EDSR: analyse de discrétisation et résolution par méthodes de Monte Carlo adaptatives;Perturbation de domaines pour les options américaines

Celine Labart

Thèse présentée pour obtenir le titre de

DOCTEUR DE L'ECOLE POLYTECHNIQUE

Spécialité: Mathématiques appliquées

par

Céline Labart

EDSR: analyse de discrétisation et résolution par méthodes de Monte Carlo adaptatives; Perturbation de domaines pour les options américaines

Thèse soutenue le 22 octobre 2007 devant le jury composé de:

Vlad BALLY	*Rapporteur*
Bruno BOUCHARD	*Examinateur*
Cristina COSTANTINI	*Examinateur*
Nicole EL KAROUI	*Président du jury*
Emmanuel GOBET	*Directeur de thèse*
Arturo KOHATSU-HIGA	*Rapporteur*
Damien LAMBERTON	*Examinateur*
Nizar TOUZI	*Examinateur*

Quand tu arrives en haut de la montagne, continues de grimper.
Proverbe tibétain

Remerciements

Mes premiers remerciements vont tout naturellement à Emmanuel Gobet, qui a été un directeur de thèse exemplaire, tant par ses qualités humaines que professionelles. Je le remercie pour son encadrement si régulier pendant ces trois ans, malgré la distance qui nous a séparés dès la seconde année. Je le remercie aussi pour la qualité de son encadrement, si sérieux et si rigoureux.

Je remercie tous les membres du jury pour le temps qu'ils m'ont consacré. Mes remerciements vont tout particulièrement à Vlad Bally et à Arturo Kohatsu-Higa, qui ont accepté de rapporter cette thèse. Je leur suis reconnaissante pour la lecture très attentive du manuscrit et leurs nombreuses remarques.

Je remercie Bruno Bouchard, Cristina Costantini, Nicole El Karoui, Damien Lamberton et Nizar Touzi d'avoir accepté de faire partie de mon jury.

I would like to thank Cristina Costantini for her warm welcome during my stay in Italy last year. I hold in high regard her great human and professional qualities.

Cette thèse a été menée au Centre de Mathématiques Appliquées de l'Ecole Polytechnique. Je tiens à remercier tous ses membres, actuels et anciens, pour leur gentillesse et leur soutien. Je pense notamment à Jeanne, Násséra et Sébastien qui m'ont toujours accueilli avec le sourire. Je pense aussi à mes collègues de l'équipe "Aléatoire, Finance et Statistique", pour leurs précieux conseils lors de mes exposés au groupe de travail, ainsi que mes amis pour leur soutien. Un grand merci donc à Ying, Elisabeth, Jie, Carlos, Caroline, Romain, Claudia, Julien, Guoshen, Mohammed, Claire et Emily.

Pour finir, je voudrais remercier ma famille et plus particulièrement mes parents pour m'avoir soutenue et aidée jusque là. Ma plus grande joie est de vous voir fiers de moi. Il ne me reste qu'à te dire merci, Jérôme, pour ton aide et ton soutien dans cette longue ascencion.

Résumé

Deux thématiques différentes des probabilités numériques et de leurs applications financières sont abordées dans ma thèse : l'une traite de l'approximation et de la simulation d'équations différentielles stochastiques rétrogrades (EDSR), l'autre est liée aux options américaines et les aborde du point de vue de l'optimisation de domaine et des perturbations de frontière.
La première partie de ma thèse revisite la question d'analyse de convergence dans la discrétisation en temps d' EDSR markoviennes (Y, Z) en une équation de programmation dynamique de n pas de temps. Nous établissons un développement limité à l'ordre 1 de l'erreur sur (Y, Z) : précisément, l'erreur trajectorielle sur X se transfère intégralement sur l'EDSR et montre ainsi que si X est approché avec précision ou simulé exactement, de meilleures vitesses sont possibles (en $1/n$).
La seconde partie de ma thèse s'intéresse à la résolution des EDSR via le procédé de Picard et les méthodes de Monte Carlo séquentielles. Nous avons montré que la convergence de notre algorithme a lieu à vitesse géométrique et avec une précision indépendante au 1er ordre du nombre de simulations.
La dernière partie de ma thèse regroupe des premiers résultats sur la valorisation d'options américaines par optimisation de la frontière d'exercice. La clé de voûte de ce type d'approche est la capacité à évaluer un gradient par rapport à la frontière. Le temps continu a été traité par Costantini et al (2006) et cette thèse couvre le cas discret des options Bermuda.

Abstract

My thesis deals with two different themes of numerical probabilities and their financial applications : the first one is the approximation and the simulation of backward stochastic differential equations (BSDE). The second one concerns the American options and tackle their pricing using domain optimization and boundary perturbations.
The first part of my thesis analyzes the convergence of the time discretization (via a n steps dynamic programming equation) of markovian BSDE (Y, Z). We establish a Taylor expansion for the error on (Y, Z) : it strongly depends on the error on X. Had we been able to perfectly simulate X, we would have obtained an error on (Y, Z) of order $1/n$.
The second part of my thesis is devoted to solving BSDE using Picard's procedure and a sequential Monte Carlo method. We prove that our algorithm converges geometrically fast. Moreover, the accuracy is independent (at the first order) of the number of Monte Carlo simulations.
The last part of my thesis presents basic results on the pricing of American options using an optimization of the exercise region. The keystone of such an approach is the ability of computing a gradient w.r.t the boundary. In continuous time, this work has been done by Costantini et al (2006). This thesis deals with the discrete time.

Table des matières

Introduction

La thèse que je présente se décompose en quatre parties plus ou moins distinctes, mais qui appartiennent toutes au vaste domaine des probabilités numériques et de leur utilisation en finance. Les deux thématiques principales sont les équations différentielles stochastiques rétrogrades (notées EDSR par la suite) et les options américaines. La première partie est consacrée au développement de l'erreur commise lors de la discrétisation d' EDSR. Dans la seconde partie, nous rappelons et établissons des résultats techniques liés à la densité de transition d'un processus de diffusion et de son schéma d' Euler, résultats que nous utiliserons dans la troisième partie. Celle-ci présente un algorithme de résolution numérique des EDSR basé sur une méthode de Monte Carlo séquentielle. Dans la dernière partie, nous développons une technique de valorisation des options américaines via un calcul de sensibilité sur des domaines.

Méthodes numériques pour les EDSR

Les EDSR ont été introduites par J.M. Bismut en 1973 dans l'article Bismut [14]. Cet article concernait le contrôle stochastique optimal et la version probabiliste du principe du maximum de Pontryagin. Le premier résultat général concernant les EDSR date de 1990, il est dû à E. Pardoux et S. Peng (voir Pardoux and Peng [81]). C'est en 1997 que N. El Karoui, S. Peng et M.C. Quenez écrivent l'article fondateur de l'application des EDSR à la finance (voir El Karoui et al. [27]), article que nous citerons maintes fois au cours de ce manuscrit. Dans cette thèse, nous nous intéresserons plus particulièrement aux méthodes numériques de simulation des EDSR, à la vitesse de convergence des schémas numériques et nous présenterons un algorithme de résolution des EDSR qui utilise une technique de Monte Carlo séquentielle. Dans les parties I, II, et III, nous considérerons des EDSR markoviennes du type

$$-\,dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = \Phi(X_T), \tag{1}$$

où le processus X satisfait l'équation progressive

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{2}$$

Nous supposerons aussi que X est à valeurs dans $\mathbb{R}^d$, Y à valeurs dans $\mathbb{R}$ et Z à valeurs dans $\mathbb{R}^q$.

Développement de l'erreur commise lors de la discrétisation d' EDSR

Le but de cette première partie est d'étudier l'approximation en temps du triplet (X, Y, Z), approximation que nous noterons (X^N, Y^N, Z^N). Nous discrétisons X grâce à un schéma d'Euler en temps continu à N pas de temps $(t_k = kh)_{0\le k\le N}$, où $h = \frac{T}{N}$: $X_0^N = x$ et

$$\forall t \in [t_k, t_{k+1}],\ X_t^N = X_{t_k}^N + b(t_k, X_{t_k}^N)(t - t_k) + \sigma(t_k, X_{t_k}^N)(W_t - W_{t_k}).$$

L'équation (1) est approchée de manière rétrograde par $Y_{t_N}^N = \Phi(X_{t_N}^N)$ et

$$\begin{aligned} Y_{t_k}^N =& \mathbb{E}_{t_k}(Y_{t_{k+1}}^N) + h\mathbb{E}_{t_k} f(t_k, X_{t_k}^N, Y_{t_{k+1}}^N, Z_{t_k}^N), \\ hZ_{t_k}^N =& \mathbb{E}_{t_k}(Y_{t_{k+1}}^N \Delta W_k^*), \end{aligned} \tag{3}$$

où $\mathbb{E}_{t_k}$ désigne l'espérance conditionnelle par rapport à $\mathcal{F}_{t_k}$, et $\Delta W_k = W_{t_{k+1}} - W_{t_k}$. Il est bien connu que l'erreur $X^N - X$ en norme $\mathbb{L}^p$ est d'ordre $\frac{1}{\sqrt{N}}$. Les résultats de cette partie concernent les erreurs $(Y^N - Y, Z^N - Z)$, mesurées en norme $\mathbb{L}^p$ et presque sûrement. Les résultats en norme $\mathbb{L}^p$ présentés dans le théorème 2.3 sont une généralisation de ceux obtenus par Zhang [94] pour $p = 2$. Notre théorème établit, sous des hypothèses d'ellipticité de σ et de bornitude des fonctions b, σ, f et Φ, que l'erreur

$$e_p(N) = \left[\max_{0\le k\le N} \mathbb{E}|Y_{t_k} - Y_{t_k}^N|^p + \mathbb{E}\left(\sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} |Z_{t_k}^N - Z_t|^2 dt \right)^{\frac{p}{2}} \right]^{\frac{1}{p}},$$

est d'ordre $\frac{1}{\sqrt{N}}$.

Les résultats donnant un développement presque sûr des erreurs $(Y^N - Y, Z^N - Z)$, énoncés dans les théorèmes 2.4 et 2.5, nous assurent que sous des hypothèses plus fortes de dérivabilité et de bornitude des fonctions b, σ, f et Φ, nous avons

$$\begin{aligned} Y_{t_k}^N - Y_{t_k} =& \psi_1(t_k, X_{t_k})(X_{t_k}^N - X_{t_k}) + O\left(\frac{1}{N}\right) + O(|X_{t_k}^N - X_{t_k}|^2), \\ Z_{t_k}^N - Z_{t_k} =& \psi_2(t_k, X_{t_k})(X_{t_k}^N - X_{t_k}) + O\left(\frac{1}{N}\right) + O(|X_{t_k}^N - X_{t_k}|^2), \end{aligned}$$

pour des fonctions ψ_1 et ψ_2 explicites.

Bien sûr, puisque $|X_{t_k}^N - X_{t_k}|$ est de l'ordre de $\frac{1}{\sqrt{N}}$ en norme $\mathbb{L}^p$, nous retrouvons bien que l'erreur forte $e_p(N)$ est d'ordre $\frac{1}{\sqrt{N}}$. Mais notre résultat montre surtout que les erreurs $(Y^N - Y, Z^N - Z)$ dépendent principalement de l'erreur trajectorielle commise lors de la discrétisation de X. Cela apporte à notre avis un éclairage nouveau sur l'approximation des EDSR, car il était communément admis que l'erreur est d'ordre $\frac{1}{\sqrt{N}}$ du fait de l'équation de programmation dynamique (3). Nous montrons ainsi qu'il n' en n'est rien. Notamment dans le cas où X se simule de manière exacte aux instants $(t_k)_k$ (comme dans le cas du mouvement Brownien arithmétique ou géométrique, ou des processus d'Ornstein Uhlenbeck), les erreurs sur Y et Z sont d'ordre $\frac{1}{N}$. Cela ravive l'intérêt pour des schémas de discrétisation de X d'ordre 1, comme le schéma de Milshtein.

La première partie s'articule donc comme suit : le chapitre 1 concerne l'introduction du problème et rappelle les précédents travaux réalisés sur ce sujet. Le chapitre 2 présente les résultats principaux brièvement exposés ci-dessus, ainsi que des expériences numériques corroborant les résultats théoriques. Les chapitres 3, 4 et 5 contiennent respectivement les preuves des théorèmes 2.3, 2.4 and 2.5. Ces preuves utilisent des techniques d'analyse stochastique, combinant des propriétés sur les martingales et du calcul de Malliavin.

Algorithme numérique pour la simulation des EDSR

La troisième partie de cette thèse présente un algorithme numérique simulant une approximation de (Y, Z), solution de (1). Cet algorithme est basé sur une méthode de Monte Carlo séquentielle et le principe d'itération de Picard. Plusieurs autres méthodes ont déjà été proposées dans la littérature : Ma, Protter, and Yong [74], Bally and Pagès [6], Delarue and Menozzi [24], Bouchard and Touzi [15], Zhang [94], Gobet, Lemor, and Warin [45] et Bender and Denk [11]. Nous reviendrons plus en détails sur les différents algorithmes présentés dans ces articles en introduction de la partie III. Disons simplement qu'elles s'appuient principalement sur le principe de programmation dynamique (3) alors que nous exploitons plutôt le principe de contraction de Picard.

Description de l'algorithme

L'algorithme numérique que nous présentons ici ne résout pas à proprement parler l'EDSR (1), mais il nous renvoie une solution approchée de l'équation aux dérivées partielles (notée EDP par la suite) semilinéaire associée à cette EDSR

$$(\mathcal{E}) \begin{cases} \partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),(\partial_x u\sigma)(t,x)) = 0, \\ u(T,x) = \Phi(x), \end{cases}$$

où $\mathcal{L}$ est défini par

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

Les résultats liant les EDP semilinéaires et les EDSR seront rappelés dans la section 9.4. Ils sont dûs à Pardoux and Peng [82] entre autres. Nous pouvons aussi trouver certaines preuves dans El Karoui et al. [27]. Plus récemment, Delarue and Menozzi [24] ont présenté un résultat liant les solutions d' EDSR aux solutions d' EDP quasilinéaires. C'est de ce résultat que nous déduisons le théorème 9.14, page 99, théorème que nous utiliserons tout au long de cette partie : Si b, σ et f sont des fonctions lipschitziennes et bornées, si σ est uniformément elliptique et si Φ est de classe $C_b^{2+\alpha}$ ($\alpha \in]0,1]$), alors la solution u de l'EDP $(\mathcal{E})$ est $C_b^{1,2}$. De plus, $(Y_t, Z_t)_{0\leq t\leq T}$ solution de (1) satisfait

$$\forall t \in [0,T], \quad (Y_t, Z_t) = (u(t,X_t), \partial_x u(t,X_t)\sigma(t,X_t)). \tag{4}$$

Nous noterons u_k l'approximation de u fournie par notre algorithme à l'itération k. En simulant X via un schéma d'Euler, nous déduisons de (4) une approximation de Y, notée Y^k, et définie par $Y_t^k = u_k(t, X_t^N)$.

Nous sommes jusqu'à présent restés volontairement assez vagues sur la manière dont nous construisons pratiquement u_k. Le temps est venu de nous expliquer. A chaque étape k, nous calculons u_k sur une grille de points $(T_i, X_i)_{1\leq i\leq n}$, puis nous utilisons un opérateur $\mathcal{P}$, régulier et facilement différentiable, de type estimateur à noyaux, pour construire u_k en tout point d'un pavé de $[0,T]\times\mathbb{R}^d$. Comme $\mathcal{P}$ est régulier, il nous permet de calculer $\partial_x u_k$, et donc d'en déduire Z_t^k grâce à (4), i.e. $Z_t^k = (\partial_x u_k \sigma)(t, X_t^N)$. Les estimateurs à noyaux, et plus généralement les techniques de régression linéaire et non linéaire seront présentés aux chapitres 10 et 11, d'après les ouvrages Györfi, Kohler, Krzyżak, and Walk [47] et Härdle [50].

Pour finir la description de notre algorithme, il nous faut expliquer le passage d' une itération à l'autre. Au lieu de décrire directement l'algorithme que nous avons implémenté et étudié, nous allons d'abord proposer un premier algorithme, plus simple, mais qui permettra de comprendre le cheminement des idées telles qu'elles nous sont apparues. Une première idée consiste à utiliser le principe d'itération de Picard, décrit dans le corollaire 2.1 de El Karoui et al. [27]. L'idée est la suivante : nous pouvons écrire grâce à (1)

$$Y_t = \mathbb{E}[\Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds \big| \mathcal{F}_t].$$

En utilisant (4), l'équation précédente devient

$$u(t,x) = \mathbb{E}_{t,x}[\Phi(X_T) + \int_t^T f(s, X_s, u(s, X_s), (\partial_x u\sigma)(s, X_s))ds],$$

où $\mathbb{E}_{t,x}$ signifie que l'on calcule une espérance sachant que le processus X part de x en t. En supposant qu'à l'itération $k-1$ nous connaissons u_{k-1} et sa première dérivée en temps, un bon candidat pour u_k serait

$$\mathbb{E}_{t,x}[\Phi(X_T) + \int_t^T f(s, X_s, u_{k-1}(s, X_s), (\partial_x u_{k-1}\sigma)(s, X_s))ds]. \tag{5}$$

Bien entendu, le calcul numérique de cette espérance se fait par une méthode de Monte Carlo, et le processus X est remplacé par son approximation X^N. Nous constatons ici que le calcul direct de $\partial_x u_k$ semble difficile. L'idée précédemment exposée de calculer u_k en une grille de points, de le régulariser via l'opérateur $\mathcal{P}$ semble donc être une bonne idée. Ce premier algorithme sera étudié et comparé au second algorithme, décrit quelques lignes plus bas, dans le chapitre 15.

Afin d'améliorer drastiquement la performance de l'algorithme ci-dessus, nous avons mis en oeuvre une méthode de Monte Carlo séquentielle (ou adaptative). Cette technique a été utilisée par Gobet and Maire [39] pour résoudre des EDP linéaires, liées par la formule de Feynman-Kac à des espérances de fonctionnelles de processus de Markov. Plus précisément, cette technique peut être vue comme une méthode de variable de contrôle adaptative, car elle va nous permettre, en soustrayant et en ajoutant un terme à (5), de réduire la variance du terme de (5) dont on souhaite calculer l'espérance. Puisqu'à l'étape $k-1$ nous supposons u_{k-1} connu, il sera à la base de notre variable de contrôle. Grâce à la formule d'Itô, nous pouvons écrire

$$u_{k-1}(t,x) = \mathbb{E}_{t,x}[u_{k-1}(T, X_T^N) - \int_t^T (\partial_t + \mathcal{L}^N) u_{k-1}(s, X_s^N) ds],$$

où $\mathcal{L}^N$ est l'opérateur associé à X^N. En introduisant $\pm u_{k-1}$ dans (5), le nouveau candidat pour u_k est donc

$$\begin{aligned} u_k(t,x) &= u_{k-1}(t,x) + \frac{1}{M}\sum_{m=1}^{M} [\Phi(X_T^{m,N}) - u_{k-1}(T, X_T^{m,N}) \\ &+ \int_t^T f(s, X_s^{m,N}, u_{k-1}(s, X_s^{m,N}), (\partial_x u_{k-1}\sigma)(s, X_s^{m,N})) + (\partial_t + \mathcal{L}^N) u_{k-1}(s, X_s^{m,N}) ds]. \end{aligned}$$

Pour expliquer intuitivement en quoi la variance est grandement réduite, il suffit de remarquer que lorsque u_{k-1} est proche de u, dans la moyenne empirique les deux termes ($\Phi - u_{k-1}$ et l'intégrale) sont proches de 0 du fait de l'EDP satisfaite par u. C'est l'effet variable de contrôle adaptative.

L'expression ci-dessus de u_k n'est pas tout à fait exacte, dans la mesure où u_k est calculé via l'opérateur $\mathcal{P}$, c'est à dire que nous calculons la valeur du terme de droite en $(T_i, X_i)_{1\le i\le n}$ points de la grille, puis nous utilisons l'opérateur $\mathcal{P}$ pour obtenir u_k en tout point de $[0,T]\times\mathbb{R}^d$. C'est dans un souci de clarté que tous les détails ne sont pas présentés ici, ils le seront à la section 9.6.

Convergence de l'algorithme

La convergence de cet algorithme est énoncée dans le théorème 13.10. Sous des hypothèses appropriées de dérivabilité et de bornitude des fonctions b et σ, il est prouvé la convergence à vitesse géométrique de l'erreur $Y^k - Y$ et $Z^k - Z$ dans la norme suivante

$$\|V\|^2_{\mu,\beta} := \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |V_s(x)|^2 e^{-\mu|x|} dx ds\right],$$

où V désigne un processus prévisible $V : \Omega\times[0,T]\times\mathbb{R}^d \to \mathbb{R}^q$. Dans le cas de Y et Z, x représente le point de départ du processus X. Cette norme n'est pas tout à fait banale, voici plusieurs éléments qui permettent de justifier l'utilisation de celle-ci. Le coefficient β provient de l'espace $\mathbb{H}^2_{T,\beta}(\mathbb{R}^q)$ utilisé dans El Karoui et al. [27] pour établir des estimées a priori sur la différence entre les solutions de deux équations rétrogrades (voir la proposition 2.1 de cet article), estimées dont nous avons eu besoin au cours de la preuve du théorème 13.10. Le coefficient μ provient de la combinaison de deux résultats différents, dont est issue la proposition 7.4, un des principaux résultats de la partie II, mais aussi outil indispensable aux preuves de la partie III. Le premier résultat dont est issue la proposition 7.4 provient de l'ouvrage Bensoussan and Lions [13]. Il s'agit plus précisément du théorème 6.12 page 130, et qui permet de majorer la solution v d'une EDP du type $(\partial_t + \mathcal{L})v = f$, $u(T,\cdot) = 0$ de la manière suivante

$$\|\partial_t v\|_{\mathbb{L}^p(0,T;W^{0,p,\mu})} + \|v\|_{\mathbb{L}^p(0,T;W^{2,p,\mu})} \le C\,\|f\|_{\mathbb{L}^p(0,T;W^{0,p,\mu})}.$$

Ce résultat sera présenté au chapitre 7. Le second résultat est dû à Bally and Matoussi [5]. Leur proposition 1.6 établit le résultat d'équivalence de normes suivant : sous des hypothèses fortes de dérivabilité et de bornitude de b et σ et pour toutes les fonctions $\Psi \in L^1((0,T) \times \mathbb{R}^d, dt \otimes e^{-\mu|x|}dx)$, on a

$$c \int_{\mathbb{R}^d} \int_t^T |\Psi(s,x)| ds e^{-\mu|x|} dx \leq \int_{\mathbb{R}^d} \int_t^T \mathbb{E}(|\Psi(s, X_s^{t,x})|) ds e^{-\mu|x|} dx$$
$$\leq C \int_{\mathbb{R}^d} \int_t^T |\Psi(s,x)| ds e^{-\mu|x|} dx.$$

Ce résultat sera présenté au chapitre 6.
Avant de poursuivre, nous voudrions faire remarquer que l'intégration par rapport à $e^{-\mu|x|}dx$ qui apparaît dans $\|V\|^2_{\mu,\beta}$ peut être interprétée comme une intégration par rapport à la loi initiale de la diffusion X. Ce choix de norme nous semble assez crucial pour mesurer la stabilité et la convergence de notre algorithme.

Le plan de la partie III est le suivant : le chapitre 9 présente le cadre de travail, rappelle certains résultats connus sur les EDSR et leurs liens avec les EDP, ainsi que les techniques de variables de contrôle adaptatives. La fin du chapitre est dédiée à la présentation de notre algorithme. Le chapitre 10 introduit brièvement les techniques de régression, et le chapitre 11 présente plus précisément les estimateurs à noyaux. Dans le chapitre 12, nous étudions la convergence de l'estimateur $\mathcal{P}$, inspiré des estimateurs à noyaux, c'est-à-dire que nous nous intéressons à $v - \mathcal{P}v$ et à $\partial_x v - \partial_x(\mathcal{P}v)$, où v est une fonction dérivable et bornée. Nous présentons les principaux résultats de convergence de notre algorithme dans le chapitre 13, tandis que les preuves de ces résultats sont établies au chapitre 14. Les vitesses de convergence sont standards et coïncident avec celles de la littérature. En revanche, les difficultés dans notre cas tiennent au choix d' une norme à poids $\|\cdot\|_{\mu,\beta}$ et au fait que nous travaillons en domaine infini. Pour finir, le chapitre 15 présente des expériences numériques dans lesquelles nous avons appliqué notre algorithme à des exemples financiers, comme la valorisation et la couverture d'options sous contraintes (cas d'un investisseur qui emprunte de l'argent à un taux d'intérêt supérieur au taux de placement sans risque, valorisation d'options américaines,..)

Options bermuda - Valorisation par méthode de sensibilité sur les domaines

La partie IV de cette thèse concerne la valorisation des options bermuda par méthode de sensibilité sur les domaines. Afin de mieux comprendre la problématique, nous allons tout d'abord présenter l'application de la méthode à la valorisation des options américaines.

Cas des options américaines

Une option américaine d'achat ou de vente est un instrument financier qui donne le droit mais non l'obligation d'acheter ou de vendre une certaine quantité d'actifs financiers dits risqués (actions, obligations, devises) à n'importe quelle date avant l'échéance T et à un prix convenu à l'avance (prix d'exercice). Pour obtenir ce droit, l'acheteur de l'option paie au vendeur une prime (valeur de l'option). Une question importante est la détermination de cette prime à chaque instant t précédant l'échéance T. La théorie moderne des options américaines due à Bensoussan [12] et Karatzas [55] relie la valeur d'une option américaine à la théorie de l'arrêt optimal. Dans un marché complet, la valeur à l'instant t d'une option américaine d'échéance T définie par un profit $g(X_t)$, où X est une diffusion de type (2), est donnée par $P(t, X_t)$ où

$$P(t,x) = \sup_{\tau \in [t,T]} \mathbb{E}\left[e^{-\int_t^\tau r(s,X_s^{t,x})ds} g(\tau, X_\tau^{t,x})\right], \tag{6}$$

le supremum étant pris sur l'ensemble des temps d'arrêts à valeurs dans $[t,T]$. La complétude du marché est assurée dès lors que la matrice σ est inversible. D'autre part, comme nous le rappellerons dans le chapitre 16, on sait caractériser la valeur d'une option américaine comme la solution d'une inéquation variationnelle parabolique du second ordre (voir Bensoussan and Lions [13] et Jaillet et al. [54]). Ces deux approches font apparaître la région d'exercice de l'option

$$\mathcal{E} := \{(t,x), t < T,\ x \in \mathbb{R}^d : P(t,x) = g(t,x)\}.$$

Dans l'étude des options américaines, le problème de la détermination de la région d'exercice suscite un grand intérêt, car il fait partie intégrante du problème d'évaluation et permet de déterminer une stratégie optimale pour l'acheteur de l'option, car celui-ci a tout intérêt à exercer son droit au premier instant où (t, X_t) appartient à $\mathcal{E}$. Lorsque l'ensemble $\mathcal{E}$ est vide, le problème se résume à l'évaluation d'une option européenne. L'étude de la région d'exercice en dimension 1 a été menée dans les articles suivants : Van Moerbeke [91], Kim [57], Jacka [52], Barles et al. [10], et plus récemment dans Lamberton and Villeneuve [66]. Le cas multidimensionnel a été traité dans les articles Broadie and Detemple [18] et Villeneuve [93]. Les différents résultats apportés par ces articles seront discutés dans la section 16.2.

Comme nous le verrons dans la proposition 16.8, le prix $P(t,x)$ donné par (6) peut aussi s'écrire

$$P(t,x) = \sup_{\mathcal{D} \subset]t,T[\times \mathbb{R}^d} \mathbb{E}\left[e^{-\int_t^{\tau_{\mathcal{D}}^{t,x}} r(s,X_s^{t,x})ds} g(\tau_{\mathcal{D}}^{t,x}, X_{\tau_{\mathcal{D}}^{t,x}}^{t,x})\right],$$

où $\tau_{\mathcal{D}}^{t,x}$ désigne le premier instant où $(s, X_s^{t,x})_{t \le s \le T}$ sort du domaine $\mathcal{D}$, où $\mathcal{D}$ est un ouvert. L'optimisation du terme de droite peut être faite via un algorithme de gradient qui utilise la sensibilité par rapport au domaine $\mathcal{D}$. En termes mathématiques, cela veut dire qu'il nous faut calculer la dérivée, par rapport au domaine $\mathcal{D}$, de $\mathbb{E}\left[e^{-\int_t^{\tau_{\mathcal{D}}^{t,x}} r(s,X_s^{t,x})ds} g(\tau_{\mathcal{D}}^{t,x}, X_{\tau_{\mathcal{D}}^{t,x}}^{t,x})\right]$. Cela a été étudié récemment par Costantini et al. [22]. Les auteurs introduisent la perturbation spatiale d'un domaine temps-espace $\mathcal{D}$ de la manière suivante

$$\mathcal{D}^\epsilon = \{(t,x) : (t, x + \epsilon\theta(t,x)) \in \mathcal{D}\}, \quad \epsilon \in \mathbb{R},$$

où θ est une fonction $C_b^{1,2}([0,T] \times \mathbb{R}^d)$. Ils définissent aussi $u^\epsilon(t,x) = \mathbb{E}\left(g(\tau_\epsilon^{t,x}, X_{\tau_\epsilon^{t,x}}^{t,x}) e^{-\int_t^{\tau_\epsilon^{t,x}} r(s,X_s^{t,x})ds}\right)$, où $\tau_\epsilon^{t,x}$ est le premier temps de sortie de

$(s, X_s^{t,x})$ du domaine $\mathcal{D}^\epsilon$. Ils montrent alors

$$\partial_\epsilon u^\epsilon(t,x)\big|_{\epsilon=0} = \mathbb{E}\left(e^{-\int_t^{\tau_{\mathcal{D}}^{t,x}} r(s,X_s^{t,x})ds}[(\nabla u - \nabla g)\theta](\tau_{\mathcal{D}}^{t,x}, X_{\tau_{\mathcal{D}}^{t,x}}^{t,x})\right). \tag{7}$$

Cas des options bermuda

C'est le cadre de nos contributions. Comme pratiquement le cours d'une action n'est connu qu'en un nombre fini de dates, calculer le prix d'une option américaine revient en fait à calculer le prix d'une option bermuda, c'est-à-dire que le supremum de la formule (6) est pris sur l'ensemble des temps d'arrêts à valeurs dans $\mathcal{T}_{t,T}^N := [t,T] \cap \mathcal{T}^N$ où $\mathcal{T}^N = \{t_0, \cdots, t_N : 0 = t_0 < t_1 < \cdots < t_N = T\}$. C'est pourquoi nous avons démontré, dans le théorème 17.15, une formule de sensibilité pour les options bermuda, analogue à (7). Pour cela, nous avons introduit

$$\mathcal{D}^{N,\epsilon} = \bigcup_{j=1}^{N} (t_j, \mathcal{D}_{t_j}^\epsilon),$$

qui représente l'union des perturbations appliquées à chaque section $\mathcal{D}_{t_j}$ du domaine $\mathcal{D}$. Comme précédemment, nous définissons $u_N^\epsilon(t,x) = \mathbb{E}\left(e^{-\int_t^{\tau_{N,\epsilon}^{t,x}} r(s,X_s^{t,x})ds} g(\tau_{N,\epsilon}^{t,x}, X_{\tau_{N,\epsilon}^{t,x}}^{t,x})\right)$, où $\tau_{N,\epsilon}^{t,x}$ désigne le premier instant $s \in \mathcal{T}_{t,T}^N$ où $(s, X_s^{t,x})$ sort du domaine $\mathcal{D}^{N,\epsilon}$. Nous démontrons le théorème 17.15 qui énonce dans le cas $r = 0$ la formule de sensibilité suivante

$$\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0} = \sum_{j=0, t_j > t}^{N-1} \mathbb{E}_{t,x}\left(\mathbf{1}_{\tau_N^{t,x} > t_j} \int_{\partial \mathcal{D}_{t_{j+1}}} p(t_j, X_{t_j}; t_{j+1}, m)(g - u_N)(t_{j+1}, m)\theta \cdot \overrightarrow{n}(t_{j+1}, m) d\sigma_m\right),$$

où $d\sigma_m$ représente l'intégrale de surface et p désigne la densité de transition du processus X.

Ce calcul de sensibilité peut a priori être un point de départ pour développer une méthode de valorisation d'options bermuda via une optimisation sur les domaines par un algorithme de type gradient. De nombreuses questions se posent alors, comme par exemple le calcul numérique de cette sensibilité : comment évaluer l'intégrale de surface et la valeur de $g - u_N$ au bord ? Comment représenter les domaines de manière efficace pour une modification aisée de ceux-ci au fil des différentes étapes de l'optimisation (construction de $\mathcal{D}^\varepsilon$ à partir de $\mathcal{D}$) ? Plusieurs pistes d'études ont été abordées pendant la thèse, en collaboration avec Cristina Costantini (Université "Gabriele d'Annunzio" de Chieti et Pescara, Italie). Ces travaux sont en cours de réalisation et ne sont pas présentés dans ce manuscrit.

Mentionnons toutefois que pour le calcul de la sensibilité, il semble plus pertinent de discrétiser l'espérance de l'intégrale de surface par une somme pondérée de valeurs de la fonction $g - u_N$ à la frontière, fonction évaluée elle-même efficacement par une méthode de Monte Carlo séquentielle.

Ce que nous présentons dans cette partie permet d'affiner notre compréhension de la sensibilité $\partial_\varepsilon u_N^\varepsilon(t,x)\mid_{\varepsilon=0}$ vis-à-vis de N, le nombre de dates d'exercice. En effet, il n'est pas rare que les méthodes numériques de valorisation d'options bermuda aient des comportements indésirables par rapport à N : explosion de la variance dans l'approche de type calcul de Malliavin (Lions and Régnier [70]) ; cumul des erreurs de régression dans l'approche de Longstaff and Schwartz [71]. Notre but est donc de mieux étudier la robustesse $\partial_\varepsilon u_N^\varepsilon(t,x)\mid_{\varepsilon=0}$ lorsque N est grand (asymptotiquement le cas américain). Nous établissons des résultats de convergence lorsque $N \longrightarrow \infty$, montrant que nous retrouvons la sensibilité du cas continu, ce qui est loin d'être trivial a priori. En fait, la preuve est incomplète et repose sur un résultat technique sur les overshoots de diffusion. Nous améliorons au passage certains résultats sur les overshoots, récemment étudiés dans Gobet and Menozzi [43]. En dehors de sa convergence, nous montrons que la sensibilité reste bornée uniformément en N lorsque g est de classe $H_{1+\alpha}$ ou dans le cas du put. Ces résultats eux non plus ne sont pas évidents à la vue de la formule (3) puisque que la sensibilité s'écrit comme une somme de N termes impliquant des densités de transition explosant lorsque N tend vers l'infini.

Cette partie doit être vue comme un programme de recherche en cours sur la valorisation des options bermuda/américaines par des optimisations de domaine, avec un certain nombre de résultats prometteurs.

Les différentes parties de cette thèse ont fait ou feront l'objet de publications :

La première partie correspond à un article intitulé "Error expansion for the discretization of Backward Stochastic Differential Equations". Il a été publié dans Stochastic Processes and their Applications, Volume 117, Issue 7, July 2007 pages 803 - 829. Cet article a été réalisé avec mon directeur de thèse Emmanuel Gobet.

Les secondes et troisièmes parties feront prochainement l'objet d'une publication sous le titre "Solving BSDEs with adaptive control variates", et probablement soumise au journal Annals of Applied Probability.

La dernière partie fera aussi l'objet d'une publication sous le titre "Pricing Bermudan options via boundary sensitivities". Elle sera probablement soumise au journal Finance and Stochastics.

Deux articles concernant des travaux sur les options parisiennes ont aussi été réalisés au cours de ma thèse. Ils ne sont pas joints à ce manuscrit, mais peuvent être trouvés à l'adresse http://www.cmap.polytechnique.fr/~labart/.

"Pricing double barrier Parisian Options using Laplace transforms", réalisé avec J. Lelong et soumis à Mathematical Finance depuis novembre 2006.

"Pricing Parisian Options using Laplace transforms", réalisé avec J. Lelong et soumis à Banque & Marchés depuis mai 2007.

Part I

Error expansion for the discretization of Backward Stochastic Differential Equations

This part corresponds to an article published in the journal Stochastic Processes and their Applications, Volume 117, Issue 7, July 2007 pages 803 - 829. It's a joint work with Emmanuel Gobet.

Chapter 1

Introduction

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space on which is defined a q-dimensional standard Brownian motion W, whose natural filtration, augmented with $\mathbb{P}$-null sets, is denoted by $(\mathcal{F}_t)_{0\leq t\leq T}$ (T is a fixed terminal time). We consider the solution (X, Y, Z) to a decoupled forward-backward stochastic differential equation (FBSDE in short). Namely, X is the $\mathbb{R}^d$-valued process solution of

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \tag{1.1}$$

and Y (resp. Z) is a real-valued adapted (resp. predictable $\mathbb{R}^q$-valued) process solution of

$$-dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = \Phi(X_T). \tag{1.2}$$

We assume standard Lipschitz properties on the coefficients, which ensure existence and uniqueness in appropriate L_2-spaces (see Pardoux and Peng [81], or Ma and Yong [72] for numerous references). During the last decade, more and more attention has been paid to these equations, because of their natural applications in Mathematical Finance or in the probabilistic resolution of semi-linear partial differential equations (PDE in short): see El Karoui et al. [27] or Pardoux [80].

Our aim is to study the most usual time approximation of (X, Y, Z). For X, we use the Euler scheme X^N with N discretization times $(t_k = kh)_{0\leq k\leq N}$ ($h = \frac{T}{N}$ is the time step). For convenience, set $\Delta W_k = W_{t_{k+1}} - W_{t_k}$ (ΔW_k^l component-wise). X^N is defined by $X_0^N = x$ and

$$t \in [t_k, t_{k+1}], \; X_t^N = X_{t_k}^N + b(t_k, X_{t_k}^N)(t - t_k) + \sigma(t_k, X_{t_k}^N)(W_t - W_{t_k}). \tag{1.3}$$

The backward SDE (1.2) is approximated by (Y^N, Z^N) defined in a backward manner by $Y_{t_N}^N = \Phi(X_{t_N}^N)$ and

$$Y_{t_k}^N = \mathbb{E}_{t_k}(Y_{t_{k+1}}^N) + h\mathbb{E}_{t_k} f(t_k, X_{t_k}^N, Y_{t_{k+1}}^N, Z_{t_k}^N), \tag{1.4}$$

$$hZ_{t_k}^N = \mathbb{E}_{t_k}(Y_{t_{k+1}}^N \Delta W_k^*), \tag{1.5}$$

where $\mathbb{E}_{t_k}$ is the conditional expectation w.r.t. $\mathcal{F}_{t_k}$ and $*$ is the transpose operator. Additional tools are needed to derive a fully implementable scheme, in particular for the

computations of conditional expectations. We refer to Bouchard and Touzi [15] for Malliavin calculus techniques, or to Gobet et al. [45], Lemor et al. [67] for empirical regression methods. In this work, we leave these further questions and we only address the error analysis between (Y, Z) and (Y^N, Z^N).
On the one hand, Zhang [94] proves (in a slightly different form) that the error $\max_{k\leq N} \|Y^N_{t_k} - Y_{t_k}\|_{L_2} \leq CN^{-1/2}$. This is done under rather minimal Lipschitz assumptions on b, σ, f, Φ. On the other hand, when f does not depend on z and the coefficients are smooth, one knows that $|Y^N_0 - Y_0| \leq CN^{-1}$ (see Chevance [21]). We aim at filling the gap regarding these two different rates of convergence. In the following, we prove that

- Chevance's results are extended to the case of f depending also on z.
- the rate N^{-1} holds true also for the difference $|Z^N_0 - Z_0|$.
- more generally, for the other discretization times t_k, we expand the error as

$$\left|Y^N_{t_k} - Y_{t_k} - \alpha_k \cdot (X^N_{t_k} - X_{t_k})\right| \leq CN^{-1} \vee |X^N_{t_k} - X_{t_k}|^2$$

  (for an explicit and bounded random vector α_k).
- an analogous expansion is available for Z.

Since $|X^N_{t_k} - X_{t_k}|^2$ has the same order in L_p than N^{-1}, the error on Y is mainly due to the error $X^N_{t_k} - X_{t_k}$. Thus, Zhang's results are a consequence of this expansion, and Chevance's ones as well since $X^N_0 = X_0$. The gap is filled.
In addition, we learn from this expansion that if one could perfectly simulate X (as for Brownian motion with constant drift, geometric Brownian motion or Ornstein-Uhlenbeck process), the error on the BSDE would be of order N^{-1} and not $N^{-1/2}$ as stated by Zhang's results. Also, if one could use a discretization scheme for X of order 1 for the strong error (for instance Milshtein scheme whenever possible), the error on the BSDE would be of order N^{-1} (we would need to extend our analysis to other discretization schemes, this is straightforward for the Milshtein scheme).

The paper is organized as follows. In Section 2, we define the assumptions on the coefficients, recall the connection between BSDEs and semi-linear PDEs (which is important for our analysis). Finally, we state our main results. Firstly in Theorem 2.3, we extend Zhang's results to L_p norm. Secondly in Theorem 2.4, we expand the error on Y. Lastly in Theorem 2.5, we deal with the error on Z. Naturally, stronger and stronger assumptions are required for theses theorems. Proofs of the three results are postponed to Sections 3, 4 and 5: we combine BSDE techniques, martingale estimates and Malliavin calculus.

Notation.

- **Differentiation.** If $g : \mathbb{R}^d \mapsto \mathbb{R}^q$ is a differentiable function, its gradient $\nabla_x g(x) = (\partial_{x_1} g(x), ..., \partial_{x_d} g(x))$ takes values in $\mathbb{R}^q \otimes \mathbb{R}^d$. At many places, $\nabla_x g(x)$ will simply be denoted $g'(x)$. If $g : \mathbb{R}^d \mapsto \mathbb{R}$ is a twice differentiable function, its Hessian $H_x(g)$ takes values in $\mathbb{R}^d \otimes \mathbb{R}^d$: $(H_x(g))_{i,j} = \partial^2_{x_i x_j} g$. If $g : \mathbb{R}^d \times \mathbb{R}^q \mapsto \mathbb{R}$, $g^{''}_{xy}(x, y)$ takes values in $\mathbb{R}^d \otimes \mathbb{R}^q$: $(g^{''}_{xy})_{ij} = \frac{\partial^2 g}{\partial x_i \partial y_j}$, for $1 \leq i \leq d, 1 \leq j \leq q$.

- **Function spaces.** For an integer $k \geq 1$, we denote by $C_b^{k/2,k,k,k}$ the set of continuously differentiable functions $\phi : (t,x,y,z) \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q \mapsto \phi(t,x,y,z)$ such that the partial derivatives $\partial_t^{l_0}\partial_x^{l_1}\partial_y^{l_2}\partial_z^{l_3}\phi(t,x,y,z)$ exist for $2l_0 + l_1 + l_2 + l_2 \leq k$ and are uniformly bounded. The analogous set of functions that not depend on y and z is denoted by $C_b^{k/2,k}$. This set is denoted by $C_b^{(k+\alpha)/2,k+\alpha}$ ($\alpha \in]0,1[$) if in addition the highest derivatives are Hölder continuous with index α w.r.t. x and $\alpha/2$ w.r.t. t (for a precise definition, see Ladyzenskaja et al. [64]).

- **Norm.** For a d-dimensional vector U, we set $|U|^2 = \sum_{i=1}^d U_i^2$. For a $d \times q$-dimensional matrix A, A_i denotes its i-th column, and A^i its i-th row. Moreover, $|A|^2 = \sum_{i,j=1}^{d,q} A_{i,j}^2$.

- **Constants.** Let C denote a generic constant which may depend on the coefficients b, σ, f, Φ and on the dimensions d and q. We will keep the same notation $K(T)$ for all finite, nonnegative, and nondecreasing functions w.r.t. T: they do not depend on x and h. The generic notation $K(T,x)$ stands for any function bounded by $K(T)(1 + |x|^q)$, for some $q \geq 0$.

- $\boldsymbol{O(U)}$ **and** $\boldsymbol{O_k(h)}$**.** A random vector R is such that $R = O(U)$ for a non-negative random variable U if $|R| \leq K(T,x)U$ (in particular, $R = O(h)$ means $|R| \leq K(T,x)h$). The notation $R = O_k(h^p)$ means $|R| \leq \lambda_k^N h^p$, where λ_k^N is $\mathcal{F}_{t_k}$-measurable, $\sup_N \mathbb{E}(\sup_k |\lambda_k^N|^q) \leq K(T,x)$, for $q \geq 1$.

- $\mathbb{E}_{t_k}$ **and** $\mathbf{Var}_{t_k}$**.** $\mathbb{E}_{t_k}$ is the conditional expectation w.r.t. $\mathcal{F}_{t_k}$ and $\mathrm{Var}_{t_k}(X) = \mathbb{E}_{t_k}(X^2) - (\mathbb{E}_{t_k}(X))^2$.

- **Malliavin calculus**. We use the notations of Nualart [78] for weak spaces $\mathbb{D}^{k,p}$.

- **Discretization** Let $s \in [t_k, t_{k+1}[$. We define $\eta(s) = t_k$.

Chapter 2

Main results

2.1 Hypotheses

The coefficients $b: [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$, $\sigma: [0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times q}$, $f: [0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^q \to \mathbb{R}$ and $\Phi: \mathbb{R}^d \to \mathbb{R}$ satisfy one of the following set of assumptions.

Hypothesis 2.1 *The functions b, σ, f and Φ are bounded in x, are uniformly Lipschitz continuous w.r.t. (x,y,z) and Hölder continuous of parameter $\frac{1}{2}$ w.r.t. t. In addition, Φ is of class $C_b^{2+\alpha}$ for some $\alpha \in]0,1[$ and the matrix-valued function $a = \sigma\sigma^*$ is uniformly elliptic.*

Hypothesis 2.2 *Assume Hypothesis 2.1 and that the functions b, σ are in $C_b^{\frac{3}{2},3}$, f is in $C_b^{\frac{3}{2},3,3,3}$, Φ is in $C_b^{3+\alpha}$ for some $\alpha \in]0,1[$.*

Hypothesis 2.3 *Assume Hypothesis 2.1 and that the functions b, σ are in $C_b^{2,4}$, f is in $C_b^{2,4,4,4}$, Φ is in $C_b^{4+\alpha}$ for some $\alpha \in]0,1[$.*

We do not assert that these smoothness and boundedness conditions are the weakest ones for our error analysis, but they are sufficient. Investigations regarding minimal assumptions would be certainly interesting but it is beyond the scope of the paper.

2.2 Connection between Markovian BSDE's and semi-linear parabolic PDE's

We recall classical results connecting (Y,Z) and the solution and its gradient of the following semi-linear PDE on $[0,T]\times\mathbb{R}^d$:

$$\begin{aligned}&(\partial_t + \mathcal{L}_{(t,x)})u(t,x) + f\left(t,x,u(t,x),\nabla_x u(t,x)\sigma(t,x)\right) = 0, \qquad (2.1)\\ &u(T,x) = \Phi(x),\end{aligned}$$

where $\mathcal{L}_{(t,x)}$ is the second order differential operator

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

(see for instance Ma and Zhang [73] or Pardoux [80]).

Proposition 2.1. *Under Hypothesis 2.1, one has*

$$\forall t \in [0,T],\ Y_t = u(t,X_t),\ Z_t = \nabla_x u(t,X_t)\sigma(t,X_t), \tag{2.2}$$

where u is the unique classic solution $C_b^{1,2}$ of the PDE (2.1).

In addition under Hypothesis 2.2, $u \in C_b^{\frac{3}{2},3}$, and under Hypothesis 2.3, $u \in C_b^{2,4}$.

The first result of this Proposition corresponds to Theorem 2.1 of Delarue and Menozzi [24]. The two last regularity results can be proved in the same way. In fact for this, we would only need b, σ to be in $C_b^{1+\alpha/2,2+\alpha}$; the additional smoothness is used later for Malliavin calculus computations.

2.3 Main results

We now turn to the statement of our results. Remind the following well-known upper bound on the Euler Scheme, which is useful in the sequel.

Proposition 2.2. *Let σ and b be Lipschitz continuous. Then*

$$\forall p \geq 1, [\mathbb{E}(\sup_{t\leq T} |X_t^N - X_t|^p)]^{\frac{1}{p}} \leq K(T,x)\frac{1}{\sqrt{N}}.$$

In fact, for all $p \geq 1$ one has

$$[\mathbb{E}_{t_i}(\sup_{t_i\leq t\leq T} |X_t^N - X_t|^p)]^{\frac{1}{p}} \leq K(T,X_{t_i})\frac{1}{\sqrt{N}} + |X_{t_i}^N - X_{t_i}|. \tag{2.3}$$

Our first result is an extension of the L_2 estimates in Zhang [94] to L_q estimates (see also Gobet et al. [45]).

Theorem 2.3. *Let us assume Hypothesis 2.1. Let $q > 0$. We define the error*

$$e_q(N) = \big[\max_{0\leq k\leq N} \mathbb{E}|Y_{t_k} - Y_{t_k}^N|^q + \mathbb{E}(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} |Z_{t_k}^N - Z_t|^2 dt)^{\frac{q}{2}}\big]^{\frac{1}{q}},$$

where Y^N and Z^N are defined by (1.4) and (1.5). Then $|e_q(N)| \leq K(T,x)\frac{1}{\sqrt{N}}$.

By slightly strengthening the smoothness assumptions on b, σ, f and Φ, we are able to expand the error on Y.

Theorem 2.4. *Let us assume Hypothesis 2.2. Then, the following expansion holds*

$$Y_{t_k}^N - Y_{t_k} = \nabla_x u(t_k, X_{t_k})(X_{t_k}^N - X_{t_k}) + O_k(\frac{1}{N}) + O(|X_{t_k}^N - X_{t_k}|^2).$$

In view of Proposition 2.2, $|X_{t_k}^N - X_{t_k}|^2$ and N^{-1} have the same order (in L_p). Hence it turns out that $\nabla_x u(t_k, X_{t_k})(X_{t_k}^N - X_{t_k})$ is the first order term in the error $Y_{t_k}^N - Y_{t_k}$. Obviously, this estimate implies that of Theorem 2.3. As mentioned in the introduction, the evaluation of Y_0 by Y_0^N has still an accuracy of order N^{-1} since initial conditions for X^N and X coincide. Note that if there is no discretization error for the process X, $Y^N_{t_k} - Y_{t_k} = O(\frac{1}{N})$, a fact which is not clear from equations (1.4) and (1.5). A nice situation corresponds to σ independent of x (this is a very specific situation where Euler and Milshtein schemes are equal): in that case $\|X^N_{t_k} - X_{t_k}\|_{L_p} = O(N^{-1})$ and one gets the order of accuracy N^{-1} for Y.

For Z which plays the role of a gradient relatively to Y, we get an analogous result about the error, up to increasing by 1 the degree of smoothness of the coefficients.

Theorem 2.5. *Let us assume Hypothesis 2.3. Then, the following expansion holds*

$$Z^N_{t_k} - Z_{t_k} = \big(\nabla_x[\nabla_x u\ \sigma]^*(t_k, X_{t_k})(X^N_{t_k} - X_{t_k})\big)^* + O_k(\frac{1}{N}) + O(|X^N_{t_k} - X_{t_k}|^2).$$

Remark 2.6. The above results are sufficient to derive the weak convergence of the renormalized error process $[\sqrt{N}(Y^N_t - Y_t)]_{0\le t\le T}$ and $[\sqrt{N}(Z^N_t - Z_t)]_{0\le t\le T}$, except that one has to define Y^N and Z^N between discretization times. For $t \in [t_k, t_{k+1}[$, analogously to (1.4) and (1.5) we define

$$\begin{aligned} Y^N_t &= \mathbb{E}_t\big(Y^N_{t_{k+1}} + (t_{k+1} - t) f(t, X^N_t, Y^N_{t_{k+1}}, Z^N_t)\big), \\ Z^N_t &= \frac{1}{t_{k+1} - t}\mathbb{E}_t\big(Y^N_{t_{k+1}}(W_{t_{k+1}} - W_t)^*\big). \end{aligned}$$

Theorems 2.4 and 2.5 can be extended to all $t \in [0, T]$. We have

$$\begin{aligned} Y^N_t - Y_t =& \nabla_x u(t, X_t)(X^N_t - X_t) + O_t(\frac{1}{N}) + O(|X^N_t - X_t|^2), \\ Z^N_t - Z_t =& \big(\nabla_x[\nabla_x u\ \sigma]^*(t, X_t)(X^N_t - X_t)\big)^* + O_t(\frac{1}{N}) + O(|X^N_t - X_t|^2). \end{aligned}$$

Theorem 3.5 of Kurtz and Protter [62] allows us to establish the weak convergence of the processes $\sqrt{N}(Y^N - Y)$, and $\sqrt{N}(Z^N - Z)$. Indeed, the process $[\sqrt{N}(X^N_t - X_t)]_{0\le t\le T}$ weakly converges to the solution of

$$\begin{aligned} U_t =& \sum_{i=1}^q \int_0^t \nabla_x \sigma_i(s, X_s) U_s dW^i_s + \int_0^t \nabla_x b(s, X_s) U_s ds \\ &+ \frac{1}{\sqrt{2}} \sum_{i,j=1}^q \int_0^t \sum_{k=1}^d \partial_{x_k} \sigma_i(s, X_s) \sigma_{kj}(s, X_s) dV^{ij}_s, \end{aligned}$$

where $(V^{ij})_{1\le i,j\le q}$ are independent standard Brownian motions and independent of W. Furthermore, the convergence is stable (see Jacod and Protter [53]). Hence, $[\sqrt{N}(X^N_t - X_t), \sqrt{N}(Y^N_t - Y_t), \sqrt{N}(Z^N_t - Z_t), X_t]_{0\le t\le T}$ weakly converges to $[U_t, \nabla_x u(t, X_t)U_t, ([\nabla_x[\nabla_x u\ \sigma]^*(t, X_t)]U_t)^*, X_t]_{0\le t\le T}$.

2.4 Comments

2.4.1 Weak Error

From Theorems 2.4 and 2.5 we can derive estimates related to the weak errors on Y and Z.

Theorem 2.7. *Let ψ be a three times continuously differentiable function with bounded derivatives. Let us assume Hypothesis 2.2. Then, one has*

$$\mathbb{E}(\psi(Y^N_{t_k}) - \psi(Y_{t_k})) = O(\frac{1}{N}).$$

Under Hypothesis 2.3, the same result applies to Z.

Proof : A Taylor expansion of ψ yields

$$\mathbb{E}(\psi(Y^N_{t_k}) - \psi(Y_{t_k})) = \mathbb{E}((Y^N_{t_k} - Y_{t_k})\psi'(Y_{t_k}) + O((Y^N_{t_k} - Y_{t_k})^2).$$

By using Theorem 2.4, we get

$$\begin{aligned}
&\mathbb{E}((Y^N_{t_k} - Y_{t_k})\psi'(Y_{t_k})) \\
=&\mathbb{E}[\psi'(Y_{t_k})\nabla_x u(t_k, X_{t_k})(X^N_{t_k} - X_{t_k}) + O_k(\frac{1}{N}) + O(|X^N_{t_k} - X_{t_k}|^2)], \\
=&\mathbb{E}(\psi'(u(t_k, X_{t_k}))\nabla_x u(t_k, X_{t_k})(X^N_{t_k} - X_{t_k})) + O(\frac{1}{N}).
\end{aligned}$$

Hypotheses on ψ and u enable us to apply Remark 4.4 (see later in section 4) to $\mathbb{E}(\psi'(u(t_k, X_{t_k}))\nabla_x u(t_k, X_{t_k})(X^N_{t_k} - X_{t_k}))$. The result follows. □

Analyzing weak errors on Y and Z is admittedly useful, but studying pathwise estimates can also be relevant. Actually, both estimates are complementary. For instance, practitioners in finance are interested in finding hedging strategies. This corresponds to solving BSDEs, where Y and Z respectively represent the value of the replicating portfolio and the hedging strategy. On the one hand, Theorems 2.4 and 2.5 are suitable tools to study theses quantities for computational issues. On the other hand, Theorem 2.7 enables us to quantify the error on the distribution of the portfolio value, which is relevant in a risk management perspective.

2.4.2 Global error of the numerical resolution of BSDE

As recalled in the introduction, there exist several techniques to numerically solve BSDEs. The one we present here refers to Lemor et al. [67]; it turns to be presumably the most efficient procedure. The authors propose a numerical scheme based on iterative regressions on function bases $p_{0,k}(\cdot), p_{1,k}(\cdot), \cdots, p_{q,k}(\cdot)$ (each being represented as a vector), which coefficients are evaluated using M extra independent simulations of $(X^N_{t_k})_{0\le k\le N-1}$ and of the Brownian increments $(\Delta W_k)_{0\le k\le N-1}$. Let $(y^{N,M}_k(X^N_{t_k}), z^{N,M}_{1,k}(X^N_{t_k}), \cdots, z^{N,M}_{q,k}(X^N_{t_k}))_{0\le k\le N-1}$ denote the approximation of the solution of the discretized BSDE $(Y^N_{t_k}, Z^N_{1,t_k}, \cdots, Z^N_{q,t_k})_{0\le k\le N-1}$ computed in a backward manner with the following algorithm.

- Initialization : for $k = N$ take $y^{N,M}_N(\cdot) = \Phi(\cdot)$.

- Iteration : for $k = N-1, \cdots, 0$, solve the q least-squares problems :

$$\alpha^M_{l,k} = \arg\inf_\alpha \frac{1}{M}\sum_{m=1}^M |y^{N,M}_{k+1}(X^{N,m}_{t_{k+1}})\frac{\Delta W^{l,m}_k}{h} - \alpha \cdot p_{l,k}(X^{N,m}_{t_k})|^2.$$

Then, compute $\alpha_{0,k}^M$ as the minimizer of

$$\frac{1}{M}\sum_{m=1}^{M}|y_{k+1}^{N,M}(X_{t_{k+1}}^{N,m})+hf(t_k,X_{t_k}^{N,m},y_{k+1}^{N,M}(X_{t_{k+1}}^{N,m}),\alpha_{l,k}^M\cdot p_{l,k}(X_{t_k}^{N,m}))-\alpha\cdot p_{0,k}(X_{t_k}^{N,m})|^2.$$

Thus, define $y_k^{N,M}(\cdot)$ and $z_{l,k}^{N,M}(\cdot)$ by

$$y_k^{N,M}(\cdot)=\alpha_{0,k}^M\cdot p_{0,k}(\cdot),\quad z_{l,k}^{N,M}(\cdot)=\alpha_{l,k}^M\cdot p_{l,k}(\cdot).$$

Actually, the true algorithm requires the use of additional truncation operators that we have omitted for sake of simplicity, see Lemor et al. [67] for details. The following error on the unknown regression functions $(y_k^{N,M},z_{l,k}^{N,M})_{1\le l\le q,0\le k\le N-1}$

$$\max_{0\le k\le N}\mathbb{E}|Y_{t_k}^N-y_k^{N,M}(X_{t_k}^N)|^2+h\mathbb{E}\sum_{k=0}^{N-1}|Z_{t_k}^N-Z_k^{N,M}(X_{t_k}^N)|^2$$

is essentially bounded by $NC_{M,p}$, where M and p respectively denote the number of simulated paths and the set of functions. For suitable choices of M and p, $C_{M,p}$ goes to 0 at a given rate. This result allows them to optimally tune the parameters to ensure a given accuracy. Hence, summing this numerical error and the discretization's one given by Theorems 2.4 and 2.5 leads to the global error. For example, assume that $\|X_{t_k}^N-X_{t_k}\|_{L_p}=O(\frac{1}{N})$. Then, from Theorem 2.4, we get $\mathbb{E}|Y_{t_k}-y_k^{N,M}(X_{t_k}^N)|^2\le C(\frac{1}{N^2}+NC_{M,p})$.

2.5 Numerical Experiments

In this part, we draw some graphs to illustrate the results given by Theorems 2.3 and 2.4. To do so, one needs to explicitly know X and Y. Let us consider a Call option pricing problem. We assume that X follows the Black-Scholes model in dimension $d=1$, $\frac{dX_t}{X_t}=\mu dt+\sigma dW_t$, with $\sigma=0.2,\mu=0.1$ and $X_0=100$. The driver f is defined by $f(t,x,y,z)=-ry-\theta z$, where $\theta=\frac{\mu-r}{\sigma}$ and $r=0.02$. The terminal condition $\Phi(x)$ is given by $(x-K)_+$, where $K=100$. The maturity of the option is $T=1$. The continuous backward equation can be solved, Y_t is the price of a standard Call option (see El Karoui et al. [27] for a detailed computation).

We compute X^N and Y^N by using (1.3) and (1.4) and get

$$\begin{aligned}X_{t_k}^N=&X_0\Pi_{j=0}^{k-1}(1+\mu h+\sigma\Delta W_j),\\ Y_{t_k}^N=&\mathbb{E}_{t_k}[\Phi(X_T^N)\Pi_{j=k}^{N-1}(1-rh-\theta\Delta W_j)].\end{aligned}$$

Figure 2.1 refers to Theorem 2.3. We plot the evolution of the logarithm of $e_2(N),e_3(N)$ and $e_4(N)$ w.r.t. $\log(N)$.

We use 1000 simulations to approximate the L_p-norm $e_p(N)$ and to compute each conditional expectation $Y_{t_k}^N$, we use 1000 Monte Carlo simulations. We compute $\log(e_p(N))_{p=2,3,4}$ for $N=2^j,j=1,\cdots,7$. Looking at the graph, we see that the evolutions of $\log(e_p(N))_{p=2,3,4}$ w.r.t. $\log(N)$ are almost linear. In view of Theorem 2.3, the slope should be of order $-\frac{1}{2}$. By using a linear regression method, we get the parameters

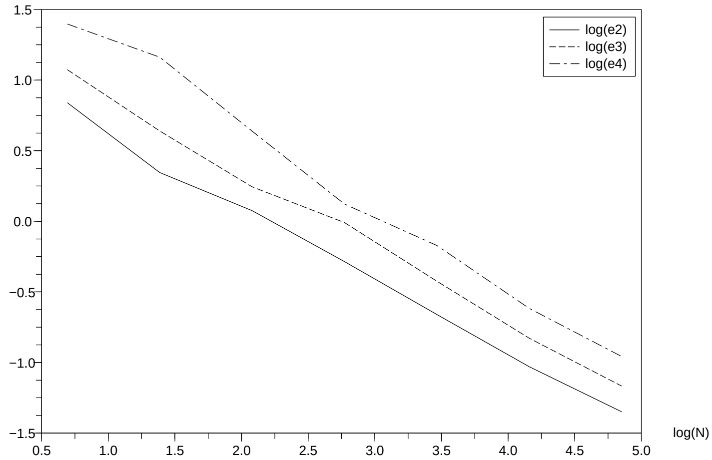

Figure 2.1: Evolution of $e_2(N), e_3(N), e_4(N)$ w.r.t. $\log(N)$

	a	b	std
L_2 error	- 0.5179119	1.14106	0.0384534
L_3 error	- 0.5321072	1.4078415	0.0367535
L_4 error	- 0.5891505	1.858531	0.0662573

Table 2.1: Coefficients of the linear regression of $\log(e_p(N))$, $p = 2, 3, 4$ w.r.t. $\log(N)$.

a, b, std where $\log(e_p(N)) = a * \log(N) + b$, and std represents the standard deviation of the residuals. Table 2.1 sums up the values of a, b, std for $p = 2, 3, 4$. Clearly a is of order $-\frac{1}{2}$.

Figure 2.2 refers to Theorem 2.4. We plot the evolution of $\log(a(N))$, where $a(N) = (\mathbb{E}[|Y^N_{t_k} - Y_{t_k} - \nabla_x u(t_k, X_{t_k})(X^N_{t_k} - X_{t_k})|^2])^{\frac{1}{2}}$ w.r.t. $\log(N)$, at time $t_k = \frac{T}{2}$. We use 100 simulations to approximate the L_2-norm and to compute each $Y^N_{t_k}$ we use 10^6 Monte Carlo simulations. N behaves as $2^j, j = 1, \cdots, 7$. We note that $\log(a(N))$ actually evolves almost linearly w.r.t $\log(N)$. Regarding Theorem 2.4, the slope should be of order -1. If we still use a linear regression,we get the slope $a = -0.9123248$, $b = -1.0172153$ and the standard deviation of the residuals equals 0.0940069.

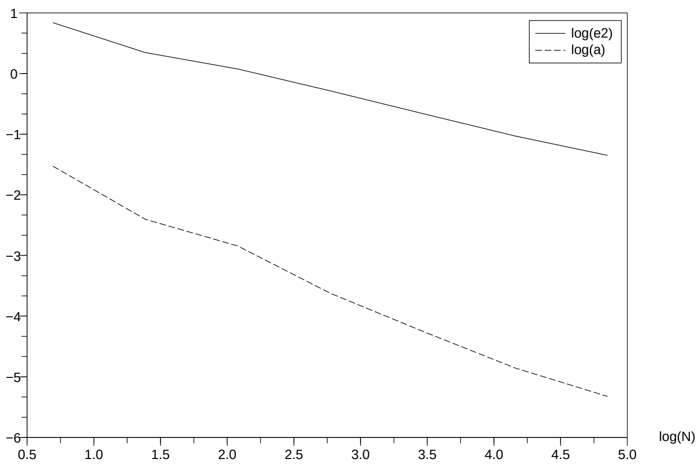

Figure 2.2: Evolution of $\log(a(N))$ and $\log(e_2(N))$ w.r.t. $\log(N)$.

Chapter 3

Proof of theorem 2.3

Extra notations for all the proofs. For any process U (except the Brownian increments ΔW_k), we define $\Delta U_k = U^N_{t_k} - U_{t_k}$. Let θ_s denote (s, X_s, Y_s, Z_s) and $f^N_{t_k}$ denote $f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k})$.
$\overline{Z}_{t_k}$ is defined as $h\overline{Z}_{t_k} := \mathbb{E}_{t_k} \int_{t_k}^{t_{k+1}} Z_s ds$ and we put $\Delta \overline{Z}_k = Z^N_{t_k} - \overline{Z}_{t_k}$.

If $q = 2$, the result has already been proved in Gobet et al. [45], under Lipschitz conditions on b, σ, f, Φ. Thanks to the inequality $\mathbb{E}|U|^q \leq (\mathbb{E}|U|^{2p})^{\frac{q}{2p}}$ for $2p \geq q$, we only need to prove the theorem for $q = 2p$, where $p \in \mathbb{N}^*$.
First, we give some estimates which can be easily established. We have, under Hypothesis 2.1, $\forall s \in [t_k, t_{k+1}]$,

$$\mathbb{E}_{t_k}(|X_s - X_{t_k}|^{2p} + |Y_s - Y_{t_k}|^{2p} + |Z_s - \overline{Z}_{t_k}|^{2p}) \leq Ch^p. \tag{3.1}$$

In the following computations, these estimates are repeatedly used.

3.1 Proof of $\max_{0\leq k\leq N} \mathbb{E}|Y_{t_k} - Y^N_{t_k}|^{2p} = O(h^p)$.

We prove the following result, which is a bit more general.

Proposition 3.1. $\max_{i\leq k\leq N} \mathbb{E}_{t_i}|Y_{t_k} - Y^N_{t_k}|^{2p} = O_i(h^p) + |\Delta X_i|^{2p}$.

By taking $i = 0$, we get $\max_{0\leq k\leq N} \mathbb{E}|Y_{t_k} - Y^N_{t_k}|^{2p} = O(h^p)$.
Assume that we have

$$|\Delta Y_k|^2 \leq (1 + Ch)\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2 + Ch|\Delta X_k|^2 + Ch^2. \tag{3.2}$$

Then, using the inequality $(a + b)^p \leq a^p(1 + \epsilon(2^{p-1} - 1)) + b^p(1 + \frac{2^{p-1}-1}{\epsilon^{p-1}})$ for $0 < \epsilon < 1$, we deduce

$$|\Delta Y_k|^{2p} \leq (1 + Ch)^{p+1}\mathbb{E}_{t_k}|\Delta Y_{k+1}|^{2p} + C^p h^p(|\Delta X_k|^2 + Ch)^p(1 + \frac{C}{h^{p-1}}).$$

Take the conditional expectation w.r.t. $\mathcal{F}_{t_i}$ to get $\mathbb{E}_{t_i}|\Delta Y_k|^{2p} \leq (1 + Ch)\mathbb{E}_{t_i}|\Delta Y_{k+1}|^{2p} + h(h^p + \mathbb{E}_{t_i}|\Delta X_k|^{2p})$. Using (2.3) for $|\Delta X_k|$ and Gronwall's lemma yields $\max_{i\leq k\leq N} \mathbb{E}_{t_i}|Y_{t_k} - Y^N_{t_k}|^{2p} = O_i(h^p) + |\Delta X_i|^{2p}$. □
Now we prove the inequality (3.2)**.** From (1.2) and (1.4) we obtain

$$\Delta Y_k = \mathbb{E}_{t_k}(\Delta Y_{k+1}) + \mathbb{E}_{t_k} \int_{t_k}^{t_{k+1}} (f^N_{t_k} - f(\theta_s))ds. \tag{3.3}$$

By applying Young's inequality, that is $(a+b)^2 \leq (1+\gamma h)a^2 + (1+\frac{1}{\gamma h})b^2$, where γ will be fixed later, and using the Lipschitz property of f, we get

$$|\Delta Y_k|^2 \leq (1+\gamma h)(\mathbb{E}_{t_k}(\Delta Y_{k+1}))^2 + C(h+\frac{1}{\gamma})[h^2 + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |X_s - X^N_{t_k}|^2 ds]$$
$$+ C(h+\frac{1}{\gamma})[\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Y_s - Y^N_{t_{k+1}}|^2 ds + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Z_s - Z^N_{t_k}|^2 ds]. \tag{3.4}$$

Let us introduce $\overline{Z}_{t_k}$ (see extra notations at the beginning of Section 3):

$$\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Z_s - Z^N_{t_k}|^2 ds = \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Z_s - \overline{Z}_{t_k}|^2 ds + h\mathbb{E}_{t_k}|\overline{Z}_{t_k} - Z^N_{t_k}|^2. \tag{3.5}$$

Thanks to the Cauchy Schwarz inequality we have

$$|\mathbb{E}_{t_k}(\Delta Y_{k+1}\Delta W^l_k)|^2 \leq h\{\mathbb{E}_{t_k}(|\Delta Y_{k+1}|^2) - |\mathbb{E}_{t_k}(\Delta Y_{k+1})|^2\}.$$

Hence, as $h\overline{Z}_{t_k} = \mathbb{E}_{t_k}(\{Y_{t_{k+1}} + \int_{t_k}^{t_{k+1}} f(\theta_s)ds\}\Delta W^*_k)$, with a bounded f, it follows that

$$h^2|\overline{Z}_{t_k} - Z^N_{t_k}|^2 \leq d\ h\left(\mathbb{E}_{t_k}(|\Delta Y_{k+1}|^2) - |\mathbb{E}_{t_k}(\Delta Y_{k+1})|^2\right) + Ch^3. \tag{3.6}$$

By plugging (3.5) and (3.6) into (3.4), we get

$$\begin{aligned}|\Delta Y_k|^2 \leq\ & (1+\gamma h)(\mathbb{E}_{t_k}(\Delta Y_{k+1}))^2 \\ & + C(h+\frac{1}{\gamma})[h^2 + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |X_s - X^N_{t_k}|^2 ds + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Y_s - Y^N_{t_{k+1}}|^2 ds] \\ & + C(h+\frac{1}{\gamma})[\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Z_s - \overline{Z}_{t_k}|^2 ds + \mathbb{E}_{t_k}(|\Delta Y_{k+1}|^2) - |\mathbb{E}_{t_k}(\Delta Y_{k+1})|^2].\end{aligned}$$

We can write $\mathbb{E}_{t_k}|Y_s - Y^N_{t_{k+1}}|^2 \leq 2\mathbb{E}_{t_k}|Y_s - Y_{t_{k+1}}|^2 + 2\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2$. By doing the same for $X_s - X^N_{t_{k+1}}$, and taking $\gamma = C$, we obtain

$$\begin{aligned}|\Delta Y_k|^2 \leq & (1+Ch)\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2 + Ch|\Delta X_k|^2 + Ch\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2 \\ & + C[h^2 + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |X_s - X_{t_k}|^2 ds + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Y_s - Y_{t_{k+1}}|^2 ds] \\ & + C[\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} |Z_s - \overline{Z}_{t_k}|^2 ds].\end{aligned}$$

Using (3.1) yields $|\Delta Y_k|^2 \leq (1+Ch)\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2 + Ch|\Delta X_k|^2 + Ch^2$. □

3.2 Proof of $\mathbb{E}(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} |Z^N_{t_k} - Z_t|^2 dt)^{\frac{p}{2}} = O(h^p)$.

First of all, we can split this summation into two terms

$$\mathbb{E}(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} |Z^N_{t_k} - Z_t|^2 dt)^p \leq C\mathbb{E}(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} |\overline{Z}_{t_k} - Z_t|^2 dt)^p + C\mathbb{E}\big(h\sum_{k=0}^{N-1}|\Delta\overline{Z}_k|^2\big)^p.$$

Thanks to (3.1), we have $\mathbb{E}\big(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} |\overline{Z}_{t_k} - Z_t|^2 dt\big)^p \leq T^{p-1}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} \mathbb{E}|\overline{Z}_{t_k} - Z_t|^{2p} dt = O(h^p)$.

Scheme of the proof of $\mathbb{E}\big(h\sum_{k=0}^{N-1}|\Delta\overline{Z}_k|^2\big)^p = O(h^p)$*.* The first key point is to slice the summation into small intervals and show that the result is true for small time intervals. The second key point is to use Rosenthal's inequality, see Theorem 2.12 page 23 of Hall and Heyde [48]. By using (3.6) and taking the expectation, we can write :

$$\mathbb{E}\big(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2\big)^p \le C\mathbb{E}\big(\sum_{k=0}^{k_1}\mathrm{Var}_{t_k}\Delta Y_{k+1}\big)^p + Ch^p. \tag{3.7}$$

We use Rosenthal's inequality to upper bound

$$\begin{aligned}\mathbb{E}\big(\sum_{k=0}^{k_1}\mathrm{Var}_{t_k}\Delta Y_{k+1}\big)^p &\le C\mathbb{E}\big(\sum_{k=0}^{k_1}\Delta Y_{k+1} - \mathbb{E}_{t_k}\Delta Y_{k+1}\big)^{2p},\\ &\le C3^{2p-1}\big[\mathbb{E}\Delta Y_{k_1+1}^{2p} + \mathbb{E}\Delta Y_0^{2p} + \mathbb{E}\big(\sum_{k=0}^{k_1}(\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1})\big)^{2p}\big].\end{aligned}$$

By plugging this inequality into (3.7) and using the previous estimate on $|\Delta Y_k|$, we get

$$\mathbb{E}\big(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2\big)^p \le O(h^p) + C\mathbb{E}\big(\sum_{k=0}^{k_1}(\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1})\big)^{2p}. \tag{3.8}$$

We now tackle the term $\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1}$. Using (3.3), we have $\sum_{k=0}^{k_1}(\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1}) = \sum_{k=0}^{k_1}\int_{t_k}^{t_{k+1}}(\mathbb{E}_{t_k}(f_{t_k}^N - f(\theta_s)))\mathrm{ds}$. By doing the same kind of proof as before, that is using the fact that f is Lipschitz and the results on $\mathbb{E}|\Delta X_k|^{2p}$ and $\mathbb{E}|\Delta Y_k|^{2p}$, we find

$$\mathbb{E}\big(\sum_{k=0}^{k_1}(\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1})\big)^{2p} \le O(h^p) + C(hk_1)^p\mathbb{E}\big(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2\big)^p.$$

By plugging this term back into (3.8), we can write $\big(1 - C(hk_1)^p\big)\mathbb{E}\big(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2\big)^p = O(h^p)$. Consequently, if we choose $k_1 \le \frac{1}{(2C)^{\frac{1}{p}}h}$ we come up with $\mathbb{E}\big(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2\big)^p = O(h^p)$. This result can be extended to any summation involving at most Δk terms, where $\Delta k \le \frac{1}{(2C)^{\frac{1}{p}}h}$. We can cover the interval $\{0,\cdots,N-1\}$ with a finite number of elementary intervals of size Δk and we get $\mathbb{E}\big(h\sum_{k=0}^{N-1}|\Delta\overline{Z}_k|^2\big)^p = O(h^p)$, which completes our proof. $\square$

From this result and (3.1), we also deduce

$$\mathbb{E}\big(h\sum_{k=0}^{N-1}|\Delta Z_k|^2\big)^p = O(h^p), \tag{3.9}$$

which is very useful in the following.

Chapter 4

Proof of Theorem 2.4.

To expand the error, we use usual techniques of stochastic analysis, combining martingale estimates and Malliavin calculus tools.

4.1 Preliminary estimates

Sections 4 and 5 contain proofs with similar calculations, which are quite technical. In order to be as clear as possible, we state two results really useful in the sequel, which are related to Malliavin calculus (see Nualart [78]). The results give sufficient conditions for expectations and conditional expectations to be small w.r.t. the time step h. They are based on ideas from Kohatsu-Higa and Pettersson [59] and Gobet and Munos [44].

Proposition 4.1. *Let $F \in \Delta^{1,2}$ with $\mathbb{E}_{t_k}|F|^2 + \sup_{t_k \leq s \leq T} \mathbb{E}_{t_k}|D_s F|^2 < \infty$ and let U be an Itô process of the form $U_t = U_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s dW_s$, with $\sup_{t_k \leq s \leq T} \mathbb{E}_{t_k}|\alpha_s|^2 + \sup_{t_k \leq s \leq T} \mathbb{E}_{t_k}|\beta_s|^2 < \infty$. Then, $\forall (t,t')$ such that $t_k \leq t \leq t' \leq t_{k+1}$,*

$$|\mathbb{E}_{t_k}[F(U_t - U_{t'})]| \leq (t'-t)\big[(\mathbb{E}_{t_k}|F|^2)^{\frac{1}{2}}(\sup_{t \leq s \leq t'} \mathbb{E}_{t_k}|\alpha_s|^2)^{\frac{1}{2}} + (\sup_{t \leq s \leq t'} \mathbb{E}_{t_k}|D_s F|^2)^{\frac{1}{2}}(\sup_{t \leq s \leq t'} \mathbb{E}_{t_k}|\beta_s|^2)^{\frac{1}{2}}\big].$$

This proposition can be easily proved. Assume without loss of generality that F and U are one-dimensional. From the duality formula, we have $\mathbb{E}_{t_k}[F(\int_t^{t'} \alpha_s ds + \int_t^{t'} \beta_s dW_s)] = \mathbb{E}_{t_k}[\int_t^{t'}(F\alpha_s + D_s F \cdot \beta_s)ds]$. Thanks to Cauchy Schwarz inequality and hypotheses on α and β, we get the result.

Definition 4.2. F satisfies the condition R_k if $F \in \mathbb{D}^{k,\infty}$ and if $\mathcal{C}_{k,p}(F) := \|F\|_{L_p} + \sum_{j \leq k} \sup_{0 \leq s_1,\ldots,s_j \leq T} \|D_{s_1,..,s_j} F\|_{L_p} < \infty$.

Proposition 4.3. *Let F satisfy the condition R_2. For simplicity we set $dW_s^0 = ds$. Assume that $U_t \in \mathbb{R}^d$ satisfies the following stochastic expansion property*

$$U_t = \sum_{i,j=0}^{q} c_{i,j}^{U,0}(t) \int_0^t c_{i,j}^{U,1}(s)\Big(\int_{\eta(s)}^s c_{i,j}^{U,2}(r) dW_r^i\Big) dW_s^j, \qquad (\mathcal{P})$$

where $\{(c_{i,j}^{U,i_1}(t))_{t \geq 0} : 0 \leq i,j \leq q, 0 \leq i_1 \leq 2\}$ are adapted processes satisfying

- *$\forall (i,j), 1 \leq i,j \leq q$, $\forall t \in [0,T]$, $c^{U,0}_{i,j}(t)$ satisfies R_2, and $\mathcal{C}^U_{2,p} := \sup_{0\leq t\leq T}\sup_{1\leq i,j\leq q}\mathcal{C}_{2,p}(c^{U,0}_{i,j}(t)) < \infty$, $p \geq 1$.*

- *$\forall (i,j), 1 \leq i,j \leq q$, $\forall t \in [0,T]$, $c^{U,1}_{i,j}(t), c^{U,0}_{0,j}(t), c^{U,0}_{i,0}(t), c^{U,1}_{i,0}(t)$ satisfy R_1, and $\mathcal{C}^U_{1,p} := \sup_{0\leq t\leq T}\sup_{1\leq i,j\leq q}\{\mathcal{C}_{1,p}(c^{U,1}_{i,j}(t)) + \mathcal{C}_{1,p}(c^{U,0}_{0,j}(t)) + \mathcal{C}_{1,p}(c^{U,0}_{i,0}(t)) + \mathcal{C}_{1,p}(c^{U,1}_{i,0}(t))\}$ $< \infty$, $p \geq 1$.*

- *$\forall (i,j), 0 \leq i,j \leq q$, $\forall t \in [0,T]$, $c^{U,2}_{i,j}(t), c^{U,1}_{0,j}(t), c^{U,0}_{0,0}(t)$ satisfy R_0, and $\mathcal{C}^U_{0,p} := \sup_{0\leq t\leq T}\sup_{0\leq i,j\leq q}\{\mathcal{C}_{0,p}(c^{U,2}_{i,j}(t)) + \mathcal{C}_{0,p}(c^{U,1}_{0,j}(t)) + \mathcal{C}_{0,p}(c^{U,0}_{0,0}(t))\} < \infty$, $p \geq 1$.*

Thus, there is a constant $K(T)$ which depends polynomially on $\mathcal{C}_{2,p}(F), \mathcal{C}^U_{2,p}, \mathcal{C}^U_{1,p}, \mathcal{C}^U_{0,p}$ (for some $p \geq 1$) such that $|\mathbb{E}[FU_t]| \leq K(T)h$.

Indeed, we have

$$\begin{aligned}
\mathbb{E}(FU_t) &= \sum_{i,j=0}^{q} \mathbb{E}\big(Fc^{U,0}_{i,j}(t)\int_0^t c^{U,1}_{i,j}(s)(\int_{\eta(s)}^{s} c^{U,2}_{i,j}(r)dW^i_r)dW^j_s\big) \\
&= \sum_{i,j=1}^{q}\int_0^t\int_{\eta(s)}^{s} \mathbb{E}\big(D^i_r\big[D^j_s\{Fc^{U,0}_{i,j}(t)\}c^{U,1}_{i,j}(s)\big]c^{U,2}_{i,j}(r)\big)\mathrm{dr\ ds} \\
&\quad + \sum_{j=1}^{q}\int_0^t\int_{\eta(s)}^{s} \mathbb{E}\big[D^j_s\{Fc^{U,0}_{0,j}(t)\}c^{U,1}_{0,j}(s)c^{U,2}_{0,j}(r)\big]\mathrm{dr\ ds} \\
&\quad + \sum_{i=1}^{q}\int_0^t\int_{\eta(s)}^{s} \mathbb{E}\big[D^i_r\{Fc^{U,0}_{i,0}(t)c^{U,1}_{i,0}(s)\}c^{U,2}_{i,0}(r)\big]\mathrm{dr\ ds} \\
&\quad + \int_0^t\int_{\eta(s)}^{s} \mathbb{E}\big(Fc^{U,0}_{0,0}(t)c^{U,1}_{0,0}(s)c^{U,2}_{0,0}(r)\big)\mathrm{dr\ ds}.
\end{aligned}$$

Then, the result readily follows.

Remark 4.4. Under Hypothesis 2.2, we can show (see later the proof of (5.9)) that for each t, $X^N_t - X_t$ satisfies the expansion $\mathcal{P}$. Hence, if F satisfies R_2, Proposition 4.3 yields

$$|\mathbb{E}[F(X^N_t - X_t)]| = O(h)$$

uniformly in $t \in [0,T]$, which is a very useful result for the sequel.

4.2 Expansion of $Y^N_{t_k} - Y_{t_k}$

In the following, we assume that Hypothesis 2.2 is in force. This implies in particular that u is bounded, of class $C^{3/2,3}_b$ (see Theorem 2.1). We also easily prove that $\forall p \geq$

$1, \forall k \in \{0, \cdots, N-1\}$ (see Nualart [78] e.g.)

- $\mathbb{E}_{t_k}(\sup_{t_k \le t \le T} |X_t|^{2p}) < K(T)(1 + |X_{t_k}|^{2p}), \ \sup_{t_k \le s \le T} \mathbb{E}_{t_k}(\sup_{t_k \le t \le T} |D_s X_t|^p) \le C,$

$$\sup_{t_k \le s,r \le T} \mathbb{E}_{t_k}(\sup_{t_k \le t \le T} |D_r D_s X_t|^p) + \sup_{t_k \le s,r,v \le T} \mathbb{E}_{t_k}(\sup_{t_k \le t \le T} |D_v D_r D_s X_t|^p) \le C, \tag{4.1}$$

- $\mathbb{E}_{t_k}(\sup_{t_k \le t \le T} |X^N_t|^{2p}) < K(T)(1 + |X^N_{t_k}|^{2p}), \ \sup_{N, t_k \le s \le T} \mathbb{E}_{t_k}(\sup_{t_k \le t \le T} |D_s X^N_t|^p) \le C,$

$$\sup_{N, t_k \le s,r \le T} \mathbb{E}_{t_k}(\sup_{t_k \le t \le T} |D_r D_s X^N_t|^p) + \sup_{N, t_k \le s,r,v \le T} \mathbb{E}_{t_k}(\sup_{t_k \le t \le T} |D_v D_r D_s X^N_t|^p) \le C. \tag{4.2}$$

Due to the Markov property of $(X^N_{t_k})_k$, one has $Y^N_{t_k} = u^N(t_k, X^N_{t_k})$ for some Lipschitz function $u^N(t_k, \cdot)$ (see Gobet et al. [45]) with an obvious definition of u^N. Actually, under our assumptions, this function is even three times differentiable w.r.t. x. Thus, the difference ΔY_k can be written as follows:

$$\Delta Y_k = (u^N(t_k, X^N_{t_k}) - u(t_k, X^N_{t_k})) + (u(t_k, X^N_{t_k}) - u(t_k, X_{t_k})).$$

Since u is of class $C^{3/2,3}_b$, the last term of the previous inequality becomes

$$u(t_k, X^N_{t_k}) - u(t_k, X_{t_k}) = \nabla_x u(t_k, X_{t_k}) \Delta X_k + O(|\Delta X_k|^2). \tag{4.3}$$

To complete the proof, we apply the following lemma

Lemma 4.5. *Under Hypothesis 2.2,* $|u^N(t_k, x) - u(t_k, x)| \le K(T, x) h$.

The result above is new but not so surprising. Indeed, if f is identically zero, the difference is only related to the weak approximation of $\Phi(X_T)$ by $\Phi(X^N_T)$: from Bally and Talay [7], one knows that this is of order h.

The rest of this section is devoted to the proof of the lemma. We only give the proof for $t_k = 0$. We want to find an upper bound for $|u^N(0, x) - u(0, x)| = |\Delta Y_0|$.

For the sake of clarity, we split the proof into several steps.

Step 1 : linearization of the error. We show that

$$\Delta Y_k = \mathbb{E}_{t_k}(\Delta Y_{k+1} \xi_k + h f'_x(\theta_{t_k}) \Delta X_k + h \chi_k), \tag{4.4}$$

with

$$\xi_k = (1 + h f'_y(\theta_{t_k}) + f'_z(\theta_{t_k}) \Delta W_k), \tag{4.5}$$

$$\begin{aligned} \chi_k = &\int_{t_k}^{t_{k+1}} (G_0(s, X_s) + f'_y(\theta_{t_k}) G_y(s, X_s) + f'_z(\theta_{t_k}) G_z(s, X_s)) ds \\ &+ \int_0^1 (1-\lambda) \big[\Delta X^*_k f''_{xx}(\theta^\lambda_{t_k}) \Delta X_k + f''_{yy}(\theta^\lambda_{t_k})(Y^N_{t_{k+1}} - Y_{t_k})^2 + \Delta Z_k f''_{zz}(\theta^\lambda_{t_k}) \Delta Z^*_k \\ &+ 2 \Delta X^*_k f''_{xy}(\theta^\lambda_{t_k})(Y^N_{t_{k+1}} - Y_{t_k}) + 2 \Delta X^*_k f''_{xz}(\theta^\lambda_{t_k}) \Delta Z^*_k + 2(Y^N_{t_{k+1}} - Y_{t_k}) f''_{yz}(\theta^\lambda_{t_k}) \Delta Z^*_k \big] d\lambda, \end{aligned} \tag{4.6}$$

where $\theta^\lambda_{t_k} = \lambda(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k}) + (1-\lambda)\theta_{t_k}$ and G_0, G_y, G_z are bounded functions. ¿From (3.3) and by introducing $f(\theta_{t_k})$, we have

$$\Delta Y_k = \mathbb{E}_{t_k}\big(\Delta Y_{k+1} + h(f^N_{t_k} - f(\theta_{t_k})) + \int_{t_k}^{t_{k+1}} (f(\theta_{t_k}) - f(\theta_s)) \mathrm{d}s\big). \tag{4.7}$$

By applying Itô's formula to $f(\theta_u)$ between t_k and s we show that, under Hypothesis 2.2, $\int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k}(f(\theta_{t_k}) - f(\theta_s))ds = h\int_{t_k}^{t_{k+1}} E_{t_k}(G_0(s,X_s))ds$, where G_0 is a bounded function. In the second term, perform a second order expansion of f around θ_{t_k} to get

$$\begin{aligned}
f^N_{t_k} - f(\theta_{t_k}) &= f'_x(\theta_{t_k})\Delta X_k + f'_y(\theta_{t_k})\Delta Y_{k+1} + f'_z(\theta_{t_k})\Delta Z^*_k + f'_y(\theta_{t_k})(Y_{t_{k+1}} - Y_{t_k}) \\
&+ \int_0^1 (1-\lambda)\big[\Delta X^*_k f''_{xx}(\theta^\lambda_{t_k})\Delta X_k + f''_{yy}(\theta^\lambda_{t_k})(Y^N_{t_{k+1}} - Y_{t_k})^2 + \Delta Z_k f''_{zz}(\theta^\lambda_{t_k})\Delta Z^*_k \\
&+ 2\Delta X^*_k f''_{xy}(\theta^\lambda_{t_k})(Y^N_{t_{k+1}} - Y_{t_k}) + 2\Delta X^*_k f''_{xz}(\theta^\lambda_{t_k})\Delta Z^*_k + 2(Y^N_{t_{k+1}} - Y_{t_k}) f''_{yz}(\theta^\lambda_{t_k})\Delta Z^*_k\big]d\lambda.
\end{aligned} \tag{4.8}$$

Note that $\mathbb{E}_{t_k}(Y_{t_{k+1}} - Y_{t_k}) = \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} G_y(s,X_s)ds$. If we closely look at (4.8), we can see that we need to develop ΔZ_k. By using (1.5), we can write

$$Z^N_{t_k} = \frac{1}{h}\mathbb{E}_{t_k}(\Delta Y_{k+1}\Delta W^*_k) + \frac{1}{h}\mathbb{E}_{t_k}(u(t_{k+1}, X_{t_{k+1}})\Delta W^*_k).$$

Introducing the weak derivative of $X_{t_{k+1}}$ (see Nualart [78] p.109), the second term of this summation equals $\frac{1}{h}\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} \nabla_x u(t_{k+1}, X_{t_{k+1}}) D_t X_{t_{k+1}}\mathrm{d}t$, where $D_t X_{t_{k+1}} = \nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1}\sigma(t,X_t)$. Since $Z_{t_k} = \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k})$, one gets

$$\begin{aligned}
\Delta Z_k &= \frac{1}{h}\mathbb{E}_{t_k}(\Delta Y_{k+1}\Delta W^*_k) \\
&+ \frac{1}{h}\int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k}\big(\nabla_x u(t_{k+1}, X_{t_{k+1}})\nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1}\sigma(t,X_t) - \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k})\big)\mathrm{d}t.
\end{aligned}$$

The term in the second conditional expectation is equal to $\nabla_x u(t_{k+1}, X_{t_{k+1}})$ $\nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1}\sigma(t,X_t) \pm \nabla_x u(t,X_t)\sigma(t,X_t) - \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k})$: hence, two applications of Itô's formula (for the first contribution between t and t_{k+1}, for the second one between t_k and t) prove that

$$\Delta Z^*_k = \int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k}(G_z(s,X_s))\mathrm{d}s + \frac{1}{h}\mathbb{E}_{t_k}(\Delta Y_{k+1}\Delta W_k), \tag{4.9}$$

for a bounded function G_z. Plugging this equality and (4.8) into (4.7) yields (4.4).

Step 2 : another formula of ΔY_0. First of all, we replace $Y^N_{t_{k+1}} - Y_{t_k}$ by $\Delta Y_{k+1} + Y_{t_{k+1}} - Y_{t_k}$ in the expression of χ_k. Then, easy computations combining Proposition 3.1 and estimates (3.1) show that

$$\tilde{\chi}_k = \mathbb{E}_{t_k}(\chi_k) = O_k(h) + O(|\Delta X_k|^2 + |\Delta Z_k|^2). \tag{4.10}$$

From (4.4), we deduce the following equality

$$\Delta Y_0 = \mathbb{E}(\Delta Y_N \xi_0\cdots\xi_{N-1} + h\sum_{i=0}^{N-1}(f'_x(\theta_{t_i})\Delta X_i + \tilde{\chi}_i)\xi_0\cdots\xi_{i-1}). \tag{4.11}$$

Now it is enough to show that all terms of this summation are $O(h)$. In the following, $\eta_0 = 1$ and $\eta_i = \xi_0\cdots\xi_{i-1}$ for $i \le N$.

Step 3 : some results on $\eta_N = \xi_0\cdots\xi_{N-1}$.
We establish the following results on η_N:

$$\begin{aligned}
&\eta_k \text{ satisfies the condition } R_2 \text{ uniformly in } k, \text{ i.e. } \forall k, \eta_k \in \mathbb{D}^{2,\infty} \\
&\text{and } \max_{k\le N}\mathcal{C}_{2,p}(\eta_k) < \infty, \forall p \ge 1,
\end{aligned} \tag{4.12}$$

$$\mathbb{E}(\max_{0\le k\le N}|\eta_k|^p) + \sup_{r\le T}\mathbb{E}(\max_{0\le k\le N}|D_r\eta_k|^p) + \sup_{r,s\le T}\mathbb{E}(\max_{0\le k\le N}|D_rD_s\eta_k|^p) < \infty. \tag{4.13}$$

Proof of (4.12). We have $\eta_0 = 1$, and for $i \geq 1$

$$\eta_i = \eta_{i-1}(1 + hf'_y(\theta_{t_{i-1}}) + f'_z(\theta_{t_{i-1}})\Delta W_{i-1}). \tag{4.14}$$

We begin to show that $\max_{k\leq N} \|\eta_k\|_{L_p} = O(1)$ for $p \geq 1$. Since f'_y and f'_z are bounded, we easily prove that $\mathbb{E}_{t_{i-1}}(1 + hf'_y(\theta_{t_{i-1}}) + f'_z(\theta_{t_{i-1}})\Delta W_{i-1})^{2p} \leq (1 + Ch)$, whence $\mathbb{E}|\eta_i|^{2p} \leq (1 + Ch)\mathbb{E}|\eta_{i-1}|^{2p}$. We deduce that $\max_{k\leq N} \|\eta_k\|_{L_p} = O(1)$.

Now, let us show that $\max_{k\leq N} \mathbb{E}|D_r\eta_k|^p = O(1)$, uniformly in r. Let r be such that $t_{k-1} < r \leq t_k$. $\forall i \leq k-1, D_r\eta_i = 0$. We note that $D_r\eta_k = \eta_{k-1}f'_z(\theta_{t_{k-1}})$. For $i \geq k+1$, we have

$$\begin{aligned}
D_r\eta_i &= D_r\eta_{i-1} + hD_r(\eta_{i-1}f'_y(\theta_{t_{i-1}})) + \sum_{l=1}^{q} D_r(\eta_{i-1}f'_{z_l}(\theta_{t_{i-1}}))\Delta W^l_{i-1}, \\
&= \eta_{k-1}f'_z(\theta_{t_{k-1}}) + h\sum_{j=k}^{i-1} D_r(\eta_j f'_y(\theta_{t_j})) + \sum_{l=1}^{q}\sum_{j=k}^{i-1} D_r(\eta_j f'_{z_l}(\theta_{t_j}))\Delta W^l_j.
\end{aligned} \tag{4.15}$$

Applying Burkholder-Davis-Gundy's inequality to the martingale $\sum_{j=k}^{i-1} D_r(\eta_j f'_{z_l}(\theta_{t_j}))\Delta W^l_j$ yields

$$\begin{aligned}
\mathbb{E}|D_r\eta_i|^p &\leq C\mathbb{E}|\eta_{k-1}|^p + C_p h\sum_{j=k}^{i-1}\mathbb{E}|D_r(\eta_j f'_y(\theta_{t_j}))|^p + C\sum_{l=1}^{q}\mathbb{E}|h\sum_{j=k}^{i-1}|D_r(\eta_j f'_{z_l}(\theta_{t_j}))|^2|^{\frac{p}{2}} \\
&\leq C\mathbb{E}|\eta_{k-1}|^p + Ch\sum_{j=k}^{i-1}\mathbb{E}|D_r(\eta_j f'_y(\theta_{t_j}))|^p + C\sum_{l=1}^{q} h\sum_{j=k}^{i-1}\mathbb{E}|D_r(\eta_j f'_{z_l}(\theta_{t_j}))|^p \\
&\leq C(1 + \mathbb{E}|\eta_{k-1}|^p) + Ch\sum_{j=k+1}^{i-1}\mathbb{E}|D_r\eta_j|^p,
\end{aligned}$$

using the boundedness of the derivatives of f, $\max_{j\leq N}\|\eta_j\|_q = O(1)$, identity (2.2), $u, \sigma \in C^{1,2}_b$, and estimates (4.1). By applying Gronwall's lemma, we get $\max_{k\leq i\leq N}\mathbb{E}|D_r\eta_i|^p \leq C(1 + \mathbb{E}|\eta_{k-1}|^p), t_{k-1} < r \leq t_k$.
Then, $\max_{k\leq N}\mathbb{E}|D_r\eta_k|^p = O(1)$, uniformly in $r \in [0, T]$. The proof concerning the derivative of order 2 can be done following the same scheme. □

Proof of (4.13). We begin to show that $\mathbb{E}(\max_{k\leq N}|\eta_k|^p) < \infty$. The idea is to use a martingale property in order to apply Doob's inequality. Since $\eta_i = \eta_{i-1} + h\eta_{i-1}f'_y(\theta_{t_{i-1}}) + \eta_{i-1}f'_z(\theta_{t_{i-1}})\Delta W_{i-1}$, one has $\eta_k = 1 + \sum_{i=1}^{k}(h\eta_{i-1}f'_y(\theta_{t_{i-1}}) + \eta_{i-1}f'_z(\theta_{t_{i-1}})\Delta W_{i-1})$. Thus,

$$\mathbb{E}(\max_{k\leq N}|\eta_k|^p) \leq C\big(1 + \mathbb{E}(\sum_{i=1}^{N} h|\eta_{i-1}||f'_y(\theta_{t_{i-1}})|)^p + \mathbb{E}(\max_{k\leq N}|\sum_{i=1}^{k}\eta_{i-1}f'_z(\theta_{t_{i-1}})\Delta W_{i-1}|^p)\big).$$

The last term is upper bounded by $C\mathbb{E}(h\sum_{i=1}^{N}|\eta_{i-1}f'_z(\theta_{t_{i-1}})|^2)^{\frac{p}{2}} \leq Ch \sum_{i=1}^{N}\mathbb{E}|\eta_{i-1}f'_z(\theta_{t_{i-1}})|^p$. Using the estimate (4.12), we get $\mathbb{E}(\max_{k\leq N}|\eta_k|^p) < \infty$.
To prove that $\sup_{r\leq T}\mathbb{E}(\max_{k\leq N}|D_r\eta_k|^p) < \infty$, we proceed in the same way, by starting from (4.15). For the second derivative, this is analogous.

Step 4 : we prove that $\mathbb{E}(\Delta Y_N\eta_N) = O(h)$.
If η_N were equal to 1, the results of Bally and Talay [7] would directly apply. Here

the approach has to be different and we use techniques of Malliavin calculus. We have $\mathbb{E}(\Delta Y_N \eta_N) = \mathbb{E}(\eta_N \Phi(X_T^N) - \eta_N \Phi(X_T))$. Let us introduce $X_t^{N,\lambda} = (1-\lambda)X_t + \lambda X_t^N$. Thus, we have

$$\mathbb{E}(\Delta Y_N \eta_N) = \int_0^1 \mathbb{E}\big(\eta_N \Phi'_x(X_T^{N,\lambda})(X_T^N - X_T)\big)\mathrm{d}\lambda.$$

As $\Phi \in C^{3+\alpha}$, by using (4.12), (4.1) and (4.2), we note that $\eta_N \Phi'_x(X_T^{N,\lambda})$ satisfies R_2. By applying Remark 4.4, we deduce that $\mathbb{E}(\Delta Y_N \eta_N) = O(h)$.

Step 5 : we prove that $\mathbb{E}(f'_x(\theta_{t_i})\Delta X_i \eta_i) = O(h)$. This is a very similar proof to Step 4, in a case where $\Phi(x) = x$.

Conclusion. We now work on $h\mathbb{E}(\sum_{i=0}^{N-1} \tilde{\chi}_i \eta_i)$, where $|\tilde{\chi}_k| \le \lambda_k^N h + K(T,x)|\Delta X_k|^2 + K(T,x)|\Delta Z_k|^2$. Hence,

$$\begin{aligned}
|h \sum_{i=0}^{N-1} \mathbb{E}(\tilde{\chi}_i \eta_i)| \le& C \sum_{i=0}^{N-1} \mathbb{E}(\lambda_i^N |\eta_i|)h^2 + K(T,x) \sum_{i=0}^{N-1} h\mathbb{E}\big(|\eta_i|(|\Delta X_i|^2 + |\Delta Z_i|^2)\big) \\
\le& K(T,x)h + K(T,x) \sum_{i=0}^{N-1} h\mathbb{E}(|\eta_i||\Delta Z_i|^2) \\
\le& K(T,x)h + K(T,x)\big(\mathbb{E}(\max_{0\le i\le N-1} |\eta_i|)^2\big)^{\frac{1}{2}} \big(\mathbb{E}(h\sum_{i=0}^{N-1} |\Delta Z_i|^2)^2\big)^{\frac{1}{2}}.
\end{aligned}$$

By using (4.13) on $(\eta_i)_i$ and the upper bound (3.9) we get that $|h\mathbb{E}(\sum_{i=0}^{N-1} \tilde{\chi}_i \eta_i)| \le K(T,x)h$. By combining this result and the results of Step 4 and Step 5, (4.11) shows that $|\Delta Y_0| \le K(T,x)h$. Lemma 4.5 is proved. $\square$

Chapter 5

Proof of Theorem 2.5.

As it could be expected, its proof is more difficult. The main extra ingredient is the convergence of the weak derivative of the discrete BSDE (Y^N, Z^N), with the rate of convergence $N^{-1/2}$. The next paragraph is aimed at proving this result. In the following, Hypothesis 2.3 is in force.

5.1 Proof of an intermediate result

Proposition 5.1. *Let $r \in]0, t_1[$. Under Hypothesis 2.3, we have* $\max_{1\le i\le N} \mathbb{E}|D_r \Delta Y_i|^2 + h\mathbb{E}\big(\sum_{i=1}^{N-1} |D_r \Delta Z_i^*|^2\big) = O(h)$, *uniformly in r.*

This proposition is analogous to Theorem 2.3, where $q = 2$, and the scheme of its proof as well. However, there is a significative difference: the BSDE solved by the weak derivatives (see (5.1-5.2-5.3)) has a non Lipschitz driver, which requires extra technicalities that we detail. In what follows, we fix $r \in]0, t_1[$ and introduce some specific notations. $\widehat{X_t}$ stands for $D_r X_t$. In the case of Z_t, which is a row vector, $\widehat{Z_t}$ is a matrix whose the i-th column is $D_r^i Z_t^*$. It is well-known (Proposition 5.3 of El Karoui et al. [27]) that $(\widehat{Y_t}, \widehat{Z_t})_{r\le t\le T}$ solves

$$\widehat{Y_t} = \Phi'_x(X_T)\widehat{X_T} + \int_t^T (f'_x(\theta_s)\widehat{X_s} + f'_y(\theta_s)\widehat{Y_s} + f'_z(\theta_s)\widehat{Z_s})ds - (\int_t^T \widehat{Z_s}^* dW_s)^*. \tag{5.1}$$

Regarding $(\widehat{Y^N}, \widehat{Z^N})$, one obtains

$$\widehat{Y^N_{t_k}} = \mathbb{E}_{t_k}[\widehat{Y^N_{t_{k+1}}} + h\nabla_x f^N_{t_k}\widehat{X^N_{t_k}} + h\nabla_y f^N_{t_k}\widehat{Y^N_{t_{k+1}}} + h\nabla_z f^N_{t_k}\widehat{Z^N_{t_k}}], \tag{5.2}$$

$$\widehat{Z^N_{t_k}} = \frac{1}{h}\mathbb{E}_{t_k}[\Delta W_k \widehat{Y^N_{t_{k+1}}}], \tag{5.3}$$

where we set $\nabla_x f^N_{t_k} = \nabla_x f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k})$ and analogously for $\nabla_y f^N_{t_k}$ and $\nabla_z f^N_{t_k}$. Indeed, we can start from (1.4-1.5) and interchange conditional expectations and weak derivatives (see Proposition 1.2.4 in Nualart [78]). Another way to get (5.2-5.3) is to take advantage of the Markov structure of $(X^N_{t_k})_k$ to write $Y^N_{t_k} = y^N(t_k, X^N_{t_k})$, where the function y^N is the solution of a dynamic programming equation, and then apply the chain rule. We omit further details.

From (2.2), we also have

$$\widehat{Y_t} = \nabla_x u(t, X_t)\widehat{X_t},\ \widehat{Z_t} = \nabla_x(\nabla_x u\sigma)^*(t, X_t)\widehat{X_t}. \tag{5.4}$$

For the sake of clarity, let us write, for any process V, $\widehat{\Delta V_k} = D_r V^N_{t_k} - D_r V_{t_k}$. In particular, we have $\widehat{\Delta \overline{Z}_k} = D_r(Z^{N*}_{t_k} - \overline{Z}^*_{t_k}) = \widehat{Z^N_{t_k}} - \widehat{\overline{Z}_{t_k}}$, where $\widehat{\overline{Z}_{t_k}}$ is defined as $h\widehat{\overline{Z}_{t_k}} = \mathbb{E}_{t_k} \int_{t_k}^{t_{k+1}} \widehat{Z_s} ds$ (see the beginning of Section 3).

5.1.1 Preparatory estimates

In this part we give some L_p-estimates ($p \geq 1$), which are repeatedly used in the following calculations.

- $$\sup_{i \leq j \leq N} (\mathbb{E}_{t_i} |\widehat{X^N_{t_j}}|^{2p}) \leq C |\widehat{X^N_i}|^{2p}, \tag{5.5}$$
- $$\mathbb{E}(\max_{0 \leq j \leq N} |\widehat{X^N_{t_j}}|^{2p}) = O(1), \tag{5.6}$$
- $$\forall j \in 0..N-1,\ |\widehat{Y^N_{t_j}}|^2 \leq C |\widehat{X^N_{t_j}}|^2,\ \ \mathbb{E}(\max_{0 \leq j \leq N} |\widehat{Y^N_{t_j}}|^{2p}) = O(1), \tag{5.7}$$
- $$\mathbb{E}(\sup_{0 \leq t \leq T} |\widehat{X_t}|^{2p} + \sup_{0 \leq t \leq T} |\widehat{Y_t}|^{2p} + \sup_{0 \leq t \leq T} |\widehat{Z_t}|^{2p}) = O(1), \tag{5.8}$$
- Let F satisfy R_3. Then, $|\mathbb{E}(F(\widehat{X^N_t} - \widehat{X_t}))| = O(h)$. Furthermore,
$$\sup_{0 \leq k \leq N} \mathbb{E} |\widehat{\Delta X_k}|^{2p} = O(h^p). \tag{5.9}$$
- Analogously to (3.1), $\forall s \in [t_k, t_{k+1}]$, we have
$$\mathbb{E}_{t_k} \big(|\widehat{X_s} - \widehat{X_{t_k}}|^{2p} + |\widehat{Y_s} - \widehat{Y_{t_k}}|^{2p} + |\widehat{Z_s} - \widehat{\overline{Z}_{t_k}}|^{2p} \big) = O_k(h^p). \tag{5.10}$$

Note that $\widehat{X^N_{t_1}} = \sigma(0,x)$, and $\widehat{X^N_{t_{k+1}}} = (1 + h b'_x(t_k, X^N_{t_k}) + \sum_{i=1}^q (\sigma_i)'_x(t_k, X^N_{t_k}) \Delta W^i_k) \widehat{X^N_{t_k}}$ for $1 \leq k \leq N$. Thus, we easily get $\mathbb{E}_{t_i} |\widehat{X^N_{t_j}}|^{2p} \leq (1 + Ch) \mathbb{E}_{t_i} |\widehat{X^N_{t_{j-1}}}|^{2p}$, and (5.5) follows. The proof of (5.6) can be done as the proof of (4.13).

Proof of (5.7). From (5.2), we use Young's inequality and boundedness of ∇f to get

$$|\widehat{Y^N_{t_i}}|^2 \leq (1 + \gamma h) |\mathbb{E}_{t_i} \widehat{Y^N_{t_{i+1}}}|^2 + Ch(h + \frac{1}{\gamma}) \big(|\widehat{X^N_{t_i}}|^2 + \mathbb{E}_{t_i} |\widehat{Y^N_{t_{i+1}}}|^2 + |\widehat{Z^N_{t_i}}|^2 \big). \tag{5.11}$$

From (5.3) and the Cauchy Schwarz inequality, we obtain $h |\widehat{Z^N_{t_i}}|^2 \leq C(\mathbb{E}_{t_i} |\widehat{Y^N_{t_{i+1}}}|^2 - |\mathbb{E}_{t_i} \widehat{Y^N_{t_{i+1}}}|^2)$. Hence, with an appropriate choice of γ, (5.11) is reduced to $|\widehat{Y^N_{t_i}}|^2 \leq (1 + Ch) \mathbb{E}_{t_i} |\widehat{Y^N_{t_{i+1}}}|^2 + Ch |\widehat{X^N_{t_i}}|^2$, and thus Gronwall's lemma yields

$$|\widehat{Y^N_{t_i}}|^2 \leq C \mathbb{E}_{t_i} (|\widehat{Y^N_{t_N}}|^2 + h \sum_{j=i}^{N-1} |\widehat{X^N_{t_j}}|^2) \leq C \sup_{i \leq j \leq N-1} \mathbb{E}_{t_i} |\widehat{X^N_{t_j}}|^2.$$

Finally, estimates (5.5) and (5.6) complete the proof.

Proof of (5.8). $\mathbb{E}(\sup_{0 \leq t \leq T} |\widehat{X_t}|^{2p}) = O(1)$ follows from (4.1). The other estimates come from this result and (5.4).

Proof of (5.9). Let us introduce $X'_t = \nabla_x X_t (\nabla_x X_r)^{-1} \sigma(0,x)$ and write $\widehat{X^N_t} - \widehat{X_t} = \widehat{X^N_t} - X'_t + X'_t - \widehat{X_t}$.

Since $\widehat{X_t} = \nabla_x X_t (\nabla_x X_r)^{-1} \sigma(r, X_r)$, a direct application of Proposition 4.1 with $U_t = \sigma(t, X_t)$ gives $\mathbb{E}(F(X'_t - \widehat{X_t})) = O(h)$ for F satisfying R_2. Moreover, simple increment estimates yield $\sup_{t \leq T} \mathbb{E} |X'_t - \widehat{X_t}|^{2p} = O(h^p)$.

