University of Oxford

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The computation of Greeks with multilevel Monte Carlo

Michaelmas Term 2013

Author:
Sylvestre Burgos

College:
Lady Margaret Hall

May 20, 2014

Acknowledgements

First and foremost I wish to express my deepest gratitude to my supervisor, Prof. Mike Giles for being unfailingly supportive and for being an inspiration. Without his expertise, his dedication and outstanding guidance this thesis would never have been possible.

I am extremely grateful to the Man Group plc for their financial backing and for providing me with such an amazing work environment at the Oxford-Man Institute.

Special thanks go to William Chesters for his understanding and stimulation in the final stages of my thesis.

Thanks to the University of Oxford, Lady Margaret Hall, the common rooms and clubs for making these years so unique and enriching.

Thanks to the many unsung heroes of free software without whom I wouldn't have had the tools for writing this thesis.

On a more personal level, I would also like to mention the very special people I have the privilege to know both in Oxford and across the globe. I am greatly indebted to all of them for their kindness, their joviality, their wisdom and for all the things I have learnt from them. Although not mentioned individually, they will recognise themselves. To all of them: "Thanks for being part of my life".

Finally I want to thank my family for their love and for always supporting me in times of doubt. All I have and will accomplish is only possible thanks to them, the importance of their sacrifices could never be overstated. This work is for them.

The computation of Greeks with multilevel Monte Carlo

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy.

Sylvestre Burgos

Lady Margaret Hall

Michaelmas Term 2013

Abstract

In mathematical finance, the sensitivities of option prices to various market parameters, also known as the "Greeks", reflect the exposure to different sources of risk. Computing these is essential to predict the impact of market moves on portfolios and to hedge them adequately. This is commonly done using Monte Carlo simulations. However, obtaining accurate estimates of the Greeks can be computationally costly.

Multilevel Monte Carlo offers complexity improvements over standard Monte Carlo techniques. However the idea has never been used for the computation of Greeks. In this work we answer the following questions: can multilevel Monte Carlo be useful in this setting? If so, how can we construct efficient estimators? Finally, what computational savings can we expect from these new estimators?

We develop multilevel Monte Carlo estimators for the Greeks of a range of options: European options with Lipschitz payoffs (e.g. call options), European options with discontinuous payoffs (e.g. digital options), Asian options, barrier options and lookback options. Special care is taken to construct efficient estimators for non-smooth and exotic payoffs. We obtain numerical results that demonstrate the computational benefits of our algorithms.

We discuss the issues of convergence of pathwise sensitivities estimators. We show rigorously that the differentiation of common discretisation schemes for Ito processes does result in satisfactory estimators of the the exact solutions' sensitivities. We also prove that pathwise sensitivities estimators can be used under some regularity conditions to compute the Greeks of options whose underlying asset's price is modelled as an Ito process.

We present several important results on the moments of the solutions of stochastic differential equations and their discretisations as well as the principles of the so-called "extreme path analysis". We use these to develop a rigorous analysis of the complexity of the multilevel Monte Carlo Greeks estimators constructed earlier. The resulting complexity bounds appear to be sharp and prove that our multilevel algorithms are more efficient than those derived from standard Monte Carlo.

Contents

Chapter 1

Multilevel Monte Carlo Greeks

In mathematical finance, Monte Carlo methods are often used to compute the value of an option by estimating the expected value of an appropriately discounted payoff $\mathbb{E}(P)$. Here, P is a payoff function that depends on an underlying asset's scalar price $S(t)$ whose initial value $S(0)$ is given and whose evolution SDE on the time interval $[0, T]$ is of the form

$$\mathrm{d}S(t) = a_\theta(S, t)\,\mathrm{d}t + b_\theta(S, t)\,\mathrm{d}W_t\,, \tag{1.1}$$

where W_t is a Brownian motion. We also write $S_t := S(t)$. The drift $a_\theta(S, t)$ and the volatility $b_\theta(S, t)$ depend on a certain parameter θ. For brevity, we commonly ignore the subscript and just write

$$\mathrm{d}S(t) = a(S, t)\,\mathrm{d}t + b(S, t)\,\mathrm{d}W_t\,. \tag{1.2}$$

This is just one use of Monte Carlo in finance. In practice the prices of common contracts are often quoted and used to calibrate our market models; the option's sensitivities to market parameters, the so-called Greeks, reflect the exposure to different sources of risk. Computing these is essential to hedge portfolios and is therefore even more important than pricing the option itself. However, obtaining accurate estimates of the Greeks can be computationally costly, much more so than simply estimating prices. This is why our research focuses on getting fast and accurate estimates of Greeks through Monte Carlo simulations.

