



FiNANCE FOR ENERGY MARKET RESEARCH CENTRE



Option valuation and hedging using asymmetric risk function: asymptotic optimality through fully nonlinear Partial Differential Equations

Emmanuel Gobet, Isaque Pimentel and Xavier Warin

**Working Paper
RR-FiME-18-01**

April 2018

Option valuation and hedging using asymmetric risk function: asymptotic optimality through fully nonlinear Partial Differential Equations*

Emmanuel Gobet[†] Isaque Pimentel^{‡§} Xavier Warin[¶]

April 8, 2018

Abstract

Discrete time hedging produces a residual risk, namely, the tracking error. The major problem is to get valuation/hedging policies minimizing this error. We evaluate the risk between trading dates through a function penalizing asymmetrically profits and losses. After deriving the asymptotics within a discrete time risk measurement for a large number of trading dates, we derive the optimal strategies minimizing the asymptotic risk in the continuous time setting. We characterize the optimality through a class of fully nonlinear Partial Differential Equations (PDE). Numerical experiments show that the optimal strategies associated with discrete and asymptotic approach coincides asymptotically.

Keywords: hedging; asymmetric risk; fully nonlinear parabolic PDE; regression Monte Carlo.

MSC2010: 60H30, 35K55, 91G60, 91G80.

JEL: G13, C60.

1 Introduction

Statement of the problem. The valuation and hedging of contingent claims are major concerns in finance, both from a theoretical and a practical point of view. The continuous time theory is well established (see [12], for instance). But, in practice, hedging can be performed only at discrete times, say $t_0 = 0 < t_1 < \dots < t_N = T$, yielding a residual risk. Here, we intend to hedge the claim H_T at time T using d hedging instruments with price processes $X = (X^{(1)}, \dots, X^{(d)})$. So the *local risk* $\mathcal{E}_n$ associated with the hedging times t_n and t_{n+1} writes

$$\mathcal{E}_n = V_{t_{n+1}} - V_{t_n} - \langle \vartheta_{t_n}, X_{t_{n+1}} - X_{t_n} \rangle. \quad (1.1)$$

*This research is part of the Chair *Financial Risks* of the *Risk Foundation*, the *Finance for Energy Market Research Centre* (FiME) and the ANR project *CAESARS* (ANR-15-CE05-0024).

[†]Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France, email: emmanuel.gobet@polytechnique.edu

[‡]Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France, email: isaque.santa-brigida-pimentel@polytechnique.edu

[§]Optimization, Simulation, Risques et Statistiques (OSIRIS), Electricité de France (EDF), 7 boulevard Gaspard Monge, 91120 Palaiseau, France.

[¶]Optimization, Simulation, Risques et Statistiques (OSIRIS), Electricité de France (EDF), 7 boulevard Gaspard Monge, 91120 Palaiseau, France, email: xavier.warin@edf.fr

Here, V stands for the valuation process and $\vartheta = (\vartheta^{(1)}, \dots, \vartheta^{(d)})$, for the hedging process. Also, $\vartheta^{(i)}$ denotes the number of shares invested in the i -th hedging instrument. Up to considering discounted prices, we suppose the non-risky asset has zero drift.

In high-frequency hedging, the impact of discrete time hedging compared to continuous time one is small (see, for instance, [10] for results about convergence rate). In low-frequency hedging such as in energy markets [3], the local residual risk is slightly bigger and may become an issue. Our aim is to find the valuation/hedging rules (V, ϑ) minimizing this risk. We differ from the existing results (for instance, those related to the quadratic local risk minimization [6, 19]) by dealing with a risk function ℓ penalizing asymmetrically profits ($\mathcal{E}_n < 0$) and losses ($\mathcal{E}_n > 0$). So the *integrated local risk* under study takes the form

$$\mathcal{E}_N(V, \vartheta) = \sum_{n=0}^{N-1} \mathbb{E}[\ell(\mathcal{E}_n)].$$

The simplest case of such a risk function ℓ is

$$\ell_\gamma(y) = (1 + \gamma \operatorname{Sgn}(y))^2 y^2 / 2 \quad (1.2)$$

where $\gamma \in (0, 1)$ to penalize losses further than profits (see Figure 1). We define the above sign function as $\operatorname{Sgn}(y) := \mathbf{I}_{y>0} - \mathbf{I}_{y<0}$.

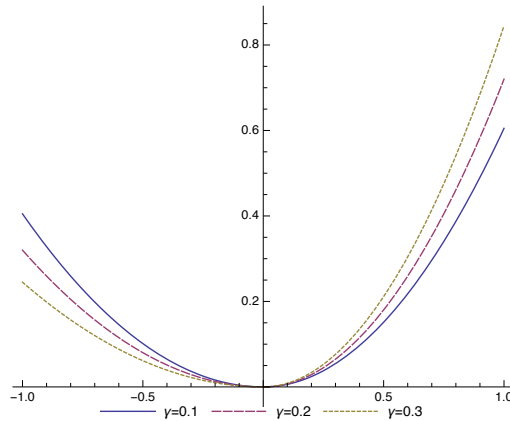


Figure 1: Plot of the risk function ℓ_γ for different γ .

In this setting, our aim is to study the asymptotics of the minimum

$$\min_{\substack{(V, \vartheta) \in \mathcal{A}_{V, \vartheta} \\ V_T = H_T}} \mathcal{E}_N(V, \vartheta) \quad (1.3)$$

as the number N of hedging dates becomes larger. To simplify we take equidistant hedging times $t_n = n\varepsilon_N$ with time step $\varepsilon_N = T/N$. The minimum (1.3) is computed over the set $\mathcal{A}_{V, \vartheta}$ of all adapted to the underlying filtration $(\mathcal{F}_t)_{t \geq 0}$ and appropriately integrable pair (V, ϑ) , under the replication constraint $V_T = H_T$.

There are a few results in that direction. In [15], the author deals with a L_p risk function of the losses and a fixed number trading dates. In [16], the authors consider expected shortfall risk function. Their research concentrates on numerics for a fixed number of dates and does not handle any asymptotic analysis. In [1], the authors study pseudo-optimal strategies and get asymptotic results under the condition that the risk function is of class C^3 . So their analysis discards the prototype risk function (1.2). Indeed, the

discontinuity of the second derivative ℓ''_γ complicates the analysis and fully changes the nature of subsequent results. In short, the existing references consider different settings and difficulties from ours.

Exogenous reference valuation and f -PDE valuation. The minimization problem (1.3) appears attractive, but its study in the asymptotic regime $N \rightarrow +\infty$ is tough in the case of asymmetric risk function (1.2). To tackle this problem, we slightly modify the approach.

First, we suppose the hedging instruments are modeled by a Stochastic Differential Equation (SDE) with drift μ and diffusion σ . We also consider contingent claims of the form $H_T = h(X_T)$. Second, we suppose that the contingent claim is evaluated exogenously by a valuation process $V_t = v(t, X_t)$ for some function v . For instance, v is given by a mark-to-model value promoted by the regulator or the central counterparty (CCP). The latter imposes its minimum margin requirement to which the hedging entity has to comply with. Given this exogenous reference valuation, the trader will determine how to hedge on each interval $[t_n, t_{n+1}]$ by choosing an adapted valuation/hedging rule $(\tilde{V}_{t_n}, \tilde{\vartheta}_{t_n})$ and considering the related *conditional local risk*

$$\mathcal{R}_n(\gamma) = \mathbb{E}[\ell_\gamma(V_{t_{n+1}} - \tilde{V}_{t_n} - \langle \tilde{\vartheta}_{t_n}, X_{t_{n+1}} - X_{t_n} \rangle) \mid \mathcal{F}_{t_n}]. \quad (1.4)$$

To clarify, the valuation/hedging rule of the trader will be parametrized by a function f , possibly nonlinear. Inspired by the connection between dynamic risk valuations, nonlinear Partial Differential Equations (PDE) and nonlinear Backward SDEs [5, 14, 4], we introduce the concept of the *f -PDE valuation*. Let

$$\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}, \quad f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d \times d} \rightarrow \mathbb{R}$$

be continuous functions. Let $\tau \in (0, T]$ be a time horizon and let $v(\tau, \cdot)$ be a reference valuation at the time τ . Given τ and $v(\tau, \cdot)$, the function $u_\tau : [0, \tau] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a solution to the *f -PDE*, if it satisfies

$$\begin{aligned} \partial_t u_\tau(t, x) + \frac{1}{2} \text{Tr}[\sigma \sigma^\top D_x^2 u_\tau](t, x) \\ + f(t, x, u_\tau(t, x), D_x u_\tau(t, x), D_x^2 u_\tau(t, x)) = 0, \end{aligned} \quad (1.5)$$

for all $(t, x) \in [0, T] \times \mathbb{R}^d$ with the terminal condition $u_\tau(\tau, x) = v(\tau, x)$ at the time τ . The *f -PDE valuation* is the mapping from $(\tau, v(\tau, \cdot))$ to the *f -PDE* (1.5) solution u_τ ; this is typically the nonlinear valuation/hedging rule of the trader. We refer to f as the kernel and $f \equiv 0$ corresponds to the usual risk-neutral valuation [12], other kernels appear in [5] for instance. Then, in the conditional local risk expression given by (1.4), we naturally set

$$\tilde{V}_{t_n} = u^{(n+1)}(t_n, X_{t_n}), \quad \tilde{\vartheta}_{t_n} = D_x u^{(n+1)}(t_n, X_{t_n}),$$

where we denote $u^{(n+1)} := u_{t_{n+1}}$.

Our contributions. Our first main result is to prove the existence (Theorem 2.1) of the following limit, called the *asymptotic risk*,

$$\mathcal{R}_\gamma(v, f) = \lim_{N \rightarrow +\infty} \frac{1}{\varepsilon_N} \sum_{n=0}^{N-1} \mathbb{E}[\mathcal{R}_n(\gamma)].$$

Moreover, we give an explicit expression for $\mathcal{R}_\gamma(v, f)$ depending on γ, v, f, σ, X and T . Then, we discuss the existence of an optimal kernel f^* such that the f^* -PDE valuation minimizes the asymptotic risk in the sense

$$\mathcal{R}_\gamma(v, f^*) \leq \mathcal{R}_\gamma(v, f), \quad (1.6)$$

for any admissible f . In the one-dimensional case, this optimal kernel f^* is explicit (see (2.15)) and it depends on the risk parameter γ , on the reference valuation second derivative and on the price process volatility.

Now, a natural choice for the reference valuation may be the solution to the f^* -PDE (1.5). Here, the payoff $h : \mathbb{R} \rightarrow \mathbb{R}$ is the f^* -PDE terminal condition at the time T . We denote by v^* the resulting valuation. In dimension one, this PDE takes the form

$$\partial_t v^*(t, x) + \frac{1}{2} \sigma^2(t, x) \partial_x^2 v^*(t, x) + c_1^* \sigma^2(t, x) (\partial_x^2 v^*(t, x))_+ - c_2^* \sigma^2(t, x) (\partial_x^2 v^*(t, x))_- = 0$$

for some constants $c_1^* \geq 0$ and $c_2^* \leq 0$ depending on the risk parameter γ . In higher dimension, v^* solves a fully nonlinear PDE with a nonlinear term depending on the Hessian $D_x^2 v^*$ (see the nonlinear PDE (2.12)).

It gives somehow a consistent way to value the claim h by accounting for local hedging errors measured with the asymmetric risk function ℓ_γ . To the best of our knowledge, this work is an original contribution, where local hedging errors are analyzed with asymmetric risk function. We perform the asymptotics of a large number of trading dates and we derive an optimal valuation/hedging policy.

Summing up, instead of minimizing (1.3) and then taking the limit in N after rescaling by ε_N , we take first the limit in N of the cumulated integrated local risk for a wide class of f -PDE valuation and then minimize over all kernels f . We do not prove that inverting minimization and limit holds true in this setting. In other words, we do not state the limit of the minimum (1.3) rescaled by ε_N corresponds to $\mathcal{R}_\gamma(v^*, f^*)$. However, our numerical tests in dimension one seem to corroborate this fact. Proving this result rigorously is, so far, an open problem, that we expect to handle in the next future.

The paper is structured as follows. Below, we present the notations and conventions used throughout the paper. In Section 2, we define the stochastic setting, then state the assumptions and the main results. The proofs are gathered in Section 3. Section 4 contains our numerical experiments. Some technical results are collected in Appendix 5.

Usual notations. Let $d \in \mathbb{N}^*$ and let a, b in $\mathbb{R}^d$. We denote by $\langle a, b \rangle = \sum_{i=1}^d a_i b_i$ the scalar product on $\mathbb{R}^d$, adopted for both row or column vectors a and b . We set $\|a\| = \sqrt{\langle a, a \rangle}$. We denote by $\mathcal{M}^d$ the set of all $d \times d$ matrices with real entries. By $\mathcal{S}^d$ we denote all symmetric matrices in $\mathcal{M}^d$. Let $A \in \mathcal{M}^d$, we denote $\text{Tr}[A]$ and $A^\top$ respectively the trace and the transpose of a matrix A . For A in $\mathcal{M}^d$, we set $\|A\| = \sqrt{\text{Tr}[AA^\top]}$.

Let E, E' be two generic Euclidean space and let $\phi : [0, T] \times E$ be a E' -valued function. In this paper, we say ϕ satisfies a *local regularity condition in time and space* if for some real $q > 0$ the coefficient

$$\|\phi\|_{H_{\text{loc}, \text{pol}}^{1/2, 1}} := \sup_{\substack{t, t' \in [0, T] \\ t \neq t'}} \sup_{\substack{x, x' \in E \\ x \neq x'}} \frac{\|\phi(t, x) - \phi(t', x')\|}{(|t - t'|^{1/2} + \|x - x'\|)(1 + \|x\|^q + \|x'\|^q)}$$

is finite, then ϕ is said to be in $H_{\text{loc}, \text{pol}}^{1/2, 1}$. We are aware that $\|\phi\|_{H_{\text{loc}, \text{pol}}^{1/2, 1}}$ depends on q but in the following, the precise value of q is unimportant and we prefer to avoid the reference to q in the notation $\|\phi\|_{H_{\text{loc}, \text{pol}}^{1/2, 1}}$ for the sake of simplicity.

Observe that $\phi \in H_{\text{loc}, \text{pol}}^{1/2, 1}$ means that ϕ is locally $1/2$ -Hölder continuous in time and Lipschitz continuous in space; and it has polynomial growth in space uniformly in time. Furthermore, we assert that for any ϕ_1 and ϕ_2 in $H_{\text{loc}, \text{pol}}^{1/2, 1}$, the product $\phi_1 \phi_2$, the pair (ϕ_1, ϕ_2) and the composition $\phi_1(t, \phi_2(t, \cdot))$ and are also in $H_{\text{loc}, \text{pol}}^{1/2, 1}$.

The set $C^{1,2}([0, T] \times E, E')$ denotes the set of functions $\phi : [0, T] \times E \rightarrow E'$ such that the partial derivatives $\partial_t \phi, \partial_{x_i} \phi, \partial_{x_i} \partial_{x_j} \phi$ exist and are continuous, for any $1 \leq i, j \leq d$. When $E = \mathbb{R}^d$ and the domain E' is unambiguous, we simply write $C^{1,2}([0, T] \times \mathbb{R}^d)$.

For every function $\phi \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, we denote its gradient in space by a row vector $D_x \phi = (\partial_{x_i} \phi)_{1 \leq i \leq d}$ and its Hessian by $D_x^2 \phi = (\partial_{x_i} \partial_{x_j} \phi)_{1 \leq i, j \leq d}$. Also, let $\mathcal{L}_t \phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ be given by

$$\mathcal{L}_t \phi(t, x) = \partial_t \phi(t, x) + \frac{1}{2} \text{Tr}[\sigma \sigma^\top D_x^2 \phi](t, x).$$

Notice that $\phi, \partial_t \phi, D_x \phi, D_x^2 \phi \in H_{\text{loc}, \text{pol}}^{1/2, 1}$ is a sufficient condition to have $\phi \in C^{1,2}$ and be able to apply Ito's formula.

2 Model, assumptions and main results

2.1 Probabilistic risk model

We fix a finite time horizon $T > 0$. Let $W = (W^{(1)}, \dots, W^{(d)}) : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ be a standard Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathbb{F} = \{\mathcal{F}_t, t \in [0, T]\}$ be the augmented and completed filtration generated by W . We consider the $\mathbb{F}$ -adapted process $X = (X^{(1)}, \dots, X^{(d)}) : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ satisfying the following stochastic differential equation (SDE)

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \quad (2.1)$$

with initial value $X_0 = x_0 \in \mathbb{R}^d$. The coefficients $\mu : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathcal{M}^d$ are Lipschitz in space uniformly in time (see Assumption A1 later).

Given $N \in \mathbb{N}^*$ equidistant hedging times $\{t_0 = 0 < t_1 < \dots < t_N = T\}$ on the interval $[0, T]$, with $t_n = n\varepsilon_N$ and $\varepsilon_N = T/N$, we write

$$\varphi_t^N := \sup \{t_n \mid t_n \leq t\}, \quad \bar{\varphi}_t^N := \inf \{t_n \mid t_n > t\}$$

and the increment of X from t_n to t_{n+1} as $\Delta X_n = X_{t_{n+1}} - X_{t_n}$.

In the following we systematically consider the risk function ℓ_γ as defined in (1.2). It is a convex and continuously differentiable function satisfying $\ell_\gamma(0) = \ell'_\gamma(0) = 0$ and $\ell_\gamma(y) = \ell_{-\gamma}(-y)$. In addition, it is symmetric if and only if $\gamma = 0$. Further, ℓ'_γ is a piecewise continuously differentiable function with ℓ''_γ being discontinuous as soon as $\gamma \neq 0$:

$$\ell'_\gamma(y) = (1 + \gamma \text{Sgn}(y))^2 y, \quad \ell''_\gamma(y) = (1 + \gamma \text{Sgn}(y))^2, \quad (2.2)$$

where ℓ''_γ is extended to zero as $\ell''_\gamma(0) = 1$, owing to $\text{Sgn}(0) = 0$ (see Figure 2). In all the sequel, we assume $\gamma \in [0, 1)$.