It remains to study the impact of the difference $\widehat{X_t^N} - X'_t$. $(\widehat{X_t^N})_{t\geq r}$ and $(X'_t)_{t\geq r}$ are solutions of

$$\begin{aligned}\widehat{X_t^N} =&\sigma(0,x) + \int_r^t b'_x(\eta(s), X^N_{\eta(s)})\widehat{X^N_{\eta(s)}}ds + \sum_{i=1}^q \int_r^t (\sigma_i)'_x(\eta(s), X^N_{\eta(s)})\widehat{X^N_{\eta(s)}}dW^i_s,\\ X'_t =&\sigma(0,x) + \int_r^t b'_x(s, X_s)X'_s ds + \sum_{i=1}^q \int_r^t (\sigma_i)'_x(s, X_s)X'_s dW^i_s. \end{aligned} \tag{5.12}$$

For the sake of simplicity, we take $b \equiv 0$ and $d = q = 1$. If we set $\sigma'(s) = \int_0^1 \sigma'_x(s, X_s + \lambda(X^N_s - X_s))d\lambda$, we observe that ΔX_t solves the linear equation $\Delta X_t = \int_0^t [\sigma(\eta(s), X^N_{\eta(s)}) - \sigma(s, X^N_s)]dW_s + \int_0^t \sigma'(s)\Delta X_s dW_s$, which solution is given by (see Theorem 56 p. 271 in Protter [87])

$$\begin{aligned}\Delta X_t =&\epsilon_t \int_0^t \epsilon_s^{-1}[\sigma(\eta(s), X^N_{\eta(s)}) - \sigma(s, X^N_s)](dW_s - \sigma'(s)ds)\\ =& -\epsilon_t \int_0^t \epsilon_s^{-1}\Big[\int_{\eta(s)}^s \sigma'_x(v, X^N_v)\sigma(\eta(v), X^N_{\eta(v)})dW_v\\ &+ (\sigma'_t(v, X^N_v) + \frac{1}{2}\sigma''_{xx}(v, X^N_v)\sigma^2(\eta(v), X^N_{\eta(v)}))dv\Big](dW_s - \sigma'(s)ds)\end{aligned}$$

where $\epsilon_t = 1 + \int_0^t \sigma'(s)\epsilon_s dW_s$. This proves that ΔX_t satisfies the property $\mathcal{P}$. Analogously, if we define $\sigma''(s) = \int_0^1 \sigma''_{xx}(s, X_s + \lambda(X^N_s - X_s))d\lambda$ and $\epsilon^N_t = 1 + \int_r^t \sigma'_x(s, X^N_s)\epsilon^N_s dW_s$, simple computations lead to

$$\begin{aligned}\widehat{X_t^N} - X'_t =&\epsilon_t^N \int_r^t (\epsilon^N_s)^{-1}([\sigma'_x(\eta(s), X^N_{\eta(s)})\widehat{X^N_{\eta(s)}} - \sigma'_x(s, X^N_s)\widehat{X^N_s}] + \sigma''(s)X'_s\Delta X_s)\\ &(dW_s - \sigma'_x(s, X^N_s)ds).\end{aligned}$$

From the above representation, it is straightforward to conclude $\sup_{t\leq T}\mathbb{E}|\widehat{\Delta X_t}|^{2p} = O(h^p)$. Now, let us upper bound $\mathbb{E}(F(\widehat{X_t^N} - X'_t))$ which can be decomposed into several terms.

- The contribution associated to $\epsilon^N_t \int_r^t (\epsilon^N_s)^{-1}[\sigma'_x(\eta(s), X^N_{\eta(s)})\widehat{X^N_{\eta(s)}} - \sigma'_x(s, X^N_s)\widehat{X^N_s}](dW_s - \sigma'_x(s, X^N_s)ds)$ satisfies property $\mathcal{P}$, thus Proposition 4.3 yields the expected result.

- The contribution $\mathbb{E}(F\epsilon^N_t \int_r^t (\epsilon^N_s)^{-1}\sigma''(s)X'_s\Delta X_s\sigma'_x(s, X^N_s)ds)$ is equal to $\int_r^t \mathbb{E}(F\epsilon^N_t(\epsilon^N_s)^{-1}\sigma''(s)X'_s\Delta X_s\sigma'_x(s, X^N_s))ds = O(h)$ in view of Remark 4.4.

- In the same way, the duality relationship ensures that the last contribution $\mathbb{E}(F\epsilon^N_t \int_r^t (\epsilon^N_s)^{-1}\sigma''(s)X'_s\Delta X_s dW_s) = \int_r^t \mathbb{E}(D_s(F\epsilon^N_t)(\epsilon^N_s)^{-1}\sigma''(s)X'_s\Delta X_s)ds$ is a $O(h)$ (using here that F satisfies R_3).

Proof of (5.10). In view of $\widehat{X_t} = D_r X_t = \nabla_x X_t(\nabla_x X_r)^{-1}\sigma(r, X_r)$, the estimate on the increments of $\widehat{X_t}$ becomes clear. The other ones easily follow. □

5.1.2 Proof of $\max_{1\le i\le N}\mathbb{E}|\widehat{\Delta Y_i}|^2=O(h)$.

Assume that for some non negative random variable $\Lambda_k=O_k(h)+|\Delta X_k|^2+|\Delta Z_k|^2$, one has

$$|\widehat{\Delta Y_k}|^2\le(1+Ch)\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2+h|\widehat{\Delta X_k}|^2+h\Lambda_k O_k(1). \tag{5.13}$$

Take the expectation on both sides, use estimates (5.9) and those of Proposition 2.2 to get

$$\mathbb{E}|\widehat{\Delta Y_k}|^2\le C\mathbb{E}|\widehat{\Delta Y_N}|^2+O(h)+Ch\sum_{k=0}^{N-1}\mathbb{E}(|\Delta Z_k|^2O_k(1)).$$

On the one hand, as $\widehat{\Delta Y_N}=\Phi'(X^N_{t_N})\widehat{X^N_{t_N}}-\Phi'(X_{t_N})\widehat{X_{t_N}}$, clearly $\mathbb{E}|\widehat{\Delta Y_N}|^2=O(h)$. On the other hand, in view of (3.9) with $p=2$, the summation above is a $O(h)$. This proves $\max_{1\le k\le N}\mathbb{E}|\widehat{\Delta Y_k}|^2=O(h)$.

Proof of (5.13). ¿From (5.1) and (5.2), we obtain

$$\begin{aligned}\widehat{\Delta Y_k}=\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})+\mathbb{E}_{t_k}\Big(\int_{t_k}^{t_{k+1}}&[\nabla_x f^N_{t_k}\widehat{X^N_{t_k}}-f'_x(\theta_s)\widehat{X_s}\\&+\nabla_y f^N_{t_k}\widehat{Y^N_{t_{k+1}}}-f'_y(\theta_s)\widehat{Y}_s+\nabla_z f^N_{t_k}\widehat{Z^N_{t_k}}-f'_z(\theta_s)\widehat{Z_s}]ds\Big).\end{aligned}$$

Since $f\in C_b^{2,4,4,4}$, it follows that for any $\gamma>0$ (to be fixed later)

$$\begin{aligned}|\widehat{\Delta Y_k}|^2\le&(1+\gamma h)|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2+C(h+\frac{1}{\gamma})\mathbb{E}_{t_k}\Big(\int_{t_k}^{t_{k+1}}[|\nabla_x f^N_{t_k}\widehat{X^N_{t_k}}-f'_x(\theta_s)\widehat{X_s}|^2\\&+|\nabla_y f^N_{t_k}\widehat{Y^N_{t_{k+1}}}-f'_y(\theta_s)\widehat{Y}_s|^2+|\nabla_z f^N_{t_k}\widehat{Z^N_{t_k}}-f'_z(\theta_s)\widehat{Z_s}|^2]ds\Big) &(5.14)\\\le&(1+\gamma h)|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2+C(h+\frac{1}{\gamma})(T^1_k+T^2_k), &(5.15)\end{aligned}$$

where we put $T^1_k=\mathbb{E}_{t_k}(\int_{t_k}^{t_{k+1}}[|\widehat{X^N_{t_k}}-\widehat{X_s}|^2+|\widehat{Y^N_{t_{k+1}}}-\widehat{Y}_s|^2+|\widehat{Z^N_{t_k}}-\widehat{Z_s}|^2]ds)$, $T^2_k=\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}(h+|X_s-X^N_{t_k}|^2+|Y_s-Y^N_{t_{k+1}}|^2+|Z_s-Z^N_{t_k}|^2)(|\widehat{X_s}|^2+|\widehat{Y}_s|^2+|\widehat{Z_s}|^2)ds$. To get (5.13), we need to simplify (5.15), by estimating T^1_k and T^2_k.

Term T^1_k. Firstly, we write $\mathbb{E}_{t_k}|\widehat{Y^N_{t_{k+1}}}-\widehat{Y}_s|^2\le 2\mathbb{E}_{t_k}|\widehat{Y_{t_{k+1}}}-\widehat{Y}_s|^2+2\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2$. We do the same for $\widehat{X^N_{t_k}}-\widehat{X_s}$. Then, the usual increment estimates yield

$$\mathbb{E}_{t_k}|\widehat{Y^N_{t_{k+1}}}-\widehat{Y}_s|^2+\mathbb{E}_{t_k}|\widehat{X^N_{t_k}}-\widehat{X_s}|^2\le O_k(h)+2|\widehat{\Delta X_k}|^2+2\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2.$$

Secondly, analogously to (3.5), we have

$$\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|\widehat{Z^N_{t_k}}-\widehat{Z_s}|^2ds=\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|\widehat{\overline{Z}_{t_k}}-\widehat{Z_s}|^2ds+h\mathbb{E}_{t_k}|\widehat{Z^N_{t_k}}-\widehat{\overline{Z}_{t_k}}|^2.$$

Finally, we obtain $T^1_k\le Ch(O_k(h)+|\widehat{\Delta X_k}|^2+\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2+|\widehat{\Delta\overline{Z}_k}|^2)$.

Term T^2_k. Easy calculations combining (3.1), Proposition 3.1 and (5.8) give $T^2_k\le(O_k(h^2)+h|\Delta X_k|^2+h|\Delta Z_k|^2)O_k(1)=h\Lambda_kO_k(1)$.

Conclusion. Plugging the estimates on T^1_k and T^2_k into (5.15), we get

$$|\widehat{\Delta Y_k}|^2 \leq (1+\gamma h)|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 + Ch(h+\frac{1}{\gamma})|\widehat{\Delta \overline{Z}_k}|^2 \\ + Ch(h+\frac{1}{\gamma})(|\widehat{\Delta X_k}|^2 + \mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2 + \Lambda_k O_k(1)). \tag{5.16}$$

Note that $h\widehat{\overline{Z}_{t_k}} = \mathbb{E}_{t_k}(\Delta W_k(\widehat{Y_{t_{k+1}}} + \int_{t_k}^{t_{k+1}}[f'_x(\theta_s)\widehat{X_s} + f'_y(\theta_s)\widehat{Y_s} + f'_z(\theta_s)\widehat{Z_s}]ds))$, whence $h\widehat{\Delta \overline{Z}_{t_k}} = \mathbb{E}_{t_k}(\Delta W_k(\widehat{\Delta Y_{k+1}} + \int_{t_k}^{t_{k+1}}[f'_x(\theta_s)\widehat{X_s} + f'_y(\theta_s)\widehat{Y_s} + f'_z(\theta_s)\widehat{Z_s}]ds))$. By proceeding as before, we easily prove

$$h|\widehat{\Delta \overline{Z}_{t_k}}|^2 \leq C(\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2 - |\mathbb{E}_{t_k}\widehat{\Delta Y_{k+1}}|^2) + O_k(h^2). \tag{5.17}$$

Combining this upper bound with (5.16) for a good choice of γ gives (5.13). □

5.1.3 Proof of $h\mathbb{E}\big(\sum_{k=1}^{N-1}|\widehat{\Delta Z_k}|^2\big) = O(h)$.

In view of (5.10), this is equivalent to prove $h\mathbb{E}\big(\sum_{k=1}^{N-1}|\widehat{\Delta \overline{Z}_k}|^2\big) = O(h)$. To establish this estimate, we start from (5.17) to get

$$h\sum_{k=1}^{N-1}\mathbb{E}|\widehat{\Delta \overline{Z}_k}|^2 \leq C\sum_{k=1}^{N-1}(\mathbb{E}|\widehat{\Delta Y_k}|^2 - \mathbb{E}|\mathbb{E}_{t_k}\widehat{\Delta Y_{k+1}}|^2) + C\mathbb{E}|\widehat{\Delta Y_N}|^2 + O(h). \tag{5.18}$$

Now, we work on $|\widehat{\Delta Y_k}|^2 - |\mathbb{E}_{t_k}\widehat{\Delta Y_{k+1}}|^2$. The choice $\gamma = 2C^2$ in (5.16) leads to

$$|\widehat{\Delta Y_k}|^2 - |\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 \leq \gamma h|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 + h(\frac{1}{2C}+Ch)|\widehat{\Delta \overline{Z}_k}|^2 \\ + h(Ch+\frac{1}{2C})(|\widehat{\Delta X_k}|^2 + \mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2 + \Lambda_k O_k(1)).$$

From (5.9) and the result from Section 5.1.2, we have $\max_{1\leq k\leq N}\mathbb{E}(|\widehat{\Delta X_k}|^2 + |\widehat{\Delta Y_k}|^2) = O(h)$. We also have $\mathbb{E}(\Lambda_k O_k(1)) = O(h) + \mathbb{E}(|\Delta Z_k|^2 O_k(1))$. Consequently, for h small enough, one has $\mathbb{E}|\widehat{\Delta Y_k}|^2 - \mathbb{E}|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 \leq \frac{2h}{3C}\mathbb{E}|\widehat{\Delta \overline{Z}_k}|^2 + O(h^2) + Ch\mathbb{E}(|\Delta Z_k|^2 O_k(1))$. Putting this estimate into (5.18) yields

$$\frac{1}{3}h\sum_{k=1}^{N-1}\mathbb{E}|\widehat{\Delta \overline{Z}_k}|^2 \leq O(h) + Ch\sum_{k=1}^{N-1}\mathbb{E}(|\Delta Z_k|^2 O_k(1)).$$

Inequality (3.9) with $p=2$ directly shows that the sum above is a $O(h)$. □

5.2 Expansion of $Z^N_{t_k} - Z_{t_k}$

We recall that $u \in C_b^{2,4}$ owing to Hypothesis 2.3. From (4.9), we have $\Delta Z_k = O(h) + \frac{1}{h}\mathbb{E}_{t_k}[(u^N(t_{k+1}, X^N_{t_{k+1}}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W^*_k]$. Let $(X_t^{s,\overline{x}})_{t\geq s}$ denote the solution of the SDE (1.1) starting at time s from $\overline{x}$. We write X_t for $X_t^{0,x}$. Note that $X_{t_{k+1}} = X_{t_{k+1}}^{t_k, X_{t_k}}$. In the same way, the Euler scheme starting at time t_k at $\overline{x}$ is denoted by $(X_{t_j}^{N,t_k,\overline{x}})_{j\geq k}$. With this notation we can rewrite ΔZ_k

$$\begin{aligned}\Delta Z_k =&\frac{1}{h}\mathbb{E}_{t_k}[(u^N(t_{k+1}, X_{t_{k+1}}^{N,t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*] + O(h),\\ =&\frac{1}{h}\mathbb{E}_{t_k}[(u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*]\\ &+ \frac{1}{h}\mathbb{E}_{t_k}[(u^N(t_{k+1}, X_{t_{k+1}}^{N,t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}))\Delta W_k^*] + O(h).\end{aligned} \tag{5.19}$$

We work on the first two terms separately by proving

Lemma 5.2. $\frac{1}{h}\mathbb{E}_{t_k}[(u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*] = O(|\Delta X_k|^2) + O(h) + [\nabla_x(\nabla_x u\ \sigma)^*(t_k, X_{t_k})\Delta X_k]^*$.

Lemma 5.3. $\frac{1}{h}\big|\mathbb{E}_{t_k}[(u^N(t_{k+1}, X_{t_{k+1}}^{N,t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}))\Delta W_k^*]\big| = O_k(h)$.

The combination of these Lemmas completes the proof of Theorem 2.5.

5.2.1 Proof of Lemma 5.2.

For the sake of simplicity, let $\Delta_N X_{k+1}$ denote $X_{t_{k+1}}^{t_k,X_{t_k}^N} - X_{t_{k+1}}$ (which is different from $\Delta X_{k+1} = X_{t_{k+1}}^{N,t_k,X_{t_k}^N} - X_{t_{k+1}}$). From a Taylor-Lagrange formula, we obtain

$$\begin{aligned}&u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}) = u'_x(t_{k+1}, X_{t_{k+1}})\Delta_N X_{k+1}\\ &+ \int_0^1 (1-\lambda)(\Delta_N X_{k+1})^* H_x(u)\big(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1}\big)\Delta_N X_{k+1} d\lambda.\end{aligned}$$

Thus, using the duality relationship, one has

$$\begin{aligned}&\mathbb{E}_{t_k}[(u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*]\\ &= \int_{t_k}^{t_{k+1}} R_k^1(t)\mathrm{d}t + \int_{t_k}^{t_{k+1}} R_k^2(t)\mathrm{d}t + \int_0^1 (1-\lambda)R_k^3(\lambda)d\lambda,\end{aligned}$$

$$\begin{aligned}\text{with } &R_k^1(t) = \mathbb{E}_{t_k}[(\Delta_N X_{k+1})^* H_x(u)(t_{k+1}, X_{t_{k+1}})D_t X_{t_{k+1}}],\\ &R_k^2(t) = \mathbb{E}_{t_k}[u'_x(t_{k+1}, X_{t_{k+1}})D_t(\Delta_N X_{k+1})],\\ &R_k^3(\lambda) = \mathbb{E}_{t_k}[(\Delta_N X_{k+1})^* H_x(u)(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1})\Delta_N X_{k+1}\Delta W_k^*].\end{aligned}$$

Expansion of $R_k^1(t)$. Clearly $\Delta_N X_{k+1} = \Delta X_k + U_{t_{k+1}} - U_{t_k}$, where U is an Itô process with drift term $\alpha_s = b(s, X_s^{t_k,X_{t_k}^N}) - b(s, X_s)$ and diffusion term $\beta_s = \sigma(s, X_s^{t_k,X_{t_k}^N}) - \sigma(s, X_s)$, both being bounded. Thus, we can apply Proposition 4.1, letting $F = H_x(u)(t_{k+1}, X_{t_{k+1}})D_t X_{t_{k+1}}$. Because $u \in C_b^{2,4}$ and in view of (4.1), we get

$$R_k^1(t) = O(h) + (\Delta X_k)^*\mathbb{E}_{t_k}[H_x(u)(t_{k+1}, X_{t_{k+1}})D_t X_{t_{k+1}}].$$

We expand the latter factor. As $D_t X_{t_{k+1}} = \nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1}\sigma(t, X_t)$, we have

$$\begin{aligned}H_x(u)(t_{k+1}, X_{t_{k+1}})D_t X_{t_{k+1}} =& (H_x(u)(t_{k+1}, X_{t_{k+1}})\sigma(t, X_t) - H_x(u)(t, X_t)\sigma(t, X_t))\\ &+ (H_x(u)(t, X_t)\sigma(t, X_t) - H_x(u)(t_k, X_{t_k})\sigma(t_k, X_{t_k}))\\ &+ (H_x(u)(t_{k+1}, X_{t_{k+1}})[\nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1} - I]\sigma(t, X_t))\\ &+ H_x(u)(t_k, X_{t_k})\sigma(t_k, X_{t_k}).\end{aligned}$$

The first three contributions in the r.h.s. above can be handled in the same way and we give a detailed proof only for the first one. It is enough to apply Proposition 4.1 with $F = \sigma(t, X_t)$ and $U_s = H_x(u)(s, X_s)$. Then, $\mathbb{E}_{t_k}[F(U_{t_{k+1}} - U_t)]$ is of order h with a constant involving b, σ, u and its derivatives up to order 4. Finally, this gives

$$R^1_k(t) = O(h) + (\Delta X_k)^* H_x(u)(t_k, X_{t_k})\sigma(t_k, X_{t_k}),$$

uniformly in $t \in [t_k, t_{k+1}]$.

Expansion of $R^2_k(t)$. For $t_k \le t \le t_{k+1}$, we have

$$\begin{aligned}
D_t(\Delta_N X_{k+1}) =& [\nabla_x X^{X^N_{t_k}, t_k}_{t_{k+1}} (\nabla_x X^{X^N_{t_k}, t_k}_t)^{-1} - I]\sigma(t, X^{X^N_{t_k}, t_k}_t) \\
& - [\nabla_x X_{t_{k+1}} (\nabla_x X_t)^{-1} - I]\sigma(t, X_t) - (\sigma(t, X_t) - \sigma(t_k, X_{t_k})) \\
& + \sigma(t, X^{X^N_{t_k}, t_k}_t) - \sigma(t_k, X^N_{t_k}) + \sigma(t_k, X^N_{t_k}) - \sigma(t_k, X_{t_k}).
\end{aligned}$$

As before, apply Proposition 4.1 to each of these terms but the last one, with $F = u'_x(t_{k+1}, X_{t_{k+1}})$, using $u, b, \sigma \in C^{2,4}_b$ and (4.1). It follows that $R^2_k(t) = O(h) + \mathbb{E}_{t_k}[u'_x(t_{k+1}, X_{t_{k+1}})](\sigma(t_k, X^N_{t_k}) - \sigma(t_k, X_{t_k}))$. An application of Itô's formula yields

$$\begin{aligned}
R^2_k(t) &= O(h) + \sum_{i=1}^{d} u'_{x_i}(t_k, X_{t_k})(\sigma^i(t_k, X^N_{t_k}) - \sigma^i(t_k, X_{t_k})) \\
&= O(h + |\Delta X_k|^2) + \sum_{i=1}^{d} u'_{x_i} \nabla_x([\sigma^i]^*)(t_k, X_{t_k})\Delta X_k,
\end{aligned}$$

uniformly in $t \in [t_k, t_{k+1}]$. Finally, simple matrix computations lead to

$$R^1_k(t) + R^2_k(t) = O(h + |\Delta X_k|^2) + [\nabla_x(\nabla_x u\sigma)^*(t_k, X_{t_k})\Delta X_k]^*.$$

Upper bound for $R^3_k(\lambda)$. To complete the proof of Lemma 5.2, note that it remains to justify that $R^3_k(\lambda) = hO(h + |\Delta X_k|^2)$ uniformly in λ. The duality formula gives

$$R^3_k(\lambda) = \mathbb{E}_{t_k}[\int_{t_k}^{t_{k+1}} D_t[(\Delta_N X_{k+1})^* H_x(u)(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1})\Delta_N X_{k+1}]\mathrm{d}t.$$

The term in the integral equals $\sum_{i,j=1}^{d}[2D_t(\Delta_N X_{k+1,i})\Delta_N X_{k+1,j}\partial^2_{x_i,x_j}u(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1}) + \Delta_N X_{k+1,i}\Delta_N X_{k+1,j} D_t(\partial^2_{x_i,x_j}u(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1}))]$. Thanks to (4.1) and (4.2) and successive applications of Proposition 4.1, we finally prove our assertion. We omit further details. □

5.2.2 Proof of Lemma 5.3

As for Lemma 4.5, we only do the proof for $t_k = 0$, i.e. we have to show $|\mathbb{E}_{t_k}[(u^N(t_1, X^{N,0,x}_{t_1}) - u(t_1, X^{0,x}_{t_1}))\Delta W^*_0]| \le K(T, x)h^2$. We have $\mathbb{E}[(u^N(t_1, X^{N,0,x}_{t_1}) - u(t_1, X^{0,x}_{t_1}))\Delta W^*_0] = \mathbb{E}[\Delta Y_1 \Delta W^*_0]$. By using (4.4), we come up with

$$\mathbb{E}[\Delta Y_1 \Delta W^*_0] = \mathbb{E}[\xi_1 ... \xi_{N-1}\Delta Y_N \Delta W^*_0] + \mathbb{E}[h\sum_{i=1}^{N-1}(f'_x(\theta_{t_i})\Delta X_i + \tilde{\chi}_i)\xi_1 ... \xi_{i-1}\Delta W^*_0],$$

where $\tilde{\chi_i} = \mathbb{E}_{t_i}(\chi_i)$ (ξ_i and χ_i are defined in (4.5) and (4.6)). In the following $\tilde{\eta_i}$ denotes $\xi_1 ... \xi_{i-1}$ and $\tilde{\eta_1} = 1$. We easily prove that $(\tilde{\eta_i})_{1 \leq i \leq N}$ has the analogous properties to $(\eta_i)_{0 \leq i \leq N}$. Estimates (4.12) and (4.13) remain valid for $\tilde{\eta}$ and under Hypothesis 2.3, the estimate (4.12) becomes

$$\tilde{\eta_k} \text{ satisfies } R_3 \text{ uniformly in } k. \tag{5.20}$$

Step 1 : Proof of $\mathbb{E}[\xi_1 ... \xi_{N-1} \Delta Y_N \Delta W_0^*] = \mathbb{E}[\tilde{\eta_N} \Delta Y_N \Delta W_0^*] = O(h^2)$**.**
As before, we use the duality formula:

$$\mathbb{E}[\tilde{\eta_N} \Delta Y_N \Delta W_0^*] = \mathbb{E} \int_0^{t_1} (D_t[\tilde{\eta_N}] \Delta Y_N + \tilde{\eta_N} D_t[\Delta Y_N]) \mathrm{dt}.$$

Since $\tilde{\eta_N}$ satisfies (5.20), we proceed as in **Step 4** of Lemma 4.5 and we get $\mathbb{E}(D_t[\tilde{\eta_N}] \Delta Y_N) = O(h)$. Furthermore, we have

$$D_t[\Delta Y_N] = (\Phi'(X_T^N) - \Phi'(X_T)) D_t X_T^N + \Phi'(X_T)(D_t X_T^N - D_t X_T).$$

On the one hand, analogously to previous computations, we establish $\mathbb{E}(\tilde{\eta_N}(\Phi'(X_T^N) - \Phi'(X_T)) D_t X_T^N) = O(h)$.
On the other hand, we prove $\mathbb{E}(\tilde{\eta_N} \Phi'(X_T)(D_t X_T^N - D_t X_T)) = O(h)$. Thanks to (4.1) and (4.12), $\tilde{\eta_N} \Phi'(X_T)$ satisfies condition R_3. Then, by applying (5.9), we get the result.

Step 2 : Proof of $\mathbb{E}[h \sum_{i=1}^{N-1} f_x'(\theta_{t_i}) \Delta X_i \xi_1 ... \xi_{i-1} \Delta W_0^*] = O(h^2)$**.**
This is a similar proof to the one done at **Step 1**, with $\Phi(x) = x$.

Step 3 : Proof of $\mathbb{E}[h \sum_{i=1}^{N-1} \tilde{\chi_i} \tilde{\eta_i} \Delta W_0^*] = O(h^2)$.
A careful inspection of the definition of G_0, G_y and G_z appearing in (4.6) shows that under Hypothesis 2.3, these functions are continuously differentiable w.r.t. the variable x (with a bounded derivative). Hence, if we write $\chi_i = \chi_i^1 + \int_0^1 (1-\lambda) \chi_i^2(\lambda) d\lambda$ with (see (4.6))

$$\begin{aligned}
\chi_i^1 &= \int_{t_i}^{t_{i+1}} (G_0(s, X_s) + f_y'(\theta_{t_i}) G_y(s, X_s) + f_z'(\theta_{t_i}) G_z(s, X_s)) ds, \\
\chi_i^2(\lambda) &= \Delta X_i^* f_{xx}''(\theta_{t_i}^\lambda) \Delta X_i + f_{yy}''(\theta_{t_i}^\lambda)(Y_{t_{i+1}}^N - Y_{t_i})^2 + \Delta Z_i f_{zz}''(\theta_{t_i}^\lambda) \Delta Z_i^* \\
&\quad + 2 \Delta X_i^* f_{xy}''(\theta_{t_i}^\lambda)(Y_{t_{i+1}}^N - Y_{t_i}) + 2 \Delta X_i^* f_{xz}''(\theta_{t_i}^\lambda) \Delta Z_i^* + 2(Y_{t_{i+1}}^N - Y_{t_i}) f_{yz}''(\theta_{t_i}^\lambda) \Delta Z_i^*,
\end{aligned}$$

we note that the random variable χ_i is in $\mathbb{D}^{1,\infty}$. Thus and because $\tilde{\chi_i} = \mathbb{E}_{t_i}(\chi_i)$, one has $\mathbb{E}[\tilde{\chi_i} \tilde{\eta_i} \Delta W_0^*] = \mathbb{E}[\chi_i \tilde{\eta_i} \Delta W_0^*] = \mathbb{E}[\int_0^{t_1} (\chi_i D_t \tilde{\eta_i} + \tilde{\eta_i} D_t \chi_i) dt]$.

The upper bound $\tilde{\chi_i} = \mathbb{E}_{t_i}(\chi_i) = O_i(h) + O(|\Delta X_i|^2 + |\Delta Z_i|^2)$ (see (4.10)) is sufficient to show $\mathbb{E}[\sum_{i=1}^{N-1} \chi_i D_t \tilde{\eta_i}] = O(1)$ uniformly in t (follow the arguments of the conclusion of the proof of Lemma 4.5 and use (4.13) with $\tilde{\eta}$).

Now, it remains to establish $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta_i} D_t \chi_i] = O(1)$. On the one hand, clearly $\mathbb{E}_{t_i}[D_t \chi_i^1] = O_i(h)$ and we conclude $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta_i} D_t \chi_i^1] = O(1)$ uniformly in t. On the other hand, χ_i^2 can be decomposed into several contributions, which can be analyzed with the same arguments. Let us detail how to handle one of them, for instance $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta_i} D_t(\Delta X_i^* f_{xz}''(\theta_{t_i}^\lambda) \Delta Z_i^*)]$ which has to be a $O(1)$. We do the proof for $d = q = 1$. Write $D_t(\Delta X_i f_{xz}''(\theta_{t_i}^\lambda) \Delta Z_i) = \Delta X_i f_{xz}''(\theta_{t_i}^\lambda) D_t(\Delta Z_i) + D_t(\Delta X_i) f_{xz}''(\theta_{t_i}^\lambda) \Delta Z_i +$

$\Delta X_i D_t(f''_{xz}(\theta^\lambda_{t_i}))\Delta Z_i$. As f'' is bounded, we have

$$\begin{aligned}\big|\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i f''_{xz}(\theta^\lambda_{t_i}) D_t(\Delta Z_i)]\big| &\leq \mathbb{E}\big[\sum_{i=1}^{N-1} |\tilde{\eta}_i||\Delta X_i||f''_{xz}(\theta^\lambda_{t_i})||D_t(\Delta Z_i)|\big] \\ &\leq C\big(\mathbb{E}(\sum_{i=1}^{N-1}(|\tilde{\eta}_i|^2|\Delta X_i|^2))\big)^{\frac{1}{2}}\big(\mathbb{E}(\sum_{i=1}^{N-1}|D_t(\Delta Z_i)|^2)\big)^{\frac{1}{2}}.\end{aligned}$$

Thanks to Proposition 5.1, (5.20) and Proposition 2.2, we get that $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i f''_{xz}(\theta^\lambda_{t_i})(D_t \Delta Z_i)] = O(1)$. Analogously, using (5.20-3.9-5.9), we obtain $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i (D_t \Delta X_i) f''_{xz}(\theta^\lambda_{t_i})\Delta Z_i] = O(1)$. It remains to demonstrate that $\big|\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i D_t(f''_{xz}(\theta^\lambda_{t_i}))\Delta Z_i]\big| = O(1)$. We have

$$\begin{aligned}&D_t(f''_{xz}(\theta^\lambda_{t_i})) = f'''_{xzx}(\theta^\lambda_{t_i})(\lambda D_t X^N_{t_i} + (1-\lambda)D_t X_{t_i}) \\ &+ f'''_{xzy}(\theta^\lambda_{t_i})(\lambda D_t Y^N_{t_{i+1}} + (1-\lambda)D_t Y_{t_i}) + f'''_{xzz}(\theta^\lambda_{t_i})(\lambda D_t Z^N_{t_i} + (1-\lambda)D_t Z_{t_i}).\end{aligned}$$

The most difficult term to bound among these three ones is the one which contains $D_t Z^N_{t_i}$. If we write $\lambda D_t Z^N_{t_i} + (1-\lambda)D_t Z_{t_i} = \lambda D_t(\Delta Z_i) + D_t Z_{t_i}$, we obtain

$$\begin{aligned}&\big|\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i f'''_{xzz}(\theta^\lambda_{t_i})\lambda D_t(\Delta Z_i)\Delta Z_i]\big| \\ &\leq C\big(\mathbb{E}(\sum_{i=1}^{N-1}|D_t(\Delta Z_i)|^2)\big)^{\frac{1}{2}}\big(\mathbb{E}(\sum_{i=1}^{N-1}(|\Delta X_i|^2|\tilde{\eta}_i|^2|\Delta Z_i|^2))\big)^{\frac{1}{2}}, \\ &\leq C\big(\mathbb{E}(\sum_{i=1}^{N-1}|D_t(\Delta Z_i)|^2)\big)^{\frac{1}{2}}\big(\mathbb{E}(\sum_{i=1}^{N-1}|\Delta Z_i|^2)^2\big)^{\frac{1}{4}}\big(\mathbb{E}(\max_{0\leq i\leq N}|\tilde{\eta}_i|^4 \max_{0\leq i\leq N}|\Delta X_i|^4)\big)^{\frac{1}{4}},\end{aligned}$$

Applying Proposition 5.1, (3.9), Proposition 2.2 and (4.13) (with $\tilde{\eta}$) lead to $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i f'''_{xz}(\theta^\lambda_{t_i})\lambda D_t(\Delta Z_i)\Delta Z_i] = O(1)$. Proposition 2.2, (3.9), (5.6), (5.7), (5.8) and (5.20) enable us to prove that the others terms of $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i D_t(f''_{xz}(\theta^\lambda_{t_i}))\Delta Z_i]$ are $O(1)$. $\square$

Part II

Some technical results

In this part, we state some technical results concerning the transition density function of a diffusion process X, the regularity of the solution of a linear PDE, and the convergence rate for the transition density function of the Euler scheme. These results will be useful in Parts III and IV.
In the first chapter, we are interested in bounding both the transition density function of X (denoted $p(t,x;s,y)$) and $\nu^t(s,y) := \int_{\mathbb{R}^d} \rho(x)p(t,x;s,y)dx$, where ρ is a weight function specified later. $p(t,x;s,y)$ is the fundamental solution of a linear parabolic PDE. Several results on bounds for the fundamental solution of a linear parabolic PDE and for its derivatives can be found in the literature (See Aronson [2], Ladyzenskaja et al. [64] and Friedman [28]). We use these bounds to prove a norm equivalence result (see Proposition 6.12). This Proposition is quite similar to Bally and Matoussi [5], Proposition 5.1 (recalled in Proposition 6.10), except for the assumptions on ρ (Proposition 6.12 is proved for $\rho(x) = e^{-\mu|x|}$, whereas Bally and Matoussi [5], Proposition 5.1 is true for a more general function ρ) and on the coefficients of the diffusion process. Beside that, Corollary 6.13 and Proposition 6.14 give an upper bound for $\frac{\nu^t(s,y)}{\nu^t(t_0,x)}$, when (t_0,x) and (s,y) are in the same neighbourhood.

The second chapter deals with the regularity of u, the solution of the parabolic PDE $(\partial_t + \mathcal{L})u(t,x) + f(t,x) = 0$, with a null terminal condition. In particular, we recall Bensoussan and Lions [13], Theorem 6.12, page 130 (see Theorem 7.1), which asserts that $\|\partial_t u\| + \|u\| + \|\partial_x u\| + \left\|\partial_x^2 u\right\|$ is bounded by $\|f\|$ in $\mathbb{L}^p(0,T;W^{0,p,\mu})$ (see below the definition of $\mathbb{L}^p(0,T;W^{0,p,\mu})$). Then, we state in Proposition 7.4 that $\|\partial_t u\| + \|u\| + \|\partial_x u\| + \left\|\partial_x^2 u\right\|$ is bounded by $\|f\|$ in $H^\mu_{\beta,X}$ (see below the definition of $H^\mu_{\beta,X}$). We also prove in Proposition 7.3, that $\partial_x u(t,x)$ is Hölder continuous w.r.t. t of order $\frac{1}{2}$.

The last chapter of this part is devoted to the study of the convergence rate of the density of the Euler scheme. We approximate the process X by its Euler scheme X^N and we study the difference between $p(t,x;s,y)$ and $p^N(t,x;s,y)$, where p^N is the transition density of X^N. Various expansions w.r.t. N for $p(t,x;s,y) - p^N(t,x;s,y)$ can be found in the literature. (See Bally and Talay [8], Konakov and Mammen [60] or Guyon [46]). Guyon [46] makes precise the way the expansion of $p(t,x;s,y) - p^N(t,x;s,y)$ explodes when s goes to t. Theorem 8.1 states an upper bound for $p(t,x;s,y) - p^N(t,x;s,y)$, and Corollary 8.2 proves that $p(t,x;T,x) \sim p^N(t,x;T,x)$ when $T \to t$.

Chapter 6

Linear parabolic PDE and diffusion process

This chapter deals with linear parabolic PDEs and the transition density function of a diffusion process. It is organised as follows. First, we study the properties of the fundamental solution of a linear parabolic PDE. We recall some results on the bounds for the fundamental solution and for its derivatives w.r.t. time and space. These results come from Ladyzenskaja et al. [64], Friedman [28], and Aronson [2]. Secondly, we deal with the transition density of a diffusion process. This transition density function is the fundamental solution of a linear PDE. Then, we state some properties of the bounds for the transition density function and for its derivatives.

6.1 Introduction

Let us introduce some notations commonly used in the sequel.

- For any matrix M in $\mathbb{R}^d \otimes \mathbb{R}^q$ and any vector V in $\mathbb{R}^d$, we define $|M|^2$ and $|V|^2$: $|M|^2 = \sum_{i=1}^d \sum_{j=1}^q |M_{ij}|^2$, $|V|^2 = \sum_{i=1}^d |V_i|^2$.
- Let $C^k_{l,b}(\mathbb{R}^p, \mathbb{R}^q)$ be the set of C^k functions from $\mathbb{R}^p$ to $\mathbb{R}^q$ with continuous and uniformly bounded derivatives up to order k. The functions themselves don't need to be bounded.
- Let $C_b^{k,l}$ be the set of continuously differentiable functions $\phi : (t, x) \in [0, T] \times \mathbb{R}^d$ with continuous and uniformly bounded derivatives w.r.t. t (resp. w.r.t. x) up to order k (resp. up to order l).
- For any function $v : \mathbb{R}^d \to \mathbb{R}$ and any x in $\mathbb{R}^d$, we define $\partial_x v, \partial_x^2 v(x)$ in the following way: $\partial_x v(x) = (\partial_{x_1} v(x), \cdots \partial_{x_d}, v(x))$, and $\partial_x^2 v(x)$ is a $d \times d$ matrix whose components are $(\partial_x^2 v(x))_{i,j} = \partial^2_{x_i, x_j} v(x),\ 1 \le i, j \le d$.
- Let $(F, \|\cdot\|_F)$ be a Banach space. We define $\mathbb{L}^2(0, T; F)$ the space of functions ϕ from $[0, T]$ into F s.t

$$\|\phi\|^p_{\mathbb{L}^p(0,T;F)} = \int_0^T \|\phi(t)\|_F^p \, dt < \infty.$$

- Let $W^{m,p,\mu}$, $m \leq 2$ define the space of functions $v : \mathbb{R}^d \to \mathbb{R}$ s.t.

$$\|v\|_{W^{m,p,\mu}} = \left(\sum_{k \leq m} \int_{\mathbb{R}^d} e^{-\mu|x|} |\partial^k v(x)|^p dx \right)^{\frac{1}{p}} < \infty.$$

We set $H^{m,\mu} = W^{m,2,\mu}$. For $m = 0$, we set $H^\mu = H^{0,\mu}$.
In particular, we get for $\phi : [0,T] \times \mathbb{R}^d \to \mathbb{R}$

$$\|\phi\|^p_{\mathbb{L}^p(0,T;W^{0,p,\mu})} = \int_0^T dt \int_{\mathbb{R}^d} dx e^{-\mu|x|} |\phi(t,x)|^p.$$

- For any $m \leq 2, \beta > 0$ and $\mu > 0$, let $H^{m,\mu}_\beta$ define the space of functions $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\|u\|^2_{H^{m,\mu}_\beta} = \int_0^T e^{\beta s} \|u(s,\cdot)\|^2_{H^{m,\mu}} ds = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \sum_{k \leq m} |\partial^k_x u(s,x)|^2 dx ds < \infty.$$

- For any $m \leq 2, \beta > 0, \mu > 0$ and any diffusion process $X_s,\ 0 \leq s \leq T$ starting from x at time 0, let $H^{m,\mu}_{\beta,X}$ define the space of functions $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\|u\|^2_{H^{m,\mu}_{\beta,X}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \sum_{k \leq m} \mathbb{E}|\partial^k_x u(s, X^x_s)|^2 dx ds < \infty.$$

Let us consider the following linear parabolic PDE on $[0,T] \times \mathbb{R}^d$:

$$(-\partial_t + \mathcal{L}_{(t,x)}) u(t,x) = 0, \tag{6.1}$$

where the second order differential operator $\mathcal{L}_{(t,x)}$ is defined by

$$\mathcal{L}_{(t,x)} u(t,x) = \sum_{i,j} a_{ij}(t,x) \partial^2_{x_i x_j} u(t,x) + \sum_i b_i(t,x) \partial_{x_i} u(t,x) + c(t,x) u(t,x), \tag{6.2}$$

and $a_{ij}(t,x) = \frac{1}{2}[\sigma\sigma^*]_{ij}(t,x)$.

Definition 6.1 (Fundamental solution)**.** [Friedman [29] p. 141]
A fundamental solution of the parabolic operator $\mathcal{L}_{t,x} - \partial_t$ on $[0,T] \times \mathbb{R}^d$ is a function $\Gamma(t,x;\tau,\xi)$ defined for all $(t,x) \in [0,T] \times \mathbb{R}^d$ and $(\tau,\xi) \in [0,T] \times \mathbb{R}^d$, with $t > \tau$, which satisfies that for any continuous function $f(x)$ with compact support, the function

$$u(t,x) = \int_{\mathbb{R}^d} \Gamma(t,x;\tau,\xi) f(\xi) d\xi$$

satisfies

- $(-\partial_t + \mathcal{L}_{(t,x)}) u(t,x) = 0$, if $x \in \mathbb{R}^d,\ \tau < t \leq T$,
- $u(t,x) \to f(x)$ if $t \searrow \tau$.

Definition 6.2 (Ellipticity condition)**.** We say that the operator $\mathcal{L}$, defined by (6.2), is **uniformly elliptic** on $[0,T] \times \mathbb{R}^d$ if there exist two positive constants σ_0, σ_1 s.t., for any vector ξ and any $(t,x) \in [0,T] \times \mathbb{R}^d$

$$\sigma_0 |\xi|^2 \leq \sum_{i,j=1}^{d} [\sigma\sigma^*]_{i,j}(t,x)\xi_i\xi_j \leq \sigma_1 |\xi|^2.$$

Remark 6.3. To be rigorous in the previous definition, we should say that $\mathcal{L}$ is **uniformly parabolic** (and not **uniformly elliptic**). By misnomer, researchers in probability commonly use the word "elliptic", whereas the right term "parabolic" is used in the literature on PDEs.

6.2 Properties of the fundamental solution of a linear parabolic PDE

We recall some classical results on the regularity of the fundamental solution of (6.1), denoted $\Gamma(t,x;\tau,\xi)$, and of its derivatives.

6.2.1 Bounds for the fundamental solution of (6.1)

This result, coming from Aronson [2], gives both upper and lower bounds for Γ. These bounds are proportional to a Gaussian kernel. We have the following proposition

Proposition 6.4 (Aronson [2])**.** *Assume that the coefficients a, b, c are bounded measurable functions of $(t,x) \in [0,T] \times \mathbb{R}^d$ and that σ satisfies ellipticity condition (Definition 6.2). There exist positive constants α_1, α_2, K s.t.*

$$K^{-1}\gamma_1(x-\xi, t-\tau) \leq \Gamma(t,x;\tau,\xi) \leq K\gamma_2(x-\xi,t-\tau), \tag{6.3}$$

for all $(t,x), (\tau,\xi) \in [0,T] \times \mathbb{R}^d$ with $t > \tau$, where $\gamma_i(t,x)$ is the fundamental solution of $\frac{\alpha_i}{2}\Delta u - \partial_t u = 0$, for $i = 1,2$. We have $\gamma_i(t,x) = \frac{1}{(2\pi\alpha_i t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{2\alpha_i t}}$. The constant K depends only on σ_0, σ_1, d, T and the suprema of the coefficients a, b, c. The constants α_1, α_2 depend on σ_0, σ_1 and d.

6.2.2 Bounds for the derivatives of $\Gamma(t,x;\xi,\tau)$ w.r.t. t, x, ξ.

The first result given here states a bound for $\partial_t^r \partial_x^s \Gamma(t,x;\tau,\xi)$ when $2r + s \leq 2$, with Hölder type assumptions on the coefficients. We refer to Ladyzenskaja et al. [64] pages 376-377. The second result has been established in Friedman [28], page 261. It gives a bound for $\partial_x^{m+a}\partial_\xi^b \Gamma(t,x;\tau,\xi), \partial_x^m \partial_\xi^b \Gamma(t,x+\xi;\tau,\xi)$ when $|a| + |b| \leq r$, r a positive integer and $m = 0,1$. Some regularity on the derivatives of the coefficients up to order r are needed.

Proposition 6.5 (Ladyzenskaja et al. [64], pages 376-377)**.** *Assume that $\mathcal{L}$ is uniformly elliptic (see Definition 6.2) and that the coefficients a, b, c are Hölder continuous of order*

α in x and $\frac{\alpha}{2}$ in t. There exist two positive constants c (depending on σ_0, σ_1) and C (depending on $\sigma_0, \sigma_1, d, \alpha$), s.t.

$$|\partial_t^r \partial_x^s \Gamma(t, x; \tau, \xi)| \leq C(t-\tau)^{-\frac{(d+2r+s)}{2}} e^{-c\frac{|x-\xi|^2}{t-\tau}}, \text{ where } 2r+s \leq 2,\ t > \tau.$$

Proposition 6.6 (Friedman [28]). *Assume that $\partial_x^k a(t,x)$, $\partial_x^k b(t,x)$, $\partial_x^k c(t,x)$ $(0 \leq |k| \leq r, r \in \mathbb{N}^*)$, exist and are bounded continuous functions of (t,x) in $[0,T] \times \mathbb{R}^d$. We also assume that $\mathcal{L}$ is uniformly elliptic. Then, for all $0 \leq |a| + |b| \leq r, m = 0,1$, $\partial_x^{m+a} \partial_\xi^b \Gamma(t,x;\tau,\xi)$ exist and are continuous functions. Moreover, there exist c, C two positive constants depending on σ_0, σ_1, d and on the bounds of $\partial_x^k a, \partial_x^k b, \partial_x^k c$ $(k \leq 2)$ s.t.*

$$|\partial_x^{m+a} \partial_\xi^b \Gamma(t,x;\tau,\xi)| \leq \frac{C}{(t-\tau)^{(|m|+|a|+|b|+d)/2}} \exp\left(-c\frac{|x-\xi|^2}{t-\tau}\right),$$
$$|\partial_x^{m} \partial_\xi^b \Gamma(t,x+\xi;\tau,\xi)| \leq \frac{C}{(t-\tau)^{(|m|+d)/2}} \exp\left(-c\frac{|x|^2}{t-\tau}\right),$$

6.3 Properties of the transition density of a diffusion process

Let us introduce the d-dimensional diffusion process X of generator $\mathcal{L}$. $X_s^{t,x}, s \in [t,T]$ is the solution of the following SDE

$$dX_s^{t,x} = b(s, X_s^{t,x})ds + \sigma(s, X_s^{t,x})dW_s, \quad X_t^{t,x} = x. \tag{6.4}$$

Let $p(t,x,s,A)$ denote the transition probability function

$$p(t,x,s,A) = \mathbb{P}_{t,x}(X_s \in A) = \mathbb{P}(X_s^{t,x} \in A)$$

of the Markov process solution of (6.4). The following theorem, from Friedman [29], gives us assumptions under which the transition probability function has density.

Theorem 6.7 (Friedman [29], page 149). *Assume that σ satisfies the ellipticity condition, and that*

Hypothesis 6.1 *The functions a_{ij}, b_i are bounded on $[0,T] \times \mathbb{R}^d$ and uniformly Lipschitz continuous on compact subsets of $[0,T] \times \mathbb{R}^d$.*
The functions a_{ij} are Hölder continuous w.r.t. x, uniformly in $(t,x) \in [0,T] \times \mathbb{R}^d$.

Then, the transition probability function of the solution of the stochastic differential equation (6.4) has a density, i.e.,

$$\mathbb{P}(X_s^{t,x} \in A) = \int_A p(t,x;s,y)dy, \quad (t < s)$$

for any Borel set A, and $p(t,x;s,y)$ is the fundamental solution of $\mathcal{L}_{t,x} + \partial_t$, which means that for any continuous function $f(x)$ with compact support, the function

$$u(t,x) = \int_{\mathbb{R}^d} p(t,x;s,y) f(y)dy$$

satisfies

- $(\partial_t + \mathcal{L}_{(t,x)})u(t,x) = 0,$ *if* $x \in \mathbb{R}^d,\ t < s \leq T,$
- $u(t,x) \to f(x)$ *if* $t \nearrow s.$

The density function of the transition probability function is called the *transition density function.* It satisfies backward and forward equations.

6.3.1 Backward and forward parabolic equations for the transition density function

We refer to Friedman [29], Chapter 6.5 and to Bensoussan and Lions [13], page 133 for more details. We deduce, from the above definition of the fundamental solution, that $p(t,x;s,y)$ satisfies in (t,x) the *backward parabolic equation*

$$\begin{cases} \partial_t p(t,x;s,y) + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(t,x)\partial^2_{x_i x_j} p(t,x;s,y) + \sum_{i=1}^d b_i(t,x)\partial_{x_i} p(t,x;s,y) = 0. \\ p(s,x;s,y) = \delta(x-y). \end{cases} \tag{6.5}$$

Under stronger hypotheses, $p(t,x;s,y)$ satisfies in (s,y) a *forward parabolic equation.* (See Bensoussan and Lions [13], page 134 for a proof.)

Hypothesis 6.2

1. *σ is elliptic,*
2. *The functions $a_{ij}, \partial_{x_i} a_{ij}, \partial^2_{x_i x_j} a_{ij}, b_i$ are bounded on $[0,T]\times\mathbb{R}^d$ and are Hölder continuous (of order α) w.r.t. x uniformly in $(t,x) \in [0,T]\times\mathbb{R}^d$.*

Then, $p(t,x;s,y)$ satisfies in (s,y) the forward equation

$$\begin{cases} -\partial_s p(t,x;s,y) + \frac{1}{2}\sum_{ij=1}^d \partial^2_{y_i y_j}[a_{ij}(s,y)p(t,x;s,y)] - \sum_{i=1}^d \partial_{y_i}[b_i(s,y)p(t,x;s,y)] = 0. \\ p(t,x;t,y) = \delta(x-y). \end{cases} \tag{6.6}$$

6.3.2 Bounds for the transition density function

Since $p(t,x;s,y), s > t$ is the fundamental solution of the operator $\mathcal{L}_{t,x} + \partial_t$, we get that $\Gamma_0(t,x;\tau,y) := p(T-t,x;T-\tau,y),\ \tau < t$, satisfies

$$-\partial_t \Gamma_0(t,x;\tau,y) + \sum_{i=1}^d b_i(T-t,x)\partial_{x_i}\Gamma_0(t,x;\tau,y) + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(T-t,x)\partial^2_{x_i x_j}\Gamma_0(t,x;\tau,y) = 0$$

and $\int_{\mathbb{R}^d}\Gamma_0(t,x;\tau,y)f(y)dy \to f(x)$ when $t \searrow \tau$. Hence, $\Gamma_0(t,x;\tau,y),\ \tau < t$ is the fundamental solution of $-\partial_t + \mathcal{L}_{T-t,x}$. Applying Propositions 6.4, 6.5 and 6.6 to $\Gamma_0(T-t,x;T-\tau,y)$ enables us to state results similar to Propositions 6.4, 6.5 and 6.6 for p.

By using Proposition 6.4 and the formula of γ_i, we can rewrite the inequality (6.3). We get

Proposition 6.8. *Assume that the coefficients a, b are bounded measurable functions of $(t,x) \in [0,T] \times \mathbb{R}^d$ and that σ is elliptic. There exist positive constants K, α_0, α_1 s.t.*

$$\frac{K^{-1}}{(2\pi\alpha_1(s-t))^{\frac{d}{2}}} e^{-\frac{|x-y|^2}{2\alpha_1(s-t)}} \leq p(t,x;s,y) \leq K \frac{1}{(2\pi\alpha_2(s-t))^{\frac{d}{2}}} e^{-\frac{|x-y|^2}{2\alpha_2(s-t)}}. \tag{6.7}$$

The constant K depends only on σ_0, σ_1, d, T and the suprema of the coefficients a, b. The constants α_0, α_1 depend on σ_0, σ_1 and d.

6.3.3 Bounds for the derivatives of the transition density function w.r.t. t, x, y.

As the first two results (6.8) and (6.9) of the following proposition are exactly the same as for Γ, we only briefly recall them. The third one (6.10) gives an upper bound for $\partial_t \partial_x p(t,x;s,y)$.

Proposition 6.9.

1. *Assume that $\mathcal{L}$ is uniformly elliptic and that the coefficients a, b are Hölder continuous of order α in x and $\frac{\alpha}{2}$ in t. There exist two positive constants c (depending on σ_0, σ_1) and C (depending on $\sigma_0, \sigma_1, d, \alpha$), s.t.*

$$|\partial_t^r \partial_x^s p(t,x;s,y)| \leq C(s-t)^{-\frac{(d+2r+s)}{2}} e^{-c\frac{|x-y|^2}{s-t}}, \text{ where } 2r+s \leq 2,\ s > t. \tag{6.8}$$

2. *Assume that $\partial_x^k a(t,x)$, $\partial_x^k b(t,x)$, $(0 \leq |k| \leq 2)$, exist and are bounded continuous functions on $[0,T] \times \mathbb{R}^d$. We also assume that $\mathcal{L}$ is uniformly elliptic. Then, $\partial_x^{m+a} \partial_y^b p(t,x;s,y)$ exist and are continuous functions for all $0 \leq |a| + |b| \leq 2, m = 0, 1$. Moreover, there exist c, C two positive constants depending on σ_0, σ_1, d and on the bounds of $\partial_x^k a, \partial_x^k b, \partial_x^k c$ $(k \leq 2)$ s.t.*

$$|\partial_x^{m+a} \partial_y^b p(t,x;s,y)| \leq \frac{C}{(s-t)^{(|m|+|a|+|b|+d)/2}} \exp\left(-c\frac{|y-x|^2}{s-t}\right). \tag{6.9}$$

and

$$|\partial_{tx}^2 p(t,x;s,y)| \leq C(s-t)^{-\frac{d+3}{2}} e^{-c\frac{|x-y|^2}{s-t}}. \tag{6.10}$$

Proof of (6.10). We can differentiate (6.5) w.r.t. x_k, for $k = 1, \cdots, d$. Then,

$$\begin{aligned}\partial_{x_k} \partial_t p(t,x;s,y) = &-\frac{1}{2} \sum_{i,j=1}^d \partial_{x_k} a_{ij}(t,x) \partial_{x_i x_j}^2 p(t,x;s,y) - \frac{1}{2} \sum_{i,j=1}^d a_{ij}(t,x) \partial_{x_i x_j x_k}^3 p(t,x;s,y) \\ &- \sum_{i=1}^d \partial_{x_k} b_i(t,x) \partial_{x_i} p(t,x;s,y) - \sum_{i=1}^d b_i(t,x) \partial_{x_i x_k}^2 p(t,x;s,y).\end{aligned}$$

Since $|\partial_{tx}^2 p(t,x;s,y)| \leq \sum_{k=1}^d |\partial_{tx_k}^2 p(t,x;s,y)|$ and $\partial_x^k a(t,x)$, $\partial_x^k b(t,x)$, $(0 \leq |k| \leq 2)$ exist and are bounded continuous functions, we combine the previous equality with (6.9) (with $|b| = 0, m = 1$ and $|a| = 0, 1, 2$) to get (6.10). □

6.4 A norm equivalence result

Before giving a norm equivalence result which will be useful in the sequel, we recall a norm equivalence result which has been proved by Bally and Matoussi [5]. The authors introduce $\rho(x) := \exp(F(x))$, a weight function, where $F : \mathbb{R}^d \to \mathbb{R}$ is a continuous function s.t. there exists a constant $R > 0$ s.t. the restriction of F to $\{|x| > R\}$ is $C^2_{l,b}$. (The case $\rho = 1$ had already been treated by Barles and Lesigne [9], and the case $\rho(x) = (1 + |x|^q)$ had been handed by Kunita [61].)

Proposition 6.10 (Bally and Matoussi [5]). *Assume $b \in C^2_{l,b}(\mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C^3_{l,b}(\mathbb{R}^d, \mathbb{R}^{d\times d})$. Let ρ be a weight function which satisfies the above properties. There exist two constants $c > 0$ and $C > 0$ s.t. for every $\Psi \in L^1((0,T) \times \mathbb{R}^d, \rho(x)dx \otimes dt)$*

$$c\int_{\mathbb{R}^d}\int_t^T |\Psi(s,x)|ds\rho(x)dx \le \int_{\mathbb{R}^d}\int_t^T \mathbb{E}(|\Psi(s,X_s^{t,x})|)ds\rho(x)dx$$
$$\le C\int_{\mathbb{R}^d}\int_t^T |\Psi(s,x)|ds\rho(x)dx.$$

The constants c and C depend on T, ρ, the suprema of the first derivative of b and the suprema of the first and second derivatives of σ.

The inequality we establish now in Proposition 6.12 is weaker than the one of Proposition 6.10, since we are in the particular case $\rho(x) = e^{-\mu|x|}, \mu > 0$. We also assume different hypotheses on b, σ. Before stating this Proposition, for the sake of clearness, we introduce the following notation

Definition 6.11. For any $s, t \in [0,T]$ and any $x, y \in \mathbb{R}^d$ such that $t < s$ we define

$$\nu^t(s,y) := \int_{\mathbb{R}^d} e^{-\mu|x|}p(t,x;s,y)dx,$$

where μ is a positive constant.

In the following, μ is a positive constant.

Proposition 6.12 (Norm equivalence). *Assume that the coefficients σ, b are bounded measurable functions of $(t,x) \in [0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition (Definition 6.2). We also assume that (6.4) has a unique weak solution $(X^{t,x}, W)$. There exist two constants $c > 0$ and $C > 0$ depending on $T, d, \mu, K, \alpha_1, \alpha_2$ (see Proposition 6.4 for the definitions of K, α_1, α_2), and two constants $c_i, i = 1, 2$ depending on μ, d, α_i s.t. $\forall y \in \mathbb{R}^d$*

$$\frac{1}{2^dK}e^{-\mu|y|}e^{c_1(s-t)} \le \nu^t(s,y) \le 2^dKe^{c_2(s-t)}e^{-\mu|y|}. \tag{6.11}$$

Moreover, for every $\Psi \in L^1((0,T) \times \mathbb{R}^d, e^{-\mu|x|}dx \otimes dt)$

$$c\int_{\mathbb{R}^d}\int_t^T |\Psi(s,x)|dse^{-\mu|x|}dx \le \int_{\mathbb{R}^d}\int_t^T \mathbb{E}(|\Psi(s,X_s^{t,x})|)dse^{-\mu|x|}dx$$
$$\le C\int_{\mathbb{R}^d}\int_t^T |\Psi(s,x)|dse^{-\mu|x|}dx. \tag{6.12}$$

The following Corollary and the following Proposition give an upper bound for $\frac{\nu^t(s,y)}{\nu^t(t_0,x)}$, when (s,y) belongs to a neighbourhood of (t_0,x). Corollary 6.13 ensues from Proposition 6.12. Under stronger hypotheses than in Corollary 6.13, Proposition 6.14 states a more accurate bound for $\frac{\nu^t(s,y)}{\nu^t(t_0,x)}$.

Corollary 6.13. *Under the assumptions of Proposition 6.12. For any s, t_0 belonging to $[t,T]$, for any x, y belonging to $\mathbb{R}^d$ and for any h_t, h_x satisfying $|s-t_0| \leq h_t$ and $|x-y| \leq h_x$, there exists a constant $C > 0$ depending on $d, \mu, T, K, \alpha_1, \alpha_2$ and a constant c_2 depending on μ, d, α_2 such that*

$$\frac{1}{\sqrt{C}} e^{-\mu h_x} \nu^t(s,y) \leq \nu^t(s,x) \leq \sqrt{C} e^{c_2 h_t} \nu^t(t_0,x).$$

Proposition 6.14. *Assume that σ satisfies the ellipticity condition, a, b are respectively C_b^3, C_b^2 functions in space, and $a, \partial_x a, \partial_x^2 a, b, \partial_x b$ are C_b^1 function in time. For any s, t_0 belonging to $[t,T]$, for any x, y belonging to $\mathbb{R}^d$ and for any h_t, h_x satisfying $|s-t_0| \leq h_t$ and $|x-y| \leq h_x$, there exists a constant $C > 0$ depending on $T, \mu, d, K\alpha_1, \alpha_2$ (see Proposition 6.4 for the definition of K, α_1, α_2) and on the bounds of a, b and of the first derivatives of $a, \partial_x a, \partial_x^2 a, b, \partial_x b$ in time and space such that*

$$\nu^t(s,y) \leq (1 + Ch_x)\left(1 + C\sqrt{h_t}\right) \nu^t(t_0,x).$$

Before proving Proposition 6.12, let us state the following Lemma.

Lemma 6.15. *Let I denote $\int_{\mathbb{R}^d} dx e^{-\mu|x|} \frac{1}{(s-t)^{\frac{d}{2}}} \exp\left(-c\frac{|x-y|^2}{s-t}\right)$, where $c > 0$. For any $s, t \in [0,T]$ and any $x, y \in \mathbb{R}^d$ such that $t < s$, the following assertion holds*

$$\frac{1}{2^d}\left(\frac{\pi}{c}\right)^{d/2} e^{-\frac{d\mu^2}{4c}(s-t)} e^{-\mu|y|} \leq I \leq 2^d \left(\frac{\pi}{c}\right)^{d/2} e^{\frac{d\mu^2}{4c}(s-t)} e^{-\mu|y|}.$$

Proof of Lemma 6.15. First, we prove the r.h.s. of the previous inequality. Using a change of variables in I yields

$$I = \left(\frac{\pi}{c}\right)^{d/2} \frac{1}{2\pi(s-t)^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{-\mu|\frac{1}{\sqrt{2c}}z+y|} e^{-\frac{|z|^2}{2(s-t)}} dz = \left(\frac{\pi}{c}\right)^{d/2} \mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}+y|}]. \tag{6.13}$$

Furthermore, $\mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}+y|}] \leq e^{-\mu|y|}\mathbb{E}[e^{\mu\frac{1}{\sqrt{2c}}|W_{s-t}|}]$. The components of W_{s-t} are independent, then $\mathbb{E}[e^{\mu\frac{1}{\sqrt{2c}}|W_{s-t}|}] \leq \left(\mathbb{E}[e^{\mu\frac{1}{\sqrt{2c}}|W^1_{s-t}|}]\right)^d \leq 2^d(\mathbb{E}[\mathrm{ch}(\mu\frac{1}{\sqrt{2c}}W^1_{s-t})])^d \leq 2^d e^{\frac{d\mu^2(s-t)}{4c}}$. The last term being bounded, we get the upper bound for I.
Concerning the lower bound, we use (6.13). Moreover, $\mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}+y|}] \geq e^{-\mu|y|}\mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}|}]$ and $\mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}|}] \geq \frac{1}{E[e^{\mu|\frac{1}{\sqrt{2c}}W_{s-t}|}]} \geq \frac{1}{\left(E[e^{\mu|\frac{1}{\sqrt{2c}}W^1_{s-t}|}]\right)^d} \geq \frac{1}{2^d} e^{-\frac{d\mu^2(s-t)}{4c}}$. The last term being bounded from below, we get the lower bound for I. □

Proof of Proposition 6.12. We easily deduce (6.12) from (6.11) by multiplying (6.11) by $|\Psi(s,y)|$ and integrating w.r.t. y and s.
Upper bound for $\nu^t(s,y)$. Using the right hand side inequality of (6.7), we get

$$\nu^t(s,y) \leq \frac{K}{(2\pi\alpha_2(s-t))^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{-\mu|x|} e^{-\frac{|x-y|^2}{2\alpha_2(s-t)}} dx.$$

Then, using Lemma 6.15 with $c = \frac{1}{2\alpha_2}$ leads to $\nu^t(s,y) \leq K2^d e^{\frac{d\mu^2\alpha_2(s-t)}{2}} e^{-\mu|y|}$.
Lower bound for $\nu^t(s,y)$. Using the left hand side inequality of (6.7), we get

$$\nu^t(s,y) \geq \frac{1}{K(2\pi\alpha_1(s-t))^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{-\mu|x|} e^{-\frac{|x-y|^2}{2\alpha_1(s-t)}} dx.$$

Then, using Lemma 6.15 with $c = \frac{1}{2\alpha_1}$ leads to $\nu^t(s,y) \leq \frac{K}{2^d} e^{-\frac{d\mu^2\alpha_1(s-t)}{2}} e^{-\mu|y|}$. □

Proof of Corollary 6.13 . This result ensues from (6.11). From the r.h.s. of (6.11), we get $\nu^t(s,y) \leq K2^d e^{\mu h_x} e^{c_2(s-t)} e^{-\mu|x|}$. From the l.h.s., we have $e^{-\mu|x|} \leq K2^d e^{-c_1(s-t)} \nu^t(s,x)$. Then,

$$\nu^t(s,y) \leq (K2^d)^2 e^{(c_2-c_1)(s-t)} e^{\mu h_x} \nu^t(s,x).$$

A similar computation leads to $\nu^t(s,x) \leq (K2^d)^2 e^{(c_2-c_1)T} e^{c_2 h_t} \nu^t(t_0,x)$, and the result follows. □

Proof of Proposition 6.14. Let C denote a generic constant depending on $T, \mu, d, K, \alpha_1, \alpha_2$. The proof is done in the following way

- We first show that $\nu^t(s,y)$ can be written as an expectation of a function of the path up to time s of a process Y starting in y at time t. This way of writing $\nu^t(s,y)$ is essential to carry on the proof.
- Then, using this expression of $\nu^t(s,y)$ enables us to bound $\nu^t(s,y)$ by $\nu^t(t_0,y)\left(1 + C\sqrt{h_t}\right)$ and
- to conclude the proof by showing that $\nu^t(s,y) \leq \nu^t(s,x)(1 + Ch_x)$.

We write $\nu^t(s,y)$ as an expectation
We prove that $\nu^t(s,y) = \mathbb{E}_{t,y}[e^{-\mu|Y_s|} \exp(\int_t^s c(t+s-u, Y_u)du)]$, where c is a bounded function depending on $\partial_x^2 a_{ij}, \partial_x b$. To do so, we first use the forward equation (6.6) satisfied by $p(t,x;s,y)$ to get that $\nu^t(s,y)$ satisfies the following forward equation. $\forall s \in [t,T]$, we have

$$\begin{cases} -\partial_s \nu^t(s,y) + \frac{1}{2}\sum_{ij=1}^d a_{ij}(s,y)\partial^2_{y_i y_j} \nu^t(s,y) + \sum_{i=1}^d \overline{b_i}(s,y)\partial_{y_i}\nu^t(s,y) + c(s,y)\nu^t(s,y) = 0, \\ \nu^t(t,y) = e^{-\mu|y|}. \end{cases}$$

where $\overline{b_i}(s,y) = \sum_{j=1}^d \partial_{y_j} a_{ij}(s,y) - b_i(s,y)$, $c(s,y) = \frac{1}{2}\sum_{ij=1}^d \partial^2_{y_i y_j} a_{ij}(s,y) - \sum_{i=1}^d \partial_{y_i} b_i(s,y)$.
Then, we introduce $\eta^t(r,y) := \nu^t(t+s-r,y)$, with $r \in [t+s-T, s]$. $\eta^t(r,y)$ follows the backward equation

$$\begin{cases} \partial_r \eta^t(r,y) + \frac{1}{2}\sum_{ij=1}^d a_{ij}(t+s-r,y)\partial^2_{y_i y_j}\eta^t(r,y) + \sum_{i=1}^d \overline{b_i}(t+s-r,y)\partial_{y_i}\eta^t(t,y) \\ +c(t+s-r,y)\eta^t(r,y) = 0, \\ \eta^t(s,y) = e^{-\mu|y|}. \end{cases}$$

We also introduce the diffusion process Y satisfying

$$\forall u \in [r,s] \;\; dY_u = \overline{b}(t+s-u, Y_u)du + \sigma(t+s-u, Y_u)dW_u, \;\; Y_r = y.$$

Finally, to write $\eta^t(r,y)$ (and then $\nu^t(s,y)$) as an expectation, we apply the Feynman-Kac formula (see Appendix A.2.1) to the previous PDE satisfied by $\eta^t(r,y)$. We can use Theorem A.6 since a, b and $y \longmapsto e^{-\mu|y|}$ satisfy the following conditions (which correspond to the required assumptions to use the Feynman-Kac formula)

- $y \longmapsto e^{-\mu|y|}$ satisfies the polynomial growth condition (Remark A.8).
- $a, \partial_x a, b$ are Lipschitz continuous functions in x, uniformly in time (Hypothesis A.3).
- $a, \partial_x a, \partial_x^2 a, b, \partial_x b$ are bounded functions (Hypothesis A.2 and Remark A.8) and uniformly Hölder continuous in $[0,T] \times \mathbb{R}^d$ (Remark A.8).
- σ satisfies the uniform ellipticity condition (Remark A.8).

Hence, $\eta^t(r,y) = \mathbb{E}_{r,y}[e^{-\mu|Y_s|} \exp(\int_r^s c(t+s-u, Y_u)du)]$, where $r \in [t+s-T, s]$. Since $t+s-T < t$, we can choose $r = t$ in the previous equality. Furthermore, the definition of η gives $\eta^t(t,y) = \nu^t(s,y)$. Thus,

$$\nu^t(s,y) = \mathbb{E}_{t,y}\left[e^{-\mu|Y_s|} \exp\left(\int_t^s c(t+s-u, Y_u)du\right)\right], \tag{6.14}$$

with $c(s,y) = \frac{1}{2}\sum_{ij=1}^d \partial^2_{y_i y_j} a_{ij}(s,y) - \sum_{i=1}^d \partial_{y_i} b_i(s,y)$.

Proof of $\nu^t(s,y) \leq \left(1 + C\sqrt{h_t}\right)\nu^t(t_0,y)$, when $|s - t_0| \leq h_t$

We separately deal with the two cases $s > t_0$ and $s < t_0$. Let us begin with $s > t_0$. Since $\partial_x^2 a, \partial_x b$ are bounded, (6.14) enables us to bound $\nu^t(s,y)$

$$\nu^t(s,y) \leq e^{c_\infty(s-t_0)}\mathbb{E}_{t,y}[e^{-\mu|Y_s|} e^{\int_t^{t_0} c(t+s-u,Y_u)du}]. \tag{6.15}$$

Moreover, $\partial_x^2 a, \partial_x b$ are C_b^1 in time, so $c(t+s-u, Y_u) \leq c(t+t_0-u, Y_u) + K(s-t_0)$, and $e^{-\mu|Y_s|} \leq e^{-\mu|Y_{t_0}|}e^{\mu|Y_s - Y_{t_0}|}$. Then, (6.15) becomes

$$\nu^t(s,y) \leq e^{c_\infty(s-t_0)}\mathbb{E}_{t,y}[e^{-\mu|Y_{t_0}|} e^{\int_t^{t_0} c(t+t_0-u,Y_u)du} e^{\mu|Y_s - Y_{t_0}|}]. \tag{6.16}$$

By conditioning w.r.t. $\mathcal{F}_{t_0}$ in the previous inequality, the term $\mathbb{E}[e^{\mu|Y_s^{t,y} - Y_{t_0}^{t,y}|}|\mathcal{F}_{t_0}]$ appears. We compute it in the following way

$$\mathbb{E}[e^{\mu|Y_s^{t,y} - Y_{t_0}^{t,y}|}|\mathcal{F}_{t_0}] = \mathbb{E}[e^{\mu|Y_s^{t_0, Y_{t_0}^{t,y}} - Y_{t_0}^{t,y}|}|\mathcal{F}_{t_0}] = \Phi(Y_{t_0}^{t,y}),$$

where $\Phi(x) = \mathbb{E}[e^{\mu|Y_s^{t_0,x} - x|}]$. To bound $\Phi(x)$ from above, we use Proposition A.10, with $g(x) = e^x$. We get

$$\mathbb{E}[e^{\mu|Y_s^{t_0,x} - x|}] = 1 + \int_0^\infty \mathbb{P}(\mu|Y_s^{t_0,x} - x| > \epsilon)e^\epsilon d\epsilon. \tag{6.17}$$

Since $\partial_x^2 a, \partial_x b$ are uniformly bounded, using Lemma A.14 yields $\mathbb{P}(|Y_s^{t_0,x} - x| > \frac{\epsilon}{\mu}) \leq K(T)\exp(-c\frac{\epsilon^2}{s-t_0})$. Hence, (6.17) leads to

$$\begin{aligned}\mathbb{E}[e^{\mu|Y_s^{t_0,x} - x|}] &\leq 1 + K(T)\int_0^\infty e^\epsilon e^{-c\frac{\epsilon^2}{s-t_0}} d\epsilon = 1 + K(T)\frac{\sqrt{s-t_0}}{2\sqrt{2c}} e^{\frac{s-t_0}{4c}}, \\ &\leq 1 + CK(T)\sqrt{s-t_0}. \end{aligned} \tag{6.18}$$

Combining the previous inequality with (6.16) gives

$$\nu^t(s,y) \leq (1 + C(s-t_0))(1 + CK(T)\sqrt{s-t_0})\nu^t(t_0,y).$$

As $s - t_0 \leq h_t$, we get the result.
We deal with the case $s < t_0$, whose proof is similar to the previous one. We bound $\nu^t(t_0,y)$ from below.

$$\begin{aligned}\nu^t(t_0,y) &= \mathbb{E}_{t,y}[e^{-\mu|Y_{t_0}|}\exp(\int_t^{t_0} c(t+t_0-u,Y_u)du)]\\ &\geq \mathbb{E}_{t,y}[e^{-\mu|Y_s|}e^{-\mu|Y_{t_0}-Y_s|}e^{\int_t^s c(t+t_0-u,Y_u)du}e^{\int_s^{t_0} c(t+t_0-u,Y_u)du}]\\ &\geq e^{c_\infty(t_0-s)}\nu^t(s,y)\Psi(Y_s^{t,y}),\end{aligned}$$

where $\Psi(x) = \mathbb{E}[e^{-\mu|Y_{t_0}^{s,x}-x|}]$. As before, we can write $\mathbb{E}[e^{-\mu|Y_s^{t_0,x}-x|}] = 1 - \int_0^\infty e^{-\epsilon}\mathbb{P}(|Y_{t_0}^{s,x} - x| \geq \frac{\epsilon}{\mu})d\epsilon$. By doing a similar computation to (6.18), we obtain

$$\nu^t(t_0,y) \geq (1 - C(s-t_0))(1 - CK(T)\sqrt{s-t_0})\nu^t(s,y),$$

and the result follows.
Proof of $\nu^t(s,y) \leq (1 + Ch_x)\nu^t(s,x)$, when $|y-x| \leq h_x$
To do so, we apply a Taylor expansion formula to each component of the following sum

$$\nu^t(s,y) - \nu^t(s,x) = \sum_{i=1}^d \nu^t(s,\overline{x}_{i-1}) - \nu^t(s,\overline{x}_i),$$

where $\overline{x}_i = (x_1, x_2, \cdots, x_i, y_{i+1}, \cdots, y_d), \forall i \in \{1, \cdots, d\}$, and $\overline{x}_0 = y$. For the i-th component, we get

$$\nu^t(s,\overline{x}_{i-1}) - \nu^t(s,\overline{x}_i) = (y_i - x_i)\int_0^1 \partial_{x_i}\nu^t(s,\overline{x}_i^\lambda)d\lambda,$$

where $\overline{x}_i^\lambda = (x_1, \cdots, x_{i-1}, x_i + \lambda(y_i - x_i), y_{i+1}, \cdots, y_d)$. To conclude, it remains to prove that for each $i \in \{1, \cdots, d\}$, $\partial_{x_i}\nu^t(s,\overline{x}_i^\lambda) \leq C\nu^t(s,x)$.

- Assume that $|\partial_{y_i}\nu^t(s,y)| \leq Ce^{-\mu|y|}$. Then, $\partial_{x_i}\nu^t(s,\overline{x}_i^\lambda) \leq Ce^{-\mu|\overline{x}_i^\lambda|}$. Since for each $i \in \{1, \cdots, d\}, |y_i - x_i| \leq h_x$, $\partial_{x_i}\nu^t(s,\overline{x}_i^\lambda) \leq Ce^{h_x}e^{-\mu|x|}$. The l.h.s. of (6.11) (see Proposition 6.12) ensures that $\partial_{x_i}\nu^t(s,\overline{x}_i^\lambda) \leq C\nu^t(s,x)$.

- We prove $|\partial_{y_i}\nu^t(s,y)| \leq Ce^{-\mu|y|}$. By using (6.14), we get the following formula for $\partial_{y_i}\nu^t(s,y)$

$$\begin{aligned}\partial_{y_i}\nu^t(s,y) &= E[-\mu\partial_{y_i}Y_s^{t,y}\mathrm{sgn}(Y_s^{t,y})e^{-\mu|Y_s^{t,y}|}e^{\int_t^s c(t+s-u,Y_u^{t,y})du}]\\ &\quad + \mathbb{E}[e^{-\mu|Y_s^{t,y}|}e^{\int_t^s c(t+s-u,Y_u^{t,y})du}\int_t^s \partial_{y_i}c(t+s-u,Y_u^{t,y})\partial_{y_i}Y_u^{t,y}du].\end{aligned}$$

Applying Cauchy-Schwarz inequality to the previous equation yields

$$\begin{aligned}|\partial_{y_i}\nu^t(s,y)| \leq &\left(\mathbb{E}[e^{-2\mu|Y_s^{t,y}|}e^{2\int_t^s c(t+s-u,Y_u^{t,y})du}]\right)^{\frac{1}{2}} \\ &\times \left((\mathbb{E}[(\mu\partial_{y_i}Y_s^{t,y})^2])^{\frac{1}{2}} + (\mathbb{E}[(\int_t^s \partial_{y_i}c(t+s-u,Y_u^{t,y})\partial_{y_i}Y_u^{t,y}du)^2])^{\frac{1}{2}}\right).\end{aligned} \tag{6.19}$$

Since a, b are respectively C_b^3, C_b^2, $|\partial_{y_i} c| \leq C$. Moreover, since $\partial_x a, \partial_x^2 a, \partial_x b$ are continuous and bounded, $\sup_{t \leq u \leq s} \mathbb{E}|\partial_{y_i} Y_s^{t,y}|^2 \leq C$ (see Gihman and Skorohod [35], page 59). Hence, (6.19) becomes

$$|\partial_{y_i} \nu^t(s, y)| \leq C e^{c_\infty (s-t)} \mathbb{E}[e^{-2\mu |Y_s^{t,y}|}]^{\frac{1}{2}}.$$

Writing $\mathbb{E}[e^{-2\mu|Y_s^{t,y}|}] = \mathbb{E}[e^{-2\mu|Y_s^{t,y} - y + y|}]$ and applying (6.18) imply $|\partial_{y_i} \nu^t(s, y)| \leq C e^{-\mu|y|}$.

$\square$

Chapter 7

Properties of the solution of a linear parabolic PDE

This chapter deals with the regularity of u, the solution of the parabolic PDE

$$\begin{cases} (\partial_t + \mathcal{L}_{t,x})u(t,x) + f(t,x) = 0, \\ u(T,x) = 0. \end{cases} \tag{7.1}$$

In particular, we recall Theorem 6.12, page 130 of Bensoussan and Lions [13], which asserts that $\|\partial_t u\| + \|u\| + \|\partial_x u\| + \left\|\partial_x^2 u\right\|$ is bounded by $\|f\|$ in $\mathbb{L}^p(0,T; W^{0,p,\mu})$. Then, we state in Proposition 7.4 that $\|\partial_t u\| + \|u\| + \|\partial_x u\| + \left\|\partial_x^2 u\right\|$ is bounded by $\|f\|$ in $H^{\mu}_{\beta,X}$. We also prove in Proposition 7.3, that $\partial_x u(t,x)$ is Hölder continuous w.r.t. t of order $\frac{1}{2}$.

We assume

Hypothesis 7.1 (Hypotheses on f)

1. *f satisfies the following polynomial growth condition: $\exists K > 0, \lambda \geq 1$ s.t.*

$$|f(t,x)| \leq K(1 + |x|^{2\lambda}), \ \ \forall 0 \leq t \leq T, \ \forall x \in \mathbb{R}^d.$$

2. *f is Hölder continuous on $[0,T] \times \mathbb{R}^d$.*

We also assume that the coefficients of $\mathcal{L}$ satisfy the following hypothesis

Hypothesis 7.2 (Hypotheses on σ, b)

1. *σ is uniformly elliptic,*
2. *a_{ij}, b_i are bounded functions, and uniformly Hölder continuous on $[0,T] \times \mathbb{R}^d$,*
3. *a_{ij}, b_i are continuous Lipschitz functions in space, uniformly in time.*

As recalled in Appendix A.2.1 (see Theorem A.6 and Remark A.8), we can use the Feynman-Kac formula under Hypotheses 7.1, 7.2 to write $u(t,x) = \mathbb{E}_{t,x}[\int_t^T f(s,X_s)ds]$, where X is the solution of the following d-dimensional stochastic differential equation (SDE) $dX_s = b(s,X_s)ds + \sigma(s,X_s)dW_s$.

7.1 Regularity result in $\mathbb{L}^p(0,T;W^{0,p,\mu})$

We recall Theorem 6.12 page 130 of Bensoussan and Lions [13], with a null terminal condition. Consider the following PDE

$$\begin{cases} (\partial_t + \mathcal{L}_{t,x})v(t,x) + c(t,x)v(t,x) + f(t,x) = 0, \\ v(T,x) = 0. \end{cases} \tag{7.2}$$

which is a bit more general than (7.1), since the term $c(t,x)v(t,x)$ has been added.

Theorem 7.1 (Bensoussan and Lions [13]). *Assume*

- $a, b, c \in C^1(]0,T[\times\mathbb{R}^d) \cap \mathbb{L}^\infty(]0,T[\times\mathbb{R}^d)$,
- $c(t,x) \geq C_0 > 0$,
- σ *is uniformly elliptic,*
- $(\partial_{x_k}\sigma)_{k=1\cdots d}, \partial_t\sigma$ *are bounded.*

Assume $f \in \mathbb{L}^p(0,T;W^{0,p,\mu}) \cap \mathbb{L}^2(0,T;H^\mu)$. *Then,* $\forall p \geq 2$, *the solution* v *of (7.2) is in* $\mathbb{L}^p(0,T;W^{2,p,\mu})$ *and* $\partial_t v \in \mathbb{L}^p(0,T;W^{0,p,\mu})$. *Furthermore, we have*

$$\|\partial_t v\|_{\mathbb{L}^p(0,T;W^{0,p,\mu})} + \|v\|_{\mathbb{L}^p(0,T;W^{2,p,\mu})} \leq C\,\|f\|_{\mathbb{L}^p(0,T;W^{0,p,\mu})}\,. \tag{7.3}$$

Remark 7.2. The second assumption concerning $c(t,x)$ is not restrictive. Assume v satisfies (7.2) and the coefficients a, b follow the preceding assumptions. We also assume $c(t,x) \geq C_0$, without any assumption on C_0. We get a result similar to Theorem 7.1 for u, solution of (7.1. By studying $u(t,x) = v(t,x)e^{-C_1 t}$, we show that u solves the following PDE

$$\begin{cases} (\partial_t + \mathcal{L}_{t,x})u(t,x) + (c(t,x) + C_1)u(t,x) + e^{-C_1 t}f(t,x) = 0, \\ u(T,x) = 0. \end{cases}$$

By choosing C_1 s.t. $C_1 > -C_0$, we can apply Theorem 7.1 to u and get $\|\partial_t u\|_{\mathbb{L}^p(0,T;W^{0,p,\mu})} + \|u\|_{\mathbb{L}^p(0,T;W^{2,p,\mu})} \leq C_u\,\|f\|_{\mathbb{L}^p(0,T;W^{0,p,\mu})}$. From this inequality, we deduce that v satisfies (7.3), where $C = C_u e^{C_1 T}$.

7.2 Hölder continuity of $\partial_x u(t,x)$ w.r.t. t of order $\frac{1}{2}$.

The following proposition states a regularity result w.r.t t of the derivative $\partial_x u(t,x)$, where u is the solution of (7.1).

Proposition 7.3. *Let* u *be the solution of (7.1). Assume Hypotheses 7.1, 7.2, that* σ *and* b *are in* C_b^2 *in space and that* f *is bounded. Then, for all* $t, t' \in [0,T]$, *we have*

$$|\partial_x u(t',x) - \partial_x u(t,x)| \leq C\,\|f\|_\infty \sqrt{|t'-t|},$$

where C *depends on* d, σ_0, σ_1 *and on the bounds for* $\partial_x^k a, \partial_x^k b$, *for* $k \leq 2$.

Proof. Assume without loss of generality that $t < t'$. Since $|\partial_x u(t',x) - \partial_x u(t,x)| = \sqrt{\sum_{i=1}^d |\partial_{x_i} u(t',x) - \partial_{x_i} u(t,x)|^2}$, we give an upper bound for $|\partial_{x_i} u(t',x) - \partial_{x_i} u(t,x)|$, $\forall i \leq d$. From the Feynman-Kac formula (see Theorem A.6), we deduce $u(t,x) = \mathbb{E}[\int_t^T f(s,X_s)ds]$, where $(X_s)_{s\geq t}$ satisfies (6.4). Let $p(t,x;s,y)$ be the transition density function of X_s. Then, we can write $\partial_{x_i} u(t,x) = \int_{\mathbb{R}^d} dy \int_t^T f(s,y)\partial_{x_i} p(t,x;s,y)ds$.

$$
\begin{aligned}
\partial_{x_i} u(t',x) - \partial_{x_i} u(t,x) &= \int_{\mathbb{R}^d} dy \int_{t'}^T f(s,y)[\partial_{x_i} p(t',x;s,y) - \partial_{x_i} p(t,x;s,y)]ds \\
&\quad - \int_t^{t'} f(s,y)\partial_{x_i} p(t,x;s,y)ds := A_1 - A_2.
\end{aligned}
$$

Then, we establish upper bounds for A_1 and A_2.

Upper bound for A_2. Using (6.8) yields

$$
|A_2| \leq C\|f\|_\infty \int_{\mathbb{R}^d} dy \int_t^{t'} ds(s-t)^{-\frac{(d+1)}{2}} e^{-c\frac{|x-y|^2}{s-t}} \leq C\|f\|_\infty \int_t^{t'} \frac{ds}{\sqrt{s-t}},
$$

since $\int_{\mathbb{R}^d} dy(s-t)^{-\frac{d}{2}} e^{-c\frac{|x-y|^2}{s-t}} = (\frac{\pi}{c})^{\frac{d}{2}}$. Hence, we get $|A_2| \leq C\|f\|_\infty \sqrt{t'-t}$.

Upper bound for A_1. First, we apply a Taylor formula to the difference $\partial_{x_i} p(t',x;s,y) - \partial_{x_i} p(t,x;s,y)$.

$$
\partial_{x_i} p(t',x;s,y) - \partial_{x_i} p(t,x;s,y) = (t'-t)\int_0^1 d\lambda \partial^2_{tx} p(\lambda t' + (1-\lambda)t, x; s, y).
$$

Using this formula and (6.10) yield

$$
\begin{aligned}
|A_1| &\leq C\|f\|_\infty (t'-t) \int_{\mathbb{R}^d} dy \int_{t'}^T ds \int_0^1 d\lambda \frac{C}{(s-(\lambda t' + (1-\lambda)t))^{\frac{d+3}{2}}} e^{-c\frac{|x-y|^2}{(s-(\lambda t'+(1-\lambda)t))}} \\
&\leq C\|f\|_\infty (t'-t) \int_{t'}^T ds \int_0^1 d\lambda (s-(\lambda t' + (1-\lambda)t))^{-\frac{3}{2}}.
\end{aligned}
$$

Since $\lambda t' + (1-\lambda)t \in [t,t']$, we can write

$$
(s-(\lambda t' + (1-\lambda)t))^{-\frac{3}{2}} \leq \frac{1}{(s-(\lambda t' + (1-\lambda)t))\sqrt{s-t'}}.
$$

It remains to prove that $\int_{t'}^T ds \int_0^1 d\lambda \frac{1}{(s-(\lambda t'+(1-\lambda)t))\sqrt{s-t'}} \leq \frac{C}{\sqrt{t'-t}}$ to get $|A_1| \leq C\|f\|_\infty \sqrt{t'-t}$.

$$
\begin{aligned}
\int_{t'}^T ds \frac{1}{(s-(\lambda t' + (1-\lambda)t))\sqrt{s-t'}} &= 2\int_0^{\sqrt{T-t'}} \frac{1}{u^2 + (1-\lambda)(t'-t)} du \\
&= \frac{2}{\sqrt{(1-\lambda)(t'-t)}} \int_0^{\sqrt{\frac{T-t'}{(1-\lambda)(t'-t)}}} \frac{1}{1+v^2} dv \\
&= \frac{2}{\sqrt{(1-\lambda)(t'-t)}} \arctan\left(\sqrt{\frac{T-t'}{(1-\lambda)(t'-t)}}\right).
\end{aligned}
$$

Thus, $\int_{t'}^T ds \int_0^1 d\lambda \frac{1}{(s-(\lambda t'+(1-\lambda)t))\sqrt{s-t'}} \leq \frac{\pi}{\sqrt{t'-t}} \int_0^1 \frac{d\lambda}{\sqrt{1-\lambda}}$, and we get $|A_1| \leq C\|f\|_\infty \sqrt{t'-t}$. □

7.3 Bound for $\|u\|^2_{H^{2,\mu}_{\beta,X}}$.

Proposition 7.4. *Assume σ is uniformly elliptic, $\sigma \in C^{1,1}_b$, b is $C^{1,1}$ and bounded, and $f \in L^2(0,T;H^\mu_{\beta,X})$. We also assume that u satisfies the PDE (7.1). Then,*

$$\|u\|^2_{H^{2,\mu}_{\beta,X}} + \|\partial_t u\|^2_{H^\mu_{\beta,X}} \leq C\, \|f\|^2_{H^\mu_{\beta,X}}. \tag{7.4}$$

Proof of Proposition 7.4. We do the proof in three steps, combining Theorem 7.1, page 74, and Proposition 6.12, page 67. Let us give the sketch of the proof.

- We begin by showing that $\|u\|^2_{H^{2,\mu}_{\beta,X}} + \|\partial_t u\|^2_{H^\mu_{\beta,X}} \leq C(\|u\|^2_{H^{2,\mu}_\beta} + \|\partial_t u\|^2_{H^\mu_\beta})$.
- Then, we prove $\|u\|^2_{H^{2,\mu}_\beta} + \|\partial_t u\|^2_{H^\mu_\beta} \leq C\, \|f\|^2_{H^\mu_\beta}$.
- To conclude, we show $\|f\|^2_{H^\mu_\beta} \leq C\, \|f\|^2_{H^\mu_{\beta,X}}$.

The first and third points can be proved using respectively the second and first inequality of (6.12). Concerning the first point, we apply the second inequality of (6.12) with $t = 0$ and $\Psi(s,x) = e^{\beta s}\psi(s,x)$, where $\psi(s,x) = |u(s,x)|^2$, $|\partial_s u(s,x)|^2$, $|\partial_x u(s,x)|^2$, and $|\partial^2_x u(s,x)|^2$. For the third point, we use the first inequality of (6.12), with $\Psi(s,x) = e^{\beta s} f^2(s,x)$ and $t = 0$. To prove the second point, we apply Theorem 7.1 to $v(t,x) = e^{\frac{\beta t}{2}} u(t,x)$ and $p = 2$. Since $u(t,x)$ satisfies (7.1), we get that $v(t,x)$ satisfies

$$\begin{cases} -\partial_t v(t,x) - \mathcal{L}_{(t,x)} v(t,x) + \frac{\beta}{2} v(t,x) = e^{\frac{\beta t}{2}} f(t,x), \\ v(T,x) = 0. \end{cases}$$

According to the hypotheses of Proposition 7.4, we can apply Theorem 7.1 to get

$$\left\|\partial_t (e^{\frac{\beta t}{2}} u(t,x))\right\|_{\mathbb{L}^2(0,T;H^\mu)} + \left\|e^{\frac{\beta t}{2}} u(t,x)\right\|_{\mathbb{L}^2(0,T;H^{2,\mu})} \leq C \left\|e^{\frac{\beta t}{2}} f(t,x)\right\|_{\mathbb{L}^2(0,T;H^\mu)},$$

which implies $\|u\|^2_{H^{2,\mu}\beta} + \|\partial_t u\|^2_{H^\mu_\beta} \leq C\, \|f\|^2_{H^\mu_\beta}$, and Proposition 7.4 is proved. ☐

Chapter 8

Convergence rate for the density of the Euler scheme

The last chapter of this part is devoted to the study of the convergence rate of the density of the Euler scheme. This part is devoted to the study of the difference $p(t, x; s, y) - p^N(t, x; s, y)$, where p (resp. p^N) denotes the transition density function of the process X (resp. X^N). We aim at proving Theorem 8.1, which makes precise the way the upper bound of $|p(t, x; s, y) - p^N(t, x; s, y)|$ depends on N and $(s - t)$. We also recall different expansions for $p(t, x; s, y) - p^N(t, x; s, y)$ which can be found in the literature.

8.1 Introduction

Let us consider a d-dimensional diffusion process $(X_s)_{t\leq s\leq T}$ and a q-dimensional Brownian motion $(W_s)_{t\leq s\leq T}$. X satisfies the following SDE

$$dX_s^i = b_i(s, X_s)ds + \sum_{j=1}^{q} \sigma_{ij}(s, X_s)dW_s^j, \quad X_t^i = x^i, \forall i \in \{1, \cdots, d\}. \tag{8.1}$$

We approximate X by its Euler scheme of order $N \geq 1$, say X^N defined as follows. We consider the subdivision $\{t = t_0 < t_1 < \cdots < t_N = T\}$ of the interval $[t, T]$, i.e. $t_k = t + k\frac{(T-t)}{N}$. We put $X_t^N = x$ and, for all $k \in \{0, \cdots, N-1\}$ and $u \in [t_k, t_{k+1}]$,

$$X_u^{N,i} = X_{t_k}^{N,i} + b_i(t_k, X_{t_k}^N)(u - t_k^N) + \sum_{j=1}^{q} \sigma_{ij}(t_k, X_{t_k}^N)(W_u^j - W_{t_k}^j), \forall i \in \{1, \cdots, d\}. \tag{8.2}$$

Note that the continuous Euler scheme is an Itô process verifying

$$X_u^N = x + \int_t^u b(\varphi(s), X_{\varphi(s)}^N)ds + \int_t^u \sigma(\varphi(s), X_{\varphi(s)}^N)dW_s$$

where $\varphi(u) := \sup\{t_k : t_k \leq u\}$.
If σ is uniformly elliptic, the Markov process X (resp. the process X^N) admits a transition probability density, denoted $p(t, x; s, y)$ (resp. $p^N(t, x; s, y)$). Let us state the main result of this chapter, whose proof is postponed to Section 8.4.

Theorem 8.1. *Assume σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t \sigma \in C_b^1$ in space. Then, $\forall (s,x,y) \in [t,T] \times \mathbb{R}^d \times \mathbb{R}^d$, there exist a constant $c > 0$ and a function $K(T)$ non decreasing in T and depending on the dimension d and on the upper bounds of σ, b and their derivatives s.t.*

$$|p(t,x;s,y) - p^N(t,x;s,y)| \leq \frac{K(T)(T-t)}{N(s-t)^{\frac{d+1}{2}}} \exp(-\frac{c|x-y|^2}{s-t}).$$

Corollary 8.2. *Assume the same hypotheses as in the above theorem. From the last inequality and Aronson's inequality (6.7), page 66, we deduce*

$$\left| \frac{p(t,x;T,x) - p^N(t,x;T,x)}{p(t,x;T,x)} \right| \leq \frac{K(T)}{N} \sqrt{T-t}. \tag{8.3}$$

This inequality yields $p(t,x;T,x) \sim p^N(t,x;T,x)$ when $T \to t$.

Remark 8.3. In the rest of this section, $K(T)$ denotes a non decreasing function in T and C denotes a strictly positive constant depending on the dimension d and on the upper bounds of σ, b and their derivatives up to order 1 in time, order 2 in space for b, and order 3 in space for σ.

Before giving a proof of Theorem 8.1, we first recall some other results on the upper bound of $p(t,x;s,y) - p^N(t,x;s,y)$, obtained by Bally and Talay [8], Konakov and Mammen [60], and Guyon [46]. Then, we give basic results on Malliavin calculus for elliptic Itô processes, which will be useful for the proof of Theorem 8.1.

8.2 Previous results

The difference $p(t,x;s,y) - p^N(t,x;s,y)$ has been studied a lot. We can found several results in the literature on expansions of this difference. First, we mention a result from Bally and Talay [8] (Corollary 2.7). Let us assume σ is elliptic (see Definition 6.2) (with σ only depending on x) and

Hypothesis 8.1 *b, σ are $C^\infty(\mathbb{R}^d)$ functions whose derivatives of any order greater or equal to 1 are bounded.*

Using Malliavin calculus, Bally and Talay [8] have shown that

$$p(t,x;T,y) - p^N(t,x;T,y) = \frac{1}{N}\pi_T(x,y) + \frac{1}{N^2} R_T^N(x,y), \tag{8.4}$$

with $|\pi_T(x,y)| + |R_T^N(x,y)| \leq \frac{c_1 K(T)}{T^q} \exp(-c_2 \frac{|x-y|^2}{T})$, where $c_1 \geq 0$, $c_2 > 0$, q a positive constant, and $K(\cdot)$ a non decreasing function. We point out that q is unknown, which doesn't enable us to deduce the behaviour of $p - p^N$ when $t \nearrow T$.

Beside that, to bound from above $p(0,x;1,y) - p^N(0,x;1,y)$, Konakov and Mammen [60] have proposed an analytical approach based on the so-called parametric method. Assume σ is elliptic (see Definition 6.2), and

Hypothesis 8.2 *b, σ are $C^\infty(\mathbb{R}^d)$ functions whose derivatives of any order are bounded.*

For each pair (x, y) they get an expansion of arbitrary order j of $p^N(0, x; 1, y)$. The coefficients of the expansion depend on N

$$p(0, x; 1, y) - p^N(0, x; 1, y) = \sum_{i=1}^{j-1} \frac{1}{N^i} \pi_{N,i}(0, x; 1, y) + O(\frac{1}{N^j}). \tag{8.5}$$

The coefficients have Gaussian tails : for each i they find constants $c_1 \geq 0$, $c_2 > 0$ s.t. for all $N \geq 1$ and all $x, y \in \mathbb{R}^d$, $\pi_{N,i}(0, x; 1, y) \leq c_1 \exp(-c_2|x - y|^2)$. To do so, they use upper bounds for the partial derivatives of p (coming from Friedman [28]) and prove analogous results on the derivatives of p^N. As strong as is this result, nothing is said about replacing 1 by t, for $t \in [0, 1]$. That's why we present now the work of Guyon [46].

Guyon [46] improves (8.4) and (8.5) in the following way. Assume σ is elliptic (see Definition 6.2) and Hypothesis 8.2

Definition 8.4. Let $\mathcal{G}_l(\mathbb{R}^d), l \in \mathbb{Z}$ be the set of all measurable functions $\pi : \mathbb{R}^d \times (0, 1] \times \mathbb{R}^d \to \mathbb{R}$ s.t.

- for all $t \in (0, 1], \pi(\cdot; t, \cdot)$ is infinitely differentiable,
- for all $\alpha, \beta \in \mathbb{N}^d$, there exist two constants $c_1 \geq 0$ and $c_2 > 0$ s.t. for all $t \in (0, 1]$ and $x, y \in \mathbb{R}^d$,

$$|\partial_x^\alpha \partial_y^\beta \pi(x; t, y)| \leq c_1 t^{-(|\alpha|+|\beta|+d+l)/2} \exp(-c_2|x - y|^2/t).$$

The author has proved the following expansions

$$p^N - p = \frac{\pi}{N} + \frac{\pi_N}{N^2}, \tag{8.6}$$

$$p^N - p = \sum_{i=1}^{j-1} \frac{\pi_{N,i}}{N^i} + \sum_{i=2}^{j} \left(t - \frac{\lfloor nt \rfloor}{N}\right)^i \pi'_{N,i} + \frac{\pi''_{N,j}}{N^j}, \tag{8.7}$$

where $\pi \in \mathcal{G}_1(\mathbb{R}^d)$, $(\pi_N, N \geq 1)$ is a bounded sequence in $\mathcal{G}_4(\mathbb{R}^d)$. For each $i \geq 1$, $(\pi_{N,i}, N \geq 1)$ is a bounded family in $\mathcal{G}_{2i-2}(\mathbb{R}^d)$, and $(\pi'_{N,i}, N \geq 1), (\pi''_{N,i}, N \geq 1)$ are two bounded families in $\mathcal{G}_{2i}(\mathbb{R}^d)$. These expansions can be seen as improvements of (8.4) and (8.5) : it allows infinite differentiation w.r.t. x and y and also makes precise the way the coefficients explode when t tends to 0.

Remark 8.5. Using the notations of Guyon [46], we can rewrite Theorem 8.1 as $|p(t, x; s, y) - p^N(t, x; s, y)| \leq \frac{K(T)T}{N}\pi$, where $\pi \in \mathcal{G}_1(\mathbb{R}^d)$. If we look at (8.7) with $j = 1$, we get $p^N - p = \frac{\pi''_{N,1}}{N}$, where $\pi''_{N,1} \in \mathcal{G}_2(\mathbb{R}^d)$. Then, Theorem 8.1 gives a better accuracy for $|p(t, x; s, y) - p^N(t, x; s, y)|$ when s tends to t.

Before giving the proof of Theorem 8.1, we recall a result on upper bounds for the partial derivatives w.r.t. x and y of $p(t, x; s, y)$ and on the upper bound for $p^N(t, x; s, y)$, which will be useful in the sequel. It is stated in Guyon [46], Theorem 6, but it has essentially been proved by Konakov and Mammen [60].

Theorem 8.6 (Guyon [46])**.** *Assume σ is uniformly elliptic and Hypothesis 8.2. Then, $\forall t \in (0,1]$ and $x \in \mathbb{R}^d$, $N \geq 1$, X_t^N has a density $p_N(0, x; t, y)$ and p_N is a bounded sequence in $\mathcal{G}_0(\mathbb{R}^d)$.*

Remark 8.7. The assumptions of Theorem 8.6 are stronger than the one we made in Theorem 8.1. However, the proof of Theorem 8.1 only requires the following inequality

$$\exists c > 0 \text{ s.t. } \forall (s, x, y) \in (t, T] \times \mathbb{R}^d \times \mathbb{R}^d \ \ p_N(t, x; s, y) \leq \frac{CK(T)}{(s-t)^{\frac{d}{2}}} \exp\left(-c\frac{|x-y|^2}{s-t}\right),$$

which can be proved under the assumptions of Theorem 8.1.

The next section is devoted to recalling basic results on Malliavin calculus for elliptic Itô processes. These results will be useful for the proof of Theorem 8.1.

8.3 Basic results on Malliavin calculus for elliptic Itô processes

We refer the reader to Nualart [78], section 2.2, for more details of this section. Fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and let $(W_t)_{t\geq 0}$ be a q-dimensional Brownian motion. For $h(\cdot) \in H = \mathbb{L}^2([0,T], \mathbb{R}^q)$, $W(h)$ is the Wiener stochastic integral $\int_0^T h(t)dW_t$. Let $\mathcal{S}$ denote the class of random variables of the form $F = f(W(h_1), \cdots, W(h_n))$ where f is a C^∞ function with derivatives having a polynomial growth, $(h_1, \cdots, h_n) \in H^n$ and $n \geq 1$. For $F \in \mathcal{S}$, we define its derivative $\mathcal{D}F = (\mathcal{D}_t F := (\mathcal{D}_t^1 F, \cdots, \mathcal{D}_t^q F))_{t\in[0,T]}$ as the H valued random variable given by

$$\mathcal{D}_t F = \sum_{i=1}^n \partial_{x_i} f(W(h_1), \cdots, W(h_n)) h_i(t).$$

The operator $\mathcal{D}$ is closable as an operator from $\mathbb{L}^p(\Omega)$ to $\mathbb{L}^p(\Omega; H)$, for $p \geq 1$. Its domain is denoted by $\Delta^{1,p}$ w.r.t. the norm $\|F\|_{1,p} = [\mathbb{E}|F|^p + \mathbb{E}(\|\mathcal{D}F\|_H^p)]^{1/p}$. We can define the iteration of the operator $\mathcal{D}$, in such a way that for a smooth random variable F, the derivative $\mathcal{D}^k F$ is a random variable with values on $H^{\otimes k}$. As in the case $k = 1$, the operator $\mathcal{D}^k$ is closable from $\mathcal{S} \subset \mathbb{L}^p(\Omega)$ into $\mathbb{L}^p(\Omega; H^{\otimes k}), p \geq 1$. If we define the norm

$$\|F\|_{k,p} = [\mathbb{E}|F|^p + \sum_{j=1}^k \mathbb{E}(\|\mathcal{D}^j F\|_{H^{\otimes j}}^p)]^{1/p},$$

we denote its domain by $\Delta^{k,p}$. Finally, set $\Delta^{k,\infty} = \cap_{p\geq 1} \Delta^{k,p}$, and $\Delta^\infty = \cap_{k,p\geq 1} \Delta^{k,p}$. One has the following chain rule property

Proposition 8.8. *Fix $p \geq 1$. For $f \in C_b^1(\mathbb{R}^d, \mathbb{R})$, and $F = (F_1, \cdots, F_d)^*$ a random vector whose components belong to $\Delta^{1,p}$, $f(F) \in \Delta^{1,p}$ and for $t \geq 0$, one has $\mathcal{D}_t(f(F)) = f'(F)\mathcal{D}_t F$, with the notation*

$$\mathcal{D}_t F = \begin{pmatrix} \mathcal{D}_t F_1 \\ \vdots \\ \mathcal{D}_t F_d \end{pmatrix} \in \mathbb{R}^d \otimes \mathbb{R}^q.$$

We now introduce δ, the Skorokhod integral, defined as the adjoint operator of $\mathcal{D}$.

Proposition 8.9. *δ is a linear operator on $\mathbb{L}^2([0,T] \times \Omega, \mathbb{R}^q)$ with values in $\mathbb{L}^2(\Omega)$ s.t.*

- *the domain of δ (denoted by $Dom(\delta)$) is the set of processes $u \in \mathbb{L}^2([0,T] \times \Omega, \mathbb{R}^q)$ s.t. $|\mathbb{E}(\int_0^T \mathcal{D}_t F \cdot u_t dt)| \leq c(u) \|F\|_{\mathbb{L}^2}$ for any $F \in \Delta^{1,2}$.*
- *If u belongs to $Dom(\delta)$, then $\delta(u)$ is the one element of $\mathbb{L}^2(\Omega)$ characterised by the integration-by-parts formula*

$$\forall F \in \Delta^{1,2}, \ \ \mathbb{E}(F\delta(u)) = \mathbb{E}\left(\int_0^T \mathcal{D}_t F \cdot u_t dt\right).$$

Remark 8.10. If u is an adapted process belonging to $\mathbb{L}^2([0,T]\times\Omega, \mathbb{R}^q)$, then the Skorohod integral and the Itô integral coincide : $\delta(u) = \int_0^T u_t dW_t$, and the preceding integration-by-parts formula becomes

$$\forall F \in \Delta^{1,2}, \ \ \mathbb{E}\left(F\int_0^T u_t dW_t\right) = \mathbb{E}\left(\int_0^T \mathcal{D}_t F \cdot u_t dt\right). \tag{8.8}$$

In the following, this equality is called the duality formula.

We recall some standard results related to the integration-by-parts formula. The Malliavin covariance matrix of a smooth random variable F is defined by $\gamma^F = \int_0^T \mathcal{D}_t F[\mathcal{D}_t F]^* dt$. The following proposition corresponds to the first part of Proposition 2.4 of Gobet and Munos [44], or to Proposition 3.2.1 page 160 of Nualart [79].

Proposition 8.11 (Gobet and Munos [44])**.** *Let α be a multi-index, F be a random variable in $\Delta^{k_1,\infty}$ s.t. $det(\gamma^F)$ is almost surely positive with $1/det(\gamma^F) \in \cap_{p\geq 1}\mathbb{L}^p$ and G belongs to $\Delta^{k_2,\infty}$. Then, for any smooth function g with polynomial growth, provided that k_1 and k_2 are large enough (depending on α), there exists a random variable $H_\alpha(F,G)$ in any $\mathbb{L}^p$ s.t.*

$$\mathbb{E}[\partial^\alpha g(F)G] = \mathbb{E}[g(F)H_\alpha(F,G)].$$

Now, we intend to apply such a result with $F = Y_t$, where Y is some Itô process. (For example we can take $F = X_t$ or $F = X_t^N$). We only consider a specific class of elliptic Itô processes defined in the following Proposition. The proposition we state is quite similar to Proposition 4.2 of Gobet [37]. The difference is based on the fact that in Gobet [37] the coefficients b, σ are assumed to be C^∞ and bounded. In our case, b and σ are in $C_b^{1,2}$. The following result is derived from Kusuoka and Stroock [63].

Proposition 8.12. *We assume σ is uniformly elliptic (see Definition 6.2), $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,2}$. Consider ν, a map from $\mathbb{R}^+$ into $\mathbb{R}^+$, satisfying the non-anticipative condition $0 \le \nu(s) \le s$ for any s. Let $(Y_t^\nu)_{t\ge 0}$ be the d-dimensional Itô process defined by*

$$Y_t^\nu = x + \int_0^t b(\nu(s), Y_{\nu(s)}^\nu)ds + \int_0^t \sigma(\nu(s), Y_{\nu(s)}^\nu)dW_s.$$

Then, for $t > 0$, $Y_t^\nu \in \Delta^{2,\infty}$, and for $k \le 2, p > 1$, there is a function $K(T)$ (not depending on ν), s.t.

$$\sup_{t\in[0,T]} \|Y_t^\nu(x)\|_{k,p} \le K(T)(1 + |x|).$$

The Malliavin covariance matrix of Y_t^ν is a.s. invertible and its inverse, denoted Γ_t^ν belongs to $\cap_{p>1}\mathbb{L}^p$. Moreover, we have $\|\Gamma_t^\nu(x)\|_{\mathbb{L}^p} \le \frac{K(T)}{t^d}$, uniformly in x and ν.
Integration by parts formula : for all $p > 1$, for all multi-index α s.t. $|\alpha| \le 2$, for all $s \in [0, T]$ and $t \in (0, T]$ and for any functions f and g in $C_b^{|\alpha|}(\mathbb{R}^d, \mathbb{R})$, there exist a random variable $H_\alpha(g(Y_s^\nu), Y_t^\nu) \in \mathbb{L}^p$ and a function $K(T)$ (uniform in ν, x, s, t, f and g) s.t.

$$\mathbb{E}_x[\partial_x^\alpha f(Y_t^\nu)g(Y_s^\nu)] = \mathbb{E}_x[f(Y_t^\nu)H_\alpha(g(Y_s^\nu), Y_t^\nu)], \tag{8.9}$$

with

$$[\mathbb{E}_x|H_\alpha(g(Y_s^\nu), Y_t^\nu)|^p]^{1/p} \le \frac{K(T)}{t^{\frac{|\alpha|}{2}}} \|g\|_{C_b^{|\alpha|}}. \tag{8.10}$$

All these results are given in the article of Kusuoka and Stroock [63]. The first one, concerning the estimates of Sobolev norms $\|\cdot\|_{k,p}$, is stated in Theorem 2.19. The inequality on the Malliavin covariance matrix is stated in Theorem 3.5. (8.10) is owed to Theorem 1.20 and Corollary 3.7.

