mathbb{R}_+$ be the price function of an option delivering at time t = 1 the payoff $g_1(X_1) := (5-X_1)^+$ in the absence of default of the seller, and the capped payoff $g_0(X_1) := g_1(X_1) \wedge 5$ otherwise. The time to default τ of the seller is supposed to be independent of W and to follow an exponential law of parameter c > 0.

Following the non-arbitrage pricing theory, we assume that the price at time t = 0 of the option is given by

$$u_1(0,x) = \mathbb{E}\left[g_1(X_1)\mathbf{1}_{\tau>1} + g_0(X_1)\mathbf{1}_{\tau\leq 1} \ /X_0 = x\right].$$
(II.2.15)

Let u_0 be the price of the regular option delivering the capped payoff $g_0(X_1)$, so that

$$u_0(t,x) := \mathbb{E}\left[g_0(X_T) \ / X_t = x\right] \,.$$

Using this function u_0 , the price u_1 rewrites

$$u_1(0,x) := \mathbb{E}\left[e^{-c}g_1(X_1) + \int_0^1 ce^{-cs}u_0(s,X_s)ds / X_0 = x\right].$$

Therefore, the pair function (u_0, u_1) satisfies the following system of coupled PDEs

$$\begin{cases} \mathcal{L} u_0 = 0, & u_0(T, \cdot) = g_0, \\ \mathcal{L} u_1 = c (u_1 - u_0), & u_1(T, \cdot) = g_1. \end{cases}$$

This system has an analytic solution since u_0 can be derived first and then plugged in the second equation to deduce u_1 . We therefore have a benchmark to compare to the numerical solution.

As observed by Pardoux, Pradeilles and Rao [86], the solution of this system can be interpreted by means of the solution of a FBSDE with jump. Let first introduce a Poisson mesure μ on $[0,1] \times \{1\}$, independent of the Brownian motion W, with compensator the counting measure of the jumps multiplied by any parameter λ , representing the frequency of the jumps. We denote M the pure jump process switching between values 0 and 1 at each jump. Then, for any $t \leq 1$, $u_{M_t}(t, X_t)$ coincides with Y_t , where Y is the first component of the solution of the following BSDE with jump

$$Y_t = Y_1 + \int_t^1 \left(c \mathbf{1}_{M_s=1} - \lambda \right) U_s(1) \, ds - \int_t^1 Z_s dW_s - \int_t^1 \int_E U_s(e) \, \bar{\mu}(de, ds) \, ,$$

with terminal value $Y_1 := g_1(X_1)\mathbf{1}_{M_T=1} + g_0(X_1)\mathbf{1}_{M_T=0}$.

As detailed in Section 1.2.5, our algorithm can be adapted to the resolution of this BSDE with jump. We first simulate the pure jump process M perfectly and then use the Euler scheme of X adding the random times of jumps of M in the regular grid. Once the forward process (M, X) simulated, we compute Y backward according to the scheme (II.31). The approximation of the large number of conditional expectations is accomplished by projection on the basis of Legendre polynomials, as detailed in Section 2.1. We recall that the Legendre polynomials $(L_n)_{n \in \mathbb{N}}$ are defined on \mathbb{R} by

$$L_n(x) := \frac{1}{2^n (n!)} \nabla^n L(x) \quad \text{with} \quad L : x \mapsto (x^2 - 1)^n$$

The numerical implementation of the algorithm has been done in Visual C++, but we linked our program to the well known *LaPack* library written in Fortran, in order to have an efficient computation of the classical matrix operations required for the OLS projections. A numerical trick to improve the accuracy of the estimator consists in adding the payoff function to the bases of Legendre polynomials. The results presented in Figure 2.1 produce the true and estimated price of the option for c = 0.1 and 0.5. The numerical results observed prove that the algorithm is able to estimate the true prices. Following the theoretical study, we took 50 time steps, 10 000 Monte Carlo simulations and 5 basis functions and the relative mean square error obtained with the algorithm is of the order of 3%. Observe also that the price of the option naturally decreases when the risk of default of the seller increases.

2.2.2 Fully coupled system of PDE

Since, in the previous example, the first PDE of the system was in fact decoupled from the second one, we now consider the case where the dynamics of both PDE's depends on the solution of the other. We look for the pair function (u_0, u_1) defined on $[0, 1] \times \mathbb{R}_+$ as the solution of

$$\begin{cases} \mathcal{L}u_0 = u_1, & u_0(1, \cdot) = g_0, \\ \mathcal{L}u_1 = u_0, & u_1(1, \cdot) = g_1. \end{cases}$$

Remark that the pair function $(u_0 + u_1, u_0 - u_1)$ satisfies in fact a decoupled system of PDE which allows to compute the analytical value of the pair solution. With the previous notations, the solution (u_0, u_1) is related to the solution of the following BSDE with jump

$$Y_t = Y_1 + \int_t^1 \left[-Y_s + (1+\lambda)U_s(1) \right] \, ds - \int_t^1 Z_s dW_s - \int_t^1 \int_E U_s(e) \,\bar{\mu}(de, ds) \, ,$$



Figure 2.1: Price of a Put option with default risk on the seller given by c = 0.1 or 0.5

with terminal value $Y_1 := g_1(X_1) \mathbf{1}_{M_T=1} + g_0(X_1) \mathbf{1}_{M_T=0}$.

As provided in Figure 2.2, the algorithm still allows to recover the true value functions. With 10 000 Monte Carlo simulations, 50 time steps and 5 basis functions, the integrated relative mean square error of the algorithm is of the order of 5%. Remark that, in order to obtain the solution of both PDEs, the resolution of only one BSDE with jump is necessary. It suffices to divide the Monte Carlo simulations in two sets, one where M starts from 0 and the other where it starts from 1. Considering examples of this form and letting the dynamics of X depend on M, which is possible with our algorithm, allows to price options on an underlying with two different dynamics switching from one to the other as time goes by. Then, the jump process M characterizes the trend and volatility of each dynamic of the asset. Successful numerical results were also obtained in this set up but we prefer to present now a more complex numerical example relying on the resolution of a system of semi-linear PDE's.



Figure 2.2: Solution of the fully coupled system of PDE's



Figure 2.3: Solution of the coupled system of semi-linear PDE's

2.2.3 A more complex example

We consider the following system of semi-linear PDE's

$$\begin{cases} \mathcal{L}u_0 + x\sigma\nabla_x u_0 = \sqrt{1 + (u_1)^2}, & u_0(1, \cdot) = g_0, \\ \mathcal{L}u_1 + x\sigma\nabla_x u_1 = \sqrt{1 + (u_0)^2}, & u_1(1, \cdot) = g_1. \end{cases}$$

Its particular interest relies on the necessity of estimating the component Z to solve the corresponding BSDE with jump given by

$$Y_t = Y_1 - \int_t^1 \left(\sqrt{1 + (Y_s + U_s(1))^2} + \lambda U_s(1) - Z_s \right) \, ds - \int_t^1 Z_s dW_s$$
$$- \int_t^1 \int_E U_s(e) \, \bar{\mu}(de, ds) \,,$$

with terminal value $Y_1 := g_1(X_1)\mathbf{1}_{M_T=1} + g_0(X_1)\mathbf{1}_{M_T=0}$. Indeed, the previous theoretical study of the discretization error, we observed that the required approximation of Z reduces the speed of the numerical scheme.

We report in Figure 2.3 the smoothed estimations given by the algorithm which are coherent with the expected results, even if we do not have a benchmark because we can not compute explicitly the analytical value of the pair function (u_0, u_1) solution of the system of PDE. Several tests with different set of parameters showed the convincing stability of the result. For example, we provide in Figure 2.4 the estimations obtained with 5 basis functions and different numbers of Monte Carlo simulations M and time steps n. Observe that this Figure presented with a very small scaling shows the accuracy of the estimation. As for the influence of the choice of the parameters, taking the value given by the algorithm with a large number of simulations and time steps as a benchmark, changing for example the number of simulations from 10 000 to 50 000 with a fixed number of 50 time steps induces a decrease of the integrated mean square error of the algorithm from 5% to 2%.

Finally, we observe that the parameter λ , representing the frequency of jumps, needs to be chosen carefully. If λ is too small, the process Y does not jump often enough and the algorithm has difficulties to capture the dynamics of both solutions u_0 and u_1 . If λ is too large, there are too many jumps on each time step, and both proposed solutions look like a sort of mixture between the two real ones. The choice of λ is for sure closely related to the value of the time step. The additional difficulty in the theoretical study of the influence of λ relies on the fact that the Lipschitz constant of the driver h depends on λ . The investigation on the optimal choice of λ is left for further research.



Figure 2.4: Influence of the parameters on the resolution of coupled system of semi-linear of PDE's

Part III

Optimal consumption-investment strategy under drawdown constraint

Abstract

We consider the optimal consumption-investment problem under the drawdown constraint, i.e. the wealth process never falls below a fixed fraction of its running maximum. We assume that the risky asset is driven by the constant coefficients Black and Scholes model and we consider a general class of utility functions. On an infinite time horizon, we provide the value function in explicit form, and we derive closed-form expressions for the optimal consumption and investment strategy. The key ingredient for the obtention of the solution relies on the linearity of the PDE satisfied by the dual transform of the value function. On a finite time horizon, we interpret the value function as the unique viscosity solution of its corresponding Hamilton-Jacobi-Bellman equation. This leads to a consistent numerical scheme of approximation and allows for a comparison with the explicit solution in infinite horizon.

Keywords: consumption-investment strategy, drawdown constraint, Fenchel transform, asymptotic elasticity, viscosity solution, comparison principle.

Note

The first chapter of this part is based on a paper, written in collaboration with Nizar Touzi, submitted to *Finance and Stochastics*.

152 CONSUMPTION-INVESTMENT UNDER DRAWDOWN CONSTRAINT

Chapter 1

Explicit solution in infinite time horizon

1.1 Introduction

Since the seminal papers of Merton [79, 80], there has been an extensive literature on the problem of optimal consumption and investment decision in financial markets subject to imperfections. The case of incomplete markets was first considered by Cox and Huang [29] and Karatzas, Lehoczky and Shreve [64]. Cvitanić and Karatzas [33] considered the case where the agent portfolio is restricted to take values in some given closed convex set. He and Pagès [62] and El Karoui and Jeanblanc [44] extended the Merton model to allow for the presence of labor income. Constantinides and Magill [27], Davis and Norman [35], and Shreve and Soner [97] considered the case where the risky asset is subject to proportional transaction costs. Ben Tahar, Soner and Touzi [10] considered the case where the sales of the risky asset are subject taxes on the capital gains.

In this chapter, we study the infinite horizon optimal consumption and investment problem when the wealth never falls below a fixed fraction of its current maximum. This is the so-called drawdown constraint. Fund managers do offer this type of guarantee in order to satisfy the aversion to deception of the investors.

The drawdown constraint on the wealth accumulation of the fund manager was first considered by Grossman and Zhou [59] for an agent maximizing the long term growth rate of the expected power utility of final wealth, with no intermediate consumption. Their main result is that the optimal investment in the risky asset is an explicit constant proportion of the difference between the current wealth and the imposed fixed fraction of its running maximum. Klass and Nowicki [66] show that the strategy proposed in Grossman and Zhou [59] does not retain its optimal long term growth property when generalized to the discrete time setting. Nevertheless, Cvitanic and Karatzas [34] developed a beautiful martingale approach to the Grossman and Zhou [59] problem which makes the analysis much simpler and allows for more general class of price processes. Their main observation is that strategies based on investment in proportions of the distance between the current wealth and its drawdown constraint, are always admissible. Besides, El Karoui and Meziou [43] recently characterized the optimal portfolio obtained by Cvitanic and Karatzas [34] in terms of Azema-Yor martingales, opening the door to the study of non linear drawdown constraints. A general criticism that one may formulate about the long term growth rate criterion is that it only provides the asymptotic optimal behavior of the fund manager. In other words, there is no penalization for using an arbitrary strategy as long as it coincides with the Grossman and Zhou [59] optimal strategy after some given fixed point in time.

In this chapter, we consider the classical Merton criterion, which consists in maximizing the infinite horizon utility of consumption, for a fund manager subject to the drawdown constraint. This problem was considered recently by Roche [93] in the context of the power utility function. Following the initial Merton approach, Roche [93] was able to guess a solution of the dynamic programming equation, and provided some numerical results which highlight some interesting consequences of the drawdown constraint on the optimal consumption-investment strategy. The homogeneity of the power utility is the key-property in order to guess the candidate solution. Notice that Roche [93] does not provide any argument to verify that his candidate solution is indeed the value function of the optimal consumption-investment problem.

In contrast with Roche [93], our analysis allows for a general class of utility functions whose asymptotic elasticity (see Kramkov and Schachermayer [70]) is bounded by some level depending on the drawdown level, and satisfying some condition related to the relative risk aversion. For any utility function in this class, we derive an explicit expression for the value function of the fund manager, together with the optimal consumption and investment strategy. The key-idea in order to guess the candidate solution is to pass from the dynamic programming equation to the partial differential equation (PDE) satisfied by the dual indirect utility function. The latter PDE being linear inside the state space domain, one can easily account for the Neumann condition related to the drawdown constraint, and derive an explicit candidate solution is indeed the value function. In order to prove that the thus derived candidate solution is indeed the value function of our optimal consumption-investment problem, we use a verification argument which requires a convenient transversality condition. The verification argument is the main technical

154

step where the above mentioned restrictions on the utility functions are required.

The solution derived in this chapter agrees with that of Roche [93] in the zero interest rate and power utility case. However, for positive interest rates, we follow Cvitanic and Karatzas [34] by defining the drawdown constraint in terms of the discounted wealth.

The chapter is organized as follows. Section 1.2 is devoted to the formulation of the problem. The main result of the chapter is provided in Section 1.3. Section 1.4 presents the formal argument that we used in order to guess our candidate solution. The rigorous proof of our main result is reported in Section 1.5.

1.2 Problem formulation

We consider a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ endowed with a Brownian motion $W = \{W_t, t \geq 0\}$ valued in \mathbb{R} , and we denote by $\mathbb{F} := \{\mathcal{F}_t, t \geq 0\}$. The financial market consists of a non-risky asset, with process normalized to unity, and one risky asset with price process defined by the Black and Scholes model :

$$dS_t = \sigma S_t \left(dW_t + \lambda dt \right) \,,$$

where $\sigma > 0$ is the volatility parameter, and $\lambda \in \mathbb{R}$ is a constant risk premium.

The normalization of the non-risky asset to unity is as usual a reduction of the model obtained by taking this asset as a numéraire. Hence, all amounts are evaluated in terms of their discounted values.

For any continuous process $\{M_t, t \ge 0\}$, we shall denote by

$$M_t^* := \sup_{0 \le r \le t} M_r, \qquad t \ge 0,$$

the corresponding running maximum process, and we recall that

$$M^*$$
 is non-decreasing and $\int_0^\infty (M_t^* - M_t) \, dM_t^* = 0.$ (III.1)

1.2.1 Consumption-portfolio strategies and the drawdown constraint

We next introduce the set of consumption-investment strategies whose induced wealth process X satisfies the drawdown constraint

$$X_t \ge \alpha X_t^*$$
, for every $t \ge 0$, a.s., (III.2)

where α is some given parameter in the interval [0, 1).

A portfolio strategy is an \mathbb{F} -adapted process $\theta = \{\theta_t, t \ge 0\}$, with values in \mathbb{R} , satisfying the integrability condition

$$\int_0^T |\theta_t|^2 dt < \infty \quad \text{a.s.} \quad , \quad \text{for all} \quad T > 0 \,. \tag{III.3}$$

A consumption strategy is an \mathbb{F} -adapted process $C = \{C_t, t \ge 0\}$, with values in \mathbb{R}_+ , satisfying

$$\int_0^T C_t dt < \infty \quad \text{a.s.} \quad , \quad \text{for all} \quad T > 0 \,. \tag{III.4}$$

Here, θ_t and C_t denote respectively the amount invested in the risky asset and the consumption rate at time t. By the self-financing condition, the wealth process induced by such a pair (C, θ) is defined by

$$X_t^{x,C,\theta} = x - \int_0^t C_r dr + \int_0^t \sigma \theta_r \left(dW_r + \lambda dr \right), \quad t \ge 0, \quad (\text{III.5})$$

where x is some given initial capital. We shall denote by $\mathcal{A}_{\alpha}(x)$ the collection of all such consumption-investment strategies whose corresponding wealth process satisfies the drawdown constraint (III.2).

Remark 1.2.1 For a given initial wealth x and an admissible consumption-investment strategy $(C, \theta) \in \mathcal{A}_{\alpha}(x)$, let $X := X^{x,C,\theta}$ and $\tau := \inf \{t > 0 : X_t = \alpha X_t^*\}$.

• Denoting \mathbb{P}^{λ} the probability measure under which the process $\{W_t^{\lambda} := W_t + \lambda t, t \ge 0\}$ is a Brownian motion, we see that, for $t \ge 0$,

$$\mathbb{E}^{\mathbb{P}^{\lambda}}\left[\int_{\tau}^{\tau+t} C_r dr |\mathcal{F}_{\tau}\right] = \mathbb{E}^{\mathbb{P}^{\lambda}}\left[\alpha X_{\tau}^* - X_{\tau+t} |\mathcal{F}_{\tau}\right] \leq 0, \quad \text{on } \{\tau < \infty\}$$

This shows that $\mathbb{E}\left[\int_{\tau}^{\infty} C_r dr\right] = 0.$

• Then $X_{\tau+t} = X_{\tau} + \int_{\tau}^{\tau+t} \sigma \theta_r dW_r^{\lambda}$ on $\{\tau < \infty\}$, and in order for the drawdown constraint to be satisfied, it is necessary that $\int_{\tau}^{\infty} |\theta_r|^2 dr = 0$.

1.2.2 A subset of admissible strategies

In order to ensure that the drawdown constraint is satisfied, one may define the consumption and the investment decisions in terms of proportions of the difference $X_t - \alpha X_t^*$:

$$C_t = c_t \left[X_t - \alpha X_t^* \right] \quad \text{and} \quad \theta_t = \pi_t \left[X_t - \alpha X_t^* \right], \tag{III.6}$$

for an \mathbb{F} -adapted pair process (c, π) with values in $\mathbb{R}_+ \times \mathbb{R}$. We shall denote in this subsection by $\{X^{x,c,\pi}_{\alpha}(t), t \geq 0\}$ the corresponding wealth process with initial capital x,

where the time variable appears in parenthesis, in order to highlight the dependence on α .

Under the self-financing condition, the dynamics of this process is given by

$$dX_{\alpha}^{x,c,\pi}(t) = (X_{\alpha}^{x,c,\pi}(t) - \alpha \{X_{\alpha}^{x,c,\pi}\}^{*}(t)) \left(\pi_{t} \frac{dS_{t}}{S_{t}} - c_{t}dt\right), \quad t \ge 0.$$
(III.7)

The following argument reported from Cvitanić and Karatzas [34] shows that for any $\alpha \in [0, 1)$, and for any \mathbb{F} -adapted processes (c, π) with values in $\mathbb{R}_+ \times \mathbb{R}$ satisfying

$$\int_{0}^{T} c_{t} dt + \int_{0}^{T} |\pi_{t}|^{2} dt < \infty, \quad \text{for any} \quad T > 0, \qquad (\text{III.8})$$

the stochastic differential equation (III.7) has a unique solution satisfying the drawdown condition (III.2), which turns out to be explicit.

First, in the absence of the drawdown constraint, i.e. $\alpha = 0$, the stochastic differential equation (III.7) is well-known to have the following unique solution

$$X_0^{x,c,\pi}(t) = x \exp\left[\int_0^t \left(-c_r + \lambda \sigma \pi_r - \frac{1}{2}|\sigma \pi_r|^2\right) dr + \int_0^t \sigma \pi_r dW_r\right], \quad t \ge 0,$$

for every initial capital x > 0 and every consumption-investment strategy (c, π) satisfying (III.8).

Now, the key ingredient for the construction of a solution to (III.7) is to introduce the process

$$\tilde{X}_{\alpha}^{x,c,\pi}(t) := \left[X_{\alpha}^{x,c,\pi}(t) - \alpha \left\{X_{\alpha}^{x,c,\pi}\right\}^{*}(t)\right] \left[\left\{X_{\alpha}^{x,c,\pi}\right\}^{*}(t)\right]^{\frac{\alpha}{1-\alpha}}, \quad t \ge 0.$$
(III.9)

By Itô's Lemma together with (III.1), it follows that

$$d\tilde{X}_{\alpha}^{x,c,\pi}(t) = [\{X_{\alpha}^{x,c,\pi}\}^{*}(t)]^{\frac{\alpha}{1-\alpha}} \left(\frac{\alpha}{1-\alpha} \left[\frac{X_{\alpha}^{x,c,\pi}(t)}{\{X_{\alpha}^{x,c,\pi}\}^{*}(t)} - 1\right] d\{X_{\alpha}^{x,c,\pi}\}^{*}(t) + dX_{\alpha}^{x,c,\pi}(t)\right)$$
$$= \tilde{X}_{\alpha}^{x,c,\pi}(t) \left[(\lambda\sigma\pi_{t} - c_{t}) dt + \sigma\pi_{t} dW_{t}\right].$$
(III.10)

Since the dynamics of $\tilde{X}^{x,c,\pi}_{\alpha}$ are independent of α , we derive

$$\tilde{X}_{\alpha}^{x,c,\pi} = \tilde{X}_{0}^{x(\alpha),c,\pi} = X_{0}^{x(\alpha),c,\pi} \text{ with } x(\alpha) := \tilde{X}_{\alpha}^{x,c,\pi}(0) = (1-\alpha)x^{1/(1-\alpha)}.$$
(III.11)

We next deduce from (III.9) that, for every $r \leq t$,

$$X_0^{x(\alpha),c,\pi}(r) \leq (1-\alpha) \{X_\alpha^{x,c,\pi}\}^*(r)^{1/(1-\alpha)} \leq (1-\alpha) \{X_\alpha^{x,c,\pi}\}^*(t)^{1/(1-\alpha)}.$$
 (III.12)

At a point of maximum r^* of the process $X^{x,c,\pi}_{\alpha}$ on [0,t], the previous inequality becomes an equality so that finally

$$\left\{X_0^{x(\alpha),c,\pi}\right\}^*(t) = (1-\alpha) \left\{X_\alpha^{x,c,\pi}\right\}^*(t)^{1/(1-\alpha)}.$$
 (III.13)

Combining (III.9), (III.11) and (III.13) finally leads to

$$X_{\alpha}^{x,c,\pi} = \left[X_{0}^{x(\alpha),c,\pi} + \frac{\alpha}{1-\alpha} \left\{X_{0}^{x(\alpha),c,\pi}\right\}^{*}\right] \left(\frac{\left\{X_{0}^{x(\alpha),c,\pi}\right\}^{*}}{1-\alpha}\right)^{-\alpha}.$$
 (III.14)

Since (c, π) satisfies (III.8), $X_0^{x(\alpha),c,\pi}$ is well defined and the above argument shows that the right hand side of (III.14) is the unique solution of (III.7), as one can check by an immediate application of Itô's lemma. Remark also from (III.10) that $\tilde{X}_{\alpha}^{x,c,\pi}$ is positive so that the solution of (III.7) necessarily satisfies the drawdown condition (III.2). Hence, for any pair (c, π) of \mathbb{F} -adapted processes, with values in $\mathbb{R}_+ \times \mathbb{R}$, and satisfying

(III.8), the pair process (C, θ) defined by (III.6) is an admissible consumption-investment strategy in $\mathcal{A}_{\alpha}(x)$.