For a payoff function $h : \mathbb{R}^d \rightarrow \mathbb{R}$, the input of our approach are a reference valuation $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that $v(T, \cdot) = h(\cdot)$ and a kernel $f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{S}^d \rightarrow \mathbb{R}$. Both are assumed to be smooth functions (see Assumptions A2 and A3). So we associate the f -PDE valuation giving rise to the family of functions $u_{t_{n+1}} : [0, t_{n+1}] \times \mathbb{R}^d \rightarrow \mathbb{R}$ indexed by hedging times t_{n+1} . These functions are the solutions to the PDE (1.5)

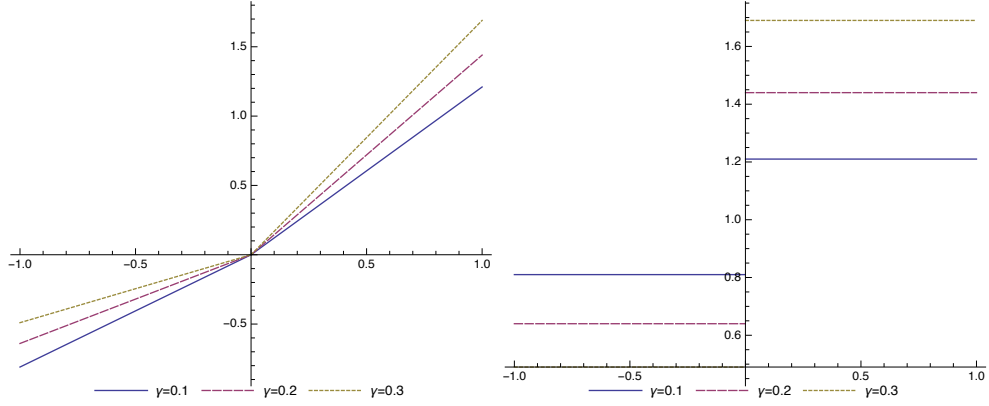


Figure 2: Plot of the risk function ℓ_γ derivatives for different γ : ℓ'_γ (left) and ℓ''_γ (right).

with Cauchy boundary condition $u_{t_{n+1}}(t_{n+1}, \cdot) = v(t_{n+1}, \cdot)$ at time horizon t_{n+1} . Also, they are assumed to be smooth in the sense of Assumption A4. In this context, we set $u^{(n+1)} = u_{t_{n+1}}$ and define the local residual risk $\mathcal{E}_n : \Omega \rightarrow \mathbb{R}$ (see (1.1)) by

$$\mathcal{E}_n = u^{(n+1)}(t_{n+1}, X_{t_{n+1}}) - u^{(n+1)}(t_n, X_{t_n}) - D_x u^{(n+1)}(t_n, X_{t_n}) \Delta X_n \quad (2.3)$$

and the conditional local risk

$$\mathcal{R}_n(\gamma) = \mathbb{E}[\ell_\gamma(\mathcal{E}_n) \mid \mathcal{F}_{t_n}]. \quad (2.4)$$

As explained in introduction, our aim is to analyze the asymptotic behavior of the integrated conditional local risk, after appropriate renormalization,

$$\mathcal{R}_{N,\gamma}(v, f) = \frac{1}{\varepsilon_N} \sum_{n=0}^{N-1} \mathbb{E}[\mathcal{R}_n(\gamma)]. \quad (2.5)$$

2.2 Asymptotic risk $\mathcal{R}_\gamma$ given a reference valuation v and a kernel f

Here, we study the asymptotic risk $\mathcal{R}_\gamma$ (as defined in (1.5)) where a reference valuation v and a kernel f are given. We state the following assumptions.

Assumption A1. The coefficients $\mu : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathcal{M}^d$ are in $H_{\text{loc}, \text{pol}}^{1/2, 1}$.

Assumption A2. The reference valuation $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is in $H_{\text{loc}, \text{pol}}^{1/2, 1}$. Further, $D_x v$ and $D_x^2 v$ exist and are in $H_{\text{loc}, \text{pol}}^{1/2, 1}$.

Assumption A3. The kernel $f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{S}^d \rightarrow \mathbb{R}$ is in $H_{\text{loc}, \text{pol}}^{1/2, 1}$.

Assumption A4. For all $\tau \in (0, T]$, there is a unique classical solution u_τ to the PDE (1.5) with the terminal condition $u_\tau(\tau, \cdot) = v(\tau, \cdot)$ at the time τ . In addition,

$$\partial_t u_\tau, \partial_{x_i} u_\tau, \partial_{x_i} \partial_{x_j} u_\tau, \partial_t \partial_{x_i} u_\tau, \partial_{x_i} \partial_{x_j} \partial_{x_k} u_\tau,$$

exist and are in $H_{\text{loc}, \text{pol}}^{1/2, 1}$.

Assumption A5. A non-degeneracy condition: the symmetric matrix $(\sigma^\top(D_x^2 v)\sigma)(t, X_t)$ is not 0 $dt \otimes d\mathbb{P}$ -a.e.

For stating the asymptotic result below, we need to introduce an extra Brownian motion B , independent of W , with the same dimension as W . All these are defined on an extended probability space with obvious definitions. Whenever necessary, the expectation w.r.t. the distribution of B , or W , or both, is denoted by $\mathbb{E}^B$, or $\mathbb{E}^W$, or $\mathbb{E}^{W \otimes B}$.

Theorem 2.1. Let $B = (B^{(1)}, \dots, B^{(d)}) : [0, 1] \times \Omega \rightarrow \mathbb{R}^d$ be another standard Brownian motion independent from W . Consider $\mathcal{R}_{N,\gamma}(v, f)$ given by (2.5) in the following form

$$\mathcal{R}_{N,\gamma}(v, f) = \frac{1}{\varepsilon_N} \sum_{n=0}^{N-1} \mathbb{E}[\ell_\gamma(\mathcal{E}_n)],$$

where $\mathcal{E}_n$ is given by (2.3). Under the Assumptions A1-A5, the limit of $\mathcal{R}_{N,\gamma}(v, f)$ as $N \rightarrow \infty$ exists and is given by,

$$\begin{aligned} \mathcal{R}_\gamma(v, f) = \mathbb{E} \left[\int_0^T \int_0^1 \ell''_\gamma \left(\int_0^\theta B_{\theta'}^\top G_t dB_{\theta'} - F_t \theta \right) \right. \\ \left. \times \left(F_t^2 \theta - F_t \int_0^\theta B_{\theta'}^\top G_t dB_{\theta'} + \|G_t B_\theta\|^2 / 2 \right) d\theta dt \right], \end{aligned} \quad (2.6)$$

where

$$\begin{aligned} F_t &= f(t, X_t, v(t, X_t), D_x v(t, X_t), D_x^2 v(t, X_t)) \in \mathbb{R}, \\ G_t &= (\sigma^\top(D_x^2 v)\sigma)(t, X_t) \in \mathcal{S}^d \end{aligned}$$

The long and delicate proof is postponed to Section 3.

2.3 Optimizing over the kernel f

Here, we study the optimization problem over the kernel f described in (1.6). To precise the definition of optimal kernel f^* , we rewrite the asymptotic risk in (2.6) as a functional $\mathcal{R}_\gamma : \Omega_v \times \Omega_v \rightarrow \mathbb{R}$ given by

$$\mathcal{R}_\gamma(v, f) = \mathbb{E} \left[\int_0^T R_\gamma(G_t, F_t) dt \right],$$

where $R_\gamma : \mathcal{S}^d \times \mathbb{R} \rightarrow \mathbb{R}$ is

$$\begin{aligned} R_\gamma(S, a) = \mathbb{E} \left[\int_0^1 \ell''_\gamma \left((B_\theta^\top S B_\theta - \text{Tr}[S]\theta) / 2 - a\theta \right) \right. \\ \left. \times \left(a^2 \theta - a (B_\theta^\top S B_\theta - \text{Tr}[S]\theta) / 2 + (B_\theta^\top S^\top S B_\theta) / 2 \right) d\theta \right], \end{aligned} \quad (2.7)$$

with

$$\Omega_v = \left\{ v \in H_{\text{loc}, \text{pol}}^{1/2, 1} \mid D_x v, D_x^2 v \in H_{\text{loc}, \text{pol}}^{1/2, 1} \right\}, \quad \Omega_f = \left\{ f \in H_{\text{loc}, \text{pol}}^{1/2, 1} \right\}.$$

We aim at proving the existence of minimizers to the variational problem

$$\min_{f \in \Omega_f} \mathcal{R}_\gamma(v, f), \quad (2.8)$$

for all $v \in \Omega_v$. Observe that the minimizer $f^\dagger(t, x, y, z, A)$ defined by (for any fixed (t, x, y, z, A))

$$f^\dagger(t, x, y, z, A) = \arg \min_{a \in \mathbb{R}} R_\gamma((\sigma^\top A \sigma)(t, x), a),$$

is also a minimizer to (2.8) under the condition to be in Ω_f . Indeed, we just need to integrate and to take the expectation in both sides of

$$R_\gamma(G(t, X_t), f^\dagger(t, X_t, v(t, X_t), D_x v(t, X_t), D_x^2 v(t, X_t))) \leq R_\gamma(G(t, X_t), F(t, X_t)).$$

This is why we seek a minimizer to $a \mapsto R_\gamma(S, a)$ for a given symmetric matrix S .

We now prove the existence of a minimizer.

Proposition 2.2. *Let $\gamma \in [0, 1)$ and $S \in \mathcal{S}^d$. Consider the minimization problem*

$$\min_{a \in \mathbb{R}} R_\gamma(S, a). \quad (2.9)$$

Under the hypothesis of Theorem 2.1, there exists a global minimizer $a^ \in \mathbb{R}$ such that $R_\gamma(S, a^*) \leq R_\gamma(S, a)$ for all $a \in \mathbb{R}$.*

If a^* is unique, a natural candidate for f^* is then given by

$$f^*(t, x, y, z, A) = a^*(\sigma^\top(t, x) A \sigma(t, x)), \quad (2.10)$$

for any $t, x, y, z, A \in [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{S}^d$.

Proof. First, we show that the function $R_\gamma(S, a)$ is coercive and continuous in a . For any $\theta \in (0, 1]$, we consider $Z_\theta^a = (B_\theta^\top S B_\theta - \text{Tr}[S|\theta])/2 - a\theta$. Through simple computations, we check that Z_θ^a is continuous in a and integrable w.r.t. $d\mathbb{P}^B \otimes d\theta$.

Regarding the coercivity, we exhibit a coercive function which bounds $R_\gamma(S, a)$ from below. Owing to the boundedness of ℓ''_γ , we estimate

$$\begin{aligned} \ell''_\gamma(Z_\theta^a) \left(-aZ_\theta^a + \|SB_\theta\|^2/2 \right) &\geq (1-\gamma)^2 \left(a^2\theta + \|SB_\theta\|^2/2 \right) \\ &\quad - (1+\gamma)^2 |a| (|B_\theta^\top S B_\theta| + |\text{Tr}[S]|\theta)/2, \end{aligned}$$

$d\mathbb{P}^B \otimes d\theta$ -almost surely. By integrating in θ and taking the expectation of the previous estimate, we get

$$R_\gamma(S, a) \geq (1-\gamma)^2 \left(a^2/2 + \text{Tr}[S^\top S]/4 \right) - (1+\gamma)^2 |a| (\mathbb{E}|G^\top S G| + |\text{Tr}[S]|)/4,$$

where G is a standard normal random vector. Then we conclude that $a \mapsto R_\gamma(S, a)$ is coercive.

Regarding the continuity, we first take $S = \mathbf{0}$ and we get

$$R_\gamma(\mathbf{0}, a) = (1 + \gamma \text{Sgn}(-a))a^2/2.$$

Therefore, $a \mapsto R_\gamma(\mathbf{0}, a)$ is a continuous and strictly convex function. Then we conclude that there is a unique global minimizer given by $a^* = 0$. Now we take $S \neq \mathbf{0}$ and decompose $R_\gamma(S, a)$ as follows

$$R_\gamma(S, a) = -\mathbb{E} \left[\int_0^1 a Z_\theta^a \ell''_\gamma(Z_\theta^a) d\theta \right] + \frac{1}{2} \mathbb{E} \left[\int_0^1 \ell''_\gamma(Z_\theta^a) \|SB_\theta\|^2 d\theta \right]. \quad (2.11)$$

By replacing the expression of ℓ''_γ (see Equation (2.2)), we get

$$a \mapsto a Z_\theta^a \ell''_\gamma(Z_\theta^a) = (1 + \gamma^2) a Z_\theta^a + 2\gamma a |Z_\theta^a|,$$

which is continuous $d\mathbb{P}^B \otimes d\theta$ -almost surely and bounded by $(1 + \gamma)^2 |a| |Z_\theta^a|$ (integrable w.r.t. $d\mathbb{P}^B \otimes d\theta$ locally uniformly in a). By the dominated convergence theorem, we conclude the first term of the decomposition in (2.11) is continuous in a . Also, we estimate $|\ell_\gamma''(Z_\theta^a) \|SB_\theta\|^2| \leq (1 + \gamma^2) \|SB_\theta\|^2$, which is integrable uniformly in a . Following that $B_\theta^\top SB_\theta$ has a density w.r.t. the Lebesgue measure (see the proof of Proposition 5.3 in Appendix), we get $Z_\theta^a \neq 0$, $d\mathbb{P}^B \otimes d\theta$ -almost surely. It holds that

$$a \mapsto \ell_\gamma''(Z_\theta^a) \|SB_\theta\|^2$$

is continuous $d\mathbb{P}^B \otimes d\theta$ -almost surely, due to the continuity of ℓ_γ'' on $\mathbb{R}^*$. Now, we conclude the second term of the decomposition in (2.11) is also continuous in a , by applying again the dominated convergence theorem. Therefore, we have proved that $R_\gamma(S, a)$ is continuous in a .

Let $\alpha \in \mathbb{R}$ large enough such that $K = \{a : R_\gamma(S, a) \leq \alpha\}$ is non-empty. Due to the continuity and coercivity of $R_\gamma(S, a)$, K is compact. Then, by Weierstrass's Theorem, we conclude the announced result. $\square$

Here, we have just shown the existence of a minimizer a^* to the Problem (2.9) for a given symmetric matrix S . The regularity of $a^*(S)$ has not been analysed, because the uniqueness has not been proved. In fact, the uniqueness and smoothness of f^* of the problem (2.8) is challenging in the general case. Certainly, if $a^*(S)$ is unique, then we could define f^* as in (2.10). Then, a natural candidate for the self-consistent valuation/hedging rule is given by the solution to nonlinear PDE

$$\begin{cases} \partial_t v^*(t, x) + \frac{1}{2} \text{Tr}[\sigma \sigma^\top D_x^2 v^*](t, x) + a^*(\sigma^\top(t, x) D_x^2 v^*(t, x) \sigma(t, x)) = 0, \\ v^*(T, x) = h(x). \end{cases} \quad (2.12)$$

This PDE is fully nonlinear with a nonlinear term depending on the Hessian. Unfortunately, in full generality, we are not able to prove the existence/uniqueness of a solution v^* satisfying Assumption 2. Also proving that the new kernel f^* fulfills Assumption 3 is not straightforward. Fortunately, the one-dimensional case provides us a quasi-explicit formulation for a^* , which hopefully is a first step in the analysis of the PDE (2.12). Further investigation is left to future research.

2.4 Quasi-explicit solution in the one-dimensional case

In this subsection, we present a quasi-explicit formulation of the optimal kernel f^* in the one-dimensional case. Here, $(B_\theta^\top SB_\theta - \text{Tr}[S]\theta)/2$ becomes $(B_\theta^2 - \theta)y/2$ for $y = S \in \mathbb{R}$. So, we rewrite the function $R_\gamma(S, a)$ given by (2.7) as

$$R_\gamma(y, a) = \mathbb{E} \left[\int_0^1 \ell_\gamma'' \left(y (B_\theta^2 - \theta) / 2 - a\theta \right) \left(a^2 \theta - ay (B_\theta^2 - \theta) / 2 + y^2 B_\theta^2 / 2 \right) d\theta \right].$$

Let $a^* \in \mathbb{R}$ a global minimizer of $\min_{a \in \mathbb{R}} R_\gamma(y, a)$. In the following proposition, we sum up some interesting properties of a^* . We denote by $\Phi_{\mathcal{N}}$ the cumulative distribution function (CDF) of the standard normal distribution and $\phi_{\mathcal{N}} = \Phi_{\mathcal{N}}'$ its density.

Proposition 2.3. *Let $\gamma \in [0, 1)$.*

- (a) *Let $c_1^* \in \mathbb{R}$ and $c_2^* \in \mathbb{R}$ be global minimizers of*

$$\min_{c \in \mathbb{R}} R_\gamma(1, c) \text{ and } \min_{c \in \mathbb{R}} R_\gamma(-1, c),$$

respectively. Then $a^*(y) = c_1^* y \mathbf{I}_{y>0} + c_2^* y \mathbf{I}_{y<0}$ is a global minimizer of

$$\min_{a \in \mathbb{R}} R_\gamma(y, a).$$

(b) The mapping

$$c \mapsto R_\gamma(1, c) \text{ and } c \mapsto R_\gamma(-1, c)$$

are strictly convex. Thus, c_1^* and c_2^* are unique characterized by

$$(1 + \gamma^2) c_1^* + \gamma T(c_1^*) = 0 \text{ and } (1 + \gamma^2) c_2^* - \gamma T(c_2^*) = 0,$$

respectively, where

$$T(c) = 2c \mathbf{I}_{2c+1 \leq 0} + \left(8c \Phi_{\mathcal{N}}(-\sqrt{2c+1}) - 4\phi_{\mathcal{N}}(\sqrt{2c+1}) \sqrt{2c+1} - 2c \right) \mathbf{I}_{2c+1 > 0}.$$

Therefore, the minimizer $a^*(y)$ is unique.