8.4 Proof of Theorem 8.1

To prove Theorem 8.1, we combine Theorem 8.6 and Proposition 8.12. The scheme of the proof is the following

- Use a PDE and Itô's calculus to write the difference $p^N(t, x; s, y) - p(t, x; s, y)$

$$= \int_t^s \mathbb{E}_{t,x}\left[\sum_{i=1}^d (b_i(\varphi(r), X_{\varphi(r)}^N) - b_i(r, X_r^N))\partial_{x_i} p(r, X_r^N; s, y)\right.$$
$$\left. + \frac{1}{2}\sum_{i,j=1}^d (a_{ij}(\varphi(r), X_{\varphi(r)}^N) - a_{ij}(r, X_r^N))\partial^2_{x_i x_j} p(r, X_r^N; s, y)\right] dr := E_1 + E_2. \tag{8.11}$$

- Prove the intermediate result $\forall (r, x, y) \in [t, s] \times \mathbb{R}^d \times \mathbb{R}^d$

$$\mathbb{E}_{t,x}\left[\exp\left(-c\frac{|y - X_r^N|^2}{s - r}\right)\right] \le K(T)\left(\frac{s - r}{s - t}\right)^{\frac{d}{2}} \exp\left(-c'\frac{|x - y|^2}{s - t}\right), \tag{8.12}$$

  where $c, c' > 0$.

- Use Malliavin calculus, Theorem 8.6 and the intermediate result, to show that each term E_1 and E_2 of the sum of the r.h.s. of (8.11) is bounded by $\frac{K(T)(T-t)}{N}\frac{1}{(s-t)^{\frac{d+1}{2}}}\exp(-c\frac{|x-y|^2}{s-t})$.

Definition 8.13. We say that a term $E(t,x,s,y)$ satisfies the property $\mathcal{P}$ if $\forall(t,x,s,y) \in [0,T]\times\mathbb{R}^d\times[t,T]\times\mathbb{R}^d$

$$E(t,x,s,y) \leq \frac{K(T)(T-t)}{N}\frac{1}{(s-t)^{\frac{d+1}{2}}}\exp\left(-c\frac{|x-y|^2}{s-t}\right). \quad (\mathcal{P})$$

8.4.1 Proof of equality (8.11)

Let us consider the following PDE, where $s \in [t,T]$.

$$\begin{cases} (\partial_r + \mathcal{L}_{(r,x)})u(r,x) = 0, & \forall r \in [t,s], \forall x \in \mathbb{R}^d \\ u(s,x) = f(x), \end{cases}$$

where $\mathcal{L}_{(r,x)}$ is defined by (6.2). We also assume that f is continuous and satisfies a polynomial growth condition. Applying Feynman-Kac formula (see Appendix A.2.1) gives $u(t,x) = \mathbb{E}_{t,x}[f(X_s)]$. By using the terminal condition of the PDE, we also have $u(s,X_s^N) = f(X_s^N)$. Then,

$$\mathbb{E}_{t,x}[f(X_s^N) - f(X_s)] = \mathbb{E}_{t,x}[u(s,X_s^N) - u(t,x)]. \tag{8.13}$$

On the other hand, $E_{t,x}[f(X_s^N) - f(X_s)]$ can also be written using the transition density $p(t,x;s,y)$ of X, and the density $p^N(t,x;s,y)$ of X^N. We get

$$E_{t,x}[f(X_s^N) - f(X_s)] = \int_{\mathbb{R}^d} f(y)[p^N(t,x;s,y) - p(t,x;s,y)]dy. \tag{8.14}$$

Besides that, using Itô's formula yields

$$\begin{aligned}\mathbb{E}_{t,x}[u(s,X_s^N) - u(t,x)] =& \mathbb{E}_{t,x}\left[\int_t^s \partial_t u(r,X_r^N)dr\right] + \mathbb{E}_{t,x}\left[\int_t^s \sum_{i=1}^d b_i(\varphi(r),X_{\varphi(r)}^N)\partial_{x_i}u(r,X_r^N)dr\right] \\ &+ \frac{1}{2}\mathbb{E}_{t,x}\left[\int_t^s \sum_{i,j=1}^d a_{ij}(\varphi(r),X_{\varphi(r)}^N)\partial^2_{x_ix_j}u(r,X_r^N)dr\right].\end{aligned}$$

From PDE we have $\partial_t u(r,X_r^N) = -\mathcal{L}u(r,X_r^N)$, then the above equality becomes

$$\begin{aligned}\mathbb{E}_{t,x}[u(s,X_s^N) - u(t,x)] =& \mathbb{E}_{t,x}\left[\int_t^s \sum_{i=1}^d (b_i(\varphi(r),X_{\varphi(r)}^N) - b_i(r,X_r^N)\partial_{x_i}u(r,X_r^N)dr\right] \\ &+ \frac{1}{2}\mathbb{E}_{t,x}\left[\int_t^s \sum_{i,j=1}^d (a_{ij}(\varphi(r),X_{\varphi(r)}^N) - a_{ij}(r,X_r^N))\partial^2_{x_ix_j}u(r,X_r^N)dr\right].\end{aligned}$$

Since $u(t,x) = \int_{\mathbb{R}^d} f(y)p(t,x;s,y)dy$ we get $\partial_{x_i} u(t,x) = \int_{\mathbb{R}^d} f(y)\partial_{x_i} p(t,x;s,y)dy$, and hence

$$\begin{aligned}
&\mathbb{E}_{t,x}[u(s,X_s^N) - u(t,x)] = \\
&\mathbb{E}_{t,x}\left[\int_{\mathbb{R}^d} dy f(y)\int_t^s \sum_{i=1}^d (b_i(\varphi(r),X_{\varphi(r)}^N) - b_i(r,X_r^N))\partial_{x_i} p(r,X_r^N;s,y)dr\right] \\
&+\frac{1}{2}\mathbb{E}_{t,x}\left[\int_{\mathbb{R}^d} dy f(y)\int_t^s \sum_{i,j=1}^d (a_{ij}(\varphi(r),X_{\varphi(r)}^N) - a_{ij}(r,X_r^N))\partial^2_{x_i x_j} p(r,X_r^N;s,y)dr\right].
\end{aligned} \tag{8.15}$$

Combining (8.13), (8.14) and (8.15) yields the following equality, which is true for all continuous functions f continuous with polynomial growth

$$\begin{aligned}
&\int_{\mathbb{R}^d} f(y)[p^N(t,x;s,y) - p(t,x;s,y)]dy = \\
&\mathbb{E}_{t,x}\left[\int_{\mathbb{R}^d} dy f(y)\int_t^s \sum_{i=1}^d (b_i(\varphi(r),X_{\varphi(r)}^N) - b_i(r,X_r^N))\partial_{x_i} p(r,X_r^N;s,y)dr\right] \\
&+\frac{1}{2}\mathbb{E}_{t,x}\left[\int_{\mathbb{R}^d} dy f(y)\int_t^s \sum_{i,j=1}^d (a_{ij}(\varphi(r),X_{\varphi(r)}^N) - a_{ij}(r,X_r^N))\partial^2_{x_i x_j} p(r,X_r^N;s,y)dr\right].
\end{aligned} \tag{8.16}$$

To get (8.11), we put the r.h.s. of (8.16) on the left one and we choose $f(y) = p^N(t,x;s,y) - p(t,x;s,y) - \mathbb{E}_{t,x}\int_t^s \sum_{i=1}^d \big(b_i(\varphi(r),X_{\varphi(r)}^N) - b_i(r,X_r^N)\big)\partial_{x_i} p(r,X_r^N;s,y)dr - \frac{1}{2}\mathbb{E}_{t,x}\int_t^s \sum_{i,j=1}^d \big(a_{ij}(\varphi(r),X_{\varphi(r)}^N) - a_{ij}(r,X_r^N)\big)\partial^2_{x_i x_j} p(r,X_r^N;s,y)dr$. We can do so since

- $p^N(t,x;s,\cdot)$ is $C^\infty(\mathbb{R}^d)$ with polynomial growth (see the Examples page 305, point 2) of Kusuoka and Stroock [63]). This result also ensues from the fact that $p^N(t,x;s,\cdot)$ can be written as a convolution product of Gaussian random variables,
- $\partial_x^k p(t,x;s,\cdot)$, $0 \le k \le 3$ are continuous with polynomial growth (see Proposition 6.9),
- the function $\Psi : y \longmapsto \mathbb{E}_{t,x}\int_t^s \sum_{i=1}^d \big(b_i(\varphi(r),X_{\varphi(r)}^N) - b_i(r,X_r^N)\big)\partial_{x_i} p(r,X_r^N;s,y)dr - \frac{1}{2}\mathbb{E}_{t,x}\int_t^s \sum_{i,j=1}^d \big(a_{ij}(\varphi(r),X_{\varphi(r)}^N) - a_{ij}(r,X_r^N)\big)\partial^2_{x_i x_j} p(r,X_r^N;s,y)dr$ is continuous. We check it by looking at the rest of the proof. In the following of the demonstration, we split this integral into four terms $E_{11}, E_{12}, E_{21}, E_{22}$. Each of these terms can be written as $\int_t^s dr \int_{\varphi(r)}^r \mathbb{E}_{t,x}[\partial_x^k p(r,X_r^N;s,y) g(\varphi(r),X_{\varphi(r)}^N,u,X_u^N)]du$, where $k \in \{0,1,2,3\}$ and g is bounded continuous (see (8.19) for example). Thus, writing the expectation as an integral and using the two preceding points lead to the continuity of Ψ,
- since each term $E_{11}, E_{12}, E_{21}, E_{22}$ satisfies property $(\mathcal{P})$ (see definition 8.13, page 83), the function $\Psi(y)$ has polynomial growth.

Consequently, the chosen function f is continuous with polynomial growth. $\int_{\mathbb{R}^d} f^2(y)dy = 0$ leads to $f(y) = 0$ a.e. As f is continuous, f is null everywhere and (8.11) follows.

8.4.2 Proof of the intermediate result

We prove inequality (8.12). $\mathbb{E}_{t,x}[\exp(-c\frac{|y-X_r^N|^2}{s-r})] = \int_{\mathbb{R}^d} \exp(-c\frac{|y-z|^2}{s-r})p^N(t,x;r,z)dz$. Using Remark 8.7, we get

$$\begin{aligned}\mathbb{E}_{t,x}\left[\exp\left(-c\frac{|y-X_r^N|^2}{s-r}\right)\right] &\leq \frac{K(T)}{(r-t)^{\frac{d}{2}}}\int_{\mathbb{R}^d}\exp\left(-c\frac{|y-z|^2}{s-r}\right)\exp\left(-c'\frac{|x-z|^2}{r-t}\right)dz\\ &\leq K(T)\Pi_{i=1}^d\int_{\mathbb{R}}\frac{1}{\sqrt{r-t}}\exp\left(-c\frac{|y_i-z_i|^2}{s-r}\right)\exp\left(-c'\frac{|x_i-z_i|^2}{r-t}\right)dz_i,\end{aligned}$$

and $\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi\frac{(s-r)}{2c}}}\exp(-c\frac{|y_i-z_i|^2}{s-r})\frac{1}{\sqrt{2\pi\frac{(r-t)}{2c'}}}\exp(-c'\frac{|x_i-z_i|^2}{r-t})dz_i$ is the convolution product of the density of two Gaussian random variables $\mathcal{N}(-x_i,\frac{r-t}{2c'})$ and $\mathcal{N}(y_i,\frac{s-r}{2c})$ computed at 0 (see Lemma A.16, page 265). The integral is equal to $\frac{1}{\sqrt{2\pi(\frac{r-t}{2c'}+\frac{s-r}{2c})}}\exp\left(-\frac{|x_i-y_i|^2}{\frac{r-t}{c'}+\frac{s-r}{c}}\right)$. Then, we get

$$\int_{\mathbb{R}}\frac{1}{\sqrt{r-t}}\exp\left(-c\frac{|y_i-z_i|^2}{s-r}\right)\exp\left(-c'\frac{|x_i-z_i|^2}{r-t}\right)dz_i \leq C\left(\frac{s-r}{s-t}\right)^{\frac{1}{2}}\exp\left(-c''\frac{|x_i-y_i|^2}{s-t}\right),$$

and (8.12) follows.

8.4.3 Upper bound for E_1

We recall that $E_1 = \int_t^s \mathbb{E}_{t,x}\left[\sum_{i=1}^d (b_i(\varphi(r),X_{\varphi(r)}^N) - b_i(r,X_r^N))\partial_{x_i}p(r,X_r^N;s,y)\right]dr$. For each i, we apply Itô's formula to $b_i(u,X_u^N)$ between $u=\varphi(r)$ and $u=r$. We get

$$b_i(\varphi(r),X_{\varphi(r)}^N) - b_i(r,X_r^N) = \int_{\varphi(r)}^r \alpha_u^i du + \int_{\varphi(r)}^r\sum_{k=1}^q \beta_u^{i,k}dW_u^k, \tag{8.17}$$

where α_u^i depends on $\partial_t b, \partial_x b, \partial_x^2 b, \sigma$, and $\beta_u^i = \nabla_x b_i(u,X_u^N)\sigma(\varphi(r),X_{\varphi(r)}^N)$. Since b,σ belong to $C_b^{1,2}, C_b^{1,3}$, α^i and $(\beta^{i,k})_{1\leq k\leq q}$ are uniformly bounded. Using (8.17) and the duality formula (8.8) yield

$$\begin{aligned}E_1 &= \sum_{i=1}^d\int_t^s\{\mathbb{E}_{t,x}[\int_{\varphi(r)}^r \partial_{x_i}p(r,X_r^N;s,y)\alpha_u^i du + \mathbb{E}_{t,x}[\int_{\varphi(r)}^r \mathcal{D}_u(\partial_{x_i}p(r,X_r^N;s,y))\cdot\beta_u^i du]\}dr\\ &:= E_{11}+E_{12},\end{aligned}\tag{8.18}$$

where β_u^i is a row vector of q components. We upper bound E_{11} and E_{12}.

Bound for $E_{11} = \sum_{i=1}^d\int_t^s \mathbb{E}_{t,x}[\int_{\varphi(r)}^r \partial_{x_i}p(r,X_r^N;s,y)\alpha_u^i du]dr$**.**
Since $\sum_{i=1}^d \partial_{x_i}p(r,X_r^N;s,y)\alpha_u^i \leq |\alpha_u||\partial_x p(r,X_r^N,s,y)|$ and α_u is uniformly bounded in u, we have

$$|E_{11}| \leq C\frac{T-t}{N}\int_t^s \mathbb{E}_{t,x}[|\partial_x p(r,X_r^N,s,y)|]dr.$$

Beside that, from Proposition 6.9, $|\partial_x p(r, X_r^N, s, y)| \leq \frac{c}{(s-r)^{\frac{d+1}{2}}} \exp\big(-c\frac{|y-X_r^N|^2}{s-r}\big)$. Then,

$$|E_{11}| \leq C\frac{T-t}{N}\int_t^s \frac{1}{(s-r)^{\frac{d+1}{2}}}\mathbb{E}_{t,x}\left[\exp\left(-c\frac{|y-X_r^N|^2}{s-r}\right)\right] dr.$$

Using the intermediate result (8.12) yields

$$\begin{aligned}|E_{11}| &\leq K(T)\frac{T-t}{N}\int_t^s \frac{1}{\sqrt{s-r}}\frac{1}{(s-t)^{\frac{d}{2}}}\exp\left(-c\frac{|x-y|^2}{s-t}\right) dr \\ &\leq K(T)\frac{T-t}{N}\frac{1}{(s-t)^{\frac{d-1}{2}}}\exp\left(-c\frac{|x-y|^2}{s-t}\right),\end{aligned}$$

and thus E_{11} satisfies property $\mathcal{P}$ (see Definition 8.13).

Bound for $E_{12} = \sum_{i=1}^d \int_t^s \mathbb{E}_{t,x}[\int_{\varphi(r)}^r \mathcal{D}_u(\partial_{x_i} p(r, X_r^N; s, y)) \cdot \beta_u^i du] dr.$

To rewrite E_{12}, we use the expression of β_u^i and Proposition 8.8, which gives $\mathcal{D}_u(\partial_{x_i} p(r, X_r^N; s, y)) = \nabla_x(\partial_{x_i} p(y, X_r^N; s, y))\sigma(\varphi(r), X_{\varphi(r)}^N)$. Then,

$$E_{12} = \int_t^s dr \int_{\varphi(r)}^r \sum_{i,k=1}^d \mathbb{E}_{t,x}[\partial^2_{x_i x_k} p(r, X_r^N; s, y)[(\sigma\sigma^*)(\varphi(r), X_{\varphi(r)}^N)(\nabla_x b_i(u, X_u^N))^*]_k] du. \tag{8.19}$$

Using the integration-by-parts formula (8.9), we get that

$$\begin{aligned}&\mathbb{E}_{t,x}[\partial^2_{x_i x_k} p(r, X_r^N; s, y)[(\sigma\sigma^*)(\varphi(r), X_{\varphi(r)}^N)(\nabla_x b_i(u, X_u^N))^*]_k] \\ &= \mathbb{E}_{t,x}[\partial_{x_i} p(r, X_r^N; s, y) H_{e_k}([(\sigma\sigma^*)(\varphi(r), X_{\varphi(r)}^N)(\nabla_x b_i(u, X_u^N))^*]_k, X_r^N)],\end{aligned}$$

where e_k is a vector whose k-th component is 1 and other components are 0. Let $H_{e_k}^i(r,u)$ denote $H_{e_k}([(\sigma\sigma^*)(\varphi(r), X_{\varphi(r)}^N)(\nabla_x b_i(u, X_u^N))^*]_k, X_r^N)$, we write

$$E_{12} = \int_t^s dr \int_{\varphi(r)}^r \sum_{k=1}^d \mathbb{E}_{t,x}[\sum_{i=1}^d \partial_{x_i} p(r, X_r^N; s, y) H_{e_k}^i(r,u)] du.$$

From (8.10), we deduce $\mathbb{E}_{t,x}[|H_{e_k}^i(r,u)|^p]^{1/p} \leq C\frac{K(T)}{(r-t)^{1/2}}$, where C only depends on $|\sigma|_\infty$, $|\partial_x \sigma|_\infty$, $|\partial_x b|_\infty$, $|\partial^2_{xx} b|_\infty$. Since $E_{t,x}[\sum_{i=1}^d \partial_{x_i} p(r, X_r^N; s, y) H_{e_k}^i] \leq \mathbb{E}_{t,x}[|\partial_x p(r, X_r^N; s, y)||H_{e_k}|]$, we apply Hölder's inequality to get

$$E_{12} \leq CK(T)\int_t^s dr \int_{\varphi(r)}^r \frac{1}{(r-t)^{1/2}}\mathbb{E}_{t,x}[|\partial_x p(r, X_r^N; s, y)|^{\frac{d+1}{d}}]^{\frac{d}{d+1}} du.$$

Using Proposition 6.9 leads to $|\partial_x p(r, X_r^N; s, y)| \leq \frac{C}{(s-r)^{\frac{d+1}{2}}}\exp(-c\frac{|y-X_r^N|^2}{s-r})$, and combining this inequality with the intermediate result yields

$$\mathbb{E}_{t,x}[|\partial_x p(r, X_r^N; s, y)|^{\frac{d+1}{d}}]^{d/(d+1)} \leq \frac{CK(T)}{(s-r)^{\frac{d+1}{2}}}\left(\frac{s-r}{s-t}\right)^{\frac{d^2}{2(d+1)}}\exp\left(-c\frac{|y-x|^2}{s-t}\right). \tag{8.20}$$

Hence, E_{12} is bounded by

$$E_{12} \leq \frac{CK(T)}{(s-t)^{\frac{d^2}{2(d+1)}}} \frac{T-t}{N} \exp\left(-c\frac{|y-x|^2}{s-t}\right) \int_t^s \frac{1}{(r-t)^{1/2}} \frac{1}{(s-r)^{\frac{d+1}{2}-\frac{d^2}{2(d+1)}}} dr.$$

It remains to compute $\int_t^s \frac{1}{(r-t)^{1/2}} \frac{1}{(s-r)^{\frac{d+1}{2}-\frac{d^2}{2(d+1)}}} dr$. To do so, we split the interval $[t,s]$ into $[t, t+\frac{s-t}{2}], [t+\frac{s-t}{2}, s]$. Thus,

$$\int_t^s \frac{1}{(r-t)^{1/2}} \frac{1}{(s-r)^{\frac{d+1}{2}-\frac{d^2}{2(d+1)}}} dr \leq \frac{1}{(\frac{s-t}{2})^{\frac{2d+1}{2d+2}}} \int_t^{t+\frac{s-t}{2}} \frac{dr}{(r-t)^{1/2}} + \frac{1}{(\frac{s-t}{2})^{\frac{1}{2}}} \int_{t+\frac{s-t}{2}}^s \frac{dr}{(s-r)^{\frac{2d+1}{2d+2}}}$$

An easy computation leads to $E_{12} \leq \frac{CK(T)}{(s-t)^{d/2}} \frac{T-t}{N} \exp(-c\frac{|y-x|^2}{s-t})$, and E_{12} satisfies property $\mathcal{P}$.

8.4.4 Upper bound for E_2

We recall $E_2 = \displaystyle\int_t^s \mathbb{E}_{t,x}[\frac{1}{2}\sum_{i,j=1}^d (a_{ij}(\varphi(r), X^N_{\varphi(r)}) - a_{ij}(r, X^N_r))\partial^2_{x_i x_j} p(y, X^N_r; s, y)]dr$. As we did for E_1, we apply Itô's formula to $a_{ij}(u, X^N_u)$ between $\varphi(r)$ and r. We get $a_{ij}(\varphi(r), X^N_{\varphi(r)}) - a_{ij}(r, X^N_r) = \int_{\varphi(r)}^r \gamma^{ij}_u du + \int_{\varphi(r)}^r \delta^{ij}_u dW_u$, where γ^{ij}_u depends on $\sigma, \partial_t\sigma, \partial_x\sigma, b, \partial^2_{xx}\sigma$ and δ^{ij}_u is a row vector of size q, whose l-th component is $(\delta^{ij}_u)_l = \sum_{k=1}^d \partial_{x_k} a_{ij}(u, X^N_u)\sigma_{kl}(\varphi(r), X^N_{\varphi(r)})$. Then, using the duality formula (8.8) leads to

$$\begin{aligned} E_2 &= \sum_{i,j=1}^d \int_t^s \{\mathbb{E}_{t,x}[\int_{\varphi(r)}^r \partial^2_{x_i x_j} p(r, X^N_r; s, y)\gamma^{ij}_u du + \mathbb{E}_{t,x}[\int_{\varphi(r)}^r \mathcal{D}_u(\partial^2_{x_i x_j} p(r, X^N_r; s, y)) \cdot \delta^{ij}_u du]\}dr \\ &:= E_{21} + E_{22}. \end{aligned}$$

Bound for $E_{21} = \sum_{ij=1}^d \int_t^s \mathbb{E}_{t,x}[\int_{\varphi(r)}^r \partial^2_{x_i x_j} p(r, X^N_r; s, y)\gamma^{ij}_u du]dr$.
We can treat this term as we did for E_{12}. As $\sigma, b, \partial_t\sigma, \partial_x\sigma, \partial^2_x\sigma$ are C^1_b in space, γ^{ij}_u has the same properties as the term $[(\sigma\sigma^*)(\varphi(r), X^N_{\varphi(r)})(\nabla_x b_i(u, X^N_u))^*]_k$ appearing in (8.19). Thus, E_{21} satisfies property $\mathcal{P}$.

Bound for $E_{22} = \sum_{i,j=1}^d \int_t^s \mathbb{E}_{t,x}[\int_{\varphi(r)}^r \mathcal{D}_u(\partial^2_{x_i x_j} p(r, X^N_r; s, y)) \cdot \delta^{ij}_u du]dr$.

To rewrite E_{22}, we use the expression of δ^{ij}_u and Proposition 8.8, which asserts $\mathcal{D}_u(\partial^2_{x_i x_j} p(r, X^N_r; s, y)) = \nabla_x(\partial^2_{x_i x_j} p(r, X^N_r; s, y))\sigma(\varphi(r), X^N_{\varphi(r)})$. Thus,

$$E_{22} = \sum_{i,j,k=1}^d \int_t^s dr \int_{\varphi(r)}^r \mathbb{E}_{t,x}[\partial^3_{x_i x_j x_k} p(r, X^N_r; s, y)[(\sigma\sigma^*)(\varphi(r), X^N_{\varphi(r)})(\nabla_x a_{ij}(u, X^N_u))^*]_k]du.$$

To complete this proof, we split E_{22} in two terms : E^1_{22} (resp E^2_{22}) corresponds to the integral in r from t to $t+\frac{s-t}{2}$ (resp. from $t+\frac{s-t}{2}$ to s).

- On $[t, t+\frac{s-t}{2}]$, E^1_{22} is bounded by

$$|E^1_{22}| \leq C\frac{T-t}{N}\int_t^{t+\frac{s-t}{2}} \mathbb{E}_{t,x}[|\partial^3_{x_i x_j x_k} p(r, X^N_r; s, y)|]dr.$$

Using Proposition 6.9 and the intermediate result gives

$$|E^1_{22}| \leq C\frac{K(T)(T-t)}{N}\frac{1}{(s-t)^{d/2}}\exp\left(-c\frac{|x-y|^2}{s-t}\right)\int_t^{t+\frac{s-t}{2}} \frac{1}{(s-r)^{3/2}}dr.$$

Hence, E_{22} satisfies $\mathcal{P}$.

- On $[t+\frac{s-t}{2}, s]$, we use the integration by part formula (8.9) of Proposition 8.12, with $|\alpha| = 2$.

$$E^2_{22} = \sum_{i,j,k=1}^{d}\int_{t+\frac{s-t}{2}}^{s} dr \int_{\varphi(r)}^{r} \mathbb{E}_{t,x}[\partial_{x_i} p(r, X^N_r; s, y) H^i_{e_{jk}}]du,$$

where $H^i_{e_{jk}} = H_{e_{jk}}([(\sigma\sigma^*)(\varphi(r), X^N_{\varphi(r)})(\nabla_x a_{ij}(u, X^N_u))^*]_k, X^N_r)$, and e_{jk} is a vector full of zeros except the j-th and the k-th components. Using Hölder's inequality and (8.10) (remember that $\sigma \in C^{1,3}_b$), we obtain

$$E^2_{22} \leq CK(T)\frac{T-t}{N}\int_{t+\frac{s-t}{2}}^{s} \frac{1}{r-t}\mathbb{E}_{t,x}[|\partial_x p(r, X^N_r; s, y)|^{\frac{d+1}{d}}]^{\frac{d}{d+1}}dr. \tag{8.21}$$

By applying (8.20), we get

$$E^2_{22} \leq CK(T)\frac{T-t}{N}\frac{1}{(s-t)^{1+\frac{d^2}{2(d+1)}}}\exp\left(-c\frac{|x-y|^2}{s-t}\right)\int_{t+\frac{s-t}{2}}^{s} \frac{1}{(s-r)^{\frac{2d+1}{2d+2}}}dr,$$

and the result follows.

Part III

Solving a BSDE with adaptive control variates

This part is devoted to the study of a numerical algorithm solving backward stochastic differential equations (BSDEs for short) of the following type

$$(E)\begin{cases} -dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \;\; Y_T = \Phi(X_T), \\ X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \end{cases} \tag{8.22}$$

Several algorithms for solving BSDEs can be found in the literature. Let us list them chronologically.
In 1994, Ma, Protter, and Yong [74] present an algorithm, called the four step scheme, which solves a class of more general BSDEs, general in the sense that b and σ, the coefficients of the diffusion, may depend on y and z. In fact, they solve the associated quasilinear PDE through a finite difference approximation. However, such an algorithm cannot run well in high dimensions. Moreover, they assume restrictive hypotheses on the coefficients.
In 1997, Chevance [21] suggests a method based on random quantization techniques. The author gives an upper bound for Y under strong regularity hypotheses for f and Φ. The error on Z is not presented.
In 2001, Briand, Delyon, and Mémin [17] provide a numerical scheme for BSDEs based on the approximation of the Brownian motion using a summation of Bernoulli random variables. This numerical scheme can easily be implemented. The authors have proved the convergence of their scheme, but no rate of convergence is given.
In 2003, Bally and Pagès [6] present an algorithm for solving reflected BSDEs by using quantization techniques. The authors replace the process X by a Markov chain with finite state space, which are the points of an optimal quantization grid. Once they have computed the transition probabilities on this grid, they easily implement a dynamic programming equation to solve the reflected BSDE.
In 2004, Delarue and Menozzi [24] provide a method for solving FBSDEs (Forward-Backward SDEs), based on quantization techniques and deterministic grids.
In 2004, Bouchard and Touzi [15] discretize the above BSDE w.r.t. the time and get a dynamic programming equation. They compute the conditional expectations appearing in the dynamic programming equation by using Malliavin calculus techniques. The authors analyse the error on Y but not on Z. Besides, the numerical scheme seems to be hard to implement and computationally demanding.
In 2005, Gobet, Lemor, and Warin [45] propose a scheme based on iterative regression functions which are approximated by projections on vector spaces of functions, with coefficients evaluated with Monte Carlo simulations.
In 2007, Bender and Denk [11] propose a forward scheme which avoids nesting of conditional expectations backwards through the time steps. Instead, it mimics the Picard type iteration for BSDEs and, consequently, has nestings of conditional expectations along the iterations. This work has some connections with our approach but it does not handle the issue of error analysis for conditional expectations. The authors use the regression based on least squares Monte Carlo method to approximate the conditional expectation, as it was suggested in Gobet et al. [45].

This part is organised as follows. In Chapter 9, we set up the framework and recall standard results on BSDEs. We also describe our algorithm based on a sequential Monte Carlo technique and on a Picard's iterations method. The algorithm also requires an estimator which computes pointwise the value of a function from its values at some points of a fixed grid. After having introduced the regression analysis in Chapter 10, based on the books of Györfi et al. [47] and Härdle [50], we present in Chapter 11 a non parametric regression technique : the kernel estimate, which enables to build the estimator $\mathcal{P}_n$ (see Section 11.3). In Chapter 12, we state some properties on the convergence of the estimator $\mathcal{P}_n$. Section 12.3 and Section 12.4 deal with the convergence rate of $\mathcal{P}_n v - v$ and $\partial_x(\mathcal{P}_n v) - \partial_x v$, where v is a $C^{1,2}$ function. In Chapter 13, we state the main convergence result of our algorithm. Chapter 14 is devoted to establish technical results needed to prove the result on the convergence of our algorithm announced in Chapter 13. Finally, we present in Chapter 15 numerical experiments based on financial issues.

Chapter 9

Framework and Hypotheses

9.1 Statement of the problem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space on which is defined a q-dimensional standard Brownian motion W, whose natural filtration, augmented with $\mathbb{P}$-null sets, is denoted $(\mathcal{F}_t)_{0\leq t\leq T}$ (T is a fixed terminal time). We aim at numerically approximating the solution (Y, Z) of the following forward backward stochastic differential equation (FBSDE) with fixed terminal time T

$$-dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = \Phi(X_T), \tag{9.1}$$

where $f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$, X is the $\mathbb{R}^d$-valued process solution of

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \tag{9.2}$$

$b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times q}$. The main focus of this work is to provide and analyse an algorithm, based on a Monte Carlo method, which approximates the solution (Y, Z) of (9.1). Section 9.3 page 95 presents some results on BSDEs stated by Pardoux and Peng [81] and El Karoui et al. [27]. In Section 9.4, page 96, we recall several results linking FBSDEs and partial differential equations (PDEs in short) and stated by Pardoux and Peng [82], Ma et al. [74] and Delarue and Menozzi [24]. More precisely, we link (Y, Z) the solution of the above BSDE to u, the solution of the following PDE :

$$(\mathcal{E}) \begin{cases} \partial_t u(t, x) + \mathcal{L}u(t, x) + f(t, x, u(t, x), (\partial_x u\sigma)(t, x)) = 0, \\ u(T, x) = \Phi(x), \end{cases}$$

where $\mathcal{L}$ is defined by

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

According to these results, we get $u \in C_b^{1,2}$ and

$$\forall t \in [0, T], \quad (Y_t, Z_t) = (u(t, X_t), \partial_x u(t, X_t)\sigma(t, X_t)).$$

In Section 9.5, we briefly recall a variance reduction technique useful in the description of our algorithm: the adaptive control variate method. In Section 9.6, page 103, we describe our algorithm, which builds u_k, an approximation of u, and deduces (Y^k, Z^k), the approximation of (Y, Z) by using the following formula

$$\forall t \in [0,T],\ Y_t^k = u_k(t, X_t^N),\ \ Z_t^k = \partial_x u_k(t, X_t^N)\sigma(t, X_t^N),$$

where X^N is an approximation of X. More precisely, we build u_k recursively in the following way

$$u_k(t,x) = \mathcal{P}_n^k(u_{k-1} + \overline{w}_k)(t,x),$$

where $\overline{w}_k(t,x)$ is given by (9.13), page 104 and $\mathcal{P}_n^k$ is a kernel estimator described in Section 11.3, page 119.

9.2 Notations

- Elements of $\mathbb{R}^d$ are encoded as row vectors and $|\cdot|$ is the Euclidean norm on $\mathbb{R}^d$.

- C_p^k denotes the set of C^{k-1} functions whose k-th derivative is piecewise continuous.

- Let $C_{l,b}^k(\mathbb{R}^p, \mathbb{R}^q)$ be the set of C^k functions from $\mathbb{R}^p$ to $\mathbb{R}^q$ with continuous and uniformly bounded derivatives up to order k. The functions themselves don't need to be bounded.

- Let $C_b^{k,l}$ be the set of continuously differentiable functions $\phi : (t,x) \in [0,T] \times \mathbb{R}^d$ with continuous and uniformly bounded derivatives w.r.t. t (resp. w.r.t. x) up to order k (resp. up to order l).

- $\mathbb{L}_t^2(\mathbb{R}^q)$ denotes the set of the q-dimensional random variables which are $\mathcal{F}_t$ measurable and square integrable. For $q = 1$, we also use the notation $\mathbb{L}_t^2$.

- $\mathbb{H}_T^2(\mathbb{R}^q)$ denotes the set of the predictable processes $\phi : \Omega \times [0,T] \to \mathbb{R}^q$ (as row vectors) such that $\mathbb{E}\int_0^T |\phi_t|^2 dt$ is finite. For $q = 1$, we also use the notation $\mathbb{H}_T^2$.

- Let $\beta > 0$ and $\Phi \in \mathbb{H}_T^2(\mathbb{R}^q)$, $\|\phi\|_\beta^2$ denotes $\mathbb{E}\int_0^T e^{\beta t}|\phi_t|^2 dt$. $\mathbb{H}_{T,\beta}^2(\mathbb{R}^q) = \{\phi \in \mathbb{H}_T^2(\mathbb{R}^q) : \|\phi\|_\beta^2 < \infty\}$.

- Let X be a solution of (9.2), $g_1 : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, and $g_2 : \mathbb{R}^d \to \mathbb{R}$. We introduce $\Psi(t, g_1(s,\cdot), g_2(\cdot), X(x)) = g_2(X_T^{t,x}) + \int_t^T g_1(s, X_s^{t,x}) ds$.

- $K(T)$ denotes a generic function non decreasing in T.

- If f is a Lipschitz function, L_f denotes its Lipschitz constant.

9.3 Results on general BSDEs

This part is devoted to recall some standard results on general BSDE of the following type

$$-dY_s = f(s, Y_s, Z_s)ds - Z_s dW_s, \ \ Y_T = \xi \tag{9.3}$$

where ξ is an $\mathcal{F}_T$ measurable random variable and $f : \Omega \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$ is $\mathcal{P} \otimes \mathcal{B} \otimes \mathcal{B}^q$ measurable. f is called the driver of the BSDE. We say that (f, ξ) are standard parameters if ξ belongs to $\mathbb{L}^2_T$, if $f(\cdot, 0, 0)$ belongs to $\mathbb{H}^2_T$ and if f is uniformly Lipschitz in the sense of there exists $C > 0$ s.t.

$$\forall (y_1, y_2, z_1, z_2), \ \ |f(\omega, t, y_1, z_1) - f(\omega, t, y_2, z_2)| \leq C(|y_1 - y_2| + |z_1 - z_2|), d\mathbb{P} \times dt \ a.e.$$

From Pardoux and Peng [81] and El Karoui et al. [27], we have

Theorem 9.1. *Let (f, ξ) be standard parameters, there exists a unique couple of processes $(Y, Z) \in \mathbb{H}^2_T \times \mathbb{H}^2_T(\mathbb{R}^q)$ which solves (9.3).*

A proof can be found in the paper of Pardoux and Peng [81]. A shorter proof using a priori estimates has been done by El Karoui et al. [27], page 20.

Proposition 9.2 (A Priori Estimates, El Karoui et al. [27], Proposition 2.1)**.** *Let $((f^i, \xi^i); i = 1, 2)$ be two standard parameters of the BSDE and $((Y^i, Z^i); i = 1, 2)$ be two square integrable solutions. Let L_f be a Lipschitz constant for f^1, and put $\delta Y_t = Y^1_t - Y^2_t$ and $\delta_2 f_t = f^1(t, Y^2_t, Z^2_t) - f^2(t, Y^2_t, Z^2_t)$. For any (λ, μ, β) such that $\mu > 0, \lambda^2 > L_f$, and $\beta \geq L_f(2 + \lambda^2) + \mu^2$, it follows that*

$$\|\delta Y\|^2_\beta \leq T \left[e^{\beta T} \mathbb{E}(|\delta Y_T|^2) + \frac{1}{\mu^2} \|\delta_2 f\|^2_\beta \right],$$

$$\|\delta Z\|^2_\beta \leq \frac{\lambda^2}{\lambda^2 - L_f} \left[e^{\beta T} \mathbb{E}(|\delta Y_T|^2) + \frac{1}{\mu^2} \|\delta_2 f\|^2_\beta \right].$$

For a proof of this Proposition, see El Karoui et al. [27], page 18. From this proof, we derive that the Picard iterative sequence converges almost surely to the solution of the BSDE.

Corollary 9.3 (Corollary 2.1, El Karoui et al. [27])**.** *Let β be such that $2(1 + T)L_f < \beta$, where L_f defines the Lipschitz constant of f. Let $(\hat{Y}^k, \hat{Z}^k)$ be the sequence defined recursively by $(\hat{Y}_0 = 0, \hat{Z}_0 = 0)$ and*

$$-d\hat{Y}^{k+1}_t = f(t, \hat{Y}^k_t, \hat{Z}^k_t)dt - (\hat{Z}^{k+1}_t)^* dW_t, \ \ \hat{Y}^{k+1}_T = \xi.$$

Then, the sequence $(\hat{Y}^k, \hat{Z}^k)$ converges to (Y, Z), $d\mathbb{P} \times dt$ a.s. (and in $\mathbb{H}^2_{T,\beta}(\mathbb{R}) \times \mathbb{H}^2_{T,\beta}(\mathbb{R}^q)$) as k goes to $+\infty$.

We refer to El Karoui et al. [27], page 21 for a proof of the previous Corollary.

Remark 9.4. FBSDEs introduced before are in a quite complex way examples of BSDEs parametrised by the initial condition x of the forward SDE (9.2). The parametrised generator is given here by $f(t, X^x_t(\omega), y, z)$ and the terminal condition by $\xi(\omega) = \Phi(X^x_T(\omega))$. We refer to El Karoui et al. [27], Section2.4, for a study of BSDEs depending on parameters.

9.4 Connection between FBSDEs and Partial Differential Equations

In this section, we study the relation between forward backward equations and partial differential equations (PDEs). First, we give a generalisation of the Feynman-Kac formula to semilinear parabolic PDEs, as stated by Pardoux and Peng [82]. Then, we recall that under smoothness conditions the function $u(t,x) := Y_t^{t,x}$ is in some sense a solution of a PDE (see also Pardoux and Peng [82]). Finally, we deal with some more general FBSDEs, in which the coefficients of the diffusion b and σ may depend on Y and Z. These FBSDEs, introduced by Antonelli [1] and then by Ma et al. [74], provide an extension of the Feynman-Kac representation to a certain class of quasilinear parabolic PDEs. We remind a result stated by Delarue and Menozzi [24].

9.4.1 FBSDEs ans semilinear parabolic differential equations

We aim at explaining the link between the following semilinear parabolic differential equation :

$$(\mathcal{E}_0)\begin{cases} \partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),(\partial_x u\sigma)(t,x)) = 0, \\ u(T,x) = \Phi(x), \end{cases}$$

where $u : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ and

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

and the decoupled FBSDE

$$(E_0)\begin{cases} \forall\ s \in [t,T],\ X_s^{t,x} = x + \int_t^s b(u,X_u^{t,x})du + \int_t^s \sigma(u,X_u^{t,x})dW_u, \\ Y_s^{t,x} = \Phi(X_T^{t,x}) + \int_s^T f(u,X_u^{t,x},Y_u^{t,x},Z_u^{t,x})du - \int_s^T Z_u^{t,x}dW_u. \end{cases}$$

The following result has been stated by Peng [84] (see also Pardoux and Peng [82], Theorem 3.1, or El Karoui et al. [27], Proposition 4.3). This is a generalisation of the Feynman-Kac formula to semi-linear parabolic PDEs.

Theorem 9.5 (Pardoux and Peng [82], Theorem 3.1)**.** *If $u \in C^{1,2}([0,T]\times\mathbb{R}^d)$ solves $(\mathcal{E}_0)$, then $\forall (t,x) \in [0,T]\times\mathbb{R}^d$, $u(t,x) = Y_t^{t,x}$, where $\{(Y_s^{t,x}, Z_s^{t,x}); t \le s \le T\}_{t\ge 0, x\in\mathbb{R}^d}$ is the unique solution of the FBSDE (E_0). Furthermore, we have*

$$\forall s \in [t,T],\ \ (Y_s^{t,x}, Z_s^{t,x}) = (u(s,X_s^{t,x}), \partial_x u(s,X_s^{t,x})\sigma(s,X_s^{t,x})). \tag{9.4}$$

Conversely, Pardoux and Peng [82] have proved that when b,σ,f and Φ are globally Lipschitz continuous w.r.t. (x,y,z) and uniformly in t (for f), the FBSDE (E_0) provides the unique viscosity solution of the semilinear parabolic PDE $(\mathcal{E}_0)$. We also refer to El Karoui et al. [27], Theorem 4.2 for a proof of this result. First, let us remind the definition of a viscosity solution.

Definition 9.6 (Viscosity solution). Let $v \in C([0,T] \times \mathbb{R}^d)$ satisfying $\forall x \in \mathbb{R}^d$, $v(T,x) = \Phi(x)$. v is said to be a viscosity sub-solution (resp. super-solution) of equation $(\mathcal{E}_0)$ if for any $(t,x) \in [0,T] \times \mathbb{R}^d$ and $\varphi \in C^{1,2}([0,T] \times \mathbb{R}^d)$ such that $\varphi(t,x) = v(t,x)$ on $\{(t,x) : (t,x) \text{ is a minimum (resp. maximum) of } (\varphi - v)\}$,

$$\partial_t \varphi(t,x) + \mathcal{L}\varphi(t,x) + f(t,x,\varphi(t,x),(\partial_x \varphi \sigma)(t,x)) \geq 0$$
$$(\text{resp. } \partial_t \varphi(t,x) + \mathcal{L}\varphi(t,x) + f(t,x,\varphi(t,x),(\partial_x \varphi \sigma)(t,x)) \leq 0).$$

v is said to be a viscosity solution of $(\mathcal{E}_0)$ if it is both a viscosity sub- and super-solution of $(\mathcal{E}_0)$.

Theorem 9.7 (Pardoux and Peng [82], Theorem 4.3). *Assume b, σ, f and Φ are globally Lipschitz continuous w.r.t. (x,y,z) uniformly in t (for f). The function $u(t,x) := Y_t^{t,x}$ is the unique viscosity solution of the backward parabolic PDE $(\mathcal{E}_0)$.*

Under stronger hypotheses on b, σ, f and Φ, we can show that $u(t,x) := Y_t^{t,x}$ solves PDE $(\mathcal{E}_0)$ in the classic sense. We recall the following result, stated by Pardoux and Peng [82] (see also El Karoui et al. [27], Proposition 4.4).

Theorem 9.8 (Pardoux and Peng [82], Theorem 3.2). *Assume b, σ, f and Φ are $C^3_{l,b}$. Then, $u(t,x) := Y_t^{t,x}$ belongs to $C^{1,2}([0,T] \times \mathbb{R}^d)$ and solves the PDE $(\mathcal{E}_0)$.*

Remark 9.9. From the beginning of this Chapter, we have assumed that Y is a one dimensional process. The previous theorem is true even if Y is a p-dimensional process, $p \geq 1$, whereas Theorem 9.7 holds true when $p = 1$.

9.4.2 FBSDEs and quasilinear parabolic differential equations

Here, we consider more general FBSDEs, in which the coefficients b and σ, appearing in (9.2), may depend on Y and Z. For a given $d \in \mathbb{N}^*$, we consider the functions $b : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$, $f : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d \times d}$, $\Phi : \mathbb{R}^d \to \mathbb{R}$. Analogously to $(\mathcal{E}_0)$ and (E_0), we define

$$(\mathcal{E}'_0) \begin{cases} \partial_t u(t,x) + \langle b(t,x,u(t,x), \partial_x u(t,x)\sigma(t,x,u(t,x))), \partial_x u(t,x) \rangle \\ +\frac{1}{2}\mathrm{tr}(a(t,x,u(t,x))\partial^2_{x,x} u(t,x)) \\ +f(t,x,u(t,x),(\partial_x u(t,x)\sigma)(t,x,u(t,x))) = 0, \\ u(T,x) = \Phi(x), \end{cases}$$

with $a(t,x,y) = (\sigma\sigma^*)(t,x,y)$, $(x,y) \in \mathbb{R}^d \times \mathbb{R}$, and the FBSDE

$$(E'_0) \begin{cases} \forall\, s \in [t,T],\ X_s^{t,x} = x + \int_t^s b(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})du + \int_t^s \sigma(u, X_u^{t,x}, Y_u^{t,x})dW_u, \\ Y_s^{t,x} = \Phi(X_T^{t,x}) + \int_s^T f(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})du - \int_s^T Z_u^{t,x} dW_u. \end{cases}$$

Ma et al. [74], Pardoux and Tang [83] and Delarue [23] have investigated in detail the link between $(\mathcal{E}'_0)$ and (E'_0). We recall some results coming from Ma et al. [74] and Delarue and Menozzi [24].

Hypothesis 9.1

1. *The functions b, σ, f and Φ are C^2 functions with first order derivatives in x, y, z being bounded by some constant $L > 0$.*

2. *The function σ satisfies*

$$\sigma(t,x,y)\sigma^*(t,x,y) \geq \nu(|y|)I, \ \ \forall\,(t,x,y) \in [0,T] \times \mathbb{R}^d \times \mathbb{R},$$

for some positive continuous function $\nu(\cdot)$.

3. *There exist a function λ and two constants $C > 0$ and $\alpha \in (0,1)$, such that Φ is bounded in $C^{2+\alpha}(\mathbb{R}^d)$, and for all $(t,x,y,z) \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$,*

$$|\sigma(t,x,y) + f(t,x,0,z)| \leq C,$$
$$|b(t,x,y,0)| \leq \lambda(y).$$

Proposition 9.10 (Ma et al. [74], Proposition 3.1)**.** *Assume Hypothesis 9.1. Then, the PDE $(\mathcal{E}_0')$ admits a unique classical solution $u(t,x)$, which is bounded and $\partial_t u(t,x), \partial_x u(t,x)$ and $\partial^2_{xx} u(t,x)$ are bounded as well.*

Theorem 9.11 (Ma et al. [74], Theorem 4.1)**.** *Assume Hypothesis 9.1. Then, the forward backward SDE (E_0') admits a unique adapted solution (X,Y,Z) where*

$$Y_s^{t,x} = u(s, X_s^{t,x}), \ \ Z_s^{t,x} = \partial_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x}, u(s, X_s^{t,x})), \tag{9.5}$$

with $u(t,x)$ being the solution of PDE $(\mathcal{E}_0')$.

Remark 9.12. Ma et al. [74] have studied a more general FBSDE, in the sense that the backward equation of (E_0') they consider is $Y_s^{t,x} = \Phi(X_T^{t,x}) + \int_s^T f(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})du - \int_s^T \hat{\sigma}(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})dW_u$, where $\hat{\sigma}$ satisfies Ma et al. [74], Hypothesis A.3. Proposition 9.10 and Theorem 9.11 are also valid when the backward equation of (E_0') is $Y_s^{t,x} = \Phi(X_T^{t,x}) + \int_s^T f(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})du - \int_s^T \hat{\sigma}(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})dW_u$.

Under Hypothesis 9.2 below, Delarue and Menozzi [24] state in Theorem 2.1 that the PDE $(\mathcal{E}_0')$ admits a unique strong solution, whose partial derivatives of order one in t and one and two in x are bounded on the whole domain by known parameters.

Hypothesis 9.2 *b, σ, f and Φ are bounded in space and have at most linear growth in the other variables, are uniformly Lipschitz continuous w.r.t. all the variables. σ is uniformly elliptic and Φ is bounded in $C^{2+\alpha}(\mathbb{R}^d)$.*

Theorem 9.13 (Delarue and Menozzi [24])**.** *Assume Hypothesis 9.2. Then, $(\mathcal{E}_0')$ admits a solution $u \in C^{1,2}([0,T] \times \mathbb{R}^d, \mathbb{R})$ and there exists a constant C, only depending on T and on known parameters deriving from Hypothesis 9.2, such that $\forall\,(t,x) \in [0,T] \times \mathbb{R}^d$,*

$$|u(t,x)| + |\partial_x u(t,x)| + |\partial^2_{xx} u(t,x)| + |\partial_t u(t,x)| + \sup_{t' \in [0,T], t \neq t'} [|t-t'|^{-\frac{1}{2}} |\partial_x u(t,x) - \partial_x u(t',x)|] \leq C.$$

Moreover, u is unique in the class of functions $\tilde{u} \in C([0,T] \times \mathbb{R}^d, \mathbb{R}) \cap C^{1,2}([0,T[\times \mathbb{R}^d, \mathbb{R})$ which satisfy $\sup_{(t,x)\in[0,T[\times\mathbb{R}^d}(|\tilde{u}(t,x)| + |\partial_x \tilde{u}(t,x)|) < +\infty$.

From Ma et al. [74], Pardoux and Tang [83] and Delarue [23], Delarue and Menozzi [24] deduce that the relationship between $(\mathcal{E}_0')$ and (E_0') can be summed up as in (9.5).

9.4.3 Application to our problem

We recall that our goal is to provide an algorithm to approximate (Y, Z), the solution of the FBSDE

$$(E)\begin{cases} -dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \\ Y_T = \Phi(X_T), \\ X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \end{cases}$$

We introduce the following semi-linear PDE

$$(\mathcal{E})\begin{cases} \partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),(\partial_x u\sigma)(t,x)) = 0, \\ u(T,x) = \Phi(x), \end{cases}$$

First, we state some hypotheses.

Hypothesis 9.3 *The driver f satisfies for all $t \in [0,T]$ and for all $(x_1, y_1, z_1), (x_2, y_2, z_2) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^{d\times q}$,*

$$|f(t, x_1, y_1, z_1) - f(t, x_2, y_2, z_2)| \leq L_f(|x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2|).$$

Hypothesis 9.4 *σ and b satisfy the following conditions*

$$\forall t \in [0,T] \text{ and } \forall x, y \in \mathbb{R}^d, \;\; |\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq C(|x-y|).$$

Hypothesis 9.5 *The functions b, σ, f and Φ are bounded in x, f satisfies Hypothesis 9.3 and σ and b satisfy Hypothesis 9.4. Φ is of class $C_b^{2+\alpha}$, $\alpha \in]0,1]$ and σ is uniformly elliptic.*

The following Theorem is an easy consequence of Theorem 9.13

Theorem 9.14. *Assume b, σ and f are Lipschitz functions in all their variables and bounded. We also assume σ is uniformly elliptic and Φ is of class $C_b^{2+\alpha}$, $\alpha \in]0,1]$. Then, the solution u of PDE $(\mathcal{E})$ belongs to $C_b^{1,2}$. Furthermore, $(Y_t, Z_t)_{0\leq t\leq T}$ solution of (E) satisfies*

$$\forall t \in [0,T], \;\; (Y_t, Z_t) = (u(t, X_t), \partial_x u(t, X_t)\sigma(t, X_t)). \tag{9.6}$$

9.5 Sequential Monte Carlo and Variance Reduction Techniques

We are often interested in computing $\mu = \mathbb{E}(X)$, where X is a random variable. In many cases we are not able to compute μ exactly. To overcome this problem we can use Monte Carlo methods. The first idea to approximate μ consists of computing $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X^i$. The Strong Law of Large Numbers and the Central Limit Theorem state that $\overline{X}_n$ converges almost surely to μ with the rate of convergence $\frac{\sigma}{\sqrt{n}}$, where σ^2 denotes the relative variance. One way to improve the rate of convergence is to reduce the value of σ^2. To do so, we can use variance reduction techniques like control variate or importance sampling. We refer to Halton [49] for a survey of the principal techniques used in implementing the Monte Carlo method and its applications.

Before introducing adaptive control variate, we recall the control variate technique. The basic idea of control variate is to write $\mathbb{E}(X)$ as

$$\mathbb{E}(X) = \mathbb{E}(X - Y) + \mathbb{E}(Y),$$

where $\mathbb{E}(Y)$ can be explicitly computed and $\mathrm{Var}(X - Y)$ is smaller than $\mathrm{Var}(X)$. In the following part we give the example of a linear adaptive control variate coming from Kim and Henderson [58] or Glynn and Szechtman [36].

9.5.1 Adaptive control Variate

Let us consider the control variate family $(Y(\theta); \theta \in \Theta)$, where Θ is an open set of $\mathbb{R}^p$. We also assume $\mathbb{E}(Y(\theta)) = 0$ for all $\theta \in \Theta$. Then, $\mathbb{E}(X - Y(\theta)) = \mu$. We aim at finding $\theta^\star$ which minimises the variance of $(X - Y(\theta))$. We restrict the problem by dealing with the linear case, e.g.

$$Y = \sum_{i=1}^{p} \theta_i C_i = \theta^T C,$$

where C is a random variable of $\mathbb{R}^p$ such that $\mathbb{E}(C) = 0$ and θ is a deterministic vector of $\mathbb{R}^p$. We approximate μ by

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} (X^i - \theta^T C^i),$$

where $(X^1, \cdots, X^n)$ (resp. $(C^1, \cdots, C^n)$) is a sample of size n following the law of X (resp. C). The strong law of large numbers gives $\mu_n \xrightarrow{p.s.} \mu$ and the Central Limit Theorem states

$$\sqrt{n}(\mu_n - \mu) \xrightarrow{law} \mathcal{N}(0, \mathrm{Var}(X - \theta^T C)),$$

and $\mathrm{Var}(X - \theta^T C) = \mathbb{E}(X - \theta^T C)^2 - (\mathbb{E}(X))^2$. Since we want to minimise the variance, we are looking for $\theta^\star$ which cancels the gradient

$$\nabla_\theta \mathrm{Var}(X - \theta^T C) = -2\mathbb{E}(XC - CC^T\theta). \tag{9.7}$$

If the covariance matrix $\Lambda = \mathrm{Cov}(C, C)$ is invertible, we get

$$\theta^\star = \Lambda^{-1}\beta, \tag{9.8}$$

where $\beta = \mathrm{Cov}(X, C)$ is defined by $\beta_i = \mathrm{Cov}(X, C_i)$, for $i = 1, \cdots, p$. Since $\mathrm{Cov}(X, C)$ and $\mathrm{Cov}(C, C)$ are usually unknown, we can estimate $\theta^\star$ with

$$\theta_n = \Lambda_n^{-1}\beta_n \tag{9.9}$$

with

$$\beta_n = \frac{1}{n} \sum_{j=1}^{n} (X^j C^j - \overline{X}_n \overline{C}_n) \quad \text{and } \Lambda_n = \frac{1}{n} \sum_{j=1}^{n} (C^j (C^j)^T - \overline{C}_n \overline{C}_n^T),$$

where $\overline{C}_n = \frac{1}{n}\sum_{j=1}^{n} C^j$ is a vector in $\mathbb{R}^p$. From this value of θ_n, we deduce two estimators of μ

$$\tilde{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}(X^i - \theta_n^T C^i), \quad \overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}(X^i - \theta_{i-1}^T C^i),$$

where θ_{i-1} is computed using $(X^1, \cdots, X^{i-1})$ and $(C^1, \cdots, C^{i-1})$. $\overline{\mu}_n$ is an adaptive estimator of μ, whereas $\tilde{\mu}_n$ is not. One can show that $\tilde{\mu}_n$ satisfies a Central Limit Theorem, e.g.

$$\sqrt{n}(\tilde{\mu}_n - \mu) \xrightarrow[n\to\infty]{law} \mathcal{N}(0, \tilde{\sigma}^2), \tag{9.10}$$

where $\tilde{\sigma}^2 = \mathrm{Var}(X - Y(\theta^\star))$. Concerning $\overline{\mu}_n$, we can state the following Theorem

Theorem 9.15. *Assume $X \in \mathbb{L}^4(\mathbb{R})$ and $C \in \mathbb{L}^4(\mathbb{R}^p)$. Then, $\overline{\mu}_n$ converges almost surely to μ and*

$$\sqrt{n}(\overline{\mu}_n - \mu) \xrightarrow[n\to\infty]{law} \mathcal{N}(0, \tilde{\sigma}^2).$$

The proof of the previous Theorem is based on the Law of Large Numbers and the Central Limit Theorem for martingales. We refer to Duflo [25], Chapter 2 or to Hall and Heyde [48] for more details on martingales theory.

Instead of approximating $\theta^\star$ defined in (9.8) by θ_n defined in (9.9), we can use the Robbins-Monro algorithm to build θ_n

$$\theta_{n+1} = \theta_n - \gamma_n F(\theta_n, X^{n+1}, C^{n+1}), \tag{9.11}$$

where

$$F(\theta_n, X^{n+1}, C^{n+1}) = 2(X^{n+1}C^{n+1} - C^{n+1}(C^{n+1})^T\theta_n),$$

and γ_n satisfies $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$. This choice for F corresponds to the value of the gradient we want to cancel in (9.7). Theorem Duflo [25], Theorem 1.4.26 enables to prove that θ_n constructed with (9.11) converges almost surely to $\theta^\star$. By plugging this value of θ_n in $\tilde{\mu}_n$ and $\overline{\mu}_n$ defined above, we also get (9.10) and Theorem 9.15.

More generally, we can use an adaptive control variate algorithm to compute $\mathbb{E}(f(X))$, where f is a multivariate smooth function. We refer to Maire [75]. The author studies regular functions f using a Fourier basis on periodised functions, Legendre and Tchebitchev polynomial bases. He reduces the dimensional effect by computing these approximations on Korobov-like spaces.

This part has been devoted to adaptive control variate, but we can also make use of adaptive importance sampling. In Baggerly et al. [4], the authors apply an adaptive importance sampling for a Markov chain with scoring. They establish conditions for exponential convergence of learning algorithms for Markov chains with scoring.

9.5.2 Application to the valuation of $\mathbb{E}(\Psi(X_s, s \geq t)|X_t = x)$

We would like to apply an adaptive control variate method to the numerical valuation of

$$\mathbb{E}(\Psi(X_s, s \geq t)|X_t = x),$$

where $(X_t)_t$ is solution of an Itô stochastic differential equation and Ψ belongs to a class of functionals related to Feynman Kac representations.
The Monte Carlo method for SDEs offers a mean of calculating solutions to certain types of parabolic partial differential equation and so has applications in various fields including stochastic control, particle physics and econometrics. They are usually used for high dimensional problems or when the functionals are complex. Before speaking about an adaptive control variate method applied to compute the above expectation, we refer to Newton [77], where the author develops methods of control variates and importance sampling for the valuation of expectations of the above type. In both cases, a perfect variate (e.g. one which is unbiased and has zero variance) is first constructed by means of the Funke-Shevlyakov-Haussmann integral representation theorem for functionals of Itô processes. These involve terms which cannot be computed exactly but which can be approximated to yield unbiased estimators of the desired integrals with reduced variances.
Typically, to evaluate $u(0, x)$ by Monte Carlo means, where $u(t, x) = \mathbb{E}_{t,x}[g(X_T)]$, the centred perfect control variate is the stochastic integral $Y = \int_0^T \nabla_x u(s, X_s)\sigma(s, X_s)dW_s$. In practice, one uses an approximation of $\nabla_x u$ and a discretization of the stochastic integral. An alternative centred control variate is $v(0, x) + \int_0^T (\partial_t + \mathcal{L})v(s, X_s)ds - v(T, X_T)$, where v approximates u. This has been introduced by Gobet and Maire [40] in an alternative way (adaptive control variate) to solve the Poisson equation with Dirichlet boundary conditions over square domains. By using the Feynman Kac formula, we can write the solution of the Poisson equation as an expectation of the above type. Thus, this method is based on Feynman Kac computations of pointwise solutions combined with a global approximation on Chebyshev polynomials. Pointwise solutions were computed using walk-on-spheres (WOS) simulations of stopped Brownian motion, which induces a simulation error due to the absorption layer thickness. The authors have observed a geometric reduction of both the simulation error and the variance with the number of steps of the algorithm. The global error is comparable to standard deterministic spectral methods while avoiding the resolution of a linear system.
In Gobet and Maire [39], the authors generalise their previous work to the computation of the above expectation where X is a linear Markov process with or without absorbing/reflecting boundary or a jump process. They use an adaptive control variate algorithm to compute Monte Carlo approximations of solutions of linear partial differential equations, which are connected through the Feynman Kac formula to linear Markov processes (with or without absorbing/reflecting boundary) and jump processes. They prove that the bias and the variance decrease geometrically with the number of steps of their algorithm.
Concerning an adaptive importance sampling method to compute such an expectation, we refer to Arouna [3]. The author presents an importance sampling scheme based on a parametric change of drift which is adaptively selected through a Monte Carlo computation by using a suitable sequence of approximation (namely, a Robbins Monro type algorithm).

More precisely, the author aims at computing option prices via Monte Carlo simulations. By the Girsanov theorem, he introduces a drift term into the expectation defining the option price. Then, he uses a truncated version of the Robbins Monro algorithm to find the drift which optimally reduces the variance. The author proves that for a large class of payoff functions, this version of the Robbins Monro algorithm converges a.s. to the optimal drift.

9.6 An algorithm for BSDEs

In the previous section we have seen that, under Hypothesis 9.5, solving the FBSDE (E) boils down to solving PDE $(\mathcal{E})$. Let u denote the solution of the PDE $(\mathcal{E})$ and (Y, Z) the solution of the FBSDE (E). The algorithm provides an approximation of u (denoted u_k) and of its gradient $\partial_x u$ (denoted $\partial_x u_k$). To get (Y^k, Z^k), the approximation of (Y, Z), we use the relation (9.6), where the process X has been replaced by X^N, the approximation of X, and u and $\partial_x u$ have been replaced by u_k and $\partial_x u_k$

$$\forall t \in [0,T],\ Y_t^k = u_k(t, X_t^N),\ \ Z_t^k = \partial_x u_k(t, X_t^N)\sigma(t, X_t^N). \tag{9.12}$$

We approximate the process X by using Euler's scheme. X^N is defined by $X_0^N = x$ and

$$\forall s \in [0,T],\ \ dX_s^N = b(\varphi(s), X_{\varphi(s)}^N)ds + \sigma(\varphi(s), X_{\varphi(s)}^N)dW_s,$$

where $\varphi(s) := \sup\{t_k : t_k \leq s\}$ and $\{0 = t_0 < t_1 < \cdots < t_N = T\}$ is a regular subdivision of the interval $[0,T]$. The time step is denoted $h = \frac{T}{N}$. Now, we describe our algorithm to compute iterative approximations $(u_k)_{k\geq 0}$ of the global solution u. These approximations rely on the computations of $\mathbb{E}[\Psi(\tilde{g_1}, \tilde{g_2}, X(x))]$, where $\tilde{g_1}, \tilde{g_2}$ depend on Φ, f and the approximations of u_{k-1} and their derivatives, at some points $(t_i^k, x_i^k)_{1\leq i\leq n} = (t_i^k, x_i^{k,1}, \cdots x_i^{k,d})_{1\leq i\leq n}$, and where $\Psi(t, g_1(s,\cdot), g_2(\cdot), X(x)) = g_2(X_T^{t,x}) + \int_t^T g_1(s, X_s^{t,x})ds$, for some functions g_1 and g_2. We point out that the choice of these points depends on the iteration. At the end of the description of the algorithm, we will be able to precise how to choose these points.

Initialisation. We begin with $u_0 \equiv 0$.

Iteration k, Step 1. Assume that an approximated solution u_{k-1} of class $C^{1,2}$ is built at stage $k-1$, and that we are able to compute $\partial_x u_{k-1}, \partial_x^2 u_{k-1}, \partial_t u_{k-1}$ at any point (t, x). Applying Itô's formula to u_{k-1} yields $u_{k-1}(t,x) = \mathbb{E}[\Psi\left(t, -(\partial_t + \mathcal{L}^N)u_{k-1}(s,\cdot), u_{k-1}(T,\cdot), X^N(x)\right)]$, where

$$\mathcal{L}^N u(s, X_s^N) = \frac{1}{2}\sum_{i,j}[\sigma\sigma^*]_{ij}(\varphi(s), X_{\varphi(s)}^N)\partial^2_{x_i x_j}u(s, X_s^N) + \sum_i b_i(\varphi(s), X_{\varphi(s)}^N)\partial_{x_i}u(s, X_s^N),$$

We would like to compute a correction $w_k = u - u_{k-1}$ on this approximation. Using Itô's formula and the PDE $(\mathcal{E})$ yield $u(t,x) = \mathbb{E}[\Psi\left(t, f(s,\cdot, u(s,\cdot), (\partial_x u\sigma)(s,\cdot)), \Phi(\cdot), X(x)\right)]$.

Then, we get

$$w_k(t,x) = \mathbb{E}\Big[\Phi(X_T^{t,x}) - u_{k-1}(T, X_T^{N,t,x}) + \int_t^T f\left(s, X_s^{t,x}, u(s,X_s^{t,x}), (\partial_x u\sigma)(s,X_s^{t,x})\right) + (\partial_t + \mathcal{L}^N)u_{k-1}(s, X_s^{N,t,x})ds|\mathcal{G}_{k-1}\Big].$$

Remark 9.16. As we will see in Definition 9.20, $\mathcal{G}_{k-1}$ is the filtration generated by the set of all random variables used to build u_{k-1}. In the above equation, we compute the expectation w.r.t. the law of X and X^N and not to the law of u_{k-1}, which is $\mathcal{G}_{k-1}$ measurable.

However, since u is unknown, we are not able to compute w_k. We introduce a new correction term $\hat{w}_k$, which corresponds to w_k after having replaced u by u_{k-1} in f.

$$\hat{w}_k(t,x) = \mathbb{E}\Big[\Phi(X_T^{t,x}) - u_{k-1}(T, X_T^{N,t,x}) + \int_t^T f\left(s, X_s^{t,x}, u_{k-1}(s,X_s^{t,x}), (\partial_x u_{k-1}\sigma)(s,X_s^{t,x})\right) + (\partial_t + \mathcal{L}^N)u_{k-1}(s, X_s^{N,t,x})ds\Big|\mathcal{G}_{k-1}\Big].$$

We intend to compute a Monte Carlo approximation of $\hat{w}_k(t_i, x_i)$. Consequently, we have

$$\overline{w}_k(t,x) = \frac{1}{M}\sum_{m=1}^{M}\Big[\Phi(X_T^{m,k,N}) - u_{k-1}(T, X_T^{m,k,N}) + \int_t^T f\Big(s, X_s^{m,k,N}, u_{k-1}(s,X_s^{m,k,N}), (\partial_x u_{k-1}\sigma)(s,X_s^{m,k,N})\Big) + (\partial_t + \mathcal{L}^N)u_{k-1}(s, X_s^{m,k,N})ds\Big], \tag{9.13}$$

using M independent simulations of the paths $X^{m,k,N}(x)$. They are also generated independently of everything else. To be as clear as possible, we have removed the superscripts (t,x) from X^N in the above expectation, but we still deal with processes X^N starting from x at time t. X^N is the approximation of X when using Euler scheme, i.e. $X_0^N = x$ and for $t \in [t_k, t_{k+1}]$

$$X_t^N = X_{t_k}^N + b(t_k, X_{t_k}^N)(t - t_k) + \sigma(t_k, X_{t_k}^N)(W_t - W_{t_k}). \tag{9.14}$$

The above relation enables to easily simulate X^N. The computation of the integral in (9.13) can be done in several ways

- we can approximate the integral with a Riemann summation,
- or with a trapeze integration method,
- we can also use that for any function f_0, $\int_t^T f_0(u)du = (T-t)\mathbb{E}(f_0(U))$, where U follows a uniform law on $[t,T]$, i.e. we can approximate $\mathbb{E}[\int_t^T Y_s ds]$ by $\frac{T-t}{M}\sum_{m=1}^{M} Y_{U^m}^m$.

The first two methods are quite long, because we need to compute the value of u_{k-1} and its derivatives in several points for each m in $\{1, \cdots, M\}$. Since we are interested in approximating an expectation, the last method is more efficient: for each m, we pick at random only one s_0 in $[t,T]$, and we compute the value of u_{k-1} and its derivatives in $(s_0, X_{s_0}^{m,k,N})$. Since we compute an average for $m = 1, \cdots, M$, this method doesn't generate any bias error. However, its variance is bigger than the one we would have obtained if we had used one of the first two methods.

Iteration k, Step 2. In order to build a global approximation w_k based on the values $\overline{w}_k(t_i^k, x_i^k)_i$, we use kernel functions. We denote by $\mathcal{P}_n$ the estimator build from kernel functions. We admit here that $\mathcal{P}_n$ is C^1 in time and C_p^2 in space. We refer to Section 11.3, page 119, for a detailed explanation of the construction of $\mathcal{P}_n$. The approximation of a function $v(t,x)$, denoted $\mathcal{P}_n v(t,x)$ can be written

$$\mathcal{P}_n v(t,x) = \frac{r_n(t,x)}{f_n(t,x)} g(2T\lambda(B) f_n(t,x)), \tag{9.15}$$

where

- $r_n(t,x) = \frac{1}{nh_t h_x^d} \sum_{i=1}^n K_t(\frac{t-T_i}{h_t}) K_x(\frac{x-X_i}{h_x}) v(T_i, X_i)$,
- $f_n(t,x) = \frac{1}{nh_t h_x^d} \sum_{i=1}^n K_t(\frac{t-T_i}{h_t}) K_x(\frac{x-X_i}{h_x})$,
- the points $(T_i, X_i)_{1 \le i \le n}$ are uniformly distributed on $[0,T] \times [-a,a]^d$,
- $\lambda(B) = (2a)^d$,
- and g is such that (see Figure 9.1)

$$g(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 & \text{if } y > 1, \\ -y^4 + 2y^2 & \text{if } y \in [0,1], \end{cases} \tag{9.16}$$

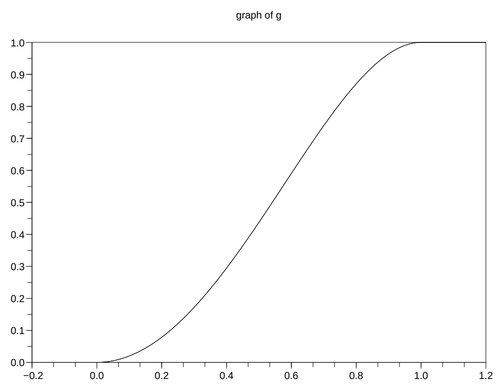

Figure 9.1: graph of g

for some positive and C_p^2 kernel functions K_x, K_t. The functions K_x, K_t can be chosen among several kernel functions (see Härdle [50]) like quartic, triweight, or Gaussian kernels. The study of these kernels is postponed to Chapter 11. Since we want to solve the PDE on $[0,T] \times \mathbb{R}^d$, we have to choose the interval $[-a,a]$ large enough.

Remark 9.17. $(t,x) \longmapsto \mathcal{P}_n v(t,x)$ is a $C_p^{1,2}$ function, i.e. $\mathcal{P}_n v$ is a $C^{1,1}$ function and $\partial_{xx}\mathcal{P}_n v$ is piecewise continuous. Then, we can apply Itô's formula to functions defined by (9.15).

The proof of the convergence of the algorithm has been done using random points $(T_i, X_i)_{1\le i\le n}$ changing at each iteration. This means that in fact $\mathcal{P}_n$ should be written $\mathcal{P}_n^k$, where k denotes the rank of the iteration. To be precise, at iteration k, we have

$$\mathcal{P}_n^k v(t,x) = \frac{r_n^k(t,x)}{f_n^k(t,x)} g(2T\lambda(B) f_n^k(t,x)), \tag{9.17}$$

where

$$r_n(t,x) = \frac{1}{nh_t h_x^d} \sum_{i=1}^n K_t\left(\frac{t-T_i^k}{h_t}\right) K_x\left(\frac{x-X_i^k}{h_x}\right) v(T_i^k, X_i^k),$$
$$f_n^k(t,x) = \frac{1}{nh_t h_x^d} \sum_{i=1}^n K_t\left(\frac{t-T_i^k}{h_t}\right) K_x\left(\frac{x-X_i^k}{h_x}\right),$$

and the points $(T_i^k, X_i^k)_{1\le i\le n}$ are chosen i.i.d. and independent from the points $(T_i^{k'}, X_i^{k'})_{1\le i\le n}$,for $k' \neq k$. We compute $u_k(t,x)$ using the following formula

$$u_k(t,x) = \mathcal{P}_n^k(u_{k-1} + \overline{w}_k)(t,x). \tag{9.18}$$

This formula enables us to get $\partial_x u_k(t,x), \partial_x^2 u_k(t,x)$, and $\partial_t u_k(t,x)$ by computing the derivative w.r.t. x,t of $\mathcal{P}_n^k(u_{k-1} + \overline{w}_k)(t,x)$. We are now able to precise a little the choice of the points $(T_i^k, X_i^k)_i$ at which $\overline{w}_k$ is to be computed. Looking at (9.18) and (9.17), we can see that we need the value of u_{k-1} and $\overline{w_k}$ at $(T_i^k, X_i^k)_i$. Thus, $(t_i^k, x_i^k)_i$ should be equal to $(T_i^k, X_i^k)_i$.

At this stage, we can proceed to the next iteration. We compute (Y^k, Z^k) with the following relation

$$\forall t \in [0,T],\ \ Y_t^k = u_k(t, X_t^N),\ \ Z_t^k = (\partial_x u_k \sigma)(t, X_t^N).$$

Remark 9.18. We need to build an estimator which is C^1 in time and C_p^2 in space since we apply Itô's formula to the functions $u_k, k \ge 0$ (see Iteration k, Step 1), and u_k is built by using $\mathcal{P}_n^k$ (see (9.18)).

Remark 9.19. Since u_k is computed using (9.18), u_k is a random variable depending on the points $\{(T_i^k, X_i^k)_{1\le i\le n}\}$ used by $\mathcal{P}_n^k$, the random variables needed to compute u_{k-1}, and the processes $X^{m,k,N}(x_i^k)$, $1 \le m \le M, 1 \le i \le n$ appearing in the computation of $\overline{y}_k$.

Definition 9.20 (Definition of the filtration $\mathcal{G}_k$)**.** Let $\mathcal{G}_k$ define the filtration generated by the set of all random variables used to build u_k. Using (9.18), we can write

$$\mathcal{G}_k = \mathcal{G}_{k-1} \vee \sigma(\mathcal{A}_k, \mathcal{S}_k),$$

where $\mathcal{A}_k := \{(T_i^k, X_i^k)_{1\le i\le n}\}$, the set of random points used at step k to build the estimator $\mathcal{P}_n^k$, $\mathcal{S}_k := \{X^{m,k,N}(x_i^k), 1 \le m \le M, 1 \le i \le n\}$, the set of independent simulations of the paths $X^{m,k,N}(x_i^k)$, and $\mathcal{G}_{k-1}$ is the filtration generated by the set of all random variables used to build u_{k-1}.

9.7 Pros and cons of our algorithm - Comparison with the other techniques

As we said in the introduction of Part III, there exist several techniques for numerically solving BSDEs. We can list them into three categories

1. the ones which solve the associated PDE with a finite difference approximation (e.g. Ma et al. [74]) ,

2. the ones which solve the dynamic programming equation with Monte Carlo methods $Y^N_{t_N} = \Phi(X^N_{t_N})$ and

$$\begin{aligned} Y^N_{t_k} =& \mathbb{E}_{t_k}(Y^N_{t_{k+1}}) + h\mathbb{E}_{t_k} f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k}), \\ hZ^N_{t_k} =& \mathbb{E}_{t_k}(Y^N_{t_{k+1}} \Delta W^*_k), \end{aligned}$$

   where $\mathbb{E}_{t_k}$ is the conditional expectation w.r.t. $\mathcal{F}_{t_k}$, and X^N satisfies (9.14). Bouchard and Touzi [15] simulates these equations using Malliavin calculus techniques, while Gobet et al. [45] use empirical regression techniques in the spirit of Longstaff-Schwartz algorithm for American options.

3. the ones which use a quantization technique (e.g. Chevance [21], Bally and Pagès [6], Delarue and Menozzi [24]) to implement the dynamic programming equation.

Let us present the advantages and drawbacks of the three methods listed above. The method consisting of solving the associated PDE through a finite difference approximation doesn't perform well in high dimension. The quantization technique presented in Delarue and Menozzi [24] can be used for solving high dimensional problems. However the algorithm provides piecewise constant solutions on cells. The algorithms using Monte Carlo tools and the above dynamic programming equation provide regular solutions in space for each time step t_k. However, the variance of the solutions are not robust in N, i.e. their values blow up when N tends to ∞. Bouchard and Touzi [15] compute the conditional expectations appearing in the dynamic programming equation by using Malliavin calculus integration by parts. The weights computed with Malliavin calculus require a lot of computational time. These algorithms based on the dynamic programming equation do not take advantage of the time regularity of the solution, and this is a major difference with ours.
Our algorithm provides a regular solution in space AND in time. Moreover, we use Monte Carlo methods which still run well in high dimensions and which are quite accurate here since we use adaptive control variates. However, the operator $\mathcal{P}$ depends on the dimension, which will probably affects the speed of the algorithm.

9.8 Glimpse of the ingredients for the proof of the convergence

For proving the convergence of the algorithm, we need to bound $\mathcal{P}_n v - v$ and $\partial_x(\mathcal{P}_n v) - \partial_x v$ in a weighted Sobolev norm $\|\cdot\|^2_{H^\mu_{\beta,X}}$, which depends on β, appearing in the definition of the space $\mathbb{H}^2_{T,\beta}$ introduced in Section 9. The $\|\cdot\|_\beta$ norm, associated to this space, enables to state a priori estimates for BSDEs:

$$\|v\|^2_{H^\mu_{\beta,X}} = \int_0^T \int_{\mathbb{R}^d} e^{-\mu|x|} e^{\beta s}\mathbb{E}|v(s, X^x_s)|^2 dxds.$$

This choice of norm is crucial. This norm also depends on μ, a parameter which appears in the technical PDE results of Part II. Since the building of $\mathcal{P}_n$ is based on a kernel estimator, the error $\|\mathcal{P}_n v - v\|^2_{H^\mu_{\beta,X}}$ will be bounded by terms of order h_t^2, h_x^4 and $\delta_n = \frac{1}{nh_t h_x^d}$. This result comes from the study of $\mathbb{E}[\frac{r_n(t,x)}{f_n(t,x)} - v(t,x)]^2$, where $\frac{r_n(t,x)}{f_n(t,x)}$ corresponds to a standard kernel estimator (see the beginning of Chapter 11). To bound $\mathbb{E}[\frac{r_n(t,x)}{f_n(t,x)} - v(t,x)]^2$, we split it in two terms: the bias error $(\mathbb{E}[\frac{r_n(t,x)}{f_n(t,x)} - v(t,x)])^2$ and the variance error $\mathrm{Var}(\frac{r_n(t,x)}{f_n(t,x)})$. The bias error is bounded by terms of order h_t^2 and h_x^4. The variance error provides an error of order $\delta_n = \frac{1}{nh_t h_x^d}$. Since $\mathcal{P}_n v(t,x)$ corresponds to $\frac{r_n(t,x)}{f_n(t,x)} g(2T\lambda(B) f_n(t,x))$, which is not an estimator studied in the literature, we detail the computations of $\|\mathcal{P}_n v - v\|^2_{H^\mu_{\beta,X}}$ in Chapter 11.

Moreover, the $\|\cdot\|^2_{H^\mu_{\beta,X}}$ norm involves the trajectory of the process X which take values in $\mathbb{R}^d$. Since we numerically choose the points $(T_i, X_i)_{1\leq i\leq n}$ in $[0,T]\times[-a,a]^d$, some errors linked to the truncation induced by the kernel function appear. They are of order h_t and h_x. They also depend on a, μ and β. We refer to Theorem 12.42 for an overview of the upper bound for $\|\mathcal{P}_n v - v\|^2_{H^\mu_{\beta,X}}$.

The study of $\|\partial_x(\mathcal{P}_n v) - \partial_x v\|^2_{H^\mu_{\beta,X}}$ leads to error terms of order h_x^2, $\delta_n = \frac{1}{nh_t h_x^{d+2}}$, coming from $\mathbb{E}[\frac{\partial_x r_n(t,x)}{f_n(t,x)} - \partial_x v(t,x)]^2$. The truncation error terms are of order $\frac{h_t}{h_x}$ and $\frac{1}{h_x}$.

Chapter 10

Regression Analysis

We refer the reader to Györfi et al. [47], Chapter 1 to get a more detailed introduction on regression analysis, and to Györfi et al. [47], Chapter 2 and Härdle [50] for more details on non parametric regression.

10.1 Introduction

In regression analysis, one considers a random vector (X, Y), where X is $\mathbb{R}^d$-valued and Y is $\mathbb{R}$-valued, and one is interested in how the value of the so-called response variable Y depends on the value of the observation vector X. This means that one wants to find a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, such that $f(X)$ is a "good approximation of Y", that is, $f(X)$ should be close to Y in some sense, which is equivalent to making $|f(X) - Y|$ "small". Since X and Y are random vectors, $|f(X) - Y|$ is random as well, therefore we introduce the so-called $\mathbb{L}^2$ *risk or mean squared error* of f,

$$\mathbb{E}|f(X) - Y|^2,$$

and we try to make it as small as possible. So, we are interested in finding a (measurable) function $m^\star : \mathbb{R}^d \to \mathbb{R}$ such that

$$\mathbb{E}|m^\star(X) - Y|^2 = \min_{f:\mathbb{R}^d \to \mathbb{R}} \mathbb{E}|f(X) - Y|^2.$$

We can easily check that

$$m(x) = \mathbb{E}(Y|X = x)$$

minimises the $\mathbb{L}^2$ risk. m is called the regression function.

Practically, the distribution of (X, Y) (and hence the regression function) is usually unknown. Therefore it is impossible to predict Y using $m(X)$. But it is often possible to observe data according to the distribution of (X, Y) and to estimate the regression function from these data. More precisely, let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be i.i.d. random variables following the law of (X, Y) with $\mathbb{E}Y^2 < \infty$. Let D_n be the set of data defined by $D_n = \{(X_1, Y_1), \cdots, (X_n, Y_n)\}$. In the regression function estimation one uses the data

D_n to construct an estimate $m_n : \mathbb{R}^d \to \mathbb{R}$ of the regression function m.
In general, estimates won't be equal to the regression function. To compare different estimates, we need an error criterion to measure the difference between the regression function and an arbitrary estimate m_n. In the literature, several distinct error criteria are used, like the pointwise error, the supremum norm error, and the $\mathbb{L}^p$ error

$$\int_C |m_n(x) - m(x)|^p dx,$$

where the integration is performed w.r.t. the Lebesgue measure, C is a fixed subset of $\mathbb{R}^d$, and $p \geq 1$ is arbitrary (often $p = 2$ is used). Recall that the main goal is to find a function f such that the $\mathbb{L}^2$ risk $\mathbb{E}|f(X) - Y|^2$ is small. The minimal value of this $\mathbb{L}^2$ risk is reached by the regression function m. One can show that the $\mathbb{L}^2$ risk $\mathbb{E}[|m_n(X) - Y|^2 \big| D_n]$ of an estimate m_n satisfies

$$\mathbb{E}[|m_n(X) - Y|^2 \big| D_n] = \int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu(dx) + \mathbb{E}|m(X) - Y|^2,$$

where μ denotes the distribution of X. Thus, the $\mathbb{L}^2$ risk of an estimate m_n is close to the optimal value if and only if the $\mathbb{L}^2$ error

$$\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu(dx) \tag{10.1}$$

is close to zero. Therefore, we will use the $\mathbb{L}^2$ error (10.1) in order to measure the quality of an estimate.

10.2 Consistency

Now, we define the modes of convergence of the regression estimates usually used. The first and weakest property an estimate should satisfy is the following: as the sample size grows, the estimator should converge to the estimated quantity, i.e., the error of the estimate should converge to zero for a sample size tending to infinity. Such estimates are called consistent. To measure the error of a regression estimate, we use the $\mathbb{L}^2$ error defined in (10.1).
The estimate m_n depends on the data D_n, so the $\mathbb{L}^2$ error (10.1) is a random variable. We are interested in the convergence in mean of this random variable to zero as well as in the almost sure convergence of this random variable to zero.

Definition 10.1. A sequence of regression function estimates $\{m_n\}$ is called **strongly consistent for a certain distribution of** (X, Y), if

$$\lim_{n\to\infty} \int (m_n(x) - m(x))^2 \mu(dx) = 0, \text{ with probability one.}$$

A regression function estimate may be consistent for a certain class of distributions of (X, Y), but not consistent for others. It is clearly desirable to have estimates that are consistent for a large class of distributions. We are interested in distribution-free or universal properties of m_n. The concept of universal consistency is important in non parametric regression because the mere use of a non parametric estimate is normally a consequence of a partial or total lack of information about the distribution of (X, Y). Since in many situations we do not have any prior information about the distribution, it is essential to have estimates that perform well for all distributions. This very strong requirement of universal goodness is formulated as follows

Definition 10.2. A sequence of regression function estimates $\{m_n\}$ is called **weakly universally consistent** (resp. **strongly universally consistent**) if it is weakly (resp. strongly) consistent for all distributions of (X, Y) with $\mathbb{E}(Y^2) < \infty$.

10.3 Rate of convergence

If an estimate is universally consistent, then the $\mathbb{L}^2$ error of the estimate converges to zero for a sample size tending to infinity, regardless of the true underlying distribution of (X, Y). But this says nothing about how fast this happens. Clearly, it is desirable to have estimates for which the $\mathbb{L}^2$ error converges to zero as fast as possible.
To study the rate of convergence of an estimate m_n, we look at the expectation of the $\mathbb{L}^2$ error

$$\mathbb{E}\int |m_n(x) - m(x)|^2 \mu(dx). \tag{10.2}$$

A natural question to ask is whether there exist estimates for which (10.2) converges to zero at some fixed, nontrivial rate for all distributions of (X, Y). Unfortunately, such estimates do not exist, i.e., for any estimate the rate of convergence may be arbitrarily slow. In order to get nontrivial rates of convergence, one has to restrict the class of distributions, e.g., by imposing some smoothness assumptions on the regression function.

10.4 Parametric Estimation

The classical approach for estimating a regression function is the so-called parametric regression function. Here, one assumes that the structure of the regression function is known and depends only on finitely many parameters, and one uses the data to estimate the (unknown) values of these parameters.
The linear regression estimate is an example of such an estimate. In linear regression, one assumes that the regression function is a linear combination of the components of $x = (x^1, \cdots, x^d)$, i.e.

$$m(x^1, \cdots, x^d) = a_0 + \sum_{i=1}^{d} a_i x^i \quad ((x^1, \cdots, x^d) \in \mathbb{R}^d)$$

for some unknown $a_0, \cdots, a_d \in \mathbb{R}$. Then, one uses the data to estimate these parameters, e.g., by applying the least square principle, where one chooses the coefficients $a_0, \cdots, a_d$ of the linear function such that it best fits the given data:

$$(\hat{a}_0, \cdots, \hat{a}_d) = \underset{a_0, \cdots, a_d \in \mathbb{R}^d}{\arg\min} \left\{ \frac{1}{n} \sum_{j=1}^{n} \left| Y_j - a_0 - \sum_{i=1}^{d} a_i X_j^i \right|^2 \right\}.$$

Here X_j^i denotes the i-th component of X_j. Finally, one defines the estimate by

$$\hat{m}_n(x) = \hat{a}_0 + \sum_{i=1}^{d} \hat{a}_i x^i \quad ((x^1, \cdots, x^d) \in \mathbb{R}^d).$$

Parametric estimates usually depend on few parameters, therefore they are suitable even for small sample sizes n, if the parametric model is appropriately chosen. Furthermore, they are often easy to interpret. For instance, in a linear model, the absolute value of the coefficient $\hat{a}_i$ indicates how much the i-th component of X affects the value of Y, and the sign of $\hat{a}_i$ describes the nature of the influence (increases or decreases the value of Y).
However, parametric estimates have a big drawback. Regardless of the data, a parametric estimate cannot approximate the regression function better than the best function which has the assumed parametric structure. For example, a linear regression estimate will produce a large error for every sample size, if the true underlying regression function is not linear and cannot be well approximated by linear functions.
For univariate X, one can often use a plot of the data to choose a proper parametric estimate. For multivariate X, there is no easy way to visualise the data. Thus, especially for multivariate X, it is not clear how to choose a proper form of a parametric estimate, and a wrong form will undoubtedly lead to a bad estimate. This inflexibility concerning the structure of the regression function is avoided by so-called non parametric regression estimates.

10.5 Non parametric Regression

In this section, we describe four paradigms of nonparametric regression : **local averaging**, **local modelling**, **global modelling** (or **least square estimation**) and **penalised modelling**. Recall that the data can be written as

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, \cdots, n.$$

where $\epsilon_i = Y_i - m(X_i)$ satisfies $\mathbb{E}(\epsilon_i \big| X_i) = 0$. Thus, Y_i can be considered as the sum of the value of the regression function at X_i and some error ϵ_i where the expected value of the error is zero. This motivates the construction of the estimates by **local averaging**, i.e. estimation of $m(x)$ by the average of those Y_i where X_i is "close" to x. Such an estimate can be written as

$$m_n(x) = \sum_{i=1}^{n} W_{n,i}(x) \cdot Y_i,$$

where the weights $W_{n,i}(x) = W_{n,i}(x, X_1, \cdots, X_n) \in \mathbb{R}$ depend on $X_1, \cdots, X_n$. Usually the weights are nonnegative and $W_{n,i}(x)$ is "small" if X_i is "far" from x.
An example of such an estimate is the *partitioning estimate*. Here, one chooses a finite or countably infinite partition $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, \cdots\}$ of $\mathbb{R}^d$ consisting of Boole sets $A_{n,j} \subseteq \mathbb{R}^d$ and defines, for $x \in A_{n,j}$, the estimate by averaging Y_i's when the corresponding X_i's is in $A_{n,j}$, i.e.

$$m_n(x) = \frac{\sum_{i=1}^{n} \mathbf{1}_{X_i \in A_{n,j}} Y_i}{\sum_{i=1}^{n} \mathbf{1}_{X_i \in A_{n,j}}} \quad \text{for } x \in A_{n,j}.$$

where $\mathbf{1}_A$ is the indicator function of the A set, so $W_{n,i}(x) = \frac{\mathbf{1}_{X_i \in A_{n,j}}}{\sum_{i=1}^n \mathbf{1}_{X_i \in A_{n,j}}}$, for $x \in A_{n,j}$.

We use the convention $\frac{0}{0} = 0$.

The second example of a local averaging estimate is the *Nadaraya-Watson estimate.* Let $K : \mathbb{R}^d \to \mathbb{R}_+$ be a so-called kernel function, and let $h > 0$ be a given bandwidth. The kernel estimate is defined by

$$m_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}, \tag{10.3}$$

so $W_{n,i}(x) = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h}\right)}$. If one uses the naive kernel $K(x) = \mathbf{1}_{|x| \leq 1}$, then

$$m_n(x) = \frac{\sum_{i=1}^n \mathbf{1}_{|x-X_i| \leq h} Y_i}{\sum_{i=1}^n \mathbf{1}_{|x-X_i| \leq h}}, \tag{10.4}$$

i.e. one estimates $m(x)$ by averaging Y_i's such that the distance between X_i and x is not greater than h. Fore more general $K : \mathbb{R}^d \to \mathbb{R}_+$ one uses a weighted average of the Y_i, where the weight of Y_i depends on the distance between X_i and x.

The kernel estimate (10.3) can be considered as locally fitting a constant to the data. In fact, it is easy to see (Györfi et al. [47], Problem 2.2, page 29) that one has

$$m_n(x) = \arg\min_{c \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) (Y_i - c)^2.$$

A generalisation of this leads to the **local modelling** paradigm: instead of locally fitting a constant to the data, one locally fits a more general function which depends on several parameters. Let $g(\cdot, \{a_k\}_{k=1}^l) : \mathbb{R}^d \to \mathbb{R}$ be a function depending on the parameters $\{a_k\}_{k=1}^l$. For each $x \in \mathbb{R}^d$, choose the values of these parameters by a local least square criterion

$$\{\hat{a}_k(x)\}_{k=1}^l = \arg\min_{\{a_k\}_{k=1}^l} \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) (Y_i - g(X_i, \{a_k\}_{k=1}^l))^2.$$

Evaluate the function g for these parameters at the point x and use this as an estimate of $m(x)$:

$$m_n(x) = g(x, \{\hat{a}_k(x)\}_{k=1}^l).$$

The most popular example of a local modelling estimate is the *local polynomial kernel estimate.* Here, one locally fits a polynomial to the data. For example, for $d = 1$, X is real-valued and $g(x, \{a_k(x)\}_{k=1}^l) = \sum_{k=1}^l a_k x^{k-1}$.

A generalisation of the partitioning estimate leads to **global modelling** or **least square estimates**. Let $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, \cdots\}$ be a partition of $\mathbb{R}^d$ and let $\mathcal{G}_n$ be the set of all piecewise constant functions with respect to the partition, i.e.,

$$\mathcal{G}_n = \left\{ \sum_j a_j \mathbf{1}_{A_{n,j}} \ : \ a_j \in \mathbb{R} \right\}. \tag{10.5}$$

Then, it is easy to see (Györfi et al. [47], Problem 2.3 page 29) that the partitioning estimate (10.4) satisfies

$$m_n(\cdot) = \arg\min_{f\in\mathcal{G}_n}\left\{\frac{1}{n}\sum_{i=1}^n |f(X_i)-Y_i|^2\right\}. \tag{10.6}$$

Hence, it minimises the empirical $\mathcal{L}^2$ risk

$$\frac{1}{n}\sum_{i=1}^n |f(X_i)-Y_i|^2 \tag{10.7}$$

over $\mathcal{G}_n$. Least square estimates are defined by minimising the empirical $\mathbb{L}^2$ risk over a general set of functions $\mathcal{G}_n$ (instead of (10.5)). Observe that it doesn't make sense to minimise (10.7) over all (measurable) functions f, because this may lead to a function which interpolates the data and hence is not a reasonable estimate. Thus, one has to restrict the set of functions over which one minimises the empirical $\mathbb{L}^2$ risk. Examples of possible choices for the set $\mathcal{G}_n$ are sets of piecewise polynomials w.r.t. a partition $\mathcal{P}_n$ or sets of smooth piecewise polynomials (splines). The use of spline spaces ensures that the estimate is a smooth function.
Instead of restricting the set of functions over which one minimises, one can also add a penalty term to the functional to be minimised. Let $J_n(f)\geq 0$ be a term penalising the "roughness" of a function f. The **penalised modelling** or **penalised least square estimate** m_n is defined by

$$m_n = \arg\min_{f}\left\{\frac{1}{n}\sum_{i=1}^n |f(X_i)-Y_i|^2 + J_n(f)\right\}, \tag{10.8}$$

where one minimises over all measurable functions f. A popular choice for $J_n(f)$ in the case $d=1$ is

$$J_n(f) = \lambda_n\int |f''(t)|^2 dt,$$

where λ_n is some positive constant.

10.6 Bias-Variance Tradeoff

Let m_n be an arbitrary estimate. For any $x\in\mathbb{R}^d$, we can write the expected error of m_n at x as

$$\begin{aligned}\mathbb{E}[|m_n(x)-m(x)|^2] &= \mathbb{E}[|m_n(x)-\mathbb{E}[m_n(x)]|^2] + |\mathbb{E}[m_n(x)]-m(x)|^2\\ &= \mathrm{Var}(m_n(x)) + |\mathrm{Bias}(m_n(x))|^2.\end{aligned}$$

This also leads to a similar decomposition of the following $\mathbb{L}^2$ norm

$$\begin{aligned}\mathbb{E}[\int |m_n(x)-m(x)|^2\mu(dx) &= \int \mathbb{E}[|m_n(x)-m(x)|^2]\mu(dx)\\ &= \int \mathrm{Var}(m_n(x))\mu(dx) + \int |\mathrm{Bias}(m_n(x))|^2\mu(dx).\end{aligned}$$

The importance of these decompositions is that the integrated variance and the integrated squared bias depend in opposite ways on the wiggliness of an estimate. If one increases the wiggliness of an estimate, then usually the integrated bias will decrease, but the integrated variance will increase (so-called **bias-variance tradeoff**). For the naive kernel (10.4) and under regularity conditions on the underlying distribution, one has

$$\int \operatorname{Var}(m_n(x))\mu(dx) = c_1 \frac{1}{nh^d} + o\left(\frac{1}{nh^d}\right), \quad \int |\operatorname{Bias}(m_n(x))|^2 \mu(dx) = c_2 h^2 + o(h^2).$$

Here, h denotes the bandwidth of the kernel estimate which controls the wiggliness of the estimate, c_1 is some constant depending on the conditional variance $\operatorname{Var}(\cdot \big| X = x)$, the regression function is assumed to be Lipschitz continuous and c_2 is some constant depending on the Lipschitz constant. The value $h^\star$ of the bandwidth for which the sum of the integrated variance and the squared bias is minimal depends on c_1 and c_2. Since the underlying distribution, and hence c_1 and c_2, are unknown, in applications it is important to have methods which choose the bandwidth automatically using only the data D_n. Such methods are described in Györfi et al. [47], Part 2.4 page 26.

Chapter 11

Kernel Estimates

The preceding chapter was devoted to introducing the regression method, and particularly the nonparametric regression. We have presented four paradigms of nonparametric regression. In this chapter, we focus on a local averaging estimate : the Nadaraya-Watson estimate. The chapter is organised as follows : first, we present different kernel functions. Then, we state a result on the consistency of kernel estimates and on the rate of convergence for a naive kernel. Finally, we build the estimator $\mathcal{P}_n$, required by our algorithm (see 9.15), from the regression function m_n described in the preceding section and give some properties on $\mathcal{P}_n$.

11.1 Introduction

The kernel estimate of a regression function takes the form, for $x \in \mathbb{R}^d$

$$m_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} = \sum_{i=1}^n W_{n,i}(x) Y_i. \tag{11.1}$$

The second equality shows that the Nadaraya-Watson estimator can be seen as a weighted (local) average of the response variables Y_i (note $\frac{1}{n}\sum_{i=1}^n W_{n,i}(x) = 1$). In fact, the Nadaraya-Watson estimator shares this weighted local average property with several other smoothing techniques, e.g. k-nearest neighbour and spline smoothing (see Härdle [50], Sections 3.2 and 3.4). Here, the bandwidth $h_n > 0$ depends only on the sample size n, and the function $K : \mathbb{R}^d \to \mathbb{R}_+$ is called a kernel. See Table 11.1 for some examples.

- Note that the bandwidth h_n determines the degree of smoothness of m_n. To see this, let h_n go to either extreme. If $h_n \to 0$, then $W_{n,i} \to n$ if $x = X_i$ and is not defined elsewhere. Hence, at an observation X_i, $m_n(X_i) \to Y_i$, i.e. we get an interpolation of the data. On the other hand, if $h_n \to \infty$, then $W_{n,i} \to 1$ for all values of x, and $m_n(X_i) \to \overline{Y}$, i.e. the estimator is a constant function that assigns the sample mean of Y to each x. Choosing h_n such that a good compromise between over and under smoothing is achieved, is a burning issue.

- If the denominator is equal to zero, the numerator is also equal to zero, and the estimate is set to zero.

Uniform	$\frac{1}{2}\mathbf{1}_{\lvert x\rvert\leq 1}$
Triangle	$(1-\lvert x\rvert)\mathbf{1}_{\lvert x\rvert\leq 1}$
Epanechnikov	$\frac{3}{4}(1-x^2)\mathbf{1}_{\lvert x\rvert\leq 1}$
Quartic	$\frac{15}{16}(1-x^2)^2\mathbf{1}_{\lvert x\rvert\leq 1}$
Triweight	$\frac{35}{32}(1-x^2)^3\mathbf{1}_{\lvert x\rvert\leq 1}$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$
Cosinus	$\frac{\pi}{4}\cos(\frac{\pi}{2}x)$

Table 11.1: Kernel Functions

- There are applications, especially in biology, where the researcher is able to control the values that the predictor variable X will take on and Y is the sole random variable. Hence X will no longer be a random variable, while Y still is. This setup is usually referred to as the *fixed design*. In that case, the Nadaraya-Watson estimator employs weights of the form $W_{n,i}(x) = \frac{K(\frac{x-x_i}{h_n})}{h_n\mu(x)}$, where μ, the density of the points $(x_i)_{1\leq i\leq n}$ is known (it is induced by the researcher). We refer to Gasser and Müller [33] and Gasser and Müller [34] for more details on fixed design model.

We recall that we deal with random design models.

11.2 Consistency

The first result, coming from Härdle [50], Proposition 3.1.1, states that $m_n(x)$ converges in probability to $m(x)$.

Proposition 11.1 (Härdle [50], Proposition 3.1.1)**.** *Assume the stochastic design model m_n defined in (11.1) with a one-dimensional predictor variable X and*

1. $\int |K(x)|dx < \infty$,
2. $\lim_{|x|\to\infty} xK(x) = 0$,
3. $\mathbb{E}Y^2 < \infty$,
4. $n \to \infty,\;\; h_n \to 0,\;\; nh_n \to \infty.$

Then, at every point of continuity of $m(x)$, $\mu(x)$ and $\sigma^2(x)$, with $\mu(x) > 0$,

$$\sum_{i=1}^{n} W_{n,i}(x)Y_i \xrightarrow{\mathbb{P}} m(x).$$

The second result, coming from Györfi et al. [47], Theorem 5.1 proves the weak universal consistency of kernel estimates under general conditions on h_n and K. The proof uses Stone's Theorem (see Györfi et al. [47], Theorem 4.1).

Theorem 11.2 (Györfi et al. [47], Theorem 5.1)**.** *Assume that there are balls $S_{0,r}$ of radius r and balls $S_{0,R}$ of radius R centred at the origin ($0 < r \leq R$), and a constant $b > 0$ such that*

$$b\mathbf{1}_{x \in S_{0,r}} \leq K(x) \leq b\mathbf{1}_{x \in S_{0,R}}$$

(boxed kernel) and consider the kernel estimate m_n. If $h_n \to 0$ and $nh_n \to \infty$, then the kernel estimate is weakly universally consistent (see Definition 10.2).

Finally, the following Theorem bounds the rate of convergence of $\mathbb{E}\int |m_n(x) - m(x)|^2\mu(dx)$ for a naive kernel and a Lipschitz continuous regression function.

Theorem 11.3 (Györfi et al. [47], Theorem 5.2)**.** *For a kernel estimate with a naive kernel assume that $Var(Y \big| X = x) \leq \sigma^2$ for all $x \in \mathbb{R}^d$, $|m(x) - m(z)| \leq C|x - z|$ for all $x, z \in \mathbb{R}^d$, and X has a compact support S^*. Then,*

$$\mathbb{E}\int |m_n(x) - m(x)|^2\mu(dx) \leq \hat{c}\frac{\sigma^2 + \sup_{z\in S^*}|m(z)|^2}{nh_n^d} + C^2h_n^2,$$

where $\hat{c}$ depends only on the diameter of S^* and on d. Thus, for

$$h_n = c'\left(\frac{\sigma^2 + \sup_{z\in S^*}|m(z)|^2}{C^2}\right)^{\frac{1}{d+2}} n^{-\frac{1}{d+2}},$$

we have

$$\mathbb{E}\int |m_n(x) - m(x)|^2\mu(dx) \leq c''(\sigma^2 + \sup_{z\in S^*}|m(z)|^2)^{\frac{2}{d+2}} C^{2d/(d+2)} n^{-\frac{2}{d+2}}.$$

Remark 11.4. Theorem 11.3 only concerns the upper bound $\mathbb{E}\int |m_n(x) - m(x)|^2\mu(dx)$. The study of $\mathbb{E}\int |m_n'(x) - m'(x)|^2\mu(dx)$ has not been done. Moreover, the integral is computed w.r.t. the measure $\mu(dx)$, whose support is compact.

11.3 Construction of our estimator

We aim at explaining the construction of our estimator $\mathcal{P}_n$ (see (9.15), page 105), from the kernel estimator m_n (see (11.1), page 117),

$$m_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} = \sum_{i=1}^n W_{n,i}(x)Y_i.$$

We recall that we want to build an estimator which approximates a function v and its derivatives $v, \partial_x v, \partial_x^2 v, \partial_t v$ at some point $(t,x) \in [0,T] \times \mathbb{R}^d$, while we only know $u(t_i, x_i^j), 1 \leq i \leq n, 1 \leq j \leq d$. This estimator should consequently be C^1 in time and C_p^2 in space. Adding time in the above expression for m_n leads to

$$m_n(t,x) = \frac{\sum_{i=1}^n K_t\left(\frac{t-T_i}{h_t}\right)K_x\left(\frac{x-X_i}{h_x}\right)Y_i}{\sum_{i=1}^n K_t\left(\frac{t-T_i}{h_t}\right)K_x\left(\frac{x-X_i}{h_x}\right)} = \sum_{i=1}^n W_{n,i}(t,x)Y_i,$$

where $K_t : \mathbb{R} \to \mathbb{R}^+$ and $K_x : \mathbb{R}^d \to \mathbb{R}^+$. We also choose $h_t \neq h_x$. m_n has been introduced to approximate the unknown function $m(x) = \mathbb{E}[Y|X = x]$, when we only have at our disposal a finite set of data $D_n = \{(T_1, X_1, Y_1), \cdots, (T_n, X_n, Y_n)\}$.
In our case, we know the relation between T, X and Y since $Y = v(T, X)$ (then, $m \equiv v$), when assuming that the statistic error $\epsilon = 0$. However, we only know the value of v at some points (T_i, X_i), and we need to quickly compute the value of $v, \partial_t v, \partial_x v, \partial_x^2 v$ at some point (t, x). That's why we have chosen to write $\mathcal{P}_n v(t, x)$ as a weighted linear combination of $v(T_i, X_i), 1 \leq i \leq n$, where the weights $W_{n,i}(t, x)$ are easily differentiable. We choose the same weights as for m_n: $W_{n,i}(t,x) = \frac{K_t(\frac{t-T_i}{ht})K_x(\frac{x-X_i}{hx})}{\sum_{i=1}^n K_t(\frac{t-T_i}{ht})K_x(\frac{x-X_i}{hx})}$, where K_t and K_x are C_p^2 positive functions.
Besides that, $\mathcal{P}_n$ should be continuous and differentiable. Therefore, we multiply $\sum_{i=1}^n W_{n,i}(t,x)v(T_i,X_i)$ by a regularising function g computed at $2T\lambda(B)\frac{1}{nh_th_x^d}\sum_{i=1}^n K_t(\frac{t-T_i}{h_t})K_x\left(\frac{x-X_i}{h_x}\right)$ to get

$$\begin{aligned}
\mathcal{P}_n v(t,x) &= \sum_{i=1}^n W_{n,i}(t,x)v(T_i,X_i)g\left(2T\lambda(B)\frac{1}{nh_th_x^d}\sum_{i=1}^n K_t\left(\frac{t-T_i}{h_t}\right)K_x\left(\frac{x-X_i}{h_x}\right)\right) \\
&= r_n(t,x)\frac{g(2T\lambda(B)f_n(t,x))}{f_n(t,x)}
\end{aligned}$$

where

$$\begin{aligned}
r_n(t,x) &= \frac{1}{nh_th_x^d}\sum_{i=1}^n K_t\left(\frac{t-T_i}{h_t}\right)K_x\left(\frac{x-X_i}{h_x}\right)v(T_i,X_i), \\
f_n(t,x) &= \frac{1}{nh_th_x^d}\sum_{i=1}^n K_t\left(\frac{t-T_i}{h_t}\right)K_x\left(\frac{x-X_i}{h_x}\right).
\end{aligned}$$

Remark 11.5. We compute g at $2T\lambda(B)\frac{1}{nh_th_x^d}\sum_{i=1}^n K_t(\frac{t-T_i}{h_t})K_x\left(\frac{x-X_i}{h_x}\right)$ for the following reason: the function $f_n(t,x)$ converges to $\frac{1}{T\lambda(B)}$ when n goes to infinity for $t \in]0,T[$ and $|x_i| < a$, $i = 1, \cdots, d$. Then, $g\left(2T\lambda(B)\frac{1}{nh_th_x^d}\sum_{i=1}^n K_t(\frac{t-T_i}{h_t})K_x\left(\frac{x-X_i}{h_x}\right)\right)$ converges to 1 when n goes to ∞. Hence, if $f_n \sim \frac{1}{T\lambda(B)}$, $\mathcal{P}_n v(t,x) = \frac{r_n(t,x)}{f_n(t,x)}$, which is a standard estimator. The function g has an impact on $\mathcal{P}_n$ only when f_n is strictly positive and small (compared $\frac{1}{T\lambda(B)}$).

The functions r_n and f_n are C_p^2 functions. We choose the following C_p^2 function g,

$$g(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 & \text{if } y > 1, \\ -y^4 + 2y^2 & \text{if } y \in [0,1]. \end{cases}$$

We refer to Figure 9.1 for a graph of g. g regularises $\sum_{i=1}^n W_{n,i}(t,x)v(T_i,X_i)$ and makes $\mathcal{P}_n$ a C_p^2 function in space and C^1 function in time, because

- g makes $\mathcal{P}_n v(t,x)$ continuous, since contrary to Härdle [50], page 25 and Györfi et al. [47], page 19, we don't need specific convention like $\frac{0}{0} = 0$,

- $\mathcal{P}_n v(t,x) = r_n(t,x)\frac{g(2T\lambda(B)f_n(t,x))}{f_n(t,x)} = 2T\lambda(B)r_n(t,x)G(2T\lambda(B)f_n(t,x))$, where $G(y) := \frac{g(y)}{y}$,

$$G(y) = \begin{cases} 0 & \text{if } y < 0, \\ \frac{1}{y} & \text{if } y > 1, \\ -y^3 + 2y & \text{if } y \in [0,1]. \end{cases}$$

(we plot on Figure 11.1 the graph of G).