1.2.3 The optimal consumption-investment problem

The previous paragraph shows in particular that, for any initial capital x, the set $\mathcal{A}_{\alpha}(x)$ contains non-trivial consumption-investment strategies.

We now formulate the optimal consumption-investment problem which will be the focus of this chapter. Throughout this chapter, we consider a utility function

$$U: \mathbb{R}_+ \to \mathbb{R}$$
 C^2 , concave, satisfying $U'(0+) = \infty$ and $U'(\infty) = 0$. (III.15)

More conditions on U will be needed for our main result, see subsection 1.3.3 below. For a given initial capital x > 0, the optimal consumption-investment problem under drawdown constraint is defined by :

$$u_0^{\alpha} := \sup_{(C,\theta)\in\mathcal{A}_{\alpha}(x)} J_0^{\alpha}(C,\theta) \quad \text{where} \quad J_0^{\alpha}(C,\theta) := \mathbb{E}\left[\int_0^{\infty} e^{-\beta t} U(C_t) dt\right], \quad (\text{III.16})$$

where $\beta > 0$ is the subjective discount factor which expresses the preference of the agent for the present. For $\alpha = 0$, u_0^0 reduces to the classical Merton optimal consumptioninvestment problem. We shall use the dynamic programming approach in order to derive an explicit solution of the problem u_0^{α} . We then need to introduce the dynamic version of this problem :

$$u^{\alpha}(x,z) := \sup_{(C,\theta)\in\mathcal{A}_{\alpha}(x,z)} J^{\alpha}(C,\theta) \quad \text{where} \quad J^{\alpha}(C,\theta) := \mathbb{E}\left[\int_{0}^{\infty} e^{-\beta t} U(C_{t}) dt\right], \quad (\text{III.17})$$

the pair (x, z), with $x \leq z$, stands for the initial condition of the state processes (X, Z) defined, for $t \geq 0$, by

$$Z_t^{x,z,C,\theta} := z \vee \left\{ X^{x,C,\theta} \right\}_t^* \quad \text{and} \quad X_t^{x,C,\theta} = x - \int_0^t C_r dr + \int_0^t \sigma \theta_r \left(dW_r + \lambda dr \right) \,, \, (\text{III.18})$$

and $\mathcal{A}_{\alpha}(x, z)$ is the collection of all \mathbb{F} -adapted processes (C, θ) satisfying (III.3)-(III.4) together with the drawdown constraint

$$X_t^{x,C,\theta} \ge \alpha Z_t^{x,z,C,\theta} \text{ a.s.}, \quad t \ge 0.$$
 (III.19)

Clearly, avoiding the trivial case x = z = 0, this restricts the pair of initial condition (x, z) to the closure $\overline{\mathbf{D}}_{\alpha}$ in $(0, \infty) \times (0, \infty)$ of the domain

$$\mathbf{D}_{\alpha} := \{ (x, z) : 0 < \alpha z < x \le z \} .$$
 (III.20)

By the same argument as in Remark 1.2.1,

$$J^{\alpha}(C,\theta) = \mathbb{E}\left[\int_0^{\tau} e^{-\beta t} U(C_t) dt + \frac{U(0)}{\beta} e^{-\beta \tau}\right], \qquad (\text{III.21})$$

where

$$\tau := \inf \left\{ t > 0 : X_t^{x,C,\theta} = \alpha Z_t^{x,z,C,\theta} \right\} .$$

In particular, this implies that

$$u^{\alpha}(x,z) = U(0)/\beta \quad \text{for} \quad (x,z) \in \overline{\mathbf{D}}_{\alpha} \setminus \mathbf{D}_{\alpha}.$$
 (III.22)

We conclude this subsection by stating the following concavity property of the value function u^{α} , as observed in [93]. This argument can be skipped by the reader as it is not needed for the proof of our main result.

Lemma 1.2.1 For any z > 0, the function $u^{\alpha}(., z)$ is concave.

Proof. Let $\nu \in [0,1]$ and a triplet (x, x', z) satisfying $(x, z) \in \overline{\mathbf{D}}_{\alpha}$ and $(x', z) \in \overline{\mathbf{D}}_{\alpha}$. Take $(C, \theta) \in \mathcal{A}_{\alpha}(x, z)$ and $(C', \theta') \in \mathcal{A}_{\alpha}(x', z)$. For any $t \ge 0$, we have

$$\nu X_t^{x,C,\theta} + (1-\nu)X^{x',C',\theta'} \geq \nu \alpha z \vee \left\{ X^{x,C,\theta} \right\}_t^* + (1-\nu) \alpha z \vee \left\{ X^{x',C',\theta'} \right\}_t^*$$
$$\geq \alpha z \vee \left\{ \nu X^{x,C,\theta} + (1-\nu)X^{x',C',\theta'} \right\}_t^*,$$

so that, from the linearity of equation (III.5), we deduce

$$\left(\nu C + (1-\nu)C', \nu\theta + (1-\nu)\theta'\right) \in \mathcal{A}_{\alpha}\left(\nu x + (1-\nu)x', z\right) .$$

Now, since J^{α} defined in (III.17) inherits the concavity of U, we get

$$\nu J^{\alpha}(C,\theta) + (1-\nu)J^{\alpha}(C',\theta') \leq J^{\alpha}(\nu C + (1-\nu)C',\nu\theta + (1-\nu)\theta')$$
$$\leq u^{\alpha}(\nu x + (1-\nu)x',z),$$

and taking the maximum over (C, θ) and (C', θ') concludes the proof.

1.3 The main results

1.3.1 The corresponding dynamic programming equation

The optimal consumption-investment problem (III.17) is in the class of stochastic control problems studied in Barles, Daher and Romano [6]. The dynamic programming equation is related to the second order operator

$$\mathcal{L}u := \beta u - \sup_{C \ge 0, \theta \in \mathbb{R}} \left[U(C) + (\theta \sigma \lambda - C)u_x + \frac{\theta^2 \sigma^2}{2} u_{xx} \right].$$
(III.23)

Defining the Legendre-Fenchel transform

$$V(y) := \sup_{x \ge 0} (U(x) - xy)$$
 (III.24)

and, recalling the concavity property of u^{α} stated in Lemma 1.2.1, the above dynamic programming equation simplifies to

$$\mathcal{L}u = \beta u - V(u_x) + \frac{\lambda^2}{2} \frac{u_x^2}{u_{xx}} \quad \text{whenever} \quad u \text{ is strictly concave.} \qquad (\text{III.25})$$

with maximizers in (III.23) given by

$$\hat{C} = -V'(u_x) = (U')^{-1}(u_x) \text{ and } \hat{\theta} := -\frac{\lambda}{\sigma} \frac{u_x}{u_{xx}}.$$
 (III.26)

Under some convenient smoothness conditions, we expect the value function u^{α} to solve the following dynamic programming equation

$$\mathcal{L}u^{\alpha}(x,z) = 0, \text{ for } (x,z) \in \mathbf{D}_{\alpha};$$
 (III.27)

$$u^{\alpha}(\alpha z, z) = 0, \quad \text{for} \quad z \ge 0; \tag{III.28}$$

$$u_z^{\alpha}(z,z) = 0, \text{ for } z > 0.$$
 (III.29)

We refer to [6] for the rigorous derivation of this dynamic programming equation in the viscosity sense. Since we will be using a verification argument in this chapter, we only need to start from this partial differential equation, and "guess" a candidate solution for it.

1.3.2 The Fenchel-Legendre dual functions

The key-ingredient in order to derive the explicit solution in this chapter is to introduce the Legendre-Fenchel transforms of the value function u^{α} with fixed z :

$$v^{\alpha}(y,z) := \sup_{x \ge 0} (u^{\alpha}(x,z) - xy) .$$
 (III.30)

Since the value function u^{α} is concave in its first variable, it can indeed be recovered from v^{α} by the duality relation

$$u^{\alpha}(x,z) = \inf_{y \in \mathbb{R}} \left(v^{\alpha}(y,z) + xy \right) .$$
 (III.31)

In the absence of drawdown constraint, the functions u^0 and v^0 are independent of the z variable and the dual function v^0 can be obtained explicitly in terms of the density of the risk-neutral measure. This can be seen by the following formal PDE argument: assuming that u^0 is smooth and satisfies the Inada conditions $(u^0)'(0+) = +\infty$, $(u^0)'(\infty) = 0$, it follows that

$$v^{0}(y) = u^{0}\left(\left[(u^{0})'\right]^{-1}(y)\right) - y\left[(u^{0})'\right]^{-1}(y), \text{ for } y \ge 0,$$
 (III.32)

and $v^0(y) = \infty$ for y < 0. Substituting in the dynamic programming equation (III.27), it follows that v^0 solves on $(0, \infty)$ the linear parabolic partial differential equation

$$\mathcal{L}^* v(y) := \beta v(y) - \beta y v_y(y) - \frac{\lambda^2}{2} y^2 v_{yy}(y) = V(y).$$
 (III.33)

Under a convenient transversality condition, this provides

$$v^{0}(y) = \mathbb{E}\left[\int_{0}^{\infty} e^{-\beta t} V\left(e^{\beta t} Y_{t}\right) dt\right] \quad \text{where} \quad Y_{t} := y \exp\left(-\lambda W_{t} - \frac{1}{2}\lambda^{2}t\right). \quad (\text{III.34})$$

In the particular case of a power utility function, this relation allows to derive explicitly v^0 and u^0 as detailed at the beginning of section 1.3.5. This result is well-known in the financial mathematics literature, and can be proved rigourously by probabilistic arguments, see e.g. [65].

In this complete market setting, it is remarkable that the Fenchel transform v^0 solves a linear PDE. This is the key-observation in order to guess a candidate solution for the optimal consumption-investment problem under drawdown constraint.

1.3.3 Assumptions

In this subsection, we collect the assumptions needed for our main result. Our first condition concerns the parameter

$$\gamma := \frac{2\beta}{\lambda^2}.$$

Assumption 1.3.1 $\frac{\gamma}{1+\gamma} < 1-\alpha.$

Observe that this condition is automatically satisfied when $\alpha = 0$. Under this condition, we may introduce the positive parameter

$$\delta := \frac{\gamma}{1 - \alpha(1 + \gamma)} \quad \text{so that} \quad \frac{\gamma}{1 + \gamma} = (1 - \alpha) \frac{\delta}{1 + \delta}, \quad (\text{III.35})$$

and we may express Assumption 1.3.1 in the equivalent form

$$\delta > 0. \tag{III.36}$$

Our next condition concerns the so-called asymptotic elasticity of the utility function U

$$\operatorname{AE}(U) := \limsup_{x \to \infty} \frac{xU'(x)}{U(x)},$$

as introduced by [70, 96].

Assumption 1.3.2
$$AE(U) < \frac{\delta}{1+\delta}$$

In view of (III.36), Assumption 1.3.2 is stronger than the usual reasonable asymptotic elasticity condition. From Lemma 6.5 in [70], we deduce the existence of a constant K_0 such that

$$U(x) \leq K_0 \left(1 + \frac{x^{\bar{p}}}{\bar{p}}\right), \quad x \geq 0, \text{ where } p := AE(U).$$
 (III.37)

Furthermore, since U and V satisfy the relation

$$U(x) = V\left([-V']^{-1}(x)\right) + x\left[-V'\right]^{-1}(x), \qquad x \ge 0,$$

where both terms on the right hand side are positive, it follows from (III.37) together with the fact that $U'(\infty) = 0$ that is

$$\limsup_{y \to 0} -V'(y)y^{\frac{1}{1-p}} < \infty \quad \text{and} \quad \limsup_{y \to 0} V(y)y^{\frac{p}{1-p}} < \infty.$$

In particular, this ensures the following integrability properties

$$\int_0^1 -V'(s)s^{\delta}ds < \infty \quad \text{and} \quad \int_0^1 V(s)s^{\delta-1}ds < \infty.$$
(III.38)

Our final assumption on the utility function is

Assumption 1.3.3
$$\inf_{y>0} \left\{ \frac{1}{yV''(y)} \int_0^y \frac{-V'(s)}{s} \left(\frac{s}{y}\right)^{1+\delta} ds \right\} > 0.$$

Remark 1.3.1 Let Assumptions 1.3.1 and 1.3.2 hold. Then, Assumption 1.3.3 is satisfied whenever the relative risk aversion of U is uniformly bounded from below. Indeed, if there exist C' > 0 such that $-xU''(x) \ge C'U'(x)$ for any x > 0, then we deduce $C'yV''(y) \le -V'(y)$, for any y > 0, and the monotonicity of V' leads to Assumption 1.3.3.

162

1.3.4 Explicit solution under drawdown constraint

According to (III.38), under Assumptions 1.3.1 and 1.3.2, the function

$$g(\zeta) := \frac{\delta}{\beta(1+\delta)} \left(\int_0^{\zeta} \frac{-V'(s)}{s} \left(\frac{s}{\zeta}\right)^{1+\delta} ds + \int_{\zeta}^{\infty} \frac{-V'(s)}{s} ds \right), \quad \zeta > 0, \quad \text{(III.39)}$$

is a well defined positive C^1 function from $(0,\infty)$ to $(0,\infty)$, with negative derivative

$$g'(\zeta) = -\frac{\delta}{\beta\zeta} \int_0^{\zeta} \frac{-V'(s)}{s} \left(\frac{s}{\zeta}\right)^{1+\delta} ds < 0, \qquad \zeta > 0.$$
(III.40)

We denote $\varphi := g^{-1}$ its inverse which is a C^1 decreasing positive function from $(0, \infty)$ to $(0, \infty)$ defined implicitly by the relation

$$z := \frac{\delta}{\beta(1+\delta)} \left(\int_0^{\varphi(z)} \frac{-V'(s)}{s} \left(\frac{s}{\varphi(z)}\right)^{1+\delta} ds + \int_{\varphi(z)}^{\infty} \frac{-V'(s)}{s} ds \right), \quad z > 0.$$
(III.41)

We now introduce the function

$$h(y,z) := \alpha z + \frac{\gamma}{\beta(1+\gamma)} \left(\frac{\varphi(z)}{y}\right)^{1+\gamma} \int_0^{\varphi(z)} \frac{-V'(s)}{s} \left(\frac{s}{\varphi(z)}\right)^{1+\delta} ds \qquad \text{(III.42)}$$
$$+ \frac{\gamma}{\beta(1+\gamma)} \left\{ \int_{\varphi(z)}^y \frac{-V'(s)}{s} \left(\frac{s}{y}\right)^{1+\gamma} ds + \int_y^\infty \frac{-V'(s)}{s} ds \right\}, \quad y \ge \varphi(z).$$

Lemma 1.3.1 Let Assumptions 1.3.1 and 1.3.2 hold. For any z > 0, the function h(.,z) is invertible and its inverse denoted f(.,z) is a strictly decreasing C^1 function from $(\alpha z, z]$ to $[\varphi(z), \infty)$ whose derivative satisfies

$$-\frac{f_x(x,z)}{f(x,z)} = \left((\gamma+1)(x-\alpha z) + \frac{\gamma}{\beta} \int_{f(x,z)}^{\infty} \frac{V'(s)}{s} ds \right)^{-1}, \quad (x,z) \in \mathbf{D}_{\alpha}.$$
(III.43)

Proof. Fix z > 0. The function h(., z) is C^1 on $(\varphi(z), \infty)$ and

$$h_y(y,z) = -\frac{\gamma}{\beta y} \left\{ \left(\frac{\varphi(z)}{y}\right)^{1+\gamma} \int_0^{\varphi(z)} \frac{-V'(s)}{s} \left(\frac{s}{\varphi(z)}\right)^{1+\delta} ds + \int_{\varphi(z)}^y \frac{-V'(s)}{s} \left(\frac{s}{y}\right)^{1+\gamma} ds \right\}$$

which is strictly negative. Therefore, since $h(\varphi(z), z) = z$ and $h(\infty, z) = \alpha z$, h is invertible and its inverse f(., z) is a strictly decreasing C^1 function from $(\alpha z, z]$ to $[\varphi(z), \infty)$. Simple computation then leads to (III.43).

We now introduce our candidate feedback solutions for the consumption-investment problem:

$$\begin{cases} \hat{C}(x,z) := -[V' \circ f](x,z) \\ \hat{\theta}(x,z) := \frac{\lambda}{\sigma} \left((\gamma+1)(x-\alpha z) - \frac{\gamma}{\beta} \int_{f(x,z)}^{\infty} \frac{-V'(s)}{s} ds \right) \end{cases}$$
(III.44)

for $(x,z) \in \mathbf{D}_{\alpha}$, and $\hat{C}(x,z) = \hat{\theta}(x,z) = 0$ on $\overline{\mathbf{D}}_{\alpha} \setminus \mathbf{D}_{\alpha}$.

Lemma 1.3.2 Let Assumptions 1.3.1, 1.3.2 and 1.3.3 hold. Then, the functions \hat{C} and $\hat{\theta}$ are Lipschitz on $\overline{\mathbf{D}}_{\alpha}$.

The proof of this lemma requires precise regularity properties of the function f and is reported in Section 1.5.2. Given an initial condition $(x, z) \in \overline{\mathbf{D}}_{\alpha}$, we consider the stochastic differential equation

$$d\hat{X}_t = -\hat{C}(\hat{X}_t, \hat{Z}_t)dt + \hat{\theta}(\hat{X}_t, \hat{Z}_t)\sigma\left(dW_t + \lambda dt\right), \qquad (\text{III.45})$$

where we used the previous notation

$$\hat{Z}_t := z \vee \hat{X}_t^*, \qquad t \ge 0.$$

Lemma 1.3.3 Let Assumptions 1.3.1, 1.3.2 and 1.3.3 hold. Then the stochastic differential equation (III.45) has a unique strong solution (\hat{X}, \hat{Z}) for any initial condition $(x, z) \in \overline{\mathbf{D}}_{\alpha}$. Moreover the pair process

$$(C^*, \theta^*) := (\hat{C}, \hat{\theta}) (\hat{X}_t, \hat{Z}_t) \in \mathcal{A}_{\alpha}(x, z),$$

so that \hat{X}_t satisfies the drawdown constraint (III.19).

Proof. We first extend continuously \hat{C} and $\hat{\theta}$ to $\{(x, z) : x \leq z\}$ by setting them equal to zero, so that they remain Lipschitz, see Lemma 1.3.2. We shall denote by K > 0 a common Lipschitz constant. For a fixed z, we consider the map G defined on $\mathbb{R}_+ \times C^0(\mathbb{R}_+)$ by $G(t, \mathbf{x}) := \hat{C}(\mathbf{x}(t), z \vee \mathbf{x}^*(t))$. Since \hat{C} is Lipschitz, We directly estimate that

$$|G(t, \mathbf{x}) - G(t, \mathbf{y})| \leq K \{ |\mathbf{x}(t) - \mathbf{y}(t)| + |z \lor \mathbf{x}^*(t) - z \lor \mathbf{y}^*(t)| \} \leq 2K |\mathbf{x} - \mathbf{y}|_t^*,$$

for $t \ge 0$ and $\mathbf{x}, \mathbf{y} \in C^0(\mathbb{R}_+)$. This proves that G is a functional Lipschitz function in the sense of Protter [92]. By a similar calculation, we also show that the diffusion coefficient of the stochastic differential equation (III.45) is also functional Lipschitz. The existence and uniqueness of a strong solution to (III.45) follows from Theorem 7 p197 in [92]. Finally, the functions \hat{c} and $\hat{\pi}$ defined by

$$\hat{c}(x,z) := \frac{\hat{C}(x,z)}{x-\alpha z}$$
 and $\hat{\pi}(x,z) := \frac{\hat{\theta}(x,z)}{x-\alpha z}$, $(x,z) \in \mathbf{D}_{\alpha}$, (III.46)

are bounded since \hat{C} and $\hat{\theta}$ are Lipschitz functions satisfying furthermore, for any z > 0, $\hat{C}(\alpha z, z) = \hat{\theta}(\alpha z, z) = 0$. Therefore, the functions \hat{c} and $\hat{\pi}$ can be arbitrary extended to $\overline{\mathbf{D}}_{\alpha}$ so that the processes $\hat{c}(X_t, Z_t)$ and $\hat{\pi}(X_t, Z_t)$ are well defined and bounded for $(X_t, Z_t) \in \overline{\mathbf{D}}_{\alpha}$. Following the same argument as in Section 1.2.2, this implies in particular that $(C^*, \theta^*) \in \mathcal{A}_{\alpha}(x, z)$.

We are now ready for the statement of our main result.

Theorem 1.3.1 Let Assumptions 1.3.1, 1.3.2, and 1.3.3 hold. Then, $u^{\alpha} = U(0)/\beta$ on $\overline{\mathbf{D}}_{\alpha} \setminus \mathbf{D}_{\alpha}$ and

$$u^{\alpha}(x,z) = f(x,z) \left(\frac{\gamma+1}{\gamma} \left(x - \alpha z \right) + \frac{1}{\beta} \int_{f(x,z)}^{\infty} \frac{V(s)}{s^2} ds \right), \quad (x,z) \in \mathbf{D}_{\alpha}, \quad (\text{III.47})$$

and the consumption-investment strategy (C^*, θ^*) is an optimal solution of the problem (III.17). Moreover, u^{α} is a $C^0(\overline{\mathbf{D}}_{\alpha}) \cap C^{2,1}(\mathbf{D}_{\alpha})$ function, and the corresponding dual function v^{α} defined in (III.30) is given by

$$v^{\alpha}(y,z) = \begin{cases} y\left(-\alpha z + \frac{1}{\gamma}h(y,z) + \frac{1}{\beta}\int_{y}^{\infty}\frac{V(s)}{s^{2}}ds\right) & \text{for} \quad y \ge \varphi(z), \\ v^{\alpha}\left(\varphi(z),z\right) + z\left(\varphi(z) - y\right) & \text{for} \quad y \le \varphi(z). \end{cases}$$

The proof of this result is reported in Section 1.5, and relies on a verification argument which requires to guess the explicit form of the theorem. The construction of the candidate explicit solution is provided for completeness in Section 1.4.

1.3.5 The power utility case

In the absence of drawdown constraint, the value function associated to a power utility function and its Fenchel transform are well-known to be explicit. The main result of this section is that, under the drawdown constraint, the Fenchel transform of the value function associated to a power utility function is completely explicit, and the expressions of the optimal strategy and the value function are considerably simplified.