Proof. (a) We start by the special case $y = 0$, we get $R_\gamma(0, a) = (1 + \gamma \text{Sgn}(-a))^2 a^2 / 2$. So, $a^*(0) = 0$. Now we consider the more interesting case $y \neq 0$. By setting $c = a/y$, we rewrite $R_\gamma(y, a)$

$$\begin{aligned} R_\gamma(y, cy) &= \mathbb{E} \left[\int_0^1 \ell''_\gamma \left((B_\theta^2 - \theta)/2 - c\theta \right) \left(c^2\theta - c(B_\theta^2 - \theta)/2 + B_\theta^2/2 \right) d\theta \right] y^2 \mathbf{I}_{y>0} \\ &\quad + \mathbb{E} \left[\int_0^1 \ell''_\gamma \left(-(B_\theta^2 - \theta)/2 + c\theta \right) \left(c^2\theta - c(B_\theta^2 - \theta)/2 + B_\theta^2/2 \right) d\theta \right] y^2 \mathbf{I}_{y<0} \end{aligned} \quad (2.13)$$

because $\ell''_\gamma(y\zeta) = \ell''_\gamma(\zeta)$ if $y > 0$ and $\ell''_\gamma(y\zeta) = \ell''_\gamma(-\zeta)$ if $y < 0$, for any $\zeta \in \mathbb{R}$.

Consider a global minimizer $c^*(y)$ of $\min_{c \in \mathbb{R}} R_\gamma(y, cy)$, then $a^*(y) = c^*(y)y$ is also a global minimizer of $\min_{a \in \mathbb{R}} R_\gamma(y, a)$. Because $(y, c) \mapsto R_\gamma(y, cy)$ is multiplicatively separable on $y > 0$ and on $y < 0$, we write $c^*(y) = c_1^* \mathbf{I}_{y>0} + c_2^* \mathbf{I}_{y<0}$, where c_1^* and c_2^* are respectively global minimizers of $\min_{c \in \mathbb{R}} R_\gamma(1, c)$ and $\min_{c \in \mathbb{R}} R_\gamma(-1, c)$.

(b) Let G be a standard normal random variable. It will be useful later to know $\mathbb{E}[G^2 \mathbf{I}_{G < \alpha}]$ for any real α : we have

$$\begin{aligned} \mathbb{E}[G^2 \mathbf{I}_{G < \alpha}] &= -\alpha \phi_{\mathcal{N}}(\alpha) + \Phi_{\mathcal{N}}(\alpha), \\ \mathbb{E}[G^2 \mathbf{I}_{G > \alpha}] &= \alpha \phi_{\mathcal{N}}(-\alpha) + \Phi_{\mathcal{N}}(-\alpha), \\ \mathbb{E}[G^2 \mathbf{I}_{-\alpha < G < \alpha}] &= -2\alpha \phi_{\mathcal{N}}(-\alpha) + (\Phi_{\mathcal{N}}(\alpha) - \Phi_{\mathcal{N}}(-\alpha)). \end{aligned} \quad (2.14)$$

It holds that $B_\theta \sim \sqrt{\theta}G$ for all θ in $[0, 1]$. From (2.13), we get

$$R_\gamma(1, c) = \frac{1 + \gamma^2}{2} T_1(c) + \gamma T_2(c), \quad R_\gamma(-1, c) = \frac{1 + \gamma^2}{2} T_1(c) - \gamma T_2(c),$$

where

$$\begin{aligned} T_1(c) &= \mathbb{E} [c^2 - c(G^2 - 1)/2 + G^2/2] = c^2 + 1/2, \\ T_2(c) &= \mathbb{E} [\text{Sgn}((G^2 - 1)/2 - c) (c^2 - c(G^2 - 1)/2 + G^2/2)]. \end{aligned}$$

Considering $\alpha(c) = \sqrt{2c+1}$, it holds

$$\text{Sgn}((G^2 - 1)/2 - c) = \mathbf{I}_{2c+1 < 0} + \mathbf{I}_{2c+1 > 0} (\mathbf{I}_{G < -\alpha(c)} + \mathbf{I}_{G > \alpha(c)} - \mathbf{I}_{-\alpha(c) < G < \alpha(c)}).$$

From the expectations in (2.14), we deduce

$$\begin{aligned}
T_2(c) &= \mathbf{I}_{2c+1 < 0} (c^2 + 1/2) \\
&\quad + \mathbf{I}_{2c+1 > 0} (c^2 + c/2) \mathbb{E} [\mathbf{I}_{G < -\alpha(c)} + \mathbf{I}_{G > \alpha(c)} - \mathbf{I}_{-\alpha(c) < G < \alpha(c)}] \\
&\quad + \mathbf{I}_{2c+1 > 0} (1/2 - c/2) \mathbb{E} [G^2 \mathbf{I}_{G < -\alpha(c)} + G^2 \mathbf{I}_{G > \alpha(c)} - G^2 \mathbf{I}_{-\alpha(c) < G < \alpha(c)}] \\
&= \mathbf{I}_{2c+1 < 0} (c^2 + 1/2) + \mathbf{I}_{2c+1 > 0} \beta(c),
\end{aligned}$$

where

$$\beta(c) = (c^2 + 1/2) (3 - 4\Phi_{\mathcal{N}}(\alpha(c))) + 2(1 - c) \alpha(c) \phi_{\mathcal{N}}(\alpha(c)).$$

We easily check that $R_\gamma(1, c)$ and $R_\gamma(-1, c)$ are C^0 and piecewise C^2 . Let us compute their first derivatives for $c < -1/2$ and $c > -1/2$

$$\begin{aligned}
\partial_c R_\gamma(1, c) &= (1 + \gamma^2) c + \gamma (\mathbf{I}_{2c+1 < 0} 2c + \mathbf{I}_{2c+1 > 0} \beta'(c)) \\
&= \mathbf{I}_{2c+1 < 0} (1 + \gamma^2)^2 c + ((1 + \gamma^2) c + \gamma \beta'(c)) \mathbf{I}_{2c+1 > 0}, \\
\partial_c R_\gamma(-1, c) &= (1 + \gamma^2) c - \gamma (\mathbf{I}_{2c+1 < 0} 2c + \mathbf{I}_{2c+1 > 0} \beta'(c)) \\
&= \mathbf{I}_{2c+1 < 0} (1 - \gamma^2)^2 c + ((1 + \gamma^2) c - \gamma \beta'(c)) \mathbf{I}_{2c+1 > 0},
\end{aligned}$$

where

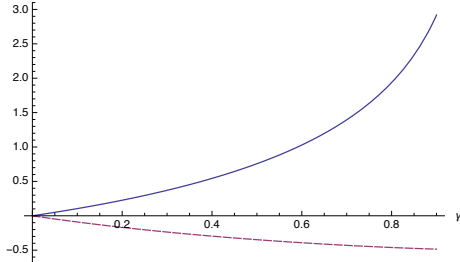
$$\beta'(c) = 8c \Phi_{\mathcal{N}}(-\sqrt{2c+1}) - 4\phi_{\mathcal{N}}(\sqrt{2c+1})\sqrt{2c+1} - 2c.$$

Standard computations show that $\partial_c R_\gamma(-1, c)$ and $\partial_c R_\gamma(1, c)$ are continuous at $c = -1/2$. Moreover, we see that $\partial_c R_\gamma(1, c)$ and $\partial_c R_\gamma(-1, c)$ are strictly increasing on c under the condition that $|\beta''(c)| \leq 2$ on $2c+1 > 0$. Indeed, we have

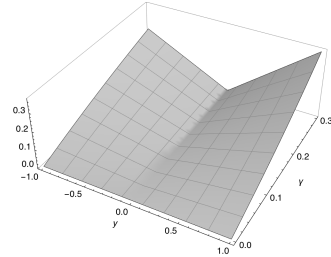
$$\beta''(c) = 6 - 8\Phi_{\mathcal{N}}(\sqrt{2c+1}) \in [-2, 2].$$

due to $\Phi_{\mathcal{N}}(\sqrt{2c+1}) \in [1/2, 1]$ for all $2c+1 > 0$. Because $R_\gamma(1, c)$, $R_\gamma(-1, c)$ are strictly convex, the optimal values c_1^* and c_2^* are unique and characterized respectively by $\partial_c R_\gamma(1, c_1^*) = 0$ and $\partial_c R_\gamma(-1, c_2^*) = 0$. $\square$

We depict the global minimizer a^* in Figure 3. We show the approximate values of c_1^* and c_2^* calculated by a root finding algorithm in Table 1.



(a) Plot of $c_1^*(\gamma)$ (blue line) and $c_2^*(\gamma)$ (purple dashed) on $[0.0, 0.9]$.



(b) Plot of $a^*(\gamma, y)$ on $[0, 0.3] \times [-1, 1]$.

Figure 3: Global minimizer $a^*(y)$.

Therefore, in the spirit of Equation (2.10), we set

$$f^*(t, x, y, z, A) = f_\gamma^*(\sigma^2(t, x) A)$$

with f_γ^* denoting the optimal kernel in dimension 1:

$$f_\gamma^*(y) := a^*(y) = c_1^* y \mathbf{I}_{y > 0} + c_2^* y \mathbf{I}_{y < 0}. \quad (2.15)$$

γ	c_1^*	c_2^*
0.1	0.1043	-0.09013
0.2	0.2262	-0.1684
0.3	0.3702	-0.2366

Table 1: Optimal slopes c_1^* and c_2^* .

3 Proof of Theorem 2.1

The proof is long and technical. For this reason, we split it into different stages.

- First, we study the conditional local risk $\mathcal{R}_n(\gamma)$ on the interval $[t_n, t_{n+1}]$, by using a time-space rescaling argument (see Subsection 3.1). This rescaling turns out to be essential to pass to the limit later.
- Second, we derive an explicit approximation of the conditional local risk $\mathcal{R}_n(\gamma)$ (see Subsection 3.2).
- Finally, we prove that the remainder terms converge almost surely towards 0. For this, we show that the Greeks of $u_\tau(t, \cdot)$ converge to those of $v(\tau, \cdot)$ as $t \uparrow \tau$ (see Subsection 3.3). Also, we show that the set of discontinuity points of ℓ''_γ has measure zero under the Assumption A5.

In the proof, we use several constants $C_{n,N}(\xi)$ depending polynomially on the space variable ξ (uniformly in the interval $[t_n, t_{n+1}]$ and in the number of time steps). To simplify, we note $C_{n,N}(\xi) \in C_{\text{pol}}$ if for some real $q > 0$,

$$\sup_{N \in \mathbb{N}^*} \sup_{0 \leq n \leq N-1} \sup_{\xi \in \mathbb{R}^d} \frac{|C_{n,N}(\xi)|}{1 + \|\xi\|^q} < +\infty.$$

This upper bound depends on the polynomial bounds on the functions μ , σ , f , v and u .

3.1 Preliminary time-space rescaling and conditioning

First, we start by a few observations.

- Thanks to the Markov property of the SDE and in view of our smoothness assumptions, $\mathcal{R}_n(\gamma)$ is a continuous function of t_n and X_{t_n} only (see Equation (2.4));
- $\mathcal{R}_n(\gamma)$ goes to zero at rate ε_N^2 , because we prove that the remainder of a second-order stochastic Taylor expansion will be inside ℓ_γ . Rescaling it by ε_N , we expect to get a non-zero limit for the aggregated value of $\mathcal{R}_n(\gamma)$ (see Equation (2.5));
- Note that ℓ''_γ has a jump discontinuity at zero (see Equation (2.2)). Then to decompose the conditional local risk, we will need to apply a stronger version of Ito's formula, known as, the Ito-Tanaka formula.

In view of a Taylor-Ito expansion, we consider the process $X^{\varepsilon_N} = \{X_\theta^{\varepsilon_N}, \theta \in [0, 1]\}$ satisfying

$$dX_\theta^{\varepsilon_N} = \varepsilon_N \mu(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) d\theta + \varepsilon_N^{1/2} \sigma(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) dB_\theta, \quad X_0^{\varepsilon_N} = \xi \in \mathbb{R}^d, \quad (3.1)$$

where B is an extra Brownian motion independent from W . It is a time-space rescaling of the original process starting from ξ at t_n .

By denoting $X^{t,\xi}$ as the SDE solution starting from ξ at t , we notice that the processes $\{X_{t_n+\theta\varepsilon_N}^{t_n,\xi}, \theta \in [0, 1]\}$ and $\{X_\theta^{\varepsilon_N}, \theta \in [0, 1]\}$ have the same distribution. This is due to the fact both processes satisfy the same SDE generated by Brownian motions both independent from $\mathcal{F}_{t_n}$. Then we can rewrite $\mathcal{R}_n(\gamma)$ (see (2.4)) as a continuous function in terms of X_{t_n} and $X_\theta^{\varepsilon_N}$. Setting

$$T^{\varepsilon_N}(t_n, \xi) = \varepsilon_N^{-2} \mathbb{E}^B \left[\ell_\gamma \left(u^{(n+1)}(t_{n+1}, X_1^{\varepsilon_N}) - u^{(n+1)}(t_n, \xi) - D_x u^{(n+1)}(t_n, \xi) (X_1^{\varepsilon_N} - \xi) \right) \right],$$

leads to

$$\mathcal{R}_n(\gamma) = \varepsilon_N^2 T^{\varepsilon_N}(t_n, X_{t_n}). \quad (3.2)$$

3.2 Stochastic expansion of the conditional local risk

Proposition 3.1 (Stochastic expansion of the conditional local risk at the time t_n). Assume notations and assumptions of Theorem 2.1. Denote $F^{(n+1)} : [0, t_{n+1}] \times \mathbb{R}^d \rightarrow \mathbb{R}$ and $G^{(n+1)} : [0, t_{n+1}] \times \mathbb{R}^d \rightarrow \mathcal{S}^d$ as

$$\begin{aligned} F^{(n+1)}(t, \cdot) &= f(t, \cdot, u^{(n+1)}(t, \cdot), D_x u^{(n+1)}(t, \cdot), D_x^2 u^{(n+1)}(t, \cdot)); \\ G^{(n+1)}(t, \cdot) &= (\sigma^\top (D_x^2 u^{(n+1)}) \sigma)(t, \cdot). \end{aligned} \quad (3.3)$$

For any t_n and $\xi \in \mathbb{R}^d$, let $X_\theta^{\varepsilon_N} : [0, 1] \times \Omega \rightarrow \mathbb{R}^d$ be the strong solution to the SDE (3.1) such that $X_0^{\varepsilon_N} = \xi$ and let $\mathcal{E}_\theta^{\varepsilon_N} : [0, 1] \times \Omega \rightarrow \mathbb{R}$ be the stochastic process defined by

$$\mathcal{E}_\theta^{\varepsilon_N} = u^{(n+1)}(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N}) - u^{(n+1)}(t_n, \xi) - \langle D_x u^{(n+1)}(t_n, \xi), X_\theta^{\varepsilon_N} - \xi \rangle \quad (3.4)$$

so that

$$T^{\varepsilon_N}(t_n, \xi) = \varepsilon_N^{-2} \mathbb{E}^B [\ell_\gamma(\mathcal{E}_1^{\varepsilon_N})]. \quad (3.5)$$

The following local risk decomposition holds

$$\begin{aligned} T^{\varepsilon_N}(t_n, \xi) &= \mathbb{E}^B \left[\int_0^1 \ell_\gamma'' \left(E_\theta(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi)) + R_\theta^{\varepsilon_N}(t_n, \xi) \right) \right. \\ &\quad \left. \times Q_\theta(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi)) d\theta \right] + C_{n,N}(\xi) \varepsilon_N^{1/2}, \end{aligned}$$

where

$$E_\theta(S, y) = \int_0^\theta B_{\theta'}^\top S dB_{\theta'} - y\theta, \quad (3.6)$$

$$Q_\theta(S, y) = y^2\theta - y \int_0^\theta B_{\theta'}^\top S dB_{\theta'} + \|SB_\theta\|^2/2, \quad (3.7)$$

$$R_\theta^{\varepsilon_N}(t_n, \xi) = \mathcal{E}_\theta^{\varepsilon_N}/\varepsilon_N - E_\theta(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi)), \quad (3.8)$$

for some constant $C_{n,N}(\xi) \in C_{pol}$.

The proof of Proposition 3.1 is delicate. We postpone it to Subsection 3.5. In order to perform a second - order stochastic expansion, we need that $u^{(n+1)}$ and $D_x u^{(n+1)}$ be in $C^{1,2}$ to apply Ito's formula. Additionally, we require σ , $D_x u^{(n+1)}$, $D_x^2 u^{(n+1)}$, $\partial_t D_x u^{(n+1)}$ and $D_x^2 D_x u^{(n+1)}$ to have polynomial growth to obtain proper integrability along the computations. Finally, we ask for σ and $D_x^2 u^{(n+1)}$ to be in $H_{loc, pol}^{1/2, 1}$, which is useful at the stochastic expansion of the gradient $D_x u^{(n+1)}$. All the above conditions are satisfied thanks to our assumptions.

3.3 Approximation of sensitivities in small time

First, notice that the above expansion of $T^{\varepsilon_N}(t_n, \xi)$ depends on $u^{(n+1)}$, solution of the PDE (1.5) on the subinterval $[t_n, t_{n+1}]$, whose size goes to 0. Therefore, by invoking a small - time approximation argument, we replace $u^{(n+1)}$ and its first - second derivatives by its terminal value $v(t_{n+1}, \cdot)$ and its first - second derivatives. Notice that the reference valuation v is independent of ε_N . This is the matter of following statement, proved in Appendix 5.2.