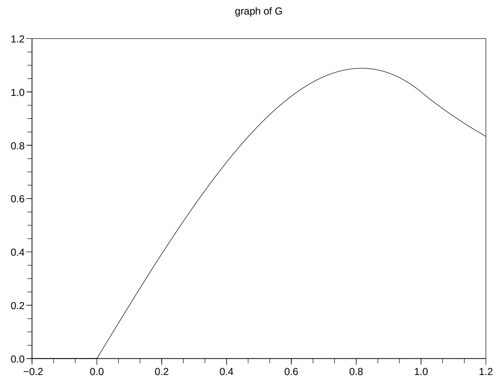

Figure 11.1: graph of G

Then, G is continuous, G' is continuous in 1 but not in 0, since $G'(0) = 0$ and $\lim_{y\searrow 0} G'(y) = 2$. $\partial_{x_i}\mathcal{P}_n v(t,x) = 2T\lambda(B)\partial_{x_i} r_n(t,x)G(2T\lambda(B)f_n(t,x)) + (2T\lambda(B))^2 r_n(t,x)\partial_{x_i} f_n(t,x)G'(2T\lambda(B)f_n(t,x))$. r_n and f_n are C_p^2 functions. Since $G'(2T\lambda(B)f_n(t,x))$ is discontinuous on $\{t,x : f_n(t,x) = 0\}$, which is also $\{t,x : r_n(t,x) = 0\}$, $(2T\lambda(B))^2 r_n(t,x)\partial_{x_i} f_n(t,x)G'(2T\lambda(B)f_n(t,x))$ is continuous in (t,x). Then $\forall i \in \{1,\cdots,d\}$, $\partial_{x_i}(\mathcal{P}_n v)(t,x)$ is continuous. We prove in the same way that $\partial_t(\mathcal{P}_n v)(t,x)$ is continuous.

- Since G is continuous, r_n and f_n are C_p^2 functions and G', G'' are piecewise continuous, $\forall (i,j) \in \{1,\cdots,d\}$, $\partial^2 x_i x_j \mathcal{P}_n v(t,x)$ is piecewise continuous.

We summarise this paragraph in the following proposition

Proposition 11.6. *The estimator $\mathcal{P}_n v(t,x)$ of a function $v : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ defined by*

$$\mathcal{P}_n v(t,x) = r_n(t,x)\frac{g(2T\lambda(B)f_n(t,x))}{f_n(t,x)} \tag{11.2}$$

where K_x and K_t are C_p^2 positive functions,

$$r_n(t,x) = \frac{1}{nh_t h_x^d}\sum_{i=1}^n K_t\left(\frac{t-T_i}{h_t}\right)K_x\left(\frac{x-X_i}{h_x}\right)v(T_i,X_i),$$
$$f_n(t,x) = \frac{1}{nh_t h_x^d}\sum_{i=1}^n K_t\left(\frac{t-T_i}{h_t}\right)K_x\left(\frac{x-X_i}{h_x}\right),$$

and $g : \mathbb{R} \longmapsto \mathbb{R}$ such that

$$g(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 & \text{if } y > 1, \\ -y^4 + 2y^2 & \text{if } y \in [0,1], \end{cases}$$

is a C_p^2 function in space and a C^1 function in time.

Remark 11.7. We give some useful upper bounds for g and its first derivative. The function $x \longmapsto \frac{g(x)}{x}$ is bounded by 2, g' is bounded by 2, $x \longmapsto \frac{g'(x)}{x}$ is bounded by 4 and $x \longmapsto \frac{g(x)}{x^2}$ is bounded by 2.

11.4 Property of our estimator

Proposition 11.8. *Assume v is bounded by a constant C_v. Then, the estimator $\mathcal{P}_n v$ defined by (11.2) satisfies*

$$\begin{aligned} |\mathcal{P}_n v(t,x)| &\leq C_v, \\ |\partial_x(\mathcal{P}_n v)(t,x)| &\leq CT\lambda(B)C_v \frac{|\partial_x K_x|_\infty |K_t|_\infty}{h_t h_x^{d+1}}, \\ |\partial_t(\mathcal{P}_n v)(t,x)| &\leq CT\lambda(B)C_v \frac{|K_x|_\infty |K_t'|_\infty}{h_t^2 h_x^d}, \\ |\partial_x^2(\mathcal{P}_n v)(t,x)| &\leq \frac{CT\lambda(B)C_v}{h_t h_x^{d+2}} |K_t|_\infty |\partial_x^2 K_x|_\infty + \frac{C(T\lambda(B))^3 C_v}{h_t^3 h_x^{3d+2}} |K_t|_\infty^3 |K_x|_\infty |\partial_x K_x|_\infty^2 \\ &\quad + \frac{C(T\lambda(B))^2 C_v}{h_t^2 h_x^{2d+2}} |K_t|_\infty^2 (|\partial_x K_x|_\infty^2 + |K_x|_\infty |\partial_x^2 K_x|_\infty). \end{aligned}$$

where C is a strictly positive constant.

Proof. Looking at (11.2), we deduce

$$|\mathcal{P}_n v(t,x)| \leq \frac{f_n(t,x)|v|_\infty}{f_n(t,x)} g(2T\lambda(B) f_n(t,x)).$$

Since v is bounded by C_v and g by 1, we get $\forall (t,x) \in [0,T] \times \mathbb{R}^d$, $|\mathcal{P}_n v(t,x)| \leq C_v$. We differentiate (11.2) to get

$$\begin{aligned} \partial_{x_i}(\mathcal{P}_n v)(t,x) =& \partial_{x_i} r_n(t,x) \frac{g(2T\lambda(B) f_n(t,x))}{f_n(t,x)} + 2T\lambda(B) r_n(t,x) \frac{g'(2T\lambda(B) f_n(t,x))}{f_n(t,x)} \partial_{x_i} f_n(t,x) \\ & - r_n(t,x) \frac{g(2T\lambda(B) f_n(t,x))}{f_n^2(t,x)} \partial_{x_i} f_n(t,x). \end{aligned}$$

As $\frac{g(x)}{x}$ is bounded by 2, $\left| \partial_{x_i} r_n(t,x) \frac{g(2T\lambda(B) f_n(t,x))}{f_n(t,x)} \right| \leq 4T\lambda(B)C_v \frac{|\partial_{x_i} K_x|_\infty |K_t|_\infty}{h_t h_x^{d+1}}$. As g' is bounded by 2, the last two terms are bounded by $CT\lambda(B)C_v \frac{|\partial_{x_i} K_x|_\infty |K_t|_\infty}{h_t h_x^{d+1}}$, where C depends on d, and the second result ensues.

The same kind of proof yields $|\partial_t(\mathcal{P}_n v)(t,x)| \leq CT\lambda(B)C_v \frac{|K_x|_\infty |K_t'|_\infty}{h_t^2 h_x^d}$. The computation of $\partial^2_{x_i x_j}(\mathcal{P}_n v)(t,x)$ leads to

$$\begin{aligned}
\partial^2_{x_i x_j}(\mathcal{P}_n v)(t,x) =& 2T\lambda(B)\partial^2_{x_i x_j} r_n(t,x) G(2T\lambda(B) f_n(t,x)) \\
&+ (2T\lambda(B))^2 \partial_{x_i} r_n(t,x) G'(2T\lambda(B) f_n(t,x)) \partial_{x_j} f_n(t,x) \\
&+ (2T\lambda(B))^2 \partial_{x_j} r_n(t,x) G'(2T\lambda(B) f_n(t,x)) \partial_{x_i} f_n(t,x) \\
&+ r_n(t,x)(2T\lambda(B))^2 \partial^2_{x_i x_j} f_n(t,x) G'(2T\lambda(B) f_n(t,x)) \\
&+ r_n(t,x)(2T\lambda(B))^3 \partial_{x_i} f_n(t,x) \partial_{x_j} f_n(t,x) G''(2T\lambda(B) f_n(t,x)),
\end{aligned}$$

where $G(y) = \frac{g(y)}{y}$. Analysing each term of the sum yields

$$\begin{aligned}
|\partial^2_x(\mathcal{P}_n v)(t,x)| \leq& \frac{CT\lambda(B)C_v}{h_t h_x^{d+2}} |K_t|_\infty |\partial^2_x K_x|_\infty + \frac{C(T\lambda(B))^3 C_v}{h_t^3 h_x^{3d+2}} |K_t|^3_\infty |K_x|_\infty |\partial_x K_x|^2_\infty \\
&+ \frac{C(T\lambda(B))^2 C_v}{h_t^2 h_x^{2d+2}} |K_t|^2_\infty (|\partial_x K_x|^2_\infty + |K_x|_\infty |\partial^2_x K_x|_\infty).
\end{aligned}$$

□

11.4.1 Application to our algorithm

In our algorithm, we build u_k in the following way

$$u_k(t,x) = \mathcal{P}^k_n(u_{k-1} + \overline{w}_k)(t,x),$$

where $\mathcal{P}^k_n$ is defined by (11.2) and

$$\begin{aligned}
&\overline{w}_k(t,x) = \frac{1}{M}\sum_{m=1}^{M} \Big[\Phi(X_T^{m,k,N}) - u_{k-1}(T, X_T^{m,k,N}) \\
&+ \int_t^T f\left(s, X_s^{m,k,N}, u_{k-1}(s, X_s^{m,k,N}), (\partial_x u_{k-1}\sigma)(s, X_s^{m,k,N})\right) + (\partial_t + \mathcal{L}^N) u_{k-1}(s, X_s^{m,k,N}) ds\Big].
\end{aligned}$$

Proposition 11.9. *Assume f, σ, b and Φ are bounded. Then, $\forall k \in \mathbb{N}$, there exists a constant C_k depending on $|f|_\infty, |\Phi|_\infty, |\sigma|_\infty, |b|_\infty, a, h_t, h_x, T$ and all the bounds for K_t, K_x and their derivatives up to order 1 for K_t and 2 for K_x such that*

$$\forall (t,x) \in [0,T] \times \mathbb{R}^d, \quad |u_k(t,x)| + |\partial_t u_k(t,x)| + |\partial_x u_k(t,x)| + |\partial^2_x u_k(t,x)| \leq C_k. \tag{11.3}$$

Proof of Proposition 11.9. We do it recursively. For $k = 0$, (11.3) is true since $u_0 \equiv 0$. Assume (11.3) is true at step $k-1$ and let C_{k-1} denote the constant which bounds u_{k-1} and its derivatives as in (11.3). Since $f, \Phi, b, \sigma, u_{k-1}, \partial_t u_{k-1}, \partial_x u_{k-1}, \partial^2_x u_{k-1}$ are bounded, we get that $(u_{k-1} + \overline{w}_k)(t,x)$ is bounded. Using Proposition 11.8 yields the result. □

Chapter 12

Convergence rate of our estimator

This chapter is devoted to study the rate convergence of $\mathcal{P}_n v - v$, where $\mathcal{P}_n v$ is the estimator of a function v, which has been defined in (9.15), page 105. We recall

$$\mathcal{P}_n v(s,y) = \frac{r_n(s,y)}{f_n(s,y)} g(2T\lambda(B) f_n(s,y)), \text{ where}$$

- $r_n(s,y) = \frac{1}{nh_t h_x^d} \sum_{i=1}^n K_t(\frac{s-T_i}{h_t}) K_x(\frac{y-X_i}{h_x}) v(T_i, X_i)$,
- $f_n(s,y) = \frac{1}{nh_t h_x^d} \sum_{i=1}^n K_t(\frac{s-T_i}{h_t}) K_x(\frac{y-X_i}{h_x})$.

We aim at studying $\mathcal{P}_n v - v$ and $\partial_x(\mathcal{P}_n v) - \partial_x v$, in norm $\|\cdot\|_{H^{\mu}_{\psi,X}}$ (see the definition of the norm below), where ψ denote a function from $[0,T]$ into $\mathbb{R}^+$ such that $\psi(s) = \frac{e^{\gamma s}}{s^{\alpha}}$, where $\gamma > 0$ and $\alpha \in [0,1[$. X is the diffusion process defined in (9.2), page 93 and its transition density is denoted p. The scheme of the chapter is the following: in Section 12.1, we recall some notations and give new definitions. In Section 12.2, we state some preliminary results on $f_n, \partial_x f_n, r_n, \partial_x r_n$ in $\mathbb{L}^2$ norm. Section 12.3 states the convergence for $\mathcal{P}_n v - v$ and Section 12.4 states the convergence for $\partial_x(\mathcal{P}_n v) - \partial_x v$.

12.1 Notations and Assumptions

First, let us recall some Definitions given in Part II.

Definition 12.1 (Definition of K, α_1, α_2 and ellipticity condition)**.** We recall the definition of some constants used in this Section and introduced in Part II.

1. σ_0 and σ_1 denote the constants appearing in Definition 6.2, page 63: σ is **uniformly elliptic** on $[0,T] \times \mathbb{R}^d$ if there exist two positive constants σ_0, σ_1 s.t., for any vector ξ and any $(t,x) \in [0,T] \times \mathbb{R}^d$

$$\sigma_0 |\xi|^2 \leq \sum_{i,j=1}^d [\sigma\sigma^*]_{i,j}(t,x)\xi_i\xi_j \leq \sigma_1 |\xi|^2.$$

2. K, α_1, and α_2 denote the constants introduced in Proposition 6.4, page 63:
Assume that the coefficients a, b are bounded measurable functions of $(t,x) \in [0,T] \times$

$\mathbb{R}^d$ and that σ is elliptic. There exist positive constants K, α_1, α_2 s.t.

$$\frac{K^{-1}}{(2\pi\alpha_1(s-t))^{\frac{d}{2}}} e^{-\frac{|x-y|^2}{2\alpha_1(s-t)}} \leq p(t,x;s,y) \leq K \frac{1}{(2\pi\alpha_2(s-t))^{\frac{d}{2}}} e^{-\frac{|x-y|^2}{2\alpha_2(s-t)}}.$$

The constant K depends only on σ_0, σ_1, d, T and the suprema of the coefficients a, b, where a denotes $\sigma\sigma^*$. The constants α_0, α_1 depend on σ_0, σ_1 and d.

Let us recall the definition of $\nu^t(s,y)$ introduced in Definition 6.11, page 67

Definition 12.2. For any $s, t \in [0,T]$ and any $x, y \in \mathbb{R}^d$ such that $t < s$ we define

$$\nu^t(s,y) := \int_{\mathbb{R}^d} e^{-\mu|x|} p(t,x;s,y)dx,$$

where p is the transition density function of the process X defined by (9.2).

We recall the definition of the norm $\|\cdot\|_{H^\mu_{\beta,X}}$, introduced in Part II :

Definition 12.3. For any $\beta > 0, \mu > 0$ and any diffusion process $(X_s)_{0 \leq s \leq T}$ starting from x at time 0, let $H^\mu_{\beta,X}$ define the space of deterministic functions $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\|v\|^2_{H^\mu_{\beta,X}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \mathbb{E}|v(s, X^x_s)|^2 dxds < \infty.$$

Using the above definition of ν, we also get $\|v\|^2_{H^\mu_{\beta,X}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y)|v(s,y)|^2$.

We also recall the definition of $H^{m,\mu}_\beta$ introduced at the beginning of in Part II.

Definition 12.4. For any $\beta > 0$, $m \leq 2$, and $\mu > 0$, let $H^{m,\mu}_\beta$ define the space of deterministic functions $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\|v\|^2_{H^{m,\mu}_\beta} = \int_0^T e^{\beta s} \|v(s,\cdot)\|^2_{H^{m,\mu}} ds = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \sum_{k \leq m} |\partial^k_x v(s,x)|^2 dxds < \infty.$$

For $m = 0$, we set $H^\mu_\beta = H^{0,\mu}_\beta$.

In the whole section we consider the following assumption

Hypothesis 12.1

1. *The set of points* $\{(T_i, X_i), 1 \leq i \leq n\}$ *are uniformly distributed on* $[0,T] \times [-a,a]^d$, $B := B_\infty(0,a) = [-a,a]^d$ *and* $\lambda(B) := (2a)^d$.

2. *Kernel function* K_t *is defined on the compact support* $[-1,1]$, *is bounded by* $|K_t|_\infty$, *is even, positive,* C^2_p *and* $\int_{\mathbb{R}} K_t(u)du = 1$.

3. *Kernel function* K_x *is defined on the compact support* $[-1,1]^d$, *is bounded by* $|K_x|_\infty$, *and is such that*

$$\forall y = (y_1, \cdots, y_d) \in \mathbb{R}^d, \quad K_x(y) = \Pi_{j=1}^d K^j_x(y_j),$$

where for $j = 1, \cdots, d$ $K^j_x : \mathbb{R} \to \mathbb{R}$ *is an even positive* C^2_p *function and* $\int_{\mathbb{R}} K^j_x(u)du = 1$.

4. *δ_n denotes $\frac{1}{nh_t h_x^d}$, and $T\lambda(B)\delta_n << 1$.*

5. *$h_x << a$ and $h_t << \frac{T}{2}$. Since we study the convergence when h_t and h_x tend to 0, we assume in the following that $h_t \leq 1$ and $h_x \leq 1$.*

Definition 12.5. Let $\overline{f}$ denote the uniform density on $[0,T] \times [-a,a]^d$. $\overline{f}(s,y) = \frac{1}{T\lambda(B)}\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in B}$.

12.2 Preliminary Results

First, we recall a well known result really used in the sequel

Lemma 12.6 (Bias-Variance decomposition)**.** *Let X be an $\mathbb{R}^d$ square integrable random variable*

$$\mathbb{E}|X|^2 = |\mathbb{E}X|^2 + \sum_{i=1}^{d} VarX_i.$$

Lemma 12.7. *Let $\phi : \mathbb{R} \to \mathbb{R}^+$ be a bounded function with compact support $[-1,1]^d$, and $h > 0$. Then, $\int_{\mathbb{R}^d} \phi\left(\frac{x}{h}\right) dx \leq (2h)^d |\phi|_\infty$.*
Moreover, if ϕ satisfies $\int_{\mathbb{R}^d} \phi(u)du = 1$, $\int_{\mathbb{R}^d} \phi\left(\frac{x}{h}\right) dx = h^d$.

Lemma 12.8. *For all $(s,y) \in [0,T] \times \mathbb{R}^d$, we define*

$$a(s,y) = \frac{C}{n}\sum_{i=1}^{n} g_t\left(\frac{s-T_i}{h_t}\right) g_x\left(\frac{y-X_i}{h_x}\right),$$

where g_t is a positive bounded function with compact support in $[-1,1]$, g_x is a positive bounded function with compact support in $[-1,1]^d$, and C is a positive constant.

$$0 \leq \mathbb{E}(a(s,y)) \leq C2^{d+1}|g_t|_\infty |g_x|_\infty \frac{h_t h_x^d}{T\lambda(B)}\mathbf{1}_{y\in B_\infty(0,a+h_x)}.$$

If $\int_{\mathbb{R}} g_t(u)du = 1$ and $\int_{\mathbb{R}^d} g_x(u)du = 1$, $\mathbb{E}(a(s,y)) \leq C\frac{h_t h_x^d}{T\lambda(B)}\mathbf{1}_{y\in B_\infty(0,a+h_x)}$.

$$Var(a(s,y)) \leq C^2 2^{d+1}|g_t|_\infty^2 |g_x|_\infty^2 \frac{h_t h_x^d}{nT\lambda(B)}\mathbf{1}_{y\in B_\infty(0,a+h_x)}.$$

Using $T\lambda(B)\delta_n << 1$ yields

$$\mathbb{E}(a^2(s,y)) \leq C^2 2^{2d+3}|g_t|_\infty^2 |g_x|_\infty^2 \frac{h_t^2 h_x^{2d}}{(T\lambda(B))^2}\mathbf{1}_{y\in B_\infty(0,a+h_x)}.$$

If $\int_{\mathbb{R}} g_t(u)du = 1$ and $\int_{\mathbb{R}^d} g_x(u)du = 1$, $(\mathbb{E}(a(s,y)))^2 \leq C^2\frac{h_t^2 h_x^{2d}}{(T\lambda(B))^2}\mathbf{1}_{y\in B_\infty(0,a+h_x)}$ and $\mathbb{E}(a^2(s,y)) \leq C^2 2^{d+2}|g_t|_\infty^2 |g_x|_\infty^2 \frac{h_t^2 h_x^{2d}}{(T\lambda(B))^2}\mathbf{1}_{y\in B_\infty(0,a+h_x)}$.

Proof of Lemma 12.8. We do the proof for $C = 1$.

$$\mathbb{E}(a(s,y)) = \mathbb{E}\left(g_t\left(\frac{s-T_1}{h_t}\right) g_x\left(\frac{y-X_1}{h_x}\right)\right) = \frac{1}{T\lambda(B)} \int_0^T dr g_t\left(\frac{s-r}{h_t}\right) \int_B dz g_x\left(\frac{y-z}{h_x}\right),$$

and Lemma 12.7 ends the first part of the proof.

$$\mathrm{Var}(a(s,y)) = \frac{1}{n}\mathrm{Var}\left(g_t\left(\frac{s-T_1}{h_t}\right) g_x\left(\frac{y-X_1}{h_x}\right)\right) \leq \frac{1}{n}\mathbb{E}\left(g_t^2\left(\frac{s-T_1}{h_t}\right) g_x^2\left(\frac{y-X_1}{h_x}\right)\right).$$

Then, $\mathrm{Var}(a(s,y)) \leq \frac{1}{nT\lambda(B)} \int_0^T dr g_t^2\left(\frac{s-r}{h_t}\right) \int_B dz g_x^2\left(\frac{y-z}{h_x}\right)$. Lemmas 12.7 and 12.6 end the proof. □

Lemma 12.9. *For all $(s,y) \in [0,T] \times \mathbb{R}^d$, we define*

$$a_\phi(s,y) = \frac{C}{n}\sum_{i=1}^n g_t\left(\frac{s-T_i}{h_t}\right) g_x\left(\frac{y-X_i}{h_x}\right)\phi(T_i, X_i),$$

where g_t is a bounded function with compact support in $[-1,1]$, g_x is a bounded function with compact support in $[-1,1]^d$ and ϕ is a continuous function from $[0,T]\times\mathbb{R}^d$ in $\mathbb{R}$.

$$\mathbb{E}(a_\phi(s,y))^2 \leq C^2\frac{2^{d+1}h_t h_x^d}{(T\lambda(B))^2} \int_0^T dr g_t^2\left(\frac{s-r}{h_t}\right) \int_B dz g_x^2\left(\frac{y-z}{h_x}\right)\phi^2(r,z),$$
$$Var(a_\phi(s,y)) \leq \frac{C^2}{nT\lambda(B)} \int_0^T dr g_t^2\left(\frac{s-r}{h_t}\right) \int_B dz g_x^2\left(\frac{y-z}{h_x}\right)\phi^2(r,z).$$

Using $T\lambda(B)\delta_n << 1$ yields

$$\mathbb{E}(a_\phi^2(s,y)) \leq \frac{2^{d+2}C^2 h_t h_x^d}{(T\lambda(B))^2} \int_0^T dr g_t^2\left(\frac{s-r}{h_t}\right) \int_B dz g_x^2\left(\frac{y-z}{h_x}\right)\phi^2(r,z).$$

Proof of Lemma 12.9. We do the proof for $C = 1$.

$$\begin{aligned}\mathbb{E}(a_\phi(s,y)) &= \mathbb{E}\left(g_t\left(\frac{s-T_1}{h_t}\right) g_x\left(\frac{y-X_1}{h_x}\right)\phi(T_1,X_1)\right)\\ &= \frac{1}{T\lambda(B)} \int_0^T dr g_t\left(\frac{s-r}{h_t}\right) \int_B dz g_x\left(\frac{y-z}{h_x}\right)\phi(r,z).\end{aligned}$$

Since g_t (resp. g_x) has its support in $[-1,1]$ (resp. $[-1,1]^d$), the first result follows.

$$\begin{aligned}\mathrm{Var}(a_\phi(s,y)) &= \frac{1}{n}\mathrm{Var}\bigg(g_t\left(\frac{s-T_1}{h_t}\right) g_x\left(\frac{y-X_1}{h_x}\right)\phi(T_1,X_1)\bigg).\\ &\leq \frac{1}{n}\mathbb{E}\left[g_t^2\left(\frac{s-T_1}{h_t}\right) g_x^2\left(\frac{y-X_1}{h_x}\right)\phi^2(T_1,X_1)\right].\end{aligned}$$

The third result comes from Lemma 12.6. □

Lemma 12.10. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. Let f be a function* *from $[0,T]\times\mathbb{R}^d$ into $\mathbb{R}^+$, g_t a positive bounded function with compact support in $[-1,1]$ and g_x a positive bounded function with compact support in $[-1,1]^d$. Then,*

$$\int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)\int_0^T drg_t\left(\frac{s-r}{h_t}\right)\int_{\mathbb{R}^d} dzg_x\left(\frac{y-z}{h_x}\right)f(r,z)\leq$$
$$K(T)h_th_x^d\int_0^T dre^{\beta r}\int_{\mathbb{R}^d} dz\nu^0(r,z)f(r,z),$$

where $K(T)$ depends on $d,\mu,|g_t|_\infty,|g_x|_\infty,K,\alpha_1$ and α_2.

Proof of Lemma 12.10. Since g_t and g_x have their support in $[-1,1]$, $|s-r|\leq h_t$ and $|y-z|\leq h_x$, $e^{\beta s}\leq e^{\beta h_t}e^{\beta r}$, we can apply Corollary 6.13, page 68, to get $\nu^0(s,y)\leq Ce^{d\mu h_x}e^{c_2h_t}\nu^0(r,z)$. We obtain

$$\int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)\int_0^T drg_t\left(\frac{s-r}{h_t}\right)\int_{\mathbb{R}^d} dzg_x\left(\frac{y-z}{h_x}\right)f(r,z)\leq$$
$$Ce^{d\mu h_x}e^{(\beta+c_2)h_t}\int_0^T dre^{\beta r}\int_{\mathbb{R}^d} dz\nu^0(r,z)f(r,z)\int_0^T dsg_t\left(\frac{s-r}{h_t}\right)\int_{\mathbb{R}^d} dyg_x\left(\frac{y-z}{h_x}\right).$$

Using Lemma 12.7, page 127 yields $\int_{\mathbb{R}^d} dyg_x\left(\frac{y-z}{h_x}\right)\leq(2h_x)^d|g_x|_\infty$. Since $|s-r|\leq h_t$, $\int_0^T dsg_t\left(\frac{s-r}{h_t}\right)\leq 2|g_t|_\infty h_t$. Then, we get

$$\int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)\int_0^T drg_t\left(\frac{s-r}{h_t}\right)\int_{\mathbb{R}^d} dzg_x\left(\frac{y-z}{h_x}\right)f(r,z)$$
$$\leq C|g_t|_\infty|g_x|_\infty h_th_x^de^{d\mu h_x}e^{(\beta+c_2)h_t}\int_0^T dre^{\beta r}\int_{\mathbb{R}^d} dz\nu^0(r,z)f(r,z),$$

and the result follows. □

Lemma 12.11. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. Let g_0 be a function from $[0,T]\times\mathbb{R}^d$ into $\mathbb{R}$ and $g:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ defined such that*

$$g(s,y):=\mathbb{E}\left(\int_s^T g_0(r,X_r^{s,y})dr\right),$$

where X is the diffusion process satisfying (9.2), page 93 and p is its transition density function. Let $\overline{\psi}:s\longmapsto e^{\beta s}\xi(s)$, where $\beta>0$, and ξ is a bounded function from $[0,T]$ into $\mathbb{R}^+$. Then,

$$\int_0^T ds\overline{\psi}(s)\int_{\mathbb{R}^d} dy\nu^0(s,y)\mathbb{E}\left(\int_s^T g_0^2(r,X_r^{s,y})dr\right)\leq\left(\int_0^T\xi(s)ds\right)\|g_0\|^2_{H^\mu_{\beta,X}},$$

$$\text{and } \|g\|^2_{H^\mu_{\overline{\psi},X}}\leq\left(\int_0^T(T-s)\xi(s)ds\right)\|g_0\|^2_{H^\mu_{\beta,X}}.$$

Proof. We assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition to provide a transition density function to the process X. We easily deduce the second result from the first one, since

$$\begin{aligned}\|g\|^2_{H^{\mu}_{\overline{\psi},X}} &= \int_0^T dse^{\beta s}\xi(s)\int_{\mathbb{R}^d} dy\nu^0(s,y)g^2(s,y),\\ &\leq \int_0^T ds(T-s)e^{\beta s}\xi(s)\int_{\mathbb{R}^d} dy\nu^0(s,y)\mathbb{E}\left(\int_s^T g_0^2(r,X_r^{s,y})dr\right).\end{aligned}$$

Let us show the first result. The l.h.s. of the first inequality to be proved is equal to $\int_0^T dse^{\beta s}\xi(s)\int_{\mathbb{R}^d} dy\nu^0(s,y)\int_{\mathbb{R}^d} dz\int_s^T drg_0^2(r,z)p(s,y;r,z)$. By using the definition of $\nu^0(s,y)$ and the Chapman-Kolmogorov equation, which states

$$\forall t<s<T,\ \ \forall x,\xi\in\mathbb{R}^d,\ \ p(t,x;T,\xi)=\int_{\mathbb{R}^d} p(t,x;s,\eta)p(s,\eta;T,\xi)d\eta,$$

we get $\int_{\mathbb{R}^d} dy\nu^0(s,y)p(s,y;r,z)=\nu^0(r,z)$. Since $s\leq r$, we obtain

$$\begin{aligned}&\int_0^T dse^{\beta s}\xi(s)\int_{\mathbb{R}^d} dy\nu^0(s,y)\int_{\mathbb{R}^d} dz\int_s^T drg_0^2(r,z)p(s,y;r,z)\\ &\qquad\leq \int_0^T \xi(s)ds\int_0^T dre^{\beta r}\int_{\mathbb{R}^d} dz\nu^0(r,z)g_0^2(r,z),\end{aligned}$$

and the proof is over. □

12.2.1 Results on f_n

Let $f_n(s,y)$ be defined in the following way

$$f_n(s,y)=\frac{1}{nh_th_x^d}\sum_{i=1}^n K_t\left(\frac{s-T_i}{h_t}\right)K_x\left(\frac{y-X_i}{h_x}\right).$$

We plot in Figure 12.1 the graph of f_n on $[0.5,1.5]\times[-1.5,1.5]$. We choose $n=500$, $h_t=\frac{1}{n^{1/3}}$ and $h_x=\frac{2}{n^{1/3}}$. These values of h_t and h_x are consistent with Theorem 11.3, page 119. The random points $(T_i,X_i)_{i=1,\cdots,n}$ are chosen on $[0,1]\times[-1,1]$ with a uniform law. The kernel functions K_x and K_t are triweight functions (see Table 11.1, page 118).

Remark 12.12. The function $f_n(s,y)$ approximates $\frac{1}{T\lambda(B)}\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in[-a,a]^d}$. Looking at Figure 12.1, we see that the approximation is quite bad close to the boundary of $[0,1]\times[-1,1]$. More generally, one can say that f_n approximates quite well $\overline{f}(s,y)$ in $[h_t,T-h_t]\times[-a+h_x,a-h_x]^d$. However, the approximation on $[0,T]\times[-a,a]^d\setminus[h_t,T-h_t]\times[-a+h_x,a-h_x]^d$ is quite poor.

Lemma 12.13. *For all* $(s,y)\in[0,T]\times\mathbb{R}^d$,

$$\mathbb{E}[f_n(s,y)]=\frac{1}{T\lambda(B)}\int_{-1\vee\frac{-s}{h_t}}^{1\wedge\frac{T-s}{h_t}} K_t(r)dr\Pi_{j=1}^d\int_{-1\vee\frac{-a-y_j}{h_x}}^{1\wedge\frac{a-y_j}{h_x}} K_x^j(x_j)dx_j, \tag{12.1}$$

and $\mathbb{E}[f_n(s,y)]\leq\frac{1}{T\lambda(B)}$. *Moreover, for* $(s,y)\in[h_t,T-h_t]\times B_\infty(0,a-h_x)$, $\mathbb{E}[f_n(s,y)]=\frac{1}{T\lambda(B)}=\overline{f}(s,y)$.

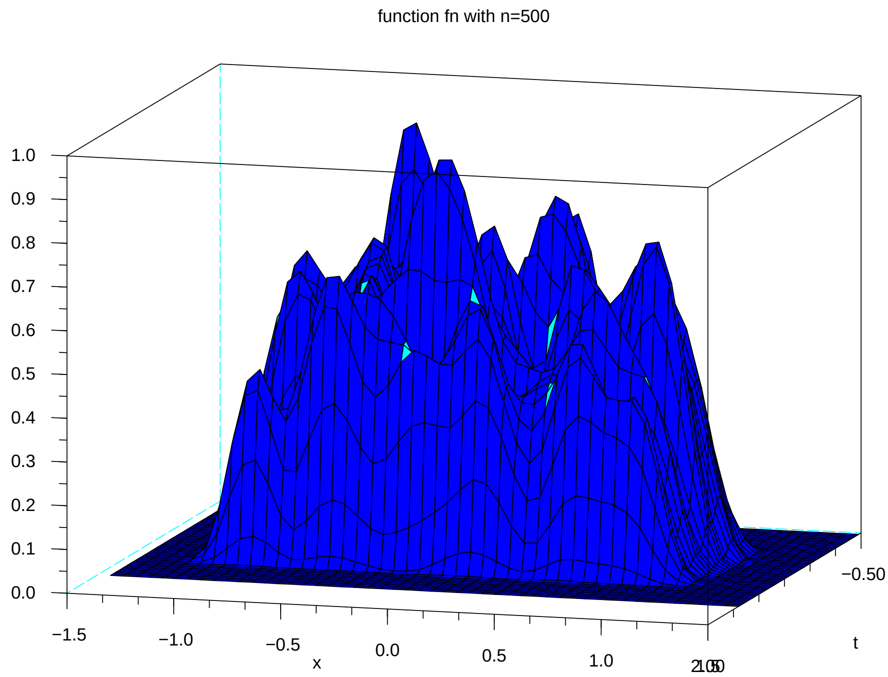

Figure 12.1: graph of f_n on $[0,1]\times[-1,1]$ with $n=500$

Proof of Lemma 12.13. A simple change of variables leads to (12.1). Let $(s,y)\in[h_t,T-h_t]\times B_\infty(0,a-h_x)$. Then,

$$\mathbb{E}[f_n(s,y)] = \frac{1}{T\lambda(B)}\int_{-1}^{1}K_t(r)dr\Pi_{j=1}^{d}\int_{-1}^{1}K_x^j(x_j)dx_j,$$

and the result follows. □

Lemma 12.14. *For all* $(s,y)\in[0,T]\times\mathbb{R}^d$, $\mathbb{E}[f_n(s,y)]^2\le 2^{d+2}\frac{|K_t|_\infty^2|K_x|_\infty^2}{(T\lambda(B))^2}\mathbf{1}_{y\in B_\infty(0,a+h_x)}$.

Proof. We refer to Lemma 12.8, page 127 with $C=\frac{1}{h_th_x^d}$, $g_t=K_t$ and $g_x=K_x$.. □

Proposition 12.15. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition.*

$$\int_0^T dse^{\beta s}\int_B dy\nu^0(s,y)|\mathbb{E}(f_n(s,y))-\overline{f}(s,y)|^2\le\frac{K(T)}{(T\lambda(B))^2}[h_t+e^{-\mu a}a^{d-1}h_x],$$

where $K(T)$ is a function non decreasing in T and depending on d,β,μ,K and α_2.

Proof. From Lemma 12.13, we deduce $\mathbb{E}(f_n(s,y))-\overline{f}(s,y)=0$ for $(s,y)\in[h_t,T-h_t]\times B_\infty(0,a-h_x)$, and $|\mathbb{E}(f_n(s,y))-\overline{f}(s,y)|\le\frac{1}{T\lambda(B)}$ on $[0,h_t]\times B$, $[T-h_t,T]\times B$ and

$[h_t, T-h_t] \times B \setminus B_\infty(0, a-h_x)$. Hence,

$$\begin{aligned}\int_0^T dse^{\beta s}\int_B dy\nu^0(s,y)|\mathbb{E}(f_n(s,y))-\overline{f}(s,y)|^2 \leq \frac{1}{(T\lambda(B))^2}\int_0^{h_t} dse^{\beta s}\int_B dy\nu^0(s,y)\\ + \frac{1}{(T\lambda(B))^2}\int_{T-h_t}^T dse^{\beta s}\int_B dy\nu^0(s,y)\\ + \frac{1}{(T\lambda(B))^2}\int_{h_t}^{T-h_t} dse^{\beta s}\int_{B\setminus B_\infty(0,a-h_x)} dy\nu^0(s,y).\end{aligned} \tag{12.2}$$

Since $\nu^0(s,y) \leq 2^d K e^{c_2 s} e^{-\mu|y|}$ (see Proposition 6.12, page 67), we get

$$\int_0^{h_t} dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y) \leq 2^d K\int_0^{h_t} dse^{(\beta+c_2)s}\int_{\mathbb{R}^d} dye^{-\mu|y|} \leq 2^{2d}d^{\frac{d}{2}}\frac{K}{\mu^d}e^{(\beta+c_2)h_t}h_t, \tag{12.3}$$

$$\int_{T-h_t}^T dse^{\beta s}\int_B dy\nu^0(s,y) \leq \int_{T-h_t}^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y) \leq \frac{2^{2d}}{d^{\frac{d}{2}}}\frac{K}{\mu^d}e^{(\beta+c_2)T}h_t. \tag{12.4}$$

It remains to bound $\int_{h_t}^{T-h_t} dse^{\beta s}\int_{B\setminus B_\infty(0,a-h_x)} dy\nu^0(s,y)$. Since

$$\int_{B\setminus B_\infty(0,a-h_x)} dy\nu^0(s,y) \leq 2^d K e^{c_2 s}e^{-\mu(a-h_x)}\int_{y\in B\setminus B_\infty(0,a-h_x)} dy \leq 2^{2d}Ke^{c_2 s}e^{-\mu(a-h_x)}a^{d-1}h_x, \tag{12.5}$$

we obtain $\int_{h_t}^{T-h_t} dse^{\beta s}\int_{B\setminus B_\infty(0,a-h_x)} dy\nu^0(s,y) \leq 2^{2d}Ke^{(\beta+c_2)T}Te^{-\mu(a-h_x)}a^{d-1}h_x$. Plugging the previous result, (12.3) and (12.4) in (12.2) ends the proof. □

Proposition 12.16. *There exist three constants c, C and C' depending on d, $|K_t|_\infty$ and $|K_x|_\infty$ such that for all $\epsilon^2 \leq (T\lambda(B))^{-2}$ and $(s,y) \in [0,T]\times\mathbb{R}^d$, the following assertions hold*

$$\begin{aligned}\mathbb{P}(|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon) &\leq 2\exp\left(-\frac{c\epsilon^2 T\lambda(B)}{\delta_n}\right),\\ \mathbb{E}\left(|f_n(s,y)-\mathbb{E}(f_n(s,y))|^2\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon}\right) &\leq C'\left(\epsilon^2+\frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon^2 T\lambda(B)}{\delta_n}\right).\\ \mathbb{E}\left(|f_n(s,y)-\mathbb{E}(f_n(s,y))|^4\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon}\right) &\leq C\left(\epsilon^2+\frac{\delta_n}{T\lambda(B)}\right)^2\exp\left(-\frac{c\epsilon^2 T\lambda(B)}{\delta_n}\right)\\ &\quad + C\frac{\delta_n}{(T\lambda(B))^3}\exp\left(-\frac{c}{T\lambda(B)\delta_n}\right).\end{aligned}$$

Proof of Proposition 12.16. We begin to prove the first assertion. To do so, we apply Berstein' s inequality (see Lemma A.11, page 264) to $\mathbb{P}(|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon)$. According to this Lemma, the random variable X_i corresponds to $\frac{1}{h_t h_x^d}K_t\left(\frac{s-T_i}{h_t}\right)K_x\left(\frac{y-X_i}{h_x}\right)$, $a = 0$, $b = \frac{1}{h_t h_x^d}|K_t|_\infty|K_x|_\infty$ and $\sigma^2 = \frac{1}{(h_t h_x^d)^2}\mathrm{Var}\left(K_t\left(\frac{s-T_i}{h_t}\right)K_x\left(\frac{y-X_i}{h_x}\right)\right) \leq 2^{d+1}\frac{|K_t|_\infty^2|K_x|_\infty^2}{T\lambda(B)h_t h_x^d}$ (see Lemma 12.8, page 127 with $g_t = K_t$ and $g_x = K_x$). Then, for all

$\epsilon \geq 0$

$$\mathbb{P}(|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \epsilon) \leq 2\exp\left(-\frac{n\epsilon^2}{2^{d+2}\frac{|K_t|_\infty^2|K_x|_\infty^2}{T\lambda(B)h_th_x^d} + \frac{2\epsilon|K_t|_\infty|K_x|_\infty}{3h_th_x^d}}\right),$$

$$\leq 2\exp\left(-\frac{c_0\epsilon^2}{\delta_n(T\lambda(B)^{-1}+\epsilon)}\right),$$

where $c_0 = \left(2^{d+2}|K_t|_\infty^2|K_x|_\infty^2 \vee \frac{2|K_t|_\infty|K_x|_\infty}{3}\right)^{-1}$.
Then, let us prove the second assertion. Using (A.9) leads to

$$\begin{aligned} E_1 :&= \mathbb{E}\left(|f_n(s,y) - \mathbb{E}(f_n(s,y))|^2\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon}\right) \\ &= \epsilon^2\mathbb{P}(|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \epsilon) + \int_{\epsilon^2}^\infty \mathbb{P}(|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \sqrt{x})dx. \end{aligned}$$

We use the first assertion to bound $\mathbb{P}(|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \epsilon)$, $\mathbb{P}(|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \sqrt{x})$ and the fact that $\epsilon^2 \leq (T\lambda(B))^{-2}$ yields

$$E_1 \leq 2\epsilon^2\exp\left(-\frac{c_0T\lambda(B)\epsilon^2}{2\delta_n}\right) + 2\int_{\epsilon^2}^\infty \exp\left(-\frac{c_0x}{\delta_n(T\lambda(B)^{-1}+\sqrt{x})}\right)dx. \tag{12.6}$$

Let us compute $I := \int_{\epsilon^2}^\infty \exp\left(-\frac{c_0x}{\delta_n(T\lambda(B)^{-1}+\sqrt{x})}\right)dx$.

$$\begin{aligned} I &= \\ &\int_{\epsilon^2}^{(T\lambda(B))^{-2}} \exp\left(-\frac{c_0x}{\delta_n(T\lambda(B)^{-1}+\sqrt{x})}\right)dx + \int_{(T\lambda(B))^{-2}}^\infty \exp\left(-\frac{c_0x}{\delta_n(T\lambda(B)^{-1}+\sqrt{x})}\right)dx, \\ :&= I_1 + I_2. \end{aligned}$$

We easily get

$$I_1 \leq \int_{\epsilon^2}^{(T\lambda(B))^{-2}} \exp\left(-\frac{c_0xT\lambda(B)}{2\delta_n}\right)dx \leq \frac{2\delta_n}{c_0T\lambda(B)}\exp\left(-\frac{c_0\epsilon^2T\lambda(B)}{2\delta_n}\right).$$

Concerning I_2, we obtain

$$I_2 \leq \int_{(T\lambda(B))^{-2}}^\infty \exp\left(-\frac{c_0\sqrt{x}}{2\delta_n}\right)dx = \int_{(T\lambda(B))^{-1}}^\infty 2u\exp\left(-\frac{c_0u}{2\delta_n}\right)du,$$

and $\int_{(T\lambda(B))^{-1}}^\infty 2u\exp\left(-\frac{c_0u}{2\delta_n}\right)du = \left(\frac{4\delta_n}{c_0T\lambda(B)} + \frac{8\delta_n^2}{c_0^2}\right)\exp\left(-\frac{c_0}{2T\lambda(B)\delta_n}\right)$. As $T\lambda(B)\delta_n \leq 1$ and $\epsilon^2 \leq (T\lambda(B))^{-2}$, we obtain

$$I_1 + I_2 \leq C_1\frac{\delta_n}{T\lambda(B)}\exp\left(-\frac{c_0\epsilon^2T\lambda(B)}{2\delta_n}\right),$$

where $C_1 = \frac{8}{c_0}(1+\frac{1}{c_0})$. Combining this result and (12.6) yields the result.
We prove the third assertion. Using (A.9), page 264 leads to

$$\begin{aligned} E_2 :&= \mathbb{E}\left(|f_n(s,y) - \mathbb{E}(f_n(s,y))|^4\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon}\right) \\ &= \epsilon^4\mathbb{P}(|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \epsilon) + \int_{\epsilon^4}^\infty \mathbb{P}(|f_n(s,y) - \mathbb{E}(f_n(s,y))| > x^{1/4})dx. \end{aligned}$$

Then, we apply the first assertion to bound $\mathbb{P}(|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon)$ and $\mathbb{P}(|f_n(s,y)-\mathbb{E}(f_n(s,y))|>x^{1/4})$, and we get

$$E_2 \leq 2\epsilon^4 \exp\left(-\frac{c_0 T\lambda(B)\epsilon^2}{2\delta_n}\right) + 2\int_{\epsilon^4}^{\infty} \exp\left(-\frac{c_0\sqrt{x}}{\delta_n(T\lambda(B)^{-1}+x^{1/4})}\right) dx. \tag{12.7}$$

Let us compute $I := \int_{\epsilon^4}^{\infty} \exp\left(-\frac{c_0\sqrt{x}}{\delta_n(T\lambda(B)^{-1}+x^{1/4})}\right) dx$.

$$\begin{aligned}
J =& \\
&\int_{\epsilon^4}^{(T\lambda(B))^{-4}} \exp\left(-\frac{c_0\sqrt{x}}{\delta_n(T\lambda(B)^{-1}+x^{1/4})}\right) dx + \int_{(T\lambda(B))^{-4}}^{\infty} \exp\left(-\frac{c_0\sqrt{x}}{\delta_n(T\lambda(B)^{-1}+x^{1/4})}\right) dx,\\
:=& J_1 + J_2.
\end{aligned}$$

We get

$$J_1 \leq \int_{\epsilon^4}^{(T\lambda(B))^{-4}} \exp\left(-\frac{c_0\sqrt{x}T\lambda(B)}{2\delta_n}\right) dx = \int_{\epsilon^2}^{(T\lambda(B))^{-2}} 2u\exp\left(-\frac{c_0T\lambda(B)u}{2\delta_n}\right) du.$$

By integration by parts we get

$$\begin{aligned}
&\int_{\epsilon^2}^{(T\lambda(B))^{-2}} 2u\exp\left(-\frac{c_0T\lambda(B)u}{2\delta_n}\right) du\\
&= \left[\frac{-4u\delta_n}{c_0T\lambda(B)}\exp\left(-\frac{c_0T\lambda(B)u}{2\delta_n}\right)\right]_{\epsilon^2}^{(T\lambda(B))^{-2}} + \frac{4\delta_n}{c_0T\lambda(B)}\int_{\epsilon^2}^{(T\lambda(B))^{-2}} \exp\left(-\frac{c_0T\lambda(B)u}{2\delta_n}\right) du\\
&= -\exp\left(-\frac{c_0}{2\delta_nT\lambda(B)}\right)\left[\frac{4\delta_n}{c_0(T\lambda(B))^3} + \frac{8\delta_n^2}{c_0^2(T\lambda(B))^2}\right]\\
&\qquad + \exp\left(-\frac{c_0\epsilon^2T\lambda(B)}{2\delta_n}\right)\left[\epsilon^2\frac{4\delta_n}{c_0T\lambda(B)} + \frac{8\delta_n^2}{c_0^2(T\lambda(B))^2}\right].
\end{aligned}$$

Thus, $J_1 \leq C_1\frac{\delta_n}{T\lambda(B)}\exp\left(-\frac{c_0\epsilon^2T\lambda(B)}{2\delta_n}\right)(\epsilon^2+\frac{\delta_n}{T\lambda(B)})$, where C_1 has been defined above. Concerning J_2, we obtain

$$J_2 \leq \int_{(T\lambda(B))^{-4}}^{\infty} \exp\left(-\frac{c_0x^{1/4}}{2\delta_n}\right) dx = \int_{(T\lambda(B))^{-1}}^{\infty} 4u^3\exp\left(-\frac{c_0u}{2\delta_n}\right) du.$$

An integration by parts formula leads to

$$\begin{aligned}
\int_{(T\lambda(B))^{-1}}^{\infty} 4u^3\exp\left(-\frac{c_0u}{2\delta_n}\right) dx =& -\left[\exp\left(-\frac{c_0u}{2\delta_n}\right)\left(\frac{8\delta_nu^3}{c_0} + \frac{48\delta_n^2u^2}{c_0^2} + \frac{192\delta_n^3u}{c_0^3}\right)\right]_{(T\lambda(B))^{-1}}^{\infty}\\
&+ \frac{192\delta_n^3}{c_0^3}\int_{(T\lambda(B))^{-1}}^{\infty} \exp\left(-\frac{c_0u}{2\delta_n}\right) du.
\end{aligned}$$

As $T\lambda(B)\delta_n \leq 1$, among these terms the largest one is $\frac{8\delta_n}{c_0(T\lambda(B))^3}\exp\left(-\frac{c_0}{2\delta_nT\lambda(B)}\right)$. Then, $J_2 \leq C_2\frac{\delta_n}{(T\lambda(B))^3}\exp\left(-\frac{c_0}{2\delta_nT\lambda(B)}\right)$, where $C_2 = \frac{8}{c_0}(1+\frac{6}{c_0}+\frac{24}{c_0^2}+\frac{48}{c_0^3})$. Combining (12.7) and the upper bounds for J_1 and J_2 yields the result. □

12.2.2 Results on $\partial_x f_n$

For all $(s,y) \in [0,T] \times \mathbb{R}^d$, for all $i \in \{1,\cdots,d\}$,

$$\partial_{x_i} f_n(s,y) = \frac{1}{nh_t h_x^{d+1}} \sum_{i=1}^{n} K_t\left(\frac{s-T_i}{h_t}\right) \partial_{x_i} K_x\left(\frac{y-X_i}{h_x}\right).$$

Lemma 12.17. *For all $(s,y) \in [0,T] \times \mathbb{R}^d$, for all $i \in \{1,\cdots,d\}$, the following assertion holds*

$$\mathbb{E}(\partial_{x_i} f_n(s,y))^2 \leq 2^{2d+3} \frac{|K_t|_\infty^2 |\partial_{x_i} K_x^i|_\infty^2}{h_x^2 (T\lambda(B))^2} \mathbf{1}_{y \in B_\infty(0,a+h_x)}.$$

Proof. We apply Lemma 12.8, page 127 to $C = \frac{1}{h_t h_x^{d+1}}$, $g_t = K_t$ and $g_x = \partial_{x_i} K_x$. □

Proposition 12.18. *There exist two constants c and C depending on d, $|K_t|_\infty$ and $|\partial_{x_i} K_x|_\infty$ such that for all $\epsilon^2 \leq (h_x T\lambda(B))^{-2}$ and $(s,y) \in [0,T] \times \mathbb{R}^d$, the following assertion holds*

$$\mathbb{E}\left(|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))|^2 \ \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon}\right)$$
$$\leq C\left(\epsilon^2 + \frac{\delta_n}{T\lambda(B) h_x^2}\right) \exp\left(-\frac{c\epsilon^2 h_x^2 T\lambda(B)}{\delta_n}\right).$$

Proof of Proposition 12.18. Let us introduce $\tilde{f}_n^i(s,y) = \frac{1}{h_x} \partial_{x_i} f_n(s,y)$. Then,

$$\mathbb{E}\left(|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))|^2 \ \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon}\right)$$
$$= \frac{1}{h_x^2} \mathbb{E}\left(|\tilde{f}_n^i(s,y) - \mathbb{E}(\tilde{f}_n^i(s,y))|^2 \mathbf{1}_{|\tilde{f}_n^i(s,y) - \mathbb{E}(\tilde{f}_n^i(s,y))| > \epsilon h_x}\right).$$

We apply Proposition 12.16 page 132 to $\tilde{f}_n^i$ (instead of f_n) and to ϵh_x (instead of ϵ), and the result follows. □

Proposition 12.19. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition.*

$$\int_0^T ds e^{\beta s} \int_B dy \nu^0(s,y) |\mathbb{E}(\partial_{x_i} f_n(s,y))|^2 \leq \frac{K(T)}{h_x (T\lambda(B))^2} e^{-\mu a} a^{d-1},$$

where $K(T)$ is a function non decreasing in T and depending on $d, \beta, \mu, |K_x^i|_\infty, K$ and α_2.

Proof of Proposition 12.19. We recall that $h_x << a$, $h_t << \frac{T}{2}$ and

$$\mathbb{E}(\partial_{x_i} f_n(s,y)) =$$
$$\frac{1}{h_t h_x^{d+1}} \frac{1}{T\lambda(B)} \int_0^T dr K_t\left(\frac{s-r}{h_t}\right) \prod_{j=1, j\neq i}^{d} \int_{-a}^{a} dx_j K_x^j\left(\frac{y_j - x_j}{h_x}\right) \int_{-a}^{a} dx_i \partial_{x_i} K_x^i\left(\frac{y_i - x_i}{h_x}\right).$$

We integrate $\int_{-a}^{a} dx_i \partial_{x_i} K_x^i\left(\frac{y_i - x_i}{h_x}\right)$ and we get

$$\frac{1}{h_x} \int_{-a}^{a} dx_i \partial_{x_i} K_x^i\left(\frac{y_i - x_i}{h_x}\right) = -K_x^i\left(\frac{y_i - a}{h_x}\right) + K_x^i\left(\frac{y_i + a}{h_x}\right).$$

If y_i belongs to $[-a+h_x, a-h_x]$, $K_x^i\left(\frac{y_i \pm a}{h_x}\right) = 0$. Then

$$\left|\frac{1}{h_x}\int_{-a}^{a} dx_i \partial_{x_i} K_x^i\left(\frac{y_i - x_i}{h_x}\right)\right| \leq |K_x^i|_\infty \mathbf{1}_{y\in B\setminus B_\infty(0,a-h_x)}.$$

Hence, $|\mathbb{E}(\partial_{x_i} f_n(s,y))| \leq \frac{|K_x^i|_\infty}{T\lambda(B)h_x}\mathbf{1}_{y\in B\setminus B_\infty(0,a-h_x)}$, and

$$\int_0^T ds e^{\beta s}\int_B dy \nu^0(s,y)|\mathbb{E}(\partial_{x_i} f_n(s,y))|^2 \leq \frac{|K_x^i|_\infty^2}{(T\lambda(B))^2 h_x^2}\int_0^T ds e^{\beta s}\int_{B\setminus B_\infty(0,a-h_x)} dy \nu^0(s,y).$$

We use (12.5), page 132, to end the proof.

□

12.2.3 Results on r_n

Let $r_n(s,y)$ be defined in the following way

$$r_n(s,y) = \frac{1}{nh_t h_x^d}\sum_{i=1}^n K_t\left(\frac{s-T_i}{h_t}\right)K_x\left(\frac{y-X_i}{h_x}\right)\phi(T_i,X_i),$$

where ϕ is a function from $[0,T]\times\mathbb{R}^d$ to $\mathbb{R}$.

Proposition 12.20. *Assume $\phi : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ is a function bounded by C_ϕ. Then, for all $(s,y)\in[0,T]\times\mathbb{R}^d$,*

$$(\mathbb{E}[r_n(s,y)])^2 \leq \frac{C_\phi^2}{(T\lambda(B))^2}\mathbf{1}_{y\in B_\infty(0,a+h_x)},\ \ Var(r_n(s,y)) \leq 2^{d+1}\frac{|K_t|_\infty^2|K_x|_\infty^2 C_\phi^2\delta_n}{T\lambda(B)}\mathbf{1}_{y\in B_\infty(0,a+h_x)}.$$

Using $T\lambda(B)\delta_n << 1$ yields $\mathbb{E}(r_n^2(s,y)) \leq 2^{d+2}\frac{|K_t|_\infty^2|K_x|_\infty^2 C_\phi^2}{(T\lambda(B))^2}\mathbf{1}_{y\in B_\infty(0,a+h_x)}$.
Assume $\phi : [0,T]\times\mathbb{R}^d \to \mathbb{R}^+$ is a continuous function. Then,

$$\mathbb{E}(r_n(s,y))^2 \leq \frac{2^{d+1}}{h_t h_x^d (T\lambda(B))^2}\int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right)\int_B dz K_x^2\left(\frac{y-z}{h_x}\right)\phi^2(r,z),$$

$$Var(r_n(s,y)) \leq \frac{1}{nh_t^2 h_x^{2d} T\lambda(B)}\int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right)\int_B dz K_x^2\left(\frac{y-z}{h_x}\right)\phi^2(r,z).$$

Using $T\lambda(B)\delta_n << 1$ yields

$$\mathbb{E}(r_n^2(s,y)) \leq \frac{2^{d+2}}{h_t h_x^d (T\lambda(B))^2}\int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right)\int_B dz K_x^2\left(\frac{y-z}{h_x}\right)\phi^2(r,z).$$

Proof of Proposition 12.20. First, we consider the case $\phi : [0,T]\times\mathbb{R}^d \to \mathbb{R}^+$ is a function bounded by C_ϕ.

$$|r_n(s,y)| \leq C_\phi \frac{1}{nh_t h_x^d}\sum_{i=1}^n K_t\left(\frac{s-T_i}{h_t}\right)K_x\left(\frac{y-X_i}{h_x}\right). \tag{12.8}$$

By using Lemma 12.8, page 127 with $C = \frac{C_\phi}{h_t h_x^d}$, $g_t = K_t$ and $g_x = K_x$ yields the result.
Finally, we assume ψ is a continuous function. Using Lemma 12.9, page 128 with $C = \frac{1}{h_t h_x^d}$, $g_t = K_t$ and $g_x = K_x$ yields the result. □

12.2.4 Results on $\partial_{x_i} r_n$

For all $(s,y) \in [0,T] \times \mathbb{R}^d$, for all $i \in \{1, \cdots, d\}$,

$$\partial_{x_i} r_n(s,y) = \frac{1}{nh_t h_x^{d+1}} \sum_{i=1}^{n} K_t\left(\frac{s-T_i}{h_t}\right) \partial_{x_i} K_x\left(\frac{y-X_i}{h_x}\right) \phi(T_i, X_i),$$

where ϕ is a function from $[0,T] \times \mathbb{R}^d$ to $\mathbb{R}$.

Proposition 12.21. *Assume $\phi : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ is a function bounded by C_ϕ. For all $(s,y) \in [0,T] \times \mathbb{R}^d$, for all $i \in \{1, \cdots, d\}$,*

$$\mathbb{E}[\partial_{x_i} r_n(s,y)]^2 \leq \frac{2^{d+3} C_\phi^2 |K_t|_\infty^2 |\partial_{x_i} K_x|^2}{h_x^2 (T\lambda(B))^2} \mathbf{1}_{y \in B_\infty(0, a+h_x)}.$$

Assume $\phi : [0,T] \times \mathbb{R}^d \to \mathbb{R}^+$ is a bounded function, continuously differentiable in space. Then,

$$\begin{aligned}
|\mathbb{E}(\partial_{x_i} r_n(s,y))| \leq & \, 2\frac{C_\phi |K_x^i|_\infty}{h_x T\lambda(B)} (\mathbf{1}_{|y_i+a| \leq h_x} + \mathbf{1}_{|y_i-a| \leq h_x}) \\
& + \frac{1}{h_t h_x^d} \frac{1}{T\lambda(B)} \left| \int_0^T dr K_t\left(\frac{s-r}{h_t}\right) \int_B dz K_x\left(\frac{y-z}{h_x}\right) \partial_{x_i} \phi(r,z) \right|, \\
Var(\partial_{x_i} r_n(s,y)) \leq & \, \frac{1}{nh_t h_x^{d+2} T\lambda(B)} \int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right) \int_B dz (\partial_{x_i} K_x)^2 \left(\frac{y-z}{h_x}\right) \phi^2(r,z).
\end{aligned}$$

Proof of Proposition 12.21. Assume ϕ is bounded by C_ϕ, we apply Lemma 12.8, page 127 with $C = \frac{C_\phi}{h_t h_x^{d+1}}$, $g_t = K_t$ and $g_x = \partial_{x_i} K_x$, and we use $T\lambda(B)\delta_n << 1$.
In the case ϕ differentiable in space, we write

$$\mathbb{E}(\partial_{x_i} r_n(s,y)) = \frac{1}{h_t h_x^{d+1}} \frac{1}{T\lambda(B)} \int_0^T dr K_t\left(\frac{s-r}{h_t}\right) \int_B dz \partial_{x_i} K_x\left(\frac{y-z}{h_x}\right) \phi(r,z).$$

We integrate by parts $\int_{-a}^{a} dz_i \partial_{x_i} K_x^i \left(\frac{y_i - z_i}{h_x}\right) \phi(r,z)$ and we get

$$\begin{aligned}
\frac{1}{h_x} \int_{-a}^{a} dz_i \partial_{x_i} K_x^i \left(\frac{y_i-z_i}{h_x}\right) \phi(r,z) = & - K_x^i\left(\frac{y_i-a}{h_x}\right) \phi(r, z_a^i) + K_x^i\left(\frac{y_i+a}{h_x}\right) \phi(r, z_{-a}^i) \\
& + \int_{-a}^{a} dz_i \partial_{x_i} \phi(r,z) K_x^i\left(\frac{y_i-z_i}{h_x}\right),
\end{aligned}$$

where z_y^i denotes the vector $(z_1, \cdots, z_{i-1}, y, z_{i+1}, \cdots, z_d)$. Since ϕ is bounded, Lemma 12.7, page 127 ends the proof for $\mathbb{E}(\partial_{x_i} r_n(s,y))$. Concerning $\mathrm{Var}(\partial_{x_i} r_n(s,y))$, we apply Lemma 12.9, page 128 $C = \frac{1}{h_t h_x^{d+1}}$, $g_t = K_t$ and $g_x = \partial_{x_i} K_x$.

□

Proposition 12.22. *Let f_n be defined as above. Let $\overline{r}_n$ be such that $\overline{r}_n(s,y) = \frac{1}{nh_t h_x^d} \sum_{i=1}^{n} K_t^2\left(\frac{s-T_i}{h_t}\right) K_x^2\left(\frac{y-X_i}{h_x}\right) \phi(T_i, X_i)$, where ϕ is a positive function. For all $(s,y) \in$*

$[0,T]\times\mathbb{R}^d$, for all $i\in\{1,\cdots,d\}$,

$$\mathbb{E}((\partial_{x_i} f_n(s,y))^2\overline{r}_n(s,y)) \leq \frac{\delta_n^2}{h_t h_x^{d+2}} A_n^1(s,y) + \frac{C}{h_t h_x^{d+2}}\left(\frac{\delta_n}{T\lambda(B)} + \frac{1}{(T\lambda(B))^2}\right) B_n^1(s,y) + \frac{C\delta_n}{h_t h_x^{d+2} T\lambda(B)} C_n^1(s,y),$$

where C depends on $|K_t|_\infty$ and $|\partial_{x_i} K_x|_\infty$ and

$$\begin{aligned} A_n^1(s,y) =& \mathbb{E}\left(K_t^4\left(\frac{s-T_1}{h_t}\right)(K_x\partial_{x_i}K_x)^2\left(\frac{y-X_1}{h_x}\right)\phi(T_1,X_1)\right),\\ B_n^1(s,y) =& \mathbb{E}\left(K_t^2\left(\frac{s-T_1}{h_t}\right)K_x^2\left(\frac{y-X_1}{h_x}\right)\phi(T_1,X_1)\right),\\ C_n^1(s,y) =& \mathbb{E}\left(K_t^3\left(\frac{s-T_1}{h_t}\right)(K_x^2\partial_{x_i}K_x)\left(\frac{y-X_1}{h_x}\right)\phi(T_1,X_1)\right). \end{aligned}$$

Proof. For the sake of clearness, we use the following notations: for all $j\in\{1,\cdots,d\}$, $K_t(j) := K_t\left(\frac{s-T_j}{h_t}\right)$, $K_x(j) := K_x\left(\frac{y-X_j}{h_x}\right)$, and $(\partial_{x_i}K_x)(j) := (\partial_{x_i}K_x)\left(\frac{y-X_j}{h_x}\right)$. Using the definition of $\partial_{x_i} f_n(s,y)$, we get

$$(\partial_{x_i} f_n(s,y))^2 = \frac{1}{n^2 h_t^2 h_x^{2d+2}}\left(\sum_{j=1}^n K_t^2(j)(\partial_{x_i}K_x)^2(j) + \sum_{i,j=1,i\neq j}^n K_t(i)K_t(j)(\partial_{x_i}K_x)(i)(\partial_{x_i}K_x)(j)\right).$$

First, we develop $(\partial_{x_i} f_n(s,y))^2\overline{r}_n(s,y)$

$$\begin{aligned}(\partial_{x_i} f_n(s,y))^2\overline{r}_n(s,y) =& \frac{1}{n^3 h_t^3 h_x^{3d+2}}\sum_{k=1}^n K_t^2(k)K_x^2(k)\phi(T_k,X_k)\\ &\left[\sum_{j=1}^n K_t^2(j)(\partial_{x_i}K_x)^2(j) + \sum_{i,j=1,i\neq j}^n K_t(i)K_t(j)(\partial_{x_i}K_x)(i)(\partial_{x_i}K_x)(j)\right].\end{aligned} \tag{12.9}$$

We write

$$\begin{aligned}\sum_{j=1}^n K_t^2(j)(\partial_{x_i}K_x)^2(j)\sum_{k=1}^n K_t^2(k)K_x^2(k)\phi(T_k,X_k) =& \sum_{j=1}^n K_t^4(j)K_x^2(j)(\partial_{x_i}K_x)^2(j)\phi(T_j,X_j)\\ &+ \sum_{j,k=1,j\neq k}^n K_t^2(j)(\partial_{x_i}K_x)^2(j)K_t^2(k)K_x^2(k)\phi(T_k,X_k).\end{aligned} \tag{12.10}$$

Then,

$$\begin{aligned}&\sum_{i,j=1,i\neq j}^n K_t(i)K_t(j)(\partial_{x_i}K_x)(i)(\partial_{x_i}K_x)(j)\sum_{k=1}^n K_t^2(k)K_x^2(k)\phi(T_k,X_k) = \\ &\sum_{i,j,k=1,i\neq j,i\neq k,j\neq k}^n K_t(i)K_t(j)(\partial_{x_i}K_x)(i)(\partial_{x_i}K_x)(j)K_t^2(k)K_x^2(k)\phi(T_k,X_k)\\ &+ 2\sum_{i,j=1,i\neq j}^n K_t^3(i)K_t(j)(\partial_{x_i}K_x)(i)(\partial_{x_i}K_x)(j)K_x^2(i)\phi(T_i,X_i).\end{aligned} \tag{12.11}$$

Combining (12.9), (12.10) and (12.11) and taking the expectation leads to

$$\begin{aligned}\mathbb{E}((\partial_{x_i} f_n(s,y))^2 \overline{r}_n(s,y)) = \frac{1}{n^3 h_t^3 h_x^{3d+2}} &\left[n\mathbb{E}(K_t^4(1) K_x^2(1) (\partial_{x_i} K_x)^2(1) \phi(T_1, X_1)) \right. \\ &+ n(n-1)\mathbb{E}(K_t^2(1)(\partial_{x_i} K_x)^2(1))\mathbb{E}(K_t^2(1)K_x^2(1)\phi(T_1,X_1)) \\ &+ n(n-1)(n-2)(\mathbb{E}(K_t(1)(\partial_{x_i} K_x)(1)))^2 \mathbb{E}(K_t^2(1)K_x^2(1)\phi(T_1,X_1)) \\ &\left. + 2n(n-1)\mathbb{E}(K_t(1)(\partial_{x_i} K_x)(1))\mathbb{E}(K_t^3(1)K_x^2(1)(\partial_{x_i} K_x)(1)\phi(T_1,X_1)) \right].\end{aligned}$$

Applying Lemma 12.8, page 127 ends the proof. □

12.3 Convergence of $\mathcal{P}_n v - v$

In this Section we assume v is a deterministic function. Let us recall (9.15): $\mathcal{P}_n v(s,y) = \dfrac{r_n(s,y)}{f_n(s,y)} g(2T\lambda(B) f_n(s,y))$, where

- $r_n(s,y) = \frac{1}{n h_t h_x^d} \sum_{i=1}^n K_t(\frac{s-T_i}{h_t}) K_x(\frac{y-X_i}{h_x}) v(T_i, X_i)$,
- $f_n(s,y) = \frac{1}{n h_t h_x^d} \sum_{i=1}^n K_t(\frac{s-T_i}{h_t}) K_x(\frac{y-X_i}{h_x})$.

We aim at studying $\|\mathcal{P}_n v - v\|^2_{H^\mu_{\beta,X}}$. X is the diffusion process defined in (9.2), page 93 and its transition density is denoted p.

Remark 12.23. For all random functions F and G from $[0,T] \times \mathbb{R}^d$ into $\mathbb{R}^q$ independent of X, we have

$$\begin{aligned}\|F - G\|^2_{H^\mu_{\beta,X}} &= \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dx e^{-\mu|x|} \mathbb{E}[|F(s, X_s^x) - G(s, X_s^x)|^2 | F, G], \\ &= \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y) |F(s,y) - G(s,y)|^2.\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{E}\,\|F - G\|^2_{H^\mu_{\beta,X}} &= \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dx e^{-\mu|x|} \mathbb{E}|F(s, X_s^x) - G(s, X_s^x)|^2, \\ &= \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y) \mathbb{E}|F(s,y) - G(s,y)|^2,\end{aligned}$$

where $\nu^0(s,y)$ is defined in Definition 12.2, page 126.

Proposition 12.24 (Bias-Variance decomposition)**.** *Let F and G be two random functions from $[0,T] \times \mathbb{R}^d$ into $\mathbb{R}^q$ independent of X.*

$$\mathbb{E}\,\|F - G\|^2_{H^\mu_{\beta,X}} = \|\mathbb{E}(F - G)\|^2_{H^\mu_{\beta,X}} + \sum_{i=1}^q \|Std(F_i - G_i)\|^2_{H^\mu_{\beta,X}},$$

where $Std(Y(s,y)) = (VarY(s,y))^{1/2}$ corresponds to the standard deviation of the real r.v. $Y(s,y)$.

Proof. From Remark 12.23 and using the bias-variance decomposition (see Lemma 12.6), we get

$$\mathbb{E}\,\|F-G\|^2_{H^\mu_{\beta,X}} = \int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)|\mathbb{E}(F(s,y)-G(s,y))|^2 + \int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)\sum_{i=1}^q \mathrm{Var}(F_i(s,y)-G_i(s,y)).$$

□

From Remark 12.23, we get

$$\mathbb{E}\,\|\mathcal{P}_n v - v\|^2_{H^\mu_{\beta,X}} = \int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)\mathbb{E}|\mathcal{P}_n v(s,y)-v(s,y)|^2. \tag{12.12}$$

We won't study $\mathcal{P}_n v(s,y)-v(s,y)$ directly. We introduce $A_n(s,y)$ and $B_n(s,y)$, defined by

Definition 12.25. $A_n(s,y)$ and $B_n(s,y)$ are defined in the following way

$$A_n(s,y) = T\lambda(B)r_n(s,y)\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in B}, \quad B_n(s,y) = T\lambda(B)r_n(s,y).$$

Remark 12.26. The functions $A_n(s,y)$ and $B_n(s,y)$ are "standard" estimators of $v(s,y)$. The functions A_n and B_n are close to each other, they only differ on $[-h_t,0]\cap[T,T+h_t]\times B_\infty(a+h_x)\backslash B$. The function $(s,y)\longmapsto T\lambda(B)\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in B}$ corresponds to $\frac{1}{f(s,y)}$. The function $(s,y)\longmapsto \frac{1}{f_n(s,y)}g(2T\lambda(B)f_n(s,y))$, appearing in the definition of $\mathcal{P}_n$, approximates $\frac{1}{f(s,y)}$. The study of $\mathcal{P}_n v - A_n$ corresponds to the study of $\frac{1}{f_n(s,y)}g(2T\lambda(B)f_n(s,y)) - \frac{1}{f(s,y)}$.

The study of $\mathbb{E}\,\|v-\mathcal{P}_n v\|^2_{H^\mu_{\beta,X}}$ will be done in three steps. Introducing A_n and B_n in $\mathbb{E}\,\|\mathcal{P}_n v - v\|^2_{H^\mu_{\beta,X}}$ yields

$$\mathbb{E}\,\|\mathcal{P}_n v - v\|^2_{H^\mu_{\beta,X}} \le 3\mathbb{E}\,\|\mathcal{P}_n v - A_n\|^2_{H^\mu_{\beta,X}} + 3\mathbb{E}\,\|A_n - B_n\|^2_{H^\mu_{\beta,X}} + 3\mathbb{E}\,\|B_n - v\|^2_{H^\mu_{\beta,X}}. \tag{12.13}$$

In Section 12.3.1, we study $\mathbb{E}\,\|A_n-B_n\|^2_{H^\mu_{\beta,X}}$. In Section 12.3.2, we deal with $\mathbb{E}\,\|\mathcal{P}_n v - A_n\|^2_{H^\mu_{\beta,X}}$, and in Section 12.3.3 we consider $\mathbb{E}\,\|B_n - v\|^2_{H^\mu_{\beta,X}}$. Section 12.3.4 ends the study of $\mathbb{E}\,\|v-\mathcal{P}_n v\|^2_{H^\mu_{\beta,X}}$.

12.3.1 Study of $\mathbb{E}\,\|A_n-B_n\|^2_{H^\mu_{\beta,X}}$

Using the definitions of $A_n(s,y)$ and $B_n(s,y)$, we get

$$A_n(s,y)-B_n(s,y) = T\lambda(B)r_n(s,y)(1-\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in B}) = -T\lambda(B)r_n(s,y)\mathbf{1}_{\{s\notin[0,T]\}\cup\{y\notin B\}}.$$

From Remark 12.23, we write

$$\mathbb{E}\,\|A_n-B_n\|^2_{H^\mu_{\beta,X}} = \int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)\mathbb{E}|A_n(s,y)-B_n(s,y)|^2.$$

Theorem 12.27. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume v is bounded by C_v.*

$$\mathbb{E} \left\| A_n - B_n \right\|^2_{H^\mu_{\beta,X}} \leq K(T) e^{-\mu a} a^{d-1} h_x.$$

$K(T)$ is a function non decreasing in T and depending on $d, \beta, |K_t|_\infty, |K_x|_\infty, C_v$ and on K, α_2.

Proof of Theorem 12.27. Using the above expression of $A_n(s,y) - B_n(s,y)$, we get

$$\mathbb{E}|A_n(s,y) - B_n(s,y)|^2 = (T\lambda(B))^2 \mathbb{E}|r_n(s,y)|^2 \mathbf{1}_{\{s \notin [0,T]\} \cup \{y \notin B\}}.$$

If v is bounded, using Proposition 12.20, page 136, yields $\mathbb{E}|A_n(s,y) - B_n(s,y)|^2 \leq C_v^2 2^{d+2} |K_t|^2_\infty |K_x|^2_\infty \mathbf{1}_{\{s \notin [0,T]\} \cup \{y \notin B\}} \mathbf{1}_{y \in B_\infty(0,a+h_x)}$. Moreover, using (12.5), page 132 leads to

$$\begin{aligned}\int_0^T dse^{\beta s} \int_{y \in B_\infty(0,a+h_x) \setminus B} dy \nu^0(s,y) &\leq 2^{2d} K e^{-\mu a} a^{d-1} h_x \int_0^T e^{(\beta + c_2)s} \\ &\leq K_0(T) e^{-\mu a} a^{d-1} h_x, \qquad (12.14)\end{aligned}$$

where $K_0(T)$ is a function non decreasing in T and depending on d, α, β, K and α_2. □

12.3.2 Study of $\mathbb{E} \left\| \mathcal{P}_n v - A_n \right\|^2_{H^\mu_{\beta,X}}$

By using the definition of $\mathcal{P}_n v(s,y)$ and $A_n(s,y)$, we write

$$\mathcal{P}_n v(s,y) - A_n(s,y) = r_n(s,y) \left[\frac{1}{f_n(s,y)} g(2T\lambda(B) f_n(s,y)) - T\lambda(B) \mathbf{1}_{s \in [0,T]} \mathbf{1}_{y \in B} \right]. \qquad (12.15)$$

Before integrating w.r.t. $e^{\beta s} \nu^0(s,y) dsdy$, we study $\mathcal{P}_n v(s,y) - A_n(s,y)$ w.r.t. the value of y in the Lemmas 12.28, 12.29, 12.30 and 12.31 (as we integrate w.r.t. $s \in [0,T]$, we never study the case $s \notin [0,T]$). Then, we bound $\mathbb{E}|\mathcal{P}_n v(s,y) - A_n(s,y)|^2$ in Proposition 12.32. To conclude, we bound $\mathbb{E} \left\| \mathcal{P}_n - A_n \right\|^2_{H^\mu_{\beta,X}}$ in Theorem 12.33.

Lemma 12.28. *On the set $\{y \notin B_\infty(0, a + h_x)\}$, $\mathcal{P}_n v(s,y) - A_n(s,y) = 0$.*

Proof. If $y \notin B_\infty(0, a + h_x)$, the second term in (12.15) is null. Moreover, there exists at least one $j \in \{1, \cdots, d\}$ such that $d(y_j, B) \geq h_x$. thus, $\forall z \in B$, $K_x^j(\frac{y_j - z}{h_x}) = 0$. Since $\forall i = 1, \cdots, n$, $X_i \in B$, we get $f_n(s,y) = 0$. Looking at the definition of g in Proposition 11.6 (see page 121) yields $\frac{1}{f_n(s,y)} g(2T\lambda(B) f_n(s,y)) = 0$. Hence, the result follows. □

Lemma 12.29. *On the set $\{y \in B_\infty(0, a + h_x) \setminus B\}$, if v is bounded by C_v, $|\mathcal{P}_n v(s,y) - A_n(s,y)| \leq C_v$.*

Proof. If $y \in B_\infty(0, a + h_x) \setminus B$, the second term in (12.15) is null. $\mathcal{P}_n v(s,y) - A_n(s,y) = \frac{r_n(s,y)}{f_n(s,y)} g(2T\lambda(B) f_n(s,y))$. Since g is bounded by 1 and

$$|r_n(s,y)| \leq f_n(s,y) \sup_{(s,y) \in [0,T] \times B_\infty(0,a+h_x) \setminus B} |v(s,y)|.$$

The result follows. □

On the set $\{y \in B\}$, we bound from above $\mathcal{P}_n v(s,y) - A_n(s,y)$ of two manners (see Lemmas 12.30 and 12.31)

Lemma 12.30. *On the set $\{y \in B\}$, if v is bounded by C_v, $|\mathcal{P}_n v(s,y) - A_n(s,y)| \leq 3C_v T\lambda(B)|f_n(s,y) - \overline{f}(s,y)|$.*

Proof of Lemma 12.30. If $y \in B$ and $s \in [0,T]$ we write $\mathcal{P}_n v(s,y) - A_n(s,y) = r_n(s,y)\left[\frac{1}{f_n(s,y)} g(2T\lambda(B)f_n(s,y)) - T\lambda(B)\right]$. Since $y \in B$, $\overline{f}(s,y) = \frac{1}{T\lambda(B)}$ (we refer to Definition 12.5, page 127, for the definition of $\overline{f}$). We write $\mathcal{P}_n v(s,y) - A_n(s,y) = \frac{r_n(s,y)}{f_n(s,y)}(g(2T\lambda(B)f_n(s,y)) - T\lambda(B)f_n(s,y))$. Then, we study $\tilde{g}(x) := g(2T\lambda(B)x) - T\lambda(B)x$. $\tilde{g}(\overline{f}(s,y)) = \tilde{g}(\frac{1}{T\lambda(B)}) = 0$.

$$|\tilde{g}(f_n(s,y)) - \tilde{g}(\overline{f}(s,y))| \leq |f_n(s,y) - \overline{f}(s,y)||\tilde{g}'|_\infty \leq 3T\lambda(B)|f_n(s,y) - \overline{f}(s,y)|.$$

As before, we use $|r_n(s,y)| \leq f_n(s,y)\sup_{(s,y)\in[0,T]\times B}|v(s,y)|$, and the result follows. □

Lemma 12.31. *On the set $\{y \in B\}$, $|\mathcal{P}_n v(s,y) - A_n(s,y)| \leq 8(T\lambda(B))^2|r_n(s,y)||f_n(s,y) - \overline{f}(s,y)|$.*

Proof of Lemma 12.31. $|\mathcal{P}_n v(s,y) - A_n(s,y)| \leq |r_n(s,y)||\frac{1}{f_n(s,y)}g(2T\lambda(B)f_n(s,y)) - T\lambda(B)|$. We introduce $\overline{g} : x \longmapsto \frac{g(2T\lambda(B)x)}{x}$. $\overline{g}(\overline{f}(s,y)) = T\lambda(B)$. Thus, $|\overline{g}(f_n(s,y)) - \overline{g}(\overline{f}(s,y))| \leq |\overline{g}'|_\infty |f_n(s,y) - \overline{f}(s,y)|$, and $|\overline{g}'|_\infty \leq 8(T\lambda(B))^2$. □

Proposition 12.32. *Let ϵ be such that $0 \leq \epsilon \leq (T\lambda(B))^{-1}$. On the set $\{y \in B\}$ and for v bounded by C_v, we get*

$$\begin{aligned}\mathbb{E}|\mathcal{P}_n v(s,y) - A_n(s,y)|^2 \leq & 128\epsilon^2(T\lambda(B))^4\mathbb{E}(r_n^2(s,y)) \\ & + C(T\lambda(B))^2|\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|^2 \\ & + C(T\lambda(B))^2\left(\epsilon^2 + \frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon^2 T\lambda(B)}{\delta_n}\right),\end{aligned}$$

where C depends on $C_v, d, |K_t|_\infty$ and $|K_x|_\infty$ and c depends on $d, |K_t|_\infty$ and $|K_x|_\infty$.

Proof. We introduce $\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon}$ and $\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|\leq\epsilon}$, and we split $\mathcal{P}_n v(s,y) - A_n(s,y)$ w.r.t. these indicators. On the set $\{|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \epsilon\}$, we use Lemma 12.30 to bound $|\mathcal{P}_n v(s,y) - A_n(s,y)|$. On the other set, we use Lemma 12.31. We get

$$\begin{aligned}|\mathcal{P}_n v(s,y) - A_n(s,y)|^2 \leq & 9C_v^2(T\lambda(B))^2|f_n(s,y) - \overline{f}(s,y)|^2\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon} \\ & + 64(T\lambda(B))^4 r_n^2(s,y)|f_n(s,y) - \overline{f}(s,y)|^2\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|\leq\epsilon}.\end{aligned}$$

Then, we use $|f_n(s,y) - \overline{f}(s,y)| \leq |f_n(s,y) - \mathbb{E}(f_n(s,y))| + |\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|$. Using $|r_n(s,y)| \leq C_v f_n(s,y)$ yields

$$\begin{aligned}\mathbb{E}|\mathcal{P}_n v(s,y) - A_n(s,y)|^2 \leq & 128\epsilon^2(T\lambda(B))^4\mathbb{E}(r_n^2(s,y)) \\ & + 128C_v^2(T\lambda(B))^4\mathbb{E}(f_n^2(s,y))|\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|^2 \\ & + 18C_v^2(T\lambda(B))^2\mathbb{E}[|f_n(s,y) - \mathbb{E}(f_n(s,y))|^2\mathbf{1}_{|f_n(s,y)-\mathbb{E}(f_n(s,y))|>\epsilon}] \\ & + 18C_v^2(T\lambda(B))^2|\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|^2\mathbb{P}(|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \epsilon).\end{aligned}$$

We bound the second term of the r.h.s. of the above inequality using Lemma 12.14, page 131. We use Proposition 12.16, page 132 to bound the third and fourth terms. The result follows. □

Using Proposition 12.32 and Lemmas 12.28 and 12.29 enables us to state the following Theorem

Theorem 12.33. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. If v is bounded by C_v, $\forall\epsilon \geq 0$ such that $\epsilon^2 \leq (T\lambda(B))^{-2}$,*

$$\begin{aligned}\mathbb{E}\left\|\mathcal{P}_n v - A_n\right\|^2_{H^\mu_{\beta,X}} \leq & K_0(T)\epsilon^2(T\lambda(B))^2 \left\|v\right\|^2_{H^\mu_{\beta,X}} \\ & + K_1(T)(h_t + e^{-\mu a}a^{d-1}h_x) \\ & + K_2(T)(T\lambda(B))^2\left(\epsilon^2 + \frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon^2 T\lambda(B)}{\delta_n}\right),\end{aligned}$$

where $K_0(T), K_1(T)$ and $K_2(T)$ are functions non decreasing in T. $K_0(T)$ depends on $|K_t|_\infty, |K_x|_\infty, \beta, \alpha_2, d, \mu$ and K. $K_1(T)$ depends on C_v, d, β, K and α_2. $K_2(T)$ depends on $C_v, d, |K_t|_\infty, |K_x|_\infty, K, \mu, \beta$ and α_2.
In particular, for $\epsilon = 0$ we get

$$\mathbb{E}\left\|\mathcal{P}_n v - A_n\right\|^2_{H^\mu_{\beta,X}} \leq K_1(T)(h_t + e^{-\mu a}a^{d-1}h_x) + K_2(T)T\lambda(B)\delta_n.$$

Proof of Theorem 12.33.

$$\mathbb{E}\left\|\mathcal{P}_n v - A_n\right\|^2_{H^\mu_{\beta,X}} = \int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)\mathbb{E}|\mathcal{P}_n v(s,y) - A_n(s,y)|^2.$$

If v is bounded by C_v, we combine Lemmas 12.28, 12.29 and Proposition 12.32 to get

$$\begin{aligned}\mathbb{E}\|\mathcal{P}_n v - A_n\|^2_{H^\mu_{\beta,X}} & \leq C_v^2\int_0^T dse^{\beta s}\int_{B_\infty(0,a+h_x)\setminus B} dy\nu^0(s,y) \\ & + 128\epsilon^2(T\lambda(B))^4\int_0^T dse^{\beta s}\int_B dy\nu^0(s,y)\mathbb{E}(r_n^2(s,y)) \\ & + C(T\lambda(B))^2\int_0^T dse^{\beta s}\int_B dy\nu^0(s,y)|\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|^2 \\ & + C(T\lambda(B))^2\left(\epsilon^2 + \frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon^2 T\lambda(B)}{\delta_n}\right)\int_0^T dse^{\beta s}\int_B dy\nu^0(s,y),\end{aligned}$$

where C depends on $C_v, d, |K_t|_\infty$ and $|K_x|_\infty$ and c depends on $d, |K_t|_\infty$ and $|K_x|_\infty$. We use the last result of Proposition 12.20, page 136 to bound $\mathbb{E}(r_n^2(s,y))$. We use (12.14) to bound $\int_0^T dse^{\beta s}\int_{B_\infty(0,a+h_x)\setminus B} dy\nu^0(s,y)$ and Proposition 12.15, page 131 to bound the third term. Beside that, (6.11), page 67 leads to

$$\int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y) \leq 2^d K\int_0^T dse^{(\beta+c_2)s}\int_{\mathbb{R}^d} dye^{-\mu|y|} \leq 2^{2d}d^{\frac{d}{2}}\frac{K}{\mu^d}Te^{(\beta+c_2)T}. \tag{12.16}$$

We use (12.16) to bound the last term. We get

$$\begin{aligned}\mathbb{E}\|\mathcal{P}_n v - A_n\|^2_{H^\mu_{\beta,X}} & \leq K_0(T)e^{-\mu a}a^{d-1}h_x \\ & + 2^{d+9}\frac{\epsilon^2(T\lambda(B))^2}{h_t h_x^d}\int_0^T dse^{\beta s}\int_B dy\nu^0(s,y)\int_0^T drK_t^2\left(\frac{s-r}{h_t}\right)\int_B dzK_x^2\left(\frac{y-z}{h_x}\right)v^2(r,z) \\ & + K(T)(h_t + e^{-\mu a}a^{d-1}h_x) \\ & + K(T)(T\lambda(B))^2\left(\epsilon^2 + \frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon^2 T\lambda(B)}{\delta_n}\right),\end{aligned}$$

where $K_0(T)$ depends on C_v, d, β, K and α_2. $K(T)$ depends on $C_v, d, |K_t|_\infty, |K_x|_\infty, K, \mu, \beta$ and α_2. To conclude, we apply Lemma 12.10, page 128 with $g_t = K_t^2$, $g_x = K_x^2$ and $f \equiv v$. $\square$

12.3.3 Study of $\mathbb{E}\left\|B_n - v\right\|^2_{H^\mu_{\beta,X}}$

By using the definition of B_n we get

$$B_n(s,y) - v(s,y) = T\lambda(B) r_n(s,y) - v(s,y).$$

Proposition 12.24, page 139 gives us $\mathbb{E}\left\|B_n - v\right\|^2_{H^\mu_{\beta,X}} = \|\mathbb{E}(B_n - v)\|^2_{H^\mu_{\beta,X}} + \|\mathrm{Std}(B_n - v)\|^2_{H^\mu_{\beta,X}}$. The scheme of the Section is the following. We bound $\|\mathrm{Std}(B_n - v)\|^2_{H^\mu_{\beta,X}}$ in Proposition 12.34. Then, we study $\|\mathbb{E}(B_n - v)\|^2_{H^\mu_{\beta,X}}$. To do so, Lemma 12.35 splits $\mathbb{E}(B_n(s,y)) - v(s,y)$ in four terms. From Lemma 12.36 to Lemma 12.39, we study these four terms. Using these results we bound $\|\mathbb{E}(B_n - v)\|^2_{H^\mu_{\beta,X}}$ in Proposition 12.40. Theorem 12.41 ends the Section by bounding $\mathbb{E}\left\|B_n - v\right\|^2_{H^\mu_{\beta,X}}$.

Proposition 12.34. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition.*

$$\|Std(B_n - v)\|^2_{H^\mu_{\beta,X}} \leq K(T) T\lambda(B) \delta_n \left\|v\right\|^2_{H^\mu_{\beta,X}},$$

where $K(T)$ depends on $\mu, d, \beta, K, \alpha_1$ and α_2. If v is bounded by C_v, we get $\|Std(B_n - v)\|^2_{H^\mu_{\beta,X}} \leq K'(T) T\lambda(B)\delta_n$, where $K'(T)$ depends on $C_v, |K_t|_\infty, |K_x|_\infty, \mu, d, \beta, K, \alpha_1$ and α_2.

Proof of Proposition 12.34.

$$\begin{aligned}\|\mathrm{Std}(B_n - v)\|^2_{H^\mu_{\beta,X}} &= \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y) \mathrm{Var}(B_n(s,y) - v(s,y)) \\ &= (T\lambda(B))^2 \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y) \mathrm{Var}(r_n(s,y)).\end{aligned}$$

Using Proposition 12.20, page 136 yields

$$\|\mathrm{Std}(B_n - v)\|^2_{H^\mu_{\beta,X}} \leq \frac{T\lambda(B)}{n h_t^2 h_x^{2d}} \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y) \int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right) \int_B dz K_x^2 \left(\frac{y-z}{h_x}\right) v^2(r,z).$$

As above, we apply Lemma 12.10, page 128 with $g_t = K_t^2$, $g_x = K_x^2$ and $f \equiv v$ to get the result. If v is bounded, we combine Proposition 12.20, page 136 and (12.16) to get the result. $\square$

Second, we study $\|\mathbb{E}(B_n - v)\|^2_{H^\mu_{\beta,X}}$. First, let us split $\mathbb{E}(B_n(s,y)) - v(s,y)$.

Lemma 12.35. *For all $(s,y) \in [0,T] \times \mathbb{R}^d$,*

$$\begin{aligned}\mathbb{E}(B_n(s,y)) - v(s,y) =& \frac{1}{h_t h_x^d} \int_{\mathbb{R}^d} dz \int_0^T dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right)(v(r,z)-v(s,z)) \\ &+ \frac{1}{h_t h_x^d} \int_{\mathbb{R}^d} dz \int_0^T dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right)(v(s,z)-v(s,y)) \\ &- \frac{1}{h_t h_x^d} \int_{B^c} dz \int_0^T dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right) v(r,z) \\ &- \frac{1}{h_t h_x^d} \int_{\mathbb{R}^d} dz \int_{[-h_t,0]\cup[T,T+h_t]} dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right) v(s,y) \\ &:= B_1(s,y) + B_2(s,y) + B_3(s,y) + B_4(s,y).\end{aligned}$$

Proof. We can write $v(s,y) = v(s,y)\frac{1}{h_t h_x^d}\int_{\mathbb{R}^d} dz K_x(\frac{y-z}{h_x}) \int_{-h_t}^{T+h_t} dr K_t(\frac{s-r}{h_t})$. Since $\mathbb{E}(B_n(s,y)) = \frac{1}{h_t h_x^d}\int_B dz K_x(\frac{y-z}{h_x}) \int_0^T dr K_t(\frac{s-r}{h_t}) v(r,z)$, the result follows. □

Through the following four Lemmas, we study the norms $\|B_1(s,y)\|^2_{H^\mu_{\beta,X}}$, $\|B_2(s,y)\|^2_{H^\mu_{\beta,X}}$, $\|B_3(s,y)\|^2_{H^\mu_{\beta,X}}$ and $\|B_4(s,y)\|^2_{H^\mu_{\beta,X}}$.

Lemma 12.36. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume v is a function C^1 in time. There exists a function $K(T)$ such that*

$$\|B_1(s,y)\|^2_{H^\mu_{\beta,X}} \le K(T) h_t^2 \|\partial_t v\|^2_{H^\mu_{\beta,X}},$$

where $K(T)$ depends on $d, \mu, |K_t|_\infty, |K_x|_\infty, \beta, K, \alpha_1$, and α_2. If $\partial_t v$ is bounded, we get $\|B_1(s,y)\|^2_{H^\mu_{\beta,X}} \le K'(T) h_t^2$, where $K'(T)$ depends on $|\partial_t v|^2_\infty$ and on the same parameters as $K(T)$.

Proof. We recall

$$B_1(s,y) = \frac{1}{h_t h_x^d} \int_{\mathbb{R}^d} dz \int_0^T dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right)(v(r,z)-v(s,z)).$$

Write $v(r,z) - v(s,z) = \int_s^r \partial_t v(t,z) dt$. Since $|r-s| \le h_t$, we get

$$B_1^2(s,y) \le \frac{2^{d+2}}{h_x^d} \int_{\mathbb{R}^d} dz K_x^2\left(\frac{y-z}{h_x}\right) \int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right) \int_{r-h_t}^{r+h_t} (\partial_t v(t,z))^2 dt. \qquad (12.17)$$

Applying Lemma 12.10, page 128 gives

$$\|B_1(s,y)\|^2_{H^\mu_{\beta,X}} \le K_0(T) e^{d\mu h_x} h_t \int_0^T dr e^{\beta r} \int_{\mathbb{R}^d} dz \nu^0(r,z) \int_{r-h_t}^{r+h_t} (\partial_t v(t,z))^2 dt, \qquad (12.18)$$

where $K_0(T)$ depends on $|K_t|_\infty, |K_x|_\infty, \beta, K$ and α_2. We write $\int_{r-h_t}^{r+h_t}(\partial_t v(t,z))^2 dt = \int_0^T \mathbf{1}_{t-h_t \le r \le t+h_t} (\partial_t v(t,z))^2 dt$ and $e^{\beta r} \le e^{\beta h_t} e^{\beta t}$. Moreover, Corollary 6.13 (see page 68) gives $\nu^0(r,z) \le C e^{c_2 h_t} \nu^0(t,z)$, where C depends on d, μ, T, K, α_1 and α_2. Equality (12.18) becomes

$$\|B_1(s,y)\|^2_{H^\mu_{\beta,X}} \le C K_0(T) e^{d\mu h_x} e^{(\beta+c_2)h_t} h_t^2 \int_0^T dt e^{\beta t} \int_{\mathbb{R}^d} dz \nu^0(t,z)(\partial_t v(t,z))^2,$$

and the result follows.
If $\partial_t v$ is bounded, (12.17) becomes

$$B_1^2(s,y) \leq 2^{2(d+2)}|K_t|_\infty^2 |K_x|_\infty^2 |\partial_t v|_\infty^2 h_t^2.$$

We use (12.16) to conclude. □

Lemma 12.37. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume v is a function C^2 in space. There exists a function $K(T)$ such that*

$$\|B_2(s,y)\|_{H^\mu_{\beta,X}}^2 \leq K(T)h_x^4 \left\|\partial_x^2 v\right\|_{H^\mu_{\beta,X}}^2,$$

where $K(T)$ depends on $d,\mu,|K_x|_\infty,\beta,K,\alpha_1$, and α_2. In particular, if $\partial_x^2 v$ is bounded, we get $\|B_2(s,y)\|_{H^\mu_{\beta,X}}^2 \leq K'(T)h_x^4$, where $K'(T)$ depends on $|\partial_x^2 v|_\infty^2$ and on the same parameters as $K(T)$.

Proof of Lemma 12.37. The proof is analogous to the one of Lemma 12.36, except that we exploit the symmetry of the kernel to get more accurate estimates (h_x^4 instead of h_x^2 only). We recall

$$B_2(s,y) = \frac{1}{h_t h_x^d}\int_{\mathbb{R}^d} dz \int_0^T dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right)(v(s,z)-v(s,y)).$$

We apply a Taylor expansion formula to each component of the following sum

$$v(s,z)-v(s,y) = \sum_{i=1}^d v(s,\overline{z}_i) - v(s,\overline{z}_{i-1}),$$

where $\overline{z}_i = (z_1,z_2,\cdots,z_i,y_{i+1},\cdots,y_d), \forall i\in\{1,\cdots,d\}$, and $\overline{z}_0 = y$. For all $i\in\{1,\cdots,d\}$, we get

$$v(s,\overline{z}_i) - v(s,\overline{z}_{i-1}) = (z_i-y_i)\partial_{x_i}v(s,\overline{z}_{i-1}) + \int_{y_i}^{z_i} dl(z_i-l)\partial_{x_i}^2 v(s,\overline{z}_i^l),$$

where $\overline{z}_i^l = (z_1,\cdots,z_{i-1},l,y_{i+1},\cdots,y_d)$. Plugging this result in the definition of $B_2(s,y)$ leads to

$$\begin{aligned} B_2(s,y) =& \frac{1}{h_t h_x^d}\sum_{i=1}^d \int_{\mathbb{R}^d} dz \int_0^T dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right)(z_i-y_i)\partial_{x_i}v(s,\overline{z}_{i-1}) \\ &+ \frac{1}{h_t h_x^d}\sum_{i=1}^d \int_{\mathbb{R}^d} dz \int_0^T dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right)\int_{y_i}^{z_i} dl(z_i-l)\partial_{x_i}^2 v(s,\overline{z}_i^l), \end{aligned}$$

The first integral of the r.h.s. of the above expression is null since for all $i\in\{1,\cdots,d\}$, $\int_{\mathbb{R}} dz_i K_x^i\left(\frac{y_i-z_i}{h_x}\right)(z_i-y_i) = 0$ (K_x^i is an even function). Moreover, $|z_i-l|\leq h_x$, $\int_0^T dr K_t(\frac{s-r}{h_t}) \leq h_t$ and $\int_{\mathbb{R}^d} dz K_x(\frac{y-z}{h_x})\int_{y_i}^{z_i} dl|\partial_{x_i}^2 v(s,\overline{z}_i^l)| \leq$

$(2h_x)^{d-i+1}|K_x|_\infty \int_{y_1-h_x}^{y_1+h_x} dz_1 \cdots \int_{y_{i-1}-h_x}^{y_{i-1}+h_x} dz_{i-1} \int_{y_i-h_x}^{y_i+h_x} dl|\partial^2_{x_i} v(s, \overline{z}^l_i)|$. Hence,

$$\begin{aligned}|B_2(s,y)| &\leq |K_x|_\infty \sum_{i=1}^d \frac{2^{d-i+1}}{h_x^{i-2}} \int_{y_1-h_x}^{y_1+h_x} dz_1 \cdots \int_{y_{i-1}-h_x}^{y_{i-1}+h_x} dz_{i-1} \int_{y_i-h_x}^{y_i+h_x} dl|\partial^2_{x_i} v|(s, \overline{z}^l_i) \\ &\leq 2^d |K_x|_\infty \sum_{i=1}^d B_2^i(s,y),\end{aligned}$$

where $B_2^i(s,y) := \frac{1}{h_x^{i-2}} \int_{y_1-h_x}^{y_1+h_x} dz_1 \cdots \int_{y_{i-1}-h_x}^{y_{i-1}+h_x} dz_{i-1} \int_{y_i-h_x}^{y_i+h_x} dl|\partial^2_{x_i} v|(s, \overline{z}^l_i)$, and

$$(B_2^i(s,y))^2 \leq 2^i h_x^{4-i} \int_{y_1-h_x}^{y_1+h_x} dz_1 \cdots \int_{y_{i-1}-h_x}^{y_{i-1}+h_x} dz_{i-1} \int_{y_i-h_x}^{y_i+h_x} dl|\partial^2_{x_i} v|^2(s, \overline{z}^l_i).$$

Beside that,

$$\|B_2(s,y)\|^2_{H^\mu_{\beta,X}} \leq d2^{2d}|K_x|^2_\infty \sum_{i=1}^d \left\|B_2^i(s,y)\right\|^2_{H^\mu_{\beta,X}}. \tag{12.19}$$

It remains to upper bound $\left\|B_2^i(s,y)\right\|^2_{H^\mu_{\beta,X}}$:

$$\begin{aligned}&\left\|B_2^i(s,y)\right\|^2_{H^\mu_{\beta,X}} = \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y) (B_2^i(s,y))^2 \\ &\leq 2^i h_x^{4-i} \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y) \int_{y_1-h_x}^{y_1+h_x} dz_1 \cdots \int_{y_{i-1}-h_x}^{y_{i-1}+h_x} dz_{i-1} \int_{y_i-h_x}^{y_i+h_x} dl|\partial^2_{x_i} v|^2(s, \overline{z}^l_i).\end{aligned}$$

Corollary 6.13, page 68 gives $\nu^0(s,y) \leq Ce^{\mu i h_x} \nu^0(s, \overline{z}^l_i)$ where C depends on $T, \mu, d, K, \alpha_1, \alpha_2$. Then

$$\begin{aligned}\left\|B_2^i(s,y)\right\|^2_{H^\mu_{\beta,X}} \leq C2^i h_x^{4-i} e^{\mu i h_x} &\int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} d\overline{z}^l_i \nu^0(s, \overline{z}^l_i) |\partial^2_{x_i} v|^2(s, \overline{z}^l_i) \\ &\times \int_{\mathbb{R}} dy_1 \mathbf{1}_{z_1-h_x \leq y_1 \leq z_1+h_x} \cdots \int_{\mathbb{R}} dy_i \mathbf{1}_{l-h_x \leq y_i \leq l+h_x}.\end{aligned}$$

Thus, $\left\|B_2^i(s,y)\right\|^2_{H^\mu_{\beta,X}} \leq C2^{2i} h_x^4 e^{\mu i h_x} \left\|\partial^2_{x_i} v\right\|^2_{H^\mu_{\beta,X}}$. Plugging this result in (12.19) ends the proof. $\square$

Lemma 12.38. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition. If v is bounded by C_v, we get $\|B_3(s,y)\|^2_{H^\mu_{\beta,X}} \leq K(T)e^{-\frac{\mu}{\sqrt{d}}a}$, where $K(T)$ depends on $d, \mu, K, \beta, \alpha_2, C_v$ and $|K_x|_\infty$.*

Proof of Lemma 12.38. Assume that v is bounded. $B_3(s,y) = -\frac{1}{h_t h_x^d} \int_{B^c} dz \int_0^T dr K_x\left(\frac{y-z}{h_x}\right) K_t\left(\frac{s-r}{h_t}\right) v(r,z)$.

$$(B_3(s,y))^2 \leq 2^{d+2} \frac{|K_x|^2_\infty C_v^2}{h_x^d} \int_{B^c} dz \mathbf{1}_{|y_1-z_1| \leq h_x} \cdots \mathbf{1}_{|y_d-z_d| \leq h_x}.$$

Moreover, $\int_{B^c} dz\mathbf{1}_{|y_1-z_1|\le h_x}\cdots\mathbf{1}_{|y_d-z_d|\le h_x} \le \mathbf{1}_{y\notin B_\infty(0,a-h_x)}(2h_x)^d$. Then,

$$\|B_3(s,y)\|^2_{H^\mu_{\beta,X}} \le 2^{2d+2}|K_x|^2_\infty|K_t|^2_\infty C_v^2\int_0^T dse^{\beta s}\int_{\overline{B_\infty(0,a-h_x)}} dy\nu^0(s,y).$$

Using (6.11) page 67 leads to $\int_0^T dse^{\beta s}\int_{\overline{B_\infty(0,a-h_x)}} dy\nu^0(s,y) \le 2^d Ke^{(\beta+c_2)T}T\int_{\overline{B_\infty(0,a-h_x)}} dye^{-\mu|y|}$. It remains to bound $\int_{\overline{B_\infty(0,a-h_x)}} dye^{-\mu|y|}$:

$$\int_{\overline{B_\infty(0,a-h_x)}} dye^{-\mu|y|} \le d\int_{|y_1|\ge a-h_x} dy_1e^{-\frac{\mu}{\sqrt{d}}|y_1|}\int_{\mathbb{R}} dy_2e^{-\frac{\mu}{\sqrt{d}}|y_2|}\cdots\int_{\mathbb{R}} dy_de^{-\frac{\mu}{\sqrt{d}}|y_d|}. \quad (12.20)$$

Thus, $\int_{\overline{B_\infty(0,a-h_x)}} dye^{-\mu|y|} \le \frac{2^d d^{\frac{d}{2}}}{\mu^d}e^{-\frac{\mu}{\sqrt{d}}(a-h_x)}$, which ends the proof.

□

Lemma 12.39. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. If v is bounded by C_v, we get $\|B_4(s,y)\|^2_{H^\mu_{\beta,X}} \le K(T)h_t$, where $K(T)$ depends on $d,\mu,K,\beta,\alpha_2,C_v$ and $|K_t|_\infty$.*

Proof of Lemma 12.39. We recall

$$B_4(s,y) = -\frac{1}{h_th_x^d}\int_B dz\int_{[-h_t,0]\cup[T,T+h_t]} drK_x\left(\frac{y-z}{h_x}\right)K_t\left(\frac{s-r}{h_t}\right)v(s,y).$$

Since v is bounded, $|B_4(s,y)| \le \frac{C_v}{h_th_x^d}\int_B dzK_x(\frac{y-z}{h_x})\int_{[-h_t,0]\cup[T,T+h_t]} drK_t(\frac{s-r}{h_t})$. Using Lemma 12.7, page 127 yields $\int_B dzK_x(\frac{y-z}{h_x}) \le h_x^d$. Since $|s-r| \le h_t$, $\int_{-h_t}^0 drK_t(\frac{s-r}{h_t}) \le |K_t|_\infty\int_{-h_t}^0 dr\mathbf{1}_{r-h_t\le s\le r+h_t} \le h_t|K_t|_\infty\mathbf{1}_{-2h_t\le s\le h_t}$ and $\int_T^{T+h_t} drK_t(\frac{s-r}{h_t}) \le h_t|K_t|_\infty\mathbf{1}_{T-h_t\le s\le T+2h_t}$. We easily get

$$\forall(s,y)\in[0,T]\times\mathbb{R}^d,\quad |B_4(s,y)| \le C_v|K_t|_\infty\mathbf{1}_{s\in[0,h_t]\cup[T-h_t,T]},$$

and we deduce

$$\|B_4(s,y)\|^2_{H^\mu_{\beta,X}} \le C_v^2|K_t|^2_\infty\left(\int_0^{h_t}e^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y) + \int_{T-h_t}^T e^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\right).$$

Inequalities (12.3) and (12.4) page 132 end the proof.

□

Combining Lemmas 12.35, 12.36, 12.37, 12.38 and 12.39 leads to the following Proposition.

Proposition 12.40. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition.*

1. *Assume v is a $C^{1,2}$ function and v is bounded by C_v. Then,*

$$\|\mathbb{E}(B_n-v)\|^2_{H^\mu_{\beta,X}} \le K(T)(h_t^2\|\partial_tv\|^2_{H^\mu_{\beta,X}}+h_x^4\left\|\partial_x^2v\right\|^2_{H^\mu_{\beta,X}}) + K(T)(e^{-\frac{\mu}{\sqrt{d}}a}+h_t).$$

where $K(T)$ depends on $d,\mu,K,\alpha_1,\alpha_2,\beta,C_v,|K_t|_\infty$ and $|K_x|_\infty$.

2. *Assume v is a $C_b^{1,2}$ function and let C_v denote the constant bounding $v, \partial_t v, \partial_x v$ and $\partial_x^2 v$. Then,*

$$\|\mathbb{E}(B_n - v)\|^2_{H^\mu_{\beta,X}} \leq K(T)\left(h_t + h_x^4 + e^{-\frac{\mu}{\sqrt{d}}a}\right),$$

where $K(T)$ depends on $d, \mu, K, \alpha_1, \alpha_2, \beta, C_v, |K_t|_\infty$ and $|K_x|_\infty$.

Combining Propositions 12.34 page 144 and Proposition 12.40 leads to the following Theorem

Theorem 12.41. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition.*

1. *Assume v is a $C^{1,2}$ function and v is bounded by C_v. Then,*

$$\begin{aligned}\mathbb{E}\left\|B_n - v\right\|^2_{H^\mu_{\beta,X}} \leq & K(T)(T\lambda(B)\delta_n \left\|v\right\|^2_{H^\mu_{\beta,X}} + h_t^2 \left\|\partial_t v\right\|^2_{H^\mu_{\beta,X}} + h_x^4 \left\|\partial_x^2 v\right\|^2_{H^\mu_{\beta,X}}) \\ & + K(T)(e^{-\frac{\mu}{\sqrt{d}}a} + h_t),\end{aligned}$$

where $K(T)$ depends on $d, \mu, K, \alpha_1, \alpha_2, \beta, C_v, |K_t|_\infty$ and $|K_x|_\infty$.

2. *Assume v is a $C_b^{1,2}$ function and let C_v denote the constant bounding $v, \partial_t v, \partial_x v$ and $\partial_x^2 v$. Then,*

$$\mathbb{E}\left\|B_n - v\right\|^2_{H^\mu_{\beta,X}} \leq K(T)\left(T\lambda(B)\delta_n + h_x^4 + h_t + e^{-\frac{\mu}{\sqrt{d}}a}\right),$$

where $K(T)$ depends on $d, \mu, K, \alpha_1, \alpha_2, \beta, C_v, |K_t|_\infty$ and $|K_x|_\infty$.

12.3.4 Conclusion

Combining (12.13) page 140, Theorem 12.27 page 141, Theorem 12.33 page 143 and Theorem 12.41 yields the following Theorem

Theorem 12.42. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition.*

1. *Assume that v is a $C^{1,2}$ function and v is bounded by C_v. Then, $\forall \epsilon \geq 0$ such that $\epsilon^2 \leq (T\lambda(B))^{-2}$*

$$\begin{aligned}\mathbb{E}\left\|\mathcal{P}_n v - v\right\|^2_{H^\mu_{\beta,X}} \leq & K(T)\left((\epsilon^2(T\lambda(B))^2 + T\lambda(B)\delta_n + h_x^4)\left\|v\right\|^2_{H^{2,\mu}_{\beta,X}} + h_t^2 \left\|\partial_t v\right\|^2_{H^\mu_{\beta,X}}\right) \\ & + K(T)\left[e^{-\frac{\mu}{\sqrt{d}}a} + h_t + e^{-\mu a}a^{d-1}h_x\right] \\ & + K(T)(T\lambda(B))^2\left(\epsilon^2 + \frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon^2 T\lambda(B)}{\delta_n}\right),\end{aligned}$$

where $K(T)$ depends on $d, \mu, K, \alpha_1, \alpha_2, \beta, C_v, |K_t|_\infty$ and $|K_x|_\infty$.

2. *Assume that v is a $C_b^{1,2}$ function and let C_v denote the constant bounding $v, \partial_t v, \partial_x v$ and $\partial_x^2 v$. Then,*

$$\mathbb{E}\left\|\mathcal{P}_n v - v\right\|^2_{H^\mu_{\beta,X}} \leq K(T)\left(T\lambda(B)\delta_n + h_t + h_x^4 + e^{-\frac{\mu}{\sqrt{d}}a} + e^{-\mu a}a^{d-1}h_x\right),$$

where $K(T)$ depends on $d, \mu, K, \alpha_1, \alpha_2, \beta, C_v, |K_t|_\infty$ and $|K_x|_\infty$.