A power utility function is characterized by its asymptotic elasticity $p \in (0,1)$ and is given by

$$U_p(x) := \frac{x^p}{p}, \qquad x > 0,$$

Its Fenchel transform satisfies

$$V_p(y) = \frac{y^{-q}}{q}, \quad y > 0, \text{ with } \frac{1}{p} - \frac{1}{q} = 1.$$

We first recall briefly the solution of the Merton problem in the absence of the drawdown constraint. From section 1.3.2, under a convenient transversality condition, the Fenchel

transform v_p^0 of the value function u_p^0 is given by (III.34). One immediately checks that, under the so called Merton condition

$$\frac{\gamma}{1+\gamma} > p \,, \tag{III.48}$$

the Fenchel transform v_p^0 is given by

$$v_p^0(y) = \frac{(1-p)^3}{\beta p} \left(1 - \frac{1+\gamma}{\gamma}p\right)^{-1} y^{\frac{p}{p-1}} < \infty, \qquad y > 0$$

and the value function u_p^0 is obtained by direct calculation from (III.31),

$$u_p^0(x) = \left[\frac{\beta}{(1-p)^2} \left(1 - \frac{1+\gamma}{\gamma}p\right)\right]^{p-1} \frac{x^p}{p}, \quad x > 0.$$

The optimal consumption-investment strategy is identified as the maximizer in the dynamic programming equation (III.23), and given by $\hat{C}(x) = c_0^* x$ and $\hat{\theta}(x) = \pi_0^* x$, where

$$c_0^* := \frac{\beta}{(1-p)^2} \left(1 - \frac{1+\gamma}{\gamma} p \right), \qquad \pi_0^* := \frac{\lambda}{\sigma(1-p)}. \tag{III.49}$$

We now turn to the solution of the optimal consumption-investment problem under drawdown constraint. Let

$$b_{\alpha} := \frac{\beta}{(1-p)^2} \left(1 - \frac{1+\delta}{\delta} p \right) . \tag{III.50}$$

Observe that the optimal consumption rate in the Merton problem without drawdown constraint is $c_0^* = b_0$, since $\delta = \gamma$ whenever $\alpha = 0$. Notice also from (III.35) that Assumption 1.3.2 which rewrites

$$b_{\alpha} > 0$$
, i.e. $(1-\alpha)p < \frac{\gamma}{1+\gamma}$,

is weaker than the Merton condition (III.48), and reduces to it when $\alpha = 0$. Since the relative risk aversion of the power utility function U_p is a positive constant, Assumption 1.3.3 is always satisfied under Assumptions 1.3.1 and 1.3.2, see Remark 1.3.1.

The main observation for the particular case of a power utility function, is that the function φ , defined as the inverse of g given by (III.39) is fully explicit:

$$\varphi(z) = U'_p(b_{\alpha}z) = (b_{\alpha}z)^{p-1}, \quad z > 0.$$

Furthermore, the value function u_p^{α} inherits the homogeneity property from the power utility function U_p , so that

$$u_p^{\alpha}(x,z) = z^p u_p^{\alpha}\left(\frac{x}{z},1\right), \qquad (x,z) \in \mathbf{D}_{\alpha}.$$
(III.51)
Therefore, the function \hat{C} defined in (III.26) satisfies

$$\hat{C}(x,z) = -V_p'\left(z^{p-1}\nabla_x u_p^\alpha\left(\frac{x}{z},1\right)\right) = -z V_p'\left(\nabla_x u_p^\alpha\left(\frac{x}{z},1\right)\right) = z \hat{C}\left(\frac{x}{z},1\right),$$

for $(x, z) \in \mathbf{D}_{\alpha}$, where $\nabla_x u_p^{\alpha}$ denotes the derivative of u_p^{α} with respect to its first component. As a consequence, the function $(x, z) \mapsto -[V'_p \circ f](x, z)/(x - \alpha z)$ reduces to a function of the single variable x/z. Direct calculation reveals that this function is the inverse of the function F defined by

$$F(\xi) := \alpha + \frac{b_{\alpha}}{\xi} \left(\frac{1 - \frac{b_0}{\xi}}{1 - \frac{(1 - \alpha)b_0}{b_{\alpha}}} \right)^{\frac{\lambda^2}{2(1 - p)^2} b_0^{-1}}, \qquad (\text{III.52})$$

which is a C^1 function from $[b_0^+, b_{\alpha}/(1-\alpha)]$ to $[\alpha, 1]$. By passing to the limit in (III.52), we observe that

$$F(\xi) = \alpha + \frac{b_{\alpha}}{\xi} \exp\left[\frac{1}{\alpha\gamma}\left(1 - \alpha - \frac{b_{\alpha}}{\xi}\right)\right] \quad \text{whenever} \quad b_0 = 0. \quad \text{(III.53)}$$

Indeed, under Assumptions 1.3.1 and 1.3.2, F is strictly increasing so that its inverse F^{-1} is well defined and a strictly increasing continuous function from $[\alpha, 1]$ to $[b_0^+, b_{\alpha}/(1-\alpha)]$. The functions \hat{c} and $\hat{\pi}$ defined in (III.46) are now given by

$$\hat{c}_p(x,z) := F^{-1}\left(\frac{x}{z}\right)$$
 and $\hat{\pi}_p(x,z) := \frac{\lambda}{\sigma}(\gamma+1) - \frac{2}{\sigma\lambda}(1-p)F^{-1}\left(\frac{x}{z}\right)$,

for $(x, z) \in \overline{\mathbf{D}}_{\alpha}$. As in lemma 1.3.3, under Assumptions 1.3.1 and 1.3.2, the stochastic differential equation

$$d\hat{X}_t = \left(\hat{X}_t - \alpha \hat{Z}_t\right) \left[-\hat{c}_p\left(\hat{X}_t, \hat{Z}_t\right) dt + \hat{\pi}_p\left(\hat{X}_t, \hat{Z}_t\right) \sigma\left(dW_t + \lambda dt\right)\right],$$

has a unique strong solution (\hat{X}, \hat{Z}) for any initial condition $(x, z) \in \overline{\mathbf{D}}_{\alpha}$ and the pair process

$$(C_p^*, \theta_p^*) := (\hat{X} - \alpha \hat{Z}) \left(\hat{c}_p(\hat{X}, \hat{Z}), \hat{\pi}_p(\hat{X}, \hat{Z}) \right) \in \mathcal{A}_\alpha(x, z).$$

For completeness, we restate Theorem 1.3.1 in the context of a power utility function

Theorem 1.3.2 Let $U = U_p$, Assumptions 1.3.1 and 1.3.2 hold. Then $u_p^{\alpha} = 0$ on $\overline{\mathbf{D}}_{\alpha} \setminus \mathbf{D}_{\alpha}$ and

$$u_p^{\alpha}(x,z) := \left(\frac{\gamma+1}{\gamma} + \frac{(1-p)^2}{\beta p} F^{-1}\left(\frac{x}{z}\right)\right) \left[F^{-1}\left(\frac{x}{z}\right)\right]^{p-1} (x-\alpha z)^p, \quad (x,z) \in \mathbf{D}_{\alpha},$$

and the consumption-investment strategy (C_p^*, θ_p^*) is an optimal solution of the problem (III.17). Furthermore, u_p^{α} is a $C^0(\overline{\mathbf{D}}_{\alpha}) \cap C^{2,1}(\mathbf{D}_{\alpha})$ function, and the corresponding dual function v_p^{α} is given by

$$v_{p}^{\alpha}(y,z) = \begin{cases} -\alpha zy - \frac{\alpha(b_{\alpha}z)^{p}}{b_{\alpha}\left(\gamma - (1+\gamma)p\right)} \left(\frac{(b_{\alpha}z)^{p-1}}{y}\right)^{\gamma} + \frac{1-p}{p\,b_{0}}\,y^{-\frac{p}{1-p}} & \text{for } y \ge (b_{\alpha}z)^{p-1} \\ v^{\alpha}\left((b_{\alpha}z)^{p-1}, z\right) + z\left((b_{\alpha}z)^{p-1} - y\right) & \text{for } y \le (b_{\alpha}z)^{p-1} \end{cases}$$

The above solution agrees with the candidate solution derived by [93] in the case of possibly positive interest rates. Therefore, Theorem 1.3.2 confirms that the candidate solution derived by [93] is indeed the solution of the optimal consumption-investment problem.

1.3.6 Properties of the solution

In this subsection, we analyse the behavior of an agent maximizing its lifetime power utility of consumption under the drawdown constraint (III.19). The particular case of a power utility function enables us to compare our solution to the well-known benchmark Merton solution in the absence of drawdown constraint. Remark furthermore that, since the value functions u_p^{α} and the consumption-investment strategy (C_p, θ_p) inherit the homogeneity properties of U_p and V_p , all the evaluations and comparisons can be realized in terms of fraction of wealth x/z. The results presented here are similar to the ones observed by Roche [93] and are reported here for completeness.

Considering a particular set of parameters $\{p, \sigma, \lambda, \beta\} = \{0.2, 1, 3, 3\}$ satisfying the Merton condition (III.48), we report the value functions and optimal consumptioninvestment strategies associated to different values of α satisfying Assumption 1.3.1, i.e. between 0 and 0.6. Of course, the results observed when α reaches zero coincide with the benchmark Merton one. Because these three functions equal zero whenever the drawdown constraint binds, the reader can easily identify in each of the figures the slopes associated to the different values of α .

We first observe in Figure 1.1 that the amount of wealth invested in the risky asset decreases with α . Nevertheless, when the drawdown constraint nearly binds, the marginal investment strategy does not depend on α . But, as the fraction of wealth increases, the agent is more reluctant to investment in the risky asset as α increases. Finally, when the wealth process approaches its maximum, the amount invested in the risky asset even decreases for α high enough. Conversely, the consumption of the agent reported in Figure 1.2 is decreasing in α when the proportion of wealth is close to the drawdown constraint but increases with α whenever the wealth process approaches its current maximum.



Figure 1.1: Investment θ_p versus the fraction of wealth x/z for $\alpha=0$ to 0.6



Figure 1.2: Consumption C_p versus the fraction of wealth x/z for $\alpha = 0$ to 0.6

The key intuition behind those observations is the anticipation of the agent to the possibility that the drawdown constraint may be binding in the future. Therefore its aversion to risk increases and this explains why its investment and consumption strategy decrease with α . The particular behavior of the optimal strategy of the agent when its wealth approaches its current maximum relies in the ratcheting feature of the drawdown constraint. The agent anticipates that reaching its current maximum of wealth will increase the floor imposed by the drawdown constraint, and therefore chooses to consume instead of investing in the risky asset. When $\alpha = 1/(1 + \gamma) = 0.6$, corresponding to the highest possible value of α satisfying Assumption 1.3.1, the investor even never tries to reach its maximum, so that the value of the portfolio never exceeds the initial capital. Remark that, considering an agent maximizing the long term growth rate of expected utility of its final wealth, the optimal investment strategy derived by Grossman and Zhou [59] is conversely always linearly increasing with the fraction of wealth.

Finally Figure 1.3 shows the dependence of the value function u^{α} in terms of α . Since the set of possible consumption-investment strategies decreases with α , u_{α} is decreasing in α . This effect, due to the drawdown constraint, decreases with the proximity of the wealth to its current maximum.



Figure 1.3: Value function u_p^{α} versus the fraction of wealth x/z for $\alpha = 0$ to 0.6

1.4 Guessing a candidate solution for the dual function

In this section, we show with a formal argument how the dual function v^{α} can be guessed. We shall assume throughout that, for any z > 0,

$$u^{\alpha}(.,z)$$
 is a smooth increasing function. (III.54)

From the discussion of Section 1.3.1, the dynamic programming equation for the value function u^{α} is

$$\mathcal{L}u^{\alpha} := \beta u - V(u_x) + \frac{\lambda^2}{2} \frac{u_x^2}{u_{xx}} = 0, \qquad (x, z) \in \mathbf{D}_{\alpha}; \qquad (\text{III.55})$$

$$u^{\alpha}(\alpha z, z) = U(0)/\beta, \qquad z \ge 0;$$
 (III.56)

$$u_z^{\alpha}(z,z) = 0, \qquad z > 0.$$
 (III.57)

<u>Step 1:</u> The PDE satisfied by v^{α} . We first introduce the functions

$$\varphi(z) := u_x^{\alpha}(z, z) \text{ and } \psi(z) := u_x^{\alpha}(\alpha z, z), \quad z > 0.$$

For any z > 0, by the concavity property of $u^{\alpha}(., z)$, see Lemma 1.2.1, we deduce that $\varphi(z) \leq \psi(z)$. From the definition of the dual function v^{α} , we have

$$v^{\alpha}(y,z) = u^{\alpha}(\hat{x}(y,z),z) - \hat{x}(y,z)y \quad \text{if} \ u^{\alpha}_{x}(\hat{x}(y,z),z) = y \in [\varphi(z),\psi(z)] , \quad (\text{III.58})$$

$$v^{\alpha}(y,z) = u^{\alpha}(z,z) - yz \qquad \text{if } y \le \varphi(z), \qquad (\text{III.59})$$

$$v^{\alpha}(y,z) = U(0)/\beta - \alpha yz \qquad \text{if } y \ge \psi(z), \qquad (\text{III.60})$$

where the last equality follows from (III.56). Remark that, in the situation of (III.58) where $y \in [\varphi(z), \psi(z)]$, we obtain by a direct change of variable in (III.55) that

$$\mathcal{L}^* v^{\alpha}(y, z) = V(y) \quad \text{for} \quad \varphi(z) < y < \psi(z) \,, \tag{III.61}$$

where \mathcal{L}^* is the linear operator defined in (III.33). We also observe that the Neumann boundary condition (III.57) is converted into

$$v_z^{\alpha}(y,z) = \varphi(z) - y \quad \text{for} \quad y \le \varphi(z).$$
 (III.62)

<u>Step 2</u>: From the Neumann condition to a Dirichlet condition. Let introduce the function w^{α} defined by

$$w^{\alpha}(y,z) := v_{z}^{\alpha}(y,z) \quad \text{for} \quad \varphi(z) < y < \psi(z), \qquad (\text{III.63})$$

where z > 0. Since \mathcal{L}^* is a linear operator, it follows that w^{α} satisfies

$$\mathcal{L}^* w^{\alpha} = \beta w^{\alpha} - \beta y w^{\alpha}_y - \frac{\lambda^2}{2} y^2 w^{\alpha}_{yy} = 0, \qquad \varphi(z) < y < \psi(z).$$
(III.64)

Condition (III.60) and the Neumann condition (III.62) on v^{α} provide the following Dirichlet conditions on w^{α} ,

$$w^{\alpha}(\varphi(z), z) = 0 \text{ and } w^{\alpha}(\psi(z), z) = -\alpha\psi(z), \quad z > 0.$$
 (III.65)

For every fixed z > 0, the system (III.64)-(III.65) has a unique C^2 solution $w^{\alpha}(., z)$ given by

$$w^{\alpha}(y,z) = -\alpha y \left(1 - \left(\frac{\varphi(z)}{\psi(z)}\right)^{1+\gamma} \right)^{-1} \left(1 - \left(\frac{\varphi(z)}{y}\right)^{1+\gamma} \right), \quad \varphi(z) < y < \psi(z).$$
(III.66)

Step 3: Infinite marginal utility when the drawdown constraint nearly binds.

Since we will be using a verification argument, we just need to find a solution to the dynamic programming equation (III.55)-(III.56)-(III.57). We then seek for a candidate solution satisfying

$$\psi(z) = u_x^{\alpha}(\alpha z, z) = +\infty, \qquad z > 0.$$

From the economic viewpoint, this means that the marginal indirect utility is infinite when the wealth process approaches the drawdown constraint. This is understandable as the amounts of consumption and investment reduce to zero for the remaining lifetime whenever the drawdown constraint binds, i.e. $X_t = \alpha Z_t$, see Remark 1.2.1. So, any small departure from this constraint is very important for the investor as investment on the financial market and consumption are again possible. In this case, (III.66) reduces to

$$w^{\alpha}(y,z) = -\alpha y \left(1 - \left(\frac{\varphi(z)}{y}\right)^{1+\gamma} \right), \qquad \varphi(z) < y.$$
 (III.67)

<u>Step 4:</u> Derivation of a generic form for v_y^{α} . Integrating (III.67) with respect to z leads to

$$v^{\alpha}(y,z) = -\alpha y z + \alpha y \int_{z_0}^{z} \left(\frac{\varphi(s)}{y}\right)^{1+\gamma} ds + \phi(y), \qquad \varphi(z) < y,$$

where z_0 and $\phi(.)$ are still to be determined. Differentiating now with respect to y, we get

$$v_y^{\alpha}(y,z) = -\alpha z - \alpha \gamma \int_{z_0}^{z} \left(\frac{\varphi(s)}{y}\right)^{1+\gamma} ds + \phi'(y), \qquad \varphi(z) < y, \qquad \text{(III.68)}$$

with the two boundary conditions $v_y^{\alpha}(\varphi(z), z) = -z$ and $v_y^{\alpha}(\infty, z) = -\alpha z$ given respectively by (III.59) and (III.60). In order to determine ϕ' , we observe from (III.61), that ϕ satisfies an ordinary differential equation which provides, after differentiation with respect to y,

$$(\gamma + 2)\phi'''(y) + y\phi''(y) = -\frac{\gamma}{\beta}\frac{V'(y)}{y}, \quad \varphi(z) < y.$$

We deduce

$$\phi''(y) = -\frac{\gamma}{\beta y} \int_{y_0}^y \frac{V'(s)}{s} \left(\frac{s}{y}\right)^{1+\gamma} ds, \qquad \varphi(z) < y,$$

with y_0 a constant to be determined. Integrating with respect to y, we obtain the expression of ϕ' up to a constant which is fixed by the boundary condition $\phi'(\infty) = 0$ given by $v_y^{\alpha}(\infty, z) = -\alpha z$. Reporting this expression in (III.68), we finally get

$$v_y^{\alpha}(y,z) = -\alpha z - \alpha \gamma \int_{z_0}^{z} \left(\frac{\varphi(s)}{y}\right)^{1+\gamma} ds + \frac{\gamma}{\beta(1+\gamma)} \int_{y_0}^{y} \frac{V'(s)}{s} \left(\frac{s \wedge y}{y}\right)^{1+\gamma} ds \,, \quad \text{(III.69)}$$

for $\varphi(z) < y$, with the boundary condition $v_y^{\alpha}(\varphi(z), z) = -z$.

Step 5: Implicit obtention of the marginal utility $\varphi(z)$.

The function $\varphi(z)$ will be implicitly given by the boundary condition $v_y^{\alpha}(\varphi(z), z) = -\alpha z$. Rewriting the boundary condition according to (III.69) and differentiating with respect to z, we compute

$$\varphi'(z) v_{yy}^{\alpha}(\varphi(z), z) = -\frac{\gamma}{\delta}, \qquad z > 0.$$
 (III.70)

Assuming that φ is invertible and denoting g its inverse, we notice that (III.70) rewrites as an ordinary differential equation satisfied by g

$$(1+\delta)g(\zeta) + \zeta g'(\zeta) = \frac{\delta}{\beta} \int_{\zeta}^{\infty} \frac{-V'(s)}{s} ds \,, \qquad \zeta > 0 \,,$$

whose solution is explicitly given by

$$g(\zeta) = \frac{\delta}{\beta(1+\delta)} \left(\int_{\zeta_0}^{\zeta} \frac{-V'(s)}{s} \left(\frac{s}{\zeta}\right)^{1+\delta} ds + \int_{\zeta}^{\infty} \frac{-V'(s)}{s} ds \right), \, \zeta > 0, \quad \text{(III.71)}$$

with ζ_0 a constant to be determined. From (III.35), $\delta/(1+\delta) > 0$ and since we require g to be a positive function, ζ_0 must be 0 or ∞ depending on the sign of δ . Nevertheless, in both cases, direct computation shows that g' and then φ' are negative. Since we require the dual function v^{α} to be convex, equation (III.70) imposes $\delta > 0$ which corresponds to

assumption 1.3.1. Therefore $\zeta_0 = 0$ and g coincides with (III.39) which is well-defined under Assumption 1.3.2, see (III.38). Therefore the function $\varphi(z)$ is implicitly defined by the relation

$$z = \frac{\delta}{\beta(1+\delta)} \left(\int_0^{\varphi(z)} \frac{-V'(s)}{s} \left(\frac{s}{\zeta}\right)^{1+\delta} ds + \int_{\varphi(z)}^{\infty} \frac{-V'(s)}{s} ds \right), \qquad z > 0.$$
(III.72)

Step 6: Deducing the dual function v^{α} .

174

Now, combining (III.69), (III.72) and the boundary condition $v_y^{\alpha}(\varphi(z), z) = -z$, we compute

$$-\frac{\gamma}{\beta(\gamma+1)} \int_0^{\varphi(z)} \frac{V'(s)}{s} \left(\frac{s}{\varphi(z)}\right)^{1+\delta} ds = \alpha \gamma \int_{z_0}^z \left(\frac{\varphi(s)}{\varphi(z)}\right)^{1+\gamma} ds - \frac{\gamma}{\beta(\gamma+1)} \int_{y_0}^{\varphi(z)} \frac{V'(s)}{s} \left(\frac{s}{\varphi(z)}\right)^{1+\gamma} ds,$$

for z > 0, which reported in (III.69), leads to

$$\begin{aligned} v_y^{\alpha}(y,z) &= -\alpha z - \frac{\gamma}{\beta(1+\gamma)} \left(\frac{\varphi(z)}{y}\right)^{1+\gamma} \int_0^{\varphi(z)} \frac{-V'(s)}{s} \left(\frac{s}{\varphi(z)}\right)^{1+\delta} ds \\ &- \frac{\gamma}{\beta(\gamma+1)} \left(\int_{\varphi(z)}^y \frac{-V'(s)}{s} \left(\frac{s}{y}\right)^{1+\gamma} ds + \int_y^{\infty} \frac{-V'(s)}{s} ds\right), \qquad \varphi(z) < y. \end{aligned}$$

Starting from this expression of v_y^{α} , the ordinary differential equation (III.61) directly leads to the expression of v^{α} announced in Theorem 1.3.1. In order to deduce the value function u^{α} , we simply need, for any z > 0, to invert the function $v_y^{\alpha}(., z)$, which corresponds to inverting the function h(., z) defined in (III.42).

Remark 1.4.1 In the particular case of the power utility function, u_p^{α} inherits the homogeneity property of U_p so that $\varphi(z) = \varphi(1)z^{p-1}$. Therefore, we can skip step 5 and $\varphi(1)$ is explicitly determined by the boundary condition $v_y^{\alpha}(\varphi(1), 1) = -1$.

1.5 The verification argument

This section is devoted to the proof of Lemma 1.3.2 and Theorem 1.3.1.

1.5.1 A general version of the verification theorem

We recall the definition of the operator \mathcal{L} :

$$\mathcal{L}u = \beta u - \sup_{C \ge 0, \theta \in \mathbb{R}} \left\{ U(C) + \mathcal{L}^{C, \theta}u \right\} \text{ where } \mathcal{L}^{C, \theta}u := \frac{1}{2}\theta^2 \sigma^2 u_{xx} + (\theta \sigma \lambda - C)u_x.$$

We first derive a general verification theorem adapted to our maximization under drawdown constraint problem.

Theorem 1.5.1 Let ψ be a $C^0(\overline{\mathbf{D}}_{\alpha}) \cap C^{2,1}(\mathbf{D}_{\alpha})$ function.