Multilevel Monte Carlo offers complexity improvements over standard Monte Carlo simulations. This has been proved in a range of settings, including option pricing. Nevertheless the idea has never been used for the computation of Greeks. In this work we answer the following questions: can multilevel Monte Carlo be useful in this setting? If so, how can we construct efficient estimators? Finally, what computational savings can we expect from these new estimators?

In this chapter, we begin by recalling basic principles of Monte Carlo pricing, then introduce the multilevel technique before presenting methods used to compute Greeks.

1.1 Monte Carlo Pricing

The first step to pricing an option via Monte Carlo is to simulate the underlying asset's evolution on the time interval considered.

1.1.1 Underlying asset's model and pricing

We write V_t, the value at time t of an option paying the payoff P at time T ($0 \leq t \leq T$). Basic asset pricing theory proves that under certain assumptions (absence of arbitrage, completeness of the market, see [HK04] for more details), we can associate to any numeraire (a traded asset serving as a unit of account) with a value N_t, a unique measure $\mathbb{N}$, the equivalent martingale measure such that $\frac{V_t}{N_t}$ is a martingale. We can then write

$$V_t = N_t \mathbb{E}_{\mathbb{N}} \left[\frac{V_T}{N_T} | \mathcal{F}_t \right] = N_t \mathbb{E}_{\mathbb{N}} \left[\frac{P\left(S_T\right)}{N_T} | \mathcal{F}_t \right] \tag{1.3}$$

where $\mathbb{E}_{\mathbb{N}}$ denotes the expectation with respect to the measure $\mathbb{N}$. In simple cases (simple equity options, deterministic riskless interest rate r_t), N_t is taken to be the money market account $B_t = \exp\left(\int_0^t r_s \mathrm{d}s \right)$, i.e. the value of a unit of currency invested at the riskless rate r between time 0 and t. With $\mathbb{B}$ the associated equivalent martingale measure, the present value of the option can then be written as

$$V_0 = \mathbb{E}_{\mathbb{B}} \left[\frac{B_0}{B_T} P\left(S_T\right) | \mathcal{F}_0 \right] = \frac{B_0}{B_T} \mathbb{E}_{\mathbb{B}} \left[P\left(S_T\right) | \mathcal{F}_0 \right] \tag{1.4}$$

where S_t's evolution equation under the equivalent martingale measure is (1.2), $\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W_t$ and $D_{(t,T)} = \frac{B_t}{B_T}$ can be thought of as a discount factor. That is, to price an option, we estimate the discounted expectation of the payoff (see also [WHD95] for more details).

In our study, we focus on the computation of $\mathbb{E}\left[P\left(S_T\right)|\mathcal{F}_0\right]$ and for the sake of brevity we ignore the discount factor $D_{(0,T)}$ in the payoff as it does not alter the techniques involved.

1.1.2 Path simulation and convergence modes

The Monte Carlo estimator of V_0 is obtained by simulating trajectories of the underlying asset S_t between the present time $t = 0$ and expiry T. We then compute the payoff's value corresponding to each of these simulations and estimate the expectation of the payoff by averaging the results.

To simulate the paths described by equation (1.2), we discretise the time interval $[0, T]$ into N regular time steps of width $h := \Delta t = T/N$. We associate the index n to the n-th discretisation time and write the discretisation of the solution of equation (1.2) between times $t_n = n\,h$ and $t_{n+1} = (n+1)\,h$ as

$$\widehat{S}_{n+1} = f_\theta\left(\widehat{S}_n, \Delta W_n\right) \tag{1.5}$$

where f_θ is some function dependent on the parameter θ, $\Delta W_n = W_{(n+1)h} - W_{nh}$ and $\widehat{S}_n, \widehat{S}_{n+1}$ correspond to the discretised approximations of S_{nh} and $S_{(n+1)h}$ respectively.

Two common discretisation schemes used for 1-dimensional stochastic differential equations are the Euler and the Milstein scheme (see [KP92] or [Mil95] for a detailed presentation of various discretisation schemes and their properties).