12.4 Convergence of $\partial_x(\mathcal{P}_n v) - \partial_x v$

This part is devoted to study

$$\mathbb{E}\,\|\partial_x \mathcal{P}_n v - \partial_x v\|^2_{H^\mu_{\beta,X}} = \sum_{i=1}^{d} \mathbb{E}\,\|\partial_{x_i} \mathcal{P}_n v - \partial_{x_i} v\|^2_{H^\mu_{\beta,X}}, \tag{12.21}$$

where

$$\begin{aligned}\partial_{x_i}(\mathcal{P}_n v)(s,y) =& \partial_{x_i} r_n(s,y)\frac{g(2T\lambda(B)f_n(s,y))}{f_n(s,y)} + 2T\lambda(B)r_n(s,y)\frac{g'(2T\lambda(B)f_n(s,y))}{f_n(s,y)}\partial_{x_i} f_n(s,y)\\ &- r_n(s,y)\frac{g(2T\lambda(B)f_n(s,y))}{f_n^2(s,y)}\partial_{x_i} f_n(s,y)\end{aligned}$$

and

- $r_n(s,y) = \frac{1}{nh_t h_x^d}\sum_{i=1}^n K_t(\frac{s-T_i}{h_t})K_x(\frac{y-X_i}{h_x})v(T_i,X_i)$,
- $f_n(s,y) = \frac{1}{nh_t h_x^d}\sum_{i=1}^n K_t(\frac{s-T_i}{h_t})K_x(\frac{y-X_i}{h_x})$.

The study of $\|\partial_{x_i}(\mathcal{P}_n v) - \partial_{x_i} v\|^2_{H^\mu_{\beta,X}}$ will be done in three steps. Introducing $\partial_{x_i} A_n$ and $\partial_{x_i} B_n$ in $\|\partial_{x_i} v - \partial_{x_i}(\mathcal{P}_n v)\|^2_{H^\mu_{\beta,X}}$ yields

$$\begin{aligned}\|\partial_{x_i}(\mathcal{P}_n v) - \partial_{x_i} v\|^2_{H^\mu_{\beta,X}} \leq& 3\,\|\partial_{x_i}(\mathcal{P}_n v) - \partial_{x_i} A_n\|^2_{H^\mu_{\beta,X}}\\ &+ 3\,\|\partial_{x_i} A_n - \partial_{x_i} B_n\|^2_{H^\mu_{\beta,X}} + 3\,\|\partial_{x_i} B_n - \partial_{x_i} v\|^2_{H^\mu_{\beta,X}}.\end{aligned} \tag{12.22}$$

The scheme of this Section is the following: in Section 12.4.1, we study $\mathbb{E}\,\|\partial_{x_i} A_n - \partial_{x_i} B_n\|^2_{H^\mu_{\beta,X}}$. In Section 12.4.2, we deal with $\mathbb{E}\,\|\partial_{x_i}(\mathcal{P}_n v) - \partial_{x_i} A_n\|^2_{H^\mu_{\beta,X}}$, and in Section 12.4.3 we consider $\mathbb{E}\,\|\partial_{x_i} B_n - \partial_{x_i} v\|^2_{H^\mu_{\beta,X}}$. Section 12.4.4 ends the study of $\mathbb{E}\,\|\partial_x(\mathcal{P}_n v) - \partial_x v\|^2_{H^\mu_{\beta,X}}$.

12.4.1 Study of $\mathbb{E}\,\|\partial_x A_n - \partial_x B_n\|^2_{H^\mu_{\beta,X}}$

Let us recall the definitions of $A_n(s,y)$ and $B_n(s,y)$

$$A_n(s,y) = T\lambda(B)r_n(s,y)\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in B}, \quad B_n(s,y) = T\lambda(B)r_n(s,y).$$

Remark 12.43. By misnomer, $\partial_{x_i} A_n(s,y)$ denotes $T\lambda(B)\partial_{x_i} r_n(s,y)\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in B}$, and $\partial_{x_i} B_n(s,y)$ denotes $T\lambda(B)\partial_{x_i} r_n(s,y)$.

$$\begin{aligned}\partial_{x_i} A_n(s,y) - \partial_{x_i} B_n(s,y) &= T\lambda(B)\partial_{x_i} r_n(s,y)(\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in B} - 1)\\ &= -T\lambda(B)\partial_{x_i} r_n(s,y)\mathbf{1}_{\{s\notin[0,T]\}\cup\{y\notin B\}}.\end{aligned}$$

From Remark 12.23, page 139, we write

$$\mathbb{E}\,\|\partial_{x_i} A_n - \partial_{x_i} B_n\|^2_{H^\mu_{\beta,X}} = \int_0^T ds e^{\beta s}\int_{\mathbb{R}^d} dy \nu^0(s,y)\mathbb{E}|\partial_{x_i} A_n(s,y) - \partial_{x_i} B_n(s,y)|^2.$$

Theorem 12.44. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume v is bounded by C_v. For all $i \in \{1,\cdots,d\}$,*

$$\mathbb{E}\,\|\partial_{x_i} A_n - \partial_{x_i} B_n\|^2_{H^\mu_{\beta,X}} \leq K(T)\frac{e^{-\mu a} a^{d-1}}{h_x}.$$

$K(T)$ is a function non decreasing in T and depending on $d, \beta, |K_t|_\infty, |K_x|_\infty, |\partial_{x_i} K_x|_\infty, C_v$ and on K, α_2.

Proof of Theorem 12.44. Using the above expression of $\partial_{x_i} A_n(s,y) - \partial_{x_i} B_n(s,y)$, we get

$$\mathbb{E}|\partial_{x_i} A_n(s,y) - \partial_{x_i} B_n(s,y)|^2 = (T\lambda(B))^2 \mathbb{E}|\partial_{x_i} r_n(s,y)|^2 \mathbf{1}_{\{s\notin[0,T]\}\cup\{y\notin B\}}.$$

If v is bounded, using Proposition 12.21, page 137, yields $\mathbb{E}|\partial_{x_i} A_n(s,y) - \partial_{x_i} B_n(s,y)|^2 \leq 2^{d+1} C_v^2 |K_t|^2_\infty |\partial_{x_i} K_x|^2_\infty \frac{1}{h_x^2} \mathbf{1}_{\{s\notin[0,T]\}\cup\{y\notin B\}} \mathbf{1}_{y\in B_\infty(0,a+h_x)}$. Moreover, $\mathbf{1}_{\{s\notin[0,T]\}\cup\{y\notin B\}} \mathbf{1}_{y\in B_\infty(0,a+h_x)}$ is bounded by $\mathbf{1}_{y\in B_\infty(0,a+h_x)\setminus B}$ (see Section 12.3.1 page 140 for a detailed proof). (12.14) page 141 ends the proof. □

12.4.2 Study of $\mathbb{E}\,\|\partial_x(\mathcal{P}_n v) - \partial_x A_n\|^2_{H^\mu_{\beta,X}}$

By using the definition of $\mathcal{P}_n v(s,y)$ and $A_n(s,y)$, we write

$$\partial_{x_i}\mathcal{P}_n v(s,y) - \partial_{x_i} A_n(s,y) = \partial_{x_i} r_n(s,y)\left[\frac{1}{f_n(s,y)} g(2T\lambda(B) f_n(s,y)) - T\lambda(B)\mathbf{1}_{s\in[0,T]}\mathbf{1}_{y\in B}\right] \tag{12.23}$$

$$+\, 2T\lambda(B) r_n(s,y)\frac{g'(2T\lambda(B) f_n(s,y))}{f_n(s,y)}\partial_{x_i} f_n(s,y) - r_n(s,y)\frac{g(2T\lambda(B) f_n(s,y))}{f_n^2(s,y)}\partial_{x_i} f_n(s,y).$$

Before integrating w.r.t. $e^{\beta s}\nu^0(s,y)dsdy$, we study $\partial_{x_i}\mathcal{P}_n v(s,y) - \partial_{x_i} A_n(s,y)$ w.r.t. the value of y in the Lemmas 12.45, 12.46, 12.47 and 12.48. Then, we bound $\mathbb{E}|\partial_{x_i}\mathcal{P}_n v(s,y) - \partial_{x_i} A_n(s,y)|^2$ in Proposition 12.49. To conclude, we bound $\mathbb{E}\,\|\partial_{x_i}\mathcal{P}_n - \partial_{x_i} A_n\|^2_{H^\mu_{\beta,X}}$ in Theorem 12.50.

Lemma 12.45. *On the set $\{y \notin B_\infty(0,a+h_x)\}$, $\forall i \in \{1,\cdots,d\}$, $\partial_{x_i}\mathcal{P}_n v(s,y) - \partial_{x_i} A_n(s,y) = 0$.*

Proof of Lemma 12.45. If $y \notin B_\infty(0,a+h_x)$, the indicator in (12.23) is null. Moreover, there exists at least one $j \in \{1,\cdots,d\}$ such that $d(y_j, B) \geq h_x$. Thus, $\forall z \in B$, $K_x^j(\frac{y_j - z}{h_x}) = 0$. Since $\forall i = 1,\cdots,n$, $X_i \in B$, we get $f_n(s,y) = 0$. Looking at the definition of g in Proposition 11.6 (see page 121) yields $\frac{g(2T\lambda(B)f_n(s,y))}{f_n(s,y)} = r_n(s,y)\frac{g'(2T\lambda(B)f_n(s,y))}{f_n(s,y)} = r_n(s,y)\frac{g(2T\lambda(B)f_n(s,y))}{f_n^2(s,y)} = 0$. Hence, the result follows. □

Lemma 12.46. *On the set $\{y \in B_\infty(0,a+h_x)\setminus B\}$, if v is bounded by C_v, for all $i \in \{1,\cdots,d\}$,*

$$\mathbb{E}|\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)|^2 \leq \frac{K(T)}{h_x^2},$$

where $K(T)$ depends on $d, |K_t|_\infty, |\partial_{x_i} K_x|_\infty$ and C_v.

Proof of Lemma 12.46. Let us introduce

$$\overline{f}^i_n(s,y) = \frac{1}{nh_t h_x^{d+1}} \sum_{i=1}^n K_t\left(\frac{s-T_i}{h_t}\right) |\partial_{x_i} K_x| \left(\frac{y-X_i}{h_x}\right).$$

If $y \in B_\infty(0, a+h_x) \setminus B$, the indicator in (12.23) is null. Since $\frac{g(x)}{x}$ is bounded by $\frac{4\sqrt{2}}{3\sqrt{3}} \leq 2$, $\frac{|\partial_{x_i} r_n(s,y)|}{f_n(s,y)} g(2T\lambda(B) f_n(s,y)) \leq 4T\lambda(B) C_v \overline{f}^i_n(s,y)$, and as $|r_n(s,y)| \leq C_v f_n(s,y)$, $|r_n(s,y)| \frac{g(2T\lambda(B)f_n(s,y))}{f_n^2(s,y)} |\partial_{x_i} f_n(s,y)| \leq 4T\lambda(B) C_v |\partial_{x_i} f_n(s,y)|$. Since g' is bounded by $\frac{8}{3\sqrt{3}} \leq 2$, the term $2T\lambda(B)|r_n(s,y)| \frac{g'(2T\lambda(B)f_n(s,y))}{f_n(s,y)} |\partial_{x_i} f_n(s,y)|$ is bounded by $4T\lambda(B)C_v|\partial_{x_i} f_n(s,y)|$. To conclude, we use $|\partial_{x_i} f_n(s,y)| \leq \overline{f}^i_n(s,y)$ and Lemma 12.8 page 127, which states $\mathbb{E}(\overline{f}^i_n(s,y))^2 \leq \frac{C}{h_x^2 (T\lambda(B))^2}$, where C depends on d, $|K_t|_\infty$ and $|\partial_{x_i} K_x|_\infty$. □

On the set $\{y \in B\}$, we bound $\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)$ of two manners (see Lemmas 12.47 and 12.48), which enables us to state Proposition 12.49.

Lemma 12.47. *On the set $\{y \in B\}$, if v is bounded by C_v, for all $i \in \{1, \cdots, d\}$,*

$$\begin{aligned} |\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)| \leq & 8(T\lambda(B))^2 |\partial_{x_i} r_n(s,y)| |f_n(s,y) - \overline{f}(s,y)| \\ & + 8T\lambda(B) C_v |\partial_{x_i} f_n(s,y)|. \end{aligned}$$

Proof of Lemma 12.47. If $y \in B$ and $s \in [0,T]$, the first term of $\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)$ is $\partial_{x_i} r_n(s,y)\left[\frac{1}{f_n(s,y)} g(2T\lambda(B)f_n(s,y)) - T\lambda(B)\right]$. By using the proof of Lemma 12.31, page 142, we get that

$$\left|\partial_{x_i} r_n(s,y)\left[\frac{1}{f_n(s,y)} g(2T\lambda(B)f_n(s,y)) - T\lambda(B)\right]\right| \leq 8(T\lambda(B))^2 |\partial_{x_i} r_n(s,y)| |f_n(s,y) - \overline{f}(s,y)|. \tag{12.24}$$

Since g' and $\frac{g(x)}{x}$ are bounded by 2 and $|r_n(s,y)| \leq C_v f_n(s,y)$, the last two terms of (12.23) are bounded by $4T\lambda(B)|\partial_{x_i} f_n(s,y)|$. □

Lemma 12.48. *On the set $\{y \in B\}$, for all $i \in \{1, \cdots, d\}$,*

$$\begin{aligned} |\partial_{x_i} \mathcal{P}_n v(s,y) - \partial_{x_i} A_n(s,y)| \leq & 8(T\lambda(B))^2 |\partial_{x_i} r_n(s,y)| |f_n(s,y) - \overline{f}(s,y)| \\ & + 24(T\lambda(B))^2 |r_n(s,y)| |\partial_{x_i} f_n(s,y)|. \end{aligned}$$

Proof of Lemma 12.48. We use (12.24) to bound the first term of (12.23). Since $\frac{g'(x)}{x}$ is bounded by 4 and $\frac{g(x)}{x^2}$ is bounded by 2, the last two terms of (12.23) are respectively bounded by $16(T\lambda(B))^2|r_n(s,y)||\partial_{x_i} f_n(s,y)|$ and $8(T\lambda(B))^2|r_n(s,y)||\partial_{x_i} f_n(s,y)|$. □

Proposition 12.49. *Let ϵ_0 be such that $0 \leq \epsilon_0 \leq (T\lambda(B))^{-1}$. On the set $\{y \in B\}$ and for v bounded by C_v, we get for all $i \in \{1, \cdots, d\}$,*

$$\begin{aligned} \mathbb{E}|\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)|^2 \leq & C_0 \epsilon_0^2 (T\lambda(B))^4 \mathbb{E}(\partial_{x_i} r_n(s,y))^2 + C_0 \frac{\epsilon_0^2}{h_x^2} (T\lambda(B))^4 \mathbb{E}(r_n(s,y))^2 \\ & + C(T\lambda(B))^2 \left(\frac{1}{h_x^2} |\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|^2 + (\mathbb{E}(\partial_{x_i} f_n(s,y)))^2\right) \\ & + C\frac{(T\lambda(B))^2}{h_x^2}\left(\epsilon_0^2 + \frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon_0^2 T\lambda(B)}{\delta_n}\right) + C\frac{\sqrt{T\lambda(B)\delta_n}}{h_x^2}\exp\left(-\frac{c}{T\lambda(B)\delta_n}\right), \end{aligned}$$

where C depends on $C_v, d, |K_t|_\infty$, $|K_x|_\infty$ and $|\partial_{x_i} K_x|_\infty$ and c depends on $d, |K_t|_\infty$, $|K_x|_\infty$ and $|\partial_{x_i} K_x|_\infty$, and C_0 is a strictly positive constant.

Proof of Proposition 12.49. We introduce $\mathbf{1}_{|\partial_{x_i} f_n(s,y) - \partial_{x_i}\mathbb{E}(f_n(s,y))| > \epsilon_1}$ and $\mathbf{1}_{|\partial_{x_i} f_n(s,y) - \partial_{x_i}\mathbb{E}(f_n(s,y))| \leq \epsilon_1}$, and we split $\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)$ w.r.t. these indicators. On the set $\{|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon_1\}$, we use Lemma 12.47 to bound $|\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)|$. On the other set, we use Lemma 12.48. We get

$$\begin{aligned}|\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)| \leq & 8(T\lambda(B))^2 |\partial_{x_i} r_n(s,y)| |f_n(s,y) - \overline{f}(s,y)| \\ & + 24(T\lambda(B))^2 |r_n(s,y)| |\partial_{x_i} f_n(s,y)| \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \partial_{x_i}\mathbb{E}(f_n(s,y))| \leq \epsilon_1} \\ & + 8T\lambda(B) C_v |\partial_{x_i} f_n(s,y)| \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon_1}.\end{aligned}$$

Using this formula, we study $\mathbb{E}|\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i} A_n(s,y)|^2$.
First, let us deal with $|\partial_{x_i} r_n(s,y)| |f_n(s,y) - \overline{f}(s,y)|$. We introduce $\mathbf{1}_{|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \epsilon_0}$ and $\mathbf{1}_{|f_n(s,y) - \mathbb{E}(f_n(s,y))| \leq \epsilon_0}$ in $|\partial_{x_i} r_n(s,y)| |f_n(s,y) - \overline{f}(s,y)|$ and we apply $|f_n(s,y) - \overline{f}(s,y)| \leq |f_n(s,y) - \mathbb{E}(f_n(s,y))| + |\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|$ and $|\partial_{x_i} r_n(s,y)| \leq C_v \overline{f}^i_n(s,y)$, where $\overline{f}^i_n$ has been defined in the proof of Lemma 12.46. We get

$$\begin{aligned}\mathbb{E}|\partial_{x_i} r_n(s,y)|^2 |f_n(s,y) - \overline{f}(s,y)|^2 \leq & 2\epsilon_0^2 \mathbb{E}|\partial_{x_i} r_n(s,y)|^2 \\ & + 2C_v^2 |\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|^2 \mathbb{E}(\overline{f}^i_n(s,y))^2 \\ & + 2C_v^2 \mathbb{E}[|f_n(s,y) - \mathbb{E}(f_n(s,y))|^2 \mathbf{1}_{|f_n(s,y) - \mathbb{E}(f_n(s,y))| > \epsilon_0} (\overline{f}^i_n(s,y))^2].\end{aligned}$$

Lemma 12.8 page 127 yields $\mathbb{E}(\overline{f}^i_n(s,y))^2 \leq \frac{C}{h_x^2 (T\lambda(B))^2}$, where C depends on d, $|K_t|_\infty$ and $|\partial_{x_i} K_x|_\infty$. We use Cauchy-Schwarz inequality,

$$(\overline{f}^i_n(s,y))^2 \leq 2(\mathbb{E}(\overline{f}^i_n(s,y)) - \overline{f}^i_n(s,y))^2 + 2(\mathbb{E}(\overline{f}^i_n(s,y)))^2$$

and Proposition 12.16, page 132 to bound the third term.
Second, we study $(T\lambda(B))^2 |r_n(s,y)| |\partial_{x_i} f_n(s,y)| \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \partial_{x_i}\mathbb{E}(f_n(s,y))| \leq \epsilon_1}$. From $|\partial_{x_i} f_n(s,y)| \leq |\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| + |\mathbb{E}(\partial_{x_i} f_n(s,y))|$, it follows

$$\begin{aligned}(T\lambda(B))^4 \mathbb{E}[r_n^2(s,y) |\partial_{x_i} f_n(s,y)|^2 \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \partial_{x_i}\mathbb{E}(f_n(s,y))| \leq \epsilon_1}] \leq & 2\epsilon_1^2 (T\lambda(B))^4 \mathbb{E}[r_n^2(s,y)] \\ & + 2(T\lambda(B))^4 \mathbb{E}[r_n^2(s,y)] |\mathbb{E}(\partial_{x_i} f_n(s,y))|^2,\end{aligned}$$

and we use Proposition 12.20, page 136, to bound $\mathbb{E}[r_n^2(s,y)]$ in the second term of the above inequality.

Finally, we study $T\lambda(B) |\partial_{x_i} f_n(s,y)| \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon_1}$. Writing $|\partial_{x_i} f_n(s,y)| \leq |\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| + |\mathbb{E}(\partial_{x_i} f_n(s,y))|$ leads to

$$\begin{aligned}&(T\lambda(B))^2 \mathbb{E}[|\partial_{x_i} f_n(s,y)|^2 \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon_1}] \leq \\ &\quad 2(T\lambda(B))^2 \mathbb{E}[|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))|^2 \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon_1}] \\ &\quad + 2(T\lambda(B))^2 |\mathbb{E}(\partial_{x_i} f_n(s,y))|^2 \mathbb{P}(|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon_1).\end{aligned}$$

The second term is bounded by $2(T\lambda(B))^2 |\mathbb{E}(\partial_{x_i} f_n(s,y))|^2$. Proposition 12.18, page 135 enables us to bound the first term. We get

$$\begin{aligned}&\mathbb{E}\left(|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))|^2 \ \mathbf{1}_{|\partial_{x_i} f_n(s,y) - \mathbb{E}(\partial_{x_i} f_n(s,y))| > \epsilon_1}\right) \\ &\qquad \leq C \left(\epsilon_1^2 + \frac{\delta_n}{T\lambda(B) h_x^2}\right) \exp\left(-\frac{c\epsilon_1^2 h_x^2 T\lambda(B)}{\delta_n}\right).\end{aligned}$$

By choosing $\epsilon_1 = \frac{\epsilon_0}{h_x}$, the result follows. $\square$

Lemmas 12.45, 12.46 and Proposition 12.49 enables us to state the following Theorem.

Theorem 12.50. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. If v is bounded by C_v, $\forall\epsilon_0 \geq 0$ satisfying $\epsilon_0^2 \leq (T\lambda(B))^{-2}$, for all $i \in \{1,\cdots,d\}$,*

$$\begin{aligned}\mathbb{E}\|\partial_{x_i}(\mathcal{P}_n v) - \partial_{x_i}A_n\|^2_{H^\mu_{\beta,X}} \leq{}& \frac{K_1(T)}{h_x^2}(h_t + e^{-\mu a}a^{d-1}h_x)\\ &+ K_0(T)(T\lambda(B))^2\left(\frac{\epsilon_0^2}{h_x^2}\|v\|^2_{H^\mu_{\beta,X}} + \epsilon_0^2\|\partial_x v\|^2_{H^\mu_{\beta,X}}\right)\\ &+ K_2(T)\left[\frac{(T\lambda(B))^2}{h_x^2}\left(\epsilon_0^2 + \frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon_0^2 T\lambda(B)}{\delta_n}\right) + \frac{\sqrt{T\lambda(B)\delta_n}}{h_x^2}\exp\left(-\frac{c}{T\lambda(B)\delta_n}\right)\right].\end{aligned}$$

$K_0(T)$ depends on $|K_t|_\infty, |K_x|_\infty, \beta, \alpha_2, d, \mu$ and K. $K_1(T)$ depends on $C_v, d, |K_x^i|_\infty, \beta, K$ and α_2. $K_2(T)$ depends on $C_v, d, |K_t|_\infty, |K_x|_\infty, K, \mu, \beta$ and α_2.
In particular, for $\epsilon_0 = 0$ we get

$$\begin{aligned}\mathbb{E}\|\partial_{x_i}(\mathcal{P}_n v) - \partial_{x_i}A_n\|^2_{H^\mu_{\beta,X}} \leq{}& \frac{K_1(T)}{h_x^2}(h_t + e^{-\mu a}a^{d-1}h_x)\\ &+ K_2(T)\left(\frac{1}{h_x^2}T\lambda(B)\delta_n + \frac{\sqrt{T\lambda(B)\delta_n}}{h_x^2}\exp\left(-\frac{c}{T\lambda(B)\delta_n}\right)\right).\end{aligned}$$

Proof of Theorem 12.50.

$$\mathbb{E}\|\partial_{x_i}(\mathcal{P}_n v) - \partial_{x_i}A_n\|^2_{H^\mu_{\beta,X}} = \int_0^T ds e^{\beta s}\int_{\mathbb{R}^d} dy \nu^0(s,y)\mathbb{E}|\partial_{x_i}(\mathcal{P}_n v)(s,y) - \partial_{x_i}A_n(s,y)|^2.$$

If v is bounded by C_v, we combine Lemmas 12.45, page 151, Lemme 12.46, page 151 and Proposition 12.49, page 152 to get

$$\begin{aligned}&\mathbb{E}\|\partial_{x_i}(\mathcal{P}_n v) - \partial_{x_i}A_n\|^2_{H^\mu_{\beta,X}} \leq \frac{C}{h_x^2}\int_0^T ds e^{\beta s}\int_{B_\infty(0,a+h_x)\setminus B} dy\nu^0(s,y)\\ &+ C_0(T\lambda(B))^4\int_0^T ds e^{\beta s}\int_B dy\nu^0(s,y)(\epsilon_0^2\mathbb{E}(\partial_{x_i}r_n(s,y))^2 + \frac{\epsilon_0^2}{h_x^2}\mathbb{E}(r_n(s,y))^2)\\ &+ C(T\lambda(B))^2\int_0^T ds e^{\beta s}\int_B dy\nu^0(s,y)(\frac{1}{h_x^2}|\mathbb{E}(f_n(s,y)) - \overline{f}(s,y)|^2 + (\mathbb{E}(\partial_{x_i}f_n(s,y)))^2)\\ &+ \left[C\frac{(T\lambda(B))^2}{h_x^2}\left(\epsilon_0^2 + \frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon_0^2 T\lambda(B)}{\delta_n}\right) + C\frac{\sqrt{T\lambda(B)\delta_n}}{h_x^2}\exp\left(-\frac{c}{T\lambda(B)\delta_n}\right)\right]\\ &\qquad\times\int_0^T ds e^{\beta s}\int_B dy\nu^0(s,y),\end{aligned}$$

where C depends on $C_v, d, |K_t|_\infty$ and $|\partial_{x_i}K_x|_\infty$ and c depends on $d, |K_t|_\infty$ and $|K_x|_\infty$. We use (12.14), page 141 to bound the first term, and (12.16), page 143 to bound the last one. Proposition 12.15, page 131 and Proposition 12.19, page 135 bound the third term. It remains to bound the second term. To do so, we use the last result of Proposition

12.20, page 136 to bound $\mathbb{E}(r_n^2(s,y))$, and the second part of Proposition 12.21, page 137 to bound $\mathbb{E}(\partial_{x_i} r_n(s,y))^2$. We get

$$(T\lambda(B))^4 \int_0^T dse^{\beta s} \int_B dy\nu^0(s,y) \left(\epsilon_0^2 \mathbb{E}(\partial_{x_i} r_n(s,y))^2 + \frac{\epsilon_0^2}{h_x^2} \mathbb{E}(r_n(s,y))^2 \right) \leq$$
$$\frac{\epsilon_0^2 (T\lambda(B))^2}{h_t h_x^{d+2}} \int_0^T dse^{\beta s} \int_B dy\nu^0(s,y) \int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right) \int_B dz K_x^2 \left(\frac{y-z}{h_x}\right) v^2(r,z)$$
$$+ 4\epsilon_0^2 \frac{C_v^2 |K_x^i|_\infty^2 (T\lambda(B))^2}{h_x^2} \int_0^T dse^{\beta s} \int_{B \setminus B_\infty(0,a-h_x)} dy\nu^0(s,y)$$
$$+ \epsilon_0^2 \frac{2^{d+1}(T\lambda(B))^2}{h_t h_x^d} \int_0^T dse^{\beta s} \int_B dy\nu^0(s,y) \int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right) \int_B dz K_x^2 \left(\frac{y-z}{h_x}\right) |\partial_{x_i} v|^2(r,z)$$
$$+ \epsilon_0^2 \frac{(T\lambda(B))^3}{n h_t h_x^{d+2}} \int_0^T dse^{\beta s} \int_B dy\nu^0(s,y) \int_0^T dr K_t^2\left(\frac{s-r}{h_t}\right) \int_B dz (\partial_x K_x)^2 \left(\frac{y-z}{h_x}\right) v^2(r,z).$$

To conclude, we apply Lemma 12.10, page 128 with $g_t = K_t^2$, $g_x = K_x^2$ and $f \equiv v$ and (12.5), page 132. □

12.4.3 Study of $\mathbb{E}\left\|\partial_x B_n - \partial_x v\right\|^2_{H^\mu_{\beta,X}}$

By using the definition of B_n we get

$$\partial_{x_i} B_n(s,y) - \partial_{x_i} v(s,y) = T\lambda(B)\partial_{x_i} r_n(s,y) - \partial_{x_i} v(s,y).$$

Proposition 12.24, page 139 gives us $\mathbb{E}\left\|\partial_{x_i} B_n - \partial_{x_i} v\right\|^2_{H^\mu_{\beta,X}} = \left\|\mathbb{E}(\partial_{x_i} B_n - \partial_{x_i} v)\right\|^2_{H^\mu_{\beta,X}} + \left\|\mathrm{Std}(\partial_{x_i} B_n - \partial_{x_i} v)\right\|^2_{H^\mu_{\beta,X}}$. The scheme of the Section is the following. We bound $\left\|\mathrm{Std}(\partial_{x_i} B_n - \partial_{x_i} v)\right\|^2_{H^\mu_{\beta,X}}$ in Proposition 12.51. Then, we study $\left\|\mathbb{E}(\partial_{x_i} B_n - \partial_{x_i} v)\right\|^2_{H^\mu_{\beta,X}}$. To do so, Lemma 12.52 splits $\mathbb{E}(\partial_{x_i} B_n(s,y)) - \partial_{x_i} v(s,y)$ in four terms. From Lemma 12.53 to Lemma 12.56, we study these four terms. Using these results we bound $\left\|\mathbb{E}(\partial_{x_i} B_n - \partial_{x_i} v)\right\|^2_{H^\mu_{\beta,X}}$ in Proposition 12.40. Theorem 12.41 ends the Section by bounding $\mathbb{E}\left\|\partial_{x_i} B_n - \partial_{x_i} v\right\|^2_{H^\mu_{\beta,X}}$.

Proposition 12.51. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition. For all $i \in \{1, \cdots, d\}$,*

$$\left\|Std(\partial_{x_i} B_n - \partial_{x_i} v)\right\|^2_{H^\mu_{\beta,X}} \leq K(T) \frac{T\lambda(B)\delta_n}{h_x^2} \left\|v\right\|^2_{H^\mu_{\beta,X}},$$

where $K(T)$ depends on $\mu, d, \beta, |K_t|_\infty, |\partial_{x_i} K_x|_\infty, K, \alpha_1$ and α_2. If v is bounded by C_v, we get $\left\|Std(B_n - v)\right\|^2_{H^\mu_{\beta,X}} \leq K'(T) \frac{T\lambda(B)\delta_n}{h_x^2}$, where $K'(T)$ depends on $C_v, |K_t|_\infty, |\partial_{x_i} K_x|_\infty, \mu, d, \beta, K, \alpha_1$ and α_2.

Proof of Proposition 12.51.

$$\begin{aligned}\left\|\mathrm{Std}(\partial_{x_i} B_n - \partial_{x_i} v)\right\|^2_{H^\mu_{\beta,X}} &= \int_0^T dse^{\beta s} \int_{\mathbb{R}^d} dy\nu^0(s,y) \mathrm{Var}(\partial_{x_i} B_n(s,y) - \partial_{x_i} v(s,y)), \\ &= (T\lambda(B))^2 \int_0^T dse^{\beta s} \int_{\mathbb{R}^d} dy\nu^0(s,y) \mathrm{Var}(\partial_{x_i} r_n(s,y)).\end{aligned}$$

Using the second part of Proposition 12.21, page 137 yields

$$\|\mathrm{Std}(\partial_{x_i}B_n - \partial_{x_i}v)\|^2_{H^\mu_{\beta,X}} \leq$$
$$\frac{T\lambda(B)}{nh_th_x^{d+2}}\int_0^T dse^{\beta s}\int_{\mathbb{R}^d} dy\nu^0(s,y)\int_0^T drK_t^2\left(\frac{s-r}{h_t}\right)\int_{\mathbb{R}^d} dz(\partial_{x_i}K_x)^2\left(\frac{y-z}{h_x}\right)v^2(r,z).$$

To conclude, we apply Lemma 12.10, page 128 with $g_t = K_t^2$, $g_x = (\partial_{x_i}K_x)^2$ and $f \equiv v$. If v is bounded, we combine the first part of Proposition 12.21, page 137 and (12.16), page 143 to get the result. $\square$

Second, we study $\|\mathbb{E}(\partial_{x_i}B_n - \partial_{x_i}v)\|^2_{H^\mu_{\beta,X}}$. First, let us split $\mathbb{E}(\partial_{x_i}B_n(s,y)) - \partial_{x_i}v(s,y)$.

Lemma 12.52. *For all $(s,y) \in [0,T]\times\mathbb{R}^d$, for all $i \in \{1,\cdots,d\}$,*

$$\begin{aligned}
\mathbb{E}(\partial_{x_i}B_n(s,y)) - \partial_{x_i}v(s,y) = &\\
&\frac{1}{h_th_x^d}\int_{\mathbb{R}^d} dz\int_0^T drK_x\left(\frac{y-z}{h_x}\right)K_t\left(\frac{s-r}{h_t}\right)(\partial_{x_i}v(r,z)-\partial_{x_i}v(s,z))\\
&+\frac{1}{h_th_x^d}\int_{\mathbb{R}^d} dz\int_0^T drK_x\left(\frac{y-z}{h_x}\right)K_t\left(\frac{s-r}{h_t}\right)(\partial_{x_i}v(s,z)-\partial_{x_i}v(s,y))\\
&-\frac{1}{h_th_x^{d+1}}\int_{B^c} dz\int_0^T dr\partial_{x_i}K_x\left(\frac{y-z}{h_x}\right)K_t\left(\frac{s-r}{h_t}\right)v(r,z)\\
&-\frac{1}{h_th_x^d}\int_{\mathbb{R}^d} dz\int_{[-h_t,0]\cup[T,T+h_t]} drK_x\left(\frac{y-z}{h_x}\right)K_t\left(\frac{s-r}{h_t}\right)\partial_{x_i}v(s,y)\\
&:= \tilde{B}_1^i(s,y)+\tilde{B}_2^i(s,y)+\tilde{B}_3^i(s,y)+\tilde{B}_4^i(s,y).
\end{aligned}$$

Proof of Lemma 12.52. We can write $\partial_{x_i}v(s,y) = \partial_{x_i}v(s,y)\frac{1}{h_th_x^d}\int_{\mathbb{R}^d} dzK_x\left(\frac{y-z}{h_x}\right)\int_{-h_t}^{T+h_t} drK_t\left(\frac{s-r}{h_t}\right)$. Since $\mathbb{E}(\partial_{x_i}B_n(s,y)) = \frac{1}{h_th_x^{d+1}}\int_B dz\partial_{x_i}K_x\left(\frac{y-z}{h_x}\right)\int_0^T drK_t\left(\frac{s-r}{h_t}\right)v(r,z)$, we get

$$\begin{aligned}
&\mathbb{E}(\partial_{x_i}B_n(s,y)) - \partial_{x_i}v(s,y) =\\
&\frac{1}{h_th_x^d}\int_0^T drK_t\left(\frac{s-r}{h_t}\right)\int_{\mathbb{R}^d} dz\left[\frac{1}{h_x}\partial_{x_i}K_x\left(\frac{y-z}{h_x}\right)v(r,z)-K_x\left(\frac{y-z}{h_x}\right)\partial_{x_i}v(s,y)\right]\\
&-\frac{1}{h_th_x^{d+1}}\int_{B^c} dz\int_0^T dr\partial_{x_i}K_x\left(\frac{y-z}{h_x}\right)K_t\left(\frac{s-r}{h_t}\right)v(r,z)\\
&-\frac{1}{h_th_x^d}\int_{\mathbb{R}^d} dz\int_{[-h_t,0]\cup[T,T+h_t]} drK_x\left(\frac{y-z}{h_x}\right)K_t\left(\frac{s-r}{h_t}\right)\partial_{x_i}v(s,y).
\end{aligned}$$

We integrate by parts $\int_{\mathbb{R}^d} dz_i\frac{1}{h_x}\partial_{x_i}K_x^i(\frac{y_i-z_i}{h_x})v(r,z)$, we get $[-K_x^i(\frac{y_i-z_i}{h_x})v(r,z)]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} dz_iK_x^i(\frac{y_i-z_i}{h_x})\partial_{x_i}v(r,z)$. Since K_x^i has a support in $[-1,1]$, the first term is null and the result follows. $\square$

The following four Lemmas study the norms $\|\tilde{B}_1^i(s,y)\|^2_{H^\mu_{\beta,X}}$, $\|\tilde{B}_2^i(s,y)\|^2_{H^\mu_{\beta,X}}$, $\|\tilde{B}_3^i(s,y)\|^2_{H^\mu_{\beta,X}}$ and $\|\tilde{B}_4^i(s,y)\|^2_{H^\mu_{\beta,X}}$.

Lemma 12.53. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume v satisfies $\forall t,t' \in [0,T], \forall x \in \mathbb{R}^d$, $\forall i \in \{1,\cdots,d\}$, $|\partial_{x_i}v(t,x) - \partial_{x_i}v(t',x)| \le C\sqrt{|t'-t|}$, where C is a positive constant. We get for all $i \in \{1,\cdots,d\}$, $\|\tilde{B}_1^i(s,y)\|^2_{H^\mu_{\beta,X}} \le K(T)h_t$, where $K(T)$ depends on d,μ,K,β,α_2,C and C_v.*

Proof of Lemma 12.53. We recall $\tilde{B}_1^i(s,y) = \frac{1}{h_t h_x^d}\int_{\mathbb{R}^d} dz \int_0^T dr K_x(\frac{y-z}{h_x})K_t(\frac{s-r}{h_t})(\partial_{x_i}v(r,z) - \partial_{x_i}v(s,z))$. Using the Hypothesis on v, we get $(\tilde{B}_1^i(s,y))^2 \le C^2 h_t$. We use (12.16), page 143 to bound $\|\tilde{B}_1^i(s,y)\|^2_{H^\mu_{\beta,X}}$. □

Lemma 12.54. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume v is a function C^2 in space. There exists a function $K(T)$ such that for all $i \in \{1,\cdots,d\}$,*

$$\|\tilde{B}_2^i(s,y)\|^2_{H^\mu_{\beta,X}} \le K(T)h_x^2 \left\|\partial_x^2 v\right\|^2_{H^\mu_{\beta,X}},$$

where $K(T)$ depends on $d,\mu,|K_x|_\infty,\beta,K,\alpha_1$, and α_2. If $\partial_x^2 v$ is bounded, we get $\|\tilde{B}_2^i(s,y)\|^2_{H^\mu_{\beta,X}} \le K'(T)h_x^2$, where $K'(T)$ depends on $|\partial_x^2 v|^2_\infty$ and on the same parameters as $K(T)$.

Proof of Lemma 12.54. We recall $\tilde{B}_2^i(s,y) = \frac{1}{h_t h_x^d}\int_{\mathbb{R}^d} dz \int_0^T dr K_x(\frac{y-z}{h_x})K_t(\frac{s-r}{h_t})(\partial_{x_i}v(s,z) - \partial_{x_i}v(s,y))$. We apply a Taylor expansion formula to each component of the following sum

$$\partial_{x_i}v(s,z) - \partial_{x_i}v(s,y) = \sum_{j=1}^d \partial_{x_i}v(s,\overline{z}_j) - \partial_{x_i}v(s,\overline{z}_{j-1}),$$

where $\forall j \in \{1,\cdots,d\}$, $\overline{z}_j = (z_1,z_2,\cdots,z_j,y_{j+1},\cdots,y_d)$, and $\overline{z}_0 = y$. For all $j \in \{1,\cdots,d\}$, we get

$$\partial_{x_i}v(s,\overline{z}_j) - \partial_{x_i}v(s,\overline{z}_{j-1}) = \int_{y_j}^{z_j} dl \partial^2_{x_i x_j} v(s,\overline{z}_j^l),$$

where $\overline{z}_j^l = (z_1,\cdots,z_{j-1},l,y_{j+1},\cdots,y_d)$. Plugging this result in the definition of $\tilde{B}_2^i(s,y)$ leads to

$$\tilde{B}_2^i(s,y) = \frac{1}{h_t h_x^d}\sum_{j=1}^d \int_{\mathbb{R}^d} dz K_x\left(\frac{y-z}{h_x}\right)\int_0^T dr K_t\left(\frac{s-r}{h_t}\right)\int_{y_j}^{z_j} dl \partial^2_{x_i x_j} v(s,\overline{z}_j^l),$$

Moreover, $\int_0^T dr K_t(\frac{s-r}{h_t}) \le h_t$ and $\int_{\mathbb{R}^d} dz K_x(\frac{y-z}{h_x})\int_{y_j}^{z_j} dl|\partial^2_{x_i x_j}v(s,\overline{z}_j^l)| \le (2h_x)^{d-j+1}|K_x|_\infty \int_{y_1-h_x}^{y_1+h_x} dz_1 \cdots \int_{y_{j-1}-h_x}^{y_{j-1}+h_x} dz_{j-1} \int_{y_j-h_x}^{y_j+h_x} dl|\partial^2_{x_i x_j}v(s,\overline{z}_j^l)|$. Hence,

$$\begin{aligned}
\tilde{B}_2^i(s,y) &\le |K_x|_\infty \sum_{j=1}^d \frac{2^{d-j+1}}{h_x^{j-1}} \int_{y_1-h_x}^{y_1+h_x} dz_1 \cdots \int_{y_{j-1}-h_x}^{y_{j-1}+h_x} dz_{j-1} \int_{y_j-h_x}^{y_j+h_x} dl|\partial^2_{x_i x_j}v|(s,\overline{z}_j^l) \\
&\le 2^d |K_x|_\infty \sum_{j=1}^d \tilde{B}_2^{i,j}(s,y),
\end{aligned}$$

where $\tilde{B}_2^{i,j}(s,y) := \frac{1}{h_x^{j-1}} \int_{y_1-h_x}^{y_1+h_x} dz_1 \cdots \int_{y_{j-1}-h_x}^{y_{j-1}+h_x} dz_{j-1} \int_{y_j-h_x}^{y_j+h_x} dl |\partial^2_{x_i x_j} v|(s,\overline{z}_j^l)$. We end the proof as we did for Lemma 12.37, page 146. $\square$

Lemma 12.55. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition. If v and $\partial_x v$ are bounded by C_v, we get $\|\tilde{B}_3^i(s,y)\|^2_{H^\mu_{\beta,X}} \leq K(T) \frac{1}{h_x} e^{-\frac{\mu}{\sqrt{d}} a}$, where $K(T)$ depends on $d, \mu, K, \beta, \alpha_2, C_v$ and $|K_x|_\infty$.*

Proof of Lemma 12.55. We recall $\tilde{B}_3^i(s,y) = -\frac{1}{h_t h_x^{d+1}} \int_{\overline{B}} dz \int_0^T dr \partial_{x_i} K_x(\frac{y-z}{h_x}) K_t(\frac{s-r}{h_t}) v(r,z)$. Moreover,

$$\begin{aligned} \mathbf{1}_{z \in B^c} &= \sum_{j=1}^d \mathbf{1}_{z_1 \in [-a,a]} \cdots \mathbf{1}_{z_{j-1} \in [-a,a]} \mathbf{1}_{z_j \in \mathbb{R} \setminus [-a,a]} \mathbf{1}_{z_{j+1} \in \mathbb{R}} \cdots \mathbf{1}_{z_d \in \mathbb{R}} := \sum_{j=1}^d \mathbf{1}_{A_j}, \\ &= \sum_{j=1}^{i-1} \mathbf{1}_{A_j} + \mathbf{1}_{A_i} + \sum_{j=i+1}^d \mathbf{1}_{A_j}. \end{aligned}$$

We use the above decomposition of $\mathbf{1}_{z \in \overline{B}}$ to compute $\tilde{B}_3^i(s,y)$.
First, let us compute $\tilde{B}_{31}^i(s,y) := -\frac{1}{h_t h_x^{d+1}} \sum_{j=1}^{i-1} \int_{A_j} dz \int_0^T dr \partial_{x_i} K_x(\frac{y-z}{h_x}) K_t(\frac{s-r}{h_t}) v(r,z)$. We integrate by parts $\int_{\mathbb{R}} dz_i \partial_{x_i} K_x^i \left(\frac{y_i - z_i}{h_x}\right) v(r,z)$ and we get $h_x \int_{\mathbb{R}} dz_i \partial_{x_i} v(r,z) K_x^i \left(\frac{y_i-z_i}{h_x}\right)$. Then,

$$\tilde{B}_{31}^i(s,y)^2 \leq \frac{C_v^2}{h_x^d} \sum_{j=1}^{i-1} \int_{A_j} dz K_x^2 \left(\frac{y-z}{h_x}\right) \leq C_v^2 \frac{|K_x|_\infty^2}{h_x^d} \sum_{j=1}^{i-1} \int_{A_j} dz \mathbf{1}_{|y_1 - z_1| \leq h_x} \cdots \mathbf{1}_{|y_d - z_d| \leq h_x}.$$

Since $\int_{A_j} dz \mathbf{1}_{|y_1-z_1|\leq h_x} \cdots \mathbf{1}_{|y_d - z_d| \leq h_x} \leq (2h_x)^d \mathbf{1}_{y \notin B_\infty(0,a-h_x)}$, and using the proof of Lemma 12.38, page 147 gives $\|\tilde{B}_{31}^i(s,y)\|^2_{H^\mu_{\beta,X}} \leq K(T) e^{-\frac{\mu}{\sqrt{d}}(a-h_x)}$.
Second, we compute $\tilde{B}_{32}^i(s,y) := -\frac{1}{h_t h_x^{d+1}} \int_{A_i} dz \int_0^T dr \partial_{x_i} K_x(\frac{y-z}{h_x}) K_t(\frac{s-r}{h_t}) v(r,z)$. We integrate by parts $\int_{\mathbb{R}\setminus[-a,a]} dz_i \partial_{x_i} K_x^i \left(\frac{y_i-z_i}{h_x}\right) v(r,z)$.

$$\begin{aligned} \frac{1}{h_x} \int_{\mathbb{R}\setminus[-a,a]} dz_i \partial_{x_i} K_x^i \left(\frac{y_i - z_i}{h_x}\right) v(r,z) = &- K_x^i \left(\frac{y_i + a}{h_x}\right) v(r, z_a^i) + K_x^i \left(\frac{y_i - a}{h_x}\right) v(r, z_{-a}^i) \\ &+ \int_{\mathbb{R}\setminus[-a,a]} dz_i \partial_{x_i} v(r,z) K_x^i \left(\frac{y_i - z_i}{h_x}\right), \end{aligned}$$

where z_y^i denotes the vector $(z_1, \cdots, z_{i-1}, y, z_{i+1}, \cdots, z_d)$. Then, $\tilde{B}_{32}^i(s,y)^2 \leq C_v^2 |K_x|_\infty^2 \frac{1}{h_x^2} (\mathbf{1}_{|y_i + a| \leq hx} + \mathbf{1}_{|y_i - a| \leq h_x}) + C_v^2 \frac{|K_x|_\infty^2}{h_x^d} \int_{A_i} dz \mathbf{1}_{|y_1 - z_1| \leq h_x} \cdots \mathbf{1}_{|y_d - z_d| \leq h_x}$, and then $\tilde{B}_{32}^i(s,y)^2 \leq C_v^2 |K_x|_\infty^2 \frac{1}{h_x^2} (\mathbf{1}_{|y_i + a| \leq h_x} + \mathbf{1}_{|y_i - a| \leq h_x}) + C_v^2 |K_x|_\infty^2 \mathbf{1}_{y \notin B_\infty(0,a-h_x)}$. We obtain $\|\tilde{B}_{32}^i(s,y)\|^2_{H^\mu_{\beta,X}} \leq K(T)(1 + \frac{1}{h_x}) e^{-\frac{\mu}{\sqrt{d}}(a-h_x)}$.
Finally, we compute $\tilde{B}_{33}^i(s,y) := -\frac{1}{h_t h_x^{d+1}} \sum_{j=i+1}^d \int_{A_j} dz \int_0^T dr \partial_{x_i} K_x(\frac{y-z}{h_x}) K_t(\frac{s-r}{h_t}) v(r,z)$.

We integrate by parts $\int_{[-a,a]} dz_i \partial_{x_i} K_x^i\left(\frac{y_i - z_i}{h_x}\right) v(r,z)$.

$$\frac{1}{h_x}\int_{[-a,a]} dz_i \partial_{x_i} K_x^i\left(\frac{y_i - z_i}{h_x}\right) v(r,z) = - K_x^i\left(\frac{y_i - a}{h_x}\right) v(r, z_a^i) + K_x^i\left(\frac{y_i + a}{h_x}\right) v(r, z_{-a}^i) + \int_{[-a,a]} dz_i \partial_{x_i} v(r,z) K_x^i\left(\frac{y_i - z_i}{h_x}\right).$$

Then, $\tilde{B}_{33}^i(s,y)^2 \leq C_v^2 |K_x|_\infty^2 \frac{1}{h_x^2}(\mathbf{1}_{|y_i+a|\leq h_x} + \mathbf{1}_{|y_i-a|\leq h_x}) + C_v^2 \frac{|K_x|_\infty^2}{h_x^d}\sum_{j=i+1}^d \int_{A_j} dz \mathbf{1}_{|y_1-z_1|\leq h_x}\cdots \mathbf{1}_{|y_d-z_d|\leq h_x}$. By using the previous results for $\tilde{B}_{31}^i(s,y)^2$ and $\tilde{B}_{32}^i(s,y)^2$, we get $\|\tilde{B}_{33}^i(s,y)\|_{H_{\beta,X}^\mu}^2 \leq K(T)(1+\frac{1}{h_x})e^{-\frac{\mu}{\sqrt{d}}(a-h_x)}$. □

Lemma 12.56. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. If $\partial_{x_i} v$ is bounded by C_v, we get $\|\tilde{B}_4^i(s,y)\|_{H_{\beta,X}^\mu}^2 \leq K(T)h_t$, where $K(T)$ depends on $d,\mu,K,\beta,\alpha_2,C_v$, and $|K_t|_\infty$.*

Proof of Lemma 12.56.

We recall $\tilde{B}_4^i(s,y) = -\frac{1}{h_t h_x^d}\int_{\mathbb{R}^d} dz \int_{[-h_t,0]\cup[T,T+h_t]} dr K_x(\frac{y-z}{h_x})K_t(\frac{s-r}{h_t})\partial_{x_i} v(s,y)$. Since $\partial_{x_i} v$ is bounded, $\tilde{B}_4^i(s,y) \leq \frac{C_v}{h_t h_x^d}\int_{\mathbb{R}^d} dz K_x(\frac{y-z}{h_x})\int_{[-h_t,0]\cup[T,T+h_t]} dr K_t(\frac{s-r}{h_t})$. We end the proof as we did for Lemma 12.39, page 148. □

Combining Lemmas 12.52 to 12.56 leads to the following Proposition

Proposition 12.57. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume that v and $\partial_{x_i} v$ are bounded by C_v and v satisfies $\forall t,t' \in [0,T], \forall x \in \mathbb{R}^d$, $|\partial_{x_i} v(t,x) - \partial_{x_i} v(t',x)| \leq C\sqrt{|t'-t|}$, where C is a positive constant.*

1. *Assume that v is a C^2 function is space. Then,*

$$\|\mathbb{E}(\partial_{x_i} B_n - \partial_{x_i} v)\|_{H_{\beta,X}^\mu}^2 \leq K_0(T)h_x^2\left\|\partial_x^2 v\right\|_{H_{\beta,X}^\mu}^2 + K_1(T)\left(\frac{1}{h_x}e^{-\frac{\mu}{\sqrt{d}}a} + h_t\right),$$

where $K_0(T)$ depends on $d,\mu,K,\alpha_1,\alpha_2,\beta$ and $|K_x|_\infty$ and $K_1(T)$ depends on $C,d,\mu,K,\alpha_1,\alpha_2,\beta,C_v,|K_t|_\infty$ and $|K_x|_\infty$.

2. *Assume in addition that $\partial_x^2 v$ is bounded by C_v. Then,*

$$\|\mathbb{E}(\partial_{x_i} B_n - \partial_{x_i} v)\|_{H_{\beta,X}^\mu}^2 \leq K(T)\left(h_t + h_x^2 + \frac{1}{h_x}e^{-\frac{\mu}{\sqrt{d}}a}\right),$$

where $K(T)$ depends on $C,d,\mu,K,\alpha_1,\alpha_2,\beta,C_v,|K_t|_\infty$ and $|K_x|_\infty$.

Combining Proposition 12.51, page 155 and Proposition 12.57 leads to the following Theorem

Theorem 12.58. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume that v and $\partial_{x_i}v$ are bounded by C_v and v satisfies $\forall t,t'\in[0,T], \forall x\in\mathbb{R}^d$, $\forall i\in\{1,\cdots,d\}$, $|\partial_{x_i}v(t,x)-\partial_{x_i}v(t',x)|\le C\sqrt{|t'-t|}$, where C is a positive constant.*

1. *Assume that v is a C^2 function in space. Then, for all $i\in\{1,\cdots,d\}$,*
$$\|\partial_{x_i}B_n-\partial_{x_i}v\|^2_{H^\mu_{\beta,X}} \le K_0(T)\left(\frac{T\lambda(B)\delta_n}{h_x^2}\|v\|^2_{H^\mu_{\beta,X}}+h_x^2\left\|\partial_x^2 v\right\|^2_{H^\mu_{\beta,X}}\right) + K_1(T)\left(\frac{1}{h_x}e^{-\frac{\mu}{\sqrt{d}}a}+h_t\right),$$
where $K_0(T)$ depends on $d,\mu,K,\alpha_1,\alpha_2,\beta$ and $|K_x|_\infty$ and $K_1(T)$ depends on $C,d,\mu,K,\alpha_1,\alpha_2,\beta,C_v,|K_t|_\infty$ and $|K_x|_\infty$.

2. *Assume that v is a $C_b^{1,2}$ function and let C_v denote the constant bounding $v,\partial_t v,\partial_x v$ and $\partial_x^2 v$. Then, for all $i\in\{1,\cdots,d\}$,*
$$\|\partial_{x_i}B_n-\partial_{x_i}v\|^2_{H^\mu_{\beta,X}} \le K(T)\left(\frac{T\lambda(B)\delta_n}{h_x^2}+h_t+h_x^2+\frac{1}{h_x}e^{-\frac{\mu}{\sqrt{d}}a}\right),$$
where $K(T)$ depends on $C,d,\mu,K,\alpha_1,\alpha_2,\beta,C_v,|K_t|_\infty$ and $|K_x|_\infty$.

12.4.4 Conclusion

Combining (12.21) page 150, (12.22) page 150, and Theorem 12.44 page 151, Theorem 12.50, page 154 and Theorem 12.58 yields the following Theorem

Theorem 12.59. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume that v and $\partial_x v$ are bounded by C_v and v satisfies $\forall t,t'\in[0,T],\forall x\in\mathbb{R}^d$, $|\partial_x v(t,x)-\partial_x v(t',x)|\le C\sqrt{|t'-t|}$, where C is a positive constant. For all $\epsilon_0\ge 0$ satisfying $\epsilon_0^2\le (T\lambda(B))^{-2}$,*

1. *Assume that v is a C^2 function in space. Then,*
$$\mathbb{E}\,\|\partial_x(\mathcal{P}_n v)-\partial_x v\|^2_{H^\mu_{\beta,X}} \le K(T)\left((T\lambda(B))^2\frac{\epsilon_0^2}{h_x^2}+\frac{T\lambda(B)\delta_n}{h_x^2}+h_x^2\right)\|v\|^2_{H^{2,\mu}_{\beta,X}}$$
$$+K(T)\left[h_t h_x^{-2}+\frac{1}{h_x}e^{-\mu a}a^{d-1}+h_x^{-1}e^{-\frac{\mu}{\sqrt{d}}a}\right]$$
$$+K(T)\left[\frac{(T\lambda(B))^2}{h_x^2}\left(\epsilon_0^2+\frac{\delta_n}{T\lambda(B)}\right)\exp\left(-\frac{c\epsilon_0^2 T\lambda(B)}{\delta_n}\right)+\frac{\sqrt{T\lambda(B)\delta_n}}{h_x^2}\exp\left(-\frac{c}{T\lambda(B)\delta_n}\right)\right],$$
where $K(T)$ depends on $C,d,\mu,K,\alpha_1,\alpha_2,\beta,C_v,|K_t|_\infty$ and $|K_x|_\infty$.

2. *Assume that v is a $C_b^{1,2}$ function and let C_v denote the constant bounding $v,\partial_t v,\partial_x v$ and $\partial_x^2 v$. Then,*
$$\mathbb{E}\,\|\partial_x(\mathcal{P}_n v)-\partial_x v\|^2_{H^\mu_{\beta,X}} \le K(T)\left(h_t h_x^{-2}+h_x^2+h_x^{-1}e^{-\frac{\mu}{\sqrt{d}}a}\right)$$
$$+K(T)\left(\frac{T\lambda(B)\delta_n}{h_x^2}+\frac{e^{-\mu a}a^{d-1}}{h_x}+\frac{\sqrt{T\lambda(B)\delta_n}}{h_x^2}\exp\left(-\frac{c}{T\lambda(B)\delta_n}\right)\right).$$

where $K(T)$ depends on $C, d, \mu, K, \alpha_1, \alpha_2, \beta, C_v, |K_t|_\infty$ and $|K_x|_\infty$.

Remark 12.60. Looking at the first result stated in Theorem 12.42, we notice that the coefficients in front of $\|v\|^2_{H^{2,\mu}_{\beta,X}}$ and $\|\partial_t v\|^2_{H^{\mu}_{\beta,X}}$ are proportional to h_x^4, h_t^2 and δ_n. They correspond to the error terms due to the $\mathbb{L}_2$ error $\mathbb{E}|\mathcal{P}_n v - v|^2$: h_t^2 and h_x^4 come from the bias error $|\mathbb{E}(\mathcal{P}_n v - v)|^2$ and δ_n corresponds to the variance error $\mathrm{Var}(\mathcal{P}_n v)$. In other words, had we studied $\mathbb{E}|\mathcal{P}_n v(t,x) - v(t,x)|^2$ without integrating w.r.t. $e^{\beta t} dt \otimes e^{-\mu|x|} dx$, we would only have obtained an error of order h_x^4, h_t^2 and δ_n. Since we integrate w.r.t. x, the truncation error $h_t + e^{-\frac{\mu}{\sqrt{d}} a} + e^{-\mu a} a^{d-1} h_x$ appears.
Since we study $\partial_x \mathcal{P}_n v - \partial_x v$ in Theorem 12.59, it is consistent to get errors of order h_x^2, $\frac{\delta_n}{h_x^2}$, $h_t h_x^{-2}$, $e^{-\mu a} a^{d-1} h_x^{-1}$.

Chapter 13

Convergence Results of our algorithm

We recall that our goal is to prove the convergence of our algorithm which approximates (Y,Z), the solution of the BSDE

$$(E)\begin{cases} -dY_t = f(t,X_t,Y_t,Z_t)dt - Z_t dW_t, \\ Y_T = \Phi(X_T), \\ X_t = x + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dW_s. \end{cases}$$

Theorem 9.14, page 99, links (Y,Z) to u, the solution of the following PDE :

$$(\mathcal{E})\begin{cases} \partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),(\partial_x u\sigma)(t,x)) = 0, \\ u(T,x) = \Phi(x), \end{cases}$$

where $\mathcal{L}$ is defined by

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

According to this Theorem, under Hypothesis 9.5, page 99, we get $u \in C_b^{1,2}$ and

$$\forall t \in [0,T], \;\; (Y_t,Z_t) = (u(t,X_t), \partial_x u(t,X_t)\sigma(t,X_t)).$$

Our algorithm (described in Section 9.6, page 103), builds u_k, an approximation of u, and deduces (Y^k,Z^k), the approximation of (Y,Z) by using the following formula

$$\forall t \in [0,T], \; Y_t^k = u_k(t,X_t^N), \;\; Z_t^k = \partial_x u_k(t,X_t^N)\sigma(t,X_t^N),$$

where X^N is an approximation of X. More precisely, we build u_k recursively in the following way

$$u_k(t,x) = \mathcal{P}_n^k(u_{k-1} + \overline{w}_k)(t,x),$$

where $\overline{w}_k(t,x)$ is given by (9.13), page 104 and $\mathcal{P}_n^k$ is a Kernel estimator described in Section 11.3, page 119 and studied in Chapter 12, page 125.
In this chapter, we show that $\|Y^k - Y\|^2_{\mu,\beta} + \|Z^k - Z\|^2_{\mu,\beta}$ tends to 0 when k goes to

infinity, where $\|\cdot\|_{\mu,\beta}$ is a given norm specified below. After giving some Definitions and Notations in Section 13.1, we state in Section 13.2 the main result of this chapter : Theorem 13.10, which proves the convergence of our algorithm. This theorem ensues from two Propositions (Proposition 13.11 and Proposition13.12). We prove Proposition 13.11 in Section 13.3. Proposition 13.12 page 167 is quite long to establish. We present the scheme of the proof of Proposition 13.12 in Section 13.4. This proof combines results from Chapter 12 and from the next Chapter.

13.1 Definitions and Notations

In the whole Chapter 13, we assume Hypothesis 12.1, page 126, concerning the parameters used to build $\mathcal{P}_n^k$. First, let us recall some Definitions given in Part II.

Definition 13.1 (Definition of K, α_1, α_2, σ_0, σ_1 and c)**.** We recall the definition of some constants used in this Section and introduced in Part II.

1. σ_0 and σ_1 denote the constants appearing in Definition 6.2, page 63: σ is **uniformly elliptic** on $[0,T]\times\mathbb{R}^d$ if there exist two positive constants σ_0, σ_1 s.t., for any vector ξ and any $(t,x)\in[0,T]\times\mathbb{R}^d$

$$\sigma_0|\xi|^2 \le \sum_{i,j=1}^{d}[\sigma\sigma^*]_{i,j}(t,x)\xi_i\xi_j \le \sigma_1|\xi|^2.$$

2. K, α_1, and α_2 denote the constants introduced in Proposition 6.4, page 63:
Assume that the coefficients σ and b are bounded measurable functions of $(t,x)\in[0,T]\times\mathbb{R}^d$ and that σ is elliptic. There exist positive constants K,α_1,α_2 s.t.

$$\frac{K^{-1}}{(2\pi\alpha_1(s-t))^{\frac{d}{2}}}e^{-\frac{|x-y|^2}{2\alpha_1(s-t)}} \le p(t,x;s,y) \le K\frac{1}{(2\pi\alpha_2(s-t))^{\frac{d}{2}}}e^{-\frac{|x-y|^2}{2\alpha_2(s-t)}}.$$

The constant K depends only on σ_0,σ_1,d,T and the suprema of the coefficients σ, b. The constants α_0 and α_1 depend on σ_0,σ_1 and d.

3. c denotes the constant introduced in Theorem 8.1, page 78:
Assume σ is uniformly elliptic, $b\in C_b^{1,2}$ and $\sigma\in C_b^{1,3}$, $\partial_t\sigma\in C_b^1$ in space. Then, $\forall(s,x,y)\in[t,T]\times\mathbb{R}^d\times\mathbb{R}^d$, there exist a constant $c>0$ and a function $K(T)$ non decreasing in T and depending on the dimension d and on the upper bounds of σ, b and their derivatives s.t.

$$|p(t,x;s,y)-p^N(t,x;s,y)| \le \frac{K(T)(T-t)}{N(s-t)^{\frac{d+1}{2}}}\exp(-\frac{c|x-y|^2}{s-t}).$$

Under these hypotheses, we also have

$$|p^N(t,x;s,y)| \le \frac{K(T)}{(s-t)^{\frac{d}{2}}}\exp(-\frac{c|x-y|^2}{s-t}).$$

Second, let us recall the definitions of $H^{\mu}_{\beta,X}$ and H^{μ}_{β}.

Definition 13.2. For any $\beta > 0$, for any $\mu > 0$ and any diffusion process $(\tilde{X}_s)_{0\leq s\leq T}$ starting from x at time 0, which transition density function is denoted p, $H^{\mu}_{\beta,\tilde{X}}$ defines the space of functions $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\|v\|^2_{H^{\mu}_{\beta,\tilde{X}}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \mathbb{E}|v(s, \tilde{X}^x_s)|^2 dxds < \infty.$$

Using the definition of ν, we also get $\|v\|^2_{H^{\mu}_{\beta,\tilde{X}}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} dy\nu^0(s,y)|v(s,y)|^2$, where $\nu^t(s,y) = \int_{\mathbb{R}^d} e^{-\mu|x|} p(t,x;s,y)$ (see Definition 6.11, page 67).

Definition 13.3. For any $\beta > 0$ and $\mu > 0$, let H^{μ}_{β} define the space of functions $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\|v\|^2_{H^{\mu}_{\beta}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} |v(s,x)|^2 dx.$$

From Proposition 6.12, page 67, we deduce the following Proposition

Proposition 13.4 (Norm equivalence)**.** *Assume that the coefficients σ, b are Lipschitz bounded measurable functions on $[0,T] \times \mathbb{R}^d$ and that σ satisfies the ellipticity condition. There exist two constants $c > 0$ and $C > 0$ depending on $T, d, \mu, K, \alpha_1, \alpha_2$ s.t. for every $v \in L^1((0,T) \times \mathbb{R}^d, e^{\beta t}dt \otimes e^{-\mu|x|}dx)$*

$$c \|v\|^2_{H^{\mu}_{\beta}} \leq \|v\|^2_{H^{\mu}_{\beta,X}} \leq C \|v\|^2_{H^{\mu}_{\beta}}.$$

Definition 13.5 (Definition of $\mathbb{H}^{\mu}_{\beta}(\mathbb{R}^q)$)**.**

1. For any $\beta > 0$, we define $(\mathbb{H}^{\mu}_{\beta}(\mathbb{R}^q), \|\cdot\|_{\mu,\beta})$ the set of predictable processes $V : \Omega \times [0,T] \times \mathbb{R}^d \to \mathbb{R}^q$ such that

$$\|V\|^2_{\mu,\beta} := \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |V_s(x)|^2 e^{-\mu|x|} dxds\right]$$

   is finite.

2. Let (X,Y,Z) be the solution of (E) (see page 99), u the solution of $(\mathcal{E})$, (see page 99), and (Y^k, Z^k) and X^N defined as above. By using the previous definition, we get

$$\|Y - Y^k\|^2_{\mu,\beta} = \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |u(s, X^x_s) - u_k(s, X^{N,x}_s)|^2 e^{-\mu|x|} dxds\right], \tag{13.1}$$

$$\|Z - Z^k\|^2_{\mu,\beta} = \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |(\partial_x u\sigma)(s, X^x_s) - (\partial_x u_k\sigma)(s, X^{N,x}_s)|^2 e^{-\mu|x|} dxds\right], \tag{13.2}$$

   We can do it since u, u_k, $\partial_x u$ and $\partial_x u_k$ are bounded (see Theorem 9.14, page 99 and Proposition 11.9, page 123), then $\|Y - Y^k\|^2_{\mu,\beta}$ and $\|Z - Z^k\|^2_{\mu,\beta}$ are finite.

Remark 13.6. We point out that the expectation appearing in the above definition of $\|Y^k - Y\|^2_{\mu,\beta}$ and $\|Z^k - Z\|^2_{\mu,\beta}$ is computed w.r.t. the law of X, X^N and all the random variables used to compute u_k. We refer to Remark 9.19 for a detailed survey.

Definition 13.7 (Definition of $\tilde{Y}^k$ and $\tilde{Z}^k$)**.** Let $(\tilde{Y}^k, \tilde{Z}^k)$ be two processes given by

$$\forall t \in [0,T],\ \tilde{Y}^k_t = u_k(t, X_t),\ \ \tilde{Z}^k_t = \partial_x u_k(t, X_t)\sigma(t, X_t).$$

Remark 13.8. Since Y and $\tilde{Y}^k$ are processes both depending on the process X and since u and u_k are bounded, we can write

$$\begin{aligned}\|Y - \tilde{Y}^k\|^2_{\mu,\beta} &= \mathbb{E}\left[\int_0^T e^{\beta s} \int_{\mathbb{R}^d} |u(s, X^x_s) - u_k(s, X^x_s)|^2 e^{-\mu|x|} dx ds\right] \\ &= \mathbb{E}\,\|u - u_k\|^2_{H^\mu_{\beta,X}} = \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y)\mathbb{E}|u(s,y) - u_k(s,y)|^2.\end{aligned}$$

We refer to Definition 12.2 (page 126), for the definition of $\nu^0(s, y)$. The same result holds for $\|Z - \tilde{Z}^k\|^2_{\mu,\beta}$: $\|Z - \tilde{Z}^k\|^2_{\mu,\beta} = \mathbb{E}\,\|\partial_x u\sigma - \partial_x u_k\sigma\|^2_{H^\mu_{\beta,X}} = \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y)|\partial_x u\sigma - \partial_x u_k \sigma|^2(s,y)$.

Finally, we give some notations commonly used in the sequel.

Definition 13.9 (Euler Scheme)**.** We approximate X by its Euler scheme of order $N \geq 1$, denoted X^N, and defined as follows. We consider the subdivision $\{0 = t_0 < t_1 < \cdots < t_N = T\}$ of the interval $[0,T]$, i.e. $t_k = k\frac{T}{N}$. We put $X^N_0 = x$ and, for all $k \in \{0, \cdots, N-1\}$ and $t \in [t_k, t_{k+1}]$, $\forall i \in \{1, \cdots, d\}$,

$$X^{N,i}_t = X^{N,i}_{t_k} + b_i(t_k, X^N_{t_k})(t - t_k) + \sum_{j=1}^q \sigma_{ij}(t_k, X^N_{t_k})(W^j_t - W^j_{t_k}). \tag{13.3}$$

Note that the continuous Euler scheme is an Itô process verifying

$$X^N_t = x + \int_0^t b(\varphi(s), X^N_{\varphi(s)})ds + \int_0^t \sigma(\varphi(s), X^N_{\varphi(s)})dW_s$$

where $\varphi(t) := \sup\{t_k : t_k \leq t\}$.

13.2 Results

Theorem 13.10. *Assume that σ is uniformly elliptic, $b \in C^{1,2}_b$ and $\sigma \in C^{1,3}_b$, $\partial_t\sigma \in C^1_b$ in space. We also assume that f is a bounded Lipschitz function (whose Lipschitz constant is noted L_f) and $\Phi \in C^{2+\alpha}_b$. Then, $Y - Y^k$ and $Z - Z^k$ converge to 0 in norm $\|\cdot\|_{\mu,\beta}$ when $N, n, a, h_x^{-1}, h_t^{-1}$ and M, the parameters introduced in the description of the Algorithm (see Section 9.6) related to the discretization scheme, the kernel estimator, and the Monte Carlo method, tend to $+\infty$. More precisely, we have*

$$\|Y - Y^k\|^2_{\mu,\beta} + \|Z - Z^k\|^2_{\mu,\beta} \leq C\frac{K(T)}{N} + K(T)(\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta})$$

and $\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta}$ geometrically converge to a constant C depending on the above parameters when k goes to infinity. Moreover C goes to 0 when $N, n, a, h_x^{-1}, h_t^{-1}$ and M tend to $+\infty$. $K(T)$ depends on the dimension d, μ, the upper bounds of σ, b and *their derivatives, K, c and α_1. Moreover, if $(\beta, h_x, h_t, n, M, N, a)$ are chosen such that the following η_1 is smaller than 1,*

$$\limsup_k \left(\|Y^k - Y\|^2_{\mu,\beta} + \|Z^k - Z\|^2_{\mu,\beta} \right) \leq C\frac{K(T)}{N} + \frac{K(T)}{1-\eta_1} C_1,$$

where

$$\eta_1 = \frac{4(1+T)L_f^2}{\beta} + K_\eta(T) \left(T\lambda(B)\delta_n(1 + M^{-1}h_x^{-2}) + h_x^2 + h_t^2 \right),$$
$$C_1 = \frac{K_C(T)}{N^2 h_x^2} + K_\eta(T) \left(T\lambda(B)\delta_n h_x^{-2} + (h_t + h_x e^{-\mu a} a^{d-1}) h_x^{-2} + h_x^{-1} e^{-\frac{\mu}{\sqrt{d}}a} + h_x^2 \right),$$

$K_\eta(T)$ depends on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, |\partial_x K_x|_\infty, C_u, L_f$ and the upper bounds for f, b, σ and and their derivatives, on c, K, α_1, α_2 and σ_0 and σ_1. $K_C(T)$ depends on the dimension d, μ, β, the upper bounds for f and Φ and for σ, b and their derivatives, on c, and on K, α_2 and α_1.

The proof of this Theorem ensues from the two following Propositions.

Proposition 13.11. *Assume that σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t \sigma \in C_b^1$ in space. We also assume that f is a bounded Lipschitz function and $\Phi \in C_b^{2+\alpha}$. Then, there exists a constant C depending on $|u|_\infty, |\partial_x u|_\infty, |\partial_x^2 u|_\infty, |\sigma|_\infty$ and μ and a function $K(T)$ such that*

$$\|Y - Y^k\|^2_{\mu,\beta} + \|Z - Z^k\|^2_{\mu,\beta} \leq C\frac{K(T)}{N} + K(T)(\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta}),$$

where $K(T)$ depends on the dimension d, μ, the upper bounds of σ, b and their derivatives, K, c and α_1.

Proposition 13.11 is proved in Section 13.3.

Proposition 13.12. *Assume that σ satisfies the ellipticity condition, $b \in C_b^{1,2}$, $\sigma \in C_b^{1,3}$, $\partial_t \sigma$ is C_b^1 in space. We also assume that f is a bounded Lipschitz function and $\Phi \in C_b^{2+\alpha}$. Then,*

$$\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta} \leq \eta_1 \left(\|Y - \tilde{Y}^{k-1}\|^2_{\mu,\beta} + \|Z - \tilde{Z}^{k-1}\|^2_{\mu,\beta} \right) + C_1,$$

where η_1 and C_1 are defined in Theorem 13.10. We can choose parameters $(\beta, h_x, h_t, n, M, N, a)$ such that $\eta_1 < 1$ and C_1 tends to 0 when $(h_x^{-1}, h_t^{-1}, n, M, N, a)$ go to ∞.

Proof of Theorem 13.10. Let C denote $\|Y\|^2_{\mu,\beta} + \|Z\|^2_{\mu,\beta}$. From Proposition 13.12, we get

$$\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta} \leq \eta_1^k C + \frac{1-\eta_1^k}{1-\eta_1} C_1.$$

Hence, $\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta} \leq \eta_1^k \left(C - \frac{C_1}{1-\eta_1} \right) + \frac{1}{1-\eta_1} C_1$. Then, from Proposition 13.11, we deduce

$$\|Y - Y^k\|^2_{\mu,\beta} + \|Z - Z^k\|^2_{\mu,\beta} \leq C\frac{K(T)}{N} + K(T) \left(\eta_1^k \left(C - \frac{C_1}{1-\eta_1} \right) + \frac{1}{1-\eta_1} C_1 \right).$$

Since $\eta_1 < 1$,

$$\limsup_k \left(\|Y^k - Y\|^2_{\mu,\beta} + \|Z^k - Z\|^2_{\mu,\beta} \right) \leq C\frac{K(T)}{N} + \frac{K(T)}{1-\eta_1}C_1,$$

and the r.h.s. of the previous inequality tends to 0 when $N, n, a, h_x^{-1}, h_t^{-1}$ and M tend to $+\infty$. □

Remark 13.13 (Analysis of the terms bounding $\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta}$)**.** Looking at Theorem 13.10 and more particularly to the constants C_1 and η_1, we notice that if we neglect N we get the error terms $\frac{\delta_n}{h_x^2}, h_x^2, h_t h_x^{-2}, h_x^{-1} e^{-\mu a} a^{d-1}$. These terms appear in the upper bound for $\|v - \mathcal{P}_n v\|^2_{H^\mu_{\beta,X}}$ and $\|\partial_x v - \partial_x(\mathcal{P}_n v)\|^2_{H^\mu_{\beta,X}}$ (see Theorem 12.42 page 149 and Theorem 12.59 page 160). The error terms $h_t h_x^{-2}$ and $h_x^{-1} e^{-\mu a} a^{d-1}$ appearing in C_1 are due to the boundary discontinuities (related to the gradient). We guess that a suitable modification of our algorithm might remove these errors. This is left for further research.

Remark 13.14. The constant M, which denotes the number of Monte Carlo simulations used to compute $\overline{w}_k$ (see Section 9.6), only appears in the contraction constant η_1. This deeply relies on the adaptive control variates method. Had we implemented a non adaptive method (using naive Picard iterations as presented in the introduction of the manuscript), M^{-1} would have appeared in C_1. This would have led to a choice of M quite large.

13.3 Proof of Proposition 13.11

Under the Hypotheses of Proposition 13.11, we can apply Theorem 9.13, page 98, which states $u \in C_b^{1,2}$. We recall (13.1) and (13.2), the starting point of our proof.

$$\begin{aligned}
\|Y^k - Y\|^2_{\mu,\beta} &= \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |u(s, X_s^x) - u_k(s, X_s^{N,x})|^2 e^{-\mu|x|} dx ds\right], \\
\|Z^k - Z\|^2_{\mu,\beta} &= \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |(\partial_x u \sigma)(s, X_s^x) - (\partial_x u_k \sigma)(s, X_s^{N,x})|^2 e^{-\mu|x|} dx ds\right].
\end{aligned}$$

We split $\|Y^k - Y\|^2_{\mu,\beta}$ (resp.$\|Z^k - Z\|^2_{\mu,\beta}$) into two terms, by introducing $u(s, X_s^{N,x})$ (resp. $\partial_x u\sigma(s, X_s^{N,x})$). Since

$$|u(s, X_s^x) - u_k(s, X_s^{N,x})|^2 \leq 2\left(|u(s, X_s^x) - u(s, X_s^{N,x})|^2 + |u(s, X_s^{N,x}) - u_k(s, X_s^{N,x})|^2\right)$$

and

$$\begin{aligned}
|\partial_x u\sigma(s, X_s^x) - \partial_x u_k \sigma(s, X_s^{N,x})|^2 \leq & 2\left(|\partial_x u\sigma(s, X_s^x) - \partial_x u\sigma(s, X_s^{N,x})|^2\right. \\
& \left. + |\partial_x u\sigma(s, X_s^{N,x}) - \partial_x u_k \sigma(s, X_s^{N,x})|^2\right),
\end{aligned}$$

proving Proposition 13.11 comes down to study in Lemma 13.15

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |u(s, X_s^x) - u(s, X_s^{N,x})|^2 e^{-\mu|x|} dx ds\right] \text{ and} \\
&E\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |\partial_x u\sigma(s, X_s^x) - \partial_x u\sigma(s, X_s^{N,x})|^2 e^{-\mu|x|} dx ds\right]
\end{aligned}$$

and to study in Lemma 13.16

$$\mathbb{E}\left[\int_0^T\int_{\mathbb{R}^d} e^{\beta s}|u(s,X_s^{N,x})-u_k(s,X_s^{N,x})|^2e^{-\mu|x|}dxds\right]=\mathbb{E}\,\|u-u_k\|^2_{H^\mu_{\beta,X^N}} \text{ and}$$
$$E\left[\int_0^T\int_{\mathbb{R}^d} e^{\beta s}|\partial_x u\sigma(s,X_s^{N,x})-\partial_x u_k\sigma(s,X_s^{N,x})|^2e^{-\mu|x|}dxds\right]=\mathbb{E}\,\|\partial_x u\sigma-\partial_x u_k\sigma\|^2_{H^\mu_{\beta,X^N}}.$$

Assume the following two Lemmas hold.

Lemma 13.15. *Assume u, $\partial_x u$ and $\partial_x^2 u$ are bounded by C_u and σ and b are bounded Lipschitz functions. Then, there exists a function $K(T)$ non decreasing in T and depending on the Lipschitz constant of σ and b such that*

$$\mathbb{E}\left[\int_0^T\int_{\mathbb{R}^d} e^{\beta s}|u(s,X_s^{x})-u(s,X_s^{N,x})|^2e^{-\mu|x|}dxds\right]\le C_u\frac{K(T)}{N},$$
$$E\left[\int_0^T\int_{\mathbb{R}^d} e^{\beta s}|\partial_x u\sigma(s,X_s^{x})-\partial_x u\sigma(s,X_s^{N,x})|^2e^{-\mu|x|}dxds\right]\le C_u|\sigma|_\infty\frac{K(T)}{N}.$$

Lemma 13.16. *Assume σ is uniformly elliptic, $b\in C_b^{1,2}$ and $\sigma\in C_b^{1,3}$, $\partial_t\sigma\in C_b^1$ in space. Then, there exists a function $K(T)$ non decreasing in T such that*

$$\mathbb{E}\,\|u-u_k\|^2_{H^\mu_{\beta,X^N}}\le K(T)\mathbb{E}\,\|u-u_k\|^2_{H^\mu_{\beta,X}},$$
$$\mathbb{E}\,\|\partial_x u\sigma-\partial_x u_k\sigma\|^2_{H^\mu_{\beta,X^N}}\le K(T)\mathbb{E}\,\|\partial_x u\sigma-\partial_x u_k\sigma\|^2_{H^\mu_{\beta,X}},$$

where $K(T)$ depends on the dimension d, μ the upper bounds for σ, b and their derivatives, K, c and α_1.

Then, combining Lemmas 13.15 and 13.16 and Remark 13.8, page 166 ends the proof of Proposition 13.11.

Proof of Lemma 13.15. Since $u,\partial_x u,\partial_x^2 u,\sigma$ and $\partial_x\sigma$ are bounded, $|u(s,X_s^x)-u(s,X_s^{N,x})|\le C_u|X_s^x-X_s^{N,x}|$ and $|\partial_x u\sigma(s,X_s^x)-\partial_x u\sigma(s,X_s^{N,x})|\le C_u|\sigma|_\infty|X_s^x-X_s^{N,x}|$. Using Theorem A.4, page 262, yields the result. □

Proof of Lemma 13.16.
First, we study $\mathbb{E}\left[\int_0^T\int_{\mathbb{R}^d} e^{\beta s}|u(s,X_s^{N,x})-u_k(s,X_s^{N,x})|^2e^{-\mu|x|}dxds\big|\mathcal{G}_k\right]$ and $\mathbb{E}\left[\int_0^T\int_{\mathbb{R}^d} e^{\beta s}|\partial_x u\sigma(s,X_s^{N,x})-\partial_x u_k\sigma(s,X_s^{N,x})|^2e^{-\mu|x|}dxds\big|\mathcal{G}_k\right]$, which are respectively $\|u-u_k\|^2_{H^\mu_{\beta,X^N}}$ and $\|\partial_x u\sigma-\partial_x u_k\sigma\|^2_{H^\mu_{\beta,X^N}}$.

$$\|u-u_k\|^2_{H^\mu_{\beta,X^N}}=\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}dy\int_{\mathbb{R}^d}dxe^{-\mu|x|}|u-u_k|^2(s,y)p^N(0,x;s,y).$$

Then, by using the upper bound for p^N recalled in the beginning of section 13.1, we get

$$\|u-u_k\|^2_{H^\mu_{\beta,X^N}}\le K(T)\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}dy\int_{\mathbb{R}^d}dxe^{-\mu|x|}|u-u_k|^2(s,y)\frac{1}{s^{\frac{d}{2}}}\exp\left(-c\frac{|x-y|^2}{s}\right).$$

From Lemma 6.15 page 68, and Proposition 6.12 page 67, we deduce

$$\int_{\mathbb{R}^d} dx e^{-\mu|x|} \frac{1}{s^{\frac{d}{2}}} \exp\left(-c\frac{|x-y|^2}{s}\right) \leq 2^d \left(\frac{\pi}{c}\right)^{\frac{d}{2}} e^{\frac{d\mu^2}{4c}s} e^{-\mu|y|} \leq 2^{2d} K \left(\frac{\pi}{c}\right)^{\frac{d}{2}} e^{(\frac{d\mu^2}{4c}-c_1)s} \nu^0(s,y).$$

Hence

$$\|u-u_k\|^2_{H^\mu_{\beta,X^N}} \leq K(T) e^{(\frac{d\mu^2}{4c}-c_1)T} \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \int_{\mathbb{R}^d} dx e^{-\mu|x|} |u-u_k|^2(s,y) p(0,x;s,y).$$

The procedure being the same for $\|\partial_x u\sigma - \partial_x u_k \sigma\|^2_{H^\mu_{\beta,X^N}}$ as for $\|u-u_k\|^2_{H^\mu_{\beta,X^N}}$, we deduce from this proof the second inequality of Lemma 13.16. □

13.4 Proof of Proposition 13.12

We recall the assumption under which Proposition 13.12 is stated

Hypothesis 13.1 *Assume σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t \sigma \in C_b^1$ in space. We also assume f is a bounded Lipschitz function and $\Phi \in C_b^{2+\alpha}$.*

Remark 13.17. Under Hypothesis 13.1, u defined as the solution of $(\mathcal{E})$, page 163 is a $C_b^{1,2}$ function. We refer to Theorem 9.13, page 98 for a proof of this assertion.

Definition 13.18 (Definition of C_u and L_f)**.** If f is a Lipschitz function, we recall L_f denotes its Lipschitz constant. If u is a function such that $u, \partial_t u, \partial_x u$ and $\partial_x^2 u$ are bounded, C_u denotes $|u|_\infty \vee |\partial_t u|_\infty \vee |\partial_x u|_\infty \vee |\partial_x^2 u|_\infty$.

We recall the result of Proposition 13.12

$$\|Y-\tilde{Y}^k\|^2_{\mu,\beta} + \|Z-\tilde{Z}^k\|^2_{\mu,\beta} \leq \eta_1 \left(\|Y-\tilde{Y}^{k-1}\|^2_{\mu,\beta} + \|Z-\tilde{Z}^{k-1}\|^2_{\mu,\beta}\right) + C_1,$$

where η_1 and C_1 are defined in Theorem 13.10, page 166. Before giving the scheme of the proof of Proposition 13.12, let us give some definitions.

13.4.1 Definitions

Definition 13.19 (Definition of $\overline{Y}^k, \overline{Z}^k$)**.** Consider the linear BSDE

$$\overline{Y}_t^k = \Phi(X_T) + \int_t^T f(s, X_s, \tilde{Y}_s^{k-1}, \tilde{Z}_s^{k-1}) ds - \int_t^T \overline{Z}_s^k dW_s. \tag{13.4}$$

$(\overline{Y}_s^k, \overline{Z}_s^k)_{0\leq s\leq T}$ is the unique solution to (13.4). The solution is uniquely defined because $f(s, X_s, \tilde{Y}_s^{k-1}, \tilde{Z}_s^{k-1})_s \in \mathbb{H}_T^2(\mathbb{R})$ (see Section 9.3).

We introduce the function H

Definition 13.20 (Definition of H)**.** Let $H : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ satisfy the following linear PDE

$$\begin{cases} (\partial_t + \mathcal{L})H(t,x) + f(t,x,u_{k-1}(t,x),(\partial_x u_{k-1}\sigma)(t,x)) - f(t,x,u(t,x),(\partial_x u\sigma)(t,x)) = 0, \\ H(T,x) = 0, \end{cases} \tag{13.5}$$

where u is the solution of $(\mathcal{E})$ (see page 99) and u_k is the approximated solution of u constructed with our algorithm (see Section 9.6).

Remark 13.21. We recall that $\mathcal{G}_{k-1}$ is the filtration generated by the set of all random variables used to build u_{k-1}.

Proposition 13.22. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$ and that σ satisfies the ellipticity condition. We also assume that f is a bounded Lipschitz function, and Φ is of class $C_b^{2+\alpha}$. Then, $H(t,x)$ admits the stochastic representation*

$$H(t,x) = \tag{13.6}$$
$$\mathbb{E}_{k-1}\left[\int_t^T f(s,X_s^{t,x},u_{k-1}(s,X_s^{t,x}),(\partial_x u_{k-1}\sigma)(s,X_s^{t,x})) - f(s,X_s^{t,x},u(s,X_s^{t,x}),(\partial_x u\sigma)(s,X_s^{t,x}))ds\right],$$

on $[0,T]\times\mathbb{R}^d$, $(X_s^{t,x})_{t\leq s\leq T}$ satisfies

$$X_s^{t,x} = x + \int_t^s b(r,X_r^{t,x})dr + \int_t^s \sigma(r,X_r^{t,x})dW_r.$$

$\mathbb{E}_{k-1}(\cdot)$ means $\mathbb{E}(\cdot|\mathcal{G}_{k-1})$. Moreover, H and $\partial_x H$ are bounded.

Corollary 13.23. *Assume that the coefficients σ and b are Lipschitz and bounded measurable functions on $[0,T]\times\mathbb{R}^d$, and that σ satisfies the ellipticity condition. We also assume that f is bounded.*

1. *If f is Lipschitz, σ and b are C_b^2 in space and Φ is of class $C_b^{2+\alpha}$, we get that H satisfies, for all $t,t'\in[0,T]$,*

$$|\partial_x H(t',x) - \partial_x H(t,x)| \leq C|f|_\infty\sqrt{|t'-t|}, \tag{13.7}$$

where C depends on d,σ_0,σ_1 and on the bounds for $\partial_x^k\sigma,\partial_x^k b$, for $k\leq 2$.

2. *If σ is $C_b^{1,1}$, we get*

$$\|H\|^2_{H^{2,\mu}_{\beta,X}} + \|\partial_t H\|^2_{H^{\mu}_{\beta,X}} \leq K(T),$$

where $K(T)$ depends on σ, its derivatives, $|f|_\infty,\mu,d,\beta,K$ and α_2.

Proof of Proposition 13.22. The stochastic representation of H ensues from the Feynman-Kac formula. We can apply Theorem A.6, page 263 since

- Hypotheses A.2 and A.3 page 262 are satisfied : b and σ being Lipschitz and bounded functions, they follow the linear growth condition and the above equation satisfied by $(X_s^{t,x})_{t\leq s\leq T}$ admits a unique strong solution,
- H satisfies the polynomial growth condition (A.6), page 263. We use Remark A.8, page 263 to prove it. Looking at the assumptions required in the Remark, one notice that we only have to prove

$$\tilde{f}(t,x) := f(t,x,u_{k-1}(t,x),(\partial_x u_{k-1}\sigma)(t,x)) - f(t,x,u(t,x),(\partial_x u\sigma)(t,x))$$

is Lipschitz in t and x and satisfies $|\tilde{f}(t,x)| \leq L(1+|x|^{2\eta})$, where $L>0$ and $\eta\geq 1$ (the assumptions on b and σ required in the Remark are satisfied). We easily prove that $\tilde{f}$ satisfies the polynomial growth condition since $\tilde{f}$ is bounded (by $2|f|_\infty$). We show that $\tilde{f}$ is Lipschitz in t and x, by using that f and σ are Lipschitz in all their variables and that u and u_{k-1} are $C_b^{1,2}$ (see Theorem 9.14 page 99 and Proposition 11.9 page 123).

Since f is bounded, the stochastic representation of H enables to deduce that H is bounded by $2T|f|_\infty$. To prove that $|\partial_x H|_\infty$ is bounded, we use the stochastic representation and we introduce p, the transition density function of X

$$H(t,x) = \int_{\mathbb{R}^d} dy \int_t^T ds \tilde{f}(s,y) p(t,x;s,y).$$

Since p satisfies (6.8) (see page 66), we can easily prove that the derivative of $H(t,x)$ w.r.t. x is

$$\partial_x H(t,x) = \int_{\mathbb{R}^d} dy \int_t^T ds \tilde{f}(s,y) \partial_x p(t,x;s,y).$$

By using (6.8), we obtain $|\partial_x H(t,x)| \leq 2C|f|_\infty \int_t^T \frac{ds}{\sqrt{s-t}} \int_{\mathbb{R}^d} dy e^{-\frac{c|x-y|^2}{s-t}} \frac{1}{(s-t)^{\frac{d}{2}}}$, which is bounded by a constant depending on $T, d, |f|_\infty, C$ and c (C and c are some constants defined in Proposition 6.9, page 66 and depending on d, σ_0 and σ_1). □

Proof of Corollary 13.23. To prove the first assertion, we apply Proposition 7.3, page 74. We can do it since $\tilde{f}$, defined in the proof of Proposition 13.22, is bounded and Lipschitz, σ is assumed to be elliptic and σ and b are bounded and Lipschitz.
The second assertion, stating $\|H\|^2_{H^{2,\mu}_{\beta,X}} + \|\partial_t H\|^2_{H^{\mu}_{\beta,X}} \leq K(T)$, is proved by using Theorem 7.1, page 74. Since σ is uniformly elliptic, C_b^2 in space and b is C^1, we can apply Remark 7.2 page 74 to get

$$\|H\|^2_{\mathbb{L}^2(0,T;W^{2,2,\mu})} + \|\partial_t H\|^2_{\mathbb{L}^2(0,T;W^{0,2,\mu})} \leq K(T) \, \|\tilde{f}\|^2_{\mathbb{L}^2(0,T;W^{0,2,\mu})},$$

where $K(T)$ only depends on σ and its derivatives. Since $\tilde{f}$ is bounded, we have $\|\tilde{f}\|^2_{\mathbb{L}^2(0,T;W^{0,2,\mu})}$ is bounded by $K_0(T)$, which depends on $|f|_\infty$ and μ. Then, $\|H\|^2_{\mathbb{L}^2(0,T;W^{2,2,\mu})} + \|\partial_t H\|^2_{\mathbb{L}^2(0,T;W^{0,2,\mu})} \leq K(T)$, where $K(T)$ depends on σ, its derivatives, $|f|_\infty$ and μ.
First, let us prove $\|\partial_t H\|^2_{H^{\mu}_{\beta,X}} \leq K(T)$. Using Definition 12.5 page 127 yields

$$\|\partial_t H\|^2_{H^{\mu}_{\beta,X}} = \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy \nu^0(s,y) |\partial_t H|^2(s,y).$$

Proposition 13.4, page 165 gives $\|\partial_t H\|^2_{H^{\mu}_{\beta,X}} \leq C\,\|\partial_t H\|^2_{H^{\mu}_{\beta}} = C \int_0^T ds e^{\beta s} \int_{\mathbb{R}^d} dy e^{-\mu|y|} |\partial_t H|^2(s,y)$. Then, $\|\partial_t H\|^2_{H^{\mu}_{\beta,X}} \leq C e^{\beta T} \int_0^T ds \int_{\mathbb{R}^d} dy e^{-\mu|y|} |\partial_t H|^2(s,y)$. Recalling that $\|\partial_t H\|^2_{\mathbb{L}^2(0,T;W^{0,2,\mu})} = \int_0^T ds \int_{\mathbb{R}^d} dy e^{-\mu|y|} |\partial_t H(s,y)|^2$ and $\|\partial_t H\|^2_{\mathbb{L}^2(0,T;W^{0,2,\mu})} \leq K(T)$ ends the proof of $\|\partial_t H\|^2_{H^{\mu}_{\beta,X}} \leq K(T)$. The same kind of proof states $\|H\|^2_{H^{2,\mu}_{\beta,X}} \leq K(T)$. □

13.4.2 Scheme of the proof

Studying the difference $Y - \tilde{Y}^k$ requires to split it into two terms $Y - \overline{Y}^k$ and $\overline{Y}^k - \tilde{Y}^k$, and to prove the convergence of each of them. The same argument applies for $Z - \tilde{Z}^k$. Then, the differences $Y - \tilde{Y}^k$ and $Z - \tilde{Z}^k$ are studied in the following way :

1. We show in the next Section (see Theorem 14.2 page 177) that if $\Phi(X_T)$ and f are standard parameters (see Section 9.3 page 95 for a definition of standard parameters), we get

$$\|Y - \overline{Y}^k\|^2_{\mu,\beta} + \|Z - \overline{Z}^k\|^2_{\mu,\beta} \leq \frac{2(1+T)L_f^2}{\beta}\left(\|Y - \tilde{Y}^{k-1}\|^2_{\mu,\beta} + \|Z - \tilde{Z}^{k-1}\|^2_{\mu,\beta}\right). \tag{13.8}$$

2. We prove in the next Section (see Theorem 14.4 page 178) that under Hypothesis 13.1 page 170,

$$\begin{aligned}\|\overline{Y}^k - \tilde{Y}^k\|^2_{\mu,\beta} + \|\overline{Z}^k - \tilde{Z}^k\|^2_{\mu,\beta} \leq{} & K_0(T)\frac{T\lambda(B)\delta_n}{MNh_x^2} + \frac{K_2(T)}{N^2h_x^2} \\ & + K_1(T)\frac{T\lambda(B)\delta_n}{Mh_x^2}\left(\|Y - \tilde{Y}^{k-1}\|^2_{\mu,\beta} + \|Z - \tilde{Z}^{k-1}\|^2_{\mu,\beta}\right) \\ & + 3(\mathbb{E}\,\|H - \mathcal{P}_n^k H\|^2_{H^\mu_{\beta,X}} + |\sigma|^2_\infty \mathbb{E}\,\|\partial_x H - \partial_x(\mathcal{P}_n^k H)\|^2_{H^\mu_{\beta,X}}) \\ & + 3(\mathbb{E}\,\|u - \mathcal{P}_n^k u\|^2_{H^\mu_{\beta,X}} + |\sigma|^2_\infty \mathbb{E}\,\|\partial_x u - \partial_x(\mathcal{P}_n^k u)\|^2_{H^\mu_{\beta,X}}).\end{aligned} \tag{13.9}$$

   $K_0(T)$ depends on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, |\partial_x K_x|_\infty$, C_u, L_b, L_σ and on K, α_2 and α_1. $K_1(T)$ is a function depending on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, |\partial_x K_x|_\infty, L_f$ and the upper bounds for b and σ, and their derivatives, on c, K, α_1, α_2 and σ_0 and σ_1. $K_2(T)$ depends on the dimension d, μ, β, the upper bounds for f and Φ and for σ, b and their derivatives, on c, and on K, α_2 and α_1.

3. We use the results on the error between a function and its approximation by the kernel method to bound $\mathbb{E}(\|u - \mathcal{P}_n^k u\|^2_{H^\mu_{\beta,X}})$ and $\mathbb{E}_{k-1}(\|H - \mathcal{P}_n^k H\|^2_{H^\mu_{\beta,X}})$ (see Chapter 12, Theorem 12.42, page 149). Since u is a $C_b^{1,2}$ function (see Remark 13.17 page 170)

$$\mathbb{E}\,\|\mathcal{P}_n^k u - u\|^2_{H^\mu_{\beta,X}} \leq K(T)\left(T\lambda(B)\delta_n + h_t + h_x^4 + e^{-\frac{\mu}{\sqrt{d}}a} + e^{-\mu a}a^{d-1}h_x\right),$$

   where $K(T)$ depends on $d, \mu, K, \alpha_1, \alpha_2, \beta, C_u, |K_t|_\infty$ and $|K_x|_\infty$.
   Since H is a $C^{1,2}$ function and is bounded by $2T|f|_\infty$ (see Proposition 13.22), we apply Theorem 12.42, page 149 with $\epsilon^2 = \frac{\delta_n}{T\lambda(B)}$

$$\begin{aligned}\mathbb{E}_{k-1}\,\|\mathcal{P}_n^k H - H\|^2_{H^\mu_{\beta,X}} \leq{} & K(T)\left((T\lambda(B)\delta_n + h_x^4)\,\|H\|^2_{H^{2,\mu}_{\beta,X}} + h_t^2\,\|\partial_t H\|^2_{H^\mu_{\beta,X}}\right) \\ & + K(T)\left[e^{-\frac{\mu}{\sqrt{d}}a} + h_t + e^{-\mu a}a^{d-1}h_x + T\lambda(B)\delta_n\right],\end{aligned}$$

   where $K(T)$ depends on $d, \mu, K, \alpha_1, \alpha_2, \beta, |f|_\infty, |K_t|_\infty$ and $|K_x|_\infty$.

4. We use the results on the error between the first derivative of a bounded function and the first derivative of its approximation by the kernel method to bound $\mathbb{E}\left\|\partial_x u - \partial_x(\mathcal{P}_n^k u)\right\|^2_{H^\mu_{\beta,X}}$ and $\mathbb{E}_{k-1}\left\|\partial_x H - \partial_x(\mathcal{P}_n^k H)\right\|^2_{H^\mu_{\beta,X}}$ (see Chapter 12, Theorem 12.59, page 160).

   Under Hypothesis 13.1, u is a $C_b^{1,2}$ function. Then,

$$\mathbb{E}\left\|\partial_x(\mathcal{P}_n^k u) - \partial_x u\right\|^2_{H^\mu_{\beta,X}} \leq K(T)\left(h_t h_x^{-2} + h_x^2 + h_x^{-1} e^{-\frac{\mu}{\sqrt{d}}a} + \frac{T\lambda(B)\delta_n}{h_x^2} + \frac{e^{-\mu a}a^{d-1}}{h_x}\right).$$

   where $K(T)$ depends on $C, d, \mu, K, \alpha_1, \alpha_2, \beta, C_u, |K_t|_\infty$ and $|K_x|_\infty$. We have used $e^{-\frac{1}{x}} < \sqrt{x}$.

   Under Hypothesis 13.1, Proposition 13.22 page 171 and Corollary 13.23 page 171 state that H and $\partial_x H$ are bounded by $K(T)|f|_\infty$ and H satisfies $\forall t, t' \in [0,T], \forall x \in \mathbb{R}^d$, $|\partial_x H(t,x) - \partial_x H(t',x)| \leq C\sqrt{|t'-t|}$, where C is a positive constant. Then, we can apply Theorem 12.59, page 160 with $\epsilon_0^2 = \frac{\delta_n}{T\lambda(B)}$,

$$\begin{aligned}\mathbb{E}\left\|\partial_x(\mathcal{P}_n^k H) - \partial_x H\right\|^2_{H^\mu_{\beta,X}} &\leq K(T)\left(T\lambda(B)\delta_n h_x^{-2} + h_x^2\right)\|H\|^2_{H^{2,\mu}_{\beta,X}} \\ &\quad + K(T)\left(h_t h_x^{-2} + \frac{1}{h_x}e^{-\mu a}a^{d-1} + h_x^{-1}e^{-\frac{\mu}{\sqrt{d}}a} + \frac{T\lambda(B)\delta_n}{h_x^2}\right),\end{aligned}$$

   where $K(T)$ depends on $C, d, \mu, K, \alpha_1, \alpha_2, \beta, |f|_\infty, |K_t|_\infty$ and $|K_x|_\infty$. We have used $\sqrt{x} > e^{-\frac{1}{x}}$, for $x \leq 1$. Moreover, since $\|H\|^2_{H^{2,\mu}_{\beta,X}}$ is bounded by $K(T)$ (see Corollary 13.23, page 171), we can bound $K(T)\frac{T\lambda(B)\delta_n}{h_x^2}\|H\|^2_{H^{2,\mu}_{\beta,X}}$ by $K(T)\frac{T\lambda(B)\delta_n}{h_x^2}$. We get

$$\begin{aligned}\mathbb{E}\left\|\partial_x(\mathcal{P}_n^k H) - \partial_x H\right\|^2_{H^\mu_{\beta,X}} &\leq K_0(T) h_x^2 \|H\|^2_{H^{2,\mu}_{\beta,X}} \\ &\quad + K_1(T)\left(h_t h_x^{-2} + \frac{1}{h_x}e^{-\mu a}a^{d-1} + h_x^{-1}e^{-\frac{\mu}{\sqrt{d}}a} + \frac{T\lambda(B)\delta_n}{h_x^2}\right),\end{aligned}$$

   where $K_0(T)$ depends on $C, d, \mu, K, \alpha_1, \alpha_2, \beta, |f|_\infty, |K_t|_\infty$ and $|K_x|_\infty$ and $K_1(T)$ depends on σ and its derivatives and on the preceding parameters.

5. We use Proposition 7.4, page 76 to conclude. Under Hypothesis 13.1, σ is uniformly elliptic, $\sigma \in C_b^{1,1}$ and b is $C^{1,1}$ and bounded. Then, H satisfies the PDE (13.5) (see page 170) and we get

$$\|H\|^2_{H^{2,\mu}_{\beta,X}} + \|\partial_t H\|^2_{H^\mu_{\beta,X}} \leq C\|\tilde{f}\|^2_{H^\mu_{\beta,X}},$$

   where $\tilde{f}(t,x) := f(t,x,u_{k-1}(t,x),(\partial_x u_{k-1}\sigma)(t,x)) - f(t,x,u(t,x),(\partial_x u\sigma)(t,x))$. Since $\mathbb{E}\|\tilde{f}\|^2_{H^\mu_{\beta,X}} = \int_0^T ds e^{\beta s}\int_{\mathbb{R}^d} dx e^{-\mu|x|}\mathbb{E}|\tilde{f}(s,X_s^x)|^2$ and f is Lipschitz, we get

$$\begin{aligned}&\mathbb{E}\|H\|^2_{H^{2,\mu}_{\beta,X}} + \mathbb{E}\|\partial_t H\|^2_{H^\mu_{\beta,X}} \\ &\leq CL_f^2\int_0^T ds e^{\beta s}\int_{\mathbb{R}^d} dx e^{-\mu|x|}\mathbb{E}|(u-u_{k-1})(s,X_s^x)|^2 + \mathbb{E}|\sigma(\partial_x u - \partial_x u_{k-1})(s,X_s^x)|^2,\end{aligned}$$

   and then $\mathbb{E}\|H\|^2_{H^{2,\mu}_{\beta,X}} + \mathbb{E}\|\partial_t H\|^2_{H^\mu_{\beta,X}} \leq CL_f^2(\|Y - \tilde{Y}^{k-1}\|^2_{\mu,\beta} + \|Z - \tilde{Z}^{k-1}\|^2_{\mu,\beta})$.

6. Combining the five preceding points, we get that η_1 and C_1 appearing in Proposition 13.12, page 167 are the following

$$\eta_1 = \frac{4(1+T)L_f^2}{\beta} + K_\eta(T)\left(T\lambda(B)\delta_n(1+M^{-1}h_x^{-2}) + h_x^2 + h_t^2\right),$$
$$C_1 = \frac{K_C(T)}{N^2 h_x^2} + K_\eta(T)\left(T\lambda(B)\delta_n h_x^{-2} + (h_t + h_x e^{-\mu a} a^{d-1})h_x^{-2} + h_x^{-1} e^{-\frac{\mu}{\sqrt{d}}a} + h_x^2\right),$$

   $K(T)$ depends on $L_f, \mu, \beta, d, |K_t|_\infty, |K_x|_\infty, |\partial_x K_x|_\infty, C_u$ and the upper bounds for f, b, σ and their derivatives, on c, K, α_1, α_2 and σ_0 and σ_1. $K_1(T)$ depends on the dimension d, μ, β, the upper bounds for f and Φ and for σ, b and their derivatives, on c, and on K, α_2 and α_1. We can choose parameters $(\beta, h_x, h_t, n, M, N, a)$ such that $\eta_1 < 1$ and C_1 tends to 0 when $(h_x^{-1}, h_t^{-1}, n, M, N, a)$ go to ∞.

Chapter 14

Proofs of the key results

14.1 Point 1: Definition of $\overline{Y}^k$ and $\overline{Z}^k$ and study of $Y - \overline{Y}^k$ and $Z - \overline{Z}^k$

We recall BSDE (13.4)

$$\overline{Y}_t^k = \Phi(X_T) + \int_t^T f(s, X_s, \tilde{Y}_s^{k-1}, \tilde{Z}_s^{k-1})ds - \int_t^T \overline{Z}_s^k dW_s.$$

Before stating the existence and uniqueness of the solution of (13.4), we point out

Remark 14.1. Since $(\tilde{Y}_s^{k-1}, \tilde{Z}_s^{k-1})_{0 \le s \le T} = (u_{k-1}(s, X_s), (\partial_x u_{k-1}\sigma)(s, X_s))_{0 \le s \le T}$ and σ is bounded, we deduce from Proposition 11.9, page 123, that $(\tilde{Y}^{k-1}, \tilde{Z}^{k-1})$ belongs to $\mathbb{H}^2_{T,\beta}(\mathbb{R}) \times \mathbb{H}^2_{T,\beta}(\mathbb{R}^q)$. This equation admits a unique solution in $\mathbb{H}^2_T(\mathbb{R}) \times \mathbb{H}^2_T(\mathbb{R}^q)$.

The following theorem states (13.8) (see page 173). The proof is based on Proposition 9.2.

Theorem 14.2. *Assume $\Phi(X_T)$ and f are standard parameters (see Section 9.3 for a definition of standard parameters). Let L_f denote the Lipschitz constant of f. Let $(\overline{Y}^k, \overline{Z}^k)$ be the solution of (13.4). For any $\beta > 0$ and any $\mu > 0$,*

$$\|Y - \overline{Y}^k\|^2_{\mu,\beta} + \|Z - \overline{Z}^k\|^2_{\mu,\beta} \le \frac{2(1+T)L_f^2}{\beta} \left(\|Y - \tilde{Y}^{k-1}\|^2_{\mu,\beta} + \|Z - \tilde{Z}^{k-1}\|^2_{\mu,\beta} \right).$$

Proof of Theorem 14.2. Let us consider the following BSDEs

$$\begin{aligned} &-dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \;\; Y_T = \Phi(X_T), \\ &-d\overline{Y}_t^k = f(t, X_t, \tilde{Y}_t^{k-1}, \tilde{Z}_t^{k-1})dt - \overline{Z}_t^k dW_t, \;\; \overline{Y}_T^k = \Phi(X_T). \end{aligned}$$

(Y, Z) and $(\tilde{Y}^{k-1}, \tilde{Z}^{k-1})$ are two elements of $\mathbb{H}^2_{T,\beta}(\mathbb{R}) \times \mathbb{H}^2_{T,\beta}(\mathbb{R}^q)$. By Proposition 9.2 (see page 95), with $L_f = 0$ and $\beta = \mu^2$, we obtain

$$\int_0^T e^{\beta s}\mathbb{E}[|Y_s - \overline{Y}_s^k|^2 | \mathcal{G}_{k-1}]ds \le \frac{T}{\beta} \int_0^T e^{\beta s}\mathbb{E}[|f(s, X_s, Y_s, Z_s) - f(s, X_s, \tilde{Y}_s^{k-1}, \tilde{Z}_s^{k-1})|^2 | \mathcal{G}_{k-1}]ds,$$

$$\int_0^T e^{\beta s}\mathbb{E}[|Z_s - \overline{Z}_s^k|^2 | \mathcal{G}_{k-1}]ds \le \frac{1}{\beta} \int_0^T e^{\beta s}\mathbb{E}[|f(s, X_s, Y_s, Z_s) - f(s, X_s, \tilde{Y}_s^{k-1}, \tilde{Z}_s^{k-1})|^2 | \mathcal{G}_{k-1}]ds.$$

Now, since f is Lipschitz continuous with constant L_f, we get

$$\int_0^T e^{\beta s}\mathbb{E}\left[|Y_s-\overline{Y}^k_s|^2+|Z_s-\overline{Z}^k_s|^2\Big|\mathcal{G}_{k-1}\right]ds\le \frac{2(1+T)L_f^2}{\beta}\int_0^T e^{\beta s}\mathbb{E}\left[|Y_s-\tilde{Y}^{k-1}_s|^2+|Z_s-\tilde{Z}^{k-1}_s|^2\Big|\mathcal{G}_{k-1}\right]ds.$$

By integrating w.r.t. $e^{-\mu|x|}dx$ and taking the expectation, the result follows. □

Proposition 14.3. *Let $(\overline{Y}^k,\overline{Z}^k)$ be the solution of (13.4) and $\overline{u}_k$ be the solution of the following PDE*

$$\begin{cases}(\partial_t+\mathcal{L})\overline{u}_k(t,x)+f(t,x,u_{k-1}(t,x),(\partial_x u_{k-1}\sigma)(t,x))=0,\\ \overline{u}_k(T,x)=\Phi(x).\end{cases}\tag{14.1}$$

Then,

$$\forall t\in[0,T],\ \ (\overline{Y}^k_t,\overline{Z}^k_t)=(\overline{u}_k(t,X_t),(\partial_x\overline{u}_k\sigma)(t,X_t)).\tag{14.2}$$

Moreover, $\forall(t,x)\in[0,T]\times\mathbb{R}^d$, $\overline{u}_k(t,x)=u(t,x)+H(t,x)$ where H satisfies the PDE (13.5)

$$\begin{cases}(\partial_t+\mathcal{L})H(t,x)+f(t,x,u_{k-1}(t,x),(\partial_x u_{k-1}\sigma)(t,x))-f(t,x,u(t,x),(\partial_x u\sigma)(t,x))=0,\\ H(T,x)=0.\end{cases}$$

Proof. By applying Itô's formula to $\overline{u}_k(s,X_s)$ we get

$$d\overline{u}_k(s,X_s)=(\partial_t\overline{u}_k(s,X_s)+\mathcal{L}\overline{u}_k(s,X_s))ds+(\partial_x\overline{u}_k\sigma)(s,X_s)dW_s.$$

Since $\overline{u}_k$ satisfies (14.1), it follows that

$$-d\overline{u}_k(s,X_s)=f(s,X_s,u_{k-1}(s,X_s),(\partial_x u_{k-1}\sigma)(s,X_s))ds-(\partial_x\overline{u}_k\sigma)(s,X_s)dW_s,$$

with $\overline{u}_k(T,X_T)=\Phi(X_T)$. Thus, $\{u_{k-1}(s,X_s),(\partial_x u_{k-1}\sigma)(s,X_s),s\in[0,T]\}$ is equal to the unique solution of the BSDE (13.4), and the result follows. We easily check that $\overline{u}_k(t,x)=u(t,x)+H(t,x)$ where H satisfies (13.5) page 170, since u satisfies $(\mathcal{E})$ (see page 99). □

14.2 Point 2: Study of $\overline{Y}^k-\tilde{Y}^k$ and $\overline{Z}^k-\tilde{Z}^k$

The result announced is an obvious consequence of the following Theorem.