- (i) If ψ satisfies $\mathcal{L}\psi \geq 0$ and $-\psi_z(z,z) \geq 0$, then $\psi \geq u^{\alpha}$.
- (ii) Assume in addition that
- (a) $\mathcal{L}\psi = 0$, $\psi(\alpha z, z) = U(0)/\beta$ and $-\psi_z(z, z) = 0$;
- (b) there exist K > 0 and $0 < p_0 < \delta/(1 + \delta)$ such that

$$\psi(x,z) \leq K \left(1 + z^{\alpha p_0} (x - \alpha z)^{(1-\alpha)p_0} \right), \quad (x,z) \in \overline{\mathbf{D}}_{\alpha};$$

(c) $\mathcal{L}\psi = \beta\psi - U(\tilde{C}) + \mathcal{L}^{\tilde{C},\tilde{\theta}}\psi$ where $\tilde{C}(x,z) = (x - \alpha z)\tilde{c}(x,z)$, $\tilde{\theta}(x,z) = (x - \alpha z)\tilde{\pi}(x,z)$, and the stochastic differential equation

$$d\tilde{X}_t = -\tilde{C}(\tilde{X}_t, \tilde{Z}_t)dt + \sigma \,\tilde{\theta}(\tilde{X}_t, \tilde{Z}_t) \,(dW_t + \lambda dt) \quad t \ge 0 \,,$$

has a unique strong solution (\tilde{X}, \tilde{Z}) for any initial condition $(\tilde{X}_0, \tilde{Z}_0) = (x, z) \in \overline{\mathbf{D}}_{\alpha}$ satisfying

$$\int_0^T \tilde{c}(\tilde{X}_t, \tilde{Z}_t) dt < \infty \ a.s. \quad and \quad ||\tilde{\pi}(\tilde{X}_., \tilde{Z}_.)||_{\infty} < \infty.$$

Then $\psi = u^{\alpha}$.

Proof. We first observe that $\mathcal{L}\psi \geq 0$ implies

$$\beta \psi \geq V(\psi_x) \geq U(0),$$
 (III.73)

since V is a decreasing function and $V(\infty) = U(0)$. For $(x, z) \in \overline{\mathbf{D}}_{\alpha} \setminus \mathbf{D}_{\alpha}$, we have $u^{\alpha}(x, z) = U(0)/\beta$, and therefore the statement of the theorem is trivial. From now on, we fix a pair $(x, z) \in \mathbf{D}_{\alpha}$.

(i) Let (C, θ) be an arbitrary admissible consumption-investment strategy in $\mathcal{A}_{\alpha}(x, z)$, and let $(X, Z) := (X^{x,C,\theta}, Z^{x,z,C,\theta})$ be the solution of (III.18) with initial condition $(X_0, Z_0) = (x, z)$. We define the sequence of stopping times

$$\tau_n := \inf \{t > 0 : X_t - \alpha Z_t < n^{-1} \}.$$

By Itô's formula, we obtain

$$e^{-\beta T \wedge \tau_n} \psi \left(X_{T \wedge \tau_n}, Z_{T \wedge \tau_n} \right) = \psi(x, z) + M_T + \int_0^{T \wedge \tau_n} e^{-\beta t} \psi_z(X_t, Z_t) dZ_t + \int_0^{T \wedge \tau_n} e^{-\beta t} \left[\mathcal{L}^{C_t, \theta_t} \psi - \beta \psi \right] (X_t, Z_t) dt ,$$

where

$$M_T := \int_0^{T \wedge \tau_n} e^{-\beta t} \theta_t \sigma \psi_x(X_t, Z_t) dW_t , \qquad T \ge 0.$$

Since $-\psi_z(z, z) \ge 0$, Z is an increasing process and $dZ_t = 0$ whenever $X_t < Z_t$, it follows that the integral term with respect to Z is non-negative. Using in addition the fact that $\mathcal{L}\psi \ge 0$, we get

$$\psi(x,z) \geq e^{-\beta T \wedge \tau_n} \psi\left(X_{T \wedge \tau_n}, Z_{T \wedge \tau_n}\right) + \int_0^{T \wedge \tau_n} e^{-\beta t} U(C_t) dt - M_T. \quad (\text{III.74})$$

Recall that ψ_x is continuous on \mathbf{D}_{α} . Then, it follows from the definition of τ_n that the stopped process $\psi_x(X, Z)$ is a.s. continuous on $[0, T \wedge \tau_n]$. Since $\int_0^T \theta_t^2 dt < \infty$, this implies that M is a local martingale. By the lower bound (III.73) on ψ , it follows from (III.74) that M is uniformly bounded from below. Then M is a supermartingale. Taking expected values in (III.74), and using again the lower bound (III.73) on ψ , this implies that

$$\psi(x,z) \geq \mathbb{E}\left[\int_0^{T\wedge\tau_n} e^{-\beta t} U(C_t) dt + \frac{U(0)}{\beta} e^{-\beta T\wedge\tau_n}\right]$$

By the monotone convergence theorem together with Remark III.21, this implies that

$$\psi(x,z) \geq \mathbb{E}\left[\int_0^{\tau_{\infty}} e^{-\beta t} U(C_t) dt + \frac{U(0)}{\beta} e^{-\beta \tau_{\infty}}\right] = \mathbb{E}\left[\int_0^{\infty} e^{-\beta t} U(C_t) dt\right],$$

which proves that $\psi(x, z) \ge u^{\alpha}(x, z)$ by the arbitrariness of $(C, \theta) \in \mathcal{A}_{\alpha}(x, z)$. (ii) For simplicity, we denote $(\tilde{C}_t, \tilde{\theta}_t, \tilde{c}_t, \tilde{\pi}_t) := (\tilde{C}, \tilde{\theta}, \tilde{c}, \tilde{\pi})(\tilde{X}_t, \tilde{Z}_t)$, for any $t \ge 0$. By the same argument as in (III.10), we have

$$(\tilde{X}_t - \alpha \tilde{Z}_t)\tilde{Z}_t^{\alpha/(1-\alpha)} = \exp\left\{-\int_0^t \sigma \tilde{\pi}_r dW_r - \int_0^t \left(\tilde{c}_r - \lambda \sigma \tilde{\pi}_r + \frac{(\sigma \tilde{\pi}_r)^2}{2}\right) dr\right\}.$$
 (III.75)

In particular, this implies that the sequence of stopping times

$$\tilde{\tau}_n := \inf \left\{ t > 0 : \tilde{X}_t - \alpha \tilde{Z}_t < n^{-1} \text{ or } \tilde{Z}_t > n \right\} \longrightarrow \infty, \text{ a.s.}$$

Since we have $\beta \psi - U(\tilde{C}) - \mathcal{L}^{\tilde{C},\tilde{\theta}} \psi = 0$, it follows from Itô's lemma that

$$\psi(x,z) = e^{-\beta T \wedge \tilde{\tau}_n} \psi\left(\tilde{X}_{T \wedge \tilde{\tau}_n}, \tilde{Z}_{T \wedge \tilde{\tau}_n}\right) + \int_0^{T \wedge \tilde{\tau}_n} e^{-\beta t} U(\tilde{C}_t) dt - \tilde{M}_T, \quad \text{(III.76)}$$

where

$$\tilde{M}_T := \int_0^{T \wedge \tilde{\tau}_n} e^{-\beta t} \sigma[\tilde{\theta}\psi_x](\tilde{X}_t, \tilde{Z}_t) dW_t, \qquad T \ge 0.$$

Since ψ_x is continuous on \mathbf{D}_{α} , and the stopped process (\tilde{X}, \tilde{Z}) takes values in a compact subset of \mathbf{D}_{α} , it follows that the process $\psi_x(\tilde{X}, \tilde{Z})$ is uniformly bounded on $[0, \tilde{\tau}_n]$. Using the boundedness of the process $\tilde{\pi}$, we deduce that \tilde{M} is a martingale, and

$$\psi(x,z) = \mathbb{E}\left[e^{-\beta T \wedge \tilde{\tau}_n} \psi\left(\tilde{X}_{T \wedge \tilde{\tau}_n}, \tilde{Z}_{T \wedge \tilde{\tau}_n}\right)\right] + \mathbb{E}\left[\int_0^{T \wedge \tilde{\tau}_n} e^{-\beta t} U(\tilde{C}_t) dt\right]. \quad (\text{III.77})$$

We introduce the notation $p_{\alpha} := (1 - \alpha)p_0$ where p_0 is defined in (ii-b) and recall from (III.35) that $p_{\alpha} < \gamma/(1+\gamma)$. From (III.75) together with condition (ii-b) of the theorem, we have

$$e^{-\beta t}\psi(\tilde{X}_t,\tilde{Z}_t) \le K\left(1+N_t\exp\left\{-\int_0^t\beta+p_\alpha\left(\tilde{c}_r-\lambda\sigma\tilde{\pi}_r+(1-p_\alpha)\frac{(\sigma\tilde{\pi}_r)^2}{2}\right)dr\right\}\right)\,,$$

for any t > 0, where N is the Doléans-Dade exponential of $\int_0^t \sigma p_\alpha \tilde{\pi}_s dW_s$. We next compute that

$$\eta_s := \beta + p_\alpha \left(\tilde{c}_s - \lambda \sigma \tilde{\pi}_s + (1 - p_\alpha) \frac{(\sigma \tilde{\pi}_s)^2}{2} \right)$$

$$\geq \frac{\lambda^2}{2} \left\{ \gamma + p_\alpha \left((1 - p_\alpha) \left(\frac{\sigma \tilde{\pi}_s}{\lambda} - \frac{1}{1 - p_\alpha} \right)^2 - \frac{1}{(1 - p_\alpha)} \right) \right\}$$

$$\geq \frac{\lambda^2}{2} \left\{ \gamma - \frac{p_\alpha}{1 - p_\alpha} \right\} =: \eta > 0,$$

since $p_{\alpha} < \gamma/(1+\gamma)$. Therefore, it follows that

$$\mathbb{E}\left[e^{-\beta T \wedge \tilde{\tau}_n} \psi\left(X_{T \wedge \tilde{\tau}_n}, Z_{T \wedge \tilde{\tau}_n}\right)\right] \leq K \mathbb{E}\left[e^{-\beta T \wedge \tilde{\tau}_n} + e^{-\eta T \wedge \tilde{\tau}_n} N_{T \wedge \tilde{\tau}_n}\right]. \quad (\text{III.78})$$

Furthermore, by the Cauchy-Schwarz inequality, $\mathbb{E}\left[e^{-\eta T \wedge \tilde{\tau}_n} N_{T \wedge \tilde{\tau}_n}\right]$ is bounded from above, for any $\varepsilon > 0$, by

$$\mathbb{E}\left[\exp\left\{(1+\varepsilon^{-1})\left(-\eta T\wedge\tilde{\tau}_n+\varepsilon\int_0^{T\wedge\tilde{\tau}_n}|\sigma p_\alpha\tilde{\pi}_s|^2ds\right)\right\}\right]^{\varepsilon/(1+\varepsilon)}\mathbb{E}\left[N_{T\wedge\tilde{\tau}_n}^\varepsilon\right]^{1/(1+\varepsilon)},$$

where N^{ε} is a martingale, the Doléans-Dade exponential of $\int_0^t (1+\varepsilon)p_{\alpha} \sigma \tilde{\pi}_s dW_s$. Since $\tilde{\pi}$ is uniformly bounded, by taking ε small enough, we finally deduce from (III.78) that

$$\mathbb{E}\left[e^{-\beta T\wedge\tilde{\tau}_n}\psi\left(\tilde{X}_{T\wedge\tilde{\tau}_n},\tilde{Z}_{T\wedge\tilde{\tau}_n}\right)\right] \leq K\left(\mathbb{E}\left[e^{-\beta T\wedge\tilde{\tau}_n}\right] + \mathbb{E}\left[e^{-\eta T\wedge\tilde{\tau}_n}\right]^{\varepsilon/(1+\varepsilon)}\right).$$

Therefore, sending respectively n and T to infinity in (III.77), the dominated and the monotone convergence theorem provide

$$\psi(x,z) = \mathbb{E}\left[\int_0^\infty e^{-\beta t} U(\tilde{C}_t) dt\right].$$

In view of (i), this implies that $\psi = u^{\alpha}$.

1.5.2 Proof of Theorem 1.3.1

We now turn to the proof of Theorem 1.3.1 by verifying that the explicit expression reported in there fulfills the conditions of the verification Theorem 1.5.1. One of these conditions will indeed require the proof of Lemma 1.3.2. We first need to establish additional properties of the function f.

Lemma 1.5.1 Let Assumptions 1.3.1 and 1.3.2 hold. Then $f \in C^1(\mathbf{D}_{\alpha})$ and we have

$$\frac{f_z(x,z)}{f(x,z)} = \alpha \left(\gamma \left(\frac{\varphi(z)}{f(x,z)} \right)^{\gamma+1} + 1 \right) \left((\gamma+1)(x-\alpha z) + \frac{\gamma}{\beta} \int_{f(x,z)}^{\infty} \frac{V'(s)}{s} ds \right)^{-1}, \quad \text{(III.79)}$$

for $(x,z) \in \mathbf{D}_{\alpha}$.

Proof. We recall from lemma 1.3.1 that, for any z > 0, f(., z) is a decreasing C^1 function on $(\alpha z, z]$ whose derivative is given by (III.43). Furthermore, by construction, we have

$$f[h(y,z),z] = y$$
, for $y \ge \varphi(z)$, and $h[f(x,z),z] = x$, for $(x,z) \in \mathbf{D}_{\alpha}$. (III.80)

Now, from the definition of h, see (III.42), $h \in C^{1,1}(\{(y,z), y \ge \varphi(z)\})$ and we have

$$0 \leq h_z(y,z) = \alpha \left(\gamma \left(\frac{\varphi(z)}{y} \right)^{\gamma+1} + 1 \right) \leq \alpha(1+\gamma), \qquad y \geq \varphi(z). \quad \text{(III.81)}$$

Therefore, h and f are increasing in z. Hence f is decreasing in x, increasing in z and $\varphi : z \mapsto f(z, z)$ is decreasing. In order to prove that $f \in C^1(D_\alpha)$, we shall prove that f is differentiable in each variable with continuously partial derivatives.

1. In this step, we show that $f \in C^0(\mathbf{D}_{\alpha})$, which implies that $f_x \in C^0(\mathbf{D}_{\alpha})$ by (III.43). We take $(x, z) \in \mathbf{D}_{\alpha}$ and study separately the cases where x < z and x = z.

• If x < z, for l' small enough, $(x, z + l') \in \mathbf{D}_{\alpha}$ and we deduce from (III.81) that

$$h(f(x, z + l'), z) - x = h(f(x, z + l'), z) - h(f(x, z + l'), z + l') \le \alpha(1 + \gamma) l' \xrightarrow[l' \to 0]{} 0.$$

Therefore, since $f(x, z + l') \ge \varphi(z)$ from the monotonicity of f, combining (III.80) and the continuity of f(., z), we obtain

$$f(x, z + l') - f(x, z) = f(h(f(x, z + l'), z), z) - f(x, z) \xrightarrow[l' \to 0]{} 0.$$
(III.82)

Moreover, we remark that, for ℓ small enough, $(x + l, z + l') \in \mathbf{D}_{\alpha}$ and we have

$$f(x+l,z+l') - f(x,z) = f_x(x_l,z+l') l + f(x,z+l') - f(x,z), \qquad \text{(III.83)}$$

for some $x_l \in [x, x + l]$. Now, since f is monotonic in both its variables, we deduce from (III.43) that f and f_x are bounded on any compact subset of \mathbf{D}_{α} containing (x, z). Therefore, combining (III.82) and (III.83), we deduce that f is continuous at point (x, z).

• If x = z, we have, for any l and l' satisfying $(z + l, z + l') \in \mathbf{D}_{\alpha}$,

$$f(z+l,z+l') = f_x(z_l,z+l')(l'-l) + \varphi(z+l')$$
, for some $z_l \in [z+l,z+l']$.

Therefore similar arguments as above combined with the continuity of φ lead to the continuity of f on \mathbf{D}_{α} .

2. We now prove that f is differentiable with respect to z with continuous partial derivatives. Take $(x, z) \in \mathbf{D}_{\alpha}$ and l' such that $(x, z + l') \in \mathbf{D}_{\alpha}$. Combining (III.80) with $f(x, z) \geq \varphi(z + l')$, we deduce

$$\frac{1}{l'} \{ f(x, z + l') - f(x, z) \} = \frac{1}{l'} \{ f(x, z + l') - f(h(f(x, z), z + l', z + l')) \}$$

= $f_x(x_{l'}, z + l') \frac{1}{l'} \{ h(f(x, z), z) - h(f(x, z), z + l') \},$

for some $x_{l'} \in [x, x + l']$. Since $f_x \in C^0(\mathbf{D}_{\alpha})$ and $h_z(f(x, z), .)$ is continuous, we obtain

$$\frac{1}{h'} \left\{ f(x,z+h') - f(x,z) \right\} \xrightarrow[h' \to 0]{} -f_x(x,z) h_z(f(x,z),z)$$

Finally, combining (III.43) and (III.81), simple computations lead to (III.79) and f_z inherits the continuity of f on \mathbf{D}_{α} .

We are now ready for the proof of Lemma 1.3.2 which states that the functions \hat{C} and $\hat{\theta}$ defined in (III.44) are Lipschitz on $\overline{\mathbf{D}}_{\alpha}$.

Proof of Lemma 1.3.2. Remark from lemma 1.5.1 that $\hat{\theta}$ and \hat{C} are in $C^1(\mathbf{D}_{\alpha})$. **1.**We first study $\hat{\theta}$ and, since f_x and V' are negative functions, we have

$$\hat{\theta}_x(x,z) = \frac{\lambda}{\sigma} \left(\gamma + 1 - \frac{\gamma}{\beta} \frac{f_x(x,z)}{f(x,z)} [V' \circ f](x,z) \right) \le \frac{\lambda}{\sigma} (\gamma + 1), \quad (x,z) \in \mathbf{D}_{\alpha}.$$
(III.84)

Notice that, combining the definition of f and (III.43), we get

$$\frac{\beta}{\gamma} \frac{f(x,z)}{f_x(x,z)} = \left(\frac{\varphi(z)}{f(x,z)}\right)^{1+\gamma} \int_0^{\varphi(z)} \frac{V'(s)}{s} \left(\frac{s}{\varphi(z)}\right)^{1+\delta} ds + \int_{\varphi(z)}^{f(x,z)} \frac{V'(s)}{s} \left(\frac{s}{f(x,z)}\right)^{1+\gamma} ds$$
$$\leq \int_0^{f(x,z)} \frac{V'(s)}{s} \left(\frac{s}{f(x,z)}\right)^{1+\delta} ds, \quad (x,z) \in \mathbf{D}_\alpha, \qquad \text{(III.85)}$$

since $\varphi(z) \leq f(x, z)$ and $\gamma \leq \delta$. Now, since V' is a negative increasing function, we deduce

$$\frac{f(x,z)}{f_x(x,z)[V'\circ f](x,z)} \ge \frac{\gamma}{\beta} \int_0^{f(x,z)} \frac{1}{s} \left(\frac{s}{f(x,z)}\right)^{1+\delta} ds = \frac{\gamma}{\beta(1+\delta)} > 0, \quad \text{(III.86)}$$

by Assumption 1.3.1. Combining this inequality with (III.84), we deduce that the function θ_x is bounded on \mathbf{D}_{α} . Similarly we compute that, for $(x, z) \in \mathbf{D}_{\alpha}$,

$$\hat{\theta}_z(x,z) = -\frac{\lambda}{\sigma} \left(\alpha(\gamma+1) + \frac{\gamma}{\beta} \frac{f_z(x,z)}{f(x,z)} [V' \circ f](x,z) \right) \ge -\frac{\lambda}{\sigma} \alpha(\gamma+1),$$

since f_z and -V' are positive functions. Combining (III.43) and (III.79), we compute

$$\frac{f(x,z)}{f_z(x,z)} = -\frac{1}{\alpha} \left(\gamma \left(\frac{\varphi(z)}{f(x,z)} \right)^{1+\gamma} + 1 \right)^{-1} \frac{f(x,z)}{f_x(x,z)} \ge -\frac{1}{\alpha(\gamma+1)} \frac{f(x,z)}{f_x(x,z)}, \quad \text{(III.87)}$$

for $(x, z) \in \mathbf{D}_{\alpha}$. We then deduce from (III.86) that $\hat{\theta}_z$ is bounded from above and that $\hat{\theta}$ is a Lipschitz function on \mathbf{D}_{α} . Since, for any z > 0, $\hat{\theta}(0+, z) = 0 = \hat{\theta}(0, z)$, the function $\hat{\theta}$ is in fact Lipschitz on $\overline{\mathbf{D}}_{\alpha}$.

2. We now study \hat{C} whose derivatives are given by

$$\hat{C}_x(x,z) = -f_x(x,z)[V'' \circ f](x,z) \ge 0$$
 and $\hat{C}_z(x,z) = -f_z(x,z)[V'' \circ f](x,z) \le 0$,

for $(x, z) \in \mathbf{D}_{\alpha}$. We deduce from (III.85) that

$$\hat{C}_{x}(x,z) \leq \frac{\beta}{\gamma} f(x,z) [V'' \circ f](x,z) \left(\int_{0}^{f(x,z)} \frac{-V'(s)}{s} \left(\frac{s}{f(x,z)} \right)^{1+\delta} \right)^{-1}, \quad \text{(III.88)}$$

for $(x, z) \in \mathbf{D}_{\alpha}$, so that \hat{C}_x is bounded according to Assumption 1.3.3. Combining (III.87) and (III.88), we obtain a lower bound on \hat{C}_z and therefore \hat{C} is a Lipschitz function on \mathbf{D}_{α} .

Before stating the proof of Theorem 1.3.1, we first isolate two particular properties of the candidate value function denoted \hat{u}^{α} and defined in Theorem 1.3.1 by

$$\hat{u}^{\alpha}(x,z) := f(x,z) \left(\frac{\gamma+1}{\gamma} \left(x - \alpha z \right) + \frac{1}{\beta} \int_{f(x,z)}^{\infty} \frac{V(s)}{s^2} ds \right), \quad (x,z) \in \mathbf{D}_{\alpha}, \quad (\text{III.89})$$

and $\hat{u}^{\alpha} = U(0)/\beta$ on $\overline{\mathbf{D}}_{\alpha} \setminus \mathbf{D}_{\alpha}$.

Lemma 1.5.2 Let Assumptions 1.3.1 and 1.3.2 hold. Then \hat{u}^{α} is a $C^0(\overline{\mathbf{D}}_{\alpha}) \cap C^{2,1}(\mathbf{D}_{\alpha})$ function satisfying

$$\hat{u}_x^{\alpha}(x,z) = f(x,z) \quad and \quad \hat{u}_z^{\alpha}(z,z) = 0, \qquad (x,z) \in \mathbf{D}_{\alpha}.$$
 (III.90)

Proof. Under Assumptions 1.3.1 and 1.3.2, $f \in C^1(\mathbf{D}_{\alpha})$, see lemma 1.5.1. Therefore $\hat{u}^{\alpha} \in C^1(\mathbf{D}_{\alpha})$ and by direct differentiation in (III.89), it follows from (III.43) that $\hat{u}_x^{\alpha} = f$. Then \hat{u}^{α} is a $C^{2,1}(\mathbf{D}_{\alpha})$ function and we compute from (III.79) that

$$\hat{u}_{z}^{\alpha}(x,z) = \alpha f(x,z) \left(\left(\frac{\varphi(z)}{f(x,z)} \right)^{\gamma+1} - 1 \right), \quad (x,z) \in \mathbf{D}_{\alpha}, \quad (\text{III.91})$$

which leads to (III.90).