The Euler scheme is very straightforward; it consists in approximating the drift and the volatility by assuming they are constant on each interval $[t_n, t_{n+1}[$. Applied to equation (1.2), it yields

$$\widehat{S}_{(n+1)} = \widehat{S}_n + a_\theta\left(\widehat{S}_n, t_n\right) h + b_\theta\left(\widehat{S}_n, t_n\right) \Delta W_n \tag{1.6}$$

The Milstein scheme ([Mil79]) is derived from a higher order expansion of the volatility on the interval $[t_n, t_{n+1}[$. It can be written as

$$\begin{aligned}\widehat{S}_{(n+1)} = \widehat{S}_n &+ a_\theta\left(\widehat{S}_n, t_n\right) h + b_\theta\left(\widehat{S}_n, t_n\right) \Delta W_n \\ &+ \frac{1}{2} b_\theta\left(\widehat{S}_n, t_n\right) \frac{\partial b_\theta\left(\widehat{S}_n, t_n\right)}{\partial \widehat{S}_n} \left(\Delta W_n^2 - h\right)\end{aligned} \tag{1.7}$$

As explained below, this second scheme offers better convergence properties than the simple Euler scheme. However, note that its use would not be straightforward if we were to consider equations driven by multidimensional Brownian motions. Indeed, it would then involve the problematic simulation of iterated Itô integrals known as Lévy areas (see [RW01], [Wik01], [GL94]).

The weak error of a scheme is defined as

$$\mathbb{E}\left[P\left(S_T\right)\right] - \mathbb{E}\left[P\left(\widehat{S}_{T/h}\right)\right] \tag{1.8}$$

where the payoff value $P\left(S_T\right)$ depends on the exact solution of the SDE and its approximation $P\left(\widehat{S}_{T/h}\right)$ is based on the discretisation of S on the interval $[0, T]$. The weak order of convergence corresponds to the rate at which the weak error converges to 0 as we refine the discretisation of the path. If it is defined, it is the largest value $\alpha \in \mathbb{R}$ such that

$$\mathbb{E}\left[P\left(\widehat{S}_{T/h}\right)\right] - \mathbb{E}\left[P\left(S_T\right)\right] = O\left(h^\alpha\right) \tag{1.9}$$

as $h \to 0$ for a certain family of payoffs P. It corresponds to the bias of the payoff estimator resulting from the discretisation and it is therefore the value usually considered in most financial applications. Note that we also casually refer to $O(h^\alpha)$ as being the order of weak convergence as no ambiguity arises from doing so.

In practice, note that instead of just considering the weak error, we often end up studying the behaviour of the weak approximation error, defined as

$$\mathbb{E}\left[P\left(S_T\right)\right] - \mathbb{E}\left[\widehat{P}\left(\widehat{S}_{T/h}\right)\right] \tag{1.10}$$

where the function $\widehat{P}$ is an approximation of the payoff P.

The strong error of the scheme is usually defined as

$$\sqrt{\mathbb{E}\left[\left(\widehat{S}_{T/h} - S_T\right)^2\right]} \tag{1.11}$$

and the order of strong convergence, if it exists, is the largest value of $\beta \in \mathbb{R}$ such that

$$\sqrt{\mathbb{E}\left[\left(\widehat{S}_{T/h} - S_T\right)^2\right]} = O\left(h^\beta\right) \tag{1.12}$$

that is, a measure of the average error for each individual path. We also casually refer to $O\left(h^\beta\right)$ as being the order of strong convergence.

We will see in section 1.2.2 that the strong order of convergence, and therefore the discretisation scheme used, is one of the important parameters determining the efficiency of the multilevel Monte Carlo approach.

Under some smoothness conditions on the coefficients of the SDE (1.2), the weak order of convergence offered by the Euler scheme is $\alpha_{Euler} = 1$ for payoffs P that are piecewise smooth with polynomial growth and have a finite number of discontinuities (see [Tal84] in the case of smooth functions P and [BT95] for weaker conditions on P). This is for example the case for European options with Lipschitz payoffs like European calls/puts or with discontinuous payoffs like digital calls/puts. The strong order of convergence is $\beta_{Euler} = 1/2$ (see [KP92]).

Under higher order smoothness conditions on the coefficients of (1.2), the weak and the strong orders of convergence of the Milstein scheme are $\alpha_{Milstein} = 1$ and $\beta_{Milstein} = 1$ (see Theorem 3.4.3).

Note that an alternative definition of the order of strong convergence is sometimes found in the literature. It is then defined as the largest value $\beta \in \mathbb{R}$ such that

$$\mathbb{E}\left(\left|\widehat{S}_{T/h} - S_T\right|\right) = O\left(h^\beta\right) \tag{1.13}$$

The results we just presented still hold with this definition (see [KP92]).

In section 2.1.3, we will explain in detail how those schemes are applied to the Black and Scholes evolution SDE. In section 3.1, we will also explain how they can be applied to the joint evolution equations of an asset and its sensitivities.