Theorem 14.4. *Let $(\overline{Y}^k,\overline{Z}^k)$ be the solution of (13.4) page 170 and $(\tilde{Y}^k,\tilde{Z}^k)$ be defined by (14.2) page 178. Assume σ is uniformly elliptic, $b\in C_b^{1,2}$ and $\sigma\in C_b^{1,3}$, $\partial_t\sigma\in C_b^1$ in space. We also assume that f and Φ are bounded, f is Lipschitz and $u,\partial_t u,\partial_x u$ and $\partial_x^2 u$ are bounded by a constant C_u. Then,*

$$\|\overline{Y}^k-\tilde{Y}^k\|^2_{\mu,\beta}\le 3\mathbb{E}\,\|H-\mathcal{P}^k_n H\|^2_{H^\mu_{\beta,X}}+3\mathbb{E}\,\|u-\mathcal{P}^k_n u\|^2_{H^\mu_{\beta,X}}+K_0(T)\frac{T\lambda(B)\delta_n}{MN}+\frac{K_2(T)}{N^2}+K_1(T)\frac{T\lambda(B)\delta_n}{M}\left(\|Y-\tilde{Y}^{k-1}\|^2_{\mu,\beta}+\|Z-\tilde{Z}^{k-1}\|^2_{\mu,\beta}\right).\tag{14.3}$$

If $T\lambda(B)\delta_n \le 1$, we get

$$\|\overline{Z}^k - \tilde{Z}^k\|^2_{\mu,\beta} \le 3|\sigma|^2_\infty \mathbb{E}\,\|\partial_x H - \partial_x(\mathcal{P}^k_n H)\|^2_{H^\mu_{\beta,X}} + 3|\sigma|^2_\infty \mathbb{E}\,\|\partial_x u - \partial_x(\mathcal{P}^k_n u)\|^2_{H^\mu_{\beta,X}}$$
$$+ K_0'(T)\frac{T\lambda(B)\delta_n}{MNh_x^2} + \frac{K_2'(T)}{N^2h_x^2} + K_1'(T)\frac{T\lambda(B)\delta_n}{Mh_x^2}\left(\|Y - \tilde{Y}^{k-1}\|^2_{\mu,\beta} + \|Z - \tilde{Z}^{k-1}\|^2_{\mu,\beta}\right). \tag{14.4}$$

$K_0(T)$ depends on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, C_u, L_b, L_\sigma$ and on K, α_2 and α_1. $K_1(T)$ depends on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, L_f$ and the upper bounds for b and σ, and their derivatives, on c, K, α_1, α_2 and σ_0 and σ_1. $K_2(T)$ depends on the dimension d, μ, β, the upper bounds for f and Φ and for σ, b and their derivatives, on c, and on K, α_2 and α_1. For $i = 0, 1, 2$, $K_i'(T)$ depends on $|\partial_x K_x|_\infty$ and on the same parameters as $K_i(T)$.

14.2.1 Proof of (14.3) : A first decomposition

By using (14.2) page 178 and the definition of $(\tilde{Y}^k, \tilde{Z}^k)$ (see Definition 9.20, page 106), we can apply Remark 13.8, page 166 $\|\overline{Y}^k - \tilde{Y}^k\|^2_{\mu,\beta} = \mathbb{E}\,\|\overline{u}_k - u_k\|^2_{H^\mu_{\beta,X}}$. Let p denote the transition density of the process X. Conditioning w.r.t. $\mathcal{G}_{k-1}$, which contains all the random variables used to build u_{k-1}, leads to

$$\mathbb{E}\,\|\overline{u}_k - u_k\|^2_{H^\mu_{\beta,X}} = \mathbb{E}\left[\int_0^T ds e^{\beta s}\int_{\mathbb{R}^d} dx e^{-\mu|x|}\int_{\mathbb{R}^d}\mathbb{E}[|\overline{u}_k(s,y) - u_k(s,y)|^2\big|\mathcal{G}_{k-1}]p(0,x;s,y)dy\right],$$
$$= \mathbb{E}\left[\int_0^T ds e^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\mathbb{E}[|\overline{u}_k(s,y) - u_k(s,y)|^2\big|\mathcal{G}_{k-1}]dy\right], \tag{14.5}$$

where $\nu^0(s,y)$ has been introduced in Definition 6.11, page 67. The study of $\mathbb{E}\,\|\overline{u}_k - u_k\|^2_{H^\mu_{\beta,X}}$ will be done in three steps, corresponding to the three terms appearing in the following decomposition

Lemma 14.5. *We recall that $\mathcal{A}_k$ denotes the set of points $\{(T^k_i, X^k_i), 1 \le i \le n\}$ used by the operator $\mathcal{P}^k_n$. Then,*

$$\mathbb{E}[(\overline{u}_k - u_k)^2(s,y)\big|\mathcal{G}_{k-1}] = (\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}])^2 + Var(\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k]|\mathcal{G}_{k-1})$$
$$+ \mathbb{E}\left[Var((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)|\mathcal{G}_{k-1}\right].$$

Proof. First, we introduce $\mathcal{A}_k$ in the conditional expectation

$$\mathbb{E}[(\overline{u}_k - u_k)^2(s,y)\big|\mathcal{G}_{k-1}] = \mathbb{E}\left[\mathbb{E}[(\overline{u}_k - u_k)^2(s,y)\big|\mathcal{G}_{k-1},\mathcal{A}_k]\big|\mathcal{G}_{k-1}\right]. \tag{14.6}$$

Then, we use the bias-variance decomposition

$$\mathbb{E}[(\overline{u}_k - u_k)^2(s,y)\big|\mathcal{G}_{k-1},\mathcal{A}_k] = (\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k])^2 + \mathrm{Var}((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k).$$

Plugging this result in (14.6) leads to

$$\mathbb{E}[(\overline{u}_k - u_k)^2(s,y)\big|\mathcal{G}_{k-1}] = \mathbb{E}\left[(\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k])^2\,\big|\mathcal{G}_{k-1}\right]$$
$$+ \mathbb{E}\left[\mathrm{Var}((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)\big|\mathcal{G}_{k-1}\right]. \tag{14.7}$$

Using again the bias-variance decomposition for the first term of the above r.h.s. brings about

$$\mathbb{E}\left[(\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k])^2 \,\middle|\mathcal{G}_{k-1}\right] = (\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}])^2 + \mathrm{Var}(\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k]|\mathcal{G}_{k-1}).$$

Plugging the previous result in (14.7) yields the result. □

14.2.2 Proof of (14.3) : Study of $\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}]$

Proposition 14.6. *Assume that σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t\sigma \in C_b^1$ in space. We also assume that f and Φ are bounded. Then, there exists a function $K(T)$, non decreasing in T such that*

$$\mathbb{E}\Big[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)(\mathbb{E}[(\overline{u}_k - u_k)(s,y)\big|\mathcal{G}_{k-1}])^2dy\Big] \le \frac{K(T)}{N^2} + 3\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)(\mathbb{E}[(u-\mathcal{P}_n^k u)(s,y)])^2dy\right] + 3\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)(\mathbb{E}[(H-\mathcal{P}_n^k H)(s,y)\big|\mathcal{G}_{k-1}])^2dy\right].$$

$K(T)$ depends on the dimension d, μ,β, the upper bounds for f and Φ and for σ, b and their derivatives, on c, appearing in Theorem 8.1, page 78, and on K and α_2 defined in Proposition 6.4, page 63.

First, we recall

$$u_k(s,y) = \mathcal{P}_n^k(u_{k-1} + \overline{w}_k)(s,y), \tag{14.8}$$

$$\text{where } \overline{w}_k(s,y) = \frac{1}{M}\sum_{m=1}^M\Big[\Phi(X_T^{m,k,N}) - u_{k-1}(T,X_T^{m,k,N}) \tag{14.9}$$

$$+\int_s^T [f\left(r, X_r^{m,k,N}, u_{k-1}(r,X_r^{m,k,N}), (\partial_x u_{k-1}\sigma)(r,X_r^{m,k,N})\right) + (\partial_t + \mathcal{L}^N)u_{k-1}(r,X_r^{m,k,N})]dr\Big].$$

Before bounding $\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)(\mathbb{E}[(\overline{u}_k - u_k)(s,y)\big|\mathcal{G}_{k-1}])^2dy\right]$, we develop $\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}]$ in the following Lemma.

Lemma 14.7. *The following decomposition holds*

$$\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}] = \mathbb{E}[(u-\mathcal{P}_n^k u)(s,y)] + \mathbb{E}[(H-\mathcal{P}_n^k H)(s,y)|\mathcal{G}_{k-1}] + \mathbb{E}\left[\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1}\right], \tag{14.10}$$

where $\Delta(s,y) := \Phi(X_T^{N,s,y}) - \Phi(X_T^{s,y})$

$$+\int_s^T [f(r,X_r^N,u_{k-1}(r,X_r^N),(\partial_x u_{k-1}\sigma)(r,X_r^N)) - f(r,X_r,u_{k-1}(r,X_r),(\partial_x u_{k-1}\sigma)(r,X_r))]dr.$$

Moreover, if the assumptions of Theorem 8.1, page 78, are satisfied and if f and Φ are bounded,

$$|E[\Delta(s,y)|\mathcal{G}_{k-1}]| \leq \frac{K(T)}{N}\sqrt{T-s}(|\Phi|_\infty + (T-s)|f|_\infty),$$

where $K(T)$ is a function non decreasing in T and depending on the dimension d, the upper bounds for σ, b and their derivatives, and on c.

Proof of Proposition 14.6. By using Lemma 14.7, we easily deduce the last two terms of Proposition 14.6. It remains to show that

$$\mathbb{E}\left[\int_0^T ds e^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\left(\mathbb{E}[\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1}]\right)^2 dy\right] \leq \frac{K(T)}{N^2}.$$

To do so, we write

$$\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]) = \frac{r_n^\Delta(s,y)}{f_n(s,y)}g(2T\lambda(B)f_n(s,y)), \tag{14.11}$$

where $r_n^\Delta(s,y) = \frac{1}{nh_th_x^d}\sum_{i=1}^n K_t(\frac{s-T_i^k}{h_t})K_x(\frac{y-X_i^k}{h_x})\mathbb{E}[\Delta(T_i^k,X_i^k)|\mathcal{G}_{k-1}]$. Since g is bounded and $r_n^\Delta(s,y) \leq f_n(s,y)|\mathbb{E}[\Delta|\mathcal{G}_{k-1}]|_\infty$, using the second part of Lemma 14.7 yields

$$\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]) \leq \frac{K'(T)}{N}(|\Phi|_\infty + |f|_\infty),$$

where $K'(T)$ is a function non decreasing in T and depending on d and on the upper bounds for b and σ and their derivatives and on c. Estimate (12.16), page 143, yields the result. □

Proof of Lemma 14.7. Since $\overline{u}_k$ is $\mathcal{G}_{k-1}$-measurable, we get

$$\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}] = \overline{u}_k(s,y) - \mathbb{E}[\mathcal{P}_n^k(u_{k-1}+\overline{w}_k)(s,y)|\mathcal{G}_{k-1}]. \tag{14.12}$$

Conditioning w.r.t. $\mathcal{A}_k$ in the above expectation leads to

$$\mathbb{E}[\mathcal{P}_n^k(u_{k-1}+\overline{w}_k)(s,y)|\mathcal{G}_{k-1}] = \mathbb{E}\left[\mathcal{P}_n^k(u_{k-1}+\mathbb{E}[\overline{w}_k|\mathcal{G}_{k-1}])(s,y)|\mathcal{G}_{k-1}\right]. \tag{14.13}$$

Moreover, taking the conditional expectation w.r.t. $\mathcal{G}_{k-1}$ in the formula (14.9) gives

$$\mathbb{E}[\overline{w}_k(s,y)|\mathcal{G}_{k-1}] = \mathbb{E}[\Phi(X_T^{N,s,y}) - u_{k-1}(T,X_T^{N,s,y})|\mathcal{G}_{k-1}] +$$
$$\mathbb{E}[\int_s^T [f\left(r,X_r^{N,s,y},u_{k-1}(r,X_r^{N,s,y}),(\partial_x u_{k-1}\sigma)(r,X_r^{N,s,y})\right) + (\partial_t+\mathcal{L}^N)u_{k-1}(r,X_r^{N,s,y})]dr|\mathcal{G}_{k-1}].$$

By applying Itô's formula to u_{k-1} between s and T to the process $X^{N,s,y}$, we can simplify the above expression and get

$$\begin{aligned}(u_{k-1}+\mathbb{E}[\overline{w}_k|\mathcal{G}_{k-1}])(s,y) &= \mathbb{E}[\Phi(X_T^{N,s,y})] \\ &+ \mathbb{E}\left[\int_s^T f\left(r,X_r^{N,s,y},u_{k-1}(r,X_r^{N,s,y}),(\partial_x u_{k-1}\sigma)(r,X_r^{N,s,y})\right)dr|\mathcal{G}_{k-1}\right].\end{aligned} \tag{14.14}$$

Finally, applying Itô's formula to $\overline{u}_k$ between s and T to the process $X^{s,y}$ and using the PDE (14.1) page 178 satisfied by $\overline{u}_k$ lead to

$$\overline{u}_k(s,y) = \mathbb{E}\left[\Phi(X_T^{s,y}) + \int_s^T f\left(r, X_r^{s,y}, u_{k-1}(r, X_r^{s,y}), (\partial_x u_{k-1}\sigma)(r, X_r^{s,y})\right) dr | \mathcal{G}_{k-1}\right].$$

Plugging this term in (14.14) gives

$$\begin{aligned}(u_{k-1} + \mathbb{E}[\overline{w}_k|\mathcal{G}_{k-1}])(s,y) &= \overline{u}_k(s,y) + \mathbb{E}[\Phi(X_T^{N,s,y}) - \Phi(X_T^{s,y})] \\ &\quad + \mathbb{E}_{s,y}\left[\int_s^T [f(r, X_r^N, u_{k-1}(r, X_r^N), (\partial_x u_{k-1}\sigma)(r, X_r^N)) \right. \\ &\quad \left. - f(r, X_r, u_{k-1}(r, X_r), (\partial_x u_{k-1}\sigma)(r, X_r))]dr|\mathcal{G}_{k-1}\right] \\ &= \overline{u}_k(s,y) + \mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]. \qquad (14.15)\end{aligned}$$

Recalling that $\overline{u}_k = u + H$ and combining (14.12), (14.13) and (14.15) end the proof of (14.10).
Let us show the inequality $E[\Delta(s,y)|\mathcal{G}_{k-1}] \le \frac{K(T)}{N}\sqrt{T-s}(|\Phi|_\infty + |f|_\infty)$. To do so, we write

$$\begin{aligned}\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}] =& \mathbb{E}[\Phi(X_T^{N,s,y}) - \Phi(X_T^{s,y})] + \mathbb{E}\left[\int_s^T [f(r, X_r^N, u_{k-1}(r, X_r^N), (\partial_x u_{k-1}\sigma)(r, X_r^N))\right. \\ &\left. - f(r, X_r, u_{k-1}(r, X_r), (\partial_x u_{k-1}\sigma)(r, X_r))]dr|\mathcal{G}_{k-1}\right].\end{aligned}$$

Let p^N denote the transition density of X^N. Then, $\mathbb{E}[\Phi(X_T^{N,s,y}) - \Phi(X_T^{s,y})] = \int_{\mathbb{R}^d} \Phi(u)(p^N(s,y;T,u) - p(s,y;T,u))du$. By using Theorem 8.1, page 78, we get

$$\begin{aligned}\left|\mathbb{E}[\Phi(X_T^{N,s,y}) - \Phi(X_T^{s,y})]\right| &\le \frac{K(T)}{N}\sqrt{T-s}\int_{\mathbb{R}^d} |\Phi(u)| \frac{1}{(T-s)^{\frac{d}{2}}} \exp\left(-\frac{c|u-y|^2}{T-s}\right) du \\ &\le |\Phi|_\infty \frac{K(T)}{N}\sqrt{T-s}.\end{aligned}$$

The same proof holds for $\mathbb{E}[\int_s^T [f(r, X_r^N, u_{k-1}(r, X_r^N), (\partial_x u_{k-1}\sigma)(r, X_r^N)) - f(r, X_r, u_{k-1}(r, X_r), (\partial_x u_{k-1}\sigma)(r, X_r))]dr|\mathcal{G}_{k-1}]$:

$$\begin{aligned}&\left|\mathbb{E}_{s,y}\left[f(r, X_r^N, u_{k-1}(r, X_r^N), (\partial_x u_{k-1}\sigma)(r, X_r^N)) - f(r, X_r, u_{k-1}(r, X_r), (\partial_x u_{k-1}\sigma)(r, X_r))|\mathcal{G}_{k-1}\right]\right| \\ &= \left|\int_{\mathbb{R}^d} f(r, v, u_{k-1}(r,v), (\partial_x u_{k-1}\sigma)(r,v))(p^N(s,y;r,v) - p(s,y;r,v))dv\right| \\ &\le \frac{K(T)}{N}\frac{T-s}{\sqrt{r-s}}|f|_\infty.\end{aligned}$$

Integrating both sides w.r.t. r between s and T leads to the result. □

14.2.3 Proof of (14.3) : Study of Var$(\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k]|\mathcal{G}_{k-1})$

Proposition 14.8. *Assume that σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t\sigma \in C_b^1$ in space. We also assume that f and Φ are bounded. Then, there exists a function $K(T)$*

such that

$$\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\;\; Var(\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k]|\mathcal{G}_{k-1})dy] \le \frac{K(T)}{N^2}\right.$$
$$+ 3\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\, Var((u-\mathcal{P}_n^k u)(s,y))dy\right]$$
$$+ 3\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\, Var((H-\mathcal{P}_n^k H)(s,y)|\mathcal{G}_{k-1})dy\right].$$

$K(T)$ depends on the dimension d, μ, β, the upper bounds for f and Φ and for σ, b and their derivatives, on c and on K and α_2.

Before proving Proposition 14.8, we state the following Lemma.

Lemma 14.9. *Assume that σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t\sigma \in C_b^1$ in space. We also assume that f and Φ are bounded. Then, there exists a constant $K(T)$ non decreasing in T such that*

$$Var(\mathcal{P}_n^k\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]|\mathcal{G}_{k-1}) \le \frac{K(T)}{N^2}.$$

$K(T)$ depends on the dimension d, the upper bounds for f and Φ and for σ, b and their derivatives, and on c.

Remark 14.10. The upper bound stated in the previous Lemma is not optimal: it doesn't depend on n although $\mathrm{Var}((\mathcal{P}_n^k\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1})$ tends to 0 when n goes to infinity.

Proof of Proposition 14.8. We know that $\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k] = \overline{u}_k(s,y) - \mathcal{P}_n^k(u_{k-1}(s,y) + \mathbb{E}[\overline{w}_k(s,y)|\mathcal{G}_{k-1}])$. Using (14.15) the definition of $\overline{u}_k$ (see Proposition 14.3 page 178) and the definition of u_k (see 14.8 page 180) yields

$$\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k] = (u-\mathcal{P}_n^k u)(s,y) + (H-\mathcal{P}_n^k H)(s,y) - \mathcal{P}_n^k\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}],$$

and hence,

$$\begin{aligned}\mathrm{Var}(\mathbb{E}[(\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k]|\mathcal{G}_{k-1}) \le& 3\mathrm{Var}((u-\mathcal{P}_n^k u)(s,y)) \\ &+ 3\mathrm{Var}((H-\mathcal{P}_n^k H)(s,y)|\mathcal{G}_{k-1}) \\ &+ 3\mathrm{Var}(\mathcal{P}_n^k\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]|\mathcal{G}_{k-1}).\end{aligned}$$

Then, we use Lemma 14.9 to bound $\mathrm{Var}(\mathcal{P}_n^k\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]|\mathcal{G}_{k-1})$. Using (12.16) page 143 ends the proof. □

Proof of Lemma 14.9. We use (14.11) page 181 to develop $\mathrm{Var}(\mathcal{P}_n^k\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]|\mathcal{G}_{k-1})$.

$$\begin{aligned}\mathrm{Var}(\mathcal{P}_n^k\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]|\mathcal{G}_{k-1}) &= \mathrm{Var}\left(\frac{r_n^\Delta(s,y)}{f_n(s,y)}g(2T\lambda(B)f_n(s,y))|\mathcal{G}_{k-1}\right), \\ &\le \mathbb{E}\left[\left(\frac{r_n^\Delta(s,y)}{f_n(s,y)}g(2T\lambda(B)f_n(s,y))\right)^2|\mathcal{G}_{k-1}\right].\end{aligned}$$

g is bounded and $|r_n^\Delta(s,y)| \le f_n(s,y)|\mathbb{E}[\Delta|\mathcal{G}_{k-1}]|_\infty$. Using the second part of Lemma 14.7 page 180 ends the proof. □

14.2.4 Proof of (14.3) : Study of $\mathbb{E}\left[\mathbf{Var}((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)|\mathcal{G}_{k-1}\right]$

Proposition 14.11. *Assume that σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t\sigma \in C_b^1$ in space. We also assume that f is Lipschitz and $u, \partial_t u, \partial_x u$ and $\partial_x^2 u$ are bounded by a constant C_u. It holds*

$$\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\mathbb{E}\left[Var((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)|\mathcal{G}_{k-1}\right]dy\right] \leq K_0(T)\frac{T\lambda(B)\delta_n}{MN}$$
$$+K_1(T)\frac{T\lambda(B)\delta_n}{M}(\mathbb{E}\,\|u-u_{k-1}\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\,\|\partial_x u - \partial_x u_{k-1}\|^2_{H^\mu_{\beta,X}}).$$

$K_0(T)$ is a function depending on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, C_u, L_b, L_\sigma$ and on K and α_2. $K_1(T)$ is a function depending on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, L_f$ and the upper bounds for b and σ, and their derivatives, on c and on K, α_1, and α_2.

Remark 14.12. The above proposition characterises the use of the adaptive control variate method: $\frac{1}{M}$ is multiplied by $\mathbb{E}\,\|u-u_{k-1}\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\,\|\partial_x u - \partial_x u_{k-1}\|^2_{H^\mu_{\beta,X}}$. This term, among others, enables the geometric convergence of the algorithm.

Before proving Proposition 14.11, let us state two Lemmas.

Lemma 14.13. *Assume that f, b and σ are Lipschitz functions. We also assume that $u, \partial_t u, \partial_x u$ and $\partial_x^2 u$ are bounded by a constant C_u. There exists a function $K(T)$ non decreasing in T such that*

$$\mathbb{E}\left[Var((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)|\mathcal{G}_{k-1}\right] \leq K(T)\frac{T\lambda(B)\delta_n}{MN}$$
$$+\frac{32(T\lambda(B))^2(1+L_f^2)}{Mnh_t^2h_x^{2d}}\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)K_x^2\left(\frac{y-X_1^k}{h_x}\right)\mathbb{E}((\Theta^2+\Gamma^2)(T_1^k,X_1^k)|\mathcal{G}_{k-1},\mathcal{A}_k)|\mathcal{G}_{k-1}\right),$$

where $\Gamma(s,y) = \int_s^T(\partial_x u - \partial_x u_{k-1})(r, X_r^{N,s,y})\sigma(\varphi(r), X_{\varphi(r)}^{N,s,y})dW_r$, $\Theta(s,y) = \int_s^T(|u - u_{k-1}|(r, X_r^{N,s,y}) + |\partial_x u\sigma - \partial_x u_{k-1}\sigma|(r, X_r^{N,s,y}))dr$ and $K(T)$ depends on d, C_u, L_b, L_σ and the upper bounds for K_t and K_x.

Lemma 14.14. *Assume that σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t\sigma \in C_b^1$ in space. Let $\Gamma(s,y) = \int_s^T(\partial_x u - \partial_x u_{k-1})(r, X_r^{N,s,y})\sigma(\varphi(r), X_{\varphi(r)}^{N,s,y})dW_r$ and $\Theta(s,y) = \int_s^T(|u-u_{k-1}|(r, X_r^{N,s,y}) + |\partial_x u\sigma - \partial_x u_{k-1}\sigma|(r, X_r^{N,s,y}))dr$. Then,*

$$\mathbb{E}((\Theta^2+\Gamma^2)(T_1^k,X_1^k)|\mathcal{G}_{k-1},\mathcal{A}_k) \leq (1+|\sigma|^2_\infty)V_{k-1}(T_1^k,X_1^k),$$

where

$$V_{k-1}(s,y) = \mathbb{E}\left(\int_s^T(|u-u_{k-1}|^2(r,X_r^{s,y}) + |\partial_x u - \partial_x u_{k-1}|^2(r,X_r^{s,y}))dr|\mathcal{G}_{k-1}\right) \quad (14.16)$$
$$+\frac{K(T)(T-s)}{N}\times$$
$$\int_s^T\int_{\mathbb{R}^d}(|u-u_{k-1}|^2(r,z) + |\partial_x u - \partial_x u_{k-1}|^2(r,z))\frac{1}{(r-s)^{(d+1)/2}}\exp\left(-\frac{c|z-y|^2}{r-s}\right)dzdr,$$

$K(T)$ and c have been introduced in Theorem 8.1. $K(T)$ depends on d and the upper bounds for b and σ and their derivatives.

Proof of Proposition 14.11. We combine Lemmas 14.13 and 14.14 to do the proof. The first term in the r.h.s. of the Proposition comes from $K(T)\frac{T\lambda(B)\delta_n}{MN}$, which appears in Lemma 14.13, and from (12.16) page 143. By combining Lemmas 14.13 and 14.14, we notice that the remaining term to bound from above is $\frac{32(T\lambda(B))^2(1+L_f^2)}{Mnh_t^2h_x^{2d}}\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)K_x^2\left(\frac{y-X_1^k}{h_x}\right)V_{k-1}(T_1^k,X_1^k)|\mathcal{G}_{k-1}\right)dy\right]$. Since

$$\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)K_x^2\left(\frac{y-X_1^k}{h_x}\right)V_{k-1}(T_1^k,X_1^k)|\mathcal{G}_{k-1}\right)= \frac{1}{T\lambda(B)}\int_0^T drK_t^2\left(\frac{s-r}{h_t}\right)\int_B dzK_x^2\left(\frac{y-z}{h_x}\right)V_{k-1}(r,z),$$

we can apply Lemma 12.10, page 128, to get

$$\mathbb{E}\left[\int_0^T ds\ e^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)K_x^2\left(\frac{y-X_1^k}{h_x}\right)V_{k-1}(T_1^k,X_1^k)|\mathcal{G}_{k-1}\right)dy\right] \tag{14.17}$$
$$\leq K(T)h_th_x^d\mathbb{E}\left(\int_0^T dte^{\beta t}\int_{\mathbb{R}^d}dz\nu^0(t,z)V_{k-1}(t,z)\right),$$

where $K(T)$ depends on $d,\mu,K,\alpha_1,\alpha_2$. Let us bound from above $\mathbb{E}\left(\int_0^T dt\psi(t)\int_{\mathbb{R}^d}dz\nu^0(t,z)V_{k-1}(t,z)\right)$, where $V_{k-1}(t,z)$ is defined by (14.16).
First, we compute $\int_0^T dte^{\beta t}\int_{\mathbb{R}^d}dz\nu^0(t,z)\mathbb{E}\left(\int_t^T|u-u_{k-1}|^2(r,X_r^{t,z})dr|\mathcal{G}_{k-1}\right)$. To do so, we apply Lemma 12.11 page 129 with $\xi\equiv 1$ and $g_0\equiv u-u_{k-1}$. We obtain $\int_0^T dte^{\beta t}\int_{\mathbb{R}^d}dz\nu^0(t,z)\mathbb{E}\left(\int_t^T|u-u_{k-1}|^2(r,X_r^{t,z})dr|\mathcal{G}_{k-1}\right)\leq T\left\|u-u_{k-1}\right\|^2_{H^\mu_{\beta,X}}$. Taking the expectation of the above inequality leads to

$$\mathbb{E}\left(\int_0^T dte^{\beta t}\int_{\mathbb{R}^d}dz\nu^0(t,z)\mathbb{E}\left(\int_t^T|u-u_{k-1}|^2(r,X_r^{t,z})dr|\mathcal{G}_{k-1}\right)\right)\leq T\mathbb{E}\left\|u-u_{k-1}\right\|^2_{H^\mu_{\beta,X}}. \tag{14.18}$$

The same kind of proof leads to

$$\mathbb{E}\left(\int_0^T dte^{\beta t}\int_{\mathbb{R}^d}dz\nu^0(t,z)\mathbb{E}\left(\int_t^T|\partial_x u-\partial_x u_{k-1}|^2(r,X_r^{t,z})dr|\mathcal{G}_{k-1}\right)\right)\leq T\mathbb{E}\left\|\partial_x u-\partial_x u_{k-1}\right\|^2_{H^\mu_{\beta,X}}. \tag{14.19}$$

Second, we compute $I := \int_0^T dte^{\beta t}\int_{\mathbb{R}^d}dz\nu^0(t,z)\int_t^T dr\int_{\mathbb{R}^d}dy|u-u_{k-1}|^2(r,y)\frac{1}{(r-t)^{(d+1)/2}}\exp\left(-\frac{c|y-z|^2}{r-t}\right)$. By using successively the r.h.s. of (6.11) page 67, which gives $\nu^0(t,z)\leq 2^dKe^{c_2t}e^{-\mu|z|}$, Lemma 6.15 page 68, and the l.h.s. of (6.11), we deduce

$$\begin{aligned}\int_{\mathbb{R}^d}dz\nu^0(t,z)\frac{1}{(r-t)^{\frac{d+1}{2}}}\exp\left(-\frac{c|y-z|^2}{r-t}\right)&\leq 2^{2d}\left(\frac{\pi}{c}\right)^{d/2}\frac{K}{\sqrt{r-t}}e^{c_2t+d\mu^2/(4c)(r-t)}e^{-\mu|y|},\\ &\leq 2^{3d}\left(\frac{\pi}{c}\right)^{d/2}\frac{K^2}{\sqrt{r-t}}e^{c_2t-c_1r+\frac{d\mu^2}{4c}(r-t)}\nu^0(r,y).\end{aligned} \tag{14.20}$$

Introducing (14.20) into I leads to

$$I \leq K'(T) \int_0^T dte^{\beta t} \int_t^T dr \frac{1}{\sqrt{r-t}} \int_{\mathbb{R}^d} dy \nu^0(r,y)|u - u_{k-1}|^2(r,y),$$

where $K'(T)$ is non decreasing in T and depends on d, c, K, α_1, α_2 and μ. Since $t \leq r$, $e^{\beta t} \leq e^{\beta r}$, and $I \leq K'(T) \int_0^T dt \int_t^T dr \frac{1}{\sqrt{r-t}} e^{\beta r} \int_{\mathbb{R}^d} dy \nu^0(r,y)|u - u_{k-1}|^2(r,y)$. We get $I \leq K'(T) \int_0^T dre^{\beta r} \int_{\mathbb{R}^d} dy \nu^0(r,y)|u - u_{k-1}|^2(r,y)$. Taking the expectation yields

$$\mathbb{E}(I) \leq K'(T) \mathbb{E} \left\| u - u_{k-1} \right\|^2_{H^\mu_{\beta,X}}. \tag{14.21}$$

Defining
$I_x := \int_0^T dte^{\beta t} \int_{\mathbb{R}^d} dz \nu^0(t,z) \int_t^T \int_{\mathbb{R}^d} |\partial_x u - \partial_x u_{k-1}|^2(r,y) \frac{1}{(r-t)^{(d+1)/2}} \exp\left(-\frac{c|y-z|^2}{r-t}\right) dydr$, the same proof yields $\mathbb{E}(I_x) \leq K'(T) \mathbb{E} \left\| \partial_x u - \partial_x u_{k-1} \right\|^2_{H^\mu_{\beta,X}}$. Combining the previous result, (14.17), (14.18), (14.19) and (14.21) ends the proof. □

Proof of Lemma 14.13. First, we develop $\mathrm{Var}((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k)$. Since $u_k(s,y) = \mathcal{P}^k_n(u_{k-1} + \overline{w}_k)(s,y)$, where $\overline{w}_k(s,y)$ is defined by (14.9) page 180, we get

$$\mathrm{Var}((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k) = \mathrm{Var}(u_k(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k) = \mathrm{Var}(\mathcal{P}^k_n(\overline{w}_k)(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k).$$

We write $\mathcal{P}^k_n(\overline{w}_k)(s,y) = \frac{r^w_n(s,y)}{f_n(s,y)} g(2T\lambda(B) f_n(s,y))$,

where $r^w_n(s,y) = \frac{1}{nh_t h^d_x} \sum_{i=1}^n K_t(\frac{s-T^k_i}{h_t}) K_x(\frac{y-X^k_i}{h_x}) \overline{w}_k(T^k_i, X^k_i)$. Since $\frac{g(x)}{x}$ is bounded by $\frac{4\sqrt{2}}{3\sqrt{3}}$ and $g(2T\lambda(B) f_n(s,y)) \frac{1}{f_n(s,y)}$ is $\mathcal{A}_k$-measurable, we get

$$\mathrm{Var}((\overline{u}_k - u_k)(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k) \leq 8(T\lambda(B))^2 \mathrm{Var}(r^w_n(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k), \tag{14.22}$$

Moreover,

$$\mathrm{Var}(r^w_n(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k) = \tag{14.23}$$
$$\frac{1}{n^2 h^2_t h^{2d}_x} \sum_{i=1}^n K^2_t\left(\frac{s-T^k_i}{h_t}\right) K^2_x\left(\frac{y-X^k_i}{h_x}\right) \mathrm{Var}(\overline{w}_k(T^k_i, X^k_i)|\mathcal{G}_{k-1}, \mathcal{A}_k).$$

By applying Itô's formula to u between T^k_i and T to the process $(X^{N,T^k_i,X^k_i}_r)_{r \geq T^k_i}$ and to u_{k-1} between T^k_i and T to the process $(X^{N,T^k_i,X^k_i}_r)_{r \geq T^k_i}$, we get

$$\begin{aligned}
\mathrm{Var}(\overline{w}_k(T^k_i, X^k_i)|\mathcal{G}_{k-1}, \mathcal{A}_k) = \frac{1}{M} \mathrm{Var}\Big(& (u - u_{k-1})(T^k_i, X^k_i) \\
& + \int_{T^k_i}^T [f(r, X^N_r, u_{k-1}(r, X^N_r), (\partial_x u_{k-1}\sigma)(r, X^N_r)) + (\partial_t + \mathcal{L}^N) u(r, X^N_r)] dr \\
& + \int_{T^k_i}^T (\partial_x u(r, X^N_r) - \partial_x u_{k-1}(r, X^N_r)) \sigma(\varphi(r), X^N_{\varphi(r)}) dW_r | \mathcal{G}_{k-1}, \mathcal{A}_k \Big),
\end{aligned}$$
$$\leq \frac{2}{M} \mathbb{E}\left(\left(\int_{T^k_i}^T [f(r, X^N_r, u_{k-1}(r, X^N_r), (\partial_x u_{k-1}\sigma)(r, X^N_r)) + (\partial_t + \mathcal{L}^N) u(r, X^N_r)] dr \right)^2 | \mathcal{G}_{k-1}, \mathcal{A}_k \right)$$
$$+ \frac{2}{M} \mathbb{E}\left(\left(\int_{T^k_i}^T (\partial_x u(r, X^N_r) - \partial_x u_{k-1}(r, X^N_r)) \sigma(\varphi(r), X^N_{\varphi(r)}) dW_r \right)^2 | \mathcal{G}_{k-1}, \mathcal{A}_k \right).$$

To bound the first term, we introduce $f(r, X_r^N, u(r, X_r^N), (\partial_x u\sigma)(r, X_r^N))$ in the sum $f(r, X_r^N, u_{k-1}(r, X_r^N), (\partial_x u_{k-1}\sigma)(r, X_r^N)) + (\partial_t + \mathcal{L}^N)u(r, X_r^N)$. Since f is Lipschitz continuous with constant L_f and satisfies $(\partial_t + \mathcal{L})u + f = 0$, we get

$$\begin{aligned}
\big|f(r, X_r^N, u_{k-1}(r, X_r^N), &(\partial_x u_{k-1}\sigma)(r, X_r^N)) + (\partial_t + \mathcal{L}^N)u(r, X_r^N)\big| \leq \\
&L_f(|u - u_{k-1}|(r, X_r^N) + |\partial_x u_{k-1}\sigma - \partial_x u\sigma|(r, X_r^N)) \\
&+ \sum_{i=1}^d |b_i(r, X_r^N) - b_i(\varphi(r), X_{\varphi(r)}^N)||\partial_{x_i} u(r, X_r^N)| \\
&+ \frac{1}{2}\sum_{i,j=1}^d |[\sigma\sigma^*]_{ij}(r, X_r^N) - [\sigma\sigma^*]_{ij}(\varphi(r), X_{\varphi(r)}^N)||\partial^2_{x_i x_j} u(r, X_r^N)|.
\end{aligned}$$

Since f, b and σ are Lipschitz continuous and $u, \partial_t u, \partial_x u$ and $\partial_x^2 u$ are bounded, we get

$$\begin{aligned}
&|f(r, X_r^N, u_{k-1} + (r, X_r^N), (\partial_x u_{k-1}\sigma)(r, X_r^N)) + (\partial_t + \mathcal{L}^N)u(r, X_r^N)| \leq \\
&L_f\left(|u - u_{k-1}|(r, X_r^N) + |\partial_x u_{k-1}\sigma - \partial_x u\sigma|(r, X_r^N)\right) + dC_u(L_b + \frac{1}{2}L_\sigma^2)\left(\frac{T}{N} + |X_r^N - X_{\varphi(r)}^N|\right).
\end{aligned}$$

Using the definitions of θ, Γ and Lemma A.3, page 261 yields

$$\mathrm{Var}(\overline{w}_k(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k) \leq \frac{K'(T)}{MN} + \frac{4(L_f^2+1)}{M}\mathbb{E}((\Theta^2 + \Gamma^2)(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k), \tag{14.24}$$

where $K'(T)$ depends on d, C_u, L_b and L_σ. Combining (14.22), (14.23), (14.24), taking the conditional expectation w.r.t. $\mathcal{G}_{k-1}$ and applying the first part of Lemma 12.8 page 127 end the proof. □

Proof of Lemma 14.14.

$$\begin{aligned}
\mathbb{E}(\Gamma^2(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k) &= \mathbb{E}\left[\left(\int_{T_i^k}^T (\partial_x u - \partial_x u_{k-1})(r, X_r^N)\sigma(\varphi(r), X_{\varphi(r)}^N)dW_r\right)^2 |\mathcal{G}_{k-1}, \mathcal{A}_k\right], \\
&\leq |\sigma|_\infty^2 \mathbb{E}\left[\int_{T_i^k}^T |\partial_x u - \partial_x u_{k-1}|^2(r, X_r^N)dr|\mathcal{G}_{k-1}, \mathcal{A}_k\right].
\end{aligned}$$

Let us denote $V_{k-1}^0(T_i^k, X_i^k) := \mathbb{E}_{T_i^k, X_i^k}\left(\int_{T_i^k}^T |\partial_x u - \partial_x u_{k-1}|^2(r, X_r^N)dr|\mathcal{G}_{k-1}, \mathcal{A}_k\right)$. We develop

$$\begin{aligned}
V_{k-1}^0(s, y) &= \mathbb{E}\left(\int_s^T |\partial_x u - \partial_x u_{k-1}|^2(r, X_r^{N,s,y})dr|\mathcal{G}_{k-1}\right), \\
&= \int_s^T \int_{\mathbb{R}^d} |\partial_x u - \partial_x u_{k-1}|^2(r, z)p^N(s, y; r, z)dzdr.
\end{aligned}$$

Then, we write $p^N(s,y;r,z) = p(s,y;r,z) + (p^N(s,y;r,z) - p(s,y;r,z))$, and we use Theorem 8.1, page 78, to bound from above $p^N(s,y;r,z) - p(s,y;r,z)$.

$$V^0_{k-1}(s,y) \leq \mathbb{E}\left(\int_s^T |\partial_x u - \partial_x u_{k-1}|^2(r, X_r^{s,y})dr|\mathcal{G}_{k-1}\right)$$
$$+ \frac{K(T)(T-s)}{N}\int_s^T\int_{\mathbb{R}^d}|\partial_x u - \partial_x u_{k-1}|^2(r,z)\frac{1}{(r-s)^{(d+1)/2}}\exp\left(-\frac{c|z-y|^2}{r-s}\right)dzdr,$$

and we get the bound for $\mathbb{E}(\Gamma^2(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k)$. By doing the same kind of proof to bound $\mathbb{E}(\Theta^2(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k)$, we get the result. □

14.2.5 Proof of (14.3) : Conclusion

We recall $\|\overline{Y}^k - \tilde{Y}^k\|^2_{\mu,\beta} = \mathbb{E}\|\overline{u}_k - u_k\|^2_{H^\mu_{\beta,X}}$. Combining (14.5) page 179, Lemma 14.5 page 179, Proposition 14.6 page 180, Proposition 14.8 page 182 and Proposition 14.11 page 184 leads to

$$\|\overline{Y}^k - \tilde{Y}^k\|^2_{\mu,\beta} \leq 3\mathbb{E}\|H - \mathcal{P}^k_n H\|^2_{H^\mu_{\beta,X}} + 3\mathbb{E}\|u - \mathcal{P}^k_n u\|^2_{H^\mu_{\beta,X}} + K_0(T)\frac{T\lambda(B)\delta_n}{MN}$$
$$+ \frac{K_2(T)}{N^2} + K_1(T)\frac{T\lambda(B)\delta_n}{M}\left(\mathbb{E}\|u - u_{k-1}\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\|\partial_x u - \partial_x u_{k-1}\|^2_{H^\mu_{\beta,X}}\right). \quad (14.25)$$

$K_0(T)$ depends on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, C_u, L_b, L_\sigma, K$ and α_2. $K_1(T)$ is a function depending on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, L_f$ and the upper bounds for b and σ, and their derivatives, on c and on K, α_1, and α_2. $K_2(T)$ depends on the dimension d, μ, β, the upper bounds for f and Φ and for σ, b and their derivatives, on c, and on K and α_2.
We apply Remark 13.8, page 166 and we get $\mathbb{E}\|u - u_{k-1}\|^2_{H^\mu_{\beta,X}} = \|Y - \tilde{Y}^{k-1}\|^2_{\mu,\beta}$. To conclude, it remains to bound $\mathbb{E}\|\partial_x u - \partial_x u_{k-1}\|^2_{H^\mu_{\beta,X}}$ from above by $\|Z - \tilde{Z}^{k-1}\|^2_{\mu,\beta}$ up to a factor. We recall

$$\mathbb{E}\|\partial_x u - \partial_x u_{k-1}\|^2_{H^\mu_{\beta,X}} = \mathbb{E}\left(\int_0^T dte^{\beta t}\int_{\mathbb{R}^d} dz\nu^0(t,z)|\partial_x u(t,z) - \partial_x u_{k-1}(t,z)|^2 dr\right).$$

By using the definition of the ellipticity condition (see Definition 6.2, page 63), we get

$$|\partial_x u(t,z) - \partial_x u_{k-1}(t,z)|^2 \leq \frac{\sigma_0}{\sigma_1}|\sigma(t,z)(\partial_x u(t,z) - \partial_x u_{k-1}(t,z))|^2,$$

and then,

$$\mathbb{E}\|\partial_x u - \partial_x u_{k-1}\|^2_{H^\mu_{\beta,X}} \leq \frac{2\sigma_0}{\mu\sigma_1}\|Z - \tilde{Z}^{k-1}\|^2_{\mu,\beta}. \quad (14.26)$$

Plugging this result in (14.25) ends the proof of (14.3) page 178.

Let us now prove (14.4), page 179. We study $\|\overline{Z}^k - \tilde{Z}^k\|^2_{\mu,\beta} = \mathbb{E}\|\partial_x\overline{u}_k\sigma - \partial_x u_k\sigma\|_{H^\mu_{\beta,X}}$. Then,

$$\|\overline{Z}^k - \tilde{Z}^k\|^2_{\mu,\beta} \leq |\sigma|^2_\infty\mathbb{E}\|\partial_x\overline{u}_k - \partial_x u_k\|_{H^\mu_{\beta,X}}.$$

We follow the scheme of the proof of (14.3). Some results, as Lemma 14.5 page 179 and Lemma 14.7 page 180 can be easily adapted to the study of $\partial_x\overline{u}_k - \partial_x u_k$. In such a case, we don't do the proofs again.

14.2.6 Proof of (14.4): A first decomposition

As in (14.5) page 179, we get the following equality

$$\mathbb{E}\left\|\partial_x \overline{u}_k - \partial_x u_k\right\|^2_{H^\mu_{\beta,X}} = \mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\mathbb{E}[|\partial_x\overline{u}_k(s,y)-\partial_x u_k(s,y)|^2\big|\mathcal{G}_{k-1}]dy\right]. \tag{14.27}$$

The study of $\mathbb{E}\left\|\partial_x \overline{u}_k - \partial_x u_k\right\|^2_{H^\mu_{\beta,X}}$ will be done in three steps, corresponding to the three terms appearing in the following decomposition

Lemma 14.15. *We recall that $\mathcal{A}_k$ denotes the set of points $\{(T_i^k, X_i^k), 1\le i\le n\}$ used by the approximation $\mathcal{P}_n^k$. Then,*

$$\begin{aligned}\mathbb{E}[|\partial_x\overline{u}_k-\partial_x u_k|^2(s,y)\big|\mathcal{G}_{k-1}] =& |\mathbb{E}[(\partial_x\overline{u}_k-\partial_x u_k)(s,y)|\mathcal{G}_{k-1}]|^2\\ &+\sum_{i=1}^d Var(\mathbb{E}[(\partial_{x_i}\overline{u}_k-\partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k]|\mathcal{G}_{k-1})\\ &+\sum_{i=1}^d \mathbb{E}\left[Var((\partial_{x_i}\overline{u}_k-\partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)|\mathcal{G}_{k-1}\right].\end{aligned}$$

We prove this Lemma as we have proved Lemma 14.5 page 179.

14.2.7 Proof of (14.4): Study of $\mathbb{E}[(\partial_x\overline{u}_k-\partial_x u_k)(s,y)|\mathcal{G}_{k-1}]$

Proposition 14.16. *Assume σ is uniformly elliptic, $b\in C_b^{1,2}$ and $\sigma\in C_b^{1,3}$, $\partial_t\sigma\in C_b^1$ in space. We also assume f and Φ are bounded. Then, there exists a function $K(T)$, non decreasing in T such that*

$$\begin{aligned}\mathbb{E}\Big[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)|\mathbb{E}[(\partial_x\overline{u}_k-\partial_x u_k)(s,y)\big|\mathcal{G}_{k-1}]|^2dy\Big] \le& \frac{K(T)}{N^2h_x^2}\\ &+3\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)|\mathbb{E}[(\partial_x u-\partial_x(\mathcal{P}_n^k u))(s,y)]|^2dy\right]\\ &+3\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)|\mathbb{E}[(\partial_x H-\partial_x(\mathcal{P}_n^k H))(s,y)\big|\mathcal{G}_{k-1}]|^2dy\right].\end{aligned}$$

$K(T)$ depends on the dimension d, μ,β, the upper bounds for $\partial_x K_x, K_t, f$ and Φ and for σ, b and their derivatives, on c, and on K and α_2.

Before bounding $\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)|\mathbb{E}[(\partial_x\overline{u}_k-\partial_x u_k)(s,y)\big|\mathcal{G}_{k-1}]|^2dy\right]$, we develop $\mathbb{E}[(\partial_x\overline{u}_k-\partial_x u_k)(s,y)|\mathcal{G}_{k-1}]$ in the following Lemma.

Lemma 14.17. *The following decomposition holds*

$$\begin{aligned}\mathbb{E}[(\partial_x\overline{u}_k-\partial_x u_k)(s,y)|\mathcal{G}_{k-1}] =& \mathbb{E}[(\partial_x u-\partial_x(\mathcal{P}_n^k u))(s,y)]+\mathbb{E}[(\partial_x H-\partial_x(\mathcal{P}_n^k H))(s,y)|\mathcal{G}_{k-1}]\\ &+\mathbb{E}\left[\partial_x(\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]))|\mathcal{G}_{k-1}\right],\end{aligned}$$

where $\Delta(s,y) := \Phi(X_T^{N,s,y})-\Phi(X_T^{s,y})$

$$+\int_s^T f(r,X_r^N,u_{k-1}(r,X_r^N),(\partial_x u_{k-1}\sigma)(r,X_r^N)) - f(r,X_r,u_{k-1}(r,X_r),(\partial_x u_{k-1}\sigma)(r,X_r))dr.$$

Moreover, if the assumptions of Theorem 8.1, page 78, are satisfied,

$$E[\Delta(s,y)|\mathcal{G}_{k-1}] \leq \frac{K(T)}{N}\sqrt{T-s}(|\Phi|_\infty + (T-s)|f|_\infty),$$

where $K(T)$ depends on the dimension d, the upper bounds for σ, b and their derivatives, and on c.

We prove Lemma 14.17 as Lemma 14.7 page 180.

Proof of Proposition 14.16. By using Lemma 14.17, we easily deduce the last two terms of Proposition 14.16. It remains to show that

$$\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\left|\mathbb{E}[\partial_x(\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]))|\mathcal{G}_{k-1}]\right|^2 dy\right] \leq \frac{K(T)}{N^2h_x^2}.$$

To do so, we use (14.11), page 181, and differentiate it.

$$\begin{aligned}\partial_{x_i}(\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])) &= \frac{\partial_{x_i}r_n^\Delta(s,y)}{f_n(s,y)}g(2T\lambda(B)f_n(s,y)) \qquad (14.28)\\ &+ 2T\lambda(B)\frac{r_n^\Delta(s,y)}{f_n(s,y)}g'(2T\lambda(B)f_n(s,y))\partial_{x_i}f_n(s,y) - \frac{r_n^\Delta(s,y)}{f_n^2(s,y)}g(2T\lambda(B)f_n(s,y))\partial_{x_i}f_n(s,y),\end{aligned}$$

where $r_n^\Delta(s,y) = \frac{1}{nh_th_x^d}\sum_{i=1}^n K_t(\frac{s-T_i^k}{h_t})K_x(\frac{y-X_i^k}{h_x})\mathbb{E}[\Delta(T_i^k,X_i^k)|\mathcal{G}_{k-1}]$. Since $\frac{g(x)}{x}$ and $g'(x)$ are bounded by 2, we get

$$\left|\partial_{x_i}(\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]))\right| \leq 4T\lambda(B)|\partial_{x_i}r_n^\Delta(s,y)| + 8T\lambda(B)\left|\frac{r_n^\Delta(s,y)}{f_n(s,y)}\right||\partial_{x_i}f_n(s,y)|.$$

Since $|r_n^\Delta(s,y)| \leq |f_n(s,y)||\mathbb{E}[\Delta|\mathcal{G}_{k-1}]|_\infty$ and $|\partial_{x_i}r_n^\Delta(s,y)| \leq |\partial_{x_i}f_n(s,y)||\mathbb{E}[\Delta|\mathcal{G}_{k-1}]|_\infty$, we combine Lemma 12.17 page 135 and the second part of Lemma 14.17 to get

$$\left|\mathbb{E}[\partial_x(\mathcal{P}_n^k(\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]))]\right|^2 \leq \frac{K'(T)}{N^2h_x^2}|K_t|_\infty^2|\partial_xK_x|_\infty^2(|\Phi|_\infty^2 + |f|_\infty^2),$$

where $K'(T)$ is a function non decreasing in T and depending on d and on the upper bounds for b and σ and their derivatives and on c. By using (12.16) page 143, the result easily follows. $\square$

14.2.8 Proof of (14.4): Study of Var$(\mathbb{E}[(\partial_{x_i}\overline{u}_k - \partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k]|\mathcal{G}_{k-1})$

Proposition 14.18. *Assume σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t\sigma \in C_b^1$ in space. We also assume f and Φ are bounded. Then, there exists a function $K(T)$ such that*

$$\begin{aligned}\mathbb{E}\Big[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\,Var(\mathbb{E}[(\partial_{x_i}\overline{u}_k - \partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k]|\mathcal{G}_{k-1})dy\Big] \leq \\ \frac{K(T)}{N^2h_x^2} + 3\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\,Var((\partial_{x_i}u - \partial_{x_i}(\mathcal{P}_n^ku))(s,y))dy\right] \\ + 3\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\,Var((\partial_{x_i}H - \partial_{x_i}(\mathcal{P}_n^kH))(s,y)|\mathcal{G}_{k-1})dy\right].\end{aligned}$$

$K(T)$ depends on the dimension d, μ, β, the upper bounds for $\partial_{x_i}K_x, K_t, f$ and Φ and for σ, b and their derivatives, on c, K and α_2.

Before proving Proposition 14.18, we state the following Lemma.

Lemma 14.19. *Assume σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t \sigma \in C_b^1$ in space. We also assume f and Φ are bounded. Then, there exists a constant $K(T)$ such that*

$$Var(\partial_{x_i}(\mathcal{P}_n^k \mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1}) \leq \frac{K(T)}{N^2 h_x^2}.$$

$K(T)$ depends on the dimension d, the upper bounds for $\partial_{x_i} K_x, K_t, f$ and Φ and for σ, b and their derivatives, and on c.

Remark 14.20. The upper bound stated in the previous Lemma is not optimal: it doesn't depend on n although $\text{Var}(\partial_{x_i}(\mathcal{P}_n^k \mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1})$ tends to 0 when n goes to infinity.

Proof of Proposition 14.18. We know that $\mathbb{E}[(\partial_x \overline{u}_k - \partial_x u_k)(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k] = \partial_x \overline{u}_k(s,y) - \partial_x(\mathcal{P}_n^k(u_{k-1}(s,y) + \mathbb{E}[\overline{w}_k(s,y)|\mathcal{G}_{k-1}]))$. Using (14.15), page 182, and the definition of $\overline{u}_k$ (see Proposition 14.3, page 178) yields

$$\begin{aligned}\mathbb{E}[(\partial_x \overline{u}_k - \partial_x u_k)(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k] =& (\partial_x u - \partial_x(\mathcal{P}_n^k u))(s,y) + (\partial_x H - \partial_x(\mathcal{P}_n^k H))(s,y) \\ &- \partial_x(\mathcal{P}_n^k \mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}]),\end{aligned}$$

and hence,

$$\begin{aligned}\text{Var}(\mathbb{E}[(\partial_{x_i} \overline{u}_k - \partial_{x_i} u_k)(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k]|\mathcal{G}_{k-1}) \leq& 3\text{Var}((\partial_{x_i} u - \partial_{x_i}(\mathcal{P}_n^k u))(s,y)) \\ &+ 3\text{Var}((\partial_{x_i} H - \partial_{x_i}(\mathcal{P}_n^k H))(s,y)|\mathcal{G}_{k-1}) \\ &+ 3\text{Var}(\partial_{x_i}(\mathcal{P}_n^k \mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1}).\end{aligned}$$

Then, we use Lemma 14.19 to bound $\text{Var}(\partial_{x_i}(\mathcal{P}_n^k \mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1})$. Using (12.16), page 143, ends the proof. $\square$

Proof of Lemma 14.19. We use (14.28) to develop $\text{Var}(\partial_{x_i}(\mathcal{P}_n^k \mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1})$.

$$\begin{aligned}\text{Var}(\partial_{x_i}(\mathcal{P}_n^k \mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1}) \leq& 3\text{Var}\left(\frac{\partial_{x_i} r_n^\Delta(s,y)}{f_n(s,y)} g(2T\lambda(B) f_n(s,y))|\mathcal{G}_{k-1}\right) \\ &+ 12(T\lambda(B))^2 \text{Var}\left(\frac{r_n^\Delta(s,y)}{f_n(s,y)} g'(2T\lambda(B) f_n(s,y)) \partial_{x_i} f_n(s,y)|\mathcal{G}_{k-1}\right) \\ &+ 3\text{Var}\left(\frac{r_n^\Delta(s,y)}{f_n^2(s,y)} g(2T\lambda(B) f_n(s,y)) \partial_{x_i} f_n(s,y)|\mathcal{G}_{k-1}\right),\end{aligned} \tag{14.29}$$

Since $\frac{g(x)}{x}$ is bounded by 2, we get

$$\begin{aligned}\text{Var}\left(\frac{\partial_{x_i} r_n^\Delta(s,y)}{f_n(s,y)} g(2T\lambda(B) f_n(s,y))|\mathcal{G}_{k-1}\right) &\leq \mathbb{E}\left(\left|\frac{\partial_{x_i} r_n^\Delta(s,y)}{f_n(s,y)} g(2T\lambda(B) f_n(s,y))\right|^2 |\mathcal{G}_{k-1}\right) \\ &\leq 8(T\lambda(B))^2 \mathbb{E}(|\partial_{x_i}(r_n^\Delta(s,y))|^2|\mathcal{G}_{k-1}).\end{aligned} \tag{14.30}$$

Since g' is bounded by 2 and $r_n^\Delta(s,y) \leq f_n(s,y)|\mathbb{E}[\Delta|\mathcal{G}_{k-1}]|_\infty$,

$$\text{Var}\left(\frac{r_n^\Delta(s,y)}{f_n(s,y)} g'(2T\lambda(B) f_n(s,y)) \partial_{x_i} f_n(s,y)|\mathcal{G}_{k-1}\right) \leq 4|\mathbb{E}[\Delta|\mathcal{G}_{k-1}]|_\infty^2 \mathbb{E}(\partial_{x_i} f_n(s,y))^2. \tag{14.31}$$

Analogously, one has

$$\mathrm{Var}\left(\frac{r_n^{\Delta}(s,y)}{f_n^2(s,y)}g(2T\lambda(B)f_n(s,y))\partial_{x_i}f_n(s,y)|\mathcal{G}_{k-1}\right) \leq 8(T\lambda(B))^2|\mathbb{E}[\Delta|\mathcal{G}_{k-1}]|_\infty^2\mathbb{E}(\partial_{x_i}f_n(s,y))^2. \tag{14.32}$$

Plugging (14.30), (14.31) and (14.32) into (14.29) yields

$$\begin{aligned}\mathrm{Var}(\partial_{x_i}(\mathcal{P}_n^k\mathbb{E}[\Delta(s,y)|\mathcal{G}_{k-1}])|\mathcal{G}_{k-1}) \leq & 24(T\lambda(B))^2\mathbb{E}(|\partial_{x_i}(r_n^{\Delta}(s,y))|^2|\mathcal{G}_{k-1}) \\ & + 72(T\lambda(B))^2|\mathbb{E}[\Delta|\mathcal{G}_{k-1}]|_\infty\mathbb{E}(\partial_{x_i}f_n(s,y))^2.\end{aligned}$$

Using Lemma 12.17 page 135 and the second part of Lemma 14.17 page 189 ends the proof. □

14.2.9 Proof of (14.4): Study of $\mathbb{E}\left[\mathbf{Var}((\partial_{x_i}\overline{u}_k - \partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)|\mathcal{G}_{k-1}\right]$

Proposition 14.21. *Assume σ is uniformly elliptic, $b \in C_b^{1,2}$ and $\sigma \in C_b^{1,3}$, $\partial_t\sigma \in C_b^1$ in space. We also assume that f is Lipschitz and $u, \partial_t u, \partial_x u$ and $\partial_x^2 u$ are bounded by a constant C_u. It holds*

$$\begin{aligned}&\mathbb{E}\left[\int_0^T dse^{\beta s}\int_{\mathbb{R}^d}\nu^0(s,y)\mathbb{E}\left[Var((\partial_{x_i}\overline{u}_k - \partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)|\mathcal{G}_{k-1}\right]dy\right] \leq \\ &K_0(T)\frac{T\lambda(B)\delta_n}{MNh_x^2} + K_1(T)\frac{T\lambda(B)\delta_n}{Mh_x^2}\left(\mathbb{E}\left\|u-u_{k-1}\right\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\left\|\partial_x u - \partial_x u_{k-1}\right\|^2_{H^\mu_{\beta,X}}\right).\end{aligned}$$

$K_0(T)$ is a function depending on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, |\partial_{x_i}K_x|_\infty, C_u, L_b, L_\sigma, K$ and α_2. $K_1(T)$ is a function depending on $\mu, \beta, d, \alpha, |K_t|_\infty, |K_x|_\infty$, $|\partial_{x_i}K_x|_\infty, L_f$ and the upper bounds for b and σ, and their derivatives, on c, K, α_1, and α_2.

Before proving Proposition 14.21, let us state the following Lemma.

Lemma 14.22. *Assume f, b and σ are Lipschitz, and $T\lambda(B)\delta_n << 1$. We also assume that $u, \partial_t u, \partial_x u$ and $\partial_x^2 u$ are bounded by a constant C_u. There exists a function $K(T)$ such that*

$$\begin{aligned}&\mathbb{E}\left[Var\left((\partial_{x_i}\overline{u}_k - \partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k\right)|\mathcal{G}_{k-1}\right] \leq K(T)\frac{T\lambda(B)\delta_n}{MNh_x^2} \\ &\frac{C(T\lambda(B))^2}{Mnh_t^2h_x^{2d+2}}\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)(\partial_{x_i}K_x)^2\left(\frac{y-X_1^k}{h_x}\right)\mathbb{E}((\Theta^2+\Gamma^2)(T_1^k,X_1^k)|\mathcal{G}_{k-1},\mathcal{A}_k)\right) \\ &+\frac{(T\lambda(B))^4\delta_n^2}{Mnh_t^2h_x^{2d+2}}\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)(K_x\partial_{x_i}K_x)^2\left(\frac{y-X_1^k}{h_x}\right)\mathbb{E}((\Theta^2+\Gamma^2)(T_1^k,X_1^k)|\mathcal{G}_{k-1},\mathcal{A}_k)\right) \\ &+\frac{C_0(T\lambda(B))^2}{Mnh_t^2h_x^{2d+2}}\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)K_x^2\left(\frac{y-X_1^k}{h_x}\right)\mathbb{E}((\Theta^2+\Gamma^2)(T_1^k,X_1^k)|\mathcal{G}_{k-1},\mathcal{A}_k)\right) \\ &+\frac{C_0(T\lambda(B))^3\delta_n}{Mnh_t^2h_x^{2d+2}}\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)(K_x^2\partial_{x_i}K_x)\left(\frac{y-X_1^k}{h_x}\right)\mathbb{E}((\Theta^2+\Gamma^2)(T_1^k,X_1^k)|\mathcal{G}_{k-1},\mathcal{A}_k)\right),\end{aligned}$$

where C_0 depends on $L_f, |K_t|_\infty$ and $|\partial_{x_i}K_x|_\infty$. $\Gamma(s,y) = \int_s^T \sigma(\varphi(r), X^{N,s,y}_{\varphi(r)})(\partial_x u - \partial_x u_{k-1})(r, X_r^{N,s,y})dW_r$ and $\Theta(s,y) = \int_s^T |u-u_{k-1}|(r,X_r^{N,s,y}) + |\partial_x u\sigma - \partial_x u_{k-1}\sigma|(r,X_r^{N,s,y})dr$. $K(T)$ depends on d, C_u, L_b, L_σ and the upper bounds for K_t, K_x $\partial_{x_i}K_x$, and C depends on L_f.

Proof of Proposition 14.21. We combine Lemma 14.22 and Lemma 14.14 page 184 to do the proof. The first term in the r.h.s. of the Proposition comes from $K(T)\frac{T\lambda(B)\delta_n}{MNh_x^2}$, the first term of the r.h.s. of Lemma 14.22, and from (12.16) page 143. The second term appearing in Proposition 14.21 ensues from the four other terms of the r.h.s. of Lemma 14.22. We combine Lemma 14.14, the second part of Lemma 12.9 page 128 and the proof of Proposition 14.11 page 184 to get it. □

Proof of Lemma 14.22. First, we develop $\mathrm{Var}((\partial_{x_i}\overline{u}_k - \partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)$. Since $\partial_{x_i}u_k(s,y) = \partial_{x_i}(\mathcal{P}_n^k(u_{k-1}+\overline{w}_k)(s,y))$, where $\overline{w}_k(s,y)$ is defined by (14.9) page 180, we get

$$\begin{aligned}\mathrm{Var}((\partial_{x_i}\overline{u}_k - \partial_{x_i}u_k)(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k) &= \mathrm{Var}(\partial_{x_i}u_k(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k)\\ &= \mathrm{Var}(\partial_{x_i}(\mathcal{P}_n^k(\overline{w}_k)(s,y))|\mathcal{G}_{k-1},\mathcal{A}_k).\end{aligned}$$

We use (14.28) page 190 to write

$$\begin{aligned}\partial_{x_i}(\mathcal{P}_n^k(\overline{w}_k)(s,y)) &= \frac{\partial_{x_i}r_n^w(s,y)}{f_n(s,y)}g(2T\lambda(B)f_n(s,y))\\ &+ 2T\lambda(B)\frac{r_n^w(s,y)}{f_n(s,y)}g'(2T\lambda(B)f_n(s,y))\partial_{x_i}f_n(s,y) - \frac{r_n^w(s,y)}{f_n^2(s,y)}g(2T\lambda(B)f_n(s,y))\partial_{x_i}f_n(s,y),\end{aligned}$$

where $r_n^w(s,y) = \frac{1}{nh_th_x^d}\sum_{i=1}^n K_t(\frac{s-T_i^k}{h_t})K_x(\frac{y-X_i^k}{h_x})\overline{w}_k(T_i^k,X_i^k)$. Then,

$$\begin{aligned}\mathrm{Var}(\partial_{x_i}(\mathcal{P}_n^k(\overline{w}_k)(s,y))|\mathcal{G}_{k-1},\mathcal{A}_k) &\le 3\mathrm{Var}\left(\frac{\partial_{x_i}r_n^w(s,y)}{f_n(s,y)}g(2T\lambda(B)f_n(s,y))|\mathcal{G}_{k-1},\mathcal{A}_k\right)\\ &+ 12(T\lambda(B))^2\mathrm{Var}\left(\frac{r_n^w(s,y)}{f_n(s,y)}g'(2T\lambda(B)f_n(s,y))\partial_{x_i}f_n(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k\right)\\ &+ 3\mathrm{Var}\left(\frac{r_n^w(s,y)}{f_n^2(s,y)}g(2T\lambda(B)f_n(s,y))\partial_{x_i}f_n(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k\right).\end{aligned}$$

Let us bound from above the three following terms

- $\mathrm{Var}\left(\frac{\partial_{x_i}r_n^w(s,y)}{f_n(s,y)}g(2T\lambda(B)f_n(s,y))|\mathcal{G}_{k-1},\mathcal{A}_k\right)$
  Since $\frac{g(x)}{x}$ is bounded by 2 and $\frac{g(2T\lambda(B)f_n(s,y))}{f_n(s,y)}$ is $\mathcal{A}_k$-measurable, we get

$$\mathrm{Var}\left(\frac{\partial_{x_i}r_n^w(s,y)}{f_n(s,y)}g(2T\lambda(B)f_n(s,y))|\mathcal{G}_{k-1},\mathcal{A}_k\right) \le 16(T\lambda(B))^2\mathrm{Var}(\partial_{x_i}r_n^w(s,y)|\mathcal{G}_{k-1},\mathcal{A}_k), \tag{14.33}$$

  and $\mathrm{Var}(\partial_{x_i}r_n^w(s,y)) =$

$$\frac{1}{n^2h_t^2h_x^{2d+2}}\sum_{i=1}^n K_t^2\left(\frac{s-T_i^k}{h_t}\right)(\partial_{x_i}K_x)^2\left(\frac{y-X_i^k}{h_x}\right)\mathrm{Var}(\overline{w}_k(T_i^k,X_i^k)|\mathcal{G}_{k-1},\mathcal{A}_k). \tag{14.34}$$

- $\mathrm{Var}\Big(\frac{r_n^w(s,y)}{f_n(s,y)} g'(2T\lambda(B) f_n(s,y)) \partial_{x_i} f_n(s,y) | \mathcal{G}_{k-1}, \mathcal{A}_k\Big)$
  Since $\frac{g'(x)}{x}$ is bounded by 4, we obtain

$$\mathrm{Var}\left(\frac{r_n^w(s,y)}{f_n(s,y)} g'(2T\lambda(B) f_n(s,y))\ \partial_{x_i} f_n(s,y) | \mathcal{G}_{k-1}, \mathcal{A}_k\right) \leq$$
$$64(T\lambda(B))^2 |\partial_{x_i} f_n(s,y)|^2 \mathrm{Var}(r_n^w(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k).$$

- $\mathrm{Var}\Big(\frac{r_n^w(s,y)}{f_n^2(s,y)} g(2T\lambda(B) f_n(s,y)) \partial_{x_i} f_n(s,y) | \mathcal{G}_{k-1}, \mathcal{A}_k\Big)$
  Since $\frac{g(x)}{x^2}$ is bounded by 2, we get

$$\mathrm{Var}\left(\frac{r_n^w(s,y)}{f_n^2(s,y)} g(2T\lambda(B) f_n(s,y))\ \partial_{x_i} f_n(s,y) | \mathcal{G}_{k-1}, \mathcal{A}_k\right) \leq$$
$$32(T\lambda(B))^4 |\partial_{x_i} f_n(s,y)|^2 \mathrm{Var}(r_n^w(s,y)|\mathcal{G}_{k-1}, \mathcal{A}_k).$$

Combining (14.23) page 186, (14.33) and (14.34) yields

$$\begin{aligned}
&\mathrm{Var}(\partial_{x_i}(\mathcal{P}_n^k(\overline{w}_k)(s,y))|\mathcal{G}_{k-1}, \mathcal{A}_k) \leq \\
&\frac{48(T\lambda(B))^2}{n^2 h_t^2 h_x^{2d+2}} \sum_{i=1}^n K_t^2\left(\frac{s-T_i^k}{h_t}\right)(\partial_{x_i} K_x)^2\left(\frac{y-X_i^k}{h_x}\right) \mathrm{Var}(\overline{w}_k(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k) \\
&+ \frac{864(T\lambda(B))^4 |\partial_{x_i} f_n(s,y)|^2}{n^2 h_t^2 h_x^{2d}} \sum_{i=1}^n K_t^2\left(\frac{s-T_i^k}{h_t}\right) K_x^2\left(\frac{y-X_i^k}{h_x}\right) \mathrm{Var}(\overline{w}_k(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k).
\end{aligned}$$

By taking the conditional expectation w.r.t. $\mathcal{G}_{k-1}$, using (14.24) page 187 and Lemma 12.9, page 128, we get

$$\begin{aligned}
&\mathbb{E}\Big(\mathrm{Var}(\partial_{x_i}(\mathcal{P}_n^k(\overline{w}_k)(s,y))|\mathcal{G}_{k-1}, \mathcal{A}_k)\Big) \leq K(T)\frac{T\lambda(B)\delta_n}{MNh_x^2} \\
&\frac{C(T\lambda(B))^2(1+L_f^2)}{Mnh_t^2 h_x^{2d+2}} \mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)(\partial_{x_i} K_x)^2\left(\frac{y-X_1^k}{h_x}\right) \mathbb{E}((\Theta^2+\Gamma^2)(T_1^k, X_1^k)|\mathcal{G}_{k-1}, \mathcal{A}_k)\right) \\
&+ C\frac{(T\lambda(B))^4}{n^2 h_t^2 h_x^{2d}} \mathbb{E}\left(|\partial_{x_i} f_n(s,y)|^2 \sum_{i=1}^n K_t^2\left(\frac{s-T_i^k}{h_t}\right) K_x^2\left(\frac{y-X_i^k}{h_x}\right) \mathrm{Var}(\overline{w}_k(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k)\right).
\end{aligned}$$

It remains to apply Proposition 12.22, page 137 to get

$$\begin{aligned}
&\frac{(T\lambda(B))^4}{n^2 h_t^2 h_x^{2d}} \mathbb{E}\left(|\partial_{x_i} f_n(s,y)|^2 \sum_{i=1}^n K_t^2\left(\frac{s-T_i^k}{h_t}\right) K_x^2\left(\frac{y-X_i^k}{h_x}\right) \mathrm{Var}(\overline{w}_k(T_i^k, X_i^k)|\mathcal{G}_{k-1}, \mathcal{A}_k)\right) \leq \\
&\frac{(T\lambda(B))^4 \delta_n^2}{nh_t^2 h_x^{2d+2}} \mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right)(K_x \partial_{x_i} K_x)^2\left(\frac{y-X_1^k}{h_x}\right) \mathrm{Var}(\overline{w}_k(T_1^k, X_1^k)|\mathcal{G}_{k-1}, \mathcal{A}_k)\right) \\
&+ \frac{C_0(T\lambda(B))^2}{nh_t^2 h_x^{2d+2}}(1+T\lambda(B)\delta_n)\mathbb{E}\left(K_t^2\left(\frac{s-T_1^k}{h_t}\right) K_x^2\left(\frac{y-X_1^k}{h_x}\right) \mathrm{Var}(\overline{w}_k(T_1^k, X_1^k)|\mathcal{G}_{k-1}, \mathcal{A}_k)\right) \\
&+ \frac{C_0(T\lambda(B))^3\delta_n}{nh_t^2 h_x^{2d+2}} \mathbb{E}\left(K_t^3\left(\frac{s-T_1^k}{h_t}\right)(K_x^2 \partial_{x_i} K_x)\left(\frac{y-X_1^k}{h_x}\right) \mathrm{Var}(\overline{w}_k(T_1^k, X_1^k)|\mathcal{G}_{k-1}, \mathcal{A}_k)\right),
\end{aligned}$$

where C_0 depends on $|K_t|_\infty$ and $|\partial_{x_i} K_x|_\infty$. Applying (14.24) page 187 and Lemma 12.9 page 128 ends the proof. $\square$

14.2.10 Proof of (14.4) : Conclusion

We recall $\|\overline{Z}^k - \tilde{Z}^k\|_{\mu,\beta}^2 \leq |\sigma|_\infty^2 \mathbb{E}\,\|\partial_x \overline{u}_k - \partial_x u_k\|^2_{H^\mu_{\beta,X}}$. Combining (14.27) page 189, Lemma 14.15 page 189, Proposition 14.16 page 189, Proposition 14.18 page 190 and Proposition 14.21 page 192 leads to

$$\begin{aligned}\|\overline{Z}^k - \tilde{Z}^k\|_{\mu,\beta}^2 \leq & 3|\sigma|_\infty^2 \mathbb{E}\,\|\partial_x H - \partial_x(\mathcal{P}^k_n H)\|^2_{H^\mu_{\beta,X}} + 3|\sigma|_\infty^2 \mathbb{E}\,\|\partial_x u - \partial_x(\mathcal{P}^k_n u)\|^2_{H^\mu_{\beta,X}} \\ & + K_0(T)\frac{T\lambda(B)\delta_n}{MNh_x^2} + \frac{K_2(T)}{N^2 h_x^2} \\ & + K_1(T)\frac{T\lambda(B)\delta_n}{Mh_x^2}\left(\mathbb{E}\,\|u - u_{k-1}\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\,\|\partial_x u - \partial_x u_{k-1}\|^2_{H^\mu_{\beta,X}}\right),\end{aligned}$$

$K_0(T)$ depends on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, |\partial_{x_i} K_x|_\infty, C_u, L_b, L_\sigma, K$ and α_2. $K_1(T)$ depends on $\mu, \beta, d, |K_t|_\infty, |K_x|_\infty, |\partial_{x_i} K_x|_\infty, L_f$ and the upper bounds for b and σ, and their derivatives, on c, K, α_1, and α_2. $K_2(T)$ depends on the dimension d, μ, β, the upper bounds for $|K_t|_\infty, |K_x|_\infty, |\partial_{x_i} K_x|_\infty, f$ and Φ and for σ, b and their derivatives, on c, K and α_2. We end this step as we did in section 14.2.5, page 188.

Chapter 15

Numerical Experiments

This part is devoted to the illustration of the results stated in Chapter 13, and more particularly in Theorem 13.10, page 166. Before presenting the numerical experiments we have done, we mention that the source code has been done in C++, and the graphs have been drawn using the *Scilab* and Nsp (see Chancelier et al. [19]) softwares.
First, we discuss the choice of the parameters a, n, h_x, h_t, M and N and the choice of the kernel function. Second, we focus on the numerical convergence of $Y - Y^k$ and $Z - Z^k$ to 0 for the $\|\cdot\|_{\mu,\beta}$ norm, when the parameters $N, n, a, h_x^{-1}, h_t^{-1}$ and M go to infinity. To do so, one needs to explicitly know Y and Z. That's why we study in Section 15.4 a BSDE whose solution is a Black Scholes Call option. In Section 15.5, we deal with the pricing of contingent claims with constraints on the wealth or portfolio processes. The first example concerns hedging claim with higher interest rate for borrowing (see Subsection 15.5.1). This boils down to solving a nonlinear BSDE. The second example handles the pricing of an American put option. This is equivalent to solve a reflected BSDE (RBSDE for short). According to El Karoui et al. [26], we approximate the solutions of the RBSDEs (via penalisation) by a sequence of standard BSDEs with penalizations. The third example concerns the pricing of an American Put under a portfolio constraint (see Subsection 15.5.3). Using a result from Peng [85], we link the pricing of such an option with a BSDE with penalisation on y and z.

15.1 Choice of the parameters a, n, h_x, h_t, M and N

- We recall that $[-a, a]$ is the interval where the points $X_i, 1 \leq i \leq n$ (see the Definition of $\mathcal{P}$, page 105) are chosen. The choice of a depends on the value of σ and T. Assume we want to find the value of $(Y_t, Z_t)_{0 \leq t \leq T}$ when $(X_t)_{0 \leq t \leq T}$ belongs to the interval $[x_-, x_+]$. Then, we need to compute pointwise u and $\partial_x u$ on $[x_-, x_+]$. Since X satisfies (15.3), if the starting point of X belongs to $[x_-, x_+]$, $(X_t)_{0 \leq t \leq T}$ belongs to $[x_- - 2\sigma\sqrt{T}, x_+ + (\mu - \frac{\sigma^2}{2})T + 2\sigma\sqrt{T}]$ with a probability of 0.95 (if we assume $\mu - \frac{\sigma^2}{2} \geq 0$). Hence, since u_k is given by (9.18) (see page 106) and $\overline{w}_k$ is defined by (9.13) (see page 104), we need to know the value of u_k, $\partial_x u_k$, $\partial_x^2 u_k$ and $\partial_t u_k$ on $[x_- - 2\sigma\sqrt{T}, x_+ + (\mu - \frac{\sigma^2}{2})T + 2\sigma\sqrt{T}]$. The points $\{(T_i^j, X_i^j)_{1 \leq i \leq n, 1 \leq j \leq k}\}$ used by the kernel estimator $\mathcal{P}_n^k$ to build u_k should be chosen in an interval of size $2a$ larger than $4\sigma\sqrt{T} + (\mu - \frac{\sigma^2}{2})T + x_+ - x_-$.

- We recall the values of C_1 and η_1 given by Theorem 13.10 page 166

$$\eta_1 = \frac{4(1+T)L_f^2}{\beta} + K_\eta(T)\left(T\lambda(B)\delta_n(1+M^{-1}h_x^{-2}) + h_x^2 + h_t^2\right),$$
$$C_1 = \frac{K_C(T)}{N^2h_x^2} + K_\eta(T)\left(T\lambda(B)\delta_n h_x^{-2} + (h_t + h_x e^{-\mu a}a^{d-1})h_x^{-2} + h_x^{-1}e^{-\frac{\mu}{\sqrt{d}}a} + h_x^2\right).$$

  Sufficient conditions for C_1 to converge to 0 are the following

  - $\frac{1}{N} << h_x$,
  - $h_t << h_x^2$,
  - $\frac{1}{n} << h_t h_x^{d+2}$.

  Here, we implicitly assume that a is large enough to neglect the exponential terms. The above choice of h_x, h_t and n also implies that $\eta_1 < 1$ (for β large enough) independently of M. The fact that M only appears in the contraction constant η_1 deeply relies on the adaptive control variate method. Had we considered a non adaptive method (like Picard iterations), C_1 would have depended on M. In such a case, we should choose M large. In our case, we don't need to choose M large. This will be confirmed by the numerical experiments.

To get contributions in C_1 of the same order, i.e.

$$\frac{1}{N^2h_x^2} \sim \frac{\delta_n}{h_x^2} \sim \frac{h_t}{h_x^2} \sim h_x^2,$$

we can take

$$\begin{cases} N \sim n^{\frac{2}{d+8}} \\ h_t \sim n^{\frac{-4}{d+8}} \\ h_x \sim n^{\frac{-1}{d+8}} \end{cases} \tag{15.1}$$

With this choice of parameters, we get $C_1 \sim n^{-\frac{2}{d+8}}$. We observe that N doesn't need to be large (this will be confirmed by the numerical experiments).

Let us deal with the practical choice of parameters. We choose the dimension $d = 1$ and $n \sim 10^3$ (n represents the number of random points (T_i, X_i) used by $\mathcal{P}$). We consider the case $a \sim 1$ (this is true for a normal model, see the Black Scholes Call option in Section 15.4 for details). This value of a is not very large, although the previous discussion considered a large enough to neglect the terms depending on it. The choice of h_t and h_x is not as easy as it seems to be. Indeed, the theoretical rules (15.1) ensure that the global error on Y and Z goes to 0 at a given rate. In practice, we may be more concerned by the error on Y, and in this case, the choice $h_t = h_x = \left(\frac{2aT}{n}\right)^{1/3}$ seems to be a good heuristic rule. As we said above, the value of N doesn't need to be large, in the following we choose $N \sim 100$. Concerning M, we choose $M \sim 100$. In Subsection 15.4.4, we study the convergence of the algorithm w.r.t. N and M from a numerical point of view.

15.2 Choice of the kernel function

The kernel functions K_t and K_x must satisfy Hypothesis 12.1 (see page 126), e.g. K_t is defined on the compact support $[-1,1]$, is a C^1_p positive bounded non-odd function such that $\int_{\mathbb{R}} K_t(u)du = 1$. K_x is defined on the compact support $[-1,1]^d$, is a positive bounded function such that $K_x(y) = \prod_{j=1}^d K_x^j(y_j)$. $K_x^j : \mathbb{R} \to \mathbb{R}$ is a C^2_p bounded non-odd function such that $\int_{\mathbb{R}} K_x^j(u)du = 1$, for all $j \in \{1, \cdots, d\}$. From Table 11.1, page 118, we deduce that the Quartic and the Triweight kernels satisfy the above conditions concerning K_x. In addition to the Quartic and Triweight kernels, the Epanechnikov kernel satisfies the conditions on K_t. We also test our algorithm with a truncated Gaussian kernel. Let us consider

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathbf{1}_{|x| \le 6} = K^g(x)\mathbf{1}_{|x|\le 6}.$$

From a numerical point of view, $K^g(x)$ is null for $x \le -6$ and is equal to 1 for $x \ge 6$. Then, numerically, K^g is a C^2_p function. Its support is $[-6,6]$, it is an even function and satisfies $\int_{\mathbb{R}} K^g(u)du = 1$. In the following, K^g is called the truncated Gaussian kernel. The numerical study of the different kernels is done in the following example.

15.3 BSDEs and Application to Finance

We refer to El Karoui et al. [27] for general results on the applications of the BSDEs to finance. We consider a riskless asset P^0 governed by the equation

$$dP^0_t = P^0_t r_t dt,$$

where r_t is the short rate. In addition, we consider d risky stocks satisfying for $i = 1, \cdots, d$

$$dP^i_t = P^i_t(b^i_t dt + \sum_{j=1}^d \sigma^{ij}_t dW^j_t).$$

We denote by $(\mathcal{F}_t)$ the σ-algebra generated by the Brownian motion W and augmented. An investor invests at time t an amount π^i_t of the wealth V_t in the ith stock.

A self financing trading strategy is a pair (V, π), where V is the market value and $\pi = (\pi^1, \cdots, \pi^d)^*$ is the portfolio process, such that (V, π) satisfies

$$dV_t = r_t V_t dt + \pi^*_t \sigma_t (dW_t + \theta_t dt), \int_0^T |\sigma^*_t \pi_t|^2 dt < \infty \, \mathbb{P}\, a.s,$$

where θ is a predictable and bounded-valued process, such that $b_t - r_t\mathbf{1} = \sigma_t\theta_t, \quad d\mathbb{P}\otimes dt\; a.s$, and $\mathbf{1}$ is the vector whose every component is one. From El Karoui et al. [27], Theorem 1.1, we know that if ξ is a positive square integrable contingent claim, there exists an hedging strategy (V, π) against ξ such that

$$dV_t = r_t V_t dt + \pi^*_t \sigma_t \theta_t dt + \pi^*_t \sigma_t dW_t, \quad V_T = \xi. \tag{15.2}$$

Moreover, we get

$$V_t = \mathbb{E}[H^t_T \xi | \mathcal{F}_t] \text{ a.s. where } dH^t_s = -H^t_s(r_s ds + \theta^*_s dW_s),\ H^t_t = 1.$$

Then, finding (V, π) boils down to solving the linear BSDE (15.2).

15.4 A first example : the Black Scholes Call option

We assume that X follows in dimension $d = 1$

$$dX_t = \left(\mu_0 - \frac{\sigma^2}{2}\right) dt + \sigma dW_t, \quad X_0 = x, \tag{15.3}$$

where W is a real $\mathbb{P}$-Brownian motion. (Y, Z) solves the BSDE

$$-dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = \Phi(X_T), \tag{15.4}$$

where the driver f is such that

$$f(t, x, y, z) = -ry - \theta z \text{ where } \theta = \frac{\mu_0 - r}{\sigma}$$

and the terminal condition Φ is $\Phi(x) = (e^x - K)^+$. The parameters are given in Table 15.1.

μ_0	σ	r	T	K
0.1	0.2	0.02	1	100

Table 15.1: Option Parameters

Lemma 15.1. *Solving BSDE (15.4) boils down to computing the hedging strategy against $\Phi(X_T)$ of a standard Black Scholes Call option. We get*

$$Y_t = e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}[(S_T - K)^+ | \mathcal{F}_t],$$

where $dS_t = S_t(rdt + \sigma dB_t)$, $S_0 = e^x$ and B is a $\mathbb{Q}$-Brownian motion.

Proof. In fact, applying El Karoui et al. [27], Theorem 1.1 gives $Y_t = \mathbb{E}_{\mathbb{P}}[H_T^t \Phi(X_T)|\mathcal{F}_t]$, where H_t has been defined above. Then, $H_t = e^{-rt+\theta W_t + \frac{1}{2}\theta^2 t}$ and we can write

$$Y_t = e^{-r(T-t)} \mathbb{E}_{\mathbb{Q}}[\Phi(X_T)|\mathcal{F}_t], \text{ a.s.}$$

where $\mathbb{Q}$ is the risk neutral probability measure such that $\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\theta W_T - \frac{1}{2}\theta^2 T}$. Since $dX_t = (r - \frac{\sigma^2}{2})dt + \sigma dB_t$, where B is a $\mathbb{Q}$-Brownian motion, and by defining $X_t = \log(S_t)$, we get the result. □

Remark 15.2. By using Lemma 15.1, we get a closed formula for Y_t. $Y_t = F(t, S_t)$, where F is the price function of a standard Call option. Since $Y_t = u(t, X_t)$ and $Z_t = (\partial_x u\sigma)(t, X_t)$, we deduce $Z_t = \partial_x F(t, S_t)\sigma S_t$.

15.4.1 Convergence of $\|Y - Y^k\|^2_{\mu,\beta}$ and $\|Z - Z^k\|^2_{\mu,\beta}$

We want to prove that $\|Y - Y^k\|^2_{\mu,\beta}$ and $\|Z - Z^k\|^2_{\mu,\beta}$ converge when k goes to infinity. We recall

$$\|Y - Y^k\|^2_{\mu,\beta} = \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |u(s, X_s^x) - u_k(s, X_s^{N,x})|^2 e^{-\mu|x|} dx ds\right],$$

$$\|Z - Z^k\|^2_{\mu,\beta} = \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s} |(\partial_x u\sigma)(s, X_s^x) - (\partial_x u_k\sigma)(s, X_s^{N,x})|^2 e^{-\mu|x|} dx ds\right].$$

By using Remark 13.8 page 166 and (6.11) page 67, we approximate $\|Y - Y^k\|^2_{\mu,\beta}$ and $\|Z - Z^k\|^2_{\mu,\beta}$ by

$$\|Y - Y^k\|^2_{\mu,\beta} \sim \int_0^T \int_{\mathbb{R}^d} e^{\beta s}\mathbb{E}|u(s,y) - u_k(s,y)|^2 e^{-\mu|y|} dyds, \tag{15.5}$$

$$\|Z - Z^k\|^2_{\mu,\beta} \sim \int_0^T \int_{\mathbb{R}^d} e^{\beta s}\mathbb{E}|(\partial_x u\sigma)(s,y) - (\partial_x u_k\sigma)(s,y)|^2 e^{-\mu|y|} dyds.$$

To approximate the integral w.r.t. y, we choose y in $[4.5, 4.8]$. This interval corresponds to $[x_-, x_+]$, which has been defined at the beginning of Section 15.1. This means that we compute the error $\|Y - Y^k\|^2_{\mu,\beta}$ and $\|Z - Z^k\|^2_{\mu,\beta}$ when $(S_t)_{0\leq t\leq T}$ belongs to $[90, 120]$. We also take $\mu = 1$ and $\beta = 0$. The above expectations are computed w.r.t. the points $\{(T_i^j, X_i^j)_{1\leq i\leq n, 1\leq j\leq k}\}$ used by the kernel estimator $\mathcal{P}_n^k$ to build u_k. We plot the distribution of the following errors

$$e(Y^k - Y) = \int_0^T \int_{x_-}^{x_+} |u(s,y) - u_k(s,y)|^2 e^{-|y|} dyds,$$

$$e(Z^k - Z) = \int_0^T \int_{x_-}^{x_+} |(\partial_x u\sigma)(s,y) - (\partial_x u_k\sigma)(s,y)|^2 e^{-|y|} dyds.$$

We use a truncated Gaussian kernel with the parameters listed in Table 15.2.

n	N	M	h_x	h_t	$2a$
2500	100	100	0.1	0.1	1.2

Table 15.2: Algorithm Parameters

With the value $2a = 1.2$, we choose the points $(X_i^j)_{j\geq 1, 1\leq i\leq n}$ in $[4.1, 5.28]$. Figure 15.1 represents the evolution w.r.t. k of the density of the error $e(Y^k - Y)$ and $e(Z^k - Z)$ for 50 scenarii. The means of the distributions, denoted $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$, respectively

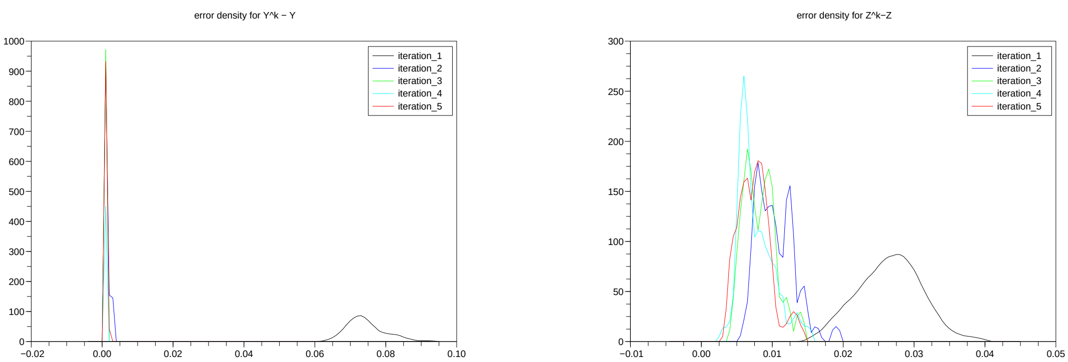

Figure 15.1: Density of the error $Y^k - Y$ and $Z^k - Z$

approximate $\|Y - Y^k\|^2_{\mu,\beta}$ and $\|Z - Z^k\|^2_{\mu,\beta}$. They are listed in Table 15.4.1. We also write down in Table 15.4.1 the variance of the distributions, denoted $e_v(Y^k - Y)$ and $e_v(Z^k - Z)$.

We notice that $e_m(Y^k - Y)$ is almost divided by 100 between the first and the fifth iterations. Moreover, it drastically decreases between the first and the second iterations. From iteration 2, the algorithm doesn't reduce so much the error on Y. $e_m(Z^k - Z)$ is almost divided by 4 between the first and the fifth iterations. A contrary to $e_m(Y^k - Y)$, $e_m(Z^k - Z)$ decreases more slowly, and seems to be stabilised from the fourth iteration. Concerning the evolution of $e_v(Y^k - Y)$ and $e_v(Z^k - Z)$ w.r.t. k, we remark that $\frac{e_v(Y^1-Y)}{e_v(Y^5-Y)} \sim 16$ and $\frac{e_v(Z^1-Z)}{e_v(Z^5-Z)} \sim 2$. We also notice that $e_v(Y^k - Y)$ drastically decreases between the first and the second iterations and decreases slowly from iteration 2. This is consistent with the evolution of $e_m(Y^k - Y)$ w.r.t. k.

	$e_m(Y^k - Y)$	$e_m(Z^k - Z)$
k=1	0.0743476	0.0265350
k=2	0.0014802	0.0104687
k=3	0.0010029	0.0082452
k=4	0.0008865	0.0076881
k= 5	0.0008373	0.0075321

Figure 15.2: Evolution of $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$ w.r.t. k

	$e_v(Y^k - Y)$	$e_v(Z^k - Z)$
k=1	0.0051774	0.0040402
k=2	0.0005904	0.0026707
k=3	0.0003638	0.0022511
k=4	0.0003268	0.0025221
k= 5	0.0003188	0.0022863

Figure 15.3: Evolution of $e_v(Y^k - Y)$ and $e_v(Z^k - Z)$ w.r.t. k

15.4.2 Pointwise convergence

In the preceding subsection, we have studied the convergence of the algorithm for the $\|\cdot\|_{\mu,\beta}$ norm. The graphs 15.4 and 15.5 represent the evolution w.r.t. k of the errors $u_k - u$ and $(\partial_x u_k - \partial_x u)\sigma$ on $[0, 1] \times [4.5, 4.81]$. The right column represents the level curves of the errors.

Looking at Figure 15.4, we notice that at the second iteration, the error is already quite small. This corroborates the remarks we have done about Table 15.4.1. A contrary to Figure 15.4, Figure 15.5, drawing the evolution of the error on $\partial_x u\sigma$ w.r.t. k, is not satisfying. We don't notice any drastic reduction of the error when k increases. If we carefully look at Figures 15.4 and 15.5, we remark that the errors on u and $\partial_x u\sigma$ are quite bad for T near to 1 and for a spot around K. This is due to the singularity of the payoff function at $S_T = K$. Since the algorithm builds a $C_p^{1,2}$ function, we cannot get a good approximation of the payoff function, which is C^0.

15.4.3 Influence of n

The asymptotic complexity of the algorithm is of order KMn^2, where K denotes the number of iterations, M the number of Monte Carlo simulations used to compute $\overline{w}_k$ (see (9.13), page 104) and n the number of random points used by the operator $\mathcal{P}$. Then, the value of n has a strong impact on the computational time of our algorithm. First, we check that the complexity of the algorithm is of order KMn^2, then we study the efficiency of the algorithm w.r.t. the computational time.

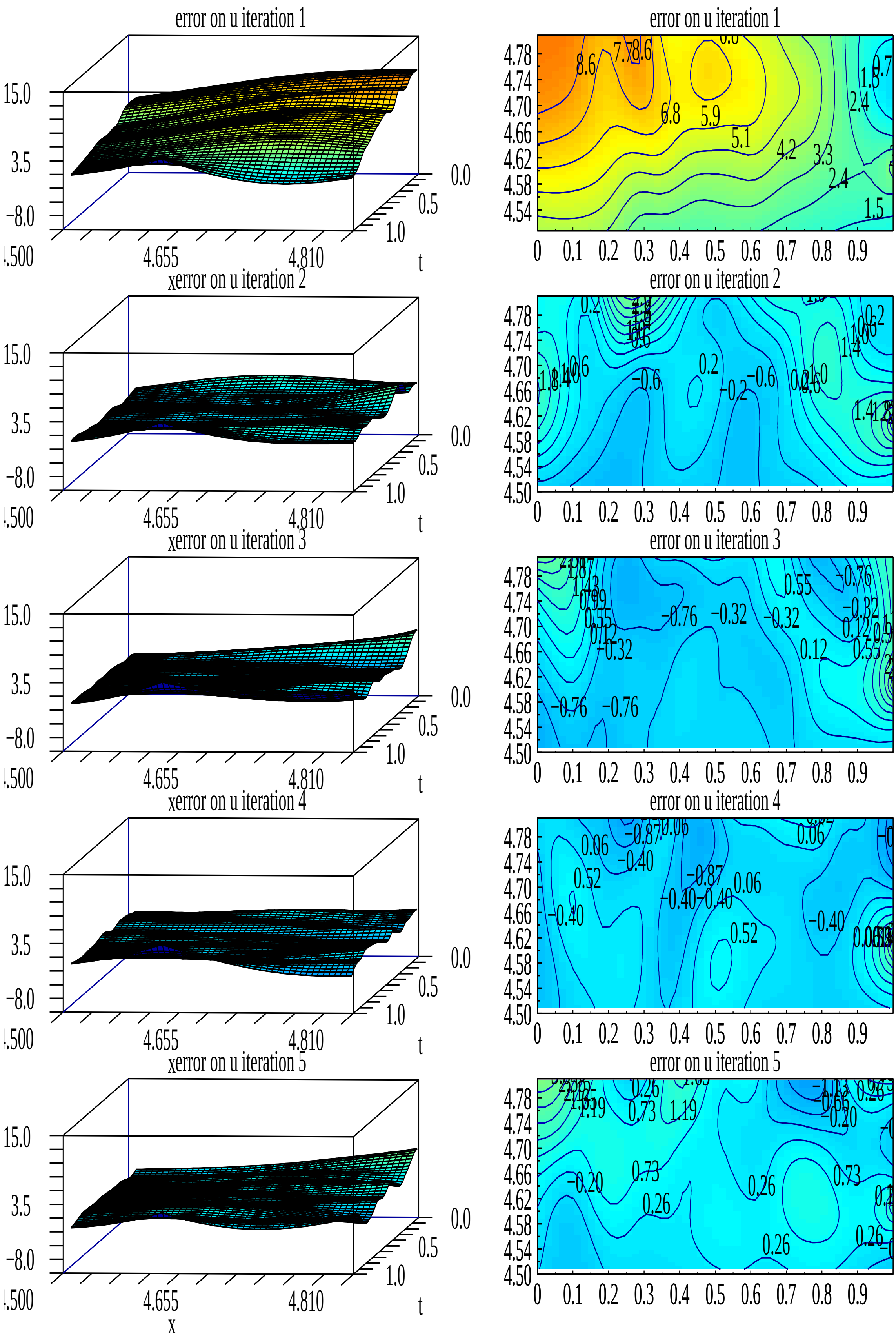

Figure 15.4: Evolution of the error on u w.r.t. k

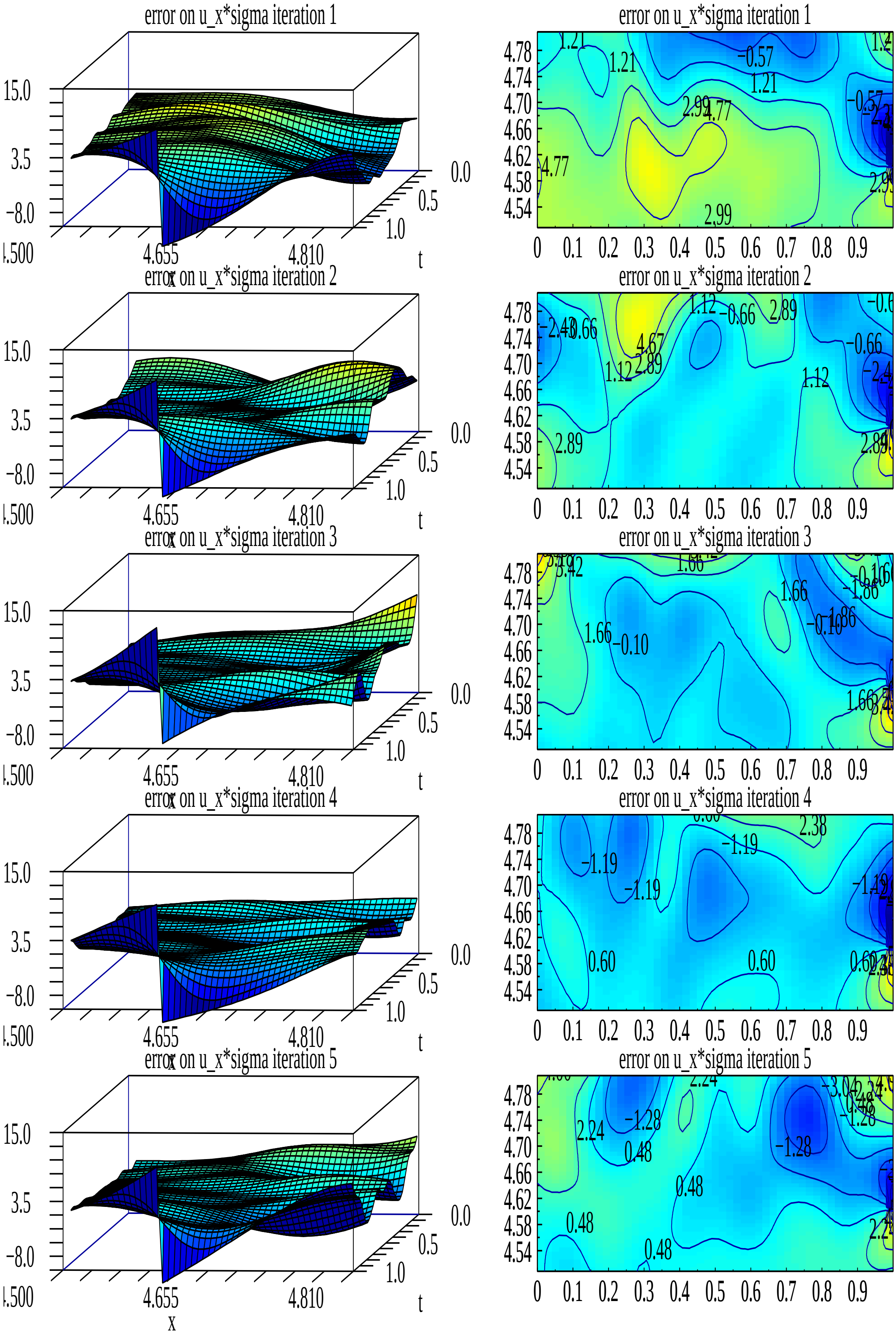

Figure 15.5: Evolution of the error on $\partial_x u\sigma$ w.r.t. k

Complexity of the algorithm

Table 15.6 represents the time spent by the algorithm to compute the value of Y^5 and Z^5, which are the approximations of Y and Z at the fifth iteration, when $n = 500$, 1000 and 2500 and for $N = M = 10$. We plot $\log(time)$ w.r.t. $log(n)$ in Figure 15.7. Since the complexity is KMn^2, the slop must be of order 2. By using a linear regression method, we get the parameters a_n, b_n, std_n, where $\log(time) = a_n * log(n) + b_n$, and std_n represents the standard deviation of the residuals: $std_n = 0.0194057$, $b_n = -9.6265698$ and $a_n = 1.7679698$.

n	500	1000	2500
time (second)	3.96	12.92	67.89

Figure 15.6: Computational time for different values of n

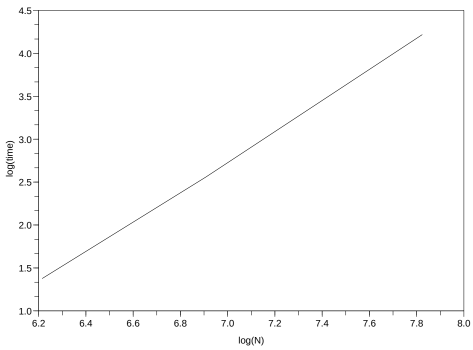

Figure 15.7: Evolution of $\log(time)$ w.r.t $\log(n)$

We also study the evolution of the computational time w.r.t. the value of M, although theoretically, as we said in Section15.1, the value of M doesn't affect the limiting error. This will be confirmed in Subsection 15.4.4, where we study the impact of M on $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$. Table 15.8 represents the time spent by the algorithm to compute the value of Y^5 and Z^5, which are the approximations of Y and Z at the fifth iteration, when $M = 10$, 50 and 100 and for $n = 500$ and $N = 10$.We plot $\log(time)$ w.r.t. $log(M)$ in Figure 15.9. Since the complexity is KMn^2, the slop must be of order 1. By using a linear regression method, we get the parameters a_M, b_M, std_M, where $\log(time) = a_M * log(M) + b_M$, and std_M represents the standard deviation of the residuals: $std_M = 0.0263865$, $b_M = -0.5323893$ and $a_M = 0.8241543$. We have got $a_N \sim 1.76$ and $a_M \sim 0.82$. These coefficients are not so far from the theoretic values given by the asymptotic complexity of the algorithm (i.e. $a_N = 2$ and $a_M = 1$).

Efficiency of the algorithm v.s. computational time

Table 15.3 represents the evolution of the error on Y (denoted $e_m(Y^k - Y)$) for different values of n, and for $N = M = 10$. Since the values of h_x and h_t change when n changes

M	10	50	100
time (second)	3.96	14.23	26.80

Figure 15.8: Computational time for different values of M

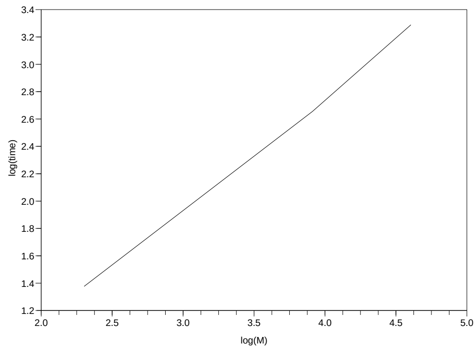

Figure 15.9: Evolution of $\log(time)$ w.r.t $\log(M)$

(see (15.1)), we specify the different values of these parameters. Figure 15.10 represents the evolution of $\log(e_m(Y^5 - Y))$ and $\log(e_m(Z^5 - Z))$ w.r.t. $\log(n)$. By using a linear regression method, we get the parameters a_y, b_y, std_y, where $\log(e_m(Y^5 - Y)) = a_y * log(n) + b_y$, and std_y represents the standard deviation of the residuals: $std_y = 0.0091393$, $b_y = 0.7361349$ and $a_y = -1.0147081$. We also get the parameters a_z, b_z, std_z, where $\log(e_m(Z^5 - Z)) = a_z * log(n) + b_z$, and std_z represents the standard deviation of the residuals: $std_z = 0.0067187$, $b_z = -3.516634$ and $a_z = -0.4784886$. We notice that the error on Y decreases linearly with n and the error on Z decreases as $\sqrt{n}$. Since the computational time increases as n^2 (see previous Subsection), improving the error on Y and especially on Z costs a lot in term of computational time.

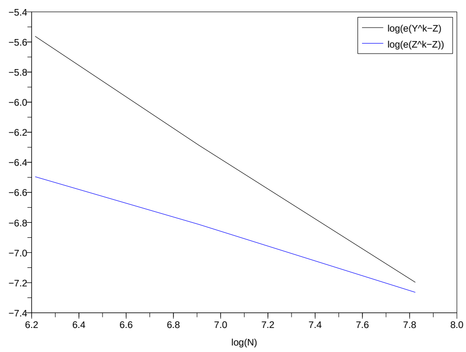

Figure 15.10: Evolution of $e_m(Y^5 - Y)$ and $e_m(Z^5 - Z)$ w.r.t $\log(n)$

n	500 ($h_x = h_t = 0.13$)	1000 ($h_x = h_t = 0.1$)	2500 ($h_x = h_t = 0.078$)
k=1	0.1067746	0.0737104	0.0621128
k=2	0.0045913	0.0020142	0.0010975
k=3	0.0037402	0.0017277	0.0007527
k=4	0.0037202	0.0017123	0.0006969
k=5	0.0038930	0.0018623	0.0007485

Table 15.3: Evolution of $e_m(Y^k - Y)$ w.r.t. k for different values of n

n	500 ($h_x = h_t = 0.13$)	1000 ($h_x = h_t = 0.1$)	2500 ($h_x = h_t = 0.078$)
k=1	0.0357527	0.0311915	0.0284791
k=2	0.0164547	0.0118549	0.0072938
k=3	0.0148545	0.0105162	0.0071240
k=4	0.0146653	0.0104761	0.0068115
k= 5	0.0151874	0.0110527	0.0070327

Table 15.4: Evolution of $e_m(Z^k - Z)$ w.r.t. k for different values of n

15.4.4 Influence of M and N

In this section, we want to show that the values of M and N do not affect $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$ much, as we have foretold it in Section 15.1. We keep the same parameters as in Tables 15.1 and 15.2, except that we change the value of N in Tables 15.5 and 15.6, and the value of M in Tables 15.7 and 15.8.

	N=10	N=50	N=100	N=500
k=1	0.0743641	0.0738299	0.0743476	0.0734905
k=2	0.0010729	0.0015602	0.0014802	0.0015361
k=3	0.0009750	0.0009144	0.0010029	0.0009863
k=4	0.0008497	0.0009315	0.0008865	0.0009587
k=5	0.0009265	0.0009624	0.0008373	0.0008692

Table 15.5: Evolution of $e_m(Y^k - Y)$ w.r.t. k and N

	N=10	N=50	N=100	N=500
k=1	0.0273184	0.0270381	0.0265350	0.0270870
k=2	0.0074741	0.0113869	0.0104687	0.0112863
k=3	0.0069694	0.0078171	0.0082452	0.0077122
k=4	0.0068796	0.0080186	0.0076881	0.0077423
k=5	0.0072260	0.0077805	0.0075321	0.0072681

Table 15.6: Evolution of $e_m(Z^k - Z)$ w.r.t. k and N

	M=10	M=50	M=100	M=500
k=1	0.0738116	0.0746727	0.0743476	0.0739790
k=2	0.0011348	0.0012501	0.0014802	0.0011098
k=3	0.0009754	0.0010027	0.0010029	0.0009499
k=4	0.0009433	0.0009787	0.0008865	0.0009293
k=5	0.0009641	0.0008824	0.0008373	0.0009633

Table 15.7: Evolution of $e_m(Y^k - Y)$ w.r.t. k and M

	M=10	M=50	M=100	M=500
k=1	0.0270811	0.0276510	0.0265350	0.0270689
k=2	0.0075221	0.0075821	0.0104687	0.0074627
k=3	0.0075100	0.0075407	0.0082452	0.0071787
k=4	0.0074765	0.0076484	0.0076881	0.0072807
k=5	0.0068578	0.0066434	0.0075321	0.0074454

Table 15.8: Evolution of $e_m(Z^k - Z)$ w.r.t. k and M

We notice from Tables 15.5, 15.6, 15.7 and 15.8 that changing the value of M or N doesn't change the values of $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$. M and N don't have a big impact on the errors $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$. Since N is always bigger that 3.16 $(= h_x^{-1/2})$, and M is bigger than 1, these results confirm what we have predicted in Section 15.1.