We now prove that $\hat{u}^{\alpha} \in C^0(\overline{\mathbf{D}}_{\alpha})$. Since V' is a negative function, we derive from (III.43),

$$\frac{f_x(x,z)}{f(x,z)} \geq \frac{1}{(\gamma+1)(x-\alpha z)}, \qquad (x,z) \in \mathbf{D}_{\alpha}.$$

Integrating this inequality on the interval [x, z], we obtain, up to the composition with the exponential function,

$$f(x,z) \ge \varphi(z)[(1-\alpha)z]^{1/(1+\gamma)} (x-\alpha z)^{-1/(1+\gamma)}, \quad (x,z) \in \mathbf{D}_{\alpha}.$$
 (III.92)

Remark now that, combining (III.89) with the definition of f, we derive, by an integration by part argument,

$$\hat{u}^{\alpha}(x,z) = \frac{\delta}{\beta} \left(\frac{\varphi(z)}{f(x,z)} \right)^{\gamma} \int_{0}^{\varphi(z)} \frac{V(s)}{s} \left(\frac{s}{\varphi(z)} \right)^{\delta} ds
+ \frac{\gamma}{\beta} \int_{\varphi(z)}^{f(x,z)} \frac{V(s)}{s} \left(\frac{s}{f(x,z)} \right)^{\gamma} ds, \qquad (x,z) \in \mathbf{D}_{\alpha}. \quad (\text{III.93})$$

Since the function V is decreasing, it is bounded from below by $V(\infty) = U(0)$, which plugged in (III.93) leads to $\hat{u}^{\alpha} \geq U(0)/\beta$. Fix now $z_0 > 0$, $\epsilon > 0$ and C_0 a compact subset of \mathbb{R}_+ containing z_0 . Remark that the existence of a constant M such that $|V(y) - U(0)| \leq \beta \epsilon/2$ for $y \geq M$.

Now, since φ and V are continuous functions and therefore bounded on compact sets, we deduce from (III.93) the existence of a constant K > 0 satisfying

$$\hat{u}^{\alpha}(x,z) \leq \left(\frac{K}{f(x,z)}\right)^{\gamma} + \frac{U(0)}{\beta} + \frac{\epsilon}{2}, \quad (x,z) \in \mathbf{D}_{\alpha}, \ z \in \mathcal{C}_{0}.$$

Observe now from (III.92) that there exists $\eta > 0$ such that, for any $(x, z) \in \mathbf{D}_{\alpha}$ with $z \in \mathcal{C}_0$ and $|x - \alpha z| < \eta$, we have $f(x, z) > K(\epsilon/2)^{-1/\gamma}$ which leads to

$$\frac{U(0)}{\beta} \leq \hat{u}^{\alpha}(x,z) \leq \frac{U(0)}{\beta} + \epsilon.$$

Therefore $\hat{u}^{\alpha} \in C^0(\overline{\mathbf{D}}_{\alpha})$ and the proof is complete.

Lemma 1.5.3 Let Assumptions 1.3.1 and 1.3.2 hold. Then, there exists K > 0 such that

$$\hat{u}^{\alpha}(x,z) \leq K \left(1+z^{\alpha p}(x-\alpha z)^{(1-\alpha)p}\right), \quad (x,z)\in\overline{\mathbf{D}}_{\alpha}.$$

Proof. First remark that this property is straightforward for $(x, z) \in \overline{\mathbf{D}}_{\alpha} \setminus \mathbf{D}_{\alpha}$. According to lemma 1.5.2, we compute

$$\hat{u}_{x}^{\alpha}(x,z) = f(x,z) = \hat{u}^{\alpha}(x,z) \left(\frac{\gamma+1}{\gamma}(x-\alpha z) + \frac{1}{\beta} \int_{f(x,z)}^{\infty} \frac{V(s)}{s^{2}} ds\right)^{-1}, \quad \text{(III.94)}$$

for $(x, z) \in \mathbf{D}_{\alpha}$.

1. We first derive (III.91) for a power utility function U_p and denote \hat{u}_p^{α} the candidate value function. As detailed in section 1.3.5, f(x,z) rewrites as $(F^{-1}(x/z)(x-\alpha z))^{p-1}$ on \mathbf{D}_{α} so that (III.94) leads to

$$\nabla_x \hat{u}_p^{\alpha}(x,z) = \hat{u}_p^{\alpha}(x,z) \left(\frac{\gamma+1}{\gamma} \left(x-\alpha z\right) + \frac{(1-p)^2}{\beta p} F^{-1}\left(\frac{x}{z}\right) \left(x-\alpha z\right)\right)^{-1}$$

for $(x, z) \in \mathbf{D}_{\alpha}$, where $\nabla_x \hat{u}_p^{\alpha}$ denotes the partial derivative of \hat{u}_p^{α} with respect to x. Since F^{-1} is an increasing function and $F^{-1}(1) = b_{\alpha}/(1-\alpha)$ where b_{α} is defined in (III.50), simple computations combined with (III.35) lead to

$$\frac{\nabla_x \hat{u}_p^{\alpha}(x,z)}{\hat{u}_p^{\alpha}(x,z)} \geq \frac{(1-\alpha)p}{x-\alpha z} , \qquad (x,z) \in \mathbf{D}_{\alpha} .$$
(III.95)

Integrating this inequality on the interval [x, z], we obtain, up to the composition with the exponential function

$$\frac{\hat{u}_p^{\alpha}(z,z)}{\hat{u}_p^{\alpha}(x,z)} \ge \left(\frac{(1-\alpha)z}{x-\alpha z}\right)^{(1-\alpha)p}, \quad (x,z) \in \mathbf{D}_{\alpha}.$$
 (III.96)

Since \hat{u}_p^{α} inherits the homogeneity property of U_p , $u_p^{\alpha}(z, z) = u_p^{\alpha}(1, 1) z^p$, for any z > 0, and we deduce from (III.96) the existence of K > 0 such that

$$\hat{u}_p^{\alpha}(x,z) \leq K z^p (x-\alpha z)^{(1-\alpha)p}, \qquad (x,z) \in \mathbf{D}_{\alpha}.$$
(III.97)

2. We next consider the case where the utility function is given by $U_p^0 = K^0(1 + U_p)$ where K^0 is the constant defined in (III.37). Observe that U_p^0 satisfies the required Assumptions 1.3.2 and 1.3.3. Simple computations show that the corresponding marginal utilities f_p^0 and f_p associated to the candidate value function \hat{u}_0^{α} and \hat{u}_p^{α} are related by $f_p^0 = K^0 f_p$. Combining (III.89) and (III.97), we easily derive

$$\hat{u}_0^{\alpha}(x,z) = K^0(1+\hat{u}_p^{\alpha}(x,z)) \le K K^0(1+z^p(x-\alpha z)^{(1-\alpha)p}), \quad (x,z) \in \mathbf{D}_{\alpha}.$$
(III.98)

3. We finally consider the general case. We recall from (III.37) that $U \leq U_p^0$ so that their Fenchel transforms satisfy also $V \leq V_p^0$. In this step, we shall prove that $\hat{u}^{\alpha} \leq \hat{u}_0^{\alpha}$, which combined with (III.98) concludes the proof.

Set $V^{\epsilon} := V + \epsilon (V_p^0 - V)$, for $0 \le \epsilon \le 1$, and denote $(V^{\epsilon})'$, φ^{ϵ} , f^{ϵ} and $\hat{u}^{\alpha,\epsilon}$ the associated functions defined in section 1.3.4. Observe first that all these functions are differentiable in ϵ . We intend to prove that $\hat{u}^{\alpha,\epsilon}$ is an increasing function of ϵ on [0, 1], which implies the required result as $V^0 = V$ and $V^1 = V_p^0$.

For ease of notation, let Υ be the operator defined for $(V, f, \varphi) \in C^1(\mathbb{R}_+, \mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}_+$ by

$$\begin{split} \Upsilon[V, f, \varphi] &:= \frac{\delta}{\beta} \left(\frac{\varphi}{f}\right)^{1+\gamma} \int_0^{\varphi} \frac{V(s)}{s^2} \left(\frac{s}{\varphi}\right)^{1+\delta} ds \\ &+ \frac{\gamma}{\beta} \int_{\varphi}^{f} \frac{V(s)}{s^2} \left(\frac{s}{f}\right)^{1+\gamma} ds - \frac{1}{\beta} \int_f^{\infty} \frac{V(s)}{s^2} ds \end{split}$$

By an integration by parts argument on (III.41), the function φ^{ϵ} is implicitly defined, for any $\epsilon \in [0, 1]$, by

$$\Upsilon[V^{\epsilon}, \varphi^{\epsilon}, \varphi^{\epsilon}](z) = \frac{1+\gamma}{\gamma}(1-\alpha)z, \quad z > 0.$$

Denoting ∇_{ϵ} the differential operator with respect to ϵ , we deduce

$$(1+\delta)\frac{\nabla_{\epsilon}\varphi^{\epsilon}}{\varphi^{\epsilon}}\left(\Upsilon[V^{\epsilon},\varphi^{\epsilon},\varphi^{\epsilon}] - \frac{1}{\beta}\int_{\varphi^{\epsilon}}^{\infty}\frac{(V^{\epsilon})'(s)}{s}ds\right) = \Upsilon[\nabla_{\epsilon}V^{\epsilon},\varphi^{\epsilon},\varphi^{\epsilon}].$$
(III.99)

Similarly f^{ϵ} is defined, for $\epsilon \in [0, 1]$, by

$$\Upsilon[V^{\epsilon}, f^{\epsilon}, \varphi^{\epsilon}](x, z) = \frac{1+\gamma}{\gamma}(x-\alpha z), \quad (x, z) \in \mathbf{D}_{\alpha},$$

and differentiation with respect to ϵ combined with (III.99) leads to

$$(1+\gamma)\frac{\nabla_{\epsilon}f^{\epsilon}}{f^{\epsilon}}\left(\Upsilon[V^{\epsilon}, f^{\epsilon}, \varphi^{\epsilon}] + \frac{1}{\beta}\int_{f^{\epsilon}}^{\infty} \frac{(V^{\epsilon})'(s)}{s}ds\right) = \Upsilon[\nabla_{\epsilon}V^{\epsilon}, f^{\epsilon}, \varphi^{\epsilon}] \quad (\text{III.100})$$
$$- \frac{\delta-\gamma}{1+\delta}\Upsilon[\nabla_{\epsilon}V^{\epsilon}, \varphi^{\epsilon}, \varphi^{\epsilon}].$$

Combining the definition of f^{ϵ} and (III.89), we rewrite $\hat{u}^{\alpha,\epsilon}$ as

$$\hat{u}^{\alpha,\epsilon} = \left(\Upsilon[V^{\epsilon}, f^{\epsilon}, \varphi^{\epsilon}] + \frac{1}{\beta} \int_{f^{\epsilon}}^{\infty} \frac{V^{\epsilon}(s)}{s^2} ds\right) f^{\epsilon}, \qquad 0 \le \epsilon \le 1.$$

Differentiating this expression with respect to ϵ , we compute from (III.99) and (III.100) that

$$\begin{split} \frac{\nabla_{\epsilon}\hat{u}^{\alpha,\epsilon}}{f^{\epsilon}} &= \frac{1}{1+\gamma}\Upsilon[\nabla_{\epsilon}V^{\epsilon}, f^{\epsilon}, \varphi^{\epsilon}] - \frac{\delta-\gamma}{(1+\gamma)(1+\delta)}\Upsilon[\nabla_{\epsilon}V^{\epsilon}, \varphi^{\epsilon}, \varphi^{\epsilon}] + \frac{1}{\beta}\int_{f^{\epsilon}}^{\infty}\frac{\nabla_{\epsilon}V^{\epsilon}(s)}{s}ds \\ &= \frac{\delta}{\beta(1+\delta)}\left(\frac{\varphi^{\epsilon}}{f^{\epsilon}}\right)^{1+\gamma}\int_{0}^{\varphi^{\epsilon}}\frac{\nabla_{\epsilon}V^{\epsilon}(s)}{s^{2}}\left(\frac{s}{\varphi^{\epsilon}}\right)^{1+\delta}ds + \frac{1}{\beta}\int_{f^{\epsilon}}^{\infty}\frac{\nabla_{\epsilon}V^{\epsilon}(s)}{s}ds \\ &+ \frac{\gamma-\delta}{\beta(1+\gamma)(1+\delta)}\int_{\varphi^{\epsilon}}^{\infty}\frac{\nabla_{\epsilon}V^{\epsilon}(s)}{s}ds + \frac{\gamma}{\beta(1+\gamma)}\int_{\varphi^{\epsilon}}^{f^{\epsilon}}\frac{\nabla_{\epsilon}V^{\epsilon}(s)}{s^{2}}\left(\frac{s}{\varphi^{\epsilon}}\right)^{1+\gamma}ds \,, \end{split}$$

for any $\epsilon \in [0, 1]$. We now observe that all the above integrals are positive since we have $\nabla_{\epsilon} V^{\epsilon} = V_p^0 - V \ge 0$. Since $\gamma \le \delta$ and $f^{\epsilon} \ge 0$, this shows that $u^{\alpha, \epsilon}$ is non-decreasing in ϵ .

We are now ready for the

Proof of Theorem 1.3.1. We will simply check that the candidate value function \hat{u}^{α} defined in (III.89) satisfies the hypothesis of Theorem 1.5.1. First, from lemma 1.5.2, $\hat{u}^{\alpha} \in C^0(\overline{\mathbf{D}}_{\alpha}) \cap C^{2,1}(\mathbf{D}_{\alpha})$. Combining (III.43) and (III.90), we easily check that \hat{u}^{α} satisfies (ii-a) in Theorem 1.5.1. Remark also that condition (ii-b) in Theorem 1.5.1 is exactly given by lemma 1.5.3. By construction, the functions $(\hat{C}, \hat{\theta})$ defined in (III.44) satisfy (III.26) so that $\mathcal{L}\hat{u}^{\alpha} = \beta\hat{u}^{\alpha} - U(\hat{C}) + \mathcal{L}^{\hat{C},\hat{\theta}}\hat{u}^{\alpha}$. Now, Lemma 1.3.3 ensures existence and uniqueness of a solution (\hat{X}, \hat{Z}) to the SDE (III.45) for any initial condition $(x, z) \in \overline{\mathbf{D}}_{\alpha}$, and, since \hat{c} and $\hat{\pi}$ defined in (III.46) are bounded functions, \hat{u}^{α} satisfies (ii-c) in Theorem 1.5.1. Therefore $\hat{u}^{\alpha} = u^{\alpha}$ and simple computations lead to the expression of the dual function of v^{α} .

Chapter 2

PDE characterization in finite time horizon

2.1 Introduction

We derived in the previous chapter the explicit solution of the optimal consumptioninvestment problem in infinite time horizon under a drawdown constraint. Instead of considering a manager handling the portfolio of investors, who may decide to recover their funding at any time, we now discuss the case where he is in charge of the portfolio over a fixed period T. We therefore study the problem of managing a portfolio subject to a drawdown constraint, with the purpose of maximizing the intertemporal utility of consumption on a finite horizon T. We seek for a better comprehension of the influence of this fixed time horizon on the behavior of the manager. In particular, we are interested in the influence of the choice of the utility function on the convergence of this optimal strategy in finite horizon T to the one obtained in the previous chapter, when T goes to infinity.

In the absence of drawdown constraint, Merton [79, 80] derived explicit solutions to this problem for particular choices of utility functions, by solving the corresponding Hamilton-Jacobi-Bellman equations. By a duality argument, Cox and Huang [29] and Karatzas, Lehoczky and Shreeve [64] extend his results to a market with non Markovian price processes. Beyond the large number of articles considering the addition of imperfections to the market, we mention the work of El Karoui, Jeanblanc and Lacoste [45], who consider a related type of constraints on the strategy. They study the behavior of a manager maximizing its finite horizon utility of wealth under the constraint that the value of the portfolio stays above a fixed floor process. Allowing the fund manager to invest in American Puts, they derive an optimal strategy. We refer also to the work of El Karoui and Meziou [46] who consider a similar minimum floor constraint, but present a very different point of view. Instead of specifying the utility function of the manager, their optimisation relies on a stochastic dominance approach, for which they prove the existence of an optimal solution.

In contrast with the infinite horizon, no explicit form of the value function is available, since the additional dependence in time of the solution makes the previous computations untractable. The purpose of this chapter is to derive a PDE characterization of the value function associated to the finite time horizon maximization. The derivation of the associated PDE relies classically on the use of the dynamic programming principle. The boundary conditions of the PDE are given by a Dirichlet condition at maturity Tand a Neumann condition when the process reaches its current maximum. Surprisingly, we do not require any Dirichlet condition on the semi real line where the drawdown constraint binds. Nevertheless, adding this Dirichlet condition allows to derive uniqueness of solution to the associated PDE in the viscosity sense under weaker assumptions. We first prove that the value function is a (discontinuous) viscosity solution of the corresponding Hamilton-Jacobi-Bellman equation. We then derive a comparison theorem for the associated PDE, which ensures the uniqueness of the solution within a particular class of functions. Since the consumption and investment controls are not bounded, the comparison result can not be obtained using classical penalization arguments. We overcame this difficulty by adapting the arguments of Zariphopoulou [103] where she studied a consumption-investment problem under general constraints. The comparison result then opens the door to the implementation of a numerical scheme, whose convergence is ensured by its stability and consistency, see Barles and Souganidis [7].

This chapter is organized as follows. The problem is formulated in Section 2.2. The main results detailing properties of the value function and its characterization as the unique viscosity solution of the associated PDE are presented in Section 2.3. A corresponding consistent numerical scheme and numerical results are provided in Section 2.4. The proofs of the viscosity property of the value function and the comparison result are respectively reported in Sections 2.5 and 2.6.

2.2 Problem formulation

We work in the same framework as in Chapter 1, that we recall briefly for convenience of the reader. The only difference lies on the finite horizon objective of the representative agent. We consider a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \le t \le T}, \mathbb{P})$ endowed

186

with a Brownian motion $W = \{W_t, 0 \le t \le T\}$ with values in \mathbb{R} , and we denote by $\mathbb{F} := \{\mathcal{F}_t, 0 \le t \le T\}$. The financial market consists of a non-risky asset, with process normalized to unity, and a risky asset with price process defined by the Black and Scholes model

$$dS_t = \sigma S_t \left(dW_t + \lambda dt \right) \,,$$

where $\sigma > 0$ is the volatility parameter, and $\lambda \in \mathbb{R}$ is a constant risk premium. For any continuous process $\{M_t, t \ge 0\}$, its current maximum is denoted M^* .

2.2.1 Consumption-portfolio strategies and the drawdown constraint

We next introduce the set of consumption-investment strategies whose induced wealth process X satisfies the drawdown constraint

$$X_t \ge \alpha X_t^*$$
 for every $0 \le t \le T$, a.s., (III.2.1)

where α is some given parameter in the interval [0, 1).

A consumption-investment strategy is an \mathbb{F} -adapted pair process $(C, \theta)_{0 \leq t \leq T}$ valued in $\mathbb{R}_+ \times \mathbb{R}$ satisfying the integrability condition

$$\int_{0}^{T} C_{s} \, ds + \int_{0}^{T} |\theta_{s}|^{2} \, ds < \infty \quad \text{a.s.} \quad . \tag{III.2.2}$$

The wealth process induced by such a pair (C, θ) is therefore defined by

$$X_t^{x,C,\theta} = x - \int_0^t C_r dr + \int_0^t \sigma \theta_r \left(dW_r + \lambda dr \right), \quad 0 \le t \le T, \quad (\text{III.2.3})$$

where x is some given initial capital. We still denote by $\mathcal{A}_{\alpha}(x)$ the collection of all such consumption-investment strategies whose corresponding wealth process satisfies the drawdown constraint (III.2.1). As in Remark 1.2.1 of Chapter 1, for a given initial wealth x and an admissible consumption-investment strategy $(C, \theta) \in \mathcal{A}_{\alpha}(x)$, we have

$$X^{x,C,\theta}_{,\nu\tau} = X^{x,C,\theta}_{\tau}, \quad \text{where } \tau := \inf \left\{ s \le T : X^{x,C,\theta}_s = \alpha \{X^{x,C,\theta}\}^*_s \right\}.$$
(III.2.4)

As in the infinite time horizon context, the set of admissible strategies consumptioninvestment strategies contains in particular the strategies of the form

$$C_t = c_t \left[X_t - \alpha X_t^* \right] \quad \text{and} \quad \theta_t = \pi_t \left[X_t - \alpha X_t^* \right], \tag{III.2.5}$$

where (c, π) is an \mathbb{F} -adapted pair process valued in $\mathbb{R}_+ \times \mathbb{R}$ satisfying the integrability condition

$$\int_{0}^{T} c_{s} ds + \int_{0}^{T} |\pi_{s}|^{2} ds < \infty.$$
 (III.2.6)

2.2.2 The finite horizon consumption-investment problem

Throughout this chapter, we consider a utility function

$$U: \mathbb{R}_+ \to \mathbb{R}$$
 C^2 , concave, satisfying $U'(0+) = \infty$ and $U'(\infty) = 0$. (III.2.7)

In addition to these properties reported from the previous chapter, we suppose without loss of generality that U(0) = 0.