1.1.3 Complexity of standard Monte Carlo Pricing

We now estimate the computational cost of pricing an option via standard Monte Carlo. Let us suppose we want to estimate an option price with an accuracy ϵ. The mean square error of the Monte Carlo estimator of the price can be decomposed into two terms. With $V = \mathbb{E}\left[P\left(S\right)\right]$ the true option value, $\widehat{V} = \mathbb{E}\left[\widehat{P}\left(\widehat{S}\right)\right]$ its discretised approximation and $\widehat{Y} = \dfrac{1}{M}\displaystyle\sum_{i=1}^{M} \widehat{P}\left(\widehat{S}\right)^{(i)}$ its Monte Carlo estimator, we can write the mean square error

$$\begin{aligned}
\mathbb{E}\left(\left(\widehat{Y} - V\right)^2\right) &= \mathbb{E}\left(\left(\widehat{Y} - \widehat{V} + \widehat{V} - V\right)^2\right) \\
&= \mathbb{E}\left[\left(\widehat{Y} - \widehat{V}\right)^2\right] + \mathbb{E}\left[\left(\widehat{V} - V\right)^2\right] + 2\,\mathbb{E}\left[\left(\widehat{Y} - \widehat{V}\right)\left(\widehat{V} - V\right)\right] \\
&= \mathbb{E}\left[\left(\widehat{Y} - \widehat{V}\right)^2\right] + \left(\widehat{V} - V\right)^2 \\
&= \frac{1}{M}\mathbb{V}\left(\widehat{P}\left(\widehat{S}\right)\right) + \left(\mathbb{E}\left[\widehat{P}\left(\widehat{S}\right)\right] - \mathbb{E}\left[P\left(S\right)\right]\right)^2
\end{aligned} \tag{1.14}$$

where the first term corresponds to the variance of the estimator and the second term to the bias due to the discretisation. The first term tends to 0 when we take an infinite number of samples M. The second term vanishes when we reduce the size of the time steps thanks to the weak convergence properties of the discretisation scheme.

In the case of a European option with a Lipschitz payoff, we usually take $\widehat{P}\left(\widehat{S}\right) = P\left(\widehat{S}\right)$. The weak order of convergence of the Euler scheme being 1 for such a payoff, it means that the mean square error is

$$\mathbb{E}\left(\left(\widehat{Y} - V\right)^2\right) = \frac{1}{M}\mathbb{V}\left(\widehat{P}\left(\widehat{S}\right)\right) + O\left(\frac{1}{N^2}\right)$$

To achieve $\left(\mathbb{E}\left[\widehat{P}\left(\widehat{S}\right)\right] - \mathbb{E}\left[P\left(S\right)\right]\right)^2 = O\left(\epsilon^2\right)$, we need $N = O\left(\epsilon^{-1}\right)$ time steps and to achieve $\dfrac{1}{M}\mathbb{V}\left(\widehat{P}\left(\widehat{S}\right)\right) = O\left(\epsilon^2\right)$, we need $M = O\left(\epsilon^{-2}\right)$. The computational complexity C being $O\left(N\,M\right)$, the total computational cost is finally $O\left(\epsilon^{-3}\right)$.

1.2 Multilevel Monte Carlo Setting

As shown in section 1.1.3, achieving a root-mean square error of $O\left(\epsilon\right)$ via standard Monte Carlo requires $O\left(\epsilon^{-2}\right)$ independent paths and $O\left(\epsilon^{-1}\right)$ time steps for discretisations with first order weak convergence, giving a total computational cost $O\left(\epsilon^{-3}\right)$. We now present Giles' multilevel Monte Carlo technique [Gil08b] and explain how it can reduce this cost to $O(\epsilon^{-2})$ under certain conditions.

1.2.1 Multilevel Monte Carlo's principle

The idea is to write the expected payoff with a fine discretisation using 2^{-L} uniform time steps as a telescopic sum. Let $\widehat{P}_l$ be the simulated payoff with a discretisation using 2^l uniform time steps. We then have

$$\mathbb{E}(\widehat{P}_L) = \mathbb{E}(\widehat{P}_0) + \sum_{l=1}^{L} \mathbb{E}(\widehat{P}_l - \widehat{P}_{l-1}) \tag{1.15}$$

We use Monte Carlo estimators using M_l independent samples to estimate each term of the sum on the right-hand side of (1.15)

$$Y_l := \mathbb{E}(\widehat{P}_l - \widehat{P}_{l-1}) \;\approx\; \widehat{Y}_l := \frac{1}{M_l} \sum_{i=1}^{M_l} \left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{(i)} \tag{1.16}$$

where $\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{(i)}$ is a small corrective term that comes from the difference between the fine and coarse discretisations of the path driven by the Brownian motion of the i-th sample. Its magnitude depends on the strong convergence properties of the scheme used.