Proposition 3.2 (Approximation of sensitivities in small time). *Assume notations and assumptions of Theorem 2.1. Then, there exists some constant $C_{n,N}(\xi) \in C_{pol}$ such that*

$$\left| u^{(n+1)}(t_n, \xi) - v(t_{n+1}, \xi) \right| \leq C_{n,N}(\xi) \varepsilon_N^{1/2}, \quad (3.9)$$

$$\left\| D_x u^{(n+1)}(t_n, \xi) - D_x v(t_{n+1}, \xi) \right\| \leq C_{n,N}(\xi) \varepsilon_N^{1/2}, \quad (3.10)$$

$$\left\| D_x^2 u^{(n+1)}(t_n, \xi) - D_x^2 v(t_{n+1}, \xi) \right\| \leq C_{n,N}(\xi) \varepsilon_N^{1/2}. \quad (3.11)$$

3.4 Aggregation of local risk and passage to the limit

We set

$$\begin{aligned} F(t, \xi) &= f(t, \xi, v(t, \xi), D_x v(t, \xi), (D_x^2 v)(t, \xi)) \in \mathbb{R}, \\ G(t, \xi) &= (\sigma^\top (D_x^2 v) \sigma)(t, \xi) \in \mathcal{S}^d. \end{aligned} \quad (3.12)$$

Replacing ξ by X_{t_n} in the expansion of $T^{\varepsilon_N}(t_n, \xi)$ in Proposition 3.1 leads to

$$\begin{aligned} T^{\varepsilon_N}(t_n, X_{t_n}) &= \mathbb{E}^B \left[\int_0^1 \ell''_\gamma \left(E_\theta \left(G^{(n+1)}(t_n, X_{t_n}), F^{(n+1)}(t_n, X_{t_n}) \right) + R_\theta^{\varepsilon_N}(t_n, X_{t_n}) \right) \right. \\ &\quad \left. \times Q_\theta \left(G^{(n+1)}(t_n, X_{t_n}), F^{(n+1)}(t_n, X_{t_n}) \right) d\theta \right] + C_{n,N}(X_{t_n}) \varepsilon_N^{1/2}, \end{aligned}$$

where $C_{n,N}(X_{t_n}) \in C_{pol}$. By substituting $u^{(n+1)}(t_n, \cdot)$ by its terminal value $v(t_{n+1}, \cdot)$ in $F^{(n+1)}(t_n, \cdot)$ and $G^{(n+1)}(t_n, \cdot)$ (see (3.3)), we get $F(t_{n+1}, \cdot)$ and $G(t_{n+1}, \cdot)$ (see (3.12)). Hence,

$$\begin{aligned} T^{\varepsilon_N}(t_n, X_{t_n}) &= \mathbb{E}^B \left[\int_0^1 \ell''_\gamma \left(E_\theta (G(t_{n+1}, X_{t_n}), F(t_{n+1}, X_{t_n})) + \bar{R}_\theta^{\varepsilon_N}(t_n, X_{t_n}) \right) \right. \\ &\quad \left. \times Q_\theta (G(t_{n+1}, X_{t_n}), F(t_{n+1}, X_{t_n})) d\theta \right] + \bar{C}^{\varepsilon_N}(t_n, X_{t_n}) + C_{n,N}(X_{t_n}) \varepsilon_N^{1/2}, \end{aligned}$$

where

$$\bar{R}_\theta^{\varepsilon_N}(t_n, \xi) := E_\theta(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi)) - E_\theta(G(t_{n+1}, \xi), F(t_{n+1}, \xi)) + R_\theta^{\varepsilon_N}(t_n, \xi), \quad (3.13)$$

$$\begin{aligned} \bar{C}^{\varepsilon_N}(t_n, \xi) &:= \mathbb{E}^B \left[\int_0^1 \ell''_\gamma \left(E_\theta(G(t_{n+1}, \xi), F(t_{n+1}, \xi)) + \bar{R}_\theta^{\varepsilon_N}(t_n, \xi) \right) \right. \\ &\quad \left. \times \left(Q_\theta(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi)) - Q_\theta(G(t_{n+1}, \xi), F(t_{n+1}, \xi)) \right) d\theta \right]. \end{aligned} \quad (3.14)$$

In the sequel, we require estimates of $\bar{R}_\theta^{\varepsilon_N}(t_n, X_{t_n})$ and $\bar{C}^{\varepsilon_N}(t_n, X_{t_n})$, summarized in the following lemma, proved later in Subsection 3.6.

Lemma 3.3. *Under the assumptions of Theorem 2.1, for any $p \geq 1$, there exists a constant K_p such that*

- (a) $\mathbb{E} \left[\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} \left| E_\theta(G^{(n+1)}(t_n, X_{t_n}), F^{(n+1)}(t_n, X_{t_n})) - E_\theta(G(t_{n+1}, X_{t_n}), F(t_{n+1}, X_{t_n})) \right|^p \right] \leq K_p \varepsilon_N^{p/2};$
- (b) $\sup_{0 \leq n \leq N-1} \mathbb{E} |\bar{C}^{\varepsilon_N}(t_n, X_{t_n})| \leq K_1 \varepsilon_N^{1/2};$
- (c) $\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} |\bar{R}_\theta^{\varepsilon_N}(t_n, X_{t_n})| \xrightarrow[N \rightarrow \infty]{} 0, \text{ d}\mathbb{P}^W \otimes \text{d}\mathbb{P}^B - \text{a.s.}$

From the definition of T^{ε_N} in (3.2), we have $\varepsilon_N^{-1} \mathcal{R}_n(\gamma) = T^{\varepsilon_N}(t_n, X_{t_n}) \varepsilon_N$. By summing it for $0 \leq n \leq N-1$, we obtain

$$\begin{aligned} \varepsilon_N^{-1} \mathbb{E} \left[\sum_{n=0}^{N-1} \mathcal{R}_n(\gamma) \right] &= \mathbb{E} \left[\sum_{n=0}^{N-1} T^{\varepsilon_N}(t_n, X_{t_n}) \varepsilon_N \right] = \mathbb{E} \left[\int_0^T T^{\varepsilon_N}(\varphi_t^N, X_{\varphi_t^N}) dt \right] \\ &= \mathbb{E}^{W \otimes B} \left[\int_0^T \int_0^1 \ell''_\gamma \left(E_\theta \left(G(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) + \bar{R}_\theta^{\varepsilon_N}(\varphi_t^N, X_{\varphi_t^N}) \right) \right. \\ &\quad \left. \times Q_\theta \left(G(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) d\theta dt \right] + \sum_{n=0}^{N-1} \mathbb{E} \left[\bar{C}^{\varepsilon_N}(t_n, X_{t_n}) \varepsilon_N + C_{n,N}(X_{t_n}) \varepsilon_N^{3/2} \right]. \end{aligned} \quad (3.15)$$

The last sum goes to 0 as $N \rightarrow +\infty$, owing to Lemma 3.3 and Proposition 3.1. It remains to determine the limit of the first term in (3.15). We achieve this result by applying the dominated convergence theorem.

1. Because of $\sigma, v, D_x v, D_x^2 v, f \in H_{\text{loc, pol}}^{1/2,1}$ (therefore, they are continuous in time and space) and the path - continuity of X , we get $\text{d}\mathbb{P}^W$ - a.s. for any t

$$\begin{aligned} &(\sigma^\top(D_x^2 v)\sigma)(\bar{\varphi}_t^N, X_{\varphi_t^N}) \xrightarrow[N \rightarrow \infty]{} (\sigma^\top(D_x^2 v)\sigma)(t, X_t), \\ &f(\bar{\varphi}_t^N, X_{\varphi_t^N}, v(\bar{\varphi}_t^N, X_{\varphi_t^N}), D_x v(\bar{\varphi}_t^N, X_{\varphi_t^N}), D_x^2 v(\bar{\varphi}_t^N, X_{\varphi_t^N})) \\ &\xrightarrow[N \rightarrow \infty]{} f(t, X_t, v(t, X_t), D_x v(t, X_t), D_x^2 v(t, X_t)). \end{aligned}$$

Hence, it holds $\text{d}\mathbb{P}^W \otimes \text{d}\mathbb{P}^B$ - a.s. for any θ, t

$$\begin{aligned} E_\theta \left(G(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) &\xrightarrow[N \rightarrow \infty]{} E_\theta(\Gamma(t, X_t), F(t, X_t)), \\ Q_\theta \left(G(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) &\xrightarrow[N \rightarrow \infty]{} Q_\theta(G(t, X_t), F(t, X_t)), \end{aligned}$$

because E_θ and Q_θ (see 3.6 - 3.7) are continuous in S, y , $\text{d}\mathbb{P}^B \otimes \text{d}\theta$ - a.s. Also, from the item (c) of Lemma 3.3, we have

$$\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} |\bar{R}_\theta^{\varepsilon_N}(t_n, X_{t_n})| \xrightarrow[N \rightarrow \infty]{} 0,$$

$\text{d}\mathbb{P}^W \otimes \text{d}\mathbb{P}^B$ - almost surely.

2. Seeing that the second derivative ℓ''_γ is discontinuous at 0 and the set

$$\mathcal{A} := \{(\omega, t, \theta) \in \Omega \times [0, T] \times [0, 1] : E_\theta(G(t, X_t(\omega)), F(t, X_t(\omega))) (\omega) = 0\} \quad (3.16)$$

has measure zero (see Proposition 5.3 in Appendix), it holds

$$\begin{aligned} \ell''_\gamma \left(E_\theta \left(G(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) + \bar{R}_\theta^{\varepsilon_N}(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) \\ \xrightarrow{N \rightarrow \infty} \ell''_\gamma (E_\theta(G(t, X_t), F(t, X_t))), \end{aligned}$$

$d\mathbb{P}^W \otimes d\mathbb{P}^B \otimes dt \otimes d\theta$ - almost surely.

3. Because of the boundedness of ℓ''_γ and the polynomial growth of $\sigma, v, D_x v, D_x^2 v$, we have

$$\begin{aligned} |\ell''_\gamma \left(E_\theta \left(\Gamma(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) + \bar{R}_\theta^{\varepsilon_N}(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) Q_\theta \left(G(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right)| \\ \leq C \left(1 + \sup_{t \in [0, T]} |X_t| + \|B_\theta\| + \left\| \int_0^\theta B_\theta dB_\theta^\top \right\| \right)^q \end{aligned}$$

for some positive constants C and q .

By the dominated convergence theorem, we conclude

$$\begin{aligned} \mathbb{E}^{W \otimes B} \left[\int_0^T \int_0^1 \ell''_\gamma \left(E_\theta \left(\Gamma(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) + \bar{R}_\theta^{\varepsilon_N}(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) \right. \\ \left. \times Q_\theta \left(\Gamma(\bar{\varphi}_t^N, X_{\varphi_t^N}), F(\bar{\varphi}_t^N, X_{\varphi_t^N}) \right) d\theta dt \right] \\ \xrightarrow{N \rightarrow \infty} \mathbb{E}^{W \otimes B} \left[\int_0^T \int_0^1 \ell''_\gamma (E_\theta(\Gamma(t, X_t), F(t, X_t))) Q_\theta(\Gamma(t, X_t), F(t, X_t)) d\theta dt \right]. \end{aligned}$$

This completes the proof of Theorem 2.1.

3.5 Proof of Proposition 3.1

For the sake of conciseness, we set $u = u^{(n+1)}$. By substituting $X_\theta^{\varepsilon_N}$ in (3.1) into $\mathcal{E}_\theta^{\varepsilon_N}$ in (3.4), we get

$$\begin{aligned} \mathcal{E}_\theta^{\varepsilon_N} &= u(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) - u(t_n, \xi) - \varepsilon_N \int_0^\theta D_x u(t_n, \xi) \mu(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) d\theta' \\ &\quad - \varepsilon_N^{1/2} \int_0^\theta D_x u(t_n, \xi) \sigma(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) dB_{\theta'}, \end{aligned} \quad (3.17)$$

where $D_x u(\cdot, \cdot)$ is a row vector.

In the proof, we use the Ito - Tanaka formula to $\ell_\gamma(\mathcal{E}_\theta^{\varepsilon_N})$ between $\theta = 0$ and $\theta = 1$ and we perform some Taylor - Ito expansions in terms of ε_N . Because $u, \partial_t u, D_x u, D_x^2 u \in H_{\text{loc, pol}}^{1/2, 1}$, then $u \in C^{1,2}([t_n, t_{n+1}] \times \mathbb{R}^d, \mathbb{R})$. Applying Ito's formula to $u(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N})$ yields

$$\begin{aligned} u(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) - u(t_n, \xi) &= \varepsilon_N^{1/2} \int_0^\theta (D_x u \sigma)(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) dB_{\theta'} \\ &\quad + \varepsilon_N \int_0^\theta (\mathcal{L}_{t_n + \theta' \varepsilon_N} u)(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) d\theta' \end{aligned}$$

$$+ \varepsilon_N \int_0^\theta ((D_x u) \mu) (t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) d\theta'. \quad (3.18)$$

Here, we denote $(\Delta D_x u)(t, \zeta) = D_x u(t, \zeta) - D_x u(t_n, \xi)$ for any $t \in [t_n, t_{n+1}]$ and $\zeta \in \mathbb{R}^d$. Replacing (3.18) in (3.17) leads to

$$\begin{aligned} \mathcal{E}_\theta^{\varepsilon_N} &= \varepsilon_N^{1/2} \int_0^\theta ((\Delta D_x u) \sigma) (t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) dB_{\theta'} \\ &\quad + \varepsilon_N \int_0^\theta (\mathcal{L}_{t_n + \theta' \varepsilon_N} u) (t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) d\theta' \\ &\quad + \varepsilon_N \int_0^\theta ((\Delta D_x u) \mu) (t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) d\theta'. \end{aligned} \quad (3.19)$$

Use that u solves the PDE (1.5) to simplify the second term above. Then, we apply the Ito - Tanaka formula to the convex function ℓ_γ (see Theorem 1.5 and Corollary 1.6 in [18, Chapter VI]) composed with the process $\mathcal{E}_\theta^{\varepsilon_N}$ between $\theta = 0$ and $\theta = 1$. Because $\ell'_\gamma(y) = \ell''_\gamma(y)y$ for all $y \in \mathbb{R}$, we get

$$\begin{aligned} \ell_\gamma(\mathcal{E}_1^{\varepsilon_N}) &= -\varepsilon_N \int_0^1 \ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) \mathcal{E}_\theta^{\varepsilon_N} F^{(n+1)}(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) d\theta \\ &\quad + \varepsilon_N^{1/2} \int_0^1 \ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) \mathcal{E}_\theta^{\varepsilon_N} ((\Delta D_x u) \sigma) (t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) dB_\theta \\ &\quad + \varepsilon_N \int_0^1 \ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) \mathcal{E}_\theta^{\varepsilon_N} ((\Delta D_x u) \mu) (t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) d\theta \\ &\quad + \frac{1}{2} \varepsilon_N \int_0^1 \ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) \|((\Delta D_x u) \sigma) (t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N})\|^2 d\theta. \end{aligned}$$

Considering $T^{\varepsilon_N}(t_n, \xi)$ in (3.5), taking the expectation of the above expression and dividing by ε_N^2 gives

$$T^{\varepsilon_N}(t_n, \xi) =: T_1^{\varepsilon_N}(t_n, \xi) + T_2^{\varepsilon_N}(t_n, \xi) + T_3^{\varepsilon_N}(t_n, \xi), \quad (3.20)$$

where

$$T_1^{\varepsilon_N}(t_n, \xi) := -\varepsilon_N^{-1} \mathbb{E}^B \left[\int_0^1 \ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) \mathcal{E}_\theta^{\varepsilon_N} F^{(n+1)}(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) d\theta \right], \quad (3.21)$$

$$T_2^{\varepsilon_N}(t_n, \xi) := \frac{1}{2} \varepsilon_N^{-1} \mathbb{E}^B \left[\int_0^1 \ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) \|((\Delta D_x u) \sigma) (t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N})\|^2 d\theta \right], \quad (3.22)$$

$$T_3^{\varepsilon_N}(t_n, \xi) := \varepsilon_N^{-1} \mathbb{E}^B \left[\int_0^1 \ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) \mathcal{E}_\theta^{\varepsilon_N} ((\Delta D_x u) \mu) (t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) d\theta \right]. \quad (3.23)$$

Here we have used that the stochastic integral in $\ell_\gamma(\mathcal{E}_1^{\varepsilon_N})$ has expectation zero, following directly from $\mathbb{E} \int_0^1 |\mathcal{E}_\theta^{\varepsilon_N}|^4 d\theta < +\infty$ and from the polynomial growth of σ and $D_x u$ (because $\sigma, D_x u \in H_{\text{loc}, \text{pol}}^{1/2, 1}$). Now we analyze the expansion of $\mathcal{E}_\theta^{\varepsilon_N}$ and then apply it to $T_i^{\varepsilon_N}(t_n, \xi)$ for $i = 1, 2, 3$.

1) Stochastic Taylor expansion of $((\Delta D_x u) \sigma) (t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N})$ and $\mathcal{E}_\theta^{\varepsilon_N}$. We approximate $\Delta D_x u$ up to order $\varepsilon_N^{1/2}$, by setting

$$(\Delta D_x u) (t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) = \varepsilon_N^{1/2} B_\theta^\top (\sigma^\top (D_x^2 u)) (t_n, \xi) + r_\theta^{\varepsilon_N}, \quad (3.24)$$

where $\Delta D_x u$ and $r_\theta^{\varepsilon_N}$ are row vectors.

Lemma 3.4. *Let $p \geq 2$. Under the assumptions of Theorem 2.1, it holds*

- (a) $\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} \mathbb{E}^B \|r_{\theta}^{\varepsilon_N}\|^p \leq C_{n,N}(\xi) \varepsilon_N^p,$
- (b) $\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} \mathbb{E}^B \|(\Delta D_x u)(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^p \leq C_{n,N}(\xi) \varepsilon_N^{p/2},$

for some constant $C_{n,N}(\xi) \in C_{pol}$.