For a given initial capital x > 0, the optimal finite-time horizon consumption-investment problem under drawdown constraint is defined by :

$$u_0 := \sup_{(C,\theta)\in\mathcal{A}_{\alpha}(x)} J_0(C,\theta) \quad \text{where} \quad J_0(C,\theta) := \mathbb{E}\left[\int_0^T e^{-\beta s} U(C_s) \, ds\right]. \quad (\text{III.2.8})$$

In order to make use of the the dynamic programming approach, we then need to introduce the dynamic version of this problem :

$$u(t,x,z) := \sup_{(C,\theta)\in\mathcal{A}_{\alpha}(t,x,z)} J(t,C,\theta) \text{ where } J(t,C,\theta) := \mathbb{E}\left[\int_{t}^{T} e^{-\beta s} U(C_{s}) ds\right],$$
(III.2.9)

the pair (x, z), with $x \leq z$, stands for the initial condition of the state processes (X, Z) defined, for $s \geq t$, by

$$Z_s^{t,x,z,C,\theta} := z \vee \left\{ X^{t,x,C,\theta} \right\}_s^* \quad \text{and} \quad X_s^{t,x,C,\theta} = x - \int_t^s C_r dr + \int_t^s \sigma \theta_r \left(dW_r + \lambda dr \right) ,$$
(III.2.10)

and $\mathcal{A}_{\alpha}(t, x, z)$ is the collection of all \mathbb{F} -adapted processes $(C_s, \theta_s)_{t \leq s \leq T}$ satisfying

$$\int_{t}^{T} C_{s} ds + \int_{t}^{T} |\theta_{s}|^{2} ds < \infty \quad \text{a.s.} \quad . \tag{III.2.11}$$

together with the drawdown constraint

$$X_s^{t,x,C,\theta} \ge \alpha Z_s^{t,x,z,C,\theta} \quad \text{a.s.}, \qquad t \le s \le T.$$
 (III.2.12)

We therefore define the value function u for any triplets (t, x, z) in the closure $\overline{\mathcal{O}}_{\alpha}$ in $[0, T] \times \mathbb{R}_+ \times \mathbb{R}_+$ of the domain

$$\mathcal{O}_{\alpha} := [0,T) \times \{(x,z) : 0 < \alpha z < x < z\}.$$

For any $y = (t, x, z) \in \overline{\mathcal{O}}_{\alpha}$ and $(C, \theta) \in \mathcal{A}_{\alpha}(y)$, we shall make use of the following notation

$$Y^{y,C,\theta}_s \ := \ (s,X^{t,x,C,\theta}_s,Z^{t,x,z,C,\theta}_s) \qquad \text{for any } s \geq t \,.$$

Remark 2.2.1 We remark first that the value function in infinite time horizon u^{α} studied in Chapter 1 provides obviously the following upper-bound

$$u(t, x, z) \le u_{\alpha}(x, z), \qquad (x, z) \in \overline{\mathcal{O}}_{\alpha}.$$
 (III.2.13)

Remark 2.2.2 Since we aim at interpreting u as a (discontinuous) viscosity solution of a PDE, one may wonder the necessity of the regularity assumptions on the utility function U adopted in (III.2.7). These assumptions are necessary to apply the results of Chapter 1 and derive the regular upper bound u_{α} to the value function u. As detailed in Lemma 2.3.3, U(0) = 0 allows the value function u to inherit continuity properties of u_{α} when the drawdown constraint nearly binds. These regularity properties are required for the proof of the general comparison result leading to Theorem 2.6.1. Nevertheless another version of the comparison result is obtained under weaker assumptions in Proposition 2.3.1 and discussed in Remark 2.6.1.

2.3 The main results

We keep similar notations as in Chapter 1, the function V still denotes Fenchel-Lengendre transform of U and we have

$$\gamma \ := \ \frac{2\beta}{\lambda^2} \,, \quad \delta \ := \ \frac{\gamma}{1 - \alpha(1 + \gamma)} \quad \text{ and } \quad p \ := \ \mathrm{AE}(U) \ = \ \lim_{x \to \infty} \frac{xU'(x)}{U(x)} \,.$$

We shall work under the following Assumptions.

Assumption 2.3.1
$$\frac{\gamma}{1+\gamma} < 1-\alpha.$$

Assumption 2.3.2 $p < \frac{\gamma}{1+\gamma}$.

Assumption 2.3.3 inf
$$\left\{\frac{1}{yV''(y)}\int_0^y \frac{-V'(s)}{s}\left(\frac{s}{y}\right)^{1+\delta}ds\right\} > 0.$$

Observe that Assumption 2.3.2 is the classical Merton condition and is is stronger than the corresponding Assumption 1.3.2 of Chapter 1. This stronger assumption is only needed for the proof of the comparison result in Theorem 2.6.1.

2.3.1 The PDE characterization

The dynamic programming equation is related to the second order operator defined for $\varphi \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R})$ by

$$\mathcal{L}_T \varphi := \sup_{C \ge 0, \theta \in \mathbb{R}} \mathcal{L}_T^{C, \theta} \varphi, \qquad (\text{III.2.14})$$

where, for any $C \geq 0$ and $\theta \in \mathbb{R}$, $\mathcal{L}_T^{C,\theta} \varphi$ is given by

$$\mathcal{L}_T^{C,\theta}\varphi := -\beta\varphi + \varphi_t + U(C) + (\sigma\lambda\theta - C)\varphi_x + \frac{(\sigma\theta)^2}{2}\varphi_{xx}.$$

Observe that the above dynamic programming equation simplifies to

$$\mathcal{L}_T \varphi = -\beta \varphi + \varphi_t + V(\varphi_x) - \frac{\lambda^2}{2} \frac{\varphi_x^2}{\varphi_{xx}} \quad \text{whenever } \varphi \text{ is strictly concave in } x. \quad (\text{III.2.15})$$

We next decompose the boundary of the domain of definition $\overline{\mathcal{O}}_{\alpha}$ of the value function u in the following four disjoint subsets

$$\begin{array}{rcl} \partial^{0}\mathcal{O}_{\alpha} & := & [0,T] \times \{(0,0)\} \,, \\ \\ \partial^{\alpha}\mathcal{O}_{\alpha} & := & [0,T] \times \{(\alpha z,z) \, : \, z > 0\} \,, \\ \\ \partial^{1}\mathcal{O}_{\alpha} & := & [0,T) \times \{(z,z) \, : \, z > 0\} \,, \\ \\ \partial^{T}\mathcal{O}_{\alpha} & := & \{T\} \times \{(x,z) \, : \, 0 < \alpha z \leq x \leq z\} \end{array}$$

The purpose of this chapter is to characterize u as the solution of the following dynamic programming equation

$$\begin{cases} -\mathcal{L}_T \varphi = 0 \quad \text{on} \quad \mathcal{O}_\alpha \cup \partial^\alpha \mathcal{O}_\alpha ,\\ -\varphi_z = 0 \quad \text{on} \quad \partial^1 \mathcal{O}_\alpha ,\\ \varphi = 0 \quad \text{on} \quad \partial^T \mathcal{O}_\alpha \cup \partial^0 \mathcal{O}_\alpha . \end{cases}$$
(III.2.16)

We now introduce the following classical notations. For any locally bounded function $v : \overline{\mathcal{O}}_{\alpha} \to \mathbb{R}$, we denote the corresponding lower and upper semi-continuous enveloppes of v by

$$v_*(y) := \liminf_{\mathcal{O}_{\alpha} \ni y' \to y} v(y') \text{ and } v^*(y) := \limsup_{\mathcal{O}_{\alpha} \ni y' \to y} v(y').$$

A viscosity solution of the PDE (III.2.16) is then defined in the following way.

Definition 2.3.1 (i) A locally bounded function v is a (discontinuous) viscosity subsolution of (III.2.16) if $v^* \leq 0$ on $\partial^T \mathcal{O}_{\alpha} \cup \partial^0 \mathcal{O}_{\alpha}$ and, for all $y_0 \in \overline{\mathcal{O}}_{\alpha}$ and $\varphi \in C^{1,2,1}(\overline{\mathcal{O}}_{\alpha})$ such that $0 = (v^* - \varphi)(y_0) = \sup_{\overline{\mathcal{O}}_{\alpha}} (v^* - \varphi)$, we have

$$-\mathcal{L}_T\varphi(y_0) \le 0 \quad if \ y_0 \in \mathcal{O}_\alpha \cup \partial^\alpha \mathcal{O}_\alpha \quad and \quad \min\{-\mathcal{L}_T\varphi, -\varphi_z\}(y_0) \le 0 \quad if \ y_0 \in \partial^1 \mathcal{O}_\alpha \,.$$

(ii) A locally bounded function v is a (discontinuous) viscosity supersolution of (III.2.16) if $v_* \geq 0$ on $\partial^T \mathcal{O}_{\alpha} \cup \partial^0 \mathcal{O}_{\alpha}$ and, for all given $y_0 \in \overline{\mathcal{O}}_{\alpha}$ and $\varphi \in C^{1,2,1}(\overline{\mathcal{O}}_{\alpha})$ such that $0 = (v_* - \varphi)(y_0) = \inf_{\overline{\mathcal{O}}_{\alpha}} (v_* - \varphi)$, we have

$$-\mathcal{L}_T \varphi(y_0) \ge 0 \quad if \ y_0 \in \mathcal{O}_\alpha \quad and \quad -\varphi_z(y_0) \ge 0 \quad if \ y_0 \in \partial^1 \mathcal{O}_\alpha \,.$$

(iii) A locally bounded function v is a (discontinuous) viscosity solution of (III.2.16) if it is both a sub- and a supersolution.

We now provide the main result of this chapter

Theorem 2.3.1 The value function u is a viscosity solution of (III.2.16). If furthermore, Assumptions 2.3.1, 2.3.2 and 2.3.3 hold, then u is the unique viscosity solution of (III.2.16) in the class of locally bounded functions v, right-continuous in the direction $\overrightarrow{e} := (0,1,1)$ on $\mathcal{O}_{\alpha} \cup \partial^{1}\mathcal{O}_{\alpha} \cup \partial^{\alpha}\mathcal{O}_{\alpha}$, equal to 0 on $\partial^{T}\mathcal{O}_{\alpha} \cup \partial^{0}\mathcal{O}_{\alpha}$ and satisfying the growth property

 $v(t, x, z) \leq K(1 + x^p), \quad (t, x, z) \in \overline{\mathcal{O}}_{\alpha}, \quad for \ some \ K > 0.$ (III.2.17)

The proof of the first part of the theorem is reported in Section 2.5. We provide some properties of the value function u in section 2.3.2, including in particular the nullity of u on $\partial^T \mathcal{O}_{\alpha} \cup \partial^0 \mathcal{O}_{\alpha}$, the growth property (III.2.17), as well as the right-continuity of u in the direction \overrightarrow{e} under Assumptions 2.3.1, 2.3.2 and 2.3.3. Finally a comparison result, ensuring uniqueness of viscosity solutions to the PDE (III.2.16) within the class of locally bounded functions, satisfying these particular growing and regularity properties, is presented in Section 2.6.

We conclude this section by stating a weaker comparison result for the solution of the PDE (III.2.16) obtained under weaker assumptions on the utility function U. Indeed, as announced in Remark 2.2.2, the imposed regularity on U allows to use the explicit solution in infinite horizon derived in Chapter 1 as a regular upper bound to the value function u_{α} , leading to the right-continuous in the direction \overrightarrow{e} of u on the boundary $\partial^{\alpha} \mathcal{O}_{\alpha}$. Nevertheless, this regularity property is not needed for the obtention of a comparaison result as long as we consider a smaller class of functions forced to equal zero on the boundary $\partial^{\alpha} \mathcal{O}_{\alpha}$. The justification of this argument is provided in Remark 2.6.1. Remark that the particular interest this second comparaison result relies on its consequences on the choice of a consistant numerical scheme as discussed in section 2.4.

Proposition 2.3.1 Let U be a C^1 , increasing, concave function satisfying U(0) = 0 as well as Assumption 2.3.2, and u be its associated value function. Then u is the unique viscosity solution of (III.2.16) in the class of locally bounded functions v, right-continuous in the direction $\vec{e} := (0, 1, 1)$ on $\mathcal{O}_{\alpha} \cup \partial^1 \mathcal{O}_{\alpha}$, equal to 0 on $\partial^T \mathcal{O}_{\alpha} \cup \partial^0 \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha}$, and satisfying the growth property (III.2.17).

2.3.2 Properties of the value function

This section collects some properties of the value function u which, in addition to their self interest, will allow us to derive precise viscosity properties of u on the boundary of the domain $\overline{\mathcal{O}}_{\alpha}$ and to restrain the class of functions for which a comparison result is required.

Lemma 2.3.1 The value function u satisfies

$$u \ge 0 \quad on \quad \overline{\mathcal{O}}_{\alpha} \quad and \quad u = 0 \quad on \quad \partial^T \mathcal{O}_{\alpha} \cup \partial^0 \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha} \,.$$
 (III.2.18)

If Assumption 2.3.2 holds, then there exists K > 0 such that

$$u(y) \le K(1+x^p), \quad y = (t, x, z) \in \overline{\mathcal{O}}_{\alpha}.$$
 (III.2.19)

Proof. Observe first that u inherits the positivity of U. Recalling (III.2.4), we remark that there is no non-trivial admissible strategy on $\partial^T \mathcal{O}_{\alpha} \cup \partial^0 \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha}$ and derive (III.2.18). Under Assumption 2.3.2, the asymptotic elasticity p of U is strictly smaller than one. We then deduce from Lemma 6.5 in [70] the existence of K > 0 such that

$$U(x) \leq K\left(1+\frac{x^p}{p}\right), \quad x \geq 0.$$
 (III.2.20)

But, in the absence of drawdown constraint, the value function u^* associated to the power utility function $x \mapsto x^p/p$ is well known to satisfy

$$u^*(t,x) \leq K'x^p, \quad t \geq 0, \quad x \geq 0,$$
 (III.2.21)

where K' is also a positive constant. Since the set of admissible strategies in the presence of drawdown constraint is smaller that the one of the classical Merton set-up, we deduce (III.2.19) from (III.2.20) and (III.2.21).

Lemma 2.3.2 The value function u is non-decreasing in its second variable x and non-increasing in its third variable z.

Proof. Take (t, x, z, z') such that $(t, x, z) \in \overline{\mathcal{O}}_{\alpha}$, $(t, x, z') \in \overline{\mathcal{O}}_{\alpha}$ and $z' \leq z$. Since

$$X^{t,x,C,\theta} \ge \alpha Z^{t,x,z,C,\theta} \ge \alpha Z^{t,x,z',C,\theta}, \quad (C,\theta) \in \mathcal{A}_{\alpha}(t,x,z),$$

we have $\mathcal{A}_{\alpha}(t, x, z) \subset \mathcal{A}_{\alpha}(t, x, z')$, which naturally leads to $u(t, x, z) \leq u(t, x, z')$. Similar arguments easily lead to the non-decreasing property of u in x.

We now derive some regularity and concavity properties of the value function u in the direction $\overrightarrow{e} = (0, 1, 1)$.

Lemma 2.3.3 The following holds.

- (i) For any $y \in \overline{\mathcal{O}}_{\alpha}$, the function $h \mapsto u[y + h \overrightarrow{e}]$ is concave on \mathbb{R}_+ .
- (ii) The function u is right-continuous in the direction \overrightarrow{e} on $\mathcal{O}_{\alpha} \cup \partial^T \mathcal{O}_{\alpha} \cup \partial^1 \mathcal{O}_{\alpha}$, i.e.

$$u[y + h\overrightarrow{e}] \underset{h \downarrow 0^+}{\longrightarrow} u[y], \quad \text{for any } y \in \mathcal{O}_{\alpha} \cup \partial^T \mathcal{O}_{\alpha} \cup \partial^1 \mathcal{O}_{\alpha}.$$

(iii) If Assumption 2.3.2 holds, then the function u is right-continuous in the direction \overrightarrow{e} on $\mathcal{O}_{\alpha} \cup \partial^{T} \mathcal{O}_{\alpha} \cup \partial^{1} \mathcal{O}_{\alpha} \cup \partial^{0} \mathcal{O}_{\alpha}$.

(iv) If furthermore Assumptions 2.3.1 and 2.3.3 hold, then the fonction u is rightcontinuous in the direction \overrightarrow{e} on $\overline{\mathcal{O}}_{\alpha}$.

Proof. Let $y = (t, x, z) \in \overline{\mathcal{O}}_{\alpha}$.

(i) Fix $\nu \in [0,1]$ and $h, h' \geq 0$. Then $(y + h\vec{e}) \in \overline{\mathcal{O}}_{\alpha}$ and $(y + h'\vec{e}) \in \overline{\mathcal{O}}_{\alpha}$. We pick any $(C,\theta) \in \mathcal{A}_{\alpha}(y + h\vec{e}), (C',\theta') \in \mathcal{A}_{\alpha}(y + h'\vec{e})$, and introduce the notation $(X, X') := (X^{t,x+h,C,\theta}, X^{t,x+h',C',\theta'})$ and $(X^*, (X')^*)$ for their current maxima. We then derive

$$\nu X + (1 - \nu)X' \geq \nu \{\alpha(z + h) \lor X^*\} + (1 - \nu)\{\alpha(z + h') \lor (X')^*\}$$

$$\geq \{\alpha(z + \nu h + (1 - \nu)h')\} \lor \{\nu X + (1 - \nu)X'\}^*.$$

Therefore $\nu(C,\theta) + (1-\nu)(C',\theta') \in \mathcal{A}_{\alpha}(y + \{\nu h + (1-\nu)h'\}\overrightarrow{e})$ and it follows from the concavity of J(t,.) inherited from U, that

$$\nu J(t,C,\theta) + (1-\nu)J(t,C',\theta') \leq u \left(y + \left\{ \nu h + (1-\nu)h' \right\} \overrightarrow{e} \right) \,.$$

The arbitrariness of (C, θ, C', θ') then leads to the concavity of $h \mapsto u[y + h \overrightarrow{e}]$.

(ii) Suppose $y \in \mathcal{O}_{\alpha} \cup \partial^T \mathcal{O}_{\alpha} \cup \partial^1 \mathcal{O}_{\alpha}$. Then, there exists $h_0 > 0$ satisfying $y - h_0 \overrightarrow{e} \in \overline{\mathcal{O}}_{\alpha}$. Recalling from (i) that the function $h \mapsto u(y + (h - h_0)\overrightarrow{e})$ is concave on \mathbb{R}_+ , it is also continuous on $(0, \infty)$ and we deduce that u is right continuous in the direction \overrightarrow{e} at point y.

(iii) Suppose now that $y \in \partial^0 \mathcal{O}_{\alpha}$. By Lemma 2.3.1, u(y) = 0. Under Assumption 2.3.2, it follows from the same arguments as in the proof of Lemma 2.3.1 that $u(y') \leq u^*(x')$, for any $(t', x', z') \in \overline{\mathcal{O}}_{\alpha}$, where u^* is the value function in the classical Merton setting (i.e. $\alpha = 0$). Thus, the required regularity result is a consequence of the continuity of u^* .

(iv) Suppose finally that $y \in \partial^{\alpha} \mathcal{O}_{\alpha}$ and Assumptions 2.3.1, 2.3.2 and 2.3.3 hold. We then recall from Chapter 1 that the value function u^{α} in the infinite time horizon is continuous on $\{(x', z'), 0 < \alpha z' \leq x' \leq z'\}$ and satisfies $u^{\alpha}(\alpha z', z') = 0$ for any z' > 0. Combining (III.2.13) with similar arguments as above completes the proof. \Box

2.4 Numerical examples

In this section, we present a numerical scheme for the resolution of the Hamilton-Jacobi-Bellman equation (III.2.16) applying the ideas of Barles, Daher and Romano [6]. The purpose of these numerical experiments is to observe the dependance of the solution in the given finite horizon T of the investor and to observe the speed of convergence of the numerical solution to the explicit solution in infinite horizon derived in chapter 1.

The partial differential equation is degenerate since the variable z only appears in the definition of the domain of the equation, and we prefer to use an explicit scheme.

We fix a value z_0 of interest and consider a regular discretization grid $(z_i)_{i \leq N_z}$ with step Δz of the interval $[0, 2z_0]$. For each z_i , we decompose the interval $[\alpha z_i, z_i]$ on a grid with step $\Delta^i x$ such that the number of points N_x does not depend of i, which is always possible as soon as $\alpha \in \mathbb{Q}$. We hence obtain a discretization of the set $\{(x, z) \in [0, 2z_0]^2 : \alpha z \leq x \leq z\}$ into a product of a matrix (x_j^i) of size $N_x \times N_z$ and a vector (z_j) of size N_z . Since we deal with a Neumann condition at each point (x_j^i, z_i) , we also add one row to the previous matrix by defining $x_j^{i+1} = z_i + \Delta_x^i$, whose use is detailed below. For a given horizon T, we decompose the interval [0, T] with a time step Δt of order $(\Delta x)^2$.

The algorithm is constructed the following way. From an approximation $(\hat{u}(t_n, x_j^i, z_i))_{i,j}$ of the value fonction $u(t_n, ., .)$, we compute an approximation of $u(t_{n+1}, ., .)$ by

$$\hat{u}(t_{n+1}, x_j^i, z_i) = \tilde{u}(t_{n+1}, x_j^i, z_i) \mathbf{1}_{x_{j+1}^i \le z_j} + \tilde{u}(t_{n+1}, x_j^i, x_j^i) \mathbf{1}_{x_{j+1}^i > z_j}$$

where \tilde{u} is defined by

$$\begin{split} \tilde{u}(t_{n+1}, x_0^i, z_i) &= 0, \\ \tilde{u}(t_{n+1}, x_j^i, z_i) &= (1 - \beta \Delta t) \hat{u}(t_n, x_j^i, z_i) + dt V \left(\frac{\hat{u}(t_n, x_{j+1}^i, z_i) - \hat{u}(t_n, x_j^i, z_i)}{\Delta^i x} \right) \\ &- \frac{\lambda^2 dt}{2} \frac{[u(t_n, x_{j+1}^i, z_i) - u(t_n, x_j^i, z_i)]^2}{u(t_n, x_{j+1}^i, z_i) + 2u(t_n, x_j^i, z_i) - u(t_n, x_{j-1}^i, z_i)}, \quad \text{for } j > 0. \end{split}$$

Observe that the previous relation $\tilde{u}(t_{n+1}, x_0^i, z_i) = 0$ corresponds simply to the condition $u(., \alpha z, z) = 0$ for $z \ge 0$. As for the initialization of the algorithm, we simply take $\hat{u}(0, ., .) = 0$.

Remark 2.4.1 The initialization of the algorithm endues a small technical problem as the previous iteration procedure can not be applied at time $t_n = 0$. This difficulty can be overcome by considering the linear form of the Hamilton-Jacobi-Bellman equation where we observe that, imposing in this time step an upper-bound c_{max} on the possible consumption strategy leads to $\hat{u}(t_1, x_0^i, z_i) = U(c_{max}(x_0^i - z_i))$. From a numerical point of view, it gives the right shape to the value function and the influence of c_{max} is still under study.

This algorithm has been implemented in Matlab and we present in Figure 2.1 numerical results obtained by considering a power utility function value function and the particular set of parameters $\{\alpha, p, \sigma, \lambda, \beta\} = \{0.5, 0.2, 1, 3, 3\}$, corresponding to the numerical examples of Chapter 1 with $\alpha = 0.5$. As the horizon T tends to infinity, we observe a pretty fast monotone convergence of the estimated value function to the solution in infinite horizon. We also report in Figures 2.2 and 2.3 the corresponding consumption and investment strategies.



Figure 2.1: Value function versus the fraction of wealth x/z for different horizon T



Figure 2.2: Consumption versus the fraction of wealth x/z for different horizon T



Figure 2.3: Investment versus the fraction of wealth x/z for different horizon T

Viscosity property 2.5

This section is devoted to the proof of the following Proposition:

Proposition 2.5.1 The value function u is a viscosity solution of the dynamic programming equation (III.2.16).

2.5.1Supersolution property

In this subsection, we prove that u is a viscosity supersolution of (III.2.16). We first observe from lemma 2.3.1 that $u_* \geq 0$ on $\partial^T \mathcal{O}_{\alpha} \cup \partial^0 \mathcal{O}_{\alpha}$. Let $y_0 := (t_0, x_0, z_0) \in \overline{\mathcal{O}}_{\alpha}$ and $\varphi \in C^{1,2,1}(\overline{\mathcal{O}}_{\alpha})$ such that

$$0 = (u_* - \varphi)(y_0) = \inf_{\overline{\mathcal{O}}_{\alpha}} (u_* - \varphi)$$

Without loss of generality, we can suppose that the previous infimum is indeed a strict minimum and we shall distinguish two different cases depending on the location of y_0 . 1. $\mathbf{y_0} \in \mathcal{O}_{\alpha}$. Let $y_n := (t_n, x_n, z_n)_n \in \mathcal{O}_\alpha$ satisfying

$$y_n \longrightarrow y_0$$
 and $u(y_n) \longrightarrow u_*(y_0)$.