15.4.5 Influence of the kernel

We plot in Figure 15.11 the graphs of the three kernels we study in this part : Quartic, Triweight and Gaussian. Table 15.4.5 gives the evolution of $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$

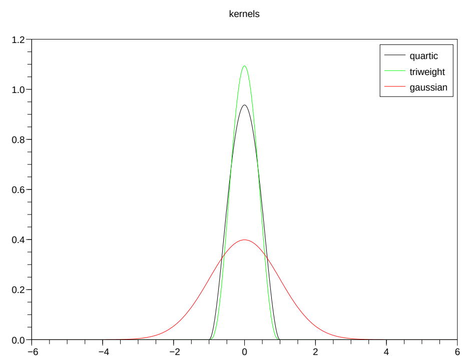

Figure 15.11: Different kernels

w.r.t. k when using a Gaussian kernel, a Triweight kernel and a Quartic kernel respectively.

We keep the same parameters as for the Gaussian kernel (Tables 15.1 and 15.2). For the Triweight and Quartic kernels, we only change h_x and h_t, which are chosen equal to 0.2.

	Gaussian	Triweight	Quartic
k=1	0.0743476	0.0569964	0.0607659
k=2	0.0014802	0.0010012	0.0010628
k=3	0.0010029	0.0007176	0.0008328
k=4	0.0008865	0.0007517	0.0008022
k=5	0.0008373	0.0007492	0.0007522

Figure 15.12: Evolution of $e_m(Y^k - Y)$ w.r.t. k

	Gaussian	Triweight	Quartic
k=1	0.0265350	0.0296810	0.0285901
k=2	0.0104687	0.0105762	0.0100127
k=3	0.0082452	0.0092400	0.0086051
k=4	0.0076881	0.0090822	0.0085337
k=5	0.0075321	0.0093429	0.0083038

Figure 15.13: Evolution of $e_m(Z^k - Z)$ w.r.t. k

We notice that the Triweight and Quartic kernels give a better accuracy than the Gaussian one when we compute $e_m(Y^k - Y)$, but this is the opposite for $e_m(Z^k - Z)$.

15.4.6 Hedging Strategy

We consider the Black Scholes Call option described at the beginning of the section whose parameters are summed up in Table 15.9.

μ_0	σ	r	T	K	S_0
0.1	0.2	0.02	1	100	100

Table 15.9: Option Parameters for the hedging

We want to check whether our algorithm provides the correct hedging strategy for such an option. We keep the algorithm parameters of Table 15.2. We use a Gaussian kernel. The strategy (wealth, portfolio) satisfies (15.2) with $\xi = \Phi(X_T)$. We can also write

$$dV_t = \delta(t, S_t)dS_t + (V_t - \delta(t, S_t)S_t)rdt,$$

where S has been introduced in Lemma 15.1 and satisfies $dS_t = S_t(\mu_0 dt + \sigma dW_t)$. $\delta(t, S_t)$ is equal to $\frac{\pi_t}{S_t}$ and represents the quantity of stocks invested. If we consider that the value of $\delta(t, S_t)$ is constant between two trading dates, we deduce from the above equation that

$$\forall i \in \{0, \cdots, N-1\}, \quad V_{t_{i+1}} = V_{t_i}e^{rh} + \delta(t_i, S_{t_i})(S_{t_{i+1}} - S_{t_i}e^{rh}),$$

where $0 := t_0 < t_1 < \cdots < t_N := T$ are the trading dates, and $h = \frac{T}{N} = t_{i+1} - t_i$, for all $i = 0, \cdots, N-1$. When h is small, we also have

$$\forall i \in \{0, \cdots, N-1\}, \quad V_{t_{i+1}} - V_{t_i} = \delta(t_i, S_{t_i})(S_{t_{i+1}} - S_{t_i}) + (V_{t_i} - \delta(t_i, S_{t_i})S_{t_i})rh. \quad (15.6)$$

The value of the wealth at time 0 is the price of the option at time 0, i.e. $V_0 = Y_0$. The value of $\delta(t, S_t)$ is related to Z_t in the following way $\delta(t, S_t) = \frac{Z_t}{\sigma S_t}$. We use our algorithm to get an approximated value of Y_0 and Z_{t_i}, $i = 1, \cdots, N$. If our algorithm performs well, the tracking error $e_T := V_T - (S_T - K)^+$, where V_T is computed using $V_0 = Y_0^k$,

$\delta(t_i, S_{t_i}) = \frac{Z_{t_i}^k}{\sigma S_{t_i}}$ and (15.6), should be small. We draw in Figure 15.14 an histogram of the value of e_T for 10000 trajectories of S and $k = 5$. We compare it with the histogram of the value of e_T when $\delta(t_i, S_{t_i})$ is computed exactly (see Remark 15.2).

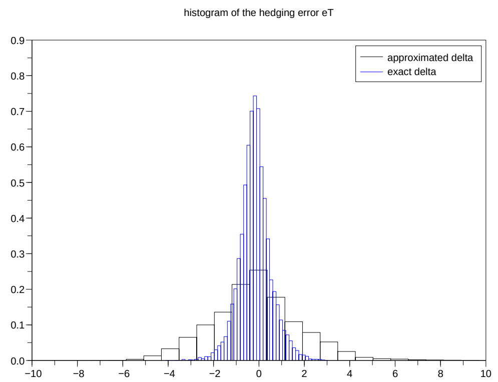

Figure 15.14: histogram of e_T

15.4.7 Comparison with Picard's Algorithm

We compare our algorithm based on a sequential Monte Carlo method (SMC algorithm for short) with an algorithm using Picard's iterations, based on Corollary 9.3, page 95 (Picard's algorithm for short). Picard's algorithm is simpler than the SMC algorithm since we don't use any adaptive control variate. We can described it in the SMC algorithm framework (see Section 9.6, page 103). The procedure is the same, except for (9.13) and (9.18). (9.13) becomes

$$\overline{w}_k^P(t,x) = \frac{1}{M}\sum_{m=1}^{M}\left[\Phi(X_T^{m,k,N}) + \int_t^T f\left(s, X_s^{m,k,N}, u_{k-1}^P(s, X_s^{m,k,N}), (\partial_x u_{k-1}^P \sigma)(s, X_s^{m,k,N})\right) ds\right].$$

Since there is no correction term in the above definition of $\overline{w}_k^P(t,x)$, (9.18) is replaced by $u_k^P(t,x) = \mathcal{P}_n^k(\overline{w}_k^P)(t,x)$. Figure 15.15 draws the evolution of $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$ w.r.t. k when using Picard's algorithm and the SMC algorithm. Figure 15.16 draws the evolution of $e_v(Y^k - Y)$ and $e_v(Z^k - Z)$.

We notice that $e_m(Y^k - Y)$, $e_m(Z^k - Z)$, $e_v(Y^k - Y)$ and $e_v(Z^k - Z)$ when using Picard's algorithm are larger than the ones ensuing from the SMC algorithm.

15.5 Second example : Constrained portfolios

Pricing of contingent claims with constraints on the wealth or portfolio processes leads to deal with nonlinear backward equations for the fair price of claims.

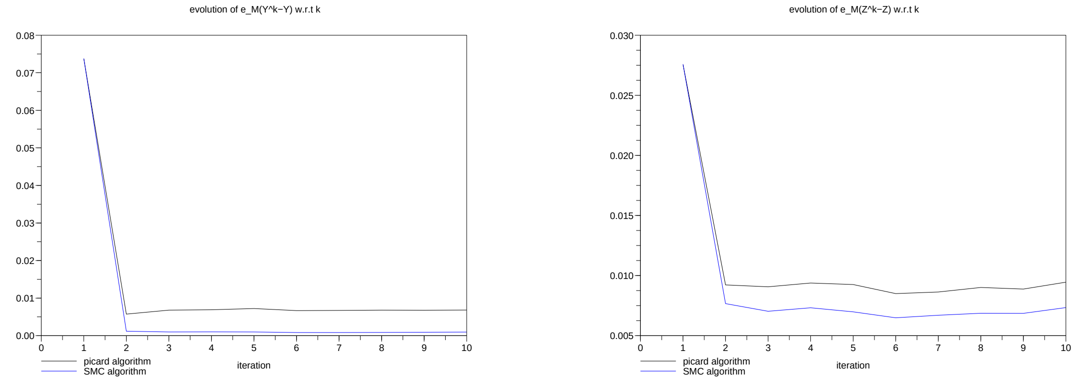

Figure 15.15: Evolution of $e_m(Y^k - Y)$ and $e_m(Z^k - Z)$ w.r.t. k

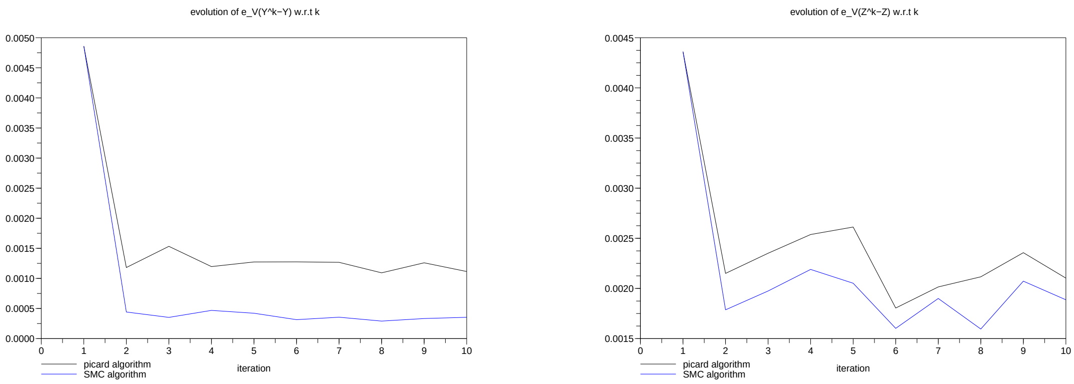

Figure 15.16: Evolution of $e_v(Y^k - Y)$ and $e_v(Z^k - Z)$ w.r.t. k

15.5.1 Hedging claim with higher interest rate for borrowing

We refer to El Karoui et al. [27] for this first example of constrained portfolio. We consider the case where the investor is allowed to borrow money at time t at an interest rate $R > r$, where r is the bond rate. We borrow and invest money in the bond at the same time, but we restrict ourselves to policies in which the amount borrowed at time t is equal to $(V_t - \sum_{i=1}^d \pi_t^i)^-$. The strategy (wealth, portfolio) (V, π) satisfies

$$dV_t = rV_t dt + \pi_t^* \sigma\theta dt + \pi_t^* \sigma dW_t - (R - r)\left(V_t - \sum_{i=1}^d \pi_t^i\right)^- dt.$$

Finding the strategy (V, π) consists in solving BSDE (15.4) with the nonlinear driver $f(t, x, y, z) = -ry - \theta z + (R - r)(y - \frac{z}{\sigma})^-$.

A first case : $\Phi(x) = (e^x - K)^+$

The option parameters are summed up in Table 15.10.

Lemma 15.3. *Let* $(Y_t^{r,R}, Z_t^{r,R})$ *denote the solution of*

$$-dY_t = f^{r,R}(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = \Phi(X_T), \tag{15.7}$$

μ_0	σ	r	R	T	K	S_0
0.05	0.2	0.04	0.06	0.5	100	100

Table 15.10: Option Parameters

where the driver $f^{r,R}$ *is such that* $f^{r,R}(t,x,y,z) = -ry - \theta z + (R-r)(y - \frac{z}{\sigma})^-$ *and* $\Phi(x) = (e^x - K)^+$*. Then,* $(Y_t^{r,R}, Z_t^{r,R})$ *is equal to* (Y_t^R, Z_t^R)*, where* (Y_t^R, Z_t^R) *is the solution of the BSDE*

$$-dY_t = f^{R,R}(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = \Phi(X_T), \tag{15.8}$$

with $f^{R,R}(t,x,y,z) = -Ry - \frac{\mu_0 - R}{\sigma} z$*.*

Proof. We can rewrite $f^{r,R}(t,x,y,z) = -Ry - \frac{\mu_0 - R}{\sigma} z + (R-r)(y - \frac{z}{\sigma})^+$. Since $Y_t^R \leq \frac{Z_t^R}{\sigma}$ for all $t \in [0,T]$ (we can prove it by using the closed formula giving the value of a standard Call option and its derivative w.r.t. S_t), we get that $f^{r,R}(t, X_t, Y_t^R, Z_t^R) = f^{R,R}(t, X_t, Y_t^R, Z_t^R)$. Applying Proposition 9.2, page 95 ends the proof. $\square$

With the option parameters of Table 15.10, we get $Y_0^R = 7.156$ and $\delta^R(0, S_0) = \frac{Z_0^R}{\sigma S_0} = 0.611$. As before, we run 50 times our algorithm and we take the mean of Y_0^k, Z_0^k over the 50 values. The algorithm parameters are summed up in Table 15.11. Figure 15.17 plots

n	N	M	h_x	h_t	$2a$
2500	100	100	0.07	0.07	0.84

Table 15.11: Algorithm Parameters

the evolution of $Y_0^{r,R,k}$ (NL price) and $\delta_0^{r,R,k}$ (NL hedging) w.r.t. k, where $(Y^{r,R,k}, Z^{r,R,k})$ is the approximated solution of (15.7). We notice that the value $Y_0^{r,R,k}$ is always below

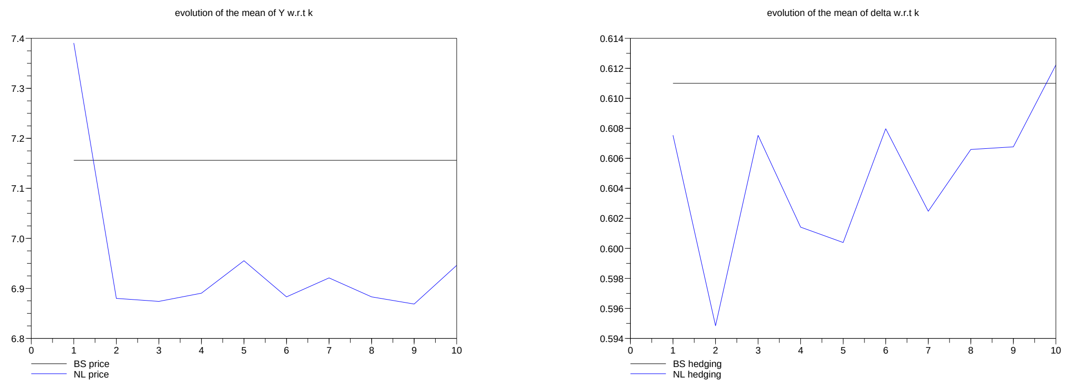

Figure 15.17: Evolution of $Y_0^{r,R,k}$ and $\delta_0^{r,R,k}$ w.r.t. k

Y_0^R (while they should be equal) and the relative error on $\delta_0^{r,R,k}$ $(\frac{\delta_0^R - \delta_0^{r,R,k}}{\delta_0^R})$ is twice smaller than the one on $Y_0^{r,R,k}$ $(\frac{Y_0^R - Y_0^{r,R,k}}{Y_0^R})$. There is a bias between Y_0^R and $Y_0^{r,R,k}$. This bias persists when we compute $Y_0^{R,R,k}$. Although the results on the convergence of $Y^k - Y$ in $\|\cdot\|_{\mu,\beta}$ norm are quite good, we notice again that the pointwise error on Y is not negligible.

A second case : $\Phi(x) = (e^x - K_1)^+ - 2(e^x - K_2)^+$

In such a case, we cannot prove a result corresponding to Lemma 15.3. Since the payoff is not $(e^x - K)^+$, we don't have $Y_t^R \leq \frac{Z_t^R}{\sigma}$ anymore, and then $Y_0 \neq Y_0^R$. Using the comparison Theorem (see El Karoui et al. [27], Theorem 2.2), we only get upper and lower bounds for Y_0.

Lemma 15.4. *Let* $(Y^{i,r}, Z^{i,r})$, $i = 1, 2$ *be the solution of*

$$-dY_t = f^{r,r}(t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = (S_T - K_i)_+.$$

We also define $(Y^{i,R}, Z^{i,R})$, $i = 1, 2$, *the solution of*

$$-dY_t = f^{R,R}(t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = (S_T - K_i)_+.$$

We recall that $(Y^{r,R}, Z^{r,R})$ *denotes the solution of (15.7), where* $\Phi(x) = (e^x - K_1)^+ - 2(e^x - K_2)^+$. *Then, we get* $Y_t^{r,R} \geq Y_t^{1,r} - 2Y_t^{2,R}$, $Y_t^{r,R} \geq Y_t^{1,r} - 2Y_t^{2,r}$, $Y_t^{r,R} \geq Y_t^{1,R} - 2Y_t^{2,R}$ *and* $Y_t^{r,R} \leq Y_t^{1,R} - 2Y_t^{2,r}$.

Proof. We only prove the first assertion, i.e. $Y_t^{r,R} \geq Y_t^{1,r} - 2Y_t^{2,R}$. The other ones can be deduced in the same way. $(Y^{1,r} - 2Y^{2,R}, Z^{1,r} - 2Z^{2,R})$ is the solution of a BSDE with terminal condition Φ and whose driver is $f(\omega, t) = f^{r,r}(t, Y_t^{1,r}, Z_t^{1,r}) - 2f^{R,R}(t, Y_t^{2,R}, Z_t^{2,R})$. To apply the comparison Theorem (see El Karoui et al. [27], Theorem 2.2), it remains to prove that $f^{r,R}(t, Y_t^{1,r} - 2Y_t^{2,R}, Z_t^{1,r} - 2Z_t^{2,R}) \geq f(\omega, t)$.

$$\begin{aligned}
&f^{r,R}(t, Y_t^{1,r} - 2Y_t^{2,R}, Z_t^{1,r} - 2Z_t^{2,R}) - f(\omega, t) = -r(Y_t^{1,r} - 2Y_t^{2,R}) - \frac{\mu_0 - r}{\sigma}(Z_t^{1,r} - 2Z_t^{2,R}) \\
&+ (R - r)\left[Y_t^{1,r} - 2Y_t^{2,R} - \frac{Z_t^{1,r} - 2Z_t^{2,R}}{\sigma}\right]^- + rY_t^{1,r} + \frac{\mu_0 - r}{\sigma}Z_t^{1,r} - 2(RY_t^{2,R} + \frac{\mu_0 - R}{\sigma}Z_t^{2,R}), \\
&= (R - r)\left(2\frac{Z_t^{2,R}}{\sigma} - 2Y_t^{2,R} + \left[Y_t^{1,r} - 2Y_t^{2,R} - \frac{Z_t^{1,r} - 2Z_t^{2,R}}{\sigma}\right]^-\right).
\end{aligned}$$

To conclude, we use that $x^- \geq -x$ and $Y_t^{1,r} \leq \frac{Z_t^{1,r}}{\sigma}$. □

By using the preceding Lemma and the option parameters of Table 15.12, we deduce that $Y_0^{r,R}$ belongs to $[2.76, 3.6]$. Table 15.14 sums up the evolution of the mean of Y_0^k (over 50 values) w.r.t. k. The algorithm parameters are summed up in Table 15.13. We notice that the values Y_0^k, for $k \geq 2$ are in the interval $[2.76, 3.6]$.

μ_0	σ	r	R	T	K_1	K_2	S_0
0.05	0.2	0.01	0.06	0.25	95	105	100

Table 15.12: Option Parameters

n	N	M	h_x	h_t	$2a$
2500	100	100	0.05	0.05	0.6

Table 15.13: Algorithm Parameters

iteration	mean of Y_0^k	variance of Y_0^k
$k=1$	2.6148572	0.0016014
$k=2$	3.1168342	0.0019487
$k=3$	3.1069206	0.0018843
$k=4$	3.1007172	0.0021258
$k=5$	3.1058564	0.0024850

Table 15.14: Evolution of mean of Y_0^k w.r.t. k

15.5.2 Application to American options

We consider an asset S following the Black Scholes model

$$dS_t = S_t(\mu_0 dt + \sigma dW_t), \quad S_0 = e^x,$$

where W is a real $\mathbb{P}$-Brownian motion. We introduce the risk neutral probability $\mathbb{Q}$, whose density w.r.t. $\mathbb{P}$ is $\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\theta W_T - \frac{1}{2}\theta^2 T}$ ($\theta = \frac{\mu_0 - r}{\sigma}$). S satisfies

$$dS_t = S_t(rdt + \sigma dB_t),$$

where B is a $\mathbb{Q}$-Brownian motion. The price of an American put option is

$$V_t = P(t, S_t), \qquad \text{where } P(t,x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}_{\mathbb{Q}} \left(e^{-r(\tau - t)}(K - S_\tau^{t,x})_+ \right),$$

and $\mathcal{T}_{t,T}$ is the set of stopping times with values in $[t,T]$. As in section 15.4, we introduce $X_t = \log(S_t)$, and we get

$$dX_t = \left(\mu_0 - \frac{\sigma^2}{2}\right) dt + \sigma dW_t, \quad X_0 = x. \tag{15.9}$$

We define $\Phi(x) = (K - e^x)_+$ and $\mathcal{A}$, the generator of X such that $\mathcal{A} = \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2} + (r - \frac{\sigma^2}{2})\frac{\partial}{\partial x} - r$. By using Lamberton and Lapeyre [65], Section 5.3.2, we get that the American Put satisfies the following partial differential equation

$$\begin{cases} \max\left(\Phi(x) - v(t,x), \partial_t v(t,x) + \mathcal{A}v(t,x)\right) = 0 \;\text{ a.e. in } [0,T] \times \mathbb{R}, \\ v(T,x) = \Phi(x). \end{cases}$$

Relation between reflected BSDEs and parabolic PDEs

We aim at linking the solution of the above PDE and the solution of the following reflected BSDE (RBSDE in short)

$$\begin{cases} -dY_t = (-rY_t - \theta Z_t)dt - Z_t dW_t + dK_t, \\ Y_T = (K - e^{X_T})_+, \; Y_t \geq (K - e^{X_t})_+, \; 0 \leq t \leq T, \\ \int_0^T \{Y_t - (K - e^{X_t})_+\} dK_t = 0, \end{cases}$$

where X has been defined in (15.9). Using El Karoui et al. [26], Theorem 8.5 yields that $Y_t^{t,x}$, solution of the above RBSDE, is equal to $v(t,x)$, the solution of the preceding PDE. Then, finding the price of an American put option boils down to solving a reflected BSDE.

Approximation via penalisation of solutions of RBSDEs

Our algorithm is not able to solve RBSDEs. That's why we present here a way of approximating solutions of RBSDEs by a sequence of standard BSDEs with penalisation. This method has been introduced by El Karoui et al. [26], Section 6 for proving the existence of a solution of RBSDE transforming the constraint $Y_t \geq (K - e^{X_t})_+$ into a penalisation.

$$
\begin{aligned}
&Y_t = g(T, X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds + K_T - K_t - \int_t^T Z_s dW_s,\\
&Y_t \geq g(t, X_t), \ \int_t^T (Y_s - g(s, X_s))dK_s = 0.
\end{aligned}
$$

Let us consider the following sequence of BSDEs

$$
Y_t^i = g(T, X_T) + \int_t^T f(s, X_s, Y_s^i, Z_s^i)ds + i \int_t^T (Y_s^i - g(s, X_s))^- ds - \int_t^T Z_s^i dW_s,
$$

whose solution is denoted (Y^i, Z^i). We define $K_t^i = i \int_t^T (Y_s^i - g(s, X_s))^- ds$. El Karoui et al. [26] prove that (Y^i, Z^i, K^i) converges to (Y, Z, K), solution of the above RBSDE, when i goes to infinity. Moreover, Y^i is an increasing sequence (converging to Y).

Application to our algorithm

Using the previous result and more particularly the definition of Y^i enables to approximate Y from below with our algorithm, (Y^i solves a standard BSDE). Then, the algorithm should give lower bounds for Y, closer and closer to Y when i increases. Consider an option whose parameters are given in Table 15.15.

μ_0	σ	r	T	K	S_0
0.05	0.4	0.05	1	100	100

Table 15.15: Option Parameters

As a reference, we choose the values given by the Kamrad Ritchken tree method (computed with PREMIA[1]): $Y_0 = 13.66$, $\delta_0 = -0.394$. The parameters of our algorithm are given in Table 15.16.

As before, we run 50 times our algorithm and we compute the mean and the variance of $Y_0^{i,k}, Z_0^{i,k}$ over all the 50 values, for $k = 1, \cdots, 10$ and $i = 0.5$, 1, 5 and 10. Figures 15.18 and 15.19 draw the evolution of $Y_0^{i,k}$ and $\delta_0^{i,k}$ (mean and variance) w.r.t. k, when $i = 0.5$, 1, 5 and 10. We notice that the variance of $Y_0^{i,k}$ and $\delta_0^{i,k}$ is very big when $i = 5, 10$ and when k is small. When k increases, the variance becomes smaller but it still oscillates.

[1]PREMIA is a pricing software developped by the MathFi team of INRIA Rocquencourt, see http://www.premia.fr

n	N	M	h_x	h_t	$2a$
2500	100	100	0.12	0.12	2.4

Table 15.16: Algorithm Parameters

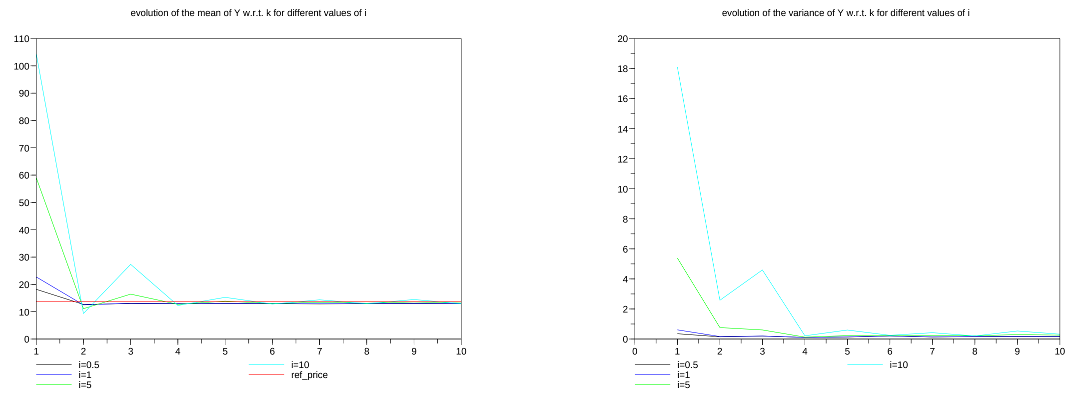

Figure 15.18: Evolution of $Y_0^{i,k}$ (mean and variance) w.r.t. i and k

Figure 15.20 zooms on the mean of $Y_0^{i,k}$ and $\delta_0^{i,k}$ for k greater than 3. According to the theory, the values of $Y_0^{i,k}$ should be smaller than Y_0 (13.66). Looking at Figure 15.20, we notice that for $i = 10$, $Y_0^{10,k}$ oscillates around Y_0 w.r.t. k. $\delta_0^{i,k}$ seems to converge around -0.39.

15.5.3 American option with constraint on the portfolio

As we have linked American options to BSDEs with penalisation related to y in Subsection 15.5.2, we can link American options with constraints on the portfolio with BSDEs with penalisations related to (y, z). First, we recall some results coming from Peng [85].

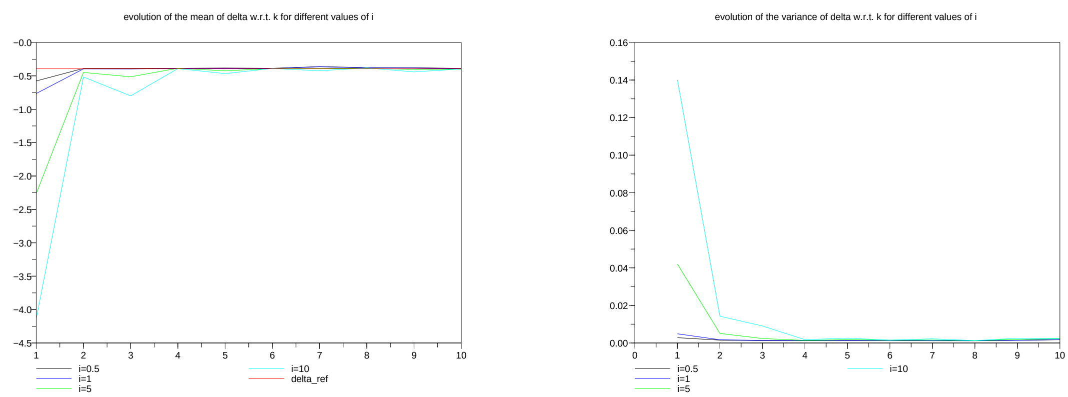

Figure 15.19: Evolution of $\delta_0^{i,k}$ (mean and variance) w.r.t. i and k

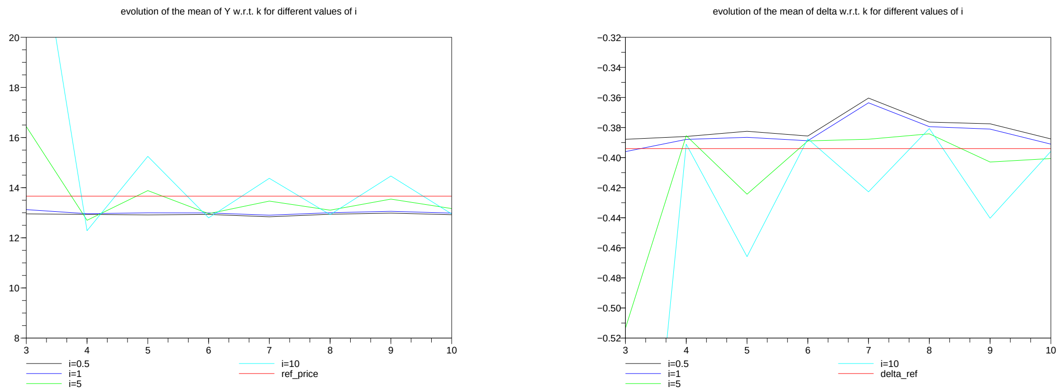

Figure 15.20: Evolution of the mean of $Y_0^{i,k}$ and δ_0^k w.r.t. k

BSDEs with constraints on (y, z)

In the previous part, we have considered BSDEs with constraints on y. This part generalises the previous one, since we are dealing with BSDEs with constraints on y and z. We consider the following BSDE

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds + (K_T - K_t) - \int_t^T Z_s dW_s, \tag{15.10}$$

such that

$$\phi(t, Y_t, Z_t) = 0, \text{ a.e., a.s.} \tag{15.11}$$

where $\phi : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R}^q$ is a non negative function, globally Lipschitz w.r.t. (y, z) and such that $\mathbb{E} \int_0^T |\phi(s, Y_s, Z_s)|^2 ds < \infty$. According to Peng [85], Section 4, such a BSDE does not have a unique solution. Peng [85] defines in Definition 4.1 the smallest f-supersolution on $[0, T]$ of BSDE (15.10) with constraint (15.11). To construct the smallest solution of the previous BSDE the author introduces the sequence

$$Y_t^i = \xi + \int_t^T f(s, Y_s^i, Z_s^i)ds + i \int_t^T \phi(s, Y_s^i, Z_s^i)ds - \int_t^T Z_s^i dW_s. \tag{15.12}$$

He proves in Theorem 4.2 that the sequence Y^i converges monotonously up to Y. Z^i and K^i converge to Z and K in $\mathbb{L}^2$. Moreover, (Y, Z, K) is the smallest f-supersolution of BSDE (15.10) with constraint (15.11).

Application to American option with constraint on the portfolio

We still consider an American Put option and we impose a constraint on the amount π involved in the asset. We refer to Peng and Xu [86], Section 7.2 for more details on the following examples.

Example 1: No borrowing

We consider the case where no borrowing is allowed. This means that π_t should be smaller than Y_t. By using Peng and Xu [86], Proposition 7.3, we know that the price process Y is the same as the one in complete market, i.e. we don't need to borrow money to replicate an American Put option. The function ϕ defined in (15.12) is

$$\phi(s, Y_s^i, Z_s^i) = (Y_s^i - g(s, X_s))^- + \left(Y_s^i - \frac{Z_s^i}{\sigma}\right)^- .$$

We consider the parameters given in Tables 15.15 and 15.16. Figures 15.21 and 15.22 show the evolution of $Y_0^{i,k}$ and $\delta_0^{i,k}$ w.r.t. i and k. We notice that for $i = 10$ there is no convergence to Y_0 and δ_0 anymore.

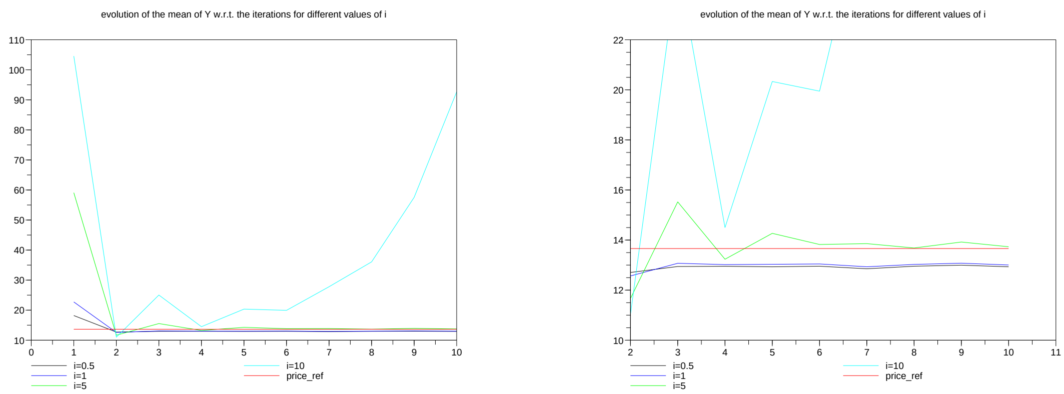

Figure 15.21: Evolution of $Y_0^{i,k}$ w.r.t. i and k

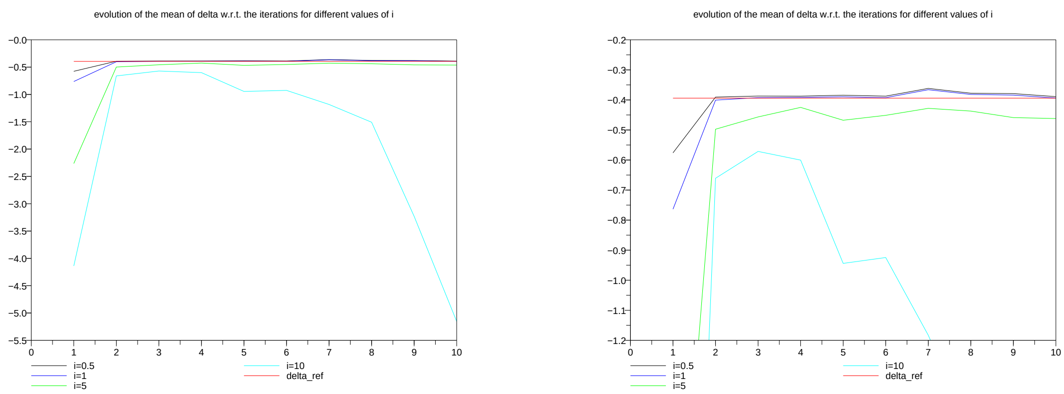

Figure 15.22: Evolution of $\delta_0^{i,k}$ w.r.t. i and k

Example 2: No short-selling

In such a case, π_t should be positive for $t \in [0, T]$. The solution of (15.10) under the constraints $Y_t \geq g(t, X_t)$ and $\pi_t \geq 0$ is

$$Y_t = \begin{cases} K, & y \in [0, T) \\ (K - e^{X_T})_+, & t = T \end{cases}$$

$$\pi_t = 0,$$

$$K_t = \begin{cases} Krt, & y \in [0, T) \\ KrT + K - (K - e^{X_T})_+, & t = T. \end{cases}$$

Then, we should get $Y_0 = K$. The function ϕ defined in (15.12) is

$$\phi(s, Y_s^i, Z_s^i) = (Y_s^i - g(s, X_s))^- + \left(\frac{Z_s^i}{\sigma}\right)^- .$$

Figure 15.23 draw the evolutions of $Y_0^{i,k}$ and $\delta_0^{i,k}$ w.r.t. k for $i = 0.5$ and $i = 1$. We notice that $Y_0^{0.5,k}$ converges to 34, $Y_0^{1,k}$ converge to 55, whereas the theoretic value of the limit is 100. $\delta_0^{0.5,k}$ converges to -0.52, $\delta_0^{1,k}$ converges to -0.43, and the theoretic value of the limit is 0. For $i \geq 2$, the evolutions of $Y_0^{i,k}$ and $\delta_0^{i,k}$ w.r.t. k diverge. The study of the sequence

$$\begin{aligned} Y_t^{i,j} =& \xi + \int_t^T f(s, Y_s^{i,j}, Z_s^{i,j})ds + i \int_t^T \phi_1(s, Y_s^{i,j}, Z_s^{i,j})ds + j \int_t^T \phi_2(s, Y_s^{i,j}, Z_s^{i,j})ds \\ & - \int_t^T Z_s^{i,j} dW_s, \end{aligned}$$

where

$$\phi_1(s, Y_s^{i,j}, Z_s^{i,j}) = (Y_s^{i,j} - g(s, X_s))^- \text{ and } \phi_2(s, Y_s^{i,j}, Z_s^{i,j}) = \left(\frac{Z_s^{i,j}}{\sigma}\right)^-$$

for different values of i and j doesn't work too. We plot in Figure 15.24 the evolution of $Y_0^{i,1,k}$ and $\delta_0^{i,1,k}$ for $i = 1, 5, 10, 15$ and 20, and for $j = 1$. We notice that the bigger i is, the higher $Y_0^{i,1,10}$ is, but still below 100. Figure 15.25 represents the evolution of $Y_0^{i,1,k}$ and $\delta_0^{i,1,k}$ w.r.t. $i \in [20, 100]$ and k for $j = 1$. We notice that for $i = 80$ and $i = 100$, $Y_0^{i,1,k}$ still converges to a value bigger than 100. Moreover, $\delta_0^{i,1,10}$ for $i \leq 40$ are less than -1, which is not compatible with the theory (we should get δ in $[-1, 1]$). If we increase j, the result is worse than for $j = 1$. As we can see in Figure 15.26, there is no convergence of $Y_0^{i,1.5,k}$ and $\delta_0^{i,1.5,k}$ when k increases, even if $i = 0.5$.

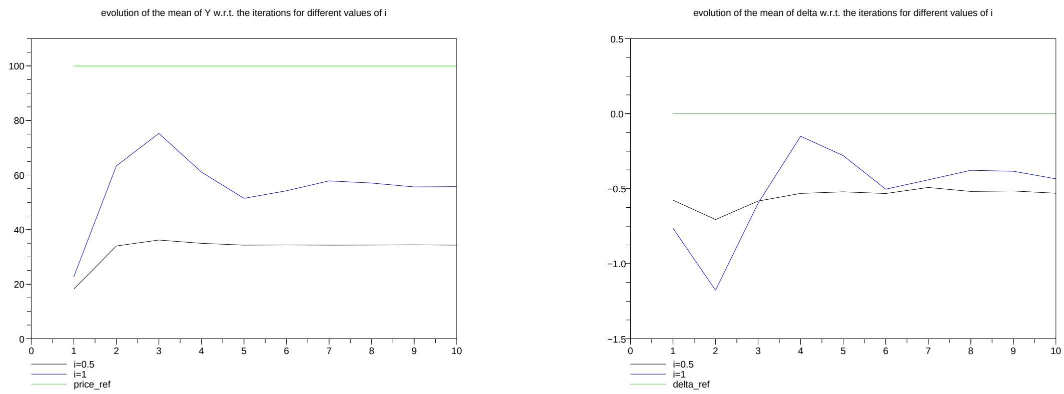

Figure 15.23: Evolution of $Y_0^{i,k}$ and $\delta_0^{i,k}$ w.r.t. i and k

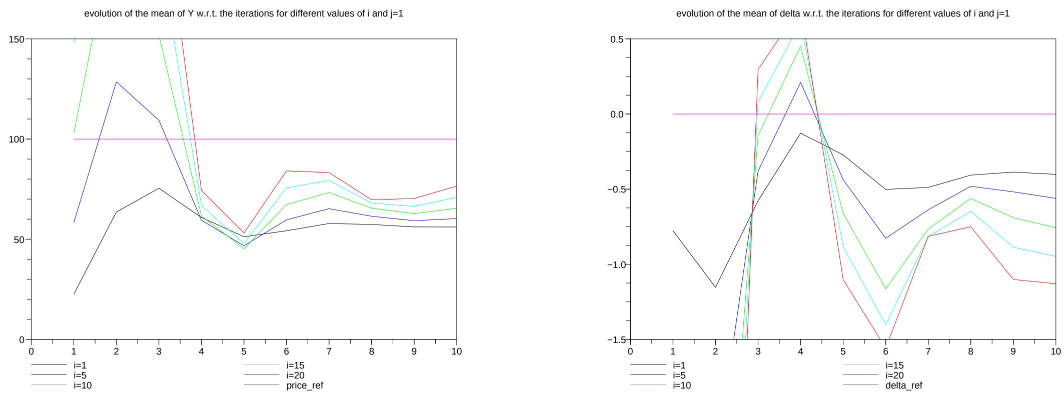

Figure 15.24: Evolution of $Y_0^{i,1,k}$ and $\delta_0^{i,1,k}$ w.r.t. $i \in [1, 15]$ and k for $j = 1$

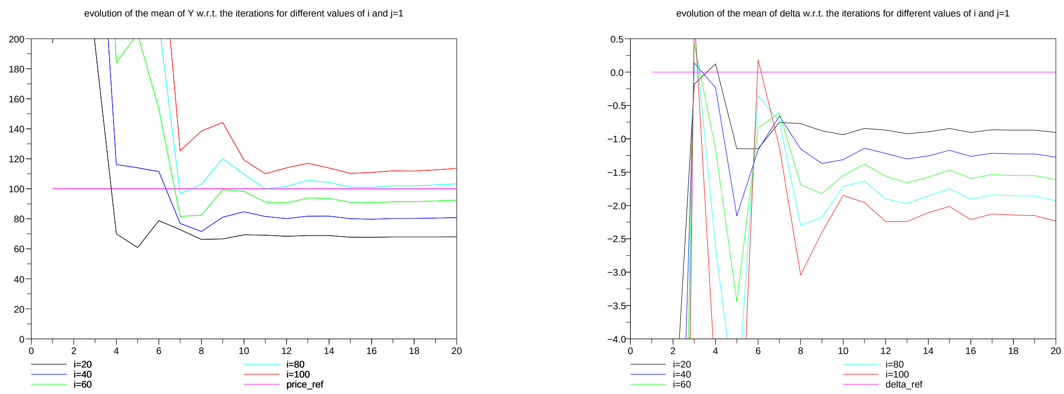

Figure 15.25: Evolution of $Y_0^{i,1,k}$ and $\delta_0^{i,1,k}$ w.r.t. $i \in [20, 100]$ and k for $j = 1$

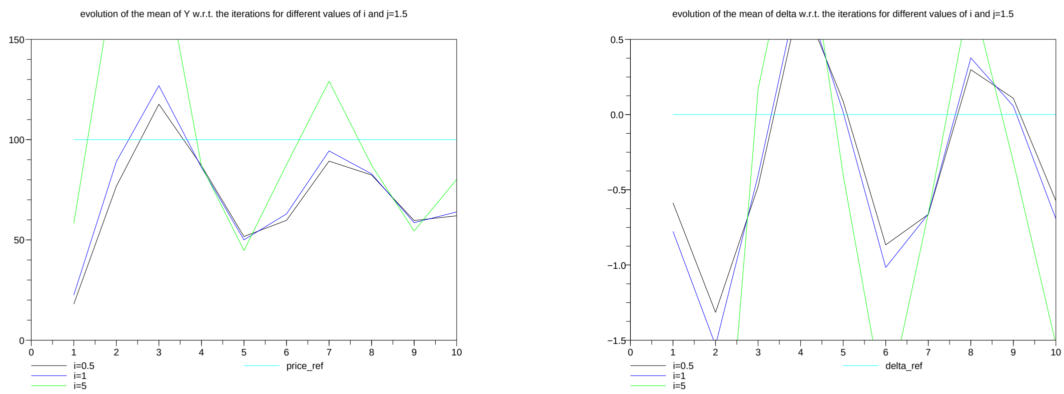

Figure 15.26: Evolution of $Y_0^{i,1.5,k}$ and $\delta_0^{i,1.5,k}$ w.r.t. i and k for $j = 1.5$

Part IV

Pricing American options with boundary sensitivities

In this part, we tackle the problem of the numerical valuation of Bermudan options using domain optimization techniques. For this, we compute domain sensitivities w.r.t. small perturbations of the boundary. After a review of known sensibility results in Section 17 in the American case, we prove new results in the Bermudan one. The design of a relevant algorithm based on this sensitivity analysis is not presented in this manuscript, although we have studied alternative schemes in collaboration with Cristina Costantini (University "Gabriele d'Annunzio", Chieti and Pescara, Italy). This is still a work in progress. Results presented in Chapter 18 concern the robustness of the sensitivity estimate as the number of exercise dates increases. We prove convergence and uniform bounds under different sets of assumptions. Our proofs are based on techniques related to the analysis of asymptotic overshoot. We mention that one of our results has not been completely proved (Conjecture 1).

The following chapter presents standard results on the valuation of American/Bermudan options. The connection with the domain optimization issue is made in Chapter 17.

Chapter 16

Motivations

16.1 Framework and Hypotheses

We consider the $\mathbb{R}^d$-valued diffusion process $(X^{t,x})$ solution of

$$X_s = x + \int_t^s b(r, X_r)dr + \int_t^s \sigma(r, X_r)dW_r, \tag{16.1}$$

where $(W_t)_{t\geq 0}$ is a q-dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying usual conditions. Hypothesis 16.1-1 ensures the existence of a unique strong solution to (16.1).

Hypothesis 16.1 *Let α be in $]0,1]$.*

1. *Smoothness of order $1+\alpha$. b and σ belong to $\mathcal{H}_{1+\alpha}$, $\alpha \in]0,1]$ (see the definition of $\mathcal{H}_{1+\alpha}$ at the beginning of Section 17),*
2. *Uniform ellipticity. For some $a_0 > 0$, it holds $\xi \cdot [\sigma\sigma^*](t,x)\xi \geq a_0|\xi|^2$ for any $(t,x,\xi) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d$,*
3. *$\mathcal{D}$ is a C^2 domain.*

The uniform ellipticity condition is satisfied when X represents the log-price of the underlying asset.

16.2 Pricing American options

The valuation of American options is still a major issue in asset pricing. The buyer of such a contract is given the right to exercise the option at any time τ between now (say t) and the maturity T. We assume that the vectors of prices or log-prices of the underlying asset X evolve according to the SDE (16.1). We also assume that the market is complete and that the instantaneous rate is of the form $(r(s, X_s))_{s\leq t\leq T}$. If the payoff at time τ is $g(\tau, X_\tau)$ (with g a continuous function satisfying suitable integrability conditions), the fair price of the option is given by (see Karatzas [55])

$$P(t,x) = \sup_{\tau\in[t,T]} \mathbb{E}\left[e^{-\int_t^\tau r(s,X_s^{t,x})ds} g(\tau, X_\tau^{t,x})\right], \tag{16.2}$$

the supremum being taken over all stopping times with values in $[t, T]$. There is no simple numerical method available to evaluate the price of American options. We refer the reader to Fu et al. [31] for a review of numerical methods to handle this issue.

16.2.1 Standard results on American options

Definition 16.1 (Continuation region and exercise region). $\mathcal{C} := \{(t,x),\ t < T, x \in \mathbb{R}^d : P(t,x) > g(t,x)\}$ is called the continuation region and $\mathcal{E} := \{(t,x),\ t < T, x \in \mathbb{R}^d : P(t,x) = g(t,x)\}$ is called the exercise region. We also define the t-sections for every $t \in [0,T[$ by $\mathcal{E}_t := \{x \in \mathbb{R}^d : P(t,x) = g(t,x)\}$. Since P and g are continuous functions, $\mathcal{C}$ is an open set of $]-\infty, T[\times\mathbb{R}^d$.

Let us define $\mathcal{A}_t$, the operator of X such that $\mathcal{A}_t = \frac{1}{2}\sum_{i,j=1}^d [\sigma\sigma^*]_{i,j}(t,x)\frac{\partial^2}{\partial_{x_i x_j}} + \sum_{i=1}^d b_i(t,x)\frac{\partial}{\partial_{x_i}} - r(t,x)$. By using Lamberton and Lapeyre [65], Section 5.3.2, we get that $P(t,x)$, defined by (16.2), solves the following partial differential equation

$$\begin{cases} \max\left(g(x) - v(t,x), \partial_t v(t,x) + \mathcal{A}_t v(t,x)\right) = 0 & \text{a.e. in } [0,T] \times \mathbb{R}^d, \\ v(T,x) = g(T,x). \end{cases}$$

In other words, one can say that P satisfies the equation

$$\begin{cases} \partial_t v(t,x) + \mathcal{A}_t v(t,x) = 0 & \text{in } \mathcal{C}, \\ v = g & \text{on } \mathcal{PC}, \end{cases} \tag{16.3}$$

where $\mathcal{PC}$ denotes the parabolic boundary of $\mathcal{C}$ (see Definition 17.1).

Let us recall some standard results on the exercise region for standard put American options (i.e. with a single underlying asset). We consider the following 1-dimensional Black Scholes model for the log-price

$$dX_t = (r - \delta - \frac{1}{2}\sigma^2)dt + \sigma dW_t,$$

where δ is the continuous dividend rate. Let $B_t := \sup\{x \in \mathbb{R} : (t,x) \in \mathcal{E}\}$ denote the immediate exercise boundary. Van Moerbeke [91] and Jacka [52] show that B_t is continuous. Kim [57] and Jacka [52] show that B_t is increasing in t. Kim [57] proves that $B_{T^-} \equiv \lim_{t\to T} B_t = \log(K \wedge \frac{r}{\delta}K)$. (i.e. $\mathcal{C}_T^- =]\log(K \wedge \frac{r}{\delta}K), +\infty[$). In the pioneer work of Van Moerbeke [91], the behaviour of B_t as t approaches T was investigated. Barles et al. [10] proved the following estimate for the case $\delta = 0$

$$\lim_{t\to T} \frac{K - B_t}{\sigma K\sqrt{(T-t)\log(\frac{1}{T-t})}} = 1. \tag{16.4}$$

The above estimate remains valid whenever $0 \le \delta < r$, as it can be proved by the method of Barles et al. [10]. Lamberton and Villeneuve [66] give rigorous results for the cases $r < \delta$ and $r = \delta$. Namely, the authors show that the parabolic behaviour stated by Van Moerbeke [91] does hold in the case $r < \delta$.

Second, we recall some results on the exercise region of American options on multiple assets. We consider the following d-dimensional Black Scholes model

$$dX_t^i = (r - \delta_i - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2)dt + \sum_{j=1}^{d} \sigma_{ij} dW_t^j, \quad i = 1, \cdots, d.$$

Broadie and Detemple [18] analyse several types of American options on two or more assets. The authors study options on the maximum of two assets, dual strike options, spread options and others. For each of these contracts they characterise the optimal exercise regions and develop valuation formulae. They also derive results for American option contracts with non convex payoffs, such as American capped exchange options. For this option, the authors explicitly identify the optimal exercise boundary. Villeneuve [93] studies the non emptiness and the shape of the exercise region of American options written on several assets. The author states an analytic theorem which characterises the non emptiness of the exercise region. He also studies a particular class of payoff functions for which he explicitly identifies the shape and the asymptotic behaviour near maturity of the associated exercise region. These results can be viewed as the multidimensional version of Kim's results.

The following Proposition sums up some properties of the exercise region $\mathcal{E}$

Proposition 16.2 (Villeneuve [93], Proposition 1.1)**.** *Let us consider the following d-dimensional Black Scholes model*

$$dX_t^i = (r - \delta_i - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2)dt + \sum_{j=1}^{d} \sigma_{ij} dW_t^j, \quad i = 1, \cdots, d.$$

where δ stands for the dividend rate and the matrix σ is invertible. We consider $P(t,x) = \sup_{\tau \in [t,T]} \mathbb{E}\left[e^{-r(\tau - t)} g(X_\tau^{t,x})\right]$, the price of an American option where the payoff g is a nonzero function. The exercise region $\mathcal{E}$ defined above satisfies the following properties

- *$\mathcal{E}$ is closed on $]-\infty, T[\times \mathbb{R}^d$,*
- *The family $(\mathcal{E}_t)_{t<T}$ is non decreasing,*
- *$\forall t < T$, $\mathcal{E}_t \subset \mathcal{O} = \{x : g(x) > 0\}$.*

Proposition 16.3 (Smooth-fit principle)**.** *Let us consider P, the price of an American option with payoff g and r constant*

$$P(t,x) = \sup_{\tau \in [t,T]} \mathbb{E}\left[e^{-r(\tau - t)} g(\tau, X_\tau^{t,x})\right],$$

Under Hypotheses on b, σ and g described in Friedman [29], page 489, we get that at the boundary of $\mathcal{C}$, $\nabla P = \nabla g$.

Remark 16.4. An important sufficient condition required to satisfy the smooth-fit condition is that g should be C^2 (see Friedman [29], page 489 or Brekke and Øksendal [16]). Therefore, if g is the payoff of a Put or a Call option, we cannot apply the above Proposition.

Remark 16.5. In the particular case of an American put option in dimension 1 and in the Black Scholes model, the smooth-fit condition is satisfied. We refer to Myneni [76], Section 4 for a proof of it.

Remark 16.6. In the recent paper Villeneuve [92], the author investigates sufficient conditions that ensure the optimality of threshold strategies for optimal stopping problems with finite or perpetual maturities. His result is based on a local-time argument that enables to give an alternative proof that the smooth-fit principle applies as soon as the payoff function is differentiable.

16.2.2 Optimal domain approach

We recall

Lemma 16.7. *The smallest optimal stopping time of (16.2) is given by* $\tau^* := \inf\{t \le s \le T : P(s, X_s^{t,x}) = g(s, X_s^{t,x})\}$.

We refer to the proof of Lamberton and Lapeyre [65], Theorem 3.2, Chapter 5 for a proof of the Lemma. Using this Lemma, we get

Proposition 16.8. *Let* $P(t,x)$ *be defined by (16.2). We can also write*

$$P(t,x) = \sup_{\mathcal{D} \subset]0,T[\times \mathbb{R}^d} \mathbb{E}\left[e^{-\int_t^{\tau_{\mathcal{D}}^{t,x}} r(s, X_s^{t,x}) ds} g(\tau_{\mathcal{D}}^{t,x}, X_{\tau_{\mathcal{D}}^{t,x}}^{t,x}) \right], \tag{16.5}$$

where $\tau_{\mathcal{D}}^{t,x} = \inf\{s > t : (s, X_s^{t,x}) \notin \mathcal{D}\}$, *i.e.* $\tau_{\mathcal{D}}^{t,x}$ *is the first exit time of* $(s, X_s^{t,x})_{t \le s \le T}$ *from* $\mathcal{D}$, *and* $\mathcal{D}$ *is an open set of* $]0,T[\times \mathbb{R}^d$.

Proof. Since the exit time of an open set by the $\mathcal{F}_t$-adapted continuous time-space process $(t, X_t)_t$ is a stopping time and $\tau_{\mathcal{D}}^{t,x} \in [t,T]$, we have $\sup_{\mathcal{D} \subset]0,T[\times \mathbb{R}^d} \mathbb{E}\left[e^{-\int_t^{\tau_{\mathcal{D}}^{t,x}} r(s, X_s^{t,x}) ds} g(\tau_{\mathcal{D}}^{t,x}, X_{\tau_{\mathcal{D}}^{t,x}}^{t,x}) \right] \le P(t,x)$.
To prove the converse inequality, we distinguish two cases.

- If $\tau^* = t$, the choice $\mathcal{D} = \emptyset$ clearly leads to $\tau_{\mathcal{D}}^{t,x} = t$. Thus, $\tau^* = \tau_{\mathcal{D}}^{t,x}$ and

$$P(t,x) \le \sup_{\mathcal{D} \subset]0,T[\times \mathbb{R}^d} \mathbb{E}\left[e^{-\int_t^{\tau_{\mathcal{D}}^{t,x}} r(s, X_s^{t,x}) ds} g(\tau_{\mathcal{D}}^{t,x}, X_{\tau_{\mathcal{D}}^{t,x}}^{t,x}) \right].$$

- If $\tau^* > t$, then $\tau^* = \inf\{s > t : (s, X_s) \notin \mathcal{C}\}$. $\mathcal{D} = \mathcal{C} \cap (]0,T[\times \mathbb{R}^d)$ is an open set of $]0,T[\times \mathbb{R}^d$ and $\tau_{\mathcal{D}}^{t,x} = \tau^*$, which ends the proof.

□

In the following, the alternative characterisation of $P(t,x)$, given by (16.5), will be useful. Let us state in the next section a result equivalent to Proposition 16.8 for Bermudan options.

16.3 Pricing Bermudan options

A Bermudan option can be exercised on prespecified days during the life of the option. It is reasonable to say that Bermudan options are an hybrid of European options, which can only be exercised on the option expiry date, and American options, which can be exercised at any time during the option life time. As a consequence, under same conditions, the value of a Bermudan option is greater than (or equal to) a European option but smaller than (or equal to) an American option. We assume the same hypotheses than in Section 16.2. We define

$$P^N(t,x) = \sup_{\tau \in \mathcal{T}^N_{t,T}} \mathbb{E}\left[e^{-\int_t^\tau r(s,X_s^{t,x})ds} g(\tau, X_\tau^{t,x})\right], \tag{16.6}$$

where $\mathcal{T}^N_{t,T} = [t,T] \cap \mathcal{T}^N$, and $\mathcal{T}^N := \{t_0, t_1, \cdots, t_N\}$ such that $0 = t_0 < t_1 < \cdots < t_N = T$. We can also prove (see Lamberton and Lapeyre [65], Chapter 2, Section 5.1) that P^N satisfies the dynamic programming equation

$$P^N(t_N, x) = g(t_N, x), \quad \forall x \in \mathbb{R}^d,$$

and for all $n \leq N-1$

$$P^N(t_n, \cdot) = \max\left(g(t_n, \cdot), \mathbb{E}\left[e^{-\int_{t_n}^{t_{n+1}} r(s,X_s^{t,\cdot})ds} P^N(t_{n+1}, X_{t_{n+1}}^{t_n,\cdot})\right]\right).$$

Before stating a result equivalent to Proposition 16.8, let us give a definition

Definition 16.9 (Definition of $\mathcal{D}_{t_j}$ and $\mathcal{D}^N$)**.** Let j be an integer of $\{0, \cdots, N\}$. $\mathcal{D}_{t_j}$ denotes the section of $\mathcal{D}$ at $t = t_j$, i.e. to $\mathcal{D}_{t_j} = \mathcal{D}\big|_{t=t_j}$. Then, $(t_j, \mathcal{D}_{t_j}) \in [0,T] \times \mathbb{R}^d$, and we define

$$\mathcal{D}^N = \bigcup_{j=1}^N (t_j, \mathcal{D}_{t_j}).$$

As for American options, the following Lemma holds

Lemma 16.10. *The smallest optimal stopping time of (16.6) is given by* $\tau^*_N := \inf\{s \in \mathcal{T}^N_{t,T} : P^N(s, X_s^{t,x}) = g(s, X_s^{t,x})\}$.

From this Lemma, we deduce the following Proposition

Proposition 16.11. *Let $P^N(t,x)$ be defined by (16.6). We can also write*

$$P^N(t,x) = \sup_{\mathcal{D}^N} \mathbb{E}\left[e^{-\int_t^{\tau_N^{t,x}} r(s,X_s^{t,x})ds} g(\tau_N^{t,x}, X^{t,x}_{\tau_N^{t,x}})\right],$$

where $\tau_N^{t,x} = \inf\{t_i > t : X_{t_i}^{t,x} \notin \mathcal{D}_{t_i}\}$, i.e. $\tau_N^{t,x}$ is the first exit time from a domain $\mathcal{D}^N$ for the time-space process $(t_i, X_{t_i}^{t,x})_{t \leq t_i \leq T}$.

This Proposition can be proved as Proposition 16.8.

16.4 Intuitive approach for pricing Bermudan options

The optimization of the r.h.s. of (16.5) (resp. (16.6)) could be carried out by a "gradient" algorithm that uses the sensitivity with respect to the domain $\mathcal{D}$ (resp. $\mathcal{D}^N$). We present in the following chapter a tractable formula for these sensitivities. The formula for the sensitivity w.r.t. $\mathcal{D}$ has been stated by Costantini et al. [22]. We give in Theorem 17.15 the formula for the sensitivity w.r.t. $\mathcal{D}^N$.

The principle of domain optimization is not new and has already appeared in the literature. García [32] and Ibáñez and Zapatero [51] have introduced parametrisations of the exercise region (in simple cases) and discussed the optimization issue (without computing the gradient). In the case of Bermudan Asian options, the frontier has specific properties which enabled Fu and Wu [30] to establish a formula for the sensitivity w.r.t the parameter characterising the domain. We go even further and prove a formula for the sensitivity in a more general case.

Chapter 17

Boundary sensitivities

17.1 A sensitivity formula in continuous time

Costantini et al. [22] have studied the sensitivity, with respect to a time dependent domain $\mathcal{D}$, of expectations of functionals of a diffusion process stopped at the exit from $\mathcal{D}$ or normally reflected at the boundary of $\mathcal{D}$. They have established a differentiability result and given an explicit expression for the gradient that allows the gradient to be computed by Monte Carlo methods. The section is organised as follows. First, we recall some smoothness definitions for the time-space domain $\mathcal{D}$. Then, we state the main results of Costantini et al. [22], by focusing on diffusion processes stopped at the boundary. We also give a way to use these results in the pricing of American options.

17.1.1 Time-space domains

In the sequel $\mathcal{D}$ stands for a bounded time-space domain in $]0,T[\times\mathbb{R}^d$, T is a fixed terminal time. The boundary of $\mathcal{D}$ is denoted, as usual, by $\partial\mathcal{D}$. Regularity assumptions on the domain $\mathcal{D}$ will be formulated in terms of Hölder spaces with time-space variables (see Lieberman [69], page 46). Let $\mathcal{D}'$ be an arbitrary time-space domain. If the index of regularity is $a = k+\alpha$ for k a non negative integer and $\alpha \in]0,1]$, then $\mathcal{H}_a(\mathcal{D}')$ is a Banach space of functions f of class $C^{\lfloor \frac{k}{2} \rfloor, k}(\mathcal{D}')$ with Hölder continuous k-th derivatives, namely with a finite norm $|f|_{a,\mathcal{D}'}$ where

$$|f|_{a,\mathcal{D}'} = \sum_{|\beta|+2j\leq k} \sup_{\mathcal{D}'} |\partial_x^\beta \partial_t^j f| + [f]_{a,\mathcal{D}'} + < f >_{a,\mathcal{D}'}$$

$$\text{with } [f]_{a,\mathcal{D}'} = \sum_{|\beta|+2j=k} \sup_{(s,y)\in\mathcal{D}'} \sup_{(t,x)\in\mathcal{D}'\setminus\{(s,y)\}} \frac{|\partial_x^\beta \partial_t^j f(t,x) - \partial_x^\beta \partial_t^j f(s,y)|}{[\mathbf{pd}((t,x),(s,y))]^\alpha}$$

$$\text{and } < f >_{a,\mathcal{D}'} = \begin{cases} \sum_{|\beta|+2j=k-1} \sup_{(s,x)\in\mathcal{D}'} \sup_{(t,x)\in\mathcal{D}'\setminus\{(s,x)\}} \frac{|\partial_x^\beta \partial_t^j f(t,x) - \partial_x^\beta \partial_t^j f(s,x)|}{|t-s|^{(\alpha+1)/2}} & \text{for } k\geq 1, \\ 0 \text{ for } k=0. \end{cases}$$

Whenever convenient, we may denote $(\mathcal{H}_a(\mathbb{R}\times\mathbb{R}^d), |\cdot|_{a,\mathbb{R}\times\mathbb{R}^d})$ by $(\mathcal{H}_a, |\cdot|_a)$. The following smoothness definition for the time-space domain $\mathcal{D}$ will be used (cf Friedman [28], page 64).

Definition 17.1. The domain $\mathcal{D}$ is of class $\mathcal{H}_a$ $(a \geq 1)$ ($\mathcal{D} \in \mathcal{H}_a$ for short) if, for every $(t_0, x_0) \in \overline{\partial\mathcal{D} \cup (]0,T[\times\mathbb{R}^d)}$, there exists a neighborhood $]t_0, t_0+\epsilon_0^2[\times B_d(x_0,\epsilon_0)$, an index i and a function $\phi_0 \in \mathcal{H}_a(]t_0,t_0+\epsilon_0^2[\times B_{d-1}((x_0^1,\cdots,x_0^{i-1},x_0^{i+1},\cdots,x_0^d),\epsilon_0))$ s.t

$$\begin{aligned}&\overline{\partial\mathcal{D}\cup(]0,T[\times\mathbb{R}^d)}\cup]t_0,t_0+\epsilon_0^2[\times B_d(x_0,\epsilon_0) :=\\ &\{(t,x)\in(]t_0,t_0+\epsilon_0^2[\cup[0,T])\times B_d(x_0,\epsilon_0) : x_i=\phi_0(t,x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_d)\}.\end{aligned}$$

Let

$$\begin{aligned}\mathcal{D}_0 &= \{x : (0,x)\in\partial\mathcal{D} - \overline{\partial\mathcal{D}\cup(]0,T[\times\mathbb{R}^d)}\},\\ \mathcal{D}_T &= \{x : (T,x)\in\partial\mathcal{D} - \overline{\partial\mathcal{D}\cup(]0,T[\times\mathbb{R}^d)}\}.\end{aligned}$$

$\mathcal{D}_0$ and $\mathcal{D}_T$ are open sets and we assume that they are non empty domains that coincide with the interior of their closure (cf Friedman [28], Section 3.2). We also assume (cf Friedman [28], Section 3.2) that the time section of $\mathcal{D}$, $\mathcal{D}_t = \{x : (t,x)\in\mathcal{D}\}$, is a domain that coincides with the interior of its closure, for every $t\in]0,T[$. If $\mathcal{D}$ is of class $\mathcal{H}_a$ $(a\geq 1)$ the set $\mathcal{PD} = \partial\mathcal{D} - \{0\}\times\mathcal{D}_0$ is the parabolic boundary of $\mathcal{D}$ (see Lieberman [69], pages 7 and 13).
If $\mathcal{D}$ is of class $\mathcal{H}_1$, $\mathcal{D}$ satisfies an exterior tusk condition, which is analogous to the exterior (Wiener's) cone condition for time independent domains. We refer to Lieberman [68] for more details on the following proposition.

Proposition 17.2 (Tusk condition)**.** *Assume $\mathcal{D}\in\mathcal{H}_1$. For some $R>0,\delta>0$, at any point $(t_0,x_0)\in\mathcal{PD}$, there is a tusk*

$$\mathcal{T} = \{(t,x) : t_0 < t < t_0+\delta, |x-x_0-\overline{x}_0\sqrt{t-t_0}|^2 < R^2(t-t_0)\},$$

for some $\overline{x}_0\in\mathbb{R}^d$, such that $\overline{\mathcal{T}}$ intersects $\overline{\mathcal{D}}$ only at (t_0,x_0).

If $\mathcal{D}$ is of class $\mathcal{H}_2$, all the domains $\mathcal{D}_t$, for $t\in[0,T]$, satisfy the uniform interior and exterior sphere condition with the same radius r_0. Moreover (see Lieberman [69], Section X.3), the signed spatial distance F given by

$$F(t,x) = \begin{cases} -\mathbf{d}(x,\partial\mathcal{D}_t), & \text{for } x\in\mathcal{D}_t^c, \mathbf{d}(x,\partial\mathcal{D}_t)\leq r_0 \quad 0<t<T,\\ \mathbf{d}(x,\partial\mathcal{D}_t), & \text{for } x\in\mathcal{D}_t, \mathbf{d}(x,\partial\mathcal{D}_t)\leq r_0 \quad 0<t<T,\end{cases}$$

belongs to $\mathcal{H}_2(\{(t,x) : 0<t<T, \mathbf{d}(x,\partial\mathcal{D}_t)<r_0\})$ and $\nabla F(t,x)$ is the unit normal vector to $\mathcal{D}_t$ at $\pi_{\partial\mathcal{D}_t}(x)$ the nearest point to x in $\partial\mathcal{D}_t$ (also called the projection of x on $\partial\mathcal{D}_t$). F can be extended as $\mathcal{H}_2([0,T]\times\mathbb{R}^d)$ function with bounded derivatives, preserving the sign. In the following, we denote $\forall r\in\mathbb{R}^+$, by $V_{\partial\mathcal{D}}(r) := \{(t,x)\in[0,T]\times\mathbb{R}^d : \mathbf{d}(x,\partial\mathcal{D}_t)\leq r\}$ a neighborhood of size r of the so called side.

Proposition 17.3 (Local diffeomorphism)**.** *Assume $\mathcal{D}'\subset\mathbb{R}^d$ is a domain of class C^2 with compact boundary $\partial\mathcal{D}'$. For all $s\in\partial\mathcal{D}'$, there are two open bounded sets U^s and V^s, a C^1-diffeomorphism F^s ($G^s=(F^s)^{-1}$) from U^s ($s\in U^s$) into $(-r_0,r_0)\times V^s$ such that*

$$F^s : \begin{cases} U^s\subset\mathbb{R}^d \to (-r_0,r_0)\times V^s\subset\mathbb{R}\times\mathbb{R}^{d-1},\\ x\longmapsto(\lambda,\overline{z}) := (\lambda,z_2,,\cdots,z_d) \textit{ such that } x = g_s(\overline{z})+\lambda\overrightarrow{n}(g_s(\overline{z})),\end{cases}$$

where g_s is a mapping of $\partial\mathcal{D}'$ in a neighborhood of s.

17.1.2 Known results

We still consider a d-dimensional process $X^{t,x}$ satisfying (16.1), which starting point (t,x) is in $\overline{\mathcal{D}}$. We introduce the infinitesimal generator of X

$$\mathcal{L}u(t,x) = \nabla u(t,x)b(t,x) + \frac{1}{2}\mathrm{Tr}(Hu(t,x)[\sigma\sigma^*](t,x)),$$

where Hu denotes the Hessian matrix of u. We also define

Definition 17.4 (Definition of $\tau^{t,x}$).

$$\tau^{t,x} := \inf\{s > t : (s, X_s^{t,x}) \notin \mathcal{D}\}$$

is the first exit time from a domain $\mathcal{D}$ for the time-space process $(s, X_s^{t,x})_{s\in[t,T]}$.

Note that $\tau^{t,x}$ is a bounded stopping time, since $\tau^{t,x} \leq T$. We focus on the expectation of functionals of the process X stopped at the exit from $\mathcal{D}$, of the form

$$u(t,x) = \mathbb{E}\left(g(\tau^{t,x}, X_{\tau^{t,x}}^{t,x})e^{-\int_t^{\tau^{t,x}} c(s,X_s^{t,x})ds} - \int_t^{\tau^{t,x}} e^{-\int_t^s c(r,X_r^{t,x})dr} f(s, X_s^{t,x})ds\right), \quad (17.1)$$

and on its sensitivity w.r.t. $\mathcal{D}$. The data f, g, c are bounded continuous functions on $\mathbb{R}^{d+1}$. First, we recall a result which relates u to the solution of a Cauchy-Dirichlet type PDE in the time-space domain $\mathcal{D}$.

Proposition 17.5 (Feynman-Kac's formula and a priori estimates on u). *Assume Hypothesis 16.1, $\mathcal{D} \in \mathcal{H}_1$, $c \in \mathcal{H}_\alpha$, $f \in \mathcal{H}_\alpha$ and $g \in C^{0,0}$ with $\alpha \in]0,1[$. Then, u is the unique solution of class $C^{1,2}(\mathcal{D}) \cap C^{0,0}(\overline{\mathcal{D}})$ to*

$$\begin{cases} \partial_t u + \mathcal{L}u - cu = f & \text{in } \mathcal{D}, \\ u = g & \text{on } \mathcal{PD}. \end{cases}$$

In addition, if $\mathcal{D}$ is of class $\mathcal{H}_{1+\alpha}$ and $g \in \mathcal{H}_{1+\alpha}$, the function u belongs to $\mathcal{H}_{1+\alpha}(\mathcal{D})$ and it holds $|u|_{1+\alpha,\mathcal{D}} \leq K(|f|_{\alpha,\mathcal{D}} + |g|_{1+\alpha,\mathcal{D}})$ (in particular ∇u is well defined and continuous up to the boundary).

Second, we recall the main contribution of Costantini et al. [22], namely the sensitivity of $\mathbb{E}\left(g(\tau^{t,x}, X_{\tau^{t,x}}^{t,x})e^{-\int_t^{\tau^{t,x}} c(s,X_s^{t,x})ds} - \int_t^{\tau^{t,x}} e^{-\int_t^s c(r,X_r^{t,x})dr} f(s, X_s^{t,x})ds\right)$ w.r.t. spatial perturbations of $\mathcal{D}$. We define a spatial perturbation of the time-space domain $\mathcal{D}$ in the following way:

Definition 17.6 (Spatial perturbation).

$$\mathcal{D}^\epsilon = \{(t,x) : (t, x + \epsilon\theta(t,x)) \in \mathcal{D}\}, \quad \epsilon \in \mathbb{R},$$

for some map $\theta : [0,T] \times \mathbb{R}^d \longmapsto \mathbb{R}^d$. Moreover, we assume θ is a function of class $C_b^{1,2}([0,T] \times \mathbb{R}^d)$.

Theorem 17.7. *Assume Hypothesis 16.1, $\mathcal{D} \in \mathcal{H}_{1+\alpha}$, $c \in \mathcal{H}_\alpha$, $f \in \mathcal{H}_\alpha$ and $g \in \mathcal{H}_{1+\alpha}$ with $\alpha \in]0,1[$. Let (t,x) be in $\mathcal{D} \cup \mathcal{D}_0$ and set*

$$\tau_\epsilon^{t,x} := \inf\{s > t : (s, X_s^{t,x}) \notin \mathcal{D}^\epsilon\}.$$

Then, $u^\epsilon(t,x) = \mathbb{E}\left(g(\tau_\epsilon^{t,x}, X_{\tau_\epsilon^{t,x}}^{t,x}) e^{-\int_t^{\tau_\epsilon^{t,x}} c(s,X_s^{t,x})ds} - \int_t^{\tau_\epsilon^{t,x}} e^{-\int_t^s c(r,X_r^{t,x})dr} f(s, X_s^{t,x})ds \right)$ *is differentiable w.r.t. ϵ at $\epsilon = 0$ and*

$$\partial_\epsilon u^\epsilon(t,x)\big|_{\epsilon=0} = \mathbb{E}\left(e^{-\int_t^{\tau^{t,x}} c(s,X_s^{t,x})ds} [(\nabla u - \nabla g)\theta](\tau^{t,x}, X_{\tau^{t,x}}^{t,x}) \right).$$

Remark 17.8. Note that ∇u in the above expression is well defined on the boundary since u is of class $\mathcal{H}_{1+\alpha}(\mathcal{D})$. In view of the formula above and because $u = g$ on $\mathcal{PD}$, only normal perturbations of θ have an impact on the derivative of $u^\epsilon(t,x)$.

17.1.3 Application to American options

Looking at (16.3), we notice that P satisfies an equation similar to the one appearing in Proposition 17.5, where $\mathcal{D}$ corresponds to the unknown continuation region $\mathcal{C}$. Finding the price of an American option can be seen as determining $\mathcal{C}$. Assume we want to solve the following equation

$$\begin{cases} \partial_t v_k(t,x) + \mathcal{A}_t v_k(t,x) = 0 & \text{in } \mathcal{D}_k, \\ v_k = g & \text{on } \mathcal{PD}_k, \end{cases}$$

where $\mathcal{D}_k \in \mathcal{H}_{1+\alpha}$ is a sequence of domains. The Feynman-Kac formula gives

$$v_k(t,x) = \mathbb{E}\left(e^{-\int_t^{\tau_k^{t,x}} r(s,X_s^{t,x})ds} g(\tau_k^{t,x}, X_{\tau_k^{t,x}}^{t,x}) \right),$$

where $\tau_k^{t,x} = \inf\{s > t : (s, X_s^{t,x}) \notin \mathcal{D}_k\}$. If $\mathcal{D}_k = \mathcal{C}$, we get $v_k = P$ and since $P(t,x)$ is also given by (16.5), $\partial_\epsilon v_k^\epsilon(t,x)\big|_{\epsilon=0}$ must be null. This enables us to give an intuitive idea of a new way for numerically pricing American options :

Algorithm 1.

- *Step 0: We start with a domain $\mathcal{D}_0$.*
- *Step k: Assume $\mathcal{D}_{k-1}$ and $\partial_\epsilon v_{k-1}^\epsilon(t,x)\big|_{\epsilon=0}$ are known. We build $\mathcal{D}_k$ from $\mathcal{D}_{k-1}$. We perturb $\mathcal{D}_k$ to get $\mathcal{D}_k^\epsilon$ in the gradient direction.*
- *Using Theorem 17.7, we get an exact expression (modulo Monte Carlo simulations) for $\partial_\epsilon v_k^\epsilon(t,x)\big|_{\epsilon=0}$, which is $\partial_\epsilon v_k^\epsilon(t,x)\big|_{\epsilon=0} = \mathbb{E}\left(e^{-\int_t^{\tau_k^{t,x}} r(s,X_s^{t,x})ds} [(\nabla v_k - \nabla g)\theta](\tau_k^{t,x}, X_{\tau_k^{t,x}}^{t,x}) \right)$.*
- *We compare $\partial_\epsilon v_k^\epsilon(t,x)\big|_{\epsilon=0}$ with $\partial_\epsilon v_{k-1}^\epsilon(t,x)\big|_{\epsilon=0}$. If the first one is smaller than the second, we keep this choice for $\mathcal{D}_k$, otherwise we change it.*

- *We continue until we find a k_0 for which $\partial_\epsilon v^\epsilon_{k_0}(t,x)\big|_{\epsilon=0}$ is smaller than a small threshold.*

Remark 17.9. This algorithm is not rigorous. Its vocation is to give a general idea of the application of Theorem 17.7 to the pricing of American options. In particular, no details concerning the construction of the sequence $\mathcal{D}_k$ are given.

Remark 17.10. The smooth-fit condition enables us to justify in an other way the fact that $\partial_\epsilon v^\epsilon_k(t,x)\big|_{\epsilon=0}$ must be null for $\mathcal{D}_k = \mathcal{C}$. If $\mathcal{D}_k = \mathcal{C}$, we get $v_k = P$ and then $(\nabla P - \nabla g)\theta](\tau^{t,x}_k, X^{t,x}_{\tau^{t,x}_k}) = 0$, which yields $\partial_\epsilon v^\epsilon_k(t,x)\big|_{\epsilon=0} = 0$.

17.2 A sensitivity formula in discrete time

As it has been done by Costantini et al. [22], we would like to get a formula for $\partial_\epsilon u^\epsilon(t,x)\big|_{\epsilon=0} = 0$ in a discrete time setting, e.g. when u, defined by (17.1), is computed for $\tau^{t,x}$ with values in $\mathcal{T}^N_{t,T}$ (see the beginning of Section 16.3 for a definition). Since we aim at using this formula for the pricing of Bermudan options, we only consider the case $f \equiv 0$.

17.2.1 Definitions and Notations

Definition 17.11 (Definition of $\mathcal{D}_{t_j}$ and $\mathcal{D}^N$)**.** Let j be an integer of $\{0,\cdots,N\}$. We recall that $\mathcal{D}_{t_j}$ belongs to $\mathbb{R}^d$ and corresponds to the section of $\mathcal{D}$ at $t = t_j$, i.e. to $\mathcal{D}_{t_j} = \mathcal{D}\big|_{t=t_j}$. Then, $(t_j, \mathcal{D}_{t_j}) \in [0,T] \times \mathbb{R}^d$, and we define

$$\mathcal{D}^N = \bigcup_{j=1}^{N} (t_j, \mathcal{D}_{t_j}).$$

Definition 17.12 (Definition of $\tau^{t,x}_N$)**.**

$$\tau^{t,x}_N = \inf\{t_i > t : X^{t,x}_{t_i} \notin \mathcal{D}_{t_i}\}$$

is the first exit time from a domain $\mathcal{D}^N$ for the time-space process $(t_i, X^{t,x}_{t_i})_{t_i \in [t,T]}$. From the definition of $\tau^{t,x}_N$, we deduce that $\tau^{t,x}_N$ takes values in $\mathcal{T}^N_{t,T}$.

In this part, we focus on the expectation of functionals of the process X stopped as soon as $(t_i, X^{t,x}_{t_i}) \notin \mathcal{D}^N$. Let us define u_N, the discrete version of u

$$u_N(t,x) = \mathbb{E}\left(e^{-\int_t^{\tau^{t,x}_N} c(s,X^{t,x}_s)ds} g(\tau^{t,x}_N, X^{t,x}_{\tau^{t,x}_N}) \right).$$

Note that u_N is the continuation value function of the Bermudan options, so that the price function is given by $P^N(t_n,\cdot) = \max(g(t_n,\cdot), u_N(t_n,\cdot))$ (see Section 16.3). We define a spatial perturbation of the space domain $\mathcal{D}_{t_j}$ in the following way:

Definition 17.13 (Definition of $\mathcal{D}^\epsilon_{t_j}$ and $\mathcal{D}^{N,\epsilon}$)**.**

$$\mathcal{D}^\epsilon_{t_j} = \{x : x + \epsilon\theta(t,x) \in \mathcal{D}_{t_j}\}, \quad \epsilon \in \mathbb{R}$$

for some map $\theta : [0,T] \times \mathbb{R}^d \longmapsto \mathbb{R}^d$. Moreover, we assume θ is a function of class $C_b^{1,2}([0,T] \times \mathbb{R}^d)$. We also define

$$\mathcal{D}^{N,\epsilon} = \bigcup_{j=1}^{N} (t_j, \mathcal{D}^{\epsilon}_{t_j}).$$

It remains to introduce $\tau^{t,x}_{N,\epsilon}$

Definition 17.14 (Definition of $\tau^{t,x}_{N,\epsilon}$)**.**

$$\tau^{t,x}_{N,\epsilon} = \inf\{t_i > t : X^{t,x}_{t_i} \notin \mathcal{D}^{\epsilon}_{t_i}\}$$

is the first exit time from a domain $\mathcal{D}^{N,\epsilon}$ for the time-space process $(t_i, X^{t,x}_{t_i})_{t_i \in [t,T]}$. Clearly, $\tau^{t,x}_{N,\epsilon}$ takes values in $\mathcal{T}^N_{t,T}$.

We aim at computing $\partial_\epsilon u^\epsilon_N(t,x)\big|_{\epsilon=0} = 0$, where

$$u^\epsilon_N(t,x) = \mathbb{E}\left(e^{-\int_t^{\tau^{t,x}_{N,\epsilon}} c(s,X^{t,x}_s)ds} g(\tau^{t,x}_{N,\epsilon}, X^{t,x}_{\tau^{t,x}_{N,\epsilon}}) \right).$$

17.2.2 Main result

Theorem 17.15. *Assume Hypothesis 16.1 and $c \equiv 0$. Let p denote the transition density of X. Then, $\forall (t,x) \in \mathcal{D}$, it holds*

$$\partial_\epsilon u^\epsilon_N(t,x)\big|_{\epsilon=0} =$$
$$\sum_{j=0,t_j>t}^{N-1} \mathbb{E}_{t,x}\left(\mathbf{1}_{\tau^{t,x}_N > t_j} \int_{\partial\mathcal{D}_{t_{j+1}}} p(t_j, X_{t_j}; t_{j+1}, m)(g - u_N)(t_{j+1}, m)\theta \cdot \overrightarrow{n}(t_{j+1}, m) d\sigma_m \right),$$

where $d\sigma_m$ denotes the boundary integral.

Remark 17.16. Theorem 17.15 only deals with the case $c \equiv 0$. However, the case $c \neq 0$ can be done following the same proof.

Remark 17.17. In the case of Bermudan options, we know that $u_N(s,y) = g(s,y)$ for y at the boundary of $\mathcal{C}$. Hence, looking at the formula stated in Theorem 17.15, we notice that the r.h.s. is null as soon as $\mathcal{D} = \mathcal{C}$. Then, $\partial_\epsilon u^\epsilon_N(t,x)\big|_{\epsilon=0}$ is null as soon as $\mathcal{D} = \mathcal{C}$. This corroborates the result of Proposition 16.11.

Lemma 17.18. *Assume Hypothesis 16.1. Let f be a continuous function. p denotes the transition density of X. Then, it holds for all $t_i, t_j \in \mathcal{T}^N_{t,T}$ such that $i < j$*

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \mathbb{E}\left(f(t_j, X^{t_i,x}_{t_j}) \mathbf{1}_{X_{t_j} \in \mathcal{D}^\epsilon_{t_j}, X_{t_j} \notin \mathcal{D}_{t_j}} \right) = \int_{\partial\mathcal{D}_{t_j}} p(t_i, x; t_j, m) f(t_j, m)(\theta \cdot \overrightarrow{n})^-(t_j, m) d\sigma_m.$$

Proof of Lemma 17.18. Note first that it is enough to prove the limit for positive functions, since any function can be written as $f = f^+ - f^-$. For the sake of simplicity, we do the proof for $i = 0$ and $j = 1$. Let us compute $I^\epsilon := \mathbb{E}\left(f(t_1, X_{t_1}^{t,x})(\mathbf{1}_{X_{t_1} \in \mathcal{D}_{t_1}^\epsilon, X_{t_1} \notin \mathcal{D}_{t_1}})\right)$. Using the transition density of X yields

$$I^\epsilon = \int_{D_{t_1}^\epsilon \setminus D_{t_1}} f(t_1, y) p(t, x; t_1, y) dy.$$

For ϵ small enough, we get $\mathcal{D}_{t_1}^\epsilon \setminus \mathcal{D}_{t_1} \subset V_{r_0}(\partial \mathcal{D}_{t_1})$. Since $\partial \mathcal{D}_{t_1}$ is compact, there exists a finite number of points $(s_i)_{1 \le i \le k}$ in $\partial \mathcal{D}_{t_1}$ such that $V_{r_0}(\partial \mathcal{D}_{t_1}) \in \bigcup_{1 \le i \le k} U^i$, where $U^i = B_\infty(s_i, \epsilon_0)$. Then, we construct a partition of unity corresponding to $(U^i)_{1 \le i \le k}$, i.e. non negative C_b^∞ functions $(\chi_i)_{1 \le i \le k}$ verifying $\mathrm{Supp}(\chi_i) \subset U^i$ and $\sum_{i=1}^k \chi_i \equiv 1$ on $\overline{\mathcal{D}_{1,r_0}}$ (i.e. $\overline{\mathcal{D}_{t_1} \cup V_{r_0}(\partial \mathcal{D}_{t_1})}$). We get

$$I^\epsilon = \sum_{i=1}^k \int_{U^i \cap \mathcal{D}_{t_1}^\epsilon \setminus \mathcal{D}_{t_1}} \chi_i(y) f(t_1, y) p(t, x; t_1, y) dy := \sum_{i=1}^k I_i^\epsilon. \tag{17.2}$$

Let us compute I_i^ϵ

$$I_i^\epsilon = \int_{U^i \cap \mathcal{D}_{t_1}^\epsilon \setminus \mathcal{D}_{t_1}} \chi_i(y) f(t_1, y) p(t, x; t_1, y) dy := \int_{U^i \cap \mathcal{D}_{t_1}^\epsilon \setminus \mathcal{D}_{t_1}} F(y) dy. \tag{17.3}$$

First, let us work on $U^i \cap \mathcal{D}_{t_1}^\epsilon \setminus \mathcal{D}_{t_1}$. Using Definition 17.1, we can assume without loss of generality that on U^i, the boundary of $\mathcal{D}_{t_1}$ is parametrised w.r.t. y as $y_1 = \phi(\overline{y})$, where $\overline{y} = (y_2, \cdots, y_d)$, $\phi \in \mathcal{H}_{1+\alpha}$. Then, $U^i \cap \partial \mathcal{D}_{t_1} = \{y \in U^i : y_1 = \phi(\overline{y})\}$, and $U^i \cap \mathcal{D}_{t_1}^c = \{y \in U^i : y_1 \ge \phi(\overline{y})\}$. The outward normal vector to $\partial \mathcal{D}_{t_1}$ at point $y = (\phi(\overline{y}), \overline{y})$ is given by

$$\overrightarrow{n}(\overline{y}) = \frac{1}{\sqrt{1 + |\nabla \phi(\overline{y})|^2}} \begin{pmatrix} 1 \\ -\partial_{y_2} \phi(\overline{y}) \\ . \\ . \\ -\partial_{y_d} \phi(\overline{y}) \end{pmatrix}.$$

Moreover, $\mathcal{D}_{t_1}^\epsilon \cap U^i = \{y \in U^i : y + \epsilon \theta(t_1, y) \in \mathcal{D}_{t_1}\} = \{y \in U^i : y_1 + \epsilon \theta_1(t_1, y) < \phi(y_2 + \theta_2(t_1, y), \cdots, y_d + \theta_d(t_1, y))\}$. Thus,

$$U^i \cap \mathcal{D}_{t_1}^\epsilon \cap \mathcal{D}_{t_1}^c = \{y \in U^i : \phi(\overline{y}) \le y_1 \le \phi(y_2 + \theta_2(t_1, y), \cdots, y_d + \theta_d(t_1, y)) - \epsilon \theta(t_1, y)\}.$$

Besides, since $\phi \in \mathcal{H}_{1+\alpha}(U^i)$, a Taylor expansion leads to

$$\begin{aligned} &\phi(\overline{y}) - \epsilon \sqrt{1 + |\nabla \phi(\overline{y})|^2} \overrightarrow{n}(\overline{y}) \cdot \theta(t_1, y) \\ &\qquad - c\epsilon^{1+\alpha} \le \phi(y_2 + \theta_2(t_1, y), \cdots, y_d + \theta_d(t_1, y)) - \epsilon \theta(t_1, y) \le \phi(\overline{y}) \\ &\qquad\qquad - \epsilon \sqrt{1 + |\nabla \phi(\overline{y})|^2} \overrightarrow{n}(\overline{y}) \cdot \theta(t_1, y) + c\epsilon^{1+\alpha}. \end{aligned}$$

In addition, on $U^i \cap \mathcal{D}_{t_1}^\epsilon \cap \mathcal{D}_{t_1}^c$, one has $|y_1 - \phi(\overline{y})| \le c\epsilon$. Thus, we can replace $\theta(t_1, y)$ by $\theta(t_1, (\phi(\overline{y}), \overline{y}))$ in the above inequality. This leads to

$$A_i^- \subset U^i \cap \mathcal{D}_{t_1}^\epsilon \cap \mathcal{D}_{t_1}^c \subset A_i^+,$$

where

$$A_i^{+/-} = \{y \in U^i : \phi(\overline{y}) \le y_1 \le \phi(\overline{y}) - \epsilon\sqrt{1+|\nabla\phi(\overline{y})|^2}\overrightarrow{n}(\overline{y}) \cdot \theta(t_1, (\phi(\overline{y}), \overline{y})) \pm c\epsilon^{1+\alpha}\}.$$

Since the function F is positive, we can use this result to bound (17.3). One gets

$$\int_{A_i^-} F(y)dy \le I_i^\epsilon \le \int_{A_i^+} F(y)dy.$$

It remains to prove that both upper and lower bounds divided by ϵ converge to the same limit as $\epsilon \to 0$. We only detail the upper bound. Because of the specific form of A_i^+, it is relevant to integrate w.r.t. y_1, which directly leads to

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_{A_i^+} F(y)dy = \int_{B_\infty^{d-1}(s^i,\epsilon_0)} F(\phi(\overline{y}),\overline{y})\sqrt{1+|\nabla\phi(\overline{y})|^2}[\overrightarrow{n}(\overline{y}) \cdot \theta(t_1,(\phi(\overline{y}),\overline{y}))]^- d\overline{y}.$$

Since
$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_{A_i^-} F(y)dy = \int_{B_\infty^{d-1}(s^i,\epsilon_0)} F(\phi(\overline{y}),\overline{y})\sqrt{1+|\nabla\phi(\overline{y})|^2}[\overrightarrow{n}(\overline{y}) \cdot \theta(t_1,(\phi(\overline{y}),\overline{y}))]^- d\overline{y}$, we get

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}I_i^\epsilon = \int_{B_\infty^{d-1}(s^i,\epsilon_0)} F(\phi(\overline{y}),\overline{y})\sqrt{1+|\nabla\phi(\overline{y})|^2}[\overrightarrow{n}(\overline{y}) \cdot \theta(t_1,(\phi(\overline{y}),\overline{y}))]^- d\overline{y}.$$

Combining this result and (17.2) ends the proof. □

Remark 17.19. Under the assumptions of Lemma 17.18, a similar proof yields to

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\mathbb{E}\left(f(t_j, X_{t_j}^{t_i,x})\mathbf{1}_{X_{t_j}\in\mathcal{D}_{t_j}, X_{t_j}\notin\mathcal{D}_{t_j}^\epsilon}\right) = \int_{\partial\mathcal{D}_{t_j}} p(t_i,x;t_j,m)f(t_j,m)(\theta\cdot\overrightarrow{n})^+(t_j,m)d\sigma_m.$$

Remark 17.20. Since $\mathbf{1}_{X_{t_j}\in\mathcal{D}_{t_j}} - \mathbf{1}_{X_{t_j}\in\mathcal{D}_{t_j}^\epsilon} = \mathbf{1}_{X_{t_j}\in\mathcal{D}_{t_j}, X_{t_j}\notin\mathcal{D}_{t_j}^\epsilon} - \mathbf{1}_{X_{t_j}\in\mathcal{D}_{t_j}^\epsilon, X_{t_j}\notin\mathcal{D}_{t_j}}$ we combine Lemma 17.18 and Remark 17.19 to get

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\mathbb{E}\left(f(t_j, X_{t_j}^{t_i,x})\mathbf{1}_{X_{t_j}\in\mathcal{D}_{t_j}} - \mathbf{1}_{X_{t_j}\in\mathcal{D}_{t_j}^\epsilon}\right) = \int_{\partial\mathcal{D}_{t_j}} p(t_i,x;t_j,m)f(t_j,m)(\theta\cdot\overrightarrow{n})(t_j,m)d\sigma_m.$$

Proof of Theorem 17.15. We aim at proving the following more general result: $\forall i \in \{0,\cdots,N-1\}$, it holds for $1 \le k \le N-i$

$$\partial_\epsilon u_{N,k}^\epsilon(t_i,x)\big|_{\epsilon=0} = \sum_{j=i}^{k+i-1}\mathbb{E}_{t_i,x}\left(\mathbf{1}_{\tau_N^{t_i,x}>t_j}\int_{\partial\mathcal{D}_{t_{j+1}}} p(t_j,X_{t_j};t_{j+1},y)(g-u_N)(t_{j+1},y)\theta\cdot\overrightarrow{n}(t_{j+1},y)d\sigma_y\right), \tag{17.4}$$

where $u_{N,k}^\epsilon(t_i,x) = \mathbb{E}_{t_i,x}\left(g(\tau_{N,k,\epsilon}^{t_i,x}, X_{\tau_{N,k,\epsilon}^{t_i,x}})\right)$, $\tau_{N,k,\epsilon}^{t_i,x} = \inf\{t_l > t_i : X_{t_l}^{t_i,x} \notin \mathcal{D}_{l,k}^{N,\epsilon}\}$ and

$$\mathcal{D}_{l,k}^{N,\epsilon} := \bigcup_{j=l+1}^{k+l}(t_j,\mathcal{D}_{t_j}^\epsilon)\bigcup_{j=l+k+1}^{N}(t_j,\mathcal{D}_{t_j}).$$

In particular, $\mathcal{D}_{0,N}^{N,\epsilon} = \mathcal{D}^{N,\epsilon}$. This result is more general than the one stated in Theorem 17.15, but it is also less readable because of several indexes, that's why we only present in Theorem 17.15 the case $k = N$. We prove Theorem 17.15 recursively.

- First, let us prove (17.4) for $k = 1$ and for any i. In that case, $\mathcal{D}_{i,1}^{N,\epsilon} = (t_{i+1}, \mathcal{D}_{i+1}^{\epsilon}) \bigcup_{j=i+2}^{N} (t_j, \mathcal{D}_{t_j})$. We have to compute

$$\frac{u_{N,1}^{\epsilon}(t_i, x) - u_N(t_i, x)}{\epsilon}$$

and to take the limit when ϵ goes to 0. We develop $u_{N,1}^{\epsilon}(t_i, x)$

$$\begin{aligned} u_{N,1}^{\epsilon}(t_i, x) &= \mathbb{E}_{t_i,x}\left(g(\tau_{N,1,\epsilon}^{t_i,x}, X_{\tau_{N,1,\epsilon}^{t_i,x}})\right), \\ &= \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{i+1}^{\epsilon}} g(\tau_{N,1,\epsilon}^{t_i,x}, X_{\tau_{N,1,\epsilon}^{t_i,x}})\right) + \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}}^{\epsilon}} g(t_{i+1}, X_{t_{i+1}})\right). \end{aligned} \tag{17.5}$$

We can write $\mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{t_{i+1}}^{\epsilon}} g(\tau_{N,1,\epsilon}^{t_i,x}, X_{\tau_{N,1,\epsilon}^{t_i,x}})\right) = \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{t_{i+1}}^{\epsilon}} \mathbb{E}[g(\tau_{N,1,\epsilon}^{t_i,x}, X_{\tau_{N,1,\epsilon}^{t_i,x}}) \big| \mathcal{F}_{t_{i+1}}]\right)$. On the indicator $\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{t_{i+1}}^{\epsilon}}$, we have $\tau_{N,1,\epsilon}^{t_i,x} = \tau_N^{t_{i+1}, X_{t_{i+1}}^{t_i,x}}$, and then $\mathbb{E}_{t_i,x}[g(\tau_{N,1,\epsilon}^{t_i,x}, X_{\tau_{N,1,\epsilon}^{t_i,x}}) \big| \mathcal{F}_{t_{i+1}}] = u_N(t_{i+1}, X_{t_{i+1}})$. Thus, (17.5) becomes

$$u_{N,1}^{\epsilon}(t_i, x) = \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{t_{i+1}}^{\epsilon}} u_N(t_{i+1}, X_{t_{i+1}})\right) + \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}}^{\epsilon}} g(t_{i+1}, X_{t_{i+1}})\right).$$

Noticing that

$$u_N(t_i, x) = \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{t_{i+1}}} u_N(t_{i+1}, X_{t_{i+1}})\right) + \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}}} g(t_{i+1}, X_{t_{i+1}})\right), \tag{17.6}$$

we get

$$u_{N,1}^{\epsilon}(t_i, x) - u_N(t_i, x) = \mathbb{E}_{t_i,x}\left((u_N - g)(t_{i+1}, X_{t_{i+1}})(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{t_{i+1}}^{\epsilon}} - \mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{t_{i+1}}})\right),$$

and Remark 17.20 enables to get

$$\begin{aligned} &\lim_{\epsilon \to 0} \frac{u_{N,1}^{\epsilon}(t_i, x) - u_N(t_i, x)}{\epsilon} = \\ &\qquad \mathbb{E}_{t_i,x}\left(\int_{\partial \mathcal{D}_{t_{i+1}}} p(t_i, X_{t_i}; t_{i+1}, m)(g - u_N)(t_{i+1}, m)\theta \cdot \overrightarrow{n}(t_{i+1}, m) d\sigma_m\right), \end{aligned} \tag{17.7}$$

and the case $k = 1$ is proved for any i.

- Assume (17.4) is true for $k - 1$. We have to compute

$$\frac{u_{N,k}^{\epsilon}(t_i, x) - u_N(t_i, x)}{\epsilon}$$

and to take the limit when ϵ goes to 0. We develop $u_{N,k}^{\epsilon}(t_i, x)$

$$\begin{aligned} u_{N,k}^{\epsilon}(t_i, x) &= \mathbb{E}_{t_i,x}\left(g(\tau_{N,k,\epsilon}^{t_i,x}, X_{\tau_{N,k,\epsilon}^{t_i,x}})\right), \\ &= \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \in \mathcal{D}_{i+1}^{\epsilon}} g(\tau_{N,k,\epsilon}^{t_i,x}, X_{\tau_{N,k,\epsilon}^{t_i,x}})\right) + \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}}^{\epsilon}} g(t_{i+1}, X_{t_{i+1}})\right). \end{aligned} \tag{17.8}$$

We can write $\mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}^\epsilon_{t_{i+1}}} g(\tau^{t_i,x}_{N,k,\epsilon}, X_{\tau^{t_i,x}_{N,k,\epsilon}})\right) = \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}^\epsilon_{t_{i+1}}} \mathbb{E}[g(\tau^{t_i,x}_{N,k,\epsilon}, X_{\tau^{t_i,x}_{N,1,\epsilon}})\big|\mathcal{F}_{t_{i+1}}]\right)$. On the event $\{X_{t_{i+1}} \in \mathcal{D}^\epsilon_{t_{i+1}}\}$, we have $\tau^{t_i,x}_{N,k,\epsilon} = \tau^{t_{i+1},X_{t_{i+1}}}_{N,k-1,\epsilon}$, and then $\mathbb{E}_{t_i,x}[g(\tau^{t_i,x}_{N,k,\epsilon}, X_{\tau^{t_i,x}_{N,1,\epsilon}})\big|\mathcal{F}_{t_{i+1}}] = u^\epsilon_{N,k-1}(t_{i+1}, X^{t_i,x}_{t_{i+1}})$. Equality (17.8) becomes

$$u^\epsilon_{N,k}(t_i,x) = \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}^\epsilon_{t_{i+1}}} u^\epsilon_{N,k-1}(t_{i+1}, X_{t_{i+1}})\right) + \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}}\notin\mathcal{D}^\epsilon_{t_{i+1}}} g(t_{i+1}, X_{t_{i+1}})\right).$$

Using (17.6) and introducing $\pm\mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}^\epsilon_{t_{i+1}}} u_N(t_{i+1}, X_{t_{i+1}})\right)$ in the difference $u^\epsilon_{N,k}(t_i,x) - u_N(t_i,x)$ yield

$$\begin{aligned} u^\epsilon_{N,k}(t_i,x) - u_N(t_i,x) =& \mathbb{E}_{t_i,x}\left((u_N - g)(t_{i+1}, X_{t_{i+1}})(\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}^\epsilon_{t_{i+1}}} - \mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}})\right) \\ &+ \mathbb{E}_{t_i,x}\left(\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}^\epsilon_{t_{i+1}}}[u^\epsilon_{N,k-1}(t_{i+1}, X_{t_{i+1}}) - u_N(t_{i+1}, X_{t_{i+1}})]\right). \end{aligned}$$

Using Remark (17.20), the hypothesis of recurrence and the equality $\mathbf{1}_{\tau_N^{t_{i+1},X_{t_{i+1}}}>t_j} \mathbf{1}_{X^{t_i,x}_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}} = \mathbf{1}_{\tau_N^{t_i,x}>t_j}$, we get

$$\begin{aligned} &\lim_{\epsilon\to 0} \frac{u^\epsilon_{N,k}(t_i,x) - u_N(t_i,x)}{\epsilon} = \\ &\mathbb{E}_{t_i,x}\left\{\int_{\partial\mathcal{D}_{t_{i+1}}} p(t_i, X_{t_i}; t_{i+1}, m)(g - u_N)(t_{i+1}, m)\theta\cdot\overrightarrow{n}(t_{i+1}, m)d\sigma_m \right. \\ &\left. + \sum_{j=i+1}^{k+i-1} \mathbb{E}_{t_{i+1},X^{t_i,x}_{t_{i+1}}}\left(\mathbf{1}_{\tau_N^{t,x}>t_j}\int_{\partial\mathcal{D}_{t_{j+1}}} p(t_j, X_{t_j}; t_{j+1}, y)(g - u_N)(t_{j+1}, y)\theta\cdot\overrightarrow{n}(t_{j+1}, y)d\sigma_y\right)\right\}, \end{aligned}$$

and the result follows.

□

Chapter 18

Robustness of $\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$ with respect to N

In the previous chapter, we have proved Theorem 17.15: under Hypothesis 16.1 and with $c \equiv 0$, $\forall (t,x) \in \mathcal{D}$, it holds

$$\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0} = \tag{18.1}$$

$$\sum_{j=0, t_j>t}^{N-1} \mathbb{E}_{t,x}\left(\mathbf{1}_{\tau_N^{t,x}>t_j} \int_{\partial \mathcal{D}_{t_{j+1}}} p(t_j, X_{t_j}; t_{j+1}, y)(g-u_N)(t_{j+1}, y)\theta \cdot \overrightarrow{n}(t_{j+1}, y) d\sigma_y\right), \tag{18.2}$$

where p denotes the transition density of X, and $u_N(t,x) = \mathbb{E}\left(g(\tau_N^{t,x}, X^{t,x}_{\tau_N^{t,x}})\right) = \mathbb{E}_{t,x}[g(\tau_N, X_{\tau_N})]$. We also recall that $\{t_0 < \cdots < t_N\}$ is a regular subdivision of $[0,T]$, with $t_0 = 0$, $t_N = T$ and step $h = \frac{T}{N}$.

At first sight, the above sensitivity estimate strongly depends on N. Indeed, it consists of a sum of N terms. Moreover, the transition density may give very large terms when $t_{j+1} - t_j \to 0$. The aim of the chapter is to analyse more precisely the dependency of $\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$ w.r.t. N. This will help us in the future implementation work: an efficient algorithm would certainly take advantage of an a priori knowledge of the behaviour of this gradient. This dependency w.r.t. N may be an important issue, since numerical algorithms for pricing Bermudan options often face problems as the exercise frequency increases. This formula can also help us to study $\lim_{N\to\infty} \partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$: do we get the American case, i.e. the sensitivity formula of Costantini et al. [22] (see Theorem 17.7), or a similar one?
In this chapter, we first address the issue of the convergence of $\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$ when N tends to infinity. It appears that this is a highly non trivial problem which is deeply related to the overshoot of the diffusion process. We present in Section 18.1 partial results to analyse the convergence of $\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$, whose limit seems to be $\partial_\epsilon u_\epsilon(t,x)\big|_{\epsilon=0}$. For this, we use recent results on overshoots stated by Gobet and Menozzi [43] and establish stronger results. Nevertheless, our proof of convergence is incomplete and relies on an unproved (but reasonable) conjecture.

In Section 18.2, we fully prove that if $g \in \mathcal{H}_{1+\alpha}$, the gradient $\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$ is uniformly bounded w.r.t. N. In Section 18.3, we extend this result to the case g is the payoff of a put option (piecewise C^1). At last, Section 18.4 brings together technical results used through the chapter. These results are essentially the adaptation to time-dependent domains of known results stated in the case of time-independent domains.

In the whole chapter, we assume Hypothesis 16.1, i.e.

Hypothesis 18.1 *Let α be in $]0,1]$.*

1. *Smoothness of order $1+\alpha$. b and σ belong to $\mathcal{H}_{1+\alpha}$, $\alpha \in]0,1]$* (see the definition of $\mathcal{H}_{1+\alpha}$ at the beginning of Section 17),
2. *Uniform ellipticity. For some $a_0 > 0$, it holds $\xi \cdot [\sigma\sigma^*](t,x)\xi \geq a_0|\xi|^2$ for any $(t,x,\xi) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d$,*
3. *$\mathcal{D}$ is a C^2 domain.*

18.1 Convergence of $\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$ when $N \to \infty$

We recall that $\tau^{t,x} := \inf\{s > t : (s, X_s^{t,x}) \notin \mathcal{D}\}$ (see Definition 17.4), and $u(t,x) = \mathbb{E}_{t,x}[g(\tau, X_\tau)]$. The aim of this section is to prove the following result

Theorem 18.1. *For $g \in \mathcal{H}_{1+\alpha}$, $\forall (t,x) \in \mathcal{D}$, one has*

$$\lim_{N\to\infty} \partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0} = \mathbb{E}_{t,x}[\mathbf{1}_{\tau<T}(\nabla u - \nabla g) \cdot \theta(\tau, X_\tau)].$$

In other words, the boundary sensitivity in discrete time asymptotically coincides with the one we get in continuous time (see Theorem 17.7). The proof consists of the combination of two new results stated below. For this, we need to introduce the Gaussian random walk $s_n = \sum_{i=1}^n G^i$, where $G^i, i = 1, \cdots, n$ are i.i.d. standard centred normal variables. We also introduce its hitting time $\tau^+ := \inf\{n \geq 0 : s_n > 0\}$ ($\tau^+ \in]0, +\infty[$). One knows that the overshoot s_{τ^+} is such that $\mathbb{E}_0[s_{\tau^+}] = \frac{1}{\sqrt{2}}$ (see Chang and Peres [20] for a proof) and that $c_0 := \frac{\mathbb{E}_0[s_{\tau^+}^2]}{2\mathbb{E}_0[s_{\tau^+}]} = -\frac{\zeta(1/2)}{\sqrt{2\pi}} = 0.5823...$ (see Siegmund [89] for a proof).

Theorem 18.2. *For any continuous function $f \in \mathcal{H}_\alpha$, for any $(t,x) \in \mathcal{D}$, one has*

$$\sum_{j=0,t_j>t}^{N-1} \sqrt{h}\mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N>t_j} \int_{\partial\mathcal{D}_{t_{j+1}}} p(t_j, X_{t_j}, t_{j+1}, y) f(t_j, y) d\sigma_y\right] \xrightarrow{N\to\infty} \frac{1}{\mathbb{E}_0[s_{\tau^+}]}\mathbb{E}_{t,x}\left[\mathbf{1}_{\tau<T}\frac{f(\tau, X_\tau)}{|\sigma^*\nabla F|(\tau, X_\tau)}\right].$$

Conjecture 1. *For $g \in \mathcal{H}_{1+\alpha}$, for $x \in \partial\mathcal{D}_t$ and $t < T$ one has*

$$\frac{1}{\sqrt{h}}(g - u_N)(t,x) \xrightarrow{N\to\infty} \mathbb{E}_0[s_{\tau^+}]\left((\nabla g - \nabla u) \cdot \nabla F|\sigma^*\nabla F|\right)(t,x).$$

The convergence is uniform in (t,x).

Proof of Theorem 18.1. In view of the above conjecture, we can replace in (18.1) $(g - u_N)(t_{j+1}, y)$ by $\sqrt{h}\mathbb{E}_0[s_{\tau^+}]\left((\nabla g - \nabla u)\cdot \nabla F|\sigma^*\nabla F|\right)(t_{j+1}, y)$. Then, we apply Theorem 18.2 with $f(t,x) = \mathbb{E}_0[s_{\tau^+}]\left((\nabla g - \nabla u)\cdot \nabla F|\sigma^*\nabla F|\right)\theta\cdot\vec{n})(t,x)$. The results readily follows. □

Recently, Gobet and Menozzi [43] have proved an analogous result to Conjecture 1.

Theorem 18.3 (Theorem 2.3, Gobet and Menozzi [43]). *For $g \in \mathcal{H}_{1+\alpha}$, for $(t,x) \in \mathcal{D}$, one has*

$$\frac{1}{\sqrt{h}}(u - u_N)(t,x) \xrightarrow{N\to\infty} c_0 \mathbb{E}_{t,x}\left(\mathbf{1}_{\tau<T}(\nabla g - \nabla u)\cdot \nabla F(\tau, X_\tau)|\sigma^*\nabla F|(\tau, X_\tau)\right).$$

By comparing Theorem 18.3 with Conjecture 1, we notice that only the factor c_0 differs (since for $x \in \partial\mathcal{D}_t$, one has $(\tau, X_\tau) = (t,x)$ and $u = g$). Then, the limit is discontinuous w.r.t. x at the boundary. As explained in Gobet and Menozzi [43], the crucial tool is the mean of the overshoot in a one dimensional Brownian case. In that setting, $(W_{t_i})_i \overset{law}{=} (\sqrt{h}s_i)_i$, and we are interested in two quantities: $\mathbb{E}_0[s_{\tau^+}]$ when the initial point x is on the boundary, and $\lim_{y\to\infty}\mathbb{E}_0[s_{\tau_y^+}] - y := c_0$ (where $\tau_y^+ := \inf\{n \geq 0 : s_n > y\}$) when the initial point is not at the boundary. This intuitively explains why Conjecture 1 differs from the results stated in Gobet and Menozzi [43]. This intuition should also help us to establish the conjecture in the general case, even if we haven't got enough time to provide a complete proof.