Proof of Lemma 3.4. (b) This item follows directly from Lemma 5.1 and from item (a), using standard computations. (a) Because $D_x u, \partial_t D_x u, D_x^2 u, D_x^2 D_x u \in H_{loc, pol}^{1/2,1}$, then $D_x u \in C^{1,2}([t_n, t_{n+1}] \times \mathbb{R}^d)$. By applying Ito's formula to $D_x u(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})$, we get

$$\begin{aligned} r_{\theta}^{\varepsilon_N} &= \varepsilon_N \int_0^{\theta} (\mathcal{L}_{t_n + \theta' \varepsilon_N} D_x u + \mu^{\top}(D_x^2 u))(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) d\theta' \\ &\quad + \varepsilon_N^{1/2} \int_0^{\theta} dB_{\theta'}^{\top} ((\sigma^{\top}(D_x^2 u))(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) - (\sigma^{\top}(D_x^2 u))(t_n, \xi)). \end{aligned}$$

Owing to the Hölder inequality, the BDG inequality and the polynomial growth conditions on the functions (because $\sigma, D_x u, \partial_t D_x u, D_x^2 u, D_x^2 D_x u \in H_{loc, pol}^{1/2,1}$), we estimate

$$\begin{aligned} \mathbb{E}^B \|r_{\theta}^{\varepsilon_N}\|^p &\leq 2^{p-1} \varepsilon_N^p \int_0^{\theta} \mathbb{E}^B \|(\mathcal{L}_{t_n + \theta' \varepsilon_N} D_x u + \mu^{\top}(D_x^2 u))(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N})\|^p d\theta' \\ &\quad + 2^{p-1} C_{BDG} \varepsilon_N^{p/2} \int_0^{\theta} \mathbb{E}^B \|(\sigma^{\top}(D_x^2 u))(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) - (\sigma^{\top}(D_x^2 u))(t_n, \xi)\|^p d\theta'. \end{aligned}$$

Using the growth conditions from the assumptions and applying bounds (5.1) in Lemma 5.1 to $\sigma^{\top}(D_x^2 u) \in H_{loc, pol}^{1/2,1}$ (because σ and $D_x^2 u$ are in $H_{loc, pol}^{1/2,1}$), we obtain the announced estimate. $\square$

By replacing the decomposition (3.24) into the expression of $\mathcal{E}_{\theta}^{\varepsilon_N}$ given in (3.19), we obtain

$$\mathcal{E}_{\theta}^{\varepsilon_N} = \varepsilon_N E_{\theta} \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + \varepsilon_N R_{\theta}^{\varepsilon_N}(t_n, \xi), \quad (3.25)$$

where

$$E_{\theta} \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) = \int_0^{\theta} B_{\theta'}^{\top} G^{(n+1)}(t_n, \xi) dB_{\theta'} - F^{(n+1)}(t_n, \xi) \theta, \quad (3.26)$$

and

$$\begin{aligned} R_{\theta}^{\varepsilon_N}(t_n, \xi) &= - \int_0^{\theta} \left(F^{(n+1)}(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) - F^{(n+1)}(t_n, \xi) \right) d\theta' \\ &\quad + \int_0^{\theta} B_{\theta'}^{\top} (\sigma^{\top}(D_x^2 u))(t_n, \xi) (\sigma(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) - \sigma(t_n, \xi)) dB_{\theta'} \\ &\quad + \varepsilon_N^{-1/2} \int_0^{\theta} r_{\theta'}^{\varepsilon_N} \sigma(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) dB_{\theta'} \\ &\quad + \varepsilon_N^{1/2} \int_0^{\theta} B_{\theta'}^{\top} (\sigma^{\top}(D_x^2 u))(t_n, \xi) \mu(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) d\theta' \\ &\quad + \int_0^{\theta} r_{\theta'}^{\varepsilon_N} \mu(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) d\theta'. \end{aligned} \quad (3.27)$$

Lemma 3.5. *Under the assumptions of Theorem 2.1, it holds*

$$(a) \sup_{\theta \in [0,1]} \mathbb{E}^B \left| E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) \right|^2 \leq C_{n,N}(\xi);$$

$$(b) \sup_{\theta \in [0,1]} \mathbb{E}^B |R_\theta^{\varepsilon_N}(t_n, \xi)|^2 \leq C_{n,N}(\xi) \varepsilon_N,$$

for some constant $C_{n,N}(\xi) \in C_{pol}$.

Proof of Lemma 3.5. (a) From (3.26) we get

$$\begin{aligned} & \mathbb{E}^B \left| E_\theta \left((\sigma^\top (D_x^2 u) \sigma)(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) \right|^2 \\ & \leq 2 |F^{(n+1)}(t_n, \xi)|^2 \theta^2 + 2 \mathbb{E}^B \left| \int_0^\theta B_{\theta'}^\top ((\sigma^\top (D_x^2 u) \sigma))(t_n, \xi) d B_{\theta'} \right|^2, \end{aligned}$$

and we conclude owing to the Ito isometry and the growth conditions on the coefficients (because $\sigma, u, D_x u, D_x^2 u, f \in H_{loc, pol}^{1/2,1}$).

(b) From (3.27) we estimate

$$\begin{aligned} \mathbb{E}^B |R_\theta^{\varepsilon_N}(t_n, \xi)|^2 & \leq 5 \int_0^1 \mathbb{E}^B \left| F^{(n+1)}(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) - F^{(n+1)}(t_n, \xi) \right|^2 d\theta' \\ & + 5 \mathbb{E}^B \int_0^1 \left\| B_{\theta'}^\top (\sigma^\top (D_x^2 u))(t_n, \xi) (\sigma(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) - \sigma(t_n, \xi)) \right\|_2^2 d\theta' \\ & + 5 \varepsilon_N^{-1} \mathbb{E}^B \int_0^1 \|r_{\theta'}^{\varepsilon_N} \sigma(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N})\|_2^2 d\theta' \\ & + 5 \varepsilon_N \int_0^1 \mathbb{E}^B \left| B_{\theta'}^\top (\sigma^\top (D_x^2 u))(t_n, \xi) \mu(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N}) \right|^2 d\theta' \\ & + 5 \int_0^1 \mathbb{E}^B |r_{\theta'}^{\varepsilon_N} \mu(t_n + \theta' \varepsilon_N, X_{\theta'}^{\varepsilon_N})|^2 d\theta', \end{aligned}$$

for all $\xi \in \mathbb{R}^d$, $n \in \{0, \dots, N-1\}$ and $\theta \in [0, 1]$. Now we conclude to the inequality (b) by using that $f, u, D_x u, D_x^2 u, \sigma$ are in $H_{loc, pol}^{1/2,1}$, and by applying Lemmas 5.1 and 3.4. $\square$

2) Expansion of $T_1^{\varepsilon_N}(t_n, \xi)$ and $T_3^{\varepsilon_N}(t_n, \xi)$. From $\ell''_\gamma(y) = \ell''_\gamma(y/\varepsilon_N)$, for all $y \in \mathbb{R}$, and the expansion of $\mathcal{E}_\theta^{\varepsilon_N}$ in (3.25), we get

$$\ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) = \ell''_\gamma \left(E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + R_\theta^{\varepsilon_N}(t_n, \xi) \right). \quad (3.28)$$

By combining this with (3.21) and (3.25), we obtain

$$\begin{aligned} T_1^{\varepsilon_N}(t_n, \xi) & = -\mathbb{E}^B \left[\int_0^1 \ell''_\gamma \left(E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + R_\theta^{\varepsilon_N}(t_n, \xi) \right) \right. \\ & \quad \left. \times F(t_n, \xi) E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) d\theta \right] + C_1^{\varepsilon_N}(t_n, \xi), \end{aligned}$$

where

$$\begin{aligned} C_1^{\varepsilon_N}(t_n, \xi) & := -\mathbb{E}^B \left[\int_0^1 \ell''_\gamma \left(E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + R_\theta^{\varepsilon_N}(t_n, \xi) \right) \right. \\ & \quad \left. \times R_\theta^{\varepsilon_N}(t_n, \xi) F^{(n+1)}(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) d\theta \right] \end{aligned}$$

$$\begin{aligned}
& -\mathbb{E}^B \left[\int_0^1 \ell''_\gamma(\mathcal{E}_\theta^{\varepsilon_N}) \left(F^{(n+1)}(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N}) - F^{(n+1)}(t_n, \xi) \right) \right. \\
& \quad \left. \times E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) d\theta \right]. \tag{3.29}
\end{aligned}$$

The estimates of $C_1^{\varepsilon_N}(t_n, \xi)$ and $T_3^{\varepsilon_N}(t_n, \xi)$ are summarized in the following lemma.

Lemma 3.6. *Under the assumptions of Theorem 2.1, it holds*

$$(a) \quad \mathbb{E}^B |C_1^{\varepsilon_N}(t_n, \xi)| \leq C_{n,N}(\xi) \varepsilon_N^{1/2},$$

$$(b) \quad \mathbb{E}^B |T_3^{\varepsilon_N}(t_n, \xi)| \leq C_{n,N}(\xi) \varepsilon_N^{1/2},$$

for some constant $C_{n,N}(\xi) \in C_{pol}$.

Proof of Lemma 3.6. (a) From (3.29), it readily follows that

$$\begin{aligned}
|C_1^{\varepsilon_N}(t_n, \xi)| & \leq K \mathbb{E}^B \left[\int_0^1 \left| F^{(n+1)}(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N}) \right| |R_\theta^{\varepsilon_N}(t_n, \xi)| d\theta \right] \\
& + K \mathbb{E}^B \left[\int_0^1 \left| F^{(n+1)}(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N}) - F^{(n+1)}(t_n, \xi) \right| \left| E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) \right| d\theta \right]
\end{aligned}$$

where K is an upper bound of ℓ''_γ . For the first term above, use that $F^{(n+1)}$ has polynomial growth in its arguments (because $u^{(n+1)}, D_x u^{(n+1)}, D_x^2 u^{(n+1)}, f \in H_{loc, pol}^{1/2, 1}$) and Lemma 3.5(b). For the second term, apply the Cauchy - Schwarz with Lemmas 5.1 and 3.5(a). It yields

$$|C_1^{\varepsilon_N}(t_n, \xi)| \leq C_{n,N}(\xi) \varepsilon_N^{1/2}$$

as announced.

(b) Similarly, from (3.23) we write

$$\begin{aligned}
& |T_3^{\varepsilon_N}(t_n, \xi)| \\
& \leq K \int_0^1 \mathbb{E}^B \left[\left| E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + R_\theta^{\varepsilon_N}(t_n, \xi) \right| |((\Delta D_x u) \mu)(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N})| \right] d\theta \\
& \leq K \int_0^1 \sqrt{\mathbb{E}^B \left[\left| E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + R_\theta^{\varepsilon_N}(t_n, \xi) \right|^2 \right]} \\
& \quad \times \sqrt{\mathbb{E}^B \left[|((\Delta D_x u) \mu)(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N})|^2 \right]} d\theta.
\end{aligned}$$

It is now straightforward to conclude that the above is bounded by $C_{n,N}(\xi) \varepsilon_N^{1/2}$, using Lemmas 5.1, 3.4 and 3.5. $\square$

3) Expansion of $C_2^{\varepsilon_N}(t_n, \xi)$. Using the expansion of $\Delta D_x u$ in (3.24), we obtain

$$\begin{aligned}
& \|((\Delta D_x u) \sigma)(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N})\|^2 \\
& = \varepsilon_N \|B_\theta^\top (\sigma^\top (D_x^2 u))(t_n, \xi) \sigma(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N})\|^2 + \|r_\theta^{\varepsilon_N} \sigma(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N})\|^2 \\
& + 2\varepsilon_N^{1/2} (B_\theta^\top (\sigma^\top (D_x^2 u))(t_n, \xi) \sigma(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N})) \cdot (r_\theta^{\varepsilon_N} \sigma(t_n + \theta\varepsilon_N, X_\theta^{\varepsilon_N})).
\end{aligned}$$

Replacing the identity $\sigma(t, \zeta) = \Delta\sigma(t, \zeta) + \sigma(t_n, \xi)$ into the first term of the previous equation, we get

$$\|((\Delta D_x u) \sigma)(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^2 = \varepsilon_N \|B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u) \sigma)(t_n, \xi)\|^2 + c_{\theta}^{\varepsilon_N}(t_n, \xi), \quad (3.30)$$

where

$$\begin{aligned} c_{\theta}^{\varepsilon_N}(t_n, \xi) &= \varepsilon_N \|B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u))(t_n, \xi) \Delta\sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^2 + \|r_{\theta}^{\varepsilon_N} \sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^2 \\ &+ 2\varepsilon_N (B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u))(t_n, \xi) \Delta\sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})) \cdot (B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u) \sigma)(t_n, \xi)) \\ &+ 2\varepsilon_N^{1/2} (B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u))(t_n, \xi) \sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})) \cdot (r_{\theta}^{\varepsilon_N} \sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})). \end{aligned} \quad (3.31)$$

From (3.30) and (3.28), the expression of $T_2^{\varepsilon_N}$ in (3.22) becomes

$$\begin{aligned} T_2^{\varepsilon_N}(t_n, \xi) &= \frac{1}{2} \mathbb{E}^B \left[\int_0^1 \ell''_{\gamma} \left(E_{\theta} \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + R_{\theta}^{\varepsilon_N}(t_n, \xi) \right) \|B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u) \sigma)(t_n, \xi)\|^2 d\theta \right] \\ &+ C_2^{\varepsilon_N}(t_n, \xi), \end{aligned}$$

where

$$C_2^{\varepsilon_N}(t_n, \xi) := \frac{1}{2} \mathbb{E}^B \left[\int_0^1 \ell''_{\gamma} \left(E_{\theta} \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + R_{\theta}^{\varepsilon_N}(t_n, \xi) \right) \varepsilon_N^{-1} c_{\theta}^{\varepsilon_N}(t_n, \xi) d\theta \right]. \quad (3.32)$$

The estimate of $C_2^{\varepsilon_N}(t_n, \xi)$ is summarized in the following lemma.

Lemma 3.7. *Under the assumptions of Theorem 2.1, it holds*

$$|C_2^{\varepsilon_N}(t_n, \xi)| \leq C_{n,N}(\xi) \varepsilon_N^{1/2},$$

for some constant $C_{n,N}(\xi) \in C_{pol}$.

Proof of Lemma 3.7. From the expression $c_{\theta}^{\varepsilon_N}(t_n, \xi)$ in (3.31), we write

$$\begin{aligned} \mathbb{E}^B |c_{\theta}^{\varepsilon_N}(t_n, \xi)| &\leq \varepsilon_N \mathbb{E}^B \|B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u))(t_n, \xi) \Delta\sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^2 \\ &+ \mathbb{E}^B \|r_{\theta}^{\varepsilon_N} \sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^2 \\ &+ 2\varepsilon_N \sqrt{\mathbb{E}^B \|B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u))(t_n, \xi) \Delta\sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^2} \\ &\quad \times \sqrt{\mathbb{E}^B \|B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u) \sigma)(t_n, \xi)\|^2} \\ &+ 2\varepsilon_N^{1/2} \sqrt{\mathbb{E}^B \|B_{\theta}^{\top} (\sigma^{\top}(D_x^2 u))(t_n, \xi) \sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^2} \\ &\quad \times \sqrt{\mathbb{E}^B \|r_{\theta}^{\varepsilon_N} \sigma(t_n + \theta \varepsilon_N, X_{\theta}^{\varepsilon_N})\|^2} \\ &\leq C_{n,N}(\xi) \varepsilon_N^{3/2}. \end{aligned}$$

Again we have used the polynomial growth condition on $\sigma, D_x^2 u$ and the local regularity condition on $\sigma \in H_{loc, pol}^{1/2, 1}$ with Lemma 5.1, and Lemma 3.4 - (a). Consequently and in view of the definition (3.32) of $C_2^{\varepsilon_N}(t_n, \xi)$, we obtain the estimate

$$|C_2^{\varepsilon_N}(t_n, \xi)| \leq \frac{1}{2} |\ell''_{\gamma}|_{\infty} \varepsilon_N^{-1} \sup_{\theta \in [0, 1]} \mathbb{E}^B |c_{\theta}^{\varepsilon_N}(t_n, \xi)|,$$

which leads to the announced result. $\square$

4) Expansion of $T^{\varepsilon_N}(t_n, \xi)$. From (3.20) and the previous expansions of $T_i^{\varepsilon_N}(t_n, \xi)$ for $i = 1, 2, 3$, we deduce

$$\begin{aligned} T^{\varepsilon_N}(t_n, \xi) &= \mathbb{E}^B \left[\int_0^1 \ell''_\gamma \left(E_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) + R_\theta^{\varepsilon_N}(t_n, \xi) \right) \right. \\ &\quad \times Q_\theta \left(G^{(n+1)}(t_n, \xi), F^{(n+1)}(t_n, \xi) \right) d\theta \Big] \\ &\quad + C_1^{\varepsilon_N}(t_n, \xi) + C_2^{\varepsilon_N}(t_n, \xi) + T_3^{\varepsilon_N}(t_n, \xi), \end{aligned}$$

where Q_θ is defined in (3.7). Since $C_1^{\varepsilon_N}(t_n, \xi)$, $C_2^{\varepsilon_N}(t_n, \xi)$ and $T_3^{\varepsilon_N}(t_n, \xi)$ satisfy $\varepsilon_N^{1/2}$ -bounds, we get the result of Proposition 3.1. $\square$

3.6 Proof of Lemma 3.3

(a) From the definition of E_θ in (3.6), it follows that

$$\begin{aligned} &\left| E_\theta \left(G^{(n+1)}(t_n, X_{t_n}), F^{(n+1)}(t_n, X_{t_n}) \right) - E_\theta \left(\Gamma(t_{n+1}, X_{t_n}), F(t_{n+1}, X_{t_n}) \right) \right| \\ &\leq \left| f(t_n, X_{t_n}, u^{(n+1)}(t_n, X_{t_n}), D_x u^{(n+1)}(t_n, X_{t_n}), D_x^2 u^{(n+1)}(t_n, X_{t_n})) \right. \\ &\quad \left. - f(t_{n+1}, X_{t_n}, v(t_{n+1}, X_{t_n}), D_x v(t_{n+1}, X_{t_n}), D_x^2 v(t_{n+1}, X_{t_n})) \right| \\ &\quad + \left\| (\sigma^\top (D_x^2 u^{(n+1)}) \sigma)(t_n, X_{t_n}) - (\sigma^\top (D_x^2 v) \sigma)(t_{n+1}, X_{t_n}) \right\| \left\| \int_0^\theta B_{\theta'} dB_{\theta'}^\top \right\| \\ &\leq C_{n,N}(X_{t_n}) \varepsilon_N^{1/2} \left(1 + \left\| \int_0^\theta B_{\theta'} dB_{\theta'}^\top \right\| \right), \end{aligned}$$

for some constant $C_{n,N}(X_{t_n}) \in C_{\text{pol}}$, where we have used Proposition 3.2 and the assumptions on coefficients, prices and greeks. Owing to the Burkholder - Davis - Gundy (BDG) inequalities, we conclude the proof of (a).