We denote $\gamma_n := u(y_n) - \varphi(y_n) \ge 0$ and $\gamma_n^* := n^{-1} \lor \sqrt{\gamma_n}$. Since $y_0 \in \mathcal{O}_{\alpha}$, there exists r > 0 such that the open ball centered at y_0 with radius r satisfies $\mathcal{B}(y_0, r) \subset \mathcal{O}_{\alpha}$. We consider the constant strategy $(C,\theta) \in \mathbb{R}_+ \times \mathbb{R}$, denote $(Y^n, Z^n) := (Y^{y_n, C, \theta}, Z^{y_n, C, \theta})$ and introduce the stopping time

$$\tau_n := \inf \left\{ s \ge t_n : Y_s^n \notin \mathcal{B}(y_0, r) \right\} \land \left(t_n + \gamma_n^* \right).$$

The dynamic programming principle implies

$$e^{-\beta t_n} u(y_n) \geq \mathbb{E}\left[\int_{t_n}^{\tau_n} e^{-\beta s} U(C) ds + e^{-\beta \tau_n} u\left(Y_{\tau_n}^n\right)\right]$$

Since $u \ge u_* \ge \varphi$, we deduce

$$\gamma_n + e^{\beta t_n} \mathbb{E}\left[e^{-\beta t_n} \varphi(y_n) - e^{-\beta \tau_n} \varphi\left(Y_{\tau_n}^n\right)\right] \geq e^{\beta t_n} \mathbb{E}\left[\int_{t_n}^{\tau_n} e^{-\beta s} U(C) ds\right]$$

Applying Itô's lemma to the regular function $e^{\beta} \varphi$, together with the previous inequality, yields

$$\gamma_n \geq e^{\beta t_n} \mathbb{E} \left[\int_{t_n}^{\tau_n} e^{-\beta s} \mathcal{L}_T^{C,\theta} \varphi\left(Y_s^n\right) ds \right] \\ + e^{\beta t_n} \mathbb{E} \left[\int_{t_n}^{\tau_n} e^{-\beta s} \varphi_z\left(Y_s^n\right) dZ_s^n + \int_{t_n}^{\tau_n} e^{-\beta s} (\sigma \lambda \theta - C) \varphi_x\left(Y_s^n\right) dW_s \right].$$

Since $\varphi_x(Y^n)$ is bounded and Z^n is a constant process on the stochastic interval $[t_n, \tau_n]$, we deduce

$$\gamma_n \ge \mathbb{E}\left[\int_{t_n}^{\tau_n} e^{\beta(t_n-s)} \mathcal{L}^{C,\theta} \varphi\left(Y_s^n\right) ds\right].$$
(III.2.22)

Dividing by γ_n^* and letting *n* go to infinity, since $\tau_n = t_n + \gamma_n^*$ for *n* large enough almost surely, the dominated convergence theorem leads to $\mathcal{L}_T^{C,\theta}\varphi(y_0) \leq 0$. From the arbitrariness of $(C,\theta) \in \mathbb{R}_+ \times \mathbb{R}$, we deduce

$$-\mathcal{L}_T\varphi(y_0) \geq 0$$
.

2. $\mathbf{y_0} \in \partial^1 \mathcal{O}_{\alpha}$.

Remark first that u_* inherits the monotony property of u derived in lemma 2.3.2. Thus, for any $z \ge z_0$ such that $y := (t_0, x_0, z) \in \overline{\mathcal{O}}_{\alpha}$, we have $\varphi(y_0) = u_*(y_0) \ge u_*(y) \ge \varphi(y)$. Since φ is a regular function, we deduce

$$-\varphi_z(y_0) \ge 0$$

2.5.2 Subsolution property

In this subsection, we prove that u is a viscosity subsolution of (III.2.16). From Lemma 2.3.1, we have $u^* \leq 0$ on $\partial^T \mathcal{O}_{\alpha} \cup \partial^0 \mathcal{O}_{\alpha}$. Let $y_0 := (t_0, x_0, z_0) \in \overline{\mathcal{O}}_{\alpha}$ and $\varphi \in C^{1,2,1}(\overline{\mathcal{O}}_{\alpha})$ such that

$$0 = (u^* - \varphi)(y_0) = \sup_{\overline{\mathcal{O}}_{\alpha}} (u^* - \varphi).$$
(III.2.23)

Once again, without loss of generality, we can suppose that the previous supremum is indeed a strict maximum, and we shall distinguish two different cases depending on the location of the maximum y_0 .

1. $\underline{\mathbf{y}_0} \in \mathcal{O}_\alpha \cup \partial^\alpha \mathcal{O}_\alpha$.

Let introduce the function $m := -\mathcal{L}_T \varphi$, suppose that $m(y_0) > 0$ and work towards a contradiction. From (III.2.23) and the regularity of u^* and φ , we deduce the existence of r > 0 and $\eta > 0$ such that $\mathcal{B}(y_0, r) \cap \partial^T \mathcal{O}_{\alpha} = \mathcal{B}(y_0, r) \cap \partial^0 \mathcal{O}_{\alpha} = \emptyset$, and

$$\min_{\overline{\mathcal{B}}(y_0,r)\cap\overline{\mathcal{O}}_{\alpha}} m > 0 \quad \text{and} \quad \max_{\partial \mathcal{B}(y_0,r)\cap\overline{\mathcal{O}}_{\alpha}} (u^* - \varphi) < -3\eta.$$
(III.2.24)

Denote $\eta_r := \eta e^{-\beta r} > 0$ and take $(y_n)_n$ a sequence valued in $\mathcal{B}(y_0, r) \cap \mathcal{O}_{\alpha}$ satisfying

$$y_n \longrightarrow y_0, \quad u(y_n) \longrightarrow u^*(y_0) \quad \text{and} \quad |u(y_n) - \varphi(y_n)| \le \eta_r, \quad n \ge 0.$$
 (III.2.25)

For any $n \ge 0$, let (C^n, θ^n) be an η_r -optimal control at point y_n and introduce the notation $(Z^n, Y^n) := (Z^{y_n, C^n, \theta^n}, Y^{y_n, C^n, \theta^n})$. We introduce the stopping time τ_n defined by

$$\tau_n := \inf\{s \ge t_n, Y_s^n \notin \mathcal{B}(y_0, r)\}$$

By construction, Y^n is valued in $\overline{\mathcal{O}}_{\alpha}$, $\tau_n - t_n \leq r$ and the η_r -optimality of (C^n, θ^n) leads to

$$u(y_n) \leq e^{\beta t_n} \mathbb{E}\left[\int_{t_n}^{\tau_n} e^{-\beta s} U(C_s^n) ds + e^{-\beta \tau_n} u\left(Y_{\tau_n}^n\right)\right] + \eta_r.$$
(III.2.26)

Applying Ito's lemma to the regular function $e^{-\beta} \varphi$, we compute

$$e^{-\beta t_n}\varphi(y_n) = \mathbb{E}[e^{-\beta\tau_n}\varphi(Y_{\tau_n}^n)] - \mathbb{E}\left[\int_{t_n}^{\tau_n} e^{-\beta s} \left(\mathcal{L}_T^{C^n,\theta^n}\varphi(Y_s^n) - U(C_s^n)\right) ds\right] \\ - \mathbb{E}\left[\int_{t_n}^{\tau_n} e^{-\beta s}\varphi_z\left(Y_s^n\right) dZ_s^n\right].$$

Combining (III.2.25) with the negativity of $\mathcal{L}_T^{C^n,\theta^n}$ on $\overline{\mathcal{B}}(y_0,r) \cap \overline{\mathcal{O}}_{\alpha}$, we deduce

$$u(y_n) \ge -\eta_r + e^{\beta t_n} \mathbb{E}\left[e^{-\beta \tau_n} \varphi(Y_{\tau_n}^n) + \int_{t_n}^{\tau_n} e^{-\beta s} U(C_s^n) ds - \int_{t_n}^{\tau_n} \varphi_z\left(Y_s^n\right) dZ_s^n\right].$$
(III.2.27)

Noticing that $Y_{\tau_n}^n \in \partial \mathcal{B}(y_0, r) \cap \overline{\mathcal{O}}_{\alpha}$ and combining (III.2.24) and $\tau_n - t_n \leq r$, we derive

$$e^{\beta t_n} \mathbb{E}\left[e^{-\beta \tau_n}\left(\varphi\left(Y_s^n\right) - u^*\left(Y_s^n\right)\right)\right] \geq 3\eta_r.$$
 (III.2.28)

We now compute from (III.2.26), (III.2.27), (III.2.28) and $u \leq u^*$, that

$$\eta_r \leq \mathbb{E}\left[\int_{t_n}^{\tau_n} \varphi_z\left(Y_s^n\right) dZ_s^n\right].$$
(III.2.29)

Since $y_0 \in \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha}$, we have $\mathcal{B}(y_0, r) \cap \partial^1 \mathcal{O}_{\alpha} = \emptyset$ for r small enough. Thus Z^n is a constant process on the random interval $[t_n, \tau_n]$ and (III.2.29) leads to a contradiction. We therefore deduce

$$-\mathcal{L}_T\varphi(y_0) \leq 0.$$

2. $\underline{\mathbf{y}_0} \in \partial^1 \mathcal{O}_{\alpha}$.

Take $m := \min\{-\mathcal{L}_T \varphi, -\varphi_z\}$ and follow the lines of the proof in the previous case. This leads to (III.2.29) and, since $-\varphi_z(Y^n) \ge m(Y^n) > 0$ on the random interval $[t_n, \tau_n]$ according to (III.2.24), we obtain a contradiction. Therefore

$$\min\{-\mathcal{L}_T\varphi,-\varphi_z\}(y_0) \leq 0.$$

2.6 A comparison result

This section is devoted to the proof of a comparison result for the PDE (III.2.16) which ensures the uniqueness of the solution. The difficulty of the proof relies on the fact that the controls are not in a compact subset. To overcome this difficulty, we adapted the arguments of Zariphopoulou [103], in particular for the choice of the penalization function. As announced, a different version of the comparison theorem is discussed in Remark 2.6.1.

Theorem 2.6.1 Let w and v be respectively an upper-continuous sub-solution and a lower-semicontinuous super-solution of (III.2.16) on $\overline{\mathcal{O}}_{\alpha}$. Suppose that the function vis right-continuous in the direction $\overrightarrow{e} = (0, 1, 1)$ on $\mathcal{O}_{\alpha} \cup \partial^{1}\mathcal{O}_{\alpha} \cup \partial^{\alpha}\mathcal{O}_{\alpha}$ and that the positive part of w and the negative part of v satisfy the following growing condition

$$[w]^+(y) + [v]^-(y) \leq K(1+x^{p'}), \quad y = (t, x, z) \in \overline{\mathcal{O}}_{\alpha}, \quad with \ p' < \frac{\gamma}{1+\gamma}, \ (\text{III.2.30})$$

and K a positive constant. Then, if $w \leq v$ on $\partial^0 \mathcal{O}_{\alpha} \cup \partial^T \mathcal{O}_{\alpha}$, we have $w \leq v$ on $\overline{\mathcal{O}}_{\alpha}$.

Proof. We do not consider the case $\alpha = 0$, already covered by the literature, see Zariphopoulou [103] for example. As a consequence, observe for later use that, for any $y = (t, x, z) \in \overline{\mathcal{O}}_{\alpha}$, we only need to control x in order to bound y, since $\alpha z \leq x \leq z$. We now suppose that

$$\sup_{y \in \overline{\mathcal{O}}_{\alpha}} [w(y) - v(y)] > 0$$
 (III.2.31)

and work towards a contradiction. For any $y \in \overline{\mathcal{O}}_{\alpha}$, we denote by (t, x, z) its components, and this convention of notation is obviously extended to elements of $\overline{\mathcal{O}}_{\alpha}$ of the form y_i^j with *i* and *j* any subcripts and superscripts.

1. We define the function ϕ by

$$\phi(y,y') := w(y) - v(y') - \delta\left(x^q + (x')^q + e^{-z} + e^{-z'}\right), \quad (y,y') \in \overline{\mathcal{O}}_{\alpha} \times \overline{\mathcal{O}}_{\alpha},$$

with $\delta > 0$ and $q := \gamma/(1+\gamma) < 1$. Choosing δ small enough and combining the growth condition (III.2.30), (III.2.31) and the semi-continuity properties of w and v, we deduce that the function $y \mapsto \phi(y, y)$ attains its suppremum on $\overline{\mathcal{O}}_{\alpha}$ and we have

$$\phi(\bar{y}, \bar{y}) := \sup_{y \in \overline{\mathcal{O}}_{\alpha}} \phi(y, y) > 0.$$
(III.2.32)

Since $w \leq v$ on $\partial^0 \mathcal{O}_{\alpha} \cup \partial^T \mathcal{O}_{\alpha}$, (III.2.32) leads to $\bar{y} \in \mathcal{O}_{\alpha} \cup \partial^1 \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha}$. Therefore, the right-continuity of v in the direction \vec{e} and the semi-continuity of w ensures that

$$\phi(\bar{y}, \bar{y} + \vec{e}/n) \underset{n \to \infty}{\longrightarrow} \phi(\bar{y}, \bar{y}) > 0.$$
 (III.2.33)

2. For any $n \ge 0$, we now define the function

$$\psi^{n}(y,y') := \left[n([x-\alpha z] - [x'-\alpha z']) + 1 - \alpha\right]^{2} + \alpha(1-\alpha)\left[n(z-z') + 1\right]^{2}$$

for $(y, y') \in \overline{\mathcal{O}}_{\alpha} \times \overline{\mathcal{O}}_{\alpha}$. Since $\psi^n(\bar{y}, \bar{y} + \vec{e}/n) = 0$, we deduce from (III.2.33) that

$$\{\phi - \psi^n\}(\bar{y}, \bar{y} + \vec{e}/n) > 0, \qquad (\text{III.2.34})$$

for *n* large enough. Therefore, according to (III.2.30), the function $\phi - \psi^n$ attains its maximum on $\overline{\mathcal{O}}_{\alpha} \times \overline{\mathcal{O}}_{\alpha}$ and we have

$$\{\phi - \psi^n\}(y_n, y'_n) := \sup_{(y, y') \in \overline{\mathcal{O}}_{\alpha} \times \overline{\mathcal{O}}_{\alpha}} \{\phi - \psi^n\}(y, y') > 0.$$
(III.2.35)

The growing assumption (III.2.30) ensures the convergence along subsequences of $(y_n)_n$ and $(y'_n)_n$ and, sending n to ∞ , we see that $\psi^n(y_n, y'_n) \to \infty$ unless $|y_n - y'_n| \to 0$. But $\phi(y_n, y'_n) - \psi^n(y_n, y'_n)$ is bounded from above according to (III.2.30) and therefore $|y_n - y'_n| \to 0$ as n goes to ∞ . Denoting y_0 the common limit of $(y_n)_n$ and $(y'_n)_n$, since $\{\phi - \psi^n\}(y_n, y'_n) \ge \phi(\bar{y}, \bar{y} + \vec{e}/n)$, we deduce from (III.2.33) and the semi-properties of w and v that

$$\phi(y_0, y_0) \geq \limsup_{n \to \infty} \{\phi - \psi^n\}(y_n, y'_n) \geq \phi(\bar{y}, \bar{y}).$$

Recalling (III.2.32), we derive

$$\phi(y_0, y_0) > 0$$
 and $\psi^n(y_n, y'_n) \underset{n \to \infty}{\longrightarrow} 0.$ (III.2.36)

3. We now discuss the location of (y_n, y'_n) and some properties of the global penalization function given by

$$\Phi^n(y,y') := \delta(x^q + (x')^q + e^{-z} + e^{-z'}) + \psi^n(y,y'), \quad (y,y') \in \overline{\mathcal{O}}_\alpha \times \overline{\mathcal{O}}_\alpha$$

Since $w \leq v$ on $\partial^0 \mathcal{O}_{\alpha} \cup \partial^T \mathcal{O}_{\alpha}$, we derive from (III.2.36) that $y_0 \in \mathcal{O}_{\alpha} \cup \partial^1 \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha}$. Furthermore, for *n* large enough, (III.2.36) implies that $x'_n - \alpha z'_n > x_n - \alpha z_n$, and we deduce that

$$y_n \in \mathcal{O}_{\alpha} \cup \partial^1 \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha} \quad \text{and} \quad y'_n \in \mathcal{O}_{\alpha} \cup \partial^1 \mathcal{O}_{\alpha} \,.$$
 (III.2.37)

In particular, since $x_n \neq 0$, Φ^n is regular on a neighborhood of (y_n, y'_n) and we denote $D_{x,z}\Phi^n$ (resp. $D_{x',z'}\Phi^n$) its gradient with respect to (x, z) (resp. (x', z')) and $H\Phi^n$ its Hessian matrix with respect to the space variables (x, z, x', z'). Observe for later use that

$$\Phi_{z}^{n}(y_{n}, y_{n}') = -\alpha n^{2}(z_{n}' - x_{n}') - \delta e^{-z_{n}} < 0, \quad \text{if } y_{n} \in \partial^{1}\mathcal{O}_{\alpha}, \quad (\text{III.2.38})$$

$$\Phi_{z'}^{n}(y_{n}, y_{n}') = -\alpha n^{2}(z_{n} - x_{n}) - \delta e^{-z_{n}'} < 0, \quad \text{if } y_{n}' \in \partial^{1} \mathcal{O}_{\alpha}, \quad (\text{III.2.39})$$

and

202

$$\Phi_x^n(y_n, y_n') + \Phi_{x'}^n(y_n, y_n') = \delta q(x_n^{q-1} + (x_n')^{q-1}) \ge 0.$$
 (III.2.40)

4. For any $\epsilon > 0$, we deduce from Theorem 8.3 in [30] the existence of $b \in \mathbb{R}$ and two real symmetric matrices Λ and Λ' such that

$$(b, D_{x,z}\Phi^n(y_n, y'_n), \Lambda) \in \overline{\mathcal{P}}^{2,+}_{\overline{\mathcal{O}}_{\alpha}}w(y_n), (b, -D_{x',z'}\Phi^n(y_n, y'_n), \Lambda') \in \overline{\mathcal{P}}^{2,-}_{\overline{\mathcal{O}}_{\alpha}}v(y'_n),$$
(III.2.41)

and

$$A := \begin{pmatrix} \Lambda & 0 \\ 0 & -\Lambda' \end{pmatrix} - H\Phi^{n}(y_{n}, y_{n}') + \epsilon \{H\Phi^{n}(y_{n}, y_{n}')\}^{2} \leq 0, \qquad (\text{III.2.42})$$

where $\overline{\mathcal{P}}_{\mathcal{O}_{\alpha}}^{2,+}$ and $\overline{\mathcal{P}}_{\mathcal{O}_{\alpha}}^{2,-}$ denotes classically the superjet and subjet operators, see [30] for the precise definition. We compute that $H\Phi^{n}(y_{n}, y'_{n})$ is explicitly given by

$$H\Phi^{n}(y_{n},y_{n}') = n^{2} \begin{pmatrix} 1 & -\alpha & -1 & \alpha \\ -\alpha & \alpha & \alpha & -\alpha \\ -1 & \alpha & 1 & -\alpha \\ \alpha & -\alpha & -\alpha & \alpha \end{pmatrix} - \delta q(1-q) \begin{pmatrix} x_{n}^{q-2} & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 \\ 0 & 0 & (x_{n}')^{q-2} & 0 \\ 0 & 0 & 0 & \delta \end{pmatrix}$$

Take X := (1, 0, 1, 0) and observe that (III.2.42) implies $XAX^T \leq 0$, which leads to

$$\Lambda_{1,1} - \Lambda'_{1,1} \le -\delta q(1-q) [x_n^{q-2} + (x_n')^{q-2}] + \epsilon [q(1-q)(x_n^{q-2} + (x_n')^{q-2})]^2 < 0, \quad (\text{III.2.43})$$

for ϵ sufficiently small.

5. According to (III.2.37), (III.2.38) and (III.2.39), it follows from (III.2.41) and the viscosity properties of w and v that

$$\beta w(y_n) \leq b + V \left[\Phi_x^n(y_n, y'_n) \right] + \sup_{\theta \in \mathbb{R}} \left\{ \sigma \lambda \theta \Phi_x^n(y_n, y'_n) + \frac{(\sigma \theta)^2}{2} \Lambda_{1,1} \right\} \,,$$
and

$$\beta v(y'_n) \geq b + V \left[-\Phi_{x'}^n(y_n, y'_n) \right] + \sup_{\theta \in \mathbb{R}} \left\{ -\sigma \lambda \theta \Phi_{x'}^n(y_n, y'_n) + \frac{(\sigma \theta)^2}{2} \Lambda'_{1,1} \right\} ,$$

where V denotes the Fenchel transform of U. Combining these inequalities with the decreasing property of V and (III.2.40), we deduce

$$\begin{split} \beta\{w(y_n) - v(y'_n)\} &\leq \sup_{\theta \in \mathbb{R}} \left\{ \sigma \lambda \theta \Phi_x^n(y_n, y'_n) + \frac{(\sigma \theta)^2}{2} \Lambda_{1,1} \right\} \\ &- \sup_{\theta \in \mathbb{R}} \left\{ -\sigma \lambda \theta \Phi_{x'}^n(y_n, y'_n) + \frac{(\sigma \theta)^2}{2} \Lambda_{1,1}' \right\} \\ &\leq \sup_{\theta \in \mathbb{R}} \left\{ \sigma \lambda \theta \left[\Phi_x^n + \Phi_{x'}^n \right](y_n, y'_n) + \frac{(\sigma \theta)^2}{2} (\Lambda_{1,1} - \Lambda_{1,1}') \right\} \,. \end{split}$$

According to (III.2.40) and (III.2.43), we then deduce

$$\beta\{w(y_n) - v(y'_n)\} \leq \frac{\lambda^2}{2} \frac{\left[\delta q(x_n^q + (x'_n)^q)\right]^2}{\delta q(1-q)(x_n^{q-2} + (x'_n)^{q-2}) - \epsilon[q(1-q)(x_n^{q-2} + (x'_n)^{q-2})]^2}$$

Since this inequality holds true for any $\epsilon > 0$, it follows that

$$w(y_n) - v(y'_n) \leq \frac{\delta q(x_n^{q-1} + (x'_n)^{q-1})^2}{\gamma(1-q)(x_n^{q-2} + (x'_n)^{q-2})}$$

Letting n go to infinity, we finally obtain

$$\phi(y_0, y_0) \leq w(y_0) - v(y_0) - 2\delta x_0^q \leq \left(\frac{q}{\gamma(1-q)} - 1\right) 2\delta x_0^q.$$

Since $q = \gamma/(1+\gamma)$, we deduce $\phi(y_0, y_0) \leq 0$ and therefore contradict (III.2.36).