Let us now prove Theorem 18.2. For this, we recall the definition of diffusion overshoots and known asymptotic results.

Definition 18.4 (Overshoot). The overshoot is related to the distance of a process to the boundary, when it exits the domain. More precisely, the overshoot of a process X is given by $F^-(\tau_N, X_{\tau_N}^{t,x})$, where τ_N is defined in Definition 17.12 and F is the signed spatial distance to $\partial\mathcal{D}$.

Since F is Lipschitz continuous in time and space, we easily get that $F^-(\tau_N, X_{\tau_N}^{t,x})$ is of order $\sqrt{h}$ in L_p norms, then it is interesting to study the asymptotics of the rescaled overshoot

$$Y_N = h^{-1/2}F^-(\tau_N, X_{\tau_N}^{t,x}).$$

We recall some properties satisfied by Y_N

Proposition 18.5 (Tightness of the overshoot, Gobet and Menozzi [43]). *For some $c > 0$, one has*

$$\sup_{h>0, s\in[t,T]} \mathbb{E}_{t,x}[\exp(c[h^{-1/2}F^-(s\wedge\tau_N, X_{s\wedge\tau_N})]^2)] < \infty.$$

Theorem 18.6 (Joint laws associated to the overshoot, Gobet and Menozzi [43]). *Let ϕ be a continuous function with compact support. For all $(t,x) \in \mathcal{D}\cup\mathcal{D}_0$, $s \in [t,T], y \geq 0$,*

$$\mathbb{E}_{t,x}[\mathbf{1}_{\tau_N\leq s}\phi(X_{\tau_N})\mathbf{1}_{F^-(\tau_N, X_{\tau_N})\geq y\sqrt{h}}] \xrightarrow{h\to 0} \mathbb{E}_{t,x}[\mathbf{1}_{\tau\leq s}\phi(X_\tau)(1 - H(y/|\sigma^*\nabla F(\tau, X_\tau)|))],$$

with $H(y) := (\mathbb{E}_0[s_{\tau^+}])^{-1}\int_0^y \mathbb{P}_0[s_{\tau^+} > z]dz$, where s_{τ^+} is defined at the beginning of the section.

In other words, $(\tau_N, X_{\tau_N}, h^{-1/2}F^-(\tau_N, X_{\tau_N}))$ weakly converges to $(\tau, X_\tau, |\sigma^*\nabla F(\tau, X_\tau)|Y)$ where Y is a random variable independent of (τ, X_τ), and whose cumulative function is equal to H. Y has the asymptotic law of the renormalised Brownian overshoot, and $\mathbb{E}(Y) = \frac{\mathbb{E}_0[s^2_{\tau^+}]}{2\mathbb{E}_0[s_{\tau^+}]} := c_0$. From Siegmund [89], we know that $c_0 = -\frac{\zeta(1/2)}{\sqrt{2\pi}} = 0.5823...$ and from Chang and Peres [20], we get $\mathbb{E}_0[s_{\tau^+}] = \frac{1}{\sqrt{2}}$.

Proof of Theorem 18.2. For any $0 \le \mu \le \overline{\mu}$ such that for any N, one has $\overline{\mu}\sqrt{\frac{T}{N}} \le \frac{r_0}{2}$, set

$$\psi_N(\mu) := \mathbb{E}_{t,x}\left(f(\tau_N, X_{\tau_N})\mathbf{1}_{F^-(\tau_N, X_{\tau_N})\in]0,\mu\sqrt{h}[}\mathbf{1}_{\tau_N<T}\right),$$

where the assumptions on f are given in Theorem 18.2. By the above weak convergence, one has

$$\begin{aligned}
\lim_{N\to\infty}\psi_N(\mu) &= \mathbb{E}_{t,x}\left(f(\tau, X_\tau)\mathbf{1}_{|\sigma^*\nabla F(\tau,X_\tau)|Y\in]0,\mu[}\mathbf{1}_{\tau<T}\right)\\
&= \mathbb{E}_{t,x}\left(f(\tau, X_\tau)\mathbf{1}_{\tau<T}\int_0^{\frac{\mu}{|\sigma^*\nabla F(\tau,X_\tau)|}}\frac{\mathbb{P}_0(s_{\tau^+} > z)}{\mathbb{E}_0[s_{\tau^+}]}dz\right)\\
&= \int_0^\mu \mathbb{E}_{t,x}\left(\frac{f(\tau, X_\tau)}{|\sigma^*\nabla F(\tau, X_\tau)|}\mathbf{1}_{\tau<T}\frac{\mathbb{P}_0(s_{\tau^+} > \frac{z}{|\sigma^*\nabla F(\tau,X_\tau)|})}{\mathbb{E}_0[s_{\tau^+}]}\right)dz. \qquad (18.3)
\end{aligned}$$

Let us define $\mathcal{D}^\mu_{t_{j+1}} := \{y \in \mathbb{R}^d : d(y, \mathcal{D}_{t_{j+1}}) < \mu\sqrt{h}\}$. From the definition of $\psi_N(\mu)$, one can easily write

$$\psi_N(\mu) = \sum_{j=0,t_j>t}^{N-1}\mathbb{E}_{t,x}\left(\mathbf{1}_{\tau_N=t_{j+1}}f(t_{j+1}, X_{t_{j+1}})\mathbf{1}_{F^-(t_{j+1},X_{t_{j+1}})\in]0,\mu\sqrt{h}[}\right).$$

Then, we write $\mathbf{1}_{\tau_N=t_{j+1}} = \mathbf{1}_{\tau_N>t_j} - \mathbf{1}_{\tau_N\le t_{j+2}}$. Since $F^-(t_{j+1}, X_{t_{j+1}})$ is null on $\mathbf{1}_{\tau_N\le t_{j+2}}$, we get

$$\psi_N(\mu) = \sum_{j=0,t_j>t}^{N-1}\mathbb{E}_{t,x}\left(\mathbf{1}_{\tau_N>t_j}\int_{\mathcal{D}^\mu_{t_{j+1}}\backslash\mathcal{D}_{t_{j+1}}} p(t_j, X_{t_j}; t_{j+1}, y)f(t_{j+1}, y)dy\right),$$

and

$$\psi'_N(\mu) = \sqrt{h}\sum_{j=0,t_j>t}^{N-1}\mathbb{E}_{t,x}\left(\mathbf{1}_{\tau_N>t_j}\int_{\partial\mathcal{D}^\mu_{t_{j+1}}} p(t_j, X_{t_j}; t_{j+1}, y)f(t_{j+1}, y)d\sigma_y\right).$$

Hence, it appears that Theorem 18.2 only consists of proving that

$$\psi'_N(\mu) \xrightarrow{N\to\infty} \mathbb{E}_{t,x}\left(\frac{f(\tau, X_\tau)}{|\sigma^*\nabla F(\tau, X_\tau)|}\mathbf{1}_{\tau<T}\frac{\mathbb{P}_0(s_{\tau^+} > \frac{\mu}{|\sigma^*\nabla F(\tau,X_\tau)|})}{\mathbb{E}_0[s_{\tau^+}]}\right),$$

for $\mu = 0$. Let us prove it for any μ. Our methodology is the following

- For all μ_1, μ_2 in $[0, \overline{\mu}]$,

$$\sup_N |\psi'_N(\mu_1) - \psi'_N(\mu_2)| \le C(\overline{\mu})|\mu_1 - \mu_2|^\alpha, \qquad (18.4)$$

- the sequence ψ'_N is uniformly bounded.

The first statement yields the sequence ψ'_N is equicontinuous. By combining it with the second statement, we apply the Arzela-Ascoli theorem: there exists a subsequence which converges uniformly. Then, the identification of the limit is achieved using the limit (18.3) of $\psi_N(\mu)$.

Step 1: ψ'_N is uniformly bounded. It is more convenient to rewrite $\psi'_N(\mu)$ using the local parametrisation. For this, let us start from $\psi_N(\mu)$.

$$\psi_N(\mu) = \mathbb{E}_{t,x}\left[\sum_{i=1,t_i\geq t}^{N} \mathbf{1}_{t_{i-1}<\tau_N} \int f(t_i,y)\mathbf{1}_{h^{-1/2}F^-(t_i,y)\in]0,\mu[}p(t_{i-1},X_{t_{i-1}};t_i,y)dy\right].$$

The integration set is in $V_{r_0}(\partial\mathcal{D}_{t_i})$. Since $\partial\mathcal{D}_{t_i}$ is compact, there exists a finite number of points $(s_{i,j})_{1\leq j\leq k}$ in $\partial\mathcal{D}_{t_i}$ such that $V_{r_0}(\partial\mathcal{D}_{t_i}) \in \bigcup_{1\leq j\leq k} U^{i,j}$. We associate to them $G^{i,j}, F^{i,j}, U^{i,j}$, $V^{i,j}$ and $g_{s_{i,j}}$ defined in Proposition 17.3. As we have done at the beginning of the proof of Lemma 17.18, we construct a partition of unity corresponding to $(U^{i,j})_{1\leq j\leq k}$, i.e. non negative C_b^∞ functions $(\chi_j)_{1\leq j\leq k}$ verifying $\mathrm{Supp}(\chi_j)\subset U^{i,j}$ and $\sum_{j=1}^k \chi_j \equiv 1$ on $\overline{\mathcal{D}_{t_i}\cup V_{r_0}(\partial\mathcal{D}_{t_i})}$. We get

$$\psi_N(\mu) = \sum_{j=1}^{k}\mathbb{E}_{t,x}\left[\sum_{i=1,t_i\geq t}^{N} \mathbf{1}_{t_{i-1}<\tau_N} \int_{U^{i,j}} \chi_j(y) f(t_i,y)\mathbf{1}_{h^{-1/2}F^-(t_i,y)\in]0,\mu[}p(t_{i-1},X_{t_{i-1}};t_i,y)dy\right],$$

and according to Proposition 17.3, for y in $U^{i,j}$ we can write $y = g_{s_{i,j}}(\overline{z}) + \lambda\overrightarrow{n}(g_{s_{i,j}}(\overline{z}))$, where $\lambda\in[-r_0,r_0]$ and $\overline{z}\in V^{i,j}$. We change variables in the above equality and we obtain

$$\psi_N(\mu) = \sum_{j=1}^{k}\sum_{i=1,t_i\geq t}^{N}\mathbb{E}_{t,x}[\mathbf{1}_{t_{i-1}<\tau_N}\int_0^{\mu h^{1/2}}d\lambda\int_{V^j}d\overline{z}$$
$$p(t_{i-1},X_{t_{i-1}};t_i,g_{s_j}(\overline{z})+\lambda\overrightarrow{n}(g_{s_j}(\overline{z})))(|\det(\mathrm{Jac}(G^{i,j}))|\chi_j f)(t_i,g_{s_{i,j}}(\overline{z})+\lambda\overrightarrow{n}(g_{s_{i,j}}(\overline{z})))],$$

A last change of variable $\nu = h^{-1/2}\lambda$ yields

$$\psi_N(\mu) = h^{1/2}\sum_{j=1}^{k}\sum_{i=1,t_i>t}^{N}\mathbb{E}_{t,x}[\mathbf{1}_{t_{i-1}<\tau_N}\int_0^{\mu}d\nu\int_{V^{i,j}}d\overline{z}$$
$$p(t_{i-1},X_{t_{i-1}};t_i,g_{s_{i,j}}(\overline{z})+h^{1/2}\nu\overrightarrow{n}(g_{s_{i,j}}(\overline{z})))(|\det(\mathrm{Jac}(G^{i,j}))|\chi_j f)(t_i,g_{s_{i,j}}(\overline{z})+h^{1/2}\nu\overrightarrow{n}(g_{s_{i,j}}(\overline{z})))].$$

Let $z_\nu^{i,j}$ denote $g_{s_{i,j}}(\overline{z})+h^{1/2}\nu\overrightarrow{n}(g_{s_{i,j}}(\overline{z}))$. This enables to deduce

$$\psi'_N(\mu) = \tag{18.5}$$
$$h^{1/2}\sum_{j=1}^{k}\sum_{i=1,t_i>t}^{N}\mathbb{E}_{t,x}[\mathbf{1}_{t_{i-1}<\tau_N}\int_{V^{i,j}}d\overline{z}p(t_{i-1},X_{t_{i-1}};t_i,z_\mu^{i,j})(|\det(\mathrm{Jac}(G^{i,j}))|\chi_j f)(t_i,z_\mu^{i,j})].$$

We use the Aronson inequality (6.7) and the definition of $z_\mu^{i,j}$ to bound $p(t_{i-1},X_{t_{i-1}};t_i,z_\mu^{i,j})$, we obtain

$$p(t_{i-1},X_{t_{i-1}};t_i,z_\mu^{i,j}) \leq \frac{C}{(2\pi h)^{d/2}}\exp\left(-c\frac{|g_{s_{i,j}}(\overline{z})-X_{t_{i-1}}|^2}{h}\right)e^{c\mu}. \tag{18.6}$$

If $|g_{s_{i,j}}(\overline{z}) - X_{t_{i-1}}|$ is larger than $\frac{r_0}{4}$ for instance, the contribution of this term in (18.5) is exponentially small w.r.t. h and it gives a negligible contribution in (18.5). Otherwise, we can assume that $g_{s_{i,j}}(\overline{z})$ and $X_{t_{i-1}}$ are both in $U^{i,j}$. If we write $F^{i,j}(X_{t_{i-1}}) = (F(t_i, X_{t_{i-1}}), \overline{z}_{i-1})$ and $(0, \overline{z}) = F^{i,j}(g_{s_{i,j}}(\overline{z}))$, we get, since $F^{i,j}$ is Lipschitz continuous, $|\overline{z} - \overline{z}_{i-1}|^2 + F^2(t_i, X_{t_{i-1}}) \leq C|g_{s_{i,j}}(\overline{z}) - X_{t_{i-1}}|^2$. Representing this lower bound in (18.5) and integrating w.r.t. $\overline{z}$ gives

$$|\psi_N'(\mu)| \leq \sum_{i=1,t_i>t}^{N} \mathbb{E}_{t,x}\left[\mathbf{1}_{t_{i-1}<\tau_N} \exp\left(-c\frac{F^2(t_i, X_{t_{i-1}})}{h}\right)\right].$$

To conclude to the boundedness of $\psi_N'(\mu)$, we use the following Lemma 18.7 and Proposition 18.8 (recall that $F^2(t_i, x) = d^2(x, \partial\mathcal{D}_{t_i})$ for x close to $\partial\mathcal{D}_{t_i}$).

Lemma 18.7. *For all $s \in [t_j, t_{j+1}]$, it holds*

$$\exp\left(-c\frac{d^2(X_{t_j}, \partial\mathcal{D}_s)}{h}\right) \leq C\exp\left(-c\frac{d^2(X_{t_j}, \partial\mathcal{D}_{t_j})}{h}\right).$$

Proof of Lemma 18.7. The above inequality is obtained by using the regularity of F, i.e. the signed spatial distance introduced in Definition 17.1.

$$d(X_{t_j}, \partial\mathcal{D}_s)^2 = F^2(s, X_{t_j}),$$

Moreover, for X_{t_j} close to $\partial\mathcal{D}_s$,

$$|F(s, X_{t_j}) - F(t_j, X_{t_j})| \leq Ch$$

and the result follows. □

Proposition 18.8. *There is a positive constant c such that*

$$\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds \leq Ch\sum_{j=0,t_j\geq t}^{N-1} \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N>t_j}\exp\left(-c\frac{d(X_{t_j}, \partial\mathcal{D}_{t_j})^2}{h}\right)\right] = O(h).$$

The proof of Proposition 18.8 is postponed to Section 18.4.

Step 2: We prove $\sup_N|\psi_N'(\mu_1) - \psi_N'(\mu_2)| \leq C(\overline{\mu})|\mu_1 - \mu_2|^\alpha$, for any $\mu_1, \mu_2 \in [0, \overline{\mu}]$. We can bound the difference by two terms I_1 and I_2, defined by

$$\begin{aligned} I_1 := h^{1/2}\sum_{j=1}^{k}\sum_{i=1,t_i\geq t}^{N} \mathbb{E}_{t,x}\Big[\mathbf{1}_{t_{i-1}<\tau_N}\int_{V^{i,j}} d\overline{z}\, p(t_{i-1}, X_{t_{i-1}}; t_i, z_{\mu_1}^{i,j}) \\ \times \left|(|\det(\operatorname{Jac}(G^{i,j}))|\chi_j f)(t_i, z_{\mu_1}^{i,j}) - (|\det(\operatorname{Jac}(G^{i,j}))|\chi_j f)(t_i, z_{\mu_2}^{i,j})\right|\Big], \end{aligned}$$

$$\begin{aligned} I_2 := h^{1/2}\sum_{j=1}^{k}\sum_{i=1,t_i\geq t}^{N} \mathbb{E}_{t,x}[\mathbf{1}_{t_{i-1}<\tau_N}\int_{V^{i,j}} d\overline{z}(|\det(\operatorname{Jac}(G^{i,j}))|\chi_j f)(t_i, z_{\mu_2}^{i,j}) \\ \times \left|p(t_{i-1}, X_{t_{i-1}}; t_i, z_{\mu_1}^{i,j}) - p(t_{i-1}, X_{t_{i-1}}; t_i, z_{\mu_2}^{i,j})\right|]. \end{aligned}$$

Bound for I_1. Since $\mu \longmapsto (|\det(\mathrm{Jac}(G^{i,j}))|\chi_j f)(t_i, z_\mu^{i,j})$ is an $\mathcal{H}_\alpha$ function, we get

$$I_1 \leq Ch|\mu_1-\mu_2|^\alpha \sum_{j=1}^k \sum_{i=1,t_i\geq t}^N \mathbb{E}_{t,x}\left[\mathbf{1}_{t_{i-1}<\tau_N}\int_{V^{i,j}} d\overline{z} p(t_{i-1},X_{t_{i-1}};t_i,z_{\mu_1}^{i,j})\right].$$

We end the computation as for the boundedness of $\psi_N'(\mu)$ and it gives $I_1 \leq C\sqrt{h}|\mu_1-\mu_2|^\alpha$.

Bound for I_2. We use a Taylor formula to develop $p(t_{i-1},X_{t_{i-1}};t_i,z_{\mu_1}^{i,j}) - p(t_{i-1},X_{t_{i-1}};t_i,z_{\mu_2}^{i,j})$

$$p(t_{i-1},X_{t_{i-1}};t_i,z_{\mu_1}^{i,j})-p(t_{i-1},X_{t_{i-1}};t_i,z_{\mu_2}^{i,j}) = \sqrt{h}(\mu_1-\mu_2)\vec{n}(g_{s_{i,j}}(\overline{z}))\int_0^1 \partial_y p(t_{i-1},X_{t_{i-1}};t_i,\lambda z_{\mu_1}^{i,j}+(1-\lambda)z_{\mu_2}^{i,j})d\lambda.$$

Then, we get

$$I_2 \leq Ch|\mu_1-\mu_2|\int_0^1 d\lambda \sum_{j=1}^k \sum_{i=1,t_i\geq t}^N \mathbb{E}_{t,x}\left[\mathbf{1}_{t_{i-1}<\tau_N}\int_{V^{i,j}} d\overline{z}|\partial_y p|(t_{i-1},X_{t_{i-1}};t_i,\lambda z_{\mu_1}^{i,j}+(1-\lambda)z_{\mu_2}^{i,j})\right],$$

where C depends on the suprema of the function $\mu \longmapsto (|\det(\mathrm{Jac}(G^{i,j}))|\chi_j f)(t_i, z_\mu^{i,j})$. We use (6.9), page 66, to bound $\partial_y p$. We obtain

$$I_2 \leq \frac{C}{h^{\frac{d-1}{2}}}|\mu_1-\mu_2| \int_0^1 d\lambda \sum_{j=1}^k \sum_{i=1,t_i\geq t}^N \mathbb{E}_{t,x}\left[\mathbf{1}_{t_{i-1}<\tau_N}\int_{V^{i,j}} d\overline{z}\exp\left(-c\frac{|\lambda z_{\mu_1}^{i,j}+(1-\lambda)z_{\mu_2}^{i,j}-X_{t_{i-1}}|^2}{h}\right)\right].$$

Since $\lambda z_{\mu_1}^{i,j}+(1-\lambda)z_{\mu_2}^{i,j} = g_{s_{i,j}}(\overline{z})+\sqrt{h}(\lambda\mu_1+(1-\lambda)\mu_2)\vec{n}(g_{s_{i,j}}(\overline{z}))$ and μ_1 and μ_2 are in $[0,\overline{\mu}]$, we bound $\exp\left(-c\frac{|\lambda z_{\mu_1}^{i,j}+(1-\lambda)z_{\mu_2}^{i,j}-X_{t_{i-1}}|^2}{h}\right)$ by $C\exp\left(-c\frac{|g_{s_{i,j}}(\overline{z})-X_{t_{i-1}}|^2}{h}\right)$. We end the proof in the same way as for I_1 and get $I_2 \leq C|\mu_1-\mu_2|$, which ends the proof. □

18.2 Boundedness of $\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$ when $g \in \mathcal{H}_{1+\alpha}$

Theorem 18.9. *Assume g is a $\mathcal{H}_{1+\alpha}$ function. Then, it holds*

$$|\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0} \leq C,$$

where C doesn't depend on N, t and x.

The proof of Theorem 18.9 is based on two propositions: Proposition 18.10 and Proposition 18.8, stated page 248 and 246.

Proof of Theorem 18.9. Using the definition of u_N yields

$$(g-u_N)(t_{j+1},y) = \mathbb{E}_{t_{j+1},y}[g(t_{j+1},y)-g(\tau_N,X_{\tau_N})].$$

Since $y \in \partial\mathcal{D}_{t_{j+1}}$, by using Proposition 18.10 page 248, we prove that $(g-u_N)(t_{j+1},y)$ is bounded by $C\sqrt{h}$. Using (18.1) and the fact that θ is bounded lead to

$$|\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0} \leq C\sqrt{h}\sum_{j=0,t_j\geq t}^{N-1}\mathbb{E}_{t,x}\left(\mathbf{1}_{\tau_N^{t,x}>t_j}\int_{\partial\mathcal{D}_{t_{j+1}}} p(t_j,X_{t_j};t_{j+1},y)d\sigma_y\right). \tag{18.7}$$

Moreover, it holds $p(t_j,X_{t_j};t_{j+1},y) \leq \frac{C}{h^{d/2}}\exp\left(-c\frac{|y-X_{t_j}|^2}{h}\right)$. Inequality (18.7) becomes

$$|\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0} \leq \frac{C}{h^{(d-1)/2}}\sum_{j=0,t_j\geq t}^{N-1}\mathbb{E}_{t,x}\left(\mathbf{1}_{\tau_N^{t,x}>t_j}\int_{\partial\mathcal{D}_{t_{j+1}}}\exp\left(-c\frac{|y-X_{t_j}|^2}{h}\right)d\sigma_y\right). \tag{18.8}$$

In the special case $\mathcal{D}_{t_{j+1}} = \{y\in\mathbb{R}^d : y_1>0\}$, we have

$$\int_{\partial\mathcal{D}_{t_{j+1}}}\exp\left(-c\frac{|y-X_{t_j}|^2}{h}\right)d\sigma_y = \exp\left(-c\frac{|d(X_{t_j},\partial\mathcal{D}_{t_{j+1}})|^2}{h}\right)\prod_{l=2}^{d}\int_{\mathbb{R}}\exp\left(-c\frac{|y_l-X_{t_j}^l|^2}{h}\right)dy_l.$$

For all $l\in\{2,\cdots,d\}$, $\int_{\mathbb{R}}\exp\left(-c\frac{|y_l-X_{t_j}^l|^2}{h}\right)dy_l \leq C\sqrt{h}$. One gets

$$\int_{\partial\mathcal{D}_{t_{j+1}}}\exp\left(-c\frac{|y-X_{t_j}|^2}{h}\right)d\sigma_y \leq Ch^{(d-1)/2}\exp\left(-c\frac{d(X_{t_j},\partial\mathcal{D}_{t_{j+1}})^2}{h}\right).$$

For the general case, we use Proposition 17.3 to map locally $\mathcal{D}_{t_{j+1}}$ as a half space: in these maps, $d(y,\partial\mathcal{D}_{t_{j+1}})$ has a simple expression which enables to boil down to the first case. For more details, we refer to Gobet [38], Lemma 3.4.5.. Plugging this result in (18.8) leads to

$$|\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0} \leq C\sum_{j=0,t_j\geq t}^{N-1}\mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N^{t,x}>t_j}\exp\left(-c\frac{d(X_{t_j},\partial\mathcal{D}_{t_{j+1}})^2}{h}\right)\right].$$

It remains to show that $\sum_{j=0,t_j\geq t}^{N-1}\mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N^{t,x}>t_j}\exp\left(-c\frac{d(X_{t_j},\partial\mathcal{D}_{t_{j+1}})^2}{h}\right)\right] = O(1)$. To do so, we use Lemma 18.7 and Proposition 18.8. □

Proposition 18.10. *For any $\mathcal{H}_{1+\alpha}$ function g, it holds*

$$\forall\, x\in\partial\mathcal{D}_t,\quad \mathbb{E}_{t,x}[g(\tau_N,X_{\tau_N})-g(t,x)] \leq C\sqrt{h},$$

where $h=\frac{T}{N}$ and $\tau_N^{t,x} = \inf\{t_i>t : (t_i,X_{t_i}^{t,x})\notin\mathcal{D}\}$.

This result has been recently proved in Gobet and Menozzi [43].

18.3 Boundedness of $\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0}$ when $g(x)=(K-e^x)_+$

This case is more involved technically because g is only piecewise C^1.

Theorem 18.11. *Assume* $g(t,x) = (K - e^x)^+$ *and* $\log(K) \in \partial\mathcal{D}_T$*. Then, it holds*

$$|\partial_\epsilon u_N^\epsilon(t,x)\big|_{\epsilon=0} \le C,$$

where C *doesn't depend on* N*.*

The additional assumption on $\partial\mathcal{D}_T$ is not a restriction because when the dividend rate δ is smaller than the interest rate r, one knows that it is satisfied by the optimal domain.

Proposition 18.12. *Let* $f(t,x) = (K - e^x)^+$*. Hence,*

$$\forall\, x \in \partial\mathcal{D}_t, \quad \mathbb{E}_{t,x}[f(\tau_N, X_{\tau_N}) - f(t,x)] \le C\sqrt{h},$$

where $h = \frac{T}{N}$ *and* $\tau_N^{t,x} = \inf\{t_i > t : (t_i, X_{t_i}^{t,x}) \notin \mathcal{D}\}$*.*

Proof of Theorem 18.11. The proof is similar to the proof of Theorem 18.9: by using Proposition 18.12, we show that $|(g - u_N)(t_{j+1}, y)| \le C\sqrt{h}$. Then, we use the proof of Theorem 18.9 to conclude. □

Proof of Proposition 18.12. One knows from Gobet and Menozzi [42] that Proposition 18.12 is a consequence of the two following lemmas. □

Lemma 18.13. *There are some positive constants* C *and* N_0 *such that for* $N \ge N_0$*, for any* $i \in \{0, \cdots, N-1\}$*, one has for* $X_{t_i} \in \mathcal{D}_{t_i}$

$$\mathbb{P}\left(\exists\, t \in [t_i, t_{i+1}] : X_t \notin \mathcal{D}_t \big| \mathcal{F}_{t_i}\right) \le C\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{t_i}).$$

The proof of this lemma is postponed to Section 18.4.

Lemma 18.14. *Assume* $g(t,x) = (K - e^x)^+$*. Define* $\tau^s = \inf\{u > s : (u, X_u) \notin \mathcal{D}\}$*. For all* $s \in [t,T]$*, we denote* $V_s := \mathbb{E}[g(\tau^s, X_{\tau^s})\big|\mathcal{F}_s]$*. For all* $i \in \{0, \cdots, N-1\}$*, on the set* $\{\tau_N > t_i, \tau^{t_i} < t_{i+1}\}$*, one has*

$$|\mathbb{E}[V_{t_{i+1}} - V_{\tau^{t_i}} \big| \mathcal{F}_{\tau^{t_i}}]| \le C\sqrt{h}.$$

Proof of Lemma 18.14. The scheme of the proof is the following

1. We prove $\mathbb{E}[V_{t_{i+1}} - V_{\tau^{t_i}} \big| \mathcal{F}_{\tau^{t_i}}] \le C\sqrt{h} + \frac{K}{2}\mathbb{E}_{\tau^{t_i}}\left[L^{\log(K)}_{\tau^{t_{i+1}}}(X)\right]$,
2. We show
$$\mathbb{E}_{\tau^{t_i}}\left[L^{\log(K)}_{\tau^{t_{i+1}}}(X)\right] \le C\sqrt{h} + \lim_{\epsilon \to 0}\frac{1}{\epsilon}\int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{s \le \tau^{t_{i+1}}}\mathbf{1}_{X_s \in [\log(K), \log(K)+\epsilon[}]ds.$$
3. We split $A_\epsilon := \int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{s \le \tau^{t_{i+1}}}\mathbf{1}_{X_s \in [\log(K), \log(K)+\epsilon[}]ds$ in two terms, namely A_ϵ^1 and A_ϵ^2, and prove that $\lim_{\epsilon\to 0}\frac{1}{\epsilon}A_\epsilon^1 \le C\sqrt{h}$.
4. We prove $\lim_{\epsilon\to 0}\frac{1}{\epsilon}A_\epsilon^2 \le C\sqrt{h} + \int_{t_{i+2}}^T ds \int_0^1 d\lambda \mathbb{E}_{\tau^{t_i}}\left[\frac{F(t_{i+1}, X_{t_{i+1}})^+}{s - t_{i+1}} \exp\left(-c\frac{|F(s,\log(K)) - \lambda F(t_{i+1}, X_{t_{i+1}})|^2}{s - t_{i+1}}\right)\right]$.
5. We deduce from the preceding point that $\lim_{\epsilon\to 0}\frac{1}{\epsilon}A_\epsilon^2 \le C\sqrt{h}$.

Step 1 By using the definition of V, we get $\mathbb{E}[V_{t_{i+1}} - V_{\tau^{t_i}} \big| \mathcal{F}_{\tau^{t_i}}] = \mathbb{E}[(K - \exp(X_{\tau^{t_{i+1}}}))^+ - (K - \exp(X_{\tau^{t_i}}))^+ \big| \mathcal{F}_{\tau^{t_i}}]$. Itô-Tanaka's formula applied to $x \longmapsto (K-x)_+$ and $(\mathrm{e}^{X_s})_s$ leads to

$$
\begin{aligned}
(K - \exp(X_{\tau^{t_{i+1}}}))^+ - (K - \exp(X_{\tau^{t_i}}))^+ = &- \int_{\tau^{t_i}}^{\tau^{t_{i+1}}} \mathbf{1}_{\exp(X_s) \le K} d(\mathrm{e}^{X_s}) \\
&+ \frac{K}{2} \int_{\tau^{t_i}}^{\tau^{t_{i+1}}} dL_s^{\log(K)}(X).
\end{aligned}
$$

Taking the conditional expectation gives

$$
\begin{aligned}
\mathbb{E}_{\tau^{t_i}}[(K - \exp(X_{\tau^{t_{i+1}}}))^+] &= (K - \exp(X_{\tau^{t_i}}))^+ \\
&- \mathbb{E}_{\tau^{t_i}} \left[\int_{\tau^{t_i}}^{\tau^{t_{i+1}}} e^{X_s} \mathbf{1}_{K \ge e^{X_s}} (b(s,X_s) + \frac{1}{2}\sigma^2(s,X_s)) ds \right] + \frac{K}{2} \mathbb{E}_{\tau^{t_i}} \left[L_{\tau^{t_{i+1}}}^{\log(K)}(X) \right].
\end{aligned}
$$

Since b and σ^2 are bounded, we get

$$
\begin{aligned}
\left| \mathbb{E}_{\tau^{t_i}}[(K - \exp(X_{\tau^{t_{i+1}}}))^+ - (K - \exp(X_{\tau^{t_i}}))^+] \right| \le & K(|b|_\infty + \frac{1}{2}|\sigma|_\infty^2) \mathbb{E}_{\tau^{t_i}}[\tau^{t_{i+1}} - \tau^{t_i}] \\
&+ \frac{K}{2} \mathbb{E}_{\tau^{t_i}}[L_{\tau^{t_{i+1}}}^{\log(K)}(X)].
\end{aligned}
$$

By using Gobet and Menozzi [42], Lemma 3.2 with $\tilde{g}(t,x) = t$, we get $\mathbb{E}_{\tau^{t_i}}[\tau^{t_{i+1}} - \tau^{t_i}] \le C\sqrt{h}$. Hence,

$$
\mathbb{E}[V_{t_{i+1}} - V_{\tau^{t_i}} \big| \mathcal{F}_{\tau^{t_i}}] \le C\sqrt{h} + \frac{K}{2} \mathbb{E}_{\tau^{t_i}} \left[L_{\tau^{t_{i+1}}}^{\log(K)}(X) \right].
$$

Step 2 is devoted to proving

$$
\mathbb{E}_{\tau^{t_i}} \left[L_{\tau^{t_{i+1}}}^{\log(K)}(X) \right] \le C\sqrt{h} + \liminf_{\epsilon \longrightarrow 0} \frac{1}{\epsilon} \int_{t_{i+1}}^{T} \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{s \le \tau^{t_{i+1}}} \mathbf{1}_{X_s \in [\log(K), \log(K)+\epsilon[}] ds.
$$

We recall

$$
L_{\tau^{t_{i+1}}}^{\log(K)}(X) = \lim_{\epsilon \longrightarrow 0} \frac{1}{\epsilon} \int_{\tau^{t_i}}^{\tau^{t_{i+1}}} \mathbf{1}_{X_s \in [\log(K), \log(K)+\epsilon[} \sigma^2(s, X_s) ds.
$$

First, by Fatou's Lemma, $\mathbb{E}_{\tau^{t_i}} \left(L_{\tau^{t_{i+1}}}^{\log(K)}(X) \right) \le \liminf_{\epsilon \longrightarrow 0} \mathbb{E}_{\tau^{t_i}}(Z_\epsilon)$, where $Z_\epsilon := \frac{1}{\epsilon} \int_{\tau^{t_i}}^{\tau^{t_{i+1}}} \mathbf{1}_{X_s \in [\log(K), \log(K)+\epsilon[} \sigma^2(s, X_s) ds$.

Second, let us prove that $\liminf_{\epsilon \longrightarrow 0} \mathbb{E}_{\tau^{t_i}}[Z_\epsilon] \le C\sqrt{h} + \liminf_{\epsilon \longrightarrow 0} \frac{1}{\epsilon} \int_{t_{i+1}}^{T} \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{s \le \tau^{t_{i+1}}} \mathbf{1}_{X_s \in [\log(K), \log(K)+\epsilon[}] ds$. Since $\tau^{t_i} < t_{i+1}$ we can write

$$
\begin{aligned}
\mathbb{E}_{\tau^{t_i}}[Z_\epsilon] &= \int_{\tau^{t_i}}^{T} \mathbb{E}_{\tau^{t_i}} \left[\mathbf{1}_{s \le \tau^{t_{i+1}}} \mathbf{1}_{X_s \in [\log(K), \log(K)+\epsilon[} \right] ds \qquad (18.9) \\
&= \int_{\tau^{t_i}}^{t_{i+1}} \mathbb{P}_{\tau^{t_i}}(X_s \in [\log(K), \log(K)+\epsilon[) ds + \int_{t_{i+1}}^{T} \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{s \le \tau^{t_{i+1}}} \mathbf{1}_{X_s \in [\log(K), \log(K)+\epsilon[}] ds.
\end{aligned}
$$

Let p denote the continuous transition density function of the process X. By the estimate (6.7) page 66, one has $p(\tau^{t_i}, X_{\tau^{t_i}}, s, y) \le \frac{c}{\sqrt{s-\tau^{t_i}}}$, which is enough to prove (by an application of the dominated convergence theorem)

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_{\tau^{t_i}}^{t_{i+1}} \mathbb{P}_{\tau^{t_i}}(X_s \in [\log(K), \log(K)+\epsilon[)ds = \int_{\tau^{t_i}}^{t_{i+1}} p(\tau^{t_i}, X_{\tau^{t_i}}, s, \log(K))ds$$
$$\le \int_{\tau^{t_i}}^{t_{i+1}} \frac{c}{\sqrt{s-\tau^{t_i}}}ds = O(\sqrt{h}).$$

Combining the previous result and (18.9) leads to

$$\mathbb{E}_{\tau^{t_i}}\left[L_{\tau^{t_{i+1}}}^{\log(K)}(X)\right] \le C\sqrt{h} + \liminf_{\epsilon\to 0}\frac{1}{\epsilon}\int_{t_{i+1}}^{T} \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds.$$

Step 3 Let us denote $A_\epsilon = \int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds$. If $X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}}$, $\tau^{t_{i+1}} = t_{i+1}$. Hence, we get $\int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{X_{t_{i+1}}\notin\mathcal{D}_{t_{i+1}}}\mathbf{1}_{s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds = 0$, and

$$A_\epsilon = \int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}}\mathbf{1}_{s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds.$$

We introduce $\tau_{r_0}^t := \inf\{s \ge t : F(s, X_s) \ge r_0\}$. We split A_ϵ in two terms depending on the position of s w.r.t. $\tau_{r_0}^{t_{i+1}}$.

$$A_\epsilon = \int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}}\mathbf{1}_{\tau_{r_0}^{t_{i+1}}\le s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds$$
$$+ \int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}}\mathbf{1}_{s\le\tau_{r_0}^{t_{i+1}}, s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds := A_\epsilon^1 + A_\epsilon^2.$$

Let us bound $\liminf_{\epsilon\to 0}\frac{1}{\epsilon}A_\epsilon^1$. To do so, we write

$$A_\epsilon^1 \le \int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}\left[\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}}\mathbf{1}_{\tau_{r_0}^{t_{i+1}}\le\tau^{t_{i+1}}\wedge s}\mathbb{E}_{\tau_{r_0}^{t_{i+1}}}[\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]\right]ds.$$

Then, using as before the upper bound for the transition density function, we easily get $\liminf_{\epsilon\to 0}\frac{1}{\epsilon}A_\epsilon^1 \le C\mathbb{E}_{\tau^{t_i}}\left[\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}}\mathbb{P}_{t_{i+1}}(\tau_{r_0}^{t_{i+1}} \le \tau^{t_{i+1}})\right]$. We apply Gobet and Menozzi [42], (4) page 14 to derive $\mathbb{P}_{t_{i+1}}(\tau_{r_0}^{t_{i+1}} \le \tau^{t_{i+1}}) \le CF^+(t_{i+1}, X_{t_{i+1}})$. Finally, an application of Gobet and Menozzi [42], Lemma 4.1 gives $\liminf_{\epsilon\to 0}\frac{1}{\epsilon}A_\epsilon^1 \le C\sqrt{h}$.

Step 4 We recall $A_\epsilon^2 = \int_{t_{i+1}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}}\mathbf{1}_{s\le\tau_{r_0}^{t_{i+1}}, s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds$. It remains to prove $\liminf_{\epsilon\to 0}\frac{1}{\epsilon}A_\epsilon^2 \le C\sqrt{h}$. First, we split A_ϵ^2 in two terms.

$$A_\epsilon^2 = \int_{t_{i+1}}^{t_{i+2}} \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}}\mathbf{1}_{s\le\tau_{r_0}^{t_{i+1}}, s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds$$
$$+ \int_{t_{i+2}}^T \mathbb{E}_{\tau^{t_i}}[\mathbf{1}_{X_{t_{i+1}}\in\mathcal{D}_{t_{i+1}}}\mathbf{1}_{s\le\tau_{r_0}^{t_{i+1}}, s\le\tau^{t_{i+1}}}\mathbf{1}_{X_s\in[\log(K),\log(K)+\epsilon[}]ds := A_\epsilon^{21} + A_\epsilon^{22}.$$

The first term A_ϵ^{21} can be easily bounded as before by $C\sqrt{h}$.

Let us bound A^{22}_ϵ. We introduce the process $(Y_s)_{s\in[t_{i+1},T]} := (F(s,X_s))_{s\in[t_{i+1},T]}$. Note that the restriction of the signed distance function $F(s,\cdot)$ to $\{x : 0 \le F(s,x) \le r_0\}$ is bijective. Then we can rewrite A^{22}_ϵ

$$A^{22}_\epsilon = \int_{t_{i+2}}^T \mathbb{E}_{\tau^{t_i}}\left[\mathbf{1}_{Y_{t_{i+1}}\ge 0}\mathbb{E}_{t_{i+1}}[\mathbf{1}_{s\le \tau^{t_{i+1}}_{Y,r_0}, s\le \tau^{t_{i+1}}_{Y,0}}\mathbf{1}_{Y_s\in[F(s,\log(K)),F(s,\log(K)+\epsilon)[}\right] ds,$$

where $\tau^{t_{i+1}}_{Y,a}$ denotes $\inf\{s\ge t_{i+1} : Y_s = a\}$. We introduce q, the transition density function of the process Y killed at 0 and r_0. Under our assumptions, q is well defined and satisfies (upper) Aronson estimates (see Gobet [37], (10) page 173) as well as its derivatives. Then, it follows

$$\liminf_{\epsilon\to 0}\frac{1}{\epsilon}A^{22}_\epsilon = \int_{t_{i+2}}^T \mathbb{E}_{\tau^{t_i}}\left[\mathbf{1}_{Y_{t_{i+1}}\ge 0}q\left(t_{i+1},Y_{t_{i+1}},s,F(s,\log(K))\right)\partial_y F(s,\log(K))\right] ds.$$

Then, we apply a Taylor formula to q between $Y_{t_{i+1}}$ and 0. Since $\forall (s,y)$, $q(t_{i+1},0,s,y) = 0$, we get $q\left(t_{i+1},Y_{t_{i+1}},s,F(s,\log(K))\right) = Y_{t_{i+1}}\int_0^1 \partial_x q\left(t_{i+1},\lambda Y_{t_{i+1}},s,F(s,\log(K))\right)d\lambda$. It gives

$$\liminf_{\epsilon\to 0}\frac{1}{\epsilon}A^{22}_\epsilon \le C\int_{t_{i+2}}^T ds\int_0^1 d\lambda \mathbb{E}_{\tau^{t_i}}\left[\frac{Y^+_{t_{i+1}}}{s-t_{i+1}}\exp\left(-c\frac{|F(s,\log(K))-\lambda Y_{t_{i+1}}|^2}{s-t_{i+1}}\right)\right].$$

Using the definition of Y leads to

$$\begin{aligned}\liminf_{\epsilon\to 0}\frac{1}{\epsilon}A^{2}_\epsilon &\le C\sqrt{h}\\ &+\int_{t_{i+2}}^T ds\int_0^1 d\lambda\mathbb{E}_{\tau^{t_i}}\left[\frac{F(t_{i+1},X_{t_{i+1}})^+}{s-t_{i+1}}\exp\left(-c\frac{|F(s,\log(K))-\lambda F(t_{i+1},X_{t_{i+1}})|^2}{s-t_{i+1}}\right)\right].\end{aligned}$$

Step 5 Let us denote $T(s,\lambda) := \mathbb{E}_{\tau^{t_i}}\left[\frac{F(t_{i+1},X_{t_{i+1}})^+}{s-t_{i+1}}\exp\left(-c\frac{|F(s,\log(K))-\lambda F(t_{i+1},X_{t_{i+1}})|^2}{s-t_{i+1}}\right)\right]$. We aim at proving that $\int_{t_{i+2}}^T ds\int_0^1 d\lambda T(s,\lambda) \le C\sqrt{h}$. We split $T(s,\lambda)$ in two terms, depending on the distance between $X_{t_{i+1}}$ and $\partial\mathcal{D}_{t_{i+1}}$.

$$\begin{aligned}T(s,\lambda) =& \mathbb{E}_{\tau^{t_i}}\left[\frac{F(t_{i+1},X_{t_{i+1}})^+}{s-t_{i+1}}\exp\left(-c\frac{|F(s,\log(K))-\lambda F(t_{i+1},X_{t_{i+1}})|^2}{s-t_{i+1}}\right)\mathbf{1}_{F(t_{i+1},X_{t_{i+1}})\ge r_0}\right]\\ &+\mathbb{E}_{\tau^{t_i}}\left[\frac{F(t_{i+1},X_{t_{i+1}})^+}{s-t_{i+1}}\exp\left(-c\frac{|F(s,\log(K))-\lambda F(t_{i+1},X_{t_{i+1}})|^2}{s-t_{i+1}}\right)\mathbf{1}_{F(t_{i+1},X_{t_{i+1}})\le r_0}\right].\end{aligned}$$

Since F is bounded and $s\in[t_{i+2},T]$, the first term of the above r.h.s. is bounded by $\frac{C}{h}\mathbb{P}_{\tau^{t_i}}(F(t_{i+1},X_{t_{i+1}})\ge r_0)$. Moreover, $F(t_{i+1},X_{t_{i+1}}) = F(\tau^{t_i},X_{t_{i+1}}) + (t_{i+1}-\tau^{t_i})\partial_t F(\bar{t},X_{t_{i+1}})$, where $\bar{t}\in[\tau^{t_i},t_{i+1}]$. Then, $F(t_{i+1},X_{t_{i+1}})\ge r_0$ leads to $F(\tau^{t_i},X_{t_{i+1}})\ge r_0 - h|\partial_t F|_\infty$, and for h small enough (such that $r_0-h|\partial_t F|_\infty > \frac{r_0}{2}$), one gets

$$\mathbb{P}_{\tau^{t_i}}(F(t_{i+1},X_{t_{i+1}})\ge r_0)\le \mathbb{P}_{\tau^{t_i}}(F(\tau^{t_i},X_{t_{i+1}})>\frac{r_0}{2})\le \mathbb{P}_{\tau^{t_i}}(|X_{\tau^{t_i}}-X_{t_{i+1}}|>\frac{r_0}{2}).$$

Since τ^{t_i} and t_{i+1} are close to each other, this probability is exponentially small (see Lemma A.14, page 265) and $\frac{C}{h}\mathbb{P}_{\tau^{t_i}}(F(t_{i+1},X_{t_{i+1}})\ge r_0) = O_{pol}(h)$. Thus, we get

$$\begin{aligned}T(s,\lambda) =& O_{pol}(h)\\ &+\mathbb{E}_{\tau^{t_i}}\left[\frac{F(t_{i+1},X_{t_{i+1}})^+}{s-t_{i+1}}\exp\left(-c\frac{|F(s,\log(K))-\lambda F(t_{i+1},X_{t_{i+1}})|^2}{s-t_{i+1}}\right)\mathbf{1}_{F(t_{i+1},X_{t_{i+1}})\le r_0}\right],\\ :=& O_{pol}(h)+T_1(s,\lambda).\end{aligned}$$

It remains to prove $\int_{t_{i+2}}^T\int_0^1 T_1(s,\lambda)d\lambda ds \le C\sqrt{h}$.

We rewrite $T_1(s,\lambda)$ by using the transition density of X

$$T_1(s,\lambda) = \int_{\mathbb{R}} \frac{F(t_{i+1},y)^+}{s-t_{i+1}} \exp\left(-c\frac{|F(s,\log(K))-\lambda F(t_{i+1},y)|^2}{s-t_{i+1}}\right)\mathbf{1}_{F(t_{i+1},y)\le r_0} p(\tau^{t_i},X_{\tau^{t_i}},t_{i+1},y)dy.$$

We use (6.7), page 66, to bound p: $p(\tau^{t_i},X_{\tau^{t_i}},t_{i+1},y) \le \frac{C}{\sqrt{t_{i+1}-\tau^{t_i}}}\exp\left(-c\frac{|y-X_{\tau^{t_i}}|^2}{t_{i+1}-\tau^{t_i}}\right)$. Then, we change variables by introducing $z=F(t_{i+1},y)$, which leads to

$$T_1(s,\lambda) \le C\int_{\mathbb{R}} \frac{z^+\mathbf{1}_{z\le r_0}}{s-t_{i+1}} \exp\left(-c\frac{|F(s,\log(K))-\lambda z|^2}{s-t_{i+1}}\right) \times \frac{1}{\sqrt{t_{i+1}-\tau^{t_i}}}\exp\left(-c\frac{|F^{-1}(t_{i+1},z)-X_{\tau^{t_i}}|^2}{t_{i+1}-\tau^{t_i}}\right)\left|\partial_x F^{-1}(t_{i+1},z)\right| dz.$$

Since $z\le r_0$, we bound $\partial_x F^{-1}(t_{i+1},z)$ and we write

$$F^{-1}(t_{i+1},z)-X_{\tau^{t_i}} = F^{-1}(t_{i+1},z)-F^{-1}(\tau^{t_i},0) = (t_{i+1}-\tau^{t_i})\partial_t F^{-1}(\bar{t},0)+z\partial_x F^{-1}(t_{i+1},\bar{z}),$$

where $\bar{t}\in[\tau^{t_i},t_{i+1}]$ and $\bar{z}\in[0,z]$. The time derivative is uniformly bounded and the space one does not vanish since $\bar{z}\in[0,r_0]$. By using the inequality $(a-b)^2\ge \frac{1}{2}a^2-2b^2$, we get

$$\exp\left(-c\frac{|F^{-1}(t_{i+1},z)-X_{\tau^{t_i}}|^2}{t_{i+1}-\tau^{t_i}}\right) \le C\exp\left(\frac{-cz^2}{t_{i+1}-\tau^{t_i}}\right).$$

We deduce

$$T_1(s,\lambda)\le C\int_{\mathbb{R}}\frac{z^+\mathbf{1}_{z\le r_0}}{s-t_{i+1}}\exp\left(-c\frac{|F(s,\log(K))-\lambda z|^2}{s-t_{i+1}}\right)\exp\left(-c'\frac{z^2}{t_{i+1}-\tau^{t_i}}\right)\frac{dz}{\sqrt{t_{i+1}-\tau^{t_i}}}.$$

We change variables in the above inequality, by setting $u=\frac{\sqrt{2c'}z}{\sqrt{t_{i+1}-\tau^{t_i}}}$.

$$T_1(s,\lambda)\le C\sqrt{t_{i+1}-\tau^{t_i}}\int_{\mathbb{R}}\frac{u^+}{s-t_{i+1}}\exp\left(-c\frac{|F(s,\log(K))-\lambda\frac{\sqrt{t_{i+1}-\tau^{t_i}}}{\sqrt{2c'}}u|^2}{s-t_{i+1}}\right)\exp(-\frac{u^2}{2})du.$$

Let us introduce x_0 and σ_0 defined by $x_0=\frac{\sqrt{2c'}F(s,\log(K))}{\lambda\sqrt{t_{i+1}-\tau^{t_i}}}$ and $\sigma_0^2=\frac{c'(s-t_{i+1})}{c\lambda^2(t_{i+1}-\tau^{t_i})}$. Then, $T_1(s,\lambda)\le CI(s,\lambda)$ where

$$I(s,\lambda) := \frac{\sqrt{t_{i+1}-\tau^{t_i}}}{s-t_{i+1}}\int_0^\infty u\exp(-\frac{|x_0-u|^2}{2\sigma_0^2})\exp(-\frac{u^2}{2})du.$$

Let us prove $\int_{t_{i+2}}^T\int_0^1 I(s,\lambda)d\lambda ds\le C\sqrt{h}$. In view of applying Lemma A.17, page 266, we rewrite I in the following way

$$I(s,\lambda)=2\pi\sqrt{\frac{c'}{c}}\frac{1}{\lambda\sqrt{s-t_{i+1}}}\int_0^\infty u\frac{\exp(-\frac{|x_0-u|^2}{2\sigma_0^2})}{\sqrt{2\pi\sigma_0^2}}\frac{\exp(-\frac{u^2}{2})}{\sqrt{2\pi}}du.$$

Using the second part of Lemma A.17 leads to

$$\begin{aligned} I(s,\lambda) \leq & 2\pi\sqrt{\frac{c'}{c}}\frac{1}{\lambda\sqrt{s-t_{i+1}}}\left(\frac{\sigma_0^2}{1+\sigma_0^2}\frac{e^{-x_0^2/(2\sigma_0^2)}}{\sqrt{2\pi\sigma_0^2}}+\frac{|x_0|}{1+\sigma_0^2}\frac{e^{-\frac{x_0^2}{2(1+\sigma_0^2)}}}{\sqrt{2\pi(1+\sigma_0^2)}}\right) \\ & := I_1(s,\lambda)+I_2(s,\lambda). \end{aligned}$$

Replacing σ_0 and x_0 by their values leads to

$$\begin{aligned} I_1(s,\lambda) &= \frac{\sqrt{2\pi}c'\sqrt{t_{i+1}-\tau^{t_i}}}{c\lambda^2(t_{i+1}-\tau^{t_i})+c'(s-t_{i+1})}\exp\left(-\frac{cF^2(s,\log(K))}{s-t_{i+1}}\right), \\ I_2(s,\lambda) &= \frac{2\sqrt{\pi}cc'|F(s,\log(K))|\lambda(t_{i+1}-\tau^{t_i})}{\sqrt{s-t_{i+1}}(c\lambda^2(t_{i+1}-\tau^{t_i})+c'(s-t_{i+1}))^{\frac{3}{2}}}\exp\left(-\frac{cc'F^2(s,\log(K))}{c\lambda^2(t_{i+1}-\tau^{t_i})+c'(s-t_{i+1})}\right). \end{aligned}$$

We aim at proving that $\int_{t_{i+2}}^T ds\int_0^1 d\lambda I_1(s,\lambda)\leq C\sqrt{h}$ and $\int_{t_{i+2}}^T ds\int_0^1 d\lambda I_2(s,\lambda)\leq C\sqrt{h}$. Let us compute $\int_0^1 d\lambda I_1(s,\lambda)$.

$$\begin{aligned} \int_0^1 I_1(s,\lambda)d\lambda &= \sqrt{2\pi}c'\sqrt{t_{i+1}-\tau^{t_i}}\exp\left(-\frac{cF^2(s,\log(K))}{s-t_{i+1}}\right)\int_0^1\frac{d\lambda}{c\lambda^2(t_{i+1}-\tau^{t_i})+c'(s-t_{i+1})}, \\ &= \sqrt{2\pi}\frac{\sqrt{t_{i+1}-\tau^{t_i}}}{s-t_{i+1}}\exp\left(-\frac{cF^2(s,\log(K))}{s-t_{i+1}}\right)\int_0^1\frac{d\lambda}{\frac{c\lambda^2(t_{i+1}-\tau^{t_i})}{c'(s-t_{i+1})}+1}. \end{aligned}$$

A change of variables leads to

$$\int_0^1 I_1(s,\lambda)d\lambda = \frac{\sqrt{2\pi}c'}{\sqrt{s-t_{i+1}}}\exp\left(-\frac{cF^2(s,\log(K))}{s-t_{i+1}}\right)\arctan\left(\sqrt{\frac{c(t_{i+1}-\tau^{t_i})}{c'(s-t_{i+1})}}\right).$$

Since for $x\geq 0$, $\arctan(x)\leq x$, we get

$$\int_{t_{i+2}}^T\int_0^1 I_1(s,\lambda)d\lambda ds \leq \int_{t_{i+2}}^T\frac{\sqrt{2\pi}\sqrt{cc'}\sqrt{t_{i+1}-\tau^{t_i}}}{s-t_{i+1}}\exp\left(-\frac{cF^2(s,\log(K))}{s-t_{i+1}}\right)ds.$$

Since $F(T,\log(K))=0$ and by assumptions, we get $c_0(T-s)\leq F(s,\log(K))\leq c_1(T-s)$. Then,

$$\int_{t_{i+2}}^T\int_0^1 I_1(s,\lambda)d\lambda ds \leq \sqrt{2\pi}\sqrt{cc'}\sqrt{t_{i+1}-\tau^{t_i}}\int_{t_{i+2}}^T\frac{1}{s-t_{i+1}}\exp\left(-C'\frac{(T-s)^2}{s-t_{i+1}}\right)ds.$$

Since $\forall u\geq 0$, $u^{1/4}e^{-u}\leq C$, we get $\exp\left(-C'\frac{(T-s)^2}{s-t_{i+1}}\right)\leq C'\frac{(s-t_{i+1})^{1/4}}{\sqrt{T-s}}$ and

$$\int_{t_{i+2}}^T\int_0^1 I_1(s,\lambda)d\lambda ds \leq C\sqrt{t_{i+1}-\tau^{t_i}}\int_{t_{i+2}}^T\frac{ds}{\sqrt{T-s}(s-t_{i+1})^{3/4}}\leq C\sqrt{h}.$$

It remains to prove $\int_{t_{i+2}}^T ds\int_0^1 d\lambda I_2(s,\lambda)\leq C\sqrt{h}$. To do so, we introduce the change of variables $r=c\lambda^2(t_{i+1}-\tau^{t_i})+c'(s-t_{i+1})$. Hence,

$$\begin{aligned} \int_0^1 I_2(s,\lambda)d\lambda &= \frac{2\sqrt{\pi}c'}{\sqrt{s-t_{i+1}}}\int_{c'(s-t_{i+1})}^{c(t_{i+1}-\tau^{t_i})+c'(s-t_{i+1})}\frac{|F(s,\log(K))|}{r^{3/2}}\exp\left(-\frac{cc'F^2(s,\log(K))}{r}\right)dr, \\ &\leq \frac{2\sqrt{\pi}c'}{\sqrt{s-t_{i+1}}}\int_{c'(s-t_{i+1})}^{c(t_{i+1}-\tau^{t_i})+c'(s-t_{i+1})}\frac{dr}{r} = \frac{2\sqrt{\pi}c'}{\sqrt{s-t_{i+1}}}\ln\left(1+\frac{c(t_{i+1}-\tau^{t_i})}{c'(s-t_{i+1})}\right). \end{aligned}$$

Then, $\int_{t_{i+2}}^T \int_0^1 I_2(s,\lambda)d\lambda ds \leq 2\sqrt{\pi}c' \int_{t_{i+2}}^T \frac{ds}{\sqrt{s-t_{i+1}}} \ln\left(1+\frac{c(t_{i+1}-\tau^{t_i})}{c'(s-t_{i+1})}\right)$. We change variables by defining $v = \frac{c(t_{i+1}-\tau^{t_i})}{c'(s-t_{i+1})}$ and we obtain

$$\int_{t_{i+2}}^T \int_0^1 I_2(s,\lambda)d\lambda ds \leq 2\sqrt{\pi}\sqrt{cc'}\sqrt{t_{i+1}-\tau^{t_i}} \int_0^\infty \frac{\ln(1+v)}{v^{3/2}} dv \leq C\sqrt{h}.$$

□

18.4 Proofs of technical results

This part is devoted to the proofs of Lemma 18.13 and Proposition 18.8

18.4.1 Proof of Lemma 18.13

We recall the result stated in Lemma 18.13:

There are some positive constants C and N_0 such that for $N \geq N_0$, for any $i \in \{0, \cdots, N-1\}$, one has for $X_{t_i} \in \mathcal{D}_{t_i}$

$$\mathbb{P}\left(\exists\, t \in [t_i, t_{i+1}] : X_t \notin \mathcal{D}_t \big| \mathcal{F}_{t_i}\right) \leq C\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{t_i}).$$

Proof of Lemma 18.13. Assume that on the set $t_i < \tau^{t_i} \leq t_{i+1}$, $\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{\tau^{t_i}}) \geq \frac{1}{C}$. Then,

$$\begin{aligned}\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{t_i}) &= \mathbb{E}[\mathbf{1}_{\tau^{t_i} \leq t_{i+1}} \mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{\tau^{t_i}}) \big| \mathcal{F}_{t_i}]\\ &\geq \frac{1}{C}\mathbb{P}(\tau^{t_i} \leq t_{i+1} \big| \mathcal{F}_{t_i}).\end{aligned}$$

One gets $\mathbb{P}(\tau^{t_i} \leq t_{i+1} \big| \mathcal{F}_{t_i}) \leq C\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{t_i})$. Since $\mathbb{P}(\tau^{t_i} \leq t_{i+1} \big| \mathcal{F}_{t_i}) = \mathbb{P}(\exists\, t \in [t_i, t_{i+1}] : X_t \notin \mathcal{D}_t \big| \mathcal{F}_{t_i})$, the result follows. It remains to prove that on the set $t_i < \tau^{t_i} \leq t_{i+1}$, $\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{\tau^{t_i}}) \geq \frac{1}{C}$. To do so, we adapt the proof of Gobet [37], Lemma 5.1: in the cited paper, the author deals with a time independent C^2 domain $\mathcal{D}$. Since $\mathcal{D}$ is regular, it satisfies an uniform exterior sphere condition, and it also satisfies an uniform truncated exterior cone condition. In our case, $\mathcal{D}$ being a regular time dependent domain, the exterior cone condition is replaced by an exterior tusk condition (see Proposition 17.2). Since $\mathcal{D} \in \mathcal{H}_1$, there exists $R > 0$, $\delta > 0$, $\overline{x}_0 \in \mathbb{R}^d$ and the tusk

$$\mathcal{T}_{\overline{x}_0} = \{(s,x) : \tau^{t_i} < s < \tau^{t_i} + \delta, |x - X_{\tau^{t_i}} - \overline{x}_0\sqrt{s-\tau^{t_i}}|^2 < R^2(s-\tau^{t_i})\},$$

such that $\overline{\mathcal{T}}_{\overline{x}_0}$ intersects $\overline{\mathcal{D}}$ only at $(\tau^{t_i}, X_{\tau^{t_i}})$. Then, for h small enough (i.e. $h \leq \delta$), we get

$$\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{\tau^{t_i}}) \geq \mathbb{P}((t_{i+1}, X_{t_{i+1}}) \in \mathcal{T}_{\overline{x}_0} \big| \mathcal{F}_{\tau^{t_i}}).$$

Under Hypothesis 16.1, we can use a Aronson type lower bound for the conditional density of $X_{t_{i+1}}$ (see (6.7), page 66). This yields

$$\begin{aligned}&\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{\tau^{t_i}}) \geq\\ &C\int_{y:|y-X_{\tau^{t_i}}-\overline{x}_0\sqrt{t_{i+1}-\tau^{t_i}}|^2<R^2(t_{i+1}-\tau^{t_i})} \exp\left(-c\frac{|X_{\tau^{t_i}}-y|^2}{t_{i+1}-\tau^{t_i}}\right)\frac{dy}{(t_{i+1}-\tau^{t_i})^{d/2}}.\end{aligned}$$

Let $z := \frac{y-X_{\tau^{t_i}}}{\sqrt{t_{i+1}-\tau^{t_i}}}$. This change of variables gives

$$\mathbb{P}(X_{t_{i+1}} \notin \mathcal{D}_{t_{i+1}} \big| \mathcal{F}_{\tau^{t_i}}) \geq C \int_{z:|z-\overline{x}_0|<R^2} \exp(-cz^2)dz.$$

The r.h.s. is equal to a strictly positive constant independent of t, which ends the proof. □

18.4.2 Proof of Proposition 18.8

We recall the result stated in Proposition 18.8:

There exists a positive constant c such that

$$\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds \leq Ch \sum_{j=0,\, t_j \geq t}^{N-1} \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > t_j} \exp\left(-c\frac{d(X_{t_j}, \partial\mathcal{D}_{t_j})^2}{h}\right)\right] = O(h).$$

For proving Proposition 18.8, we follow the proof of Gobet and Menozzi [41], Lemma 10, i.e. we prove three Lemmas, equivalent to Gobet and Menozzi [41], Lemmas 14, 15 and 17. Although in the cited paper the domain $\mathcal{D}$ doesn't depend on time, we can easily adapt their proofs to a time dependent domain. When the adaptation is straightforward, we only refer to this paper, otherwise, we do the proofs.

Lemma 18.15. *It holds $\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds \leq C\sqrt{h}$.*

Proof of Lemma 18.15.

$$\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds = \sum_{j=0,\, t_j \geq t}^{N-1} \int_{t_j}^{t_{j+1}} \mathbb{P}_{t,x}(t_j < \tau_N, X_s \notin \mathcal{D}_s)ds. \tag{18.10}$$

Moreover, $\mathbb{P}_{t,x}(t_j < \tau_N, X_s \notin \mathcal{D}_s) = \mathbb{E}_{t,x}[\mathbf{1}_{t_j<\tau_N}\mathbb{P}(X_s \notin \mathcal{D}_s \big| \mathcal{F}_{t_j})]$. We develop $\mathbb{P}(X_s \notin \mathcal{D}_s \big| \mathcal{F}_{t_j})$.

$$\mathbb{P}(X_s \notin \mathcal{D}_s \big| \mathcal{F}_{t_j}) = \mathbb{P}(X_s \notin \mathcal{D}_s, X_{t_j} \in \mathcal{D}_s \big| \mathcal{F}_{t_j}) + \mathbb{P}(X_s \notin \mathcal{D}_s, X_{t_j} \notin \mathcal{D}_s \big| \mathcal{F}_{t_j}). \tag{18.11}$$

We bound the first term of the r.h.s. of 18.11 by using Lemma A.14

$$\begin{aligned}\mathbb{P}(X_s \notin \mathcal{D}_s, X_{t_j} \in \mathcal{D}_s \big| \mathcal{F}_{t_j}) &\leq \mathbb{P}\left(|X_s - X_{t_j}| \geq d(X_{t_j}, \partial\mathcal{D}_s) \big| \mathcal{F}_{t_j}\right) \\ &\leq C\exp\left(-c\frac{d^2(X_{t_j}, \partial\mathcal{D}_s)}{h}\right).\end{aligned} \tag{18.12}$$

Using Lemma 18.7 leads to $\mathbb{P}(X_s \notin \mathcal{D}_s, X_{t_j} \in \mathcal{D}_s \big| \mathcal{F}_{t_j}) \leq C\exp\left(-c\frac{d^2(X_{t_j}, \partial\mathcal{D}_{t_j})}{h}\right)$.

We bound the second term of the r.h.s. of 18.11 in the following way

$$\mathbb{E}_{t,x}[\mathbf{1}_{t_j<\tau_N}\mathbb{P}(X_s \notin \mathcal{D}_s, X_{t_j} \notin \mathcal{D}_s \big| \mathcal{F}_{t_j})] \leq \mathbb{P}_{t,x}(X_{t_j} \notin \mathcal{D}_s, X_{t_j} \in \mathcal{D}_{t_j}) \leq \mathbb{P}_{t,x}(d(X_{t_j}, \partial\mathcal{D}_{t_j}) \leq Ch).$$

The second inequality ensues from the equality $\left|F(t_j, X_{t_j}) - F(s, X_{t_j})\right| \leq ch$. From this, we deduce

$$\begin{aligned}\mathbb{E}_{t,x}[\mathbf{1}_{t_j<\tau_N}\mathbb{P}(X_s \notin \mathcal{D}_s, X_{t_j} \notin \mathcal{D}_s \big| \mathcal{F}_{t_j})] &\leq \mathbb{E}_{t,x}[\mathbf{1}_{t_j<\tau_N}\mathbb{E}[\mathbf{1}_{d(X_{t_j},\partial\mathcal{D}_{t_j})\leq Ch} \big| \mathcal{F}_{t_j}]], \\ &\leq e^1\mathbb{E}_{t,x}\left[\mathbf{1}_{t_j<\tau_N}\mathbb{E}\left[\exp\left(-c\frac{d^2(X_{t_j}, \partial\mathcal{D}_{t_j})}{h}\right) \big| \mathcal{F}_{t_j}\right]\right].\end{aligned}$$

Combining the previous result, (18.12) and (18.11) and (18.10) leads to

$$\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds \le Ch \sum_{j=0}^{N-1} \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > t_j} \exp\left(-c\frac{d(X_{t_j}, \partial\mathcal{D}_{t_j})^2}{h}\right)\right]. \quad (18.13)$$

The above result also corresponds to the inequality of Proposition 18.8. Let us show that $h\sum_{j=0}^{N-1} \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > t_j} \exp\left(-c\frac{d(X_{t_j}, \partial\mathcal{D}_{t_j})^2}{h}\right)\right] = O(\sqrt{h})$. First, we write

$$\begin{aligned} &h \sum_{j=0}^{N-1} \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > t_j} \exp\left(-c\frac{d(X_{t_j}, \partial\mathcal{D}_{t_j})^2}{h}\right)\right] \\ &\le h + \sum_{j=2}^{N-1} \int_{t_{j-1}}^{t_j} \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > t}\mathbb{E}\left[\exp\left(-c\frac{d(X_{t_j}, \partial\mathcal{D}_{t_j})^2}{h}\right)|\mathcal{F}_t\right]\right] dt. \end{aligned}$$

Combining (18.13), Lemma A.14 and the last inequality of (18.12) with $s = t_j$ and $t_j = t$ leads to

$$\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds \le h + \int_t^T \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > s} \exp\left(-c\frac{d(X_s, \partial\mathcal{D}_s)^2}{h}\right)\right] ds. \quad (18.14)$$

Then, we split the last expectation

$$\begin{aligned} \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > s} \exp\left(-c\frac{d(X_s, \partial\mathcal{D}_s)^2}{h}\right)\right] =& \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > s}\mathbf{1}_{d(X_s,\partial\mathcal{D}_s)\le\frac{r_0}{4}} \exp\left(-c\frac{d(X_s, \partial\mathcal{D}_s)^2}{h}\right)\right] \\ &+ \mathbb{E}_{t,x}\left[\mathbf{1}_{\tau_N > s}\mathbf{1}_{d(X_s,\mathcal{D}_s)>\frac{r_0}{4}} \exp\left(-c\frac{d(X_s, \partial\mathcal{D}_s)^2}{h}\right)\right], \end{aligned}$$

where r_0 is defined in Definition 17.1. By using Lemma A.14, we show that the second term of the r.h.s. of the above equality is $O_{pol}(h)$. Plugging this result in (18.14) yields

$$\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds \le Ch + \mathbb{E}_{t,x}\left[\int_t^{\tau_N} \mathbf{1}_{F(s,X_s)\le\frac{r_0}{4}} \exp\left(-c\frac{F^2(s, X_s)}{h}\right) ds\right].$$

Applying Itô's formula to $F(s, X_s)$ gives $d\langle F(s, X_s)\rangle = |(\sigma\nabla F)(s, X_s)|^2 ds$. Since σ is uniformly elliptic and $X_s \in \partial\mathcal{D}_s(r_0/2)$, $|(\sigma\nabla F)(s, X_s)|^2 > a_0$, where a_0 is a strictly positive constant. Then, $d\langle F(s, X_s)\rangle \ge a_0 ds$, and the occupation times formula leads to

$$\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds \le Ch + C\int_{-r_0/4}^{r_0/4} \exp\left(-c\frac{y^2}{h}\right)\mathbb{E}_{t,x}[L^y_{\tau_N}(F(\cdot, X_\cdot))]dy. \quad (18.15)$$

Moreover, $\mathbb{E}_{t,x}[L^y_{\tau_N}(F(\cdot, X_\cdot))] \le \mathbb{E}_{t,x}[L^y_T(F(\cdot, X_\cdot))]$. Since σ and b are bounded, Tanaka's formula enables to write $\mathbb{E}_{t,x}[L^y_T(F(\cdot, X_\cdot))] \le \mathbb{E}_{t,x}[F^-(\cdot, X_\cdot)] + C$. Since F^- is a bounded function, we get $\mathbb{E}_{t,x}[L^y_{\tau_N}(F(\cdot, X_\cdot))] \le C$, and the proof is over. □

Lemma 18.16. *$\mathbb{E}_{t,x}[L^0_{\tau_N}(F(\cdot, X_\cdot))] \le C\sqrt{h}$.*

Proof of Lemma 18.16. We apply Itô-Tanaka's formula to $F(s,X_s)$ between t and τ_N

$$F^-(\tau_N, X_{\tau_N}) = F^-(t,x) - \int_t^{\tau_N} \mathbf{1}_{F(s,X_s)<0} dF(s,X_s) + \frac{1}{2} L^0_{\tau_N}(F(\cdot, X_\cdot)).$$

Since $(t,x) \in \mathcal{D}$, $F^-(t,x) = 0$, we get

$$\left|\mathbb{E}_{t,x}\left[\frac{1}{2}L^0_{\tau_N}(F(\cdot,X_\cdot)) - F^-(\tau_N, X_{\tau_N})\right]\right| \le \mathbb{E}_{t,x}\left[\int_t^{\tau_N} \mathbf{1}_{F(s,X_s)<0} dF(s,X_s)\right].$$

Since F is a C^2 function, we can use Itô's formula to get $\mathbb{E}_{t,x}\left[\int_t^{\tau_N} \mathbf{1}_{F(s,X_s)<0} dF(s,X_s)\right] = \int_t^T \mathbb{E}_{t,x}\left[\mathbf{1}_{s<\tau_N}\mathbf{1}_{F(s,X_s)<0}(\partial_t + \mathcal{L})F(s,X_s)\right] ds.$ Since b, σ, F and its derivatives are bounded, we obtain $\mathbb{E}_{t,x}\left[\mathbf{1}_{s<\tau_N}\mathbf{1}_{F(s,X_s)<0}(\partial_t + \mathcal{L})F(s,X_s)\right] \le C\mathbb{P}(s < \tau_N, X_s \notin \mathcal{D}_s)$, and

$$\left|\mathbb{E}_{t,x}\left[\frac{1}{2}L^0_{\tau_N}(F(\cdot,X_\cdot)) - F^-(\tau_N, X_{\tau_N})\right]\right| \le C\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s) ds. \tag{18.16}$$

Hence, since the r.h.s. is bounded by $C\sqrt{h}$ (see Lemma 18.15), it remains to control the expectation of the overshoot: $\mathbb{E}_{t,x}[F^-(\tau_N, X_{\tau_N})] = \sum_{i=1}^N \mathbb{E}_{t,x}[F^-(t_i, X_{t_i})\mathbf{1}_{\tau_N = t_i}]$. Since $\mathbf{1}_{\tau_N=t_i} = \mathbf{1}_{\tau_N > t_{i-1}} - \mathbf{1}_{\tau_N \ge t_{i+1}}$ and $\mathbb{E}_{t,x}[F^-(t_i, X_{t_i})\mathbf{1}_{\tau_N \ge t_{i+1}}] = 0$, we obtain

$$\begin{aligned}
\mathbb{E}_{t,x}[F^-(\tau_N, X_{\tau_N})] &= \sum_{i=1}^N \mathbb{E}_{t,x}[F^-(t_i, X_{t_i})\mathbf{1}_{\tau_N > t_{i-1}}] \\
&= \sum_{i=1}^N \mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_{i-1}} \in V_{\partial\mathcal{D}_{t_{i-1}}}(r_0/2)} \mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}]] \\
&+ \sum_{i=1}^N \mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_{i-1}} \notin V_{\partial\mathcal{D}_{t_{i-1}}}(r_0/2)} \mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}]],
\end{aligned} \tag{18.17}$$

Since $\mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}] = \mathbf{1}_{X_{t_i} \notin \mathcal{D}_{t_i}} \mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}]$ and $F^-(t_i, X_{t_i}) \le Ch + F^-(t_{i-1}, X_{t_i})$, we get

$$\begin{aligned}
\mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_{i-1}} \notin V_{\partial\mathcal{D}_{t_{i-1}}}(r_0/2)} \mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}]] &\le Ch\mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_i} \notin \mathcal{D}_{t_i}}] \\
&+ \mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_{i-1}} \notin V_{\partial\mathcal{D}_{t_{i-1}}}(r_0/2)} \mathbb{E}[F^-(t_{i-1}, X_{t_i}))\big|\mathcal{F}_{t_{i-1}}].
\end{aligned}$$

Moreover, $\mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_{i-1}} \notin V_{\partial\mathcal{D}_{t_{i-1}}}(r_0/2)} \mathbb{E}[F^-(t_{i-1}, X_{t_i}))\big|\mathcal{F}_{t_{i-1}}] \le C\mathbb{P}(|X_{t_i} - X_{t_{i-1}}| > r_0/2) = O_{pol}(h)$ and $\mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_i} \notin \mathcal{D}_{t_i}}] = \mathbb{E}_{t,x}[\mathbf{1}_{\tau_N = t_i}]$, then

$$\sum_{i=1}^N \mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_{i-1}} \notin V_{\partial\mathcal{D}_{t_{i-1}}}(r_0/2)} \mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}]] = O_{pol}(h).$$

Combining this result with (18.17) leads to

$$\mathbb{E}_{t,x}[F^-(\tau_N, X_{\tau_N})] = \sum_{i=1}^N \mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}} \mathbf{1}_{X_{t_{i-1}} \in V_{\partial\mathcal{D}_{t_{i-1}}}(r_0/2)} \mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}]] + O_{pol}(h). \tag{18.18}$$

It remains to bound from above $\mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}]$ on the set $\mathbf{1}_{\tau_N > t_{i-1}}\mathbf{1}_{X_{t_{i-1}} \in V_{\partial \mathcal{D}_{t_{i-1}}(r_0/2)}}$. $\mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}] = \mathbb{E}[\mathbf{1}_{\tau^{t_{i-1}} \le t_i}\mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{\tau^{t_{i-1}}}]\big|\mathcal{F}_{t_{i-1}}]$, where τ^t is defined in Definition 17.4. Since F^- is a Lipschitz function, we get

$$\mathbf{1}_{\tau^{t_{i-1}} \le t_i}\mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{\tau^{t_{i-1}}}] \le C(h + \mathbb{E}[|X_{t_i} - X_{\tau^{t_{i-1}}}|\big|\mathcal{F}_{\tau^{t_{i-1}}}])\mathbf{1}_{\tau^{t_{i-1}} \le t_i} \le C\sqrt{h}\mathbf{1}_{\tau^{t_{i-1}} \le t_i}.$$

Finally, we obtain $\mathbb{E}[F^-(t_i, X_{t_i})\big|\mathcal{F}_{t_{i-1}}] \le C\sqrt{h}\mathbb{P}(\tau^{t_{i-1}} \le t_i\big|\mathcal{F}_{t_{i-1}})$. Plugging this result in (18.18) and using Lemma 18.13 gives

$$\begin{aligned}\mathbb{E}_{t,x}[F^-(\tau_N, X_{\tau_N})] \le& C\sqrt{h}\sum_{i=1}^N \mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}}\mathbf{1}_{X_{t_{i-1}} \in V_{\partial \mathcal{D}_{t_{i-1}}(r_0/2)}}\mathbb{P}(X_{t_i} \notin \mathcal{D}_{t_i}\big|\mathcal{F}_{t_{i-1}})] + O_{pol}(h),\\ &\le C\sqrt{h}\sum_{i=1}^N \mathbb{E}_{t,x}[\mathbf{1}_{\tau_N > t_{i-1}}\mathbf{1}_{X_{t_i} \notin \mathcal{D}_{t_i}}] + O_{pol}(h) = O(\sqrt{h}),\end{aligned}$$

which ends the proof. The last equality ensues from $\mathbf{1}_{\tau_N > t_{i-1}}\mathbf{1}_{X_{t_i} \notin \mathcal{D}_{t_i}} = \mathbf{1}_{\tau_N = t_i}$. □

Lemma 18.17. *For $y \in [-r_0/4, r_0/4]$, $\mathbb{E}_{t,x}[L^y_{\tau_N}(F(\cdot, X.))] \le C(|y| + \sqrt{h})$.*

Proof of Lemma 18.17. Tanaka's formula gives

$$\begin{aligned}\mathbb{E}_{t,x}[L^y_{\tau_N}(F(\cdot, X.))] =& 2\mathbb{E}_{t,x}[(F(\tau_N, X_{\tau_N}) - y)^- - (F(t,x) - y)^-]\\ &+ 2\mathbb{E}_{t,x}\left[\int_t^T \mathbf{1}_{F(s,X_s) \le y}\mathbf{1}_{s \le \tau_N} d(F(s, X_s))\right].\end{aligned} \tag{18.19}$$

Since $(F(t,x) - y)^- \ge 0$ and $[(F(\tau_N, X_{\tau_N}) - y)^-] \le F^-(\tau_N, X_{\tau_N}) + |y|$, we get $\mathbb{E}_{t,x}[(F(\tau_N, X_{\tau_N}) - y)^- - (F(t,x) - y)^-] \le \mathbb{E}_{t,x}[F^-(\tau_N, X_{\tau_N})] + |y|$. Moreover, combining (18.16), Lemma 18.16 and Lemma 18.15 yields that $\mathbb{E}_{t,x}[F^-(\tau_N, X_{\tau_N})]$ is bounded by $\sqrt{h}$. We get

$$\mathbb{E}_{t,x}[(F(\tau_N, X_{\tau_N}) - y)^- - (F(t,x) - y)^-] \le C(\sqrt{h} + |y|).$$

Since σ, b, F and its derivatives are bounded, bounding the last term of (18.19) is equivalent to bounding $\omega(y) := \mathbb{E}_{t,x}\left[\int_t^T \mathbf{1}_{F(s,X_s) \le y}\mathbf{1}_{s \le \tau_N} ds\right]$. We aim at proving that $\omega(y) \le C(\sqrt{h} + |y|)$. For $y < 0$, since ω is increasing, Lemma 18.15 leads to $\omega(y) \le \omega(0) \le C\sqrt{h}$. For $y > 0$, it is enough to upper bound $\omega(y) - \omega(0)$ by $C(y + \sqrt{h})$. We have

$$\omega(y) - \omega(0) = \mathbb{E}_{t,x}\left[\int_t^T \mathbf{1}_{0 \le F(s,X_s) \le y}\mathbf{1}_{s \le \tau_N} ds\right].$$

Since $|y| \le \frac{r_0}{4}$, $X_s \in V_{\partial \mathcal{D}_s(r_0/2)}$. Then, $d\langle F(s, X_s)\rangle \ge a_0 ds$ and we get $\omega(y) - \omega(0) \le \frac{1}{a_0}\mathbb{E}_{t,x}\left[\int_t^T \mathbf{1}_{0 \le F(s,X_s) \le y}\mathbf{1}_{s \le \tau_N} d\langle F(s, X_s)\rangle\right]$. Using the occupation times formula leads to

$$\omega(y) - \omega(0) \le \frac{1}{a_0}\int_0^y \mathbb{E}_{t,x}[L^u_{\tau_N}(F(\cdot, X.))]du.$$

Since the expectation of the local time $L^u_{\tau_N}(F(\cdot, X.))$ is bounded for $u \in [0, \frac{r_0}{4}]$, we get $\omega(y) - \omega(0) \le Cy$, and the proof is over. □

Proof of Proposition 18.8. The first inequality corresponds to (18.13). To prove the second inequality, we use (18.15). It remains to bound $\int_{-r_0/4}^{r_0/4} \exp\left(-c\frac{y^2}{h}\right) \mathbb{E}_{t,x}[L^y_{\tau_N}(F(\cdot, X_\cdot))]dy$ by Ch. To do so, we use Lemma 18.17, and we get

$$\int_t^T \mathbb{P}_{t,x}(s < \tau_N, X_s \notin \mathcal{D}_s)ds \le C \int_{-r_0/4}^{r_0/4} \exp\left(-c\frac{y^2}{h}\right)(|y| + \sqrt{h})dy.$$

An easy computation leads to the result. $\square$

Appendix A

Appendix

A.1 Diffusions and Euler Scheme

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is defined a q-dimensional standard Brownian motion W, whose natural filtration, augmented with $\mathbb{P}$-null sets, is denoted by $(\mathcal{F}_t)_{0 \leq t \leq T}$ (T is a fixed terminal time) and $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d \times q}$.

Definition A.1. Consider the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x. \tag{A.1}$$

A solution of (A.1) is an $\mathcal{F}_t$ adapted d-dimensional process $(X_t)_{0 \leq t \leq T}$ such that

$$\int_0^T |b(s, X_s)| + |\sigma(s, X_s)|^2 ds < +\infty \ \ \mathbb{P} \text{ p.s.},$$

$$\mathbb{P} \text{ p.s.}, \forall t \in [0,T], X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s.$$

Hypothesis A.1 *There exists $K > 0$ such that $\forall t \in [0,T]$, $\forall x, y \in \mathbb{R}^d$,*

$$|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq K|x-y|,$$
$$\forall t \in [0,T], \forall x \in \mathbb{R}^d, |\sigma(t,x)| + |b(t,x)| \leq K(1+|x|).$$

Theorem A.2. *Assume Hypothesis A.1. Then, (A.1) has a unique solution X, and $\mathbb{E}[\sup_{s \leq t} |X_s|^2] < +\infty$.*

A.1.1 Euler scheme

Let $n \in \mathbb{N}, t_k = \frac{kT}{N}, 0 \leq k \leq N$. We define X^N, the approximation of X, by $X_0^N = x$ and for $t \in [t_k, t_{k+1}]$

$$X_t^N = X_{t_k}^N + b(t_k, X_{t_k}^N)(t - t_k) + \sigma(t_k, X_{t_k}^N)(W_t - W_{t_k}). \tag{A.2}$$

We easily deduce the following Lemma from (A.2).

Lemma A.3. *Assume σ and b are bounded. For all $t \in [t_k, t_{k+1}]$ and for all $p \geq 1$, $\mathbb{E}[|X_t^N - X_{t_k}^N|^2] \leq \frac{K(T)}{N^p}$, where $K(T)$ depends on p, $|b|_\infty$ and $|\sigma|_\infty$.*

Study of the convergence rate

Theorem A.4. *Assume Hypothesis A.1 and there exists $\alpha > 0$ such that $\forall\ 0 \leq s \leq t \leq T, \forall x \in \mathbb{R}^d$*

$$|\sigma(t,x) - \sigma(s,x)| + |b(t,x) - b(s,x)| \leq C(1+|x|)(t-s)^{\alpha}.$$

Then,

$$\forall p \geq 1,\ \ \mathbb{E}(\sup_{t\leq T} |X_t^N - X_t|^{2p}) \leq \frac{C}{N^{2p\beta}}(1+|x|^{2p}),$$

where $\beta = \min(\alpha, \frac{1}{2})$.

A.2 Linear PDEs

Let us consider the following linear parabolic PDE defined on $S = [0,T] \times \mathbb{R}^d$:

$$\begin{aligned} &(\partial_t + \mathcal{L}_{(t,x)})u(t,x) = f(t,x), \\ &u(T,x) = \Phi(x), \end{aligned} \tag{A.3}$$

where $\mathcal{L}_{(t,x)}$ is the second order differential operator

$$\mathcal{L}_{(t,x)}u(t,x) = \sum_{i,j} a_{ij}(t,x)\partial^2_{x_i x_j}u(t,x) + \sum_i b_i(t,x)\partial_{x_i}u(t,x) + c(t,x)u(t,x),$$

and $a_{ij}(t,x) = \frac{1}{2}[\sigma\sigma^*]_{ij}(t,x)$.

Definition A.5 (Ellipticity condition)**.** We say that the matrix σ satisfies the ellipticity condition if there exists a constant $\sigma_0 > 0$ s.t. $\forall (t,x) \in [0,T] \times \mathbb{R}^d$ $\sigma(t,x)\sigma^*(t,x) \geq \sigma_0^2 I_{\mathbb{R}^d \otimes \mathbb{R}^d}$.

A.2.1 The Cauchy Problem and a Feynman-Kac representation

We recall one version of the Feynman-Kac formula, coming from Karatzas and Shreve [56], page 366. In this section we consider a solution to the stochastic integral equation

$$X_s^{t,x} = x + \int_t^s b(\theta, X_\theta^{t,x})d\theta + \int_t^s \sigma(\theta, X_\theta^{t,x})dW_\theta;\ \ t \leq s < \infty. \tag{A.4}$$

under the standing assumptions that

Hypothesis A.2 *the coefficients $b_i(t,x), \sigma_{ij}(t,x) : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ are continuous and satisfy the following linear growth condition*

$$|b(t,x)|^2 + |\sigma(t,x)|^2 \leq K^2(1+|x|^2)$$

for every $0 \leq t < \infty$, $x, y \in \mathbb{R}^d$, where K is a positive constant.

Hypothesis A.3 *The equation (A.4) has a weak solution $(X^{t,x}, W), (\Omega, \mathcal{F}, \mathbb{P}), \{\mathcal{F}_s\}$ for every pair (t,x); and the solution is unique in the sense of probability law.*

With an arbitrary but fixed $T > 0$ and appropriate constants $L > 0, \lambda \geq 1$ we consider the functions $\Phi(x) : \mathbb{R}^d \to \mathbb{R}, f(t,x) : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ and $k(t,x) : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ which are continuous and satisfy

$$
\begin{aligned}
&|\Phi(x)| \leq L(1+|x|^{2\lambda}) \quad \text{or} \quad \Phi(x) \geq 0 \ \ \forall x \in \mathbb{R}^d \\
&|f(t,x)| \leq L(1+|x|^{2\lambda}) \quad \text{or} \quad f(t,x) \geq 0 \ \ \forall x \in \mathbb{R}^d, \ \ 0 \leq t \leq T.
\end{aligned}
$$

Theorem A.6 (Feynman-Kac formula)**.** *Under the preceding assumptions and hypotheses A.2 and A.3, suppose that $v(t,x) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is continuous, is of class $C^{1,2}([0,T) \times \mathbb{R}^d)$ and satisfies the Cauchy problem*

$$
\begin{aligned}
&\partial_t v + \mathcal{L}v + kv + f = 0 \ \ \textit{in } [0,T) \times \mathbb{R}^d; \\
&v(T,x) = \Phi(x), \ \ x \in \mathbb{R}^d,
\end{aligned}
\tag{A.5}
$$

as well as the polynomial growth condition

$$
\max_{0 \leq t \leq T} |v(t,x)| \leq M(1+|x|^{2\mu}); \ \ x \in \mathbb{R}^d,
\tag{A.6}
$$

for some $M > 0, \mu \geq 1$. Then, $v(t,x)$ admits the stochastic representation

$$
v(t,x) = \mathbb{E}_{t,x}\left[\Phi(X_T)\exp(-\int_t^T k(\theta,X_\theta)d\theta) + \int_t^T f(s,X_s)\exp(-\int_t^s k(\theta,X_\theta)d\theta)ds\right]
\tag{A.7}
$$

on $[0,T] \times \mathbb{R}^d$, in particular, such a solution is unique.

Remark A.7. Assume σ, b are continuous functions of $(t,x) \in [0,T] \times \mathbb{R}^d$ and Lipschitz functions in x, uniformly in time: $\exists K < \infty$ s.t.

$$
|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq K|x-y|, \ \ \forall t \geq 0, \forall x,y \in \mathbb{R}^d.
$$

Then, (A.4) has a unique strong (and weak) solution.

Remark A.8. For conditions under which the Cauchy problem (A.5) has a solution satisfying the polynomial growth (A.6), one should consult Friedman [28] Chapter I. One set of conditions under which there exists a unique solution v of the Cauchy problem (A.5) satisfies (A.6) are the following

- Uniform ellipticity condition of Definition A.5.
- Boundedness: the functions a_{ij}, b_i, k are bounded in $[0,T] \times \mathbb{R}^d$.
- Hölder continuity: the functions a_{ij}, b_i, k and f are uniformly Hölder continuous in $[0,T] \times \mathbb{R}^d$.
- Polynomial growth: f and Φ satisfy: $|f(t,x)| \leq L(1+|x|^{2\mu})$ in $[0,T] \times \mathbb{R}^d$, $\Phi(t,x)$ and $|\Phi(x)| \leq L(1+|x|^{2\mu})$,

where $L > 0, \mu \geq 1$. The first three conditions can be relaxed somewhat by making them local requirements. We refer the reader to Friedman [29] page 147 for more details.

Remark A.9. Conversely to the Feynman-Kac formula, define $v(t,x)$ by formula (A.7). Assume that σ, b, k are C^2 in space and their partial derivatives w.r.t. x are continuous and bounded. f, Φ are supposed to be C^2 in space and their partial derivatives up to order 2 satisfy a polynomial growth condition :

$$\exists C > 0 \text{ s.t. } |f(t,x)| + |\Phi(x)| \le C(1 + \|x\|^n),\ \forall x \in \mathbb{R}^d.$$

From Theorem 2.2 of Talay [90] we get that $v(t,x)$ defined by formula (A.7) is twice continuously differentiable w.r.t. x and is solution of (A.5).

A.3 Well-known inequalities

A.3.1 On positive random variables

Proposition A.10. *Let X be a positive random variable, and g be a C^1 function. We also assume g is monotone or g' is integrable. Then, we have*

$$\mathbb{E}[g(X)] = g(0) + \mathbb{E}[\int_0^X g'(\epsilon)d\epsilon] = g(0) + \int_0^\infty \mathbb{P}(X \ge \epsilon)g'(\epsilon)d\epsilon. \tag{A.8}$$

Let X be a positive random variable with density. We have

$$\mathbb{E}[X\mathbf{1}_{X>\epsilon}] = \epsilon\mathbb{P}(X \ge \epsilon) + \int_\epsilon^\infty \mathbb{P}(X > x)dx \tag{A.9}$$

Proof. To prove the first part of (A.8), we use that g is C^1. Concerning the second equality, we write

$$\mathbb{E}[\int_0^X g'(\epsilon)d\epsilon] = \mathbb{E}[\int_0^\infty \mathbf{1}_{X>\epsilon}g'(\epsilon)d\epsilon].$$

Since g' is integrable (or g is monotone), we apply Fubini's theorem to get the second equality.

To prove (A.9), we introduce f, the density of X, and we use Fubini's theorem.

$$\begin{aligned}\int_\epsilon^\infty \mathbb{P}(X > x)dx &= \int_\epsilon^\infty \int_x^\infty f(y)dydx = \int_\epsilon^\infty dx \int_\epsilon^\infty dy\mathbf{1}_{y\ge x}f(y)\\ &= \int_\epsilon^\infty dyf(y)(y-\epsilon) = \mathbb{E}[X\mathbf{1}_{X>\epsilon}] - \epsilon\mathbb{P}(X \ge \epsilon).\end{aligned}$$

□

A.3.2 Bernstein's inequalities

On independent random variables

We refer to Györfi et al. [47], page 594, for a proof of the following lemma.

Lemma A.11 (Bernstein (1946)). *Let $X_1, \cdots, X_n$ be independent real-valued random variables, let $a, b \in \mathbb{R}$ with $a < b$, and assume that $X_i \in [a,b]$ with probability one ($i = 1, \cdots, n$). Let $\sigma^2 = \frac{1}{n}\sum_{i=1}^n Var(X_i) > 0$. Then, for all $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}(X_i))\right| > \epsilon\right) \le 2e^{-\frac{n\epsilon^2}{2\sigma^2+2\epsilon(b-a)/3}}.$$

On martingales

We refer to Revuz and Yor [88], page 145, for the following inequality.

Proposition A.12 (Bernstein's inequality). *Let $M_t, t \geq 0$ be a martingale s.t. $M_0 = 0$, $< M >_t \leq A_t$ a.s. where A_t is deterministic. For $y \geq 0$ we have $\mathbb{P}(M_t \geq y) \leq e^{-\frac{y^2}{2A_t}}$.*

Remark A.13. Let X be a diffusion process satisfying $dX_s = b(s, X_s)ds + \sigma(s, X_s)dW_s$, $X_t = x$. We assume b, σ are uniformly bounded. From the previous proposition one easily deduce, for $y \geq 0$

$$\mathbb{P}(X_s^{t,x} \geq y) \leq K(T) \exp\left(-\frac{(x-y)^2}{2\sigma_\infty^2 (s-t)}\right).$$

On Itô processes

We refer to Gobet [37], Lemma 4.1 for a proof of the following Lemma.

Lemma A.14. *Let $(Y_t)_{t\geq 0}$ be an Itô process defined by $dY_t = b_t dt + \sigma_t dW_t$, with adapted and uniformly bounded coefficients. Let S and S' be two stopping times upper bounded by T, s.t. $0 \leq S' - S \leq \Delta \leq T$. Then, for any $p \geq 1$ and $c' > 0$, there exists a constant $c > 0$ and a function $K(T)$, s.t. for any $\eta \geq 0$, one has*

$$\mathbb{P}\left(\sup_{t\in[S,S']} \|Y_t - Y_S\|_{\mathbb{R}^d} \geq \eta \big| \mathcal{F}_S\right) \leq K(T) \exp\left(-c\frac{\eta^2}{\Delta}\right),$$

$$\mathbb{E}\left(\sup_{t\in[S,S']} \|Y_t - Y_S\|_{\mathbb{R}^d}^p \big| \mathcal{F}_S\right) \leq C\Delta^{p/2},$$

$$\mathbb{E}\left(\exp\left(-c'\frac{d^2(Y_{S'}, \partial\mathcal{D})}{\Delta}\right) \big| \mathcal{F}_S\right) \leq C \exp\left(-c\frac{d^2(Y_S, \partial\mathcal{D})}{\Delta}\right).$$

A.4 Convolutions and Integral computations

A.4.1 Convolution

Definition A.15 (Convolution). Let f and g be two functions defined on $\mathbb{R}$. The convolution of f and g is written $f * g$. It is defined as the integral of the product of the two functions after one is reversed and shifted

$$(f * g)(y) = \int_{\mathbb{R}} f(y-x)g(x)dx.$$

Lemma A.16 (Application to Gaussian density functions). *Let $f : x \longmapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$ and $f : x \longmapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. It holds*

$$(f * g)(m) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(m-x)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1+\sigma^2}} e^{-\frac{m^2}{2(1+\sigma^2)}}.$$

More generally, if f (resp. g) is the density function of $\mathcal{N}(m, \sigma)$ (resp. $\mathcal{N}(m', \sigma')$) we get

$$(f * g)(0) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma'^2 + \sigma^2}} e^{-\frac{(m+m')^2}{2(\sigma'^2+\sigma^2)}}.$$

A.4.2 Integral computations

Lemma A.17. *Let* I_n *be* $\int_0^\infty u^n \frac{e^{-u^2/2}}{\sqrt{2\pi}} \frac{e^{-(x-u)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} du$. *It holds*

$$I_1 = \frac{\sigma^2}{1+\sigma^2} \frac{e^{-x^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} + \frac{x}{1+\sigma^2} I_0.$$

Moreover, $I_1 \leq \frac{\sigma^2}{1+\sigma^2} \frac{e^{-x^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} + \frac{|x|}{1+\sigma^2} \frac{e^{-\frac{x^2}{2(1+\sigma^2)}}}{\sqrt{2\pi(1+\sigma^2)}}$.

Proof. We get the above equality by integrating I_1 by parts. The inequality ensues from the first part of Lemma A.16. $\square$

Bibliography

[1] F. Antonelli. Backward-Forward Stochastic Differential Equations. *Annals of Applied Probability*, 3-3:777–793, 1993.

[2] D. Aronson. Bounds for the fundamental solution of a parabolic equation. *Bulletin of the American Mathematical Society*, 73:890–903, 1967.

[3] B. Arouna. Robbins-Monro algorithms and variance reduction in finance. *The Journal of Computational Finance*, 7(2):35–62, Winter 2003/04.

[4] K. Baggerly, D. Cox, and R. Picard. Exponential convergence of adaptive importance sampling for Markov chains. *Journal of Applied Probability*, 37:342–358, 2000.

[5] V. Bally and A. Matoussi. Weak solutions for SPDEs and Backward Doubly Stochastic Differential Equations. *Journal of Theoretical Probability*, 14:125–164, 2001.

[6] V. Bally and G. Pagès. Error analysis of the optimal quantization algorithm for obstacle problems. *Stochastic Processes and their Applications*, 106(1):1–40, 2003.

[7] V. Bally and D. Talay. The law of Euler scheme for stochastic differential equations. *Probability Theory and related Fields*, 104:43–60, 1996.

[8] V. Bally and D. Talay. The law of Euler scheme for stochastic differential equations: convergence rate of the density. *Monte Carlo Methods Appl.*, 2(2):93–128, 1996.

[9] G. Barles and E. Lesigne. SDE, BSDE and PDE. *Pitman Research Notes in Mathematics Series*, 364:47–80, 1997.

[10] G. Barles, J. Burdeau, M. Romano, and N. Samsoen. Critical stock price near expiration. *Mathematical Finance*, 5(2):77–95, 1995.

[11] C. Bender and R. Denk. A forward scheme for backward SDEs. *Stochastic Processes and their Applications*, doi:10.1016/j.spa.2007.03.005, 2007.

[12] A. Bensoussan. On the theory of option pricing. *Acta Applicandae Mathematicae*, 2: 139–158, 1984.

[13] A. Bensoussan and J. Lions. *Applications des inéquations variationnelles en contrôle stochastique*. Dunod, 1978.

[14] J. Bismut. Conjugate Convex Functions in Optimal Stochastic Control. *J. Math. Anal. Appl.*, 44:384–404, 1973.

[15] B. Bouchard and N. Touzi. Discrete time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications*, 111:175–206, 2004.

[16] K. Brekke and B. Øksendal. The high contact principle as a sufficiency condition for optmial stopping. *Stochastic models and option values, Contribution to Economic Analysis*, 200:187–208, 1991.

[17] P. Briand, B. Delyon, and J. Mémin. Donsker-type theorem for BSDEs. *Electron. Comm. Probab*, 6:1–14, 2001.

[18] M. Broadie and J. Detemple. The valuation of American options on multiple assets. *Mathematical Finance*, 7(3):241–286, 1997.

[19] J. Chancelier, B. Pinçon, and F. Delbecque. Nsp. http://cermics.enpc.fr/~jpc/scilab/site/Scilab-Gtk/Scilab-Gtk/index.htm.

[20] J. Chang and Y. Peres. Ladder heights, Gaussian random walks and the Riemann zeta function. *Ann. Probab.*, 25(2):787–802, 1997.

[21] D. Chevance. Numerical methods for Backward Stochastic Differential Equations. In C. Rogers and D. T. editors, editors, *Newton Institute*, pages 232–244. Cambridge University Press, 1997.

[22] C. Costantini, E. Gobet, and N. El Karoui. Boundary sensitivities for diffusion processes in time dependent domains. *Applied Mathematics and Optimization*, 54(2): 159–187, 2006.

[23] F. Delarue. On the existence and uniqueness of solutions to FBSDEs in a non-degenerate case. *Stochastic Processes and their Applications*, 99:209–286, 2002.

[24] F. Delarue and S. Menozzi. A forward-backward stochastic algorithm for quasi-linear PDEs. *Annals of Applied Probability*, 16(1):140–184, 2006.

[25] M. Duflo. *Random Iterative Models*. Springer-Verlag, 1997.

[26] N. El Karoui, C. Kapoudjian, E. Pardoux, S. Peng, and M. Quenez. Reflected solutions of backward SDE's, and related obstacle problems for PDE's. *the Annals of Probability*, 25(2):702–737, 1997.

[27] N. El Karoui, S. Peng, and M. Quenez. Backward Stochastic Differential Equations in Finance. *Mathematical Finance*, 7(1):1–71, 1997.

[28] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[29] A. Friedman. *Stochastic Differential Equations and Applications*. Academic Press, New York, 1975.

[30] M. Fu and R. Wu. Optimal exercise policies and simulation-based valuation for American-Asian options. *Operations Research*, 51(1):52–66, 2003.

[31] M. Fu, S. Laprise, D. Madan, Y. Su, and R. Wu. Pricing American options: a comparison of Monte Carlo simulation approaches. *Journal of Computational Finance*, 4 (3):39–88, 2001.

[32] D. García. Convergence and biases of Monte Carlo estimates of American option prices using a parametric exercise rule. *J. Econom. Dynam. Control*, 27(10):1855–1879, 2003. ISSN 0165-1889.

[33] T. Gasser and H. Müller. Kernel estimation of regression functions. In Gasser and Rosenblatt, editors, *Smoothing Techniques for Curve Estimation*. Heidelberg: Springer-Verlag, 1979.

[34] T. Gasser and H. Müller. Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11:171–185, 1984.

[35] I. Gihman and A. Skorohod. *Stochastic Differential Equations*. Springer, Berlin, 1972.

[36] P. Glynn and R. Szechtman. Some new perspectives on the method of control variates. In H. N. ed. K.T. Fang, F.J. Hickernell, editor, *MonteCarlo and Quasi MonteCarlo Methods 2000*, pages 27–49. Berlin: Springer-Verlag, 2002.

[37] E. Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic processes and their Applications*, 87:167–197, 2000.

[38] E. Gobet. *Schémas d' Euler pour diffusion tuée. Application aux options barrière.* Ph.D. Thesis, Université Paris 7 Denis Diderot, 1998.

[39] E. Gobet and S. Maire. Sequential control variates for functionals of Markov processes. *SIAM Journal on Numerical Analysis*, 43(3):1256–1275, 2005.

[40] E. Gobet and S. Maire. A spectral Monte Carlo method for the Poisson equation. *Monte Carlo methods and applications*, 10(3-4):275–285, 2004.

[41] E. Gobet and S. Menozzi. Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme. *Stochastic Processes and their Applications*, 112: 201–223, 2004.

[42] E. Gobet and S. Menozzi. Discrete sampling of functionals of Itô processes. *Tech. report n° 559 CMAP, to appear in Séminaires de Probabilités, XL*, 2006.

[43] E. Gobet and S. Menozzi. Stopped diffusion processes: Overshoots and Boundary correction. *http://hal.archives-ouvertes.fr/hal-00157975/fr/*, 2007.

[44] E. Gobet and R. Munos. Sensitivity analysis using Itô-Malliavin calculus and martingales, and application to stochastic optimal control. *SIAM Journal on control and optimization*, 43(5):1676–1713, 2004.

[45] E. Gobet, J. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Annals of Applied Probability*, 15(3):2172–2202, 2005.

[46] J. Guyon. Euler schemes and tempered distributions. *Stochastic Processes and their Applications*, 116:877–904, 2006.

[47] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, 2002.

[48] P. Hall and C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, 1980.

[49] J. Halton. A retrospective and prospective survey of the Monte Carlo method. *SIAM Review*, 12:1–63, 1970.

[50] W. Härdle. *Applied nonparametric regression*. Cambridge University Press, 1992.

[51] A. Ibáñez and F. Zapatero. Valuation by simulation of american options through computation of the optimal exercise frontier. *Journal of Financial Quantitative Analysis*, 93:253–275, 2004.

[52] S. Jacka. Optimal Stopping and the American Put. *Mathematical Finance*, 1:1–14, 1991.

[53] J. Jacod and P. Protter. Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability*, 26(1):267–307, 1998.

[54] P. Jaillet, D. Lamberton, and B. Lapeyre. Variational inequalities and the pricing of American options. *Acta Appl. Math.*, 21:263–289, 1990.

[55] I. Karatzas. On the pricing of American options. *Appl. Math. Optim.*, 17(1):37–60, 1988.

[56] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus, Second Edition*. Springer-Verlag, New-York, 1991.

[57] I. Kim. The Analytic Valuation of American Options. *Review of financial studies*, 3: 547–572, 1990.

[58] S. Kim and S. Henderson. Adaptive control variates. In *Proceedings of the 2004 Winter Simulation Conference,*, 2004.

[59] A. Kohatsu-Higa and R. Pettersson. Variance reduction methods for simulation of densities on Wiener space. *SIAM J. Numer. Anal.*, 40(2):431–450, 2002.

[60] V. Konakov and E. Mammen. Edgeworth type expansions for Euler schemes for stochastic differential equations. *Monte Carlo Methods Appl*, 8(3):271–285, 2002.

[61] H. Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. *J. Theor. Prob.*, 7(2):247–278, 1994.

[62] T. Kurtz and P. Protter. Wong-Zakai Corrections, Random Evolutions, and Simulation Schemes for SDE's. In A. S. E. Mayer-Wolf, E. Merzbach, editor, *Stochastic Analysis*, pages 331–346. Academic Press, 1991.

[63] S. Kusuoka and D. Stroock. Applications to the Malliavin calculus I. In *Stochastic Analysis, Proceeding of the Taniguchi International Symposium on Katata and Kyoto 1982. Kinokuniya, Tokyo*, pages 271–306. Itô K. Ed, 1984.

[64] O. Ladyzenskaja, V. Solonnikov, and N. Ural'ceva. *Linear and quasi-linear equations of parabolic type.* Vol.23 of Translations of Mathematical Monographs, American Mathematical Society, Providence, 1968.

[65] D. Lamberton and B. Lapeyre. *Introduction to Stochastic Calculus Applied to Finance.* Chapman & Hall, New York and London, 1995.

[66] D. Lamberton and S. Villeneuve. Critical price near maturity for an american option on a dividend-paying stock. *Annals of Applied Probability*, 13(2):800–815, 2003.

[67] J. Lemor, E. Gobet, and X. Warin. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5), 2006.

[68] G. Lieberman. Intermediate Schauder theory for second order parabolic equations III. the tusk conditions. *Appl. Anal.*, 33(1-2):25–43, 1989.

[69] G. Lieberman. *Second order parabolic differential equations.* 1st edn. World Scientific Publishing Co. Inc., River Edge, NJ, 1996.

[70] P. Lions and H. Régnier. Calcul du prix des sensibilités d'une option Américaine par une méthode de Monte Carlo. *preprint*, 2001.

[71] F. Longstaff and R. Schwartz. Valuing American options by simulation : A simple least-square approach. *Review of Financial Studies*, 14:113–147, 2001.

[72] J. Ma and J. Yong. *Forward-Backward Stochastic Differential Equations and Their Applications*, volume 1702 of *Lecture Notes in Math.* Springer, 1999.

[73] J. Ma and J. Zhang. Representation theorems for backward stochastic differential equations. *The Annals of Applied Probability*, 12(4):1390–1418, 2002.

[74] J. Ma, P. Protter, and J. Yong. Solving forward backward stochastic differential equations explicitly-a four step scheme. *Probability Theory Related Fields*, 98(1): 339–359, 1994.

[75] S. Maire. An iterative computation of approximations on Korobov-like spaces. *J. Comput. Appl. Math*, 157(2):261–281, 2003.

[76] R. Myneni. The pricing of the American option. *The Annals of Applied Probability*, 2(1):1–23, 1992.

[77] N. Newton. Variance reduction for simulated diffusions. *SIAM Journal on Applied Mathematics*, 54(6):1780–1805, 1994.

[78] D. Nualart. *Malliavin Calculus and Related Topics.* Springer-Verlag, New York, 1995.

[79] D. Nualart. Analysis on Wiener space and anticipating stochastic calculus. In B. Springer-Verlag, editor, *Lectures on Probability Theory and Statistics (Saint- Flour, 1995)*, pages 123–227. 1998.

[80] E. Pardoux. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order. In *Stochastic analysis and related topics, VI (Geilo, 1996)*, volume 42 of *Progr. Probab.*, pages 79–127. Birkhäuser Boston, Boston, MA, 1998.

[81] E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Systems Control Lett.*, 14(1):55–61, 1990.

[82] E. Pardoux and S. Peng. Backward Stochastic Differential Equations and Quasilinear Parabolic Partial Differential Equations. *Lecture Notes in CIS*, 176(1):200–217, 1992.

[83] E. Pardoux and S. Tang. Forward-Backward Stochastic Differential Equations and Quasilinear Parabolic PDEs. *Probability Theory and Related Fields*, 114(2):123–150, 1999.

[84] S. Peng. Probabilistic Interpretation for Systems of Quasilinear Parabolic Partial Differential Equations. *Stochastics*, 37:61–74, 1991.

[85] S. Peng. Monotonic limit theorem of BSDE and nonlinear decomposition theorem of Doob-Meyer's type. *Probability Theory and Related Fields*, 113:473–499, 1999.

[86] S. Peng and M. Xu. Reflected BSDE with a Constraint and a New Doob-Meyer Nonlinear Decomposition. *arxiv:math.PR/0611869 v1*, 2007.

[87] P. Protter. *Stochastic integration and differential equations.* Springer Verlag, 1990.

[88] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer Verlag, 1991.

[89] D. Siegmund. Corrected diffusion approximations in certain random walk problems. *Adv. in Appl. Probab.*, 11(4):701–719, 1979.

[90] D. Talay. Discrétisation d'une équation différentielle stochastique et calcul approché d'espérances de fonctionnelles de la solution. *Mathematical Modelling and Numerical Analysis*, 20(1):141–179, 1986.

[91] P. Van Moerbeke. On Optimal Stopping and Free Boundary Problems. *Arch. Rational Mech. Anal*, 60:101–148, 1976.

[92] S. Villeneuve. On threshold strategies and smooth-fit principle for optimal stopping problems. *Journal of Applied Probability*, 44:181–198, 2007.

[93] S. Villeneuve. Exercise Regions of American Option on Several Assets. *Finance and Stochastics*, 3(3):295–322, 1999.

[94] J. Zhang. A numerical scheme for BSDEs. *The Annals of Applied Probability*, 14(1): 459–488, 2004.