(b) From $\bar{C}^{\varepsilon_N}(t_n, \xi)$ in (3.14) we get

$$\begin{aligned} &\mathbb{E} \left| \bar{C}^{\varepsilon_N}(t_n, X_{t_n}) \right| \leq \\ &|\ell''_\gamma|_\infty \int_0^1 \mathbb{E} \left| Q_\theta \left(G^{(n+1)}(t_n, X_{t_n}), F^{(n+1)}(t_n, X_{t_n}) \right) - Q_\theta \left(\Gamma(t_{n+1}, X_{t_n}), F(t_{n+1}, X_{t_n}) \right) \right| d\theta. \end{aligned}$$

Considering the expression of Q_θ in (3.7), we are able to apply the same arguments as for (a). Further details are left to the reader. We are done with the estimate (b).

(c) Let $p \geq 1$ and set $Z_N := \sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} |\bar{R}_\theta^{\varepsilon_N}(t_n, X_{t_n})|^p$. From the definition (3.13) of $\bar{R}_\theta^{\varepsilon_N}(t_n, \xi)$ we write

$$\begin{aligned} \mathbb{E}[Z_N] &\leq 2^{p-1} \mathbb{E} \left[\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} \left| E_\theta \left(G^{(n+1)}(t_n, X_{t_n}), F^{(n+1)}(t_n, X_{t_n}) \right) \right. \right. \\ &\quad \left. \left. - E_\theta \left(\Gamma(t_{n+1}, X_{t_n}), F(t_{n+1}, X_{t_n}) \right) \right|^p \right] \\ &\quad + 2^{p-1} \mathbb{E} \left[\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} \left| R_\theta^{\varepsilon_N}(t_n, X_{t_n}) \right|^p \right] \leq K_p N \varepsilon_N^{p/2} \end{aligned}$$

owing to (a) and Lemma 3.8 stated below. Last, apply Lemma 5.2 to the above Z_N with $p > 4$: we are done. $\square$

In the proof, we have used the following result, useful to justify the a.s. convergence to 0 of remainder terms. We postpone its proof to Appendix 5.1.

Lemma 3.8. *Let $R_\theta^{\varepsilon_N}(t_n, \xi)$ be given by (3.8) and $p \geq 1$. Under the assumptions of Theorem 2.1, there exists a finite positive constant K_p depending on the coefficients $\mu, \sigma, f, u^{(n+1)}$ and its derivatives such that*

$$\mathbb{E} \left[\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} |R_\theta^{\varepsilon_N}(t_n, X_{t_n})|^p \right] \leq K_p N \varepsilon_N^{p/2}.$$

Proof. We claim the following upper bound holds

$$\mathbb{E}^B \left[\sup_{\theta \in [0,1]} |R_\theta^{\varepsilon_N}(t_n, \xi)|_2^p \right] \leq C_{n,N}(\xi) \varepsilon_N^p \quad (3.33)$$

for some constant $C_{n,N}(\xi) \in C_{\text{pol}}$. With this control at hand, we complete the proof by using the rough inequality

$$\mathbb{E} \left[\sup_{0 \leq n \leq N-1} \sup_{\theta \in [0,1]} |R_\theta^{\varepsilon_N}(t_n, X_{t_n})|^p \right] \leq \sum_{n=0}^{N-1} \mathbb{E} \left[\sup_{\theta \in [0,1]} |R_\theta^{\varepsilon_N}(t_n, X_{t_n})|^p \right].$$

So, it is enough to show (3.33). Regarding the control with $R_\theta^{\varepsilon_N}(t_n, \xi)$, we follow the proof of Lemma 3.5 item (b). The adaptation is obvious since instead of taking $p = 2$, we take $p \geq 1$. Then we handle the supremum over θ inside the expectation using BDG inequalities. Other arguments are unchanged, leading to the announced estimate. We leave to the reader the details. $\square$

4 Numerical experiments

In this section, we compute a numerical approximation of v^* in (2.12), solution to the f -PDE in (1.5) using the optimal kernel f^* defined in (2.10). In Subsection 2.4, we obtain a quasi-explicit formulation for the optimal kernel f_γ^* (see (2.15)) in the one-dimensional case. Therefore, we only perform numerical experiments in dimension $d = 1$ with the risk parameter $\gamma \in \{0.0, 0.1, 0.2, 0.3\}$. First, in Section 4.1, we present the numerical solution for a set of European options. Then, in Section 4.2, we compute the asymptotic risk $\mathcal{R}_\gamma(v^*, f)$ for different kernels $f \in \{f_0^*, f_{0.1}^*, f_{0.2}^*, f_{0.3}^*\}$ confirming the optimality of f_γ^* . Finally, in Section 4.3, we compare numerically the solution to the f -PDE with the solution to the minimization problem (1.3). We aim to check the conjecture whether one can interchange the limit in N and the minimization over strategies in our setting. Alternatively, we verify the solution to the minimization problem in discrete time (see (1.3)) corresponds, as N large, to the solution to the nonlinear f^* -PDE (1.5).

4.1 The f^* -PDE valuation for different options

Here we show the numerical solution to (1.5) for different option payoffs under the assumption that the underlying process X satisfies the SDE (2.1) with $\sigma(t, x) = \sigma x$. We

consider the value function $U(t, x)$ as the solution to the following f -PDE valuation (in a forward form)

$$\frac{\partial U}{\partial t}(t, x) = \alpha(x) \partial_x^2 U(t, x) + f(2\alpha(x) \partial_x^2 U(t, x)), \quad (t, x) \in (0, T] \times \mathbb{R}, \quad (4.1)$$

where $\alpha(x) = \frac{1}{2}\sigma^2 x^2$ and $f : \mathbb{R} \rightarrow \mathbb{R}$ is a real-valued function to be chosen. Seeing that (4.1) has a second-order partial differential in space and first order in time, we require for a numerical resolution one initial and two boundary conditions. Also, the payoff of European options with maturity T , denoted by $h(x)$, will be used as initial time condition. We have chosen the following options:

- (i) call option with payoff $h(x) = (x - K_0)_+$ and put option with payoff $h(x) = (K_0 - x)_+$, where $x \mapsto x_+ = \max(x, 0)$ and K_0 is the strike price;
- (ii) asset-or-nothing call option with payoff $h(x) = x \mathbf{I}_{x - K_0 > 0}$ and asset-or-nothing put option with payoff $h(x) = x \mathbf{I}_{x - K_0 < 0}$, where K_0 is the strike price;
- (iii) bull spread option with payoff $h(x) = (x - K_1)_+ - (x - K_2)_+$ and bear spread option with payoff $h(x) = (K_2 - x)_+ - (K_1 - x)_+$, where K_1, K_2 are strike prices with $K_2 > K_1$;

We examine the asset-or-nothing options because of their discontinuous payoff. We analyze the spread options because of the change of convexity. We are aware these payoffs do not satisfy the assumptions of Theorem 2.1, but we believe that these hypotheses are only sufficient and the previous asymptotic analysis can also be applied to those payoffs.

In the following numerical examples, we consider:

Set	Strike	Volatility	Maturity
A (vanilla and digital)	$K_0 = 100$	$\sigma = 0.3$	$T = 1$
B (spread)	$(K_1, K_2) = (90, 110)$	$\sigma = 0.3$	$T = 1$

Space discretization. Here, we detail our numerical scheme. We look for a second-order accurate solution to the PDE in (4.1) on a finite domain $L = [0, x_{\max}]$. Let $I \in \mathbb{N}$. Therefore, we equally discretize L in $I + 1$ points $\{x_0, x_1, \dots, x_{I-1}, x_I\}$ such that $\Delta x = x_{\max}/I$ and $x_i = i\Delta x$ for each $0 \leq i \leq I$. Assuming that U is smooth enough, we get the second-order approximation of the second derivative of U

$$\frac{U_{i+1} - 2U_i + U_{i-1}}{\Delta x^2} = \partial_x^2 U|_{x_i} + \mathbf{O}(\Delta x^2),$$

for every $1 \leq i \leq I - 1$, with U_i denoting $U(x_i)$. Thanks to the second-order approximation, we obtain a semi-discretization from (4.1):

$$\partial_t U_i = \alpha_i \frac{U_{i+1} - 2U_i + U_{i-1}}{\Delta x^2} + f\left(2\alpha_i \frac{U_{i+1} - 2U_i + U_{i-1}}{\Delta x^2}\right), \quad (4.2)$$

for every $1 \leq i \leq I - 1$, where the factor α_i is $\alpha(x)$ evaluated in each x_i . Assuming that f in (4.1) is Lipschitz continuous, the system of equations (4.2) is a second-order approximation of the PDE (4.1) and can be viewed in matrix form as

$$\frac{d\mathbf{U}}{dt} = \mathbf{A}\mathbf{U} + f(2\mathbf{A}\mathbf{U}),$$

where A is the coefficient matrix and $\mathbf{U}$ the discrete solution. Besides the system in (4.2), $\mathbf{U}$ satisfies $U_0 = b_{\min}$ and $U_n = b_{\max}$, where $b_{\min}$ and $b_{\max}$ represent a Dirichlet-type boundary condition imposed to the numerical solution. Therefore, the matrix A is of form

$$A_{00} = 0, \quad A_{ii-1} = \alpha_i/\Delta x^2, \quad A_{ii} = -2\alpha_i/\Delta x^2, \quad A_{ii+1} = \alpha_i/\Delta x^2, \quad A_{II} = 0.$$

After the space discretization, it remains a system of ordinary differential equations

$$\frac{d\mathbf{U}}{dt} = A\mathbf{U} + \mathbf{F}, \quad \mathbf{U}(0) = h, \quad \mathbf{F} := f(2A\mathbf{U}).$$

Time discretization. Now we apply a second-order method in time. Let $J \in \mathbb{N}$. Divide the time interval $[0, T]$ in J intervals with a constant time step $\Delta t = T/J$.

Denote $\mathbf{U}^j$ (resp. $\mathbf{F}^j$) as the vector $\mathbf{U}$ (resp. $\mathbf{F}$) evaluated at $t = j\Delta t$. Due to the nonlinearity of $\mathbf{F}$ regarding $\mathbf{U}$, we will use Adams-Moulton (AM) methods with Adams-Bashforth (AB) methods to construct a Predictor-Corrector algorithm with AM and AB of the same order. Here we apply the second-order Adams-Bashforth (AB2) method to predict $\mathbf{F}^{j+1}$ and use $\bar{\mathbf{F}}^{j+1}$ within the second-order Adams-Moulton (AM2) method:

1. We predict $(\mathbf{U}^{j+1}, \mathbf{F}^{j+1})$ with AB2 which give us:

$$\bar{\mathbf{F}}^{j+1} = f(2A\bar{\mathbf{U}}^{j+1}), \quad \bar{\mathbf{U}}^{j+1} = \mathbf{U}^j + \Delta t \left[\frac{3}{2}(A\mathbf{U}^j + \mathbf{F}^j) - \frac{1}{2}(A\mathbf{U}^{j-1} + \mathbf{F}^{j-1}) \right].$$

2. We correct $(\mathbf{U}^{j+1}, \mathbf{F}^{j+1})$ with AM2 which gives us:

$$\mathbf{F}^{j+1} = f(2A\mathbf{U}^{j+1}), \quad \mathbf{U}^{j+1} = \mathbf{U}^j + \Delta t \left[\frac{1}{2}(A\mathbf{U}^{j+1} + \bar{\mathbf{F}}^{j+1}) + \frac{1}{2}(A\mathbf{U}^j + \mathbf{F}^j) \right].$$

Here, f is computed as the optimal kernel f_γ^* given in (2.15). Further, in Table 1, we give the constants c_1^* and c_2^* computed using a root finding algorithm. Since the algorithm looks two steps back, we will need some initialization steps. Therefore we use the AB1 (Forward Euler) and the AM1 (Backward Euler) method for the prediction and correction part, respectively,

$$\begin{aligned} \bar{\mathbf{U}}^1 &= \mathbf{U}^0 + \Delta t (A\mathbf{U}^0 + \mathbf{F}^0), & \mathbf{F}^0 &= f(2\mathbf{U}^0), \\ \mathbf{U}^1 &= \mathbf{U}^0 + \Delta t (A\mathbf{U}^1 + \bar{\mathbf{F}}^1), & \bar{\mathbf{F}}^1 &= f(2\bar{\mathbf{U}}^1). \end{aligned}$$

Initial boundary conditions. Regarding the boundary conditions, we have stipulated a space domain $L = [0, x_{\max}]$, where $x_{\max}$ is supposed to be large enough. Then we use Dirichlet boundary conditions $U(t, 0) = b_{\min}(t)$ and $U(t, x_{\max}) = b_{\max}(t)$ for any t in $[0, T]$. We set the left boundary $b_{\min}(t) = h(0)$ and the right boundary $b_{\max}(t) = h(x_{\max})$ for any t in $[0, T]$. Regarding the numerical solution, we fix $x_{\max} = 400$, $I = 200$ and $J = 200$.

In Figure 4, we show the vanilla option value plotted for different risk parameters γ . We depict analogous plot for digital and spread options in Figure 5 and 6, respectively. We remark the numerical solutions are increasing in function of γ . Intuitively, whenever the seller's risk aversion increases, it will be more reasonable to asking him for a higher option price. According to Proposition 2.3, we have $y \mapsto f_\gamma^*(y)$ is nonnegative for all $\gamma \in (0, 1)$. Therefore, the nonlinear source of PDE (4.1) is nonnegative whatever the sign of second derivatives. Our risk-aversion valuation adds a risk premium to the risk-neutral one whenever the underlying price varies too quickly, i.e., proportionally to the Greek Gamma.

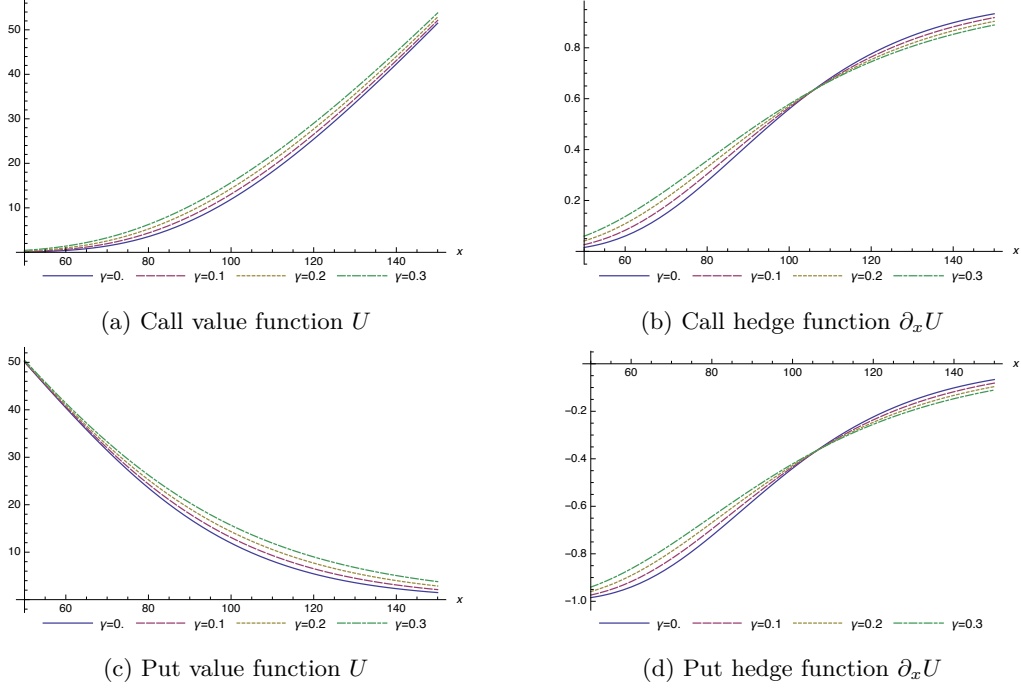


Figure 4: Vanilla options: numerical approximation of f^* -PDE solution U at final time for different γ .

4.2 The asymptotic risk $\mathcal{R}_\gamma(v, f)$ for different kernel f

Here we will test the asymptotic risk $\mathcal{R}_\gamma(v, f)$ (see in (2.6)) for different kernel f given a reference valuation v . Consider $v_\gamma^*(t, \cdot) = U(T - t, \cdot)$, where U is the solution to PDE (4.1) (in forward form) using the optimal kernel f_γ^* given by Proposition 2.3. Then we confirm numerically the optimality of f_γ^* for the reference valuation v_γ^* by computing $\mathcal{R}_\gamma(v_\gamma^*, f_{\gamma'}^*)$ for a different γ' . To achieve that, we approximate $\mathcal{R}_\gamma(v_\gamma^*, f_{\gamma'}^*)$ by forward Monte Carlo simulations of X . In addition, we use the numerical PDE solution to compute the partial derivatives of v_γ^* . We denote its estimate by $\hat{\mathcal{R}}_{N,M}(\gamma, \gamma')$, where N is the number of time steps and M is the number of paths $\{X_{t_n}\}_{n=0}^N$.

Set $\sigma = 0.3$, $N = 20$, $M = 5 \times 10^5$, and $X_0 \in \{90, 110\}$. The number of time steps used in the PDE resolution between each time step of the MC algorithm is 50. We study the following options:

- (i) call option with $K_0 = 100$ and $T = 1$;
- (ii) bear option with $K_1 = 80$, $K_2 = 120$ and $T = 1$.

Let $\gamma \in [0, 1)$ and $X_0 \in \mathbb{R}_+$ fixed. Thanks to Theorem 2.1 and Proposition 2.2, we expect the minimum of $\hat{\mathcal{R}}_{N,M}(\gamma', \gamma)$ in γ' is attained at $\gamma' = \gamma$. In Table 2, we compute the numerical approximation $\hat{\mathcal{R}}_{N,M}(\gamma, \gamma')$ for $(\gamma, \gamma') \in \{0.0, 0.1, 0.2, 0.3\}^2$ to verify this claim.

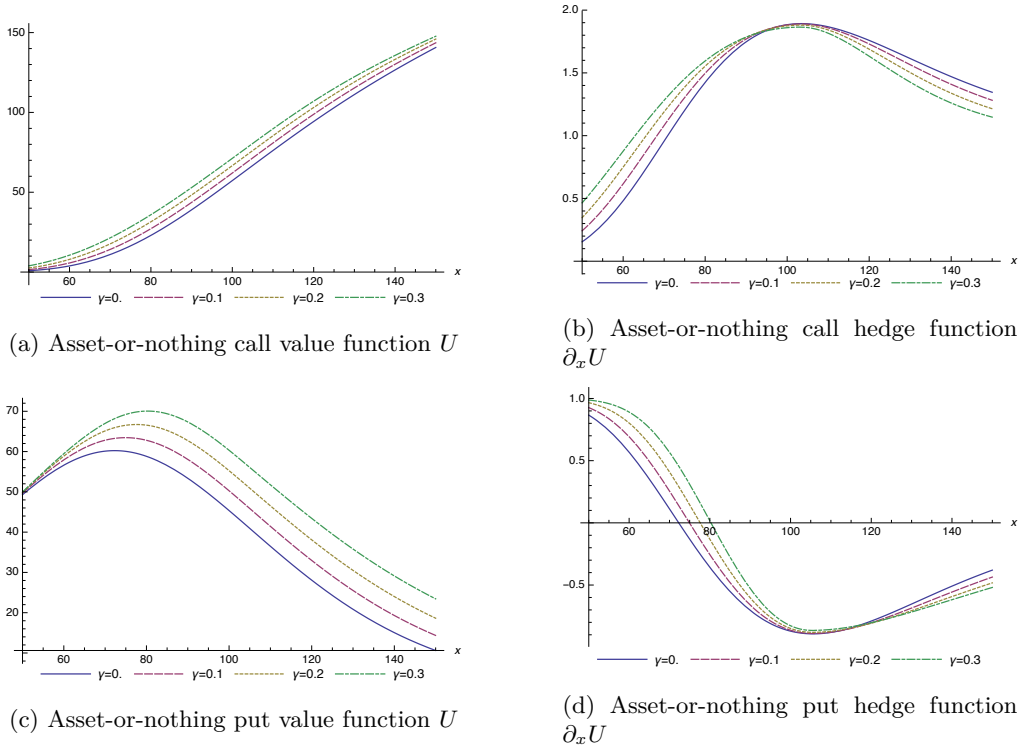


Figure 5: Digital options: numerical approximation of f^* -PDE solution U at final time for different γ .

γ	γ'	$X_0 = 90$	$X_0 = 110$
0	0	41.20±0.08	47.92±0.08
0	0.1	42.09±0.08	48.96±0.08
0	0.2	45.41±0.09	52.82±0.09
0	0.3	52.49±0.10	61.05±0.10
0.1	0	41.26±0.08	48.03±0.08
0.1	0.1	40.53±0.08	47.17±0.08
0.1	0.2	41.52±0.08	48.32±0.08
0.1	0.3	45.20±0.09	52.61±0.08
0.2	0	41.46±0.08	48.29±0.07
0.2	0.1	39.46±0.07	45.97±0.07
0.2	0.2	38.66±0.07	45.03±0.07
0.2	0.3	39.75±0.07	46.30±0.07
0.3	0	41.61±0.07	48.51±0.07
0.3	0.1	38.66±0.07	45.08±0.07
0.3	0.2	36.51±0.07	42.57±0.06
0.3	0.3	35.65±0.06	41.56±0.06

(a) $\hat{\mathcal{R}}_{N,M}$ for a call option with $K_0 = 100$.

γ	γ'	$X_0 = 90$	$X_0 = 110$
0	0	41.20±0.08	47.92±0.08
0	0.1	42.09±0.08	48.96±0.08
0	0.2	45.41±0.09	52.82±0.09
0	0.3	52.49±0.10	61.05±0.10
0.1	0	41.26±0.08	48.03±0.08
0.1	0.1	40.53±0.08	47.17±0.08
0.1	0.2	41.52±0.08	48.32±0.08
0.1	0.3	45.20±0.09	52.61±0.08
0.2	0	41.46±0.08	48.29±0.07
0.2	0.1	39.46±0.07	45.97±0.07
0.2	0.2	38.66±0.07	45.03±0.07
0.2	0.3	39.75±0.07	46.30±0.07
0.3	0	41.61±0.07	48.51±0.07
0.3	0.1	38.66±0.07	45.08±0.07
0.3	0.2	36.51±0.07	42.57±0.06
0.3	0.3	35.65±0.06	41.56±0.06

(b) $\hat{\mathcal{R}}_{N,M}$ for a bear option with $K_1 = 80$ and $K_2 = 120$.

Table 2: Asymptotic risk estimate $\hat{\mathcal{R}}_{N,M}$ for $N = 20$ and $M = 5 \times 10^5$.

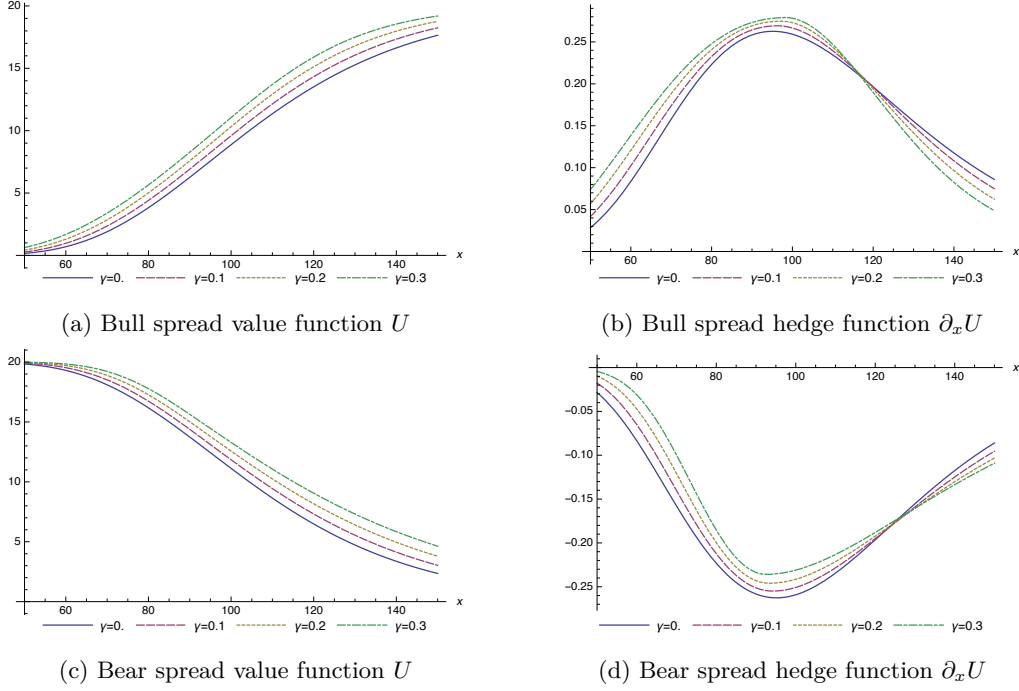


Figure 6: Spread options: numerical approximation of f^* -PDE solution U at final time for different γ .

4.3 The f^* -PDE valuation/hedging rule and the discrete time problem solution

Let U_γ be the solution to the forward f_γ^* -PDE (4.1). Here, we compare the f_γ^* -PDE valuation/hedging rule $\varphi_\gamma^*(t, \cdot) = (U_\gamma(T - t, \cdot), \partial_x U_\gamma(T - t, \cdot))$ for $t \in [0, T]$ and the discrete time problem solution $\varphi_{t_n}^N(t, \cdot) = (V_\gamma^N(t_n, \cdot), \delta_\gamma^N(t_n, \cdot))$ for $0 \leq n \leq N$, where N is the number of hedging times. We approximate φ_γ^N by $\varphi_\gamma^{N,M}$ by using a Regression Monte Carlo (RMC) algorithm, where M is the number of Monte Carlo paths.

RMC Algorithm. Here we present our RMC algorithm, which is a variation of the Hedged Monte Carlo algorithm (proposed in [17]) with a fixed point stage. We determine the option value by working step by step from $T = N\Delta t$ to the present $t = 0$, where Δt is the time interval. We denote the underlying asset price X at time $t_n = n\Delta t$ by X_n and the option value $\mathcal{V}_n(X_n)$ at time t_n only depends on the current asset price X_n . We introduce the hedge $\delta_n(X_n)$, which is the amount of the underlying asset in the portfolio at time t_n , when the asset price is X_n .

The average risk, over all paths of the underlying process, is given by

$$\mathcal{R}_n = \langle \ell_\gamma (\mathcal{V}_{n+1}(X_{n+1}) - \mathcal{V}_n(X_n) - \delta_n(X_n)(X_{n+1} - X_n)) \rangle_M,$$

where the angled brackets $\langle \dots \rangle_M$ denote the average over the sampled asset values. The functional minimization of $\mathcal{R}_n$ with respect to $\mathcal{V}_n(X_n)$ and $\delta_n(X_n)$ gives us equations which allow us to determine the option value and hedge provided that $\mathcal{V}_{n+1}$ is known. We generate a set of M paths X_n^m , where n is the time index and m the path index. We decompose $\mathcal{V}_n$ and δ_n over a set of K basis functions L_k^n and C_k^n . The use of local basis function in RMC is presented in [9]. Therefore, we set L_k^n and C_k^n , respectively, as

a piecewise linear and constant function on each partition of real line. In addition, we use adaptive breakpoints as proposed in [2]:

$$\mathcal{V}_n^K(x) := \sum_{k=1}^K a_k^n L_k^n(x), \quad \delta_n^K(x) := \sum_{k=1}^K b_k^n C_k^n(x).$$

In other words, we reduce the original functional optimization problem (find the functions $\mathcal{V}_n$ and δ_n) to a numerical optimization (find the coefficients a_k^n and b_k^n). We have a good approximation of the true functional solution conditionally to K be large enough. We then solve N minimization problems backwardly in time from maturity T , where $\mathcal{V}_N(x)$ is equal to the payoff function h . For each step n , we minimize

$$\frac{1}{M} \sum_{m=1}^M \ell_\gamma(\mathcal{E}_{n,m}^K(\mathcal{V}_{n+1}, a^n, b^n))$$

where

$$\mathcal{E}_{n,m}^K(\mathcal{V}, a, b) := \mathcal{V}(X_{n+1}^m) - \sum_{k=1}^K a_k L_k^n(X_n^m) - \sum_{k=1}^K b_k C_k^n(X_n^m) (X_{n+1}^m - X_n^m).$$

Thanks to the choice of the risk function ℓ , we write $\ell_\gamma(y) = (y w_\gamma(y))^2$ with a weight function $w_\gamma(y) = 1 + \gamma \text{Sgn}(y)$. Then, for each $n \in \{N-1, \dots, 0\}$, we solve the following fixed point problem starting from the quadratic optimal solution

$$\begin{aligned} (a^{n,0}, b^{n,0}) &:= \underset{(a,b)}{\operatorname{argmin}} \frac{1}{M} \sum_{m=1}^M (\mathcal{E}_{n,m}^K(\mathcal{V}_{n+1}, a, b))^2, \\ (a^{n,p+1}, b^{n,p+1}) &:= \underset{(a,b)}{\operatorname{argmin}} \frac{1}{M} \sum_{m=1}^M (\mathcal{E}_{n,m}^K(\mathcal{V}_{n+1}, a, b) w_\gamma(\mathcal{E}_{n,m}^K(\mathcal{V}_{n+1}, a^{n,p}, b^{n,p})))^2, \end{aligned}$$

for every $p \in \{0, \dots, P-1\}$, where $\mathcal{V}_{n+1} = \mathcal{V}_{n+1}^{K,P}$,

$$\mathcal{V}_{n+1}^{K,P} := \sum_{k=1}^K a_k^{n+1,P} L_k^{n+1}, \quad \delta_{n+1}^{K,P} := \sum_{k=1}^K b_k^{n+1,P} C_k^{n+1}.$$

The least square problem with weights is solved using standard procedures. From a practical point of view, we have used a C++ library called StOpt (see the documentation in [8]) to implement this previous RMC with local basis function and adaptative breakpoints. Even though we do not establish any theoretical convergence result, we know the previous algorithm is strongly related to a RMC method for computing generalized BSDEs proposed in [13, 11]. In following, we denote the optimal strategy $(\mathcal{V}_n^{K,P}(\cdot), \delta_n^{K,P}(\cdot))$ as $\varphi_\gamma^{N,M}(t_n, \cdot)$.

Set of parameters. Regarding the RMC algorithm, we set $M = 8 \times 10^5$, $N = 40$, $K = 80$ and $P = 20$. For the underlying process, we set $\sigma = 0.3$, $X_0 = 100$ and $T = 1$. Here, we compare the optimal valuation/hedging rule $\varphi_\gamma^*(t, \cdot)$ and the discrete time problem solution $\varphi_\gamma^N(t_n, \cdot)$ for a call option with strike $K_0 = 100$ and a bear option with strikes $K_1 = 80$, $K_2 = 120$.

Thanks to the previous algorithm, we compute the option value $V_\gamma^{N,M}(t_n, \cdot)$. Owing to the finite difference scheme in Subsection 4.1, we have the value function $U_\gamma(T - t_n, \cdot)$.

Here we consider $\gamma \in \{0.0, 0.1, 0.2, 0.3\}$ and $t_n \in \{0.1, 0.3\}$. In Figure 7, we present the relative error $V_\gamma^{N,M}(t_n, \cdot)/U_\gamma(T - t_n, \cdot) - 1$ for a call option. We show analogous plot for a bear spread option in Figure 8. We observe that relative errors seem to confirm numerically the conjecture: the optimal price in discrete time for large number of hedging times coincides asymptotically with the f^* -PDE solution.

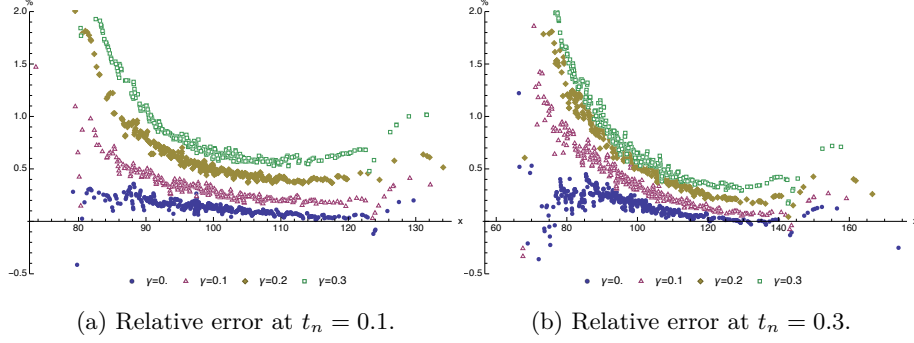


Figure 7: Relative error $V_\gamma^{N,M}(t_n, \cdot)/U_\gamma(T - t_n, \cdot) - 1$ for a call option.

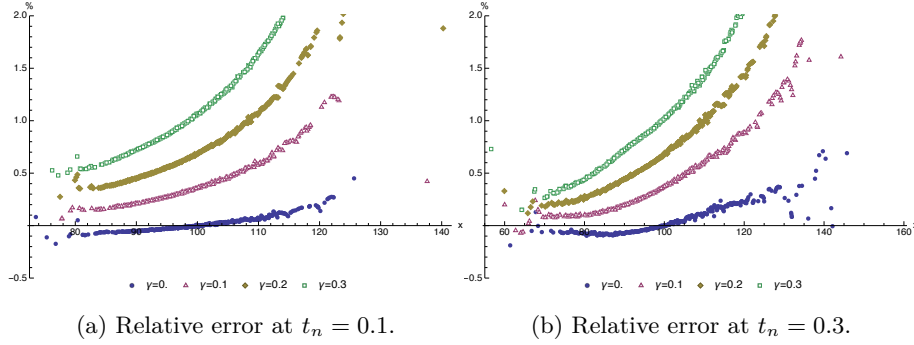


Figure 8: Relative error $V_\gamma^{N,M}(t_n, \cdot)/U_\gamma(T - t_n, \cdot) - 1$ for a bear option.

5 Appendix

5.1 Technical results

We gather the results related to the proof of Theorem 2.1. The first one is about estimating the increment of $\|X_\theta^{\varepsilon_N} - \xi\|$. This is quite standard, and the proof is left to the reader.

Lemma 5.1. *We have for any $p \geq 1$*

$$\sup_{\theta \in [0,1]} \mathbb{E}^B \|X_\theta^{\varepsilon_N} - \xi\|^p \leq C_{n,N}(\xi) \varepsilon_N^{p/2},$$

for some constant $C_{n,N}(\xi) \in C_{pol}$ depending on p . Also, for any function $\phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ in $H_{loc,pol}^{1/2,1}$, it holds for any $p \geq 1$

$$\sup_{\theta \in [0,1]} \mathbb{E}^B |\phi(t_n + \theta \varepsilon_N, X_\theta^{\varepsilon_N}) - \phi(t_n, \xi)|^p \leq C_{n,N}(\xi) \varepsilon_N^{p/2} \quad (5.1)$$

for some constant $C_{n,N}(\xi) \in C_{pol}$ depending on p .

Next, the following lemma gives a sufficient condition on the expectation of a random variable Z_N to ensure its almost sure convergence.

Lemma 5.2. *Let $(Z_N)_{N \geq 1}$ be a sequence of positive real random variables. If Z_N satisfies $\mathbb{E}[Z_N] \leq C/N^{1+\delta}$ for some finite numbers $C \geq 0$ and $\delta > 0$, then Z_N converges almost surely towards 0, i.e., $Z_N \xrightarrow[N \rightarrow \infty]{a.s.} 0$.*

Proof. The argument is quite standard, and for the sake of completeness, we give it. Summing up the expectation of Z_N , we get $\sum_{N \geq 1} \mathbb{E}[Z_N] \leq C \sum_{N \geq 1} 1/N^{1+\delta} < +\infty$ by hypothesis. Then the positive random variable $\sum_{N \geq 1} Z_N$ has a finite expectation. Owing the Fubini theorem, this implies that $\sum_{N \geq 1} Z_N < +\infty$ with probability 1. Therefore the general term Z_N converges almost surely towards 0. $\square$

This proposition is used to complete the proof of Theorem 2.1.

Proposition 5.3. *Under the assumptions of Theorem 2.1, notably Assumption A5, the set $\mathcal{A}$ has measure zero.*

Proof. From Equation (3.6), we write E_θ in the following form

$$E_\theta(S, y) = (B_\theta^\top S B_\theta - \text{Tr}[S]\theta)/2 - y\theta,$$

for any $\theta, S, y \in [0, 1] \times \mathcal{S}^d \times \mathbb{R}$. From Equation (3.12), we recall F and G

$$\begin{aligned} F(t, \xi) &= f(t, \xi, v(t, \xi), D_x v(t, \xi), D_x^2 v(t, \xi)), \\ G(t, \xi) &= (\sigma^\top(D_x^2 v)\sigma)(t, \xi), \end{aligned}$$

for any $t, \xi \in [0, 1] \times \mathbb{R}^d$. Here we show the set $\mathcal{A}$ defined in (3.16)

$$\mathcal{A} = \left\{ (\omega, t, \theta) \in \Omega \times [0, T] \times [0, 1] \mid E_\theta(G(t, X_t), F(t, X_t)) = 0 \right\}$$

has measure zero w.r.t. $d\mathbb{P}^W \otimes d\mathbb{P}^B \otimes dt \otimes d\theta$. This is equivalent to say

$$\mathbb{E}^{W \otimes B} \left[\int_0^T \int_0^1 \mathbf{I}_{E_\theta(G(t, X_t), F(t, X_t))=0} d\theta dt \right] = 0.$$

Applying Fubini's theorem, the tower property of the conditional expectation and the assumption A5, the previous condition is equivalent to

$$\int_0^T \int_0^1 \mathbb{E}^W \left[\mathbb{P}^B [E_\theta(G(t, X_t), F(t, X_t)) = 0 \mid X_t] \mathbf{I}_{G(t, X_t) \neq 0} \right] d\theta dt = 0.$$

For $S \neq \mathbf{0} \in \mathcal{S}^d$, $y \in \mathbb{R}$ and $\theta \neq 0$, we claim that the random variable $E_\theta(S, y)$ admits a density function w.r.t. the Lebesgue measure. Therefore, for such S, y, θ , we obtain $\mathbb{P}^B[E_\theta(S, y) = 0] = 0$. This proves the announced result.

Now we take $S \neq \mathbf{0} \in \mathcal{S}^d$, $y \in \mathbb{R}$ and $\theta \neq 0$. In view of the expression of $E_\theta(S, y)$, we notice that showing $E_\theta(S, y)$ admits a density function is equivalent to show that $B_\theta^\top S B_\theta / \theta$ has a density function. The latter has the same distribution as $G^\top S G$ where G is a standard normal random vector. Consider the spectral decomposition of S :

$S = \sum_{i=1}^I \lambda_i p_i p_i^\top$ where p_i are orthonormal vectors and the eigenvalues λ_i are strictly positive. Since $S \neq \mathbf{0}$, $I \geq 1$. As a consequence, setting $\bar{G}_i = p_i^\top G$ we get

$$G^\top S G = \sum_{i=1}^I \lambda_i \bar{G}_i^2.$$

The components $\bar{G}_i$ are independent and distributed as standard normal random variables. To sum up, we have decomposed $G^\top S G$ as a weighted sum of independent χ_1^2 random variables. Therefore, $G^\top S G$ has a probability density given by the convolution of χ_1^2 random variables. $\square$

5.2 Proof of Proposition 3.2

Since we make estimates on $[t_n, t_{n+1}]$, we simply write u instead of $u^{(n+1)}$. In the following, $C_{n,N}(\xi)$ denotes a generic constant with polynomial growth in the variable ξ (see the definition in Section 3).

Before starting the analysis, we mention that a strategy of proof would be to use the Feynman-Kac (FK) representations. Here, we would represent the PDE (1.5) in terms of the SDE with zero drift and diffusion coefficient σ . Although natural, this approach is cumbersome at some points, especially, when one has to deal with the derivatives of the SDE w.r.t. the initial condition.

As an alternative, we take advantage of writing a FK formula using directly a Brownian motion. Let $\xi \in \mathbb{R}^d$, we set $\tilde{W}_t^{t_n, \xi} = \xi + W_t - W_{t_n}$, for all $t \in [t_n, t_{n+1}]$, where W is a d -dimensional Brownian motion. Now, the proof consists of applying Ito's formula to $u(\cdot, \tilde{W}_t^{t_n, \xi})$ and estimating $u(t_n, \xi) - v(t_{n+1}, \xi)$. Observe that $\tilde{W}_t^{t_n, \xi}$ is quite convenient for sensitivity computations because the first (resp. the second) order derivative of $\tilde{W}_t^{t_n, \xi}$ w.r.t. ξ is the identity matrix I_d (resp. the null array $\mathbf{0}$ in $\mathbb{R}^{d \times d \times d}$).

Proof of (3.9). Applying Ito's formula to $u(\cdot, \tilde{W}_t^{t_n, \xi})$ and taking the expectation leads to

$$\mathbb{E}[u(t_{n+1}, \tilde{W}_{t_{n+1}}^{t_n, \xi})] = u(t_n, \xi) + \mathbb{E}\left[\int_{t_n}^{t_{n+1}} (\partial_t u + 1/2 \operatorname{Tr}[D_x^2 u])(t, \tilde{W}_t^{t_n, \xi}) dt\right].$$

Since $D_x u(t, x)$ has a polynomial growth in space (because $D_x u(t, x) \in H_{\text{loc}, \text{pol}}^{1/2, 1}$), we have used that the stochastic integral is a martingale. By seeing that u is the solution of PDE (1.5) with terminal condition $u(t_{n+1}, \cdot) = v(t_{n+1}, \cdot)$, we get the FK representation

$$u(t_n, \xi) = \mathbb{E}\left[v(t_{n+1}, \tilde{W}_{t_{n+1}}^{t_n, \xi}) + \int_{t_n}^{t_{n+1}} g(t, \tilde{W}_t^{t_n, \xi}) dt\right], \quad (5.2)$$

where

$$g(t, x) := f(t, x, u(t, x), D_x u(t, x), D_x^2 u(t, x)) + \frac{1}{2} \operatorname{Tr}[(\sigma \sigma^\top)(t, x) - I_d](D_x^2 u)(t, x).$$

Subtracting $v(t_{n+1}, \xi)$ from Equation (5.2) leads to

$$u(t_n, \xi) - v(t_{n+1}, \xi) = \mathbb{E}\left[v(t_{n+1}, \tilde{W}_{t_{n+1}}^{t_n, \xi}) - v(t_{n+1}, \xi)\right] + \int_{t_n}^{t_{n+1}} \mathbb{E}\left[g(t, \tilde{W}_t^{t_n, \xi})\right] dt. \quad (5.3)$$

To obtain the announced results, we need to bound the derivatives of the expectation $\mathbb{E}[g(t, \tilde{W}_t^{t_n, \xi})]$. The following lemma provides an estimate in the interval $[t_n, t_{n+1}]$.

Lemma 5.4. Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)$ be a d -dimensional multi-index with $|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_d \geq 1$. For any fonction $\phi \in H_{loc, pol}^{1/2, 1}$, it holds

$$|\partial_\xi^\alpha \mathbb{E}[\phi(t, \tilde{W}_t^{t_n, \xi})]| \leq C_{n, N}(\xi)(t - t_n)^{(1-|\alpha|)/2}, \quad t_n < t \leq t_{n+1},$$

for some constant $C_{n, N}(\xi) \in C_{pol}$.

Under our assumptions, $g(t, x)$ and v are in $H_{loc, pol}^{1/2, 1}$. By using the estimates in Lemma 5.1 with $p = 1$, we get

$$\begin{aligned} |u(t_n, \xi) - v(t_{n+1}, \xi)| &\leq \|v\|_{H_{loc, pol}^{1/2, 1}} C_{n, N}(\xi) \varepsilon_N^{1/2} + \int_{t_n}^{t_{n+1}} C_{n, N}(\xi) dt \\ &\leq C_{n, N}(\xi) \varepsilon_N^{1/2}, \end{aligned}$$

because the term related to the integral converges to zero at rate ε_N .

Proof of (3.10). Now we estimate the first derivative of $u(t_n, \xi) - v(t_{n+1}, \xi)$ w.r.t. the initial condition ξ . Differentiating (5.3) w.r.t. ξ yields

$$\begin{aligned} D_x u(t_n, \xi) - D_x v(t_{n+1}, \xi) &= \mathbb{E} \left[D_x v(t_{n+1}, \tilde{W}_{t_{n+1}}^{t_n, \xi}) - D_x v(t_{n+1}, \xi) \right] \\ &\quad + \int_{t_n}^{t_{n+1}} D_\xi \mathbb{E} \left[g(t, \tilde{W}_t^{t_n, \xi}) \right] dt. \end{aligned} \quad (5.4)$$

From our assumptions, $g(t, x)$ and $D_x v$ are in $H_{loc, pol}^{1/2, 1}$. By using the estimates in Lemma 5.4 with $|\alpha| = 1$ in Lemma 5.1 with $p = 1$, we get

$$\begin{aligned} |D_x u(t_n, \xi) - D_x v(t_{n+1}, \xi)| &\leq \|D_x v\|_{H_{loc, pol}^{1/2, 1}} C_{n, N}(\xi) \varepsilon_N^{1/2} + \int_{t_n}^{t_{n+1}} C_{n, N}(\xi) dt \\ &\leq C_{n, N}(\xi) \varepsilon_N^{1/2}, \end{aligned}$$

where the term related to the integral converges to zero at rate ε_N . This implies the announced result.

Proof of (3.11). Analogously, we estimate the second derivative of $u(t_n, \xi) - v(t_{n+1}, \xi)$ by differentiating (5.4) w.r.t. ξ

$$\begin{aligned} D_x^2 u(t_n, \xi) - D_x^2 v(t_{n+1}, \xi) &= \mathbb{E} \left[D_x^2 v(t_{n+1}, \tilde{W}_{t_{n+1}}^{t_n, \xi}) - D_x^2 v(t_{n+1}, \xi) \right] \\ &\quad + \int_{t_n}^{t_{n+1}} D_\xi^2 \mathbb{E} \left[g(t, \tilde{W}_t^{t_n, \xi}) \right] dt. \end{aligned}$$

From our assumptions, $g(t, x)$ and $D_x^2 v$ are in $H_{loc, pol}^{1/2, 1}$. By using the estimates in Lemma 5.4 with $|\alpha| = 2$ in Lemma 5.1 with $p = 1$, we get

$$\begin{aligned} |D_x^2 u(t_n, \xi) - D_x^2 v(t_{n+1}, \xi)| &\leq \|D_x^2 v\|_{H_{loc, pol}^{1/2, 1}} C_{n, N}(\xi) \varepsilon_N^{1/2} + \int_{t_n}^{t_{n+1}} \frac{C_{n, N}(\xi)}{(t - t_n)^{1/2}} dt \\ &\leq C_{n, N}(\xi) \varepsilon_N^{1/2}, \end{aligned}$$

where the integral term is of magnitude $\varepsilon_N^{1/2}$. Therefore, we obtain the announced estimates.

Proof of Lemma 5.4. Let $|\alpha| \neq 0$. Differentiating $\mathbb{E}[\phi(t, \tilde{W}_t^{t_n, \xi})]$ w.r.t ξ yields

$$\partial_\xi^\alpha \mathbb{E}[\phi(t, \tilde{W}_t^{t_n, \xi})] = \int_{\mathbb{R}^d} \phi(t, x) \partial_\xi^\alpha p(t_n, \xi; t, x) dx \quad (5.5)$$

$$= \int_{\mathbb{R}^d} (\phi(t, x) - \phi(t_n, \xi)) \partial_\xi^\alpha p(t_n, \xi; t, x) dx, \quad (5.6)$$

where

$$p(t_n, \xi; t, x) := \frac{1}{(2\pi)^{d/2} (t - t_n)^{d/2}} \exp\left(-\frac{1}{2} \frac{\|x - \xi\|^2}{t - t_n}\right)$$

is the transition density function of $\tilde{W}_t^{t_n, \xi}$. To pass from (5.6) to (5.5), we have used that for any $|\alpha| \geq 1$ it holds

$$0 = \partial_\xi^\alpha \int_{\mathbb{R}^d} p(t_n, \xi; t, x) dx = \int_{\mathbb{R}^d} \partial_\xi^\alpha p(t_n, \xi; t, x) dx.$$

According to the result in [7, Section 6, Chapter 9] related to the bounds for the transition density function $p(t_n, \xi; t, x)$, it holds that there exist two positive constants c and C depending on α such that

$$|\partial_\xi^\alpha p(t_n, \xi; t, x)| \leq \frac{C}{(t - t_n)^{(|\alpha|+d)/2}} \exp\left(-c \frac{\|x - \xi\|^2}{t - t_n}\right),$$

for any x, ξ in $\mathbb{R}^d$ and any $t_n < t \leq t_{n+1}$. Due to $\phi \in H_{\text{loc}, \text{pol}}^{1/2, 1}$ and the previous estimate, we have

$$\begin{aligned} & \left| \partial_\xi^\alpha \mathbb{E}[\phi(t, \tilde{W}_t^{t_n, \xi})] \right| \\ & \leq C \|\phi\|_{H_{\text{loc}, \text{pol}}^{1/2, 1}} \int_{\mathbb{R}^d} (1 + \|x\|^q + \|\xi\|^q) \frac{|t - t_n|^{1/2} + \|x - \xi\|}{(t - t_n)^{d/2 + |\alpha|/2}} \exp\left(-c \frac{\|x - \xi\|^2}{t - t_n}\right) dx, \end{aligned}$$

for some real $q > 0$. Using the following estimate

$$|(1 + |y|^q) e^{-y^2 c}| \leq \tilde{C} e^{-y^2 \tilde{c}}, \quad y \in \mathbb{R},$$

for some positive constants $\tilde{C}$ and $\tilde{c} < c$, we get the existence of a new constant $\bar{C}$ such that

$$\begin{aligned} \left| \partial_\xi^\alpha \mathbb{E}[\phi(t, \tilde{W}_t^{t_n, \xi})] \right| & \leq \bar{C} (1 + \|\xi\|^p) (t - t_n)^{d/2} \frac{|t - t_n|^{1/2}}{(t - t_n)^{d/2 + |\alpha|/2}} \\ & = \bar{C} (1 + \|\xi\|^p) (t - t_n)^{(1 - |\alpha|)/2}, \end{aligned}$$

for any $t_n < t \leq t_{n+1}$. Therefore, we obtain the announced result. $\square$

References

- [1] F. ABERGEL AND N. MILLOT, *Nonquadratic local risk-minimization for hedging contingent claims in incomplete markets*, SIAM Journal on Financial Mathematics, 2 (2011), pp. 342–356.
- [2] B. BOUCHARD AND X. WARIN, *Monte-Carlo Valuation of American Options: facts and new algorithms to Improve existing methods*, in Numerical methods in finance, Springer, 2012, pp. 215–255.

- [3] P. CHRISTODOULOU, N. DETERING, AND T. MEYER-BRANDIS, *Quadratic hedging with multiple assets under illiquidity with applications in energy markets*, arXiv preprint arXiv:1705.06918, (2017).
- [4] S. CREPEY, *Financial modeling: a Backward Stochastic Differential Equations perspective*, Springer, Springer-Verlag, 2013.
- [5] N. EL KAROUI, S. PENG, AND M. C. QUENEZ, *Backward Stochastic Differential Equations in finance*, Mathematical finance, 7 (1997), pp. 1–71.
- [6] H. FÖLLMER AND M. SCHWEIZER, *Hedging by sequential regression: An introduction to the mathematics of option trading*, Astin Bulletin, 18 (1988), pp. 147–160.
- [7] A. FRIEDMAN, *Partial Differential Equations of Parabolic type*, Courier Dover Publications, 2008.
- [8] H. GEVRET, J. LELONG, AND X. WARIN, *STochastic OPTimization library in C++*, tech. rep., EDF Lab, 2016.
- [9] E. GOBET, J.-P. LEMOR, AND X. WARIN, *A regression-based Monte Carlo method to solve backward stochastic differential equations*, The Annals of Applied Probability, 15 (2005), pp. 2172–2202.
- [10] E. GOBET AND E. TEMAM, *Discrete time hedging errors for options with irregular payoffs*, Finance and Stochastics, 5 (2001), pp. 357–367.
- [11] E. GOBET AND P. TURKEDJIEV, *Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions*, Math. Comp., 85 (2016), pp. 1359–1391.
- [12] I. KARATZAS AND S. SHREVE, *Methods of mathematical finance*, vol. 39, Springer Science & Business Media, 1998.
- [13] J.-P. LEMOR, E. GOBET, AND X. WARIN, *Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations*, Bernoulli, 12 (2006), pp. 889–916.
- [14] S. PENG, *Nonlinear Expectations, Nonlinear Evaluations and Risk Measures*, in Stochastic Methods in Finance, Springer, 2004, pp. 165–253.
- [15] H. PHAM, *Dynamic L_p -hedging in discrete time under cone constraints*, SIAM Journal on Control and Optimization, 38 (2000), pp. 665–682.
- [16] B. POCHART AND J.-P. BOUCHAUD, *Option pricing and hedging with minimum local expected shortfall*, Quantitative Finance, 4 (2004), pp. 607–618.
- [17] M. POTTERS, J.-P. BOUCHAUD, AND D. SESTOVIC, *Hedged Monte-Carlo: low variance derivative pricing with objective probabilities*, Physica A: Statistical Mechanics and its Applications, 289 (2001), pp. 517–525.
- [18] D. REVUZ AND M. YOR, *Continuous martingales and Brownian motion*, vol. 293 of Grundlehren der Mathematischen Wissenschaften, 1999.
- [19] M. SCHWEIZER, *A Guided Tour through Quadratic Hedging Approaches*, in Option Pricing, Interest Rates and Risk Management, M. M. E. Jouini, J. Cvitanic, ed., Cambridge University Press, 1999, pp. 538–574.

Finance for Energy Market Research Centre

Institut de Finance de Dauphine, Université Paris-Dauphine

1 place du Maréchal de Lattre de Tassigny

75775 PARIS Cedex 16

www.fime-lab.org