To ensure a better efficiency we may modify (1.16) and use different estimators of $\widehat{P}$ on the fine and coarse levels of $\widehat{Y}_l$ as long as the telescoping sum property is respected, that is

$$\mathbb{E}(\widehat{P}_L^f) = \mathbb{E}(\widehat{P}_0^c) + \sum_{l=1}^{L} \mathbb{E}(\widehat{P}_l^f - \widehat{P}_{l-1}^c) \tag{1.17}$$

with

$$\mathbb{E}(\widehat{P}_l^f) = \mathbb{E}(\widehat{P}_l^c). \tag{1.18}$$

1.2.2 Complexity theorem

Let V_l be the variance of a single sample $\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)}$. The next theorem, originally introduced in [Gil08b] and improved in [CGST11], shows that what determines the efficiency of the multilevel approach is primarily the convergence speed of V_l as $l \to \infty$.

Theorem 1.2.1. *For a real-valued random variable P, we let $\widehat{P}_l$ be the corresponding approximation using the discretisation at level l, i.e. with 2^l steps of width $h_l = 2^{-l}T$.*

If there exist independent estimators $\widehat{Y}_l$ of computational complexity C_l based on M_l samples and there are positive constants $\alpha \geq \frac{1}{2}\min(1,\beta), \beta, c_1, c_2, c_3$ such that

$$
\begin{aligned}
&A1: \ \mathbb{E}(\widehat{Y}_l) = \begin{cases} \mathbb{E}(\widehat{P}_0) & \textit{if } l = 0 \\ \mathbb{E}(\widehat{P}_l - \widehat{P}_{l-1}) & \textit{if } l > 0 \end{cases} \\
&A2: \ \left| \mathbb{E}(\widehat{P}_l - P) \right| \leq c_1 h_l^{\alpha} \\
&A3: \ \mathbb{V}(\widehat{Y}_l) \leq c_2 h_l^{\beta} M_l^{-1} \\
&A4: \ C_l \leq c_3 M_l \, h_l^{-1}
\end{aligned}
\tag{1.19}
$$

Then there is a constant c_4 such that for any $\epsilon < e^{-1}$, there are values for L and $(M_l)_{l=0,\dots,L}$ resulting in a multilevel estimator $\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l$ with a mean-square-error $MSE = \mathbb{E}((\widehat{Y} - \mathbb{E}(P))^2) < \epsilon^2$ with a complexity C bounded by

$$
C \leq \begin{cases} c_4 \epsilon^{-2} & \textit{if } \beta > 1 \\ c_4 \epsilon^{-2} (\log \epsilon)^2 & \textit{if } \beta = 1 \\ c_4 \epsilon^{-2-(1-\beta)/\alpha} & \textit{if } 0 < \beta < 1 \end{cases}
\tag{1.20}
$$

Proof. See [Gil08b]. □

Constructing estimators with properties $A1$ and $A4$ is usually straightforward. In simple cases of pricing, we know α thanks to the literature on weak convergence. Results mentioned in section 1.1.2 and [GDR13] give $\alpha = 1$ for the Milstein scheme for typical European and Asian options.

The value of β depends on the payoff shape. It can be found for the various options considered here in [GDR13]. Giles explains in [Gil08b] that it actually determines where the computational effort is primarily expended: at the coarsest levels when $\beta > 1$ and at the finest levels when $\beta < 1$. In practice it is often the parameter that determines the efficiency of the multilevel approach. The main challenge is therefore to determine the value of β and to come up when possible with estimators resulting in higher values of this parameter.

Equation (1.20) shows that the efficiency of the multilevel approach improves as the rate of convergence of $\mathbb{V}\left(\widehat{Y}_l\right)$ increases, this suggests it is advisable to use discretisation schemes with high rates of strong convergence. Giles indeed shows in [Gil08a] that the use of the Milstein scheme yields better results than the Euler scheme.

1.2.3 A short overview of multilevel Monte Carlo research

Giles introduces in [Gil08b] the method and the fundamental complexity theorem 1.2.2. He provides a complexity analysis for European options with Lipschitz payoffs using the Euler scheme and numerical evidence of the scheme's behaviour for digital, Asian, barrier and lookback options. In [Gil08a] he demonstrates numerically that the use of carefully chosen multilevel estimators of the previous options in conjunction with the Milstein discretisation scheme yield higher orders of convergence and therefore improved computational benefits.

A rigorous justification of the convergence speeds observed with the Euler scheme is provided by Giles, Higham and Mao[GHM09], and independently for digital options by Avikainen in [Avi09a] while a complete analysis of the use of the Milstein scheme is provided by Giles, Debrabant and Roessler in [GDR13].

The multilevel technique has also been applied to the pricing of other classes of option. In [Gil09a], Giles shows it can easily be applied to efficiently price basket options. Belomestny and Schoenmakers propose in [BSD13] a multilevel Monte Carlo-based method to reduce the computational complexity of the dual method for pricing American options.

Burgos and Giles explain in [BG12] how to compute Greeks of common options using multilevel techniques and provide numerical evidence of the algorithms' efficiency.

Multilevel Monte Carlo and quasi Monte Carlo have been combined by Giles and Waterhouse in [GW09] as well as by Gerstner and Noll in [GN13].

Other recent developments in the field of multilevel Monte Carlo applied to Brownian SDEs include the work of Giles and Szpruch's [GS13], where it is shown that it is possible to apply the Milstein scheme to multi-dimensional SDEs while avoiding the costly simulation of Lévy areas by constructing antithetic estimators. Hoel, von Schwerin, Szepessy and Tempone in [HSST12] as well as Gerstner and Heinz in [GH13] also investigate the possibilities offered by adaptive non uniform time discretisations in a multilevel setting.

Multilevel Monte Carlo has also been applied to discontinuous processes. In [XG12], Xia and Giles present multilevel simulations with jump-diffusion SDEs and provide numerical evidence of their behaviour. Dereich and Heidenreich introduce and analyse in [DH11] an algorithm for a Lévy-driven stochastic differential equation. Ferreiro-Castilla and Van Schaik develop in [FCKSS13] a multilevel version of the Wiener-Hopf Monte Carlo [KKPS11], which they use to simulate the joint law of the position and running maximum of a Lévy process, which they apply to determine first passage times [FCvS13]. Multilevel Monte Carlo has also been applied to SPDEs, as seen in Giles' and Reisinger's paper [GR12].

Finally, let us note that multilevel algorithms have also been used by Bujok, Hambly and Reisinger for the pricing of basket credit derivatives [BHR12].

1.3 Monte Carlo simulation of Greeks

Now that we have presented how to estimate option prices via Monte Carlo simulations and how to make these estimates more efficient via multilevel Monte Carlo, we explain how to compute the Greeks, the prices' sensitivities. We briefly recall two classic methods used to compute Greeks in a Monte Carlo setting: the pathwise sensitivities and the Likelihood Ratio Method. More details can be found in [BG96], [L'E90] or [Gla04].

We consider $\widehat{S} = \left(\widehat{S}_n\right)_{n\in[0,N]}$ the discretised solution of equation 1.2 at times $(t_n)_{n=0,\dots,N}$ and $\left(\Delta\widehat{W}_n\right)_{n\in[0,N-1]}$ the corresponding set of Brownian increments. $\widehat{P}$ is the corresponding payoff estimator and $\widehat{V}$ the estimated value of the option.

1.3.1 Pathwise sensitivities

In the pathwise sensitivity approach, also known as Infinitesimal Perturbation Analysis (IPA) [L'E90], we assume that $\widehat{P}\left(\widehat{S}\right)$ is a K-Lipschitz function (i.e. there exists a uniform constant K such that $\left|\widehat{P}\left(\widehat{S}\right) - P(S)\right| < K\left\|\widehat{S} - S\right\|$ where $\left\|\widehat{S} - S\right\|$ is defined as $\sup_{t\in[0,T]}\left\|\widehat{S}_t - S_t\right\|$ for path-dependent options and as $\left\|\widehat{S}_N - S_T\right\|$ for European options), differentiable almost everywhere and that $\widehat{S}(\theta)$ is sufficiently regular in θ (see lemma 3.2.1 for more details) and then we can write

$$\frac{\partial\widehat{V}}{\partial\theta} = \frac{\partial\mathbb{E}\left(\widehat{P}\left(\widehat{S}\right)\right)}{\partial\theta} = \mathbb{E}\left[\frac{\partial\widehat{P}\left(\widehat{S}\right)}{\partial\theta}\right] \tag{1.21}$$

For simple payoffs, the sensitivity of $\widehat{V}$ to θ comes from the sensivity of $\widehat{S}$ to this same parameter. Using the chain rule, we can then write the pathwise estimator of the Greek as

$$\frac{\partial\widehat{V}}{\partial\theta} = \mathbb{E}\left[\sum_{n=0}^{N}\left(\frac{\partial\widehat{P}\left(\widehat{S}\right)}{\partial\widehat{S}_n}\frac{\partial\widehat{S}_n}{\partial\theta}\right)\right] \tag{1.22}$$

Note that if we assume that $\widehat{P}$ also has a direct dependency on θ, equation (1.22) becomes

$$\frac{\mathrm{d}\widehat{V}}{\mathrm{d}\theta} = \mathbb{E}\left[\sum_{n=0}^{N}\left(\frac{\partial\widehat{P}_\theta\left(\widehat{S}\right)}{\partial\widehat{S}_n}\frac{\partial\widehat{S}_n}{\partial\theta}\right) + \frac{\partial\widehat{P}_\theta\left(\widehat{S}\right)}{\partial\theta}\right] \tag{1.23}$$

We typically obtain the sensitivities $\frac{\partial\widehat{S}}{\partial\theta} = \left(\frac{\partial\widehat{S}_n}{\partial\theta}\right)_{n=0,\dots,N}$ by differentiating the discretisation of the evolution SDE. Indeed, for a discretisation scheme of the form

$$\widehat{S}_{k+1} = f\left(\theta, \widehat{S}_k, \Delta W_k\right) \tag{1.24}$$

a simple differentiation leads to

$$\frac{\partial\widehat{S}_{k+1}}{\partial\theta} = \frac{\partial f\left(\theta, \widehat{S}_k, \Delta W_k\right)}{\partial\theta} + \frac{\partial f\left(\theta, \widehat{S}_k, \Delta W_k\right)}{\partial\widehat{S}_k}\frac{\partial\widehat{S}_k}{\partial\theta} \tag{1.25}$$

which we can iterate between t_0 and t_n. This is explained and justified in detail in sections 2.2 and 3.1.

Usually $\widehat{S}(\theta)$ is regular enough to be differentiated (see section 3.2.4.2) and the main limitation of this technique is that it requires the payoff to be Lipschitz, a condition not met by some simple payoffs like that of the digital call.

Note that this approach lends itself well to the use of the so-called "adjoint method" presented in [GG06], which permits the simultaneous computation of multiple Greeks at a fixed computational cost.

1.3.2 Likelihood Ratio Method

The Likelihood Ratio Method [BG96], [L'E90] consists of writing that

$$V(\theta) = \mathbb{E}\left[\widehat{P}(\widehat{S})\right] = \int \widehat{P}(\widehat{S})\, p(\theta, \widehat{S})\, \mathrm{d}\widehat{S} \tag{1.26}$$

The dependence on θ comes through the probability density function $p(\theta, \widehat{S})$. As discussed in [Gla04], under relatively benign conditions ensuring the validity of the interchange of the order of integration and differentiation, we can write for a scheme of the form (1.24) that

$$\begin{aligned} \frac{\partial V}{\partial \theta} &= \int\limits_{\widehat{S}\in\mathbb{R}^N} \widehat{P}(\widehat{S})\, \frac{\partial p(\widehat{S})}{\partial \theta}\, \mathrm{d}\widehat{S} \\ &= \int\limits_{\widehat{S}\in\mathbb{R}^N} \widehat{P}(\widehat{S})\, \frac{\partial \log p(\widehat{S})}{\partial \theta}\, p(\widehat{S})\, \mathrm{d}\widehat{S} \\ &= \mathbb{E}\left[\widehat{P}(\widehat{S})\, \frac{\partial \log p(\widehat{S})}{\partial \theta}\right] \end{aligned} \tag{1.27}$$

with

$$\mathrm{d}\widehat{S} = \prod_{n=1}^{N} \mathrm{d}\widehat{S}_n \tag{1.28}$$

and

$$p(\widehat{S}) = \prod_{n=1}^{N} p_n \tag{1.29}$$

where $p_n = p(\widehat{S}_n | \widehat{S}_{n-1})$.

The main limitation of the method is that the estimator's variance is $O(N)$, becoming infinite as we refine the discretisation. For example if we assume that S_t's evolution follows a Geometric Brownian Motion, $\mathrm{d}S_t = r\, S_t\, \mathrm{d}t + \sigma\, S_t\, \mathrm{d}W_t$, its Euler discretisation between t_n and t_{n+1} results in the following transition density

$$\log(p_n) = -\log\left(\sigma \widehat{S}_n\right) - \frac{1}{2}\log(2\pi h) - \frac{\left(\widehat{S}_{n+1} - \widehat{S}_n\,(1 + r\,h)\right)^2}{2\sigma^2\, \widehat{S}_n^{\;2}\, h} \tag{1.30}$$

Letting Z_n the unit normal random variable defined by $\widehat{S}_{n+1} - \widehat{S}_n\,(1 + rh) = \sigma \widehat{S}_n \sqrt{h} Z_n$, we have

$$\frac{\partial \log p_n}{\partial \sigma} = \frac{Z_n^2 - 1}{\sigma} \tag{1.31}$$

and the approximation of Vega is

$$\frac{\partial \mathbb{E}\left[P\left(\widehat{S}_N\right)\right]}{\partial \sigma} = \mathbb{E}\left(\sum_{n=1}^{N} \frac{Z_n^2 - 1}{\sigma} P\left(\widehat{S}_N\right)\right) \tag{1.32}$$

noting that the increments are independent and that $\mathbb{V}\left[Z_n^2 - 1\right] = 2$, we get

$$\mathbb{V}\left(\sum_{n=1}^{N} \frac{Z_n^2 - 1}{\sigma} f\left(\widehat{S}_N\right)\right) = O\left(N\right) = O\left(T/h\right) \tag{1.33}$$

that is, we have a $O\left(h^{-1}\right)$ blow-up of the variance.

1.3.3 Beyond those methods

We have seen that neither the Likelihood Ratio Method nor pathwise sensitivities can be used universally. Even simple cases like the computation of the Greeks of a digital option using a path discretisation cannot be adequately covered by either (pathwise sensitivities not applicable, large variance of the Likelihood Ratio Method). We therefore need to devise alternative techniques.

To cope with payoff discontinuities and use pathwise sensitivities in more cases, one simple idea would consist in smoothing the payoff. A crude smoothing introduces a bias in our estimators and is clearly not satisfactory, a refinement of this idea is provided in section 2.3.1, it consists of stopping the simulation before expiry and then considering the conditional expectation of the payoff. Building on this idea, Giles introduced in [Gil09b] a hybrid of the Likelihood Ratio Method and pathwise sensitivities, the Vibrato Monte Carlo, which we present in section 2.3.2. Finally, other estimators using the tools of Malliavin calculus (see [Nua05] for example) have been developed. Those were introduced in [FLL+99a] and developed further in [FLL+01], [GKH03] or [Ben03] for example. It is also shown in [CG07] that such estimators are closely related to "classical" estimators but they are nevertheless beyond the scope of this thesis.

1.4 Multilevel Monte Carlo Greeks

By combining the elements of sections 1.2 and 1.3 together, we write

$$\frac{\partial V}{\partial \theta} = \frac{\partial \mathbb{E}(P)}{\partial \theta} \approx \frac{\partial \mathbb{E}(\widehat{P}_L)}{\partial \theta} = \frac{\partial \mathbb{E}(\widehat{P}_0)}{\partial \theta} + \sum_{l=1}^{L} \frac{\partial \mathbb{E}(\widehat{P}_l - \widehat{P}_{l-1})}{\partial \theta} \tag{1.34}$$

Similarly to equation (1.16), we define the multilevel estimators

$$\begin{aligned}\widehat{Y}_0 &= M_0^{-1}\sum_{i=1}^{M}\frac{\partial \widehat{P}_0}{\partial\theta}^{(i)}\\ \widehat{Y}_l &= M_l^{-1}\sum_{i=1}^{M_l}\left(\frac{\partial \widehat{P}_l}{\partial\theta}^{(i)} - \frac{\partial \widehat{P}_{l-1}}{\partial\theta}^{(i)}\right)\end{aligned} \tag{1.35}$$

where $\frac{\partial \widehat{P}_0}{\partial\theta}$, $\frac{\partial \widehat{P}_{l-1}}{\partial\theta}$, $\frac{\partial \widehat{P}_l}{\partial\theta}$ can be computed with the techniques presented in section 1.3. The detail of how these techniques are effectively used together is presented for different cases (various payoffs, various methods) in chapter 2.

1.5 Plan of the thesis

We begin our study of multilevel Monte Carlo Greeks with an experimental part. In chapter 2, we consider a common evolution model for the asset S, the Black - Scholes model under which S is modeled as a geometric Brownian motion. We recall that in this particular case, many properties of the underlying price process are known, which facilitates the computation of option prices and their sensitivities, many of which can be expressed as closed form formulae.

We use this convenient setting to introduce and implement various multilevel algorithms for the computation of Greeks. While we often have closed form expressions for the desired quantities and we can actually perform exact path simulations (geometric Brownian motions can be integrated directly between two different dates), we avoid using these properties that are specific to the Black - Scholes model for any purpose other than verification. The point of this chapter is to introduce various estimators and get some experimental insight into their efficiency, therefore we only use methods that can be applied in the general Ito setting presented in chapter 1.

We choose to apply our techniques to various common contracts presenting different challenges: the European call option, the European digital call with a discontinuous payoff, the Asian, lookback and barrier calls for which the payoffs depend on the whole trajectory of the underlying asset's price and not just its value at expiry T.

The main focus of chapter 2