Remark 2.6.1 The results of Theorem 2.6.1 hold true if we suppose that v is rightcontinuous in the direction \overrightarrow{e} on $\mathcal{O}_{\alpha} \cup \partial^{1} \mathcal{O}_{\alpha}$ instead of $\mathcal{O}_{\alpha} \cup \partial^{1} \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha}$, but that $w \leq v$ on $\partial^{0} \mathcal{O}_{\alpha} \cup \partial^{T} \mathcal{O}_{\alpha} \cup \partial^{\alpha} \mathcal{O}_{\alpha}$ instead of $\partial^{0} \mathcal{O}_{\alpha} \cup \partial^{T} \mathcal{O}_{\alpha}$. The only modification of the previous proof relies on the obtention of (III.2.33), which remains valid since $\overline{y} \in \mathcal{O}_{\alpha} \cup \partial^{1} \mathcal{O}_{\alpha}$. Denoting furthermore that the decreasing property of V, used in part **5**. of the previous proof, relies only on the monotonicity of U, (iii) of Lemma 2.3.3 leads to Proposition 2.3.1.

204 CONSUMPTION-INVESTMENT UNDER DRAWDOWN CONSTRAINT

Bibliography

- ACHDOU Y. & O. PIRONNEAU (2005). Computational Methods for Option Pricing. Frontiers in Applied Mathematics, SIAM.
- [2] ANKIRCHNER S., P. IMKELLER & A. POPIER (2006). On measure solutions of backward stochastic differential equations. *Preprint*.
- [3] ANTONELLI F. & A. KOHATSU-HIGA (2000). Filtration stability of backward SDE's. Stochastic Analysis and Its Applications, 18, p. 11-37.
- [4] AIT-SAHALIA, Y. (1996). Non parametric pricing of interest rate derivative securities. *Econometrica*, 64, p. 527-560.
- [5] BARLES G., R. BUCKDAHN & E. PARDOUX (1997). Backward stochastic differential equations and integral-partial differential equations. *Stochastics Stochastics Reports*, 60, p. 57-83.
- [6] BARLES G., C. DAHER & M. ROMANO (1994). Optimal control of the L[∞]−norm of a diffusion process. SIAM Journal on Control and Optimization, 32, p. 612-634.
- [7] BARLES, G. & P.E. SOUGANIDIS (1991). Convergence of approximation schemes for fully nonlinear equations. *Asymptotic Analysis*, 4, p. 271-283.
- [8] BALLY V. & G. PAGES (2002). A quantization algorithm for solving discrete time multidimensional optimal stopping problems. *Bernoulli*, 9 (6), p. 1003-1049.
- [9] BECHERER D. (2005). Bounded solutions to Backward SDE's with jumps for utility optimization and indifference hedging. *Preprint, Imperial College London*.
- [10] BEN TAHAR I., M. SONER & N. TOUZI (2005). Modelling continuous-time financial markets with capital gains taxes. *Preprint*.
- [11] BENDER C. & J. ZHANG (2006). Time discretization and Markovian iteration for coupled FBSDEs. WIAS Preprint No 1160.

- [12] BICHTELER K., J.-B. GRAVEREAUX & J. JACOD (1987). Malliavin calculus for processes with jumps. Gordon and Breach Science Publishers, New York.
- [13] BICHTELER K. & J. JACOD (1983). Calcul de Malliavin pour des diffusions avec saut: existence d'une densité dans le cas unidimensionel. Séminaire de Probabilité, 17, p. 132-157.
- [14] BILLINGSLEY, P. (1968). Convergence of probability measures, Wiley.
- [15] BISMUT J. M. (1976). Théorie probabiliste du contrôle des diffusions. Mem. Amer. Math. Soc., 4-167, p. 132-157.
- [16] BISMUT J. M. (1975). Growth and optimal intertemporal allocations of risks. J. of Economic Theory, 10, p. 239-287.
- [17] BLACK F. & M. SCHOLES (1973). The Pricing of Options and Corporate Liabilities. The Journal of Political Economy, 81 (3), p. 637-654.
- [18] BOUCHARD B. & J.-F. CHASSAGNEUX (2006). Discrete time approximation for continuously and discretely reflected BSDE's. Preprint LPMA, Univ. Paris 6.
- [19] BOUCHARD B. & N. TOUZI (2004). Discrete-Time Approximation and Monte-Carlo Simulation of Backward Stochastic Differential Equations. *Stochastic Pro*cesses and their Applications, 111 (2), p. 175-206.
- [20] BRÉMAUD P. (1981). Point Processes and Queues Martingale Dynamics. Springer-Verlag, New-York.
- [21] BRIAND P. & B. DELYON, & J. MÉMIN (2001). Donsker-type theorem for BSDE's. Electronic Communications in Probability, 6, p. 1-14.
- [22] BRIAND P. & Y. HU (2006). BSDE with quadratic growth and unbounded terminal value. Probab. Theory and Related Fields, 136 (4), p. 509-660.
- [23] BROADIE M. & P. GLASSERMAN (1996). Estimating security prices using simulation. *Management Science*, 42, p. 269-285.
- [24] BRUTI-LIBERATI N. & E. PLATEN (2005). On the strong Approximation of Jump-Diffusion Processes. Technical report, Quantitative Finance Research Papers 157, University of Terchnology, Sydney.
- [25] CAI T. (2002). On adaptive wavelet estimation of a derivative and other related linear inverse problems. J. Statist. Plann. Inference, 108, p. 329-349.

- [26] CHEVANCE D. (1997). Numerical Methods for Backward Stochastic Differential Equations. In Numerical methods in finance, Edt L.C.G. Rogers and D. Talay, Cambridge University Press, p. 232-244.
- [27] CONSTANTINIDES G.M. & M.J.P. MAGILL (1976). Portfolio Selection with Transaction Costs, *Journal of Economic Theory*, 13, p. 245-263.
- [28] COQUET F., V. MACKEVIČIUS, & J. MÉMIN (1998). Stability in D of martingales and backward equations under discretization of filtration. *Stochastic Pro*cesses and their Applications, 75, p. 235-248.
- [29] COX J. & C.F. HUANG (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49, p. 33-83.
- [30] CRANDALL M.G., H. ISHII & P.L. LIONS (1992). User's guide to viscosity solutions of second order partial differential equations. Bull. Amer. Math. Soc., 27 (1), p. 1-67.
- [31] CRANDALL M.G. & P.L. LIONS (1983). Viscosity solutions of Hamilton-Jacobi Equations. Trans. Amer. Math. Soc., 277, p. 1-42.
- [32] BIELECKY T.R., S. CRÉPEY, M. JEANBLANC & M. RUTKOWSKY (2006). Valuation and hedging of defaultable game options in a hazard process model. Work in preparation.
- [33] CVITANIĆ J. & I. KARATZAS (1992). Convex duality in constrained portfolio optimization. Annals of Applied Probability, 2, p. 767-818.
- [34] CVITANIĆ, J. & I. KARATZAS (1995). On portfolio optimization under drawdown constraints. IMA volumes in Math. and its Applications, 65, p. 35-46.
- [35] DAVIS M.H.A. & A.R. NORMAN (1990). Portfolio selection with transaction costs. Mathematics of Operations Research, 15, p. 676-713.
- [36] DETEMPLE J., R. GARCIA & M. RINDISBACHER (2005). Asymptotic Properties of Monte Carlo Estimators of Derivatives. *Management Science*, 51 (11), p. 1657-1675.
- [37] DELARUE F. (2002) Equation différentielles stochastiques progressives rétrogrades, Application à l'homogénéisation des EDP Quasi-linéaires. PhD Thesis. Université de provence.

- [38] DELARUE F. & S. MENOZZI (2006). A forward-backward stochastic algorithm for quasi-linear PDEs. Annals of Applied Probability, 16 (1), p. 140-184.
- [39] DELARUE F. & S. MENOZZI (2006). An interpolated Stochastic Algorithm for Quasi-Linear PDEs. *Preprint*.
- [40] DONOHO D., I. JOHNSTONE, G. KERKYACHARIAN & D. PICARD (1996). Density estimation by wavelet thresholding. Annals of Statistics, 24 (2), p. 508-539.
- [41] DOUGLAS J. JR., J. MA & P. PROTTER (1996). Numerical Methods for Forward-Backward Stochastic Differential Equations. Annals of Applied Probability, 6, p. 940-968.
- [42] L'ECUYER P. & G. PERRON (1994). On the Convergence Rates of IPA and FDC derivative Estimators. Operations Research, 42, p. 643-656.
- [43] EL KAROUI N. (2006). Azéma-Yor martingales in finance. Invited plenary presentation at the Stochastic Processes and Applications conference, Paris.
- [44] EL KAROUI N. & M. JEANBLANC (1998). Optimization of consumption with labor income. *Finance and Stochastics*, 2, p. 409-440.
- [45] EL KAROUI N., M. JEANBLANC & V. LACOSTE (2005). Optimal portfolio management with American capital garantee. J. Econ. Dyn. Control, 29 (3), p. 409-440.
- [46] EL KAROUI N. & A. MESIOU. (2006). Constrained optimization with respect to stochastic dominance: application to portfolio insurance. *Mathematical Finance*, 16 (1), p. 103.
- [47] EL KAROUI N., S. PENG & M.-C. QUENEZ (1997). Backward stochastic differential equations in finance. Mathematical finance, 7 (1), p. 1-71.
- [48] EYRAUD-LOISEL A. (2005). Backward Stochastic Differential Equations with enlarged filtration. Option Hedging of an insider trader in a financial market with Jumps. To appear in *Stochastic processes and their Applications*.
- [49] FORSTER B., E. LÜTKEBOHMERT AND J. TEICHMANN (2005). Calculation of the greeks for jump-diffusions. *Preprint*.
- [50] FOURNIÉ E., J.M. LASRY, J. LEBUCHOUX, P.L. LIONS & N. TOUZI (1999). Applications of Malliavin Calculus to Monte Carlo Methods in Finance. *Finance and Stochastics*, 3, p. 391-412.

- [51] FOURNIÉ E., J.M. LASRY, J. LEBUCHOUX & P.L. LIONS (2000). Applications of Malliavin Calculus to Monte Carlo Methods in Finance. II. Finance and Stochastics, 5, p. 201-236.
- [52] FUJIWARA T. & H. KUNITA (1989). Stochastic differential equations of Jump type and Lévy processes in diffeomorphism group. J. Math. Kyoto Univ., 25 (1), p. 71-106.
- [53] GILES M. & P. GLASSERMAN (2006). Smoking adjoints: fast Monte Carlo Greeks. Risk, p. 92-96.
- [54] GOBET, E. (2004). Revisiting the Greeks for European and American options. In J. Akhori, S. Ogawa and S. Watanabe, editors, *Stochastic processes and applications to mathematical finance*, p. 53-71.
- [55] GOBET, E. & A. KOHATSU-HIGA (2003). Computation of Greeks for barrier and Lookback options using Malliavin Calculus. *Electronic Communications in Probability*, 8, p. 51-62.
- [56] GOBET E. & C. LABART (2006). Error expansion for the discretization of backward stochastic differential equations. To appear in Stochastic Processes and Applications.
- [57] GOBET E. & J.P. LEMOR (2006). Numerical simulation of bsdes using empirical regression methods : theory and practice. In S. Tang and S. Paeng, editors. To appear in Proceedings of the Fifth Colloquiim on BSDEs (29th May - 1st June 2005, Shangay).
- [58] GOBET E., J.P. LEMOR & X. WARIN (2005). A regression based Monte Carlo Method to solve Backward Stochastic Differential Equations. Annals of Applied Probability, 15 (3), p. 2172-2202.
- [59] GROSSMAN S.J. & Z. ZHOU (1993). Optimal investment strategies for controlling drawdowns. Math. Finance, 3 (3), p. 241-276.
- [60] GYORFI L., M. KOHLER, A. KRZYZAK & H. WALK (2002). A distribution free theory of nonparametric regression. Springer Series in Statistiques.
- [61] HAMADÈNE S. & Y. OUKNINE (2003). Reflected backward stochastic differential equation with jumps and random obstacle. *Electronic Journal of Probability*, 8 (2), p. 1-20.

- [62] HE H. & H. PAGÈS (1993). Labor income, borrowing constraints and equilibrium asset prices. *Economic Theory*, 3, p. 663-696.
- [63] HULL, J. (2002). Options, futures, and other derivatives. Prentice Hall.
- [64] KARATZAS I., J.P. LEHOCZKY & S.E. SHREVE (1987). Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. SIAM Journal on Control and Optimization, 25, p. 1557-1586.
- [65] KARATZAS I. & S.E. SHREVE (1998). Methods of Mathematical Finance, Springer-Verlag, New York.
- [66] KLASS M.J. & K. NOWICKI (2005). The Grossman and Zhou investment strategy is not always optimal. *Statistics and Probability Letters*, 74, p. 245-252.
- [67] KLOEDEN P. & E. PLATEN (2000). Numerical Solution of Stochastic Differential Equations. Springer.
- [68] KOBYLANSKI M. (2000). Backward stochastic differential equations and partial differential equations with quadratic growth. Annals of Probability, 28 (2), p. 558-602.
- [69] KOHATSU-HIGA, A. & MONTERO, M. (2004). Malliavin Calculus in Finance. Handbook of Computational and Numerical Methods in Finance, Birkhauser, p. 111-174.
- [70] KRAMKOV D. & W. SCHACHERMAYER (1999). The condition on the Asymptotic Elasticity of Utility Functions and Optimal Investment in Incomplete Markets. Annals of Applied Probability, 9, p. 904-950.
- [71] KUNITA, H. (1984). Ecole d'été de Probabilité de Saint Flour XII 1982, Stochastic differential equations and stochastic flow of diffeomorphisms. Springer-Verlag.
- [72] LEMOR, J.P. (2005). Approximation par projections et simulations Monte Carlo des equations differentielles retrogrades. PHD thesis.
- [73] LEMOR J.P., E. GOBET & X. WARIN (2006). Rate of convergence of empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12 (5), p.889-916.
- [74] LIEBSCHER E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation Stochastic processes and their applications, 65 (1), p. 69-80.

- [75] LONGSTAFF F. A. & R. S. SCHWARTZ (2001). Valuing American Options By Simulation : A simple Least-Square Approach. *Review of Financial Studies*, 14, p. 113-147.
- [76] MA J., P. PROTTER, J. SAN MARTIN & S. TORRES (2002). Numerical Method for Backward Stochastic Differential Equations. Annals of Applied Probability, 12 (1), p. 302-316.
- [77] MA J., P. PROTTER & J. YONG (1994). Solving forward-backward stochastic differential equations explicitly - a four step scheme. Probability Theory and Related Fields, 98, p. 339-359.
- [78] MA J. & ZHANG J. (2002). Path Regularity of Solutions to Backward Stochastic Differential Equations. Probability Theory and Related Fields, 122, p. 163-190.
- [79] MERTON R.C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time model. *Review of Economic Statistics*, 51, p. 247-257.
- [80] MERTON R.C. (1971). Optimum consumption and portfolio rules in a continuoustime model. *Journal of Economic Theory*, 3, p. 373-413.
- [81] MILSTEIN G. & M. TRETYAKOV (2005). Numerical Analysis of Monte Carlo Evaluation of Greeks by Finite Differences. Journal of Computational Finance, 8 (3), p. 1-34.
- [82] NUALART D. (1995). The Malliavin Calculus and Related Topics. Springer Verlag, Berlin.
- [83] NUALART D. & E. PARDOUX (1988). Stochastic calculus with anticipating integrands. Prob. Theory and Rel. Fields, 78, p. 535-581.
- [84] PARDOUX E. & S. PENG (1990). Adapted solution of a backward stochastic differential equation. Systems & Control Letters, 14 (1), p. 55-61.
- [85] PARDOUX E. & S. PENG (1992). Backward stochastic differential equations and quasilinear parabolic partial differential equations. Lecture Notes in Control and Inform. Sci, 176, p. 200-217.
- [86] PARDOUX E., F. PRADEILLES & Z. RAO (1997). Probabilistic interpretation for a system of semilinear parabolic partial differential equations. Ann. Inst. H. Poincare, 33 (4), p. 467-490.

- [87] PHAM H. (2005). On some recent aspects of stochastic control and their applications. Probability surveys, 2, p. 506-549.
- [88] PHAM H. (2006). Optimisation et Contrôle Stochastique Appliqués à la Finance. Springer Verlag.
- [89] PLISKA S.R. (1986). A stochastic calculus model of continuous trading: optimal portfolios. Math. Operations Research, 11, p. 371-382.
- [90] POLLARD, D. (1984). Convergence of stochastic processes. Springer.
- [91] PORCHET A., N. TOUZI & X. WARIN (2006). Valuation of a power plant under production constraints and market incompleteness. *Preprint*.
- [92] PROTTER P. (1990). Stochastic integration and differential equations. Springer Verlag, Berlin.
- [93] ROCHE H. (2005). Optimal consumption and investment under a drawdown constraint. *Preprint*.
- [94] ROUGE R. & N. EL KAROUI (2000). Pricing Via Utility Maximization and Entropy. Mathematical Finance, 10 (2), p. 259-276.
- [95] RONG S. (2006). BSDEs with jumps and with quadratic growth coefficients and optimal consumption. *Preprint*.
- [96] SCHACHERMAYER W. (2001). Optimal Investment in Incomplete Markets when Wealth may Become Negative. Annals of Applied Probability, 11, p. 694-734.
- [97] SHREVE S.E. & H.M. SONER (1994). Optimal investment and consumption with transaction costs. Annals of Applied Probability, 4, p. 609-692.
- [98] SOW A. B. & E. PARDOUX (2004). Probabilistic interpretation of a system of quasilinear parabolic PDEs. Stochastics and Stochastics Reports, 76 (5), p. 429-477.
- [99] SCOTT D.W. (1992). Multivariate Density estimation. Wiley.
- [100] TANG S. & X. LI (1994). Necessary conditions for optimal control of stochastic systems with random jumps. SIAM J. Control Optim., 32 (5), p. 1447-1475.
- [101] TAVELLA D. & C. RANDALL (2000). Pricing Financial Instruments: The Finite Difference Method. Wiley.

- [102] XU G.L. (1990). A duality method for optimal consumptions and investment under short-selling prohibition. Doctoral dissertation, Department of mathematics, Carnegie-Mellon University.
- [103] ZARIPHOPOULOU T. (1994). Consumption-investment models with constraints. SIAM J. control and optimization, 32 (1), p. 59-85.
- [104] ZHANG J. (2001). Some fine properties of backward stochastic differential equations. PhD thesis, Purdue University.
- [105] ZHANG J. (2004). A numerical scheme for BSDEs. Annals of Applied Probability, 14 (1), p. 459-488.

Cette thèse présente trois sujets de recherche indépendants appartenant au domaine des méthodes numériques et du contrôle stochastique avec des applications en mathématiques financières. Nous présentons dans la première partie une méthode nonparamétrique d'estimation des sensibilités des prix d'options. A l'aide d'une perturbation aléatoire du paramètre d'intérêt, nous représentons ces sensibilités sous forme d'espérance conditionnelle, que nous estimons à l'aide de simulations Monte Carlo et de régression par noyaux. Par des arguments d'intégration par parties, nous proposons des estimateurs à noyaux de ces sensibilités, qui ne nécessitent pas la connaissance de la densité du sous-jacent, et nous obtenons leurs propriétés asymptotiques. Lorsque la fonction payoff est irrégulière, ils convergent plus vite que les estimateurs par différences finies, ce que l'on vérifie numériquement. La deuxième partie s'intéresse à la résolution numérique de systèmes découplés d'équations différentielles stochastiques progressives rétrogrades. Pour des coefficients Lipschitz, nous proposons un schéma de discrétisation qui converge plus vite que $n^{-1/2+\varepsilon}$, pour tout $\varepsilon > 0$, lorsque le pas de temps 1/n tends vers 0. Lorsque les coefficients sont C_h^1 à dérivées Lipschitz, ou que le terme de saut du processus tangent de la composante progressive de l'équation satisfait une condition de non-dégénérescence, nous obtenons la vitesse optimale en $n^{-1/2}$. L'utilisation pratique de ce schéma nécessite le calcul d'un grand nombre d'espérances conditionnelles, que nous approchons à l'aide de techniques d'estimation non-paramétrique. Nous contrôlons l'erreur globale commise par l'algorithme ce qui permet le choix simultané de ses paramètres, et nous présentons des exemples de résolution numérique de systèmes couplés d'EDP semi-linéaires. Enfin, la dernière partie de cette thèse étudie le comportement d'un gestionnaire de fond, maximisant l'utilité intertemporelle de sa consommation, sous la contrainte que la valeur de son portefeuille ne descende pas en dessous d'une fraction fixée de son maximum courant. Nous considérons une classe générale de fonctions d'utilité, et un marché financier composé d'un actif risqué de dynamique Black-Scholes. Lorsque le gestionnaire se fixe un horizon de temps infini, nous obtenons sous forme explicite sa stratégie optimale d'investissement et de consommation, ainsi que la fonction valeur du problème. En horizon fini, nous caractérisons la fonction valeur comme unique solution de viscosité de l'équation d'Hamilton-Jacobi-Bellman correspondante.

Abstract

This PhD dissertation presents three independent research topics in the fields of numerical methods and stochastic control with applications to financial mathematics. The first part of this thesis is dedicated to the estimation of the sensitivities of option prices, by means of non-parametric techniques. When the density of the underlying is unknown, we propose several non-parametric estimators of the so called *Greeks*, based on the randomization of the parameter of interest combined with Monte Carlo simulations and Kernel regression techniques. We provide an asymptotic analysis of the mean squared error of these estimators, as well as their asymptotic distributions. For a discontinuous payoff function, the kernel estimators outperforms the classical finite differences one in terms of the asymptotic rate of convergence. This result is confirmed by our numerical experiments. The second part of this dissertation deals with the numerical resolution of systems of decoupled forward-backward stochastic differential equations with jumps. Assuming that the coefficients are Lipschitz-continuous, we propose a convergent discrete-time scheme whose rate of convergence is at least $n^{-1/2+\varepsilon}$, for any $\varepsilon > 0$, when the number of time steps n goes to infinity. Under the additional condition that, either all the coefficients are C_h^1 with Lipschitz derivatives, or the jump coefficient of the first variation process of the forward component satisfies a non-degeneracy condition which ensures its invertibility, we achieve the optimal convergence rate $n^{-1/2}$. The implementation of this scheme requires the computation of a large number of conditional expectations, that we approximate by means of non parametric regression techniques. We control the global error of the algorithm, allowing to calibrate all the parameters of estimation at the same time, and provide the numerical solution of systems of coupled semilinear parabolic PDE's. The third part of this thesis is concerned with the resolution of the optimal consumption-investment problem under a drawdown constraint, i.e. the wealth process never falls below a fixed fraction of its running maximum. We assume that the risky asset is driven by the constant coefficients Black and Scholes model and we consider a general class of utility functions. On an infinite time horizon, we provide the value function in explicit form, and we derive closed-form expressions for the optimal consumption and investment strategy. On a finite time horizon, we interpret the value function as the unique viscosity solution of its corresponding Hamilton-Jacobi-Bellman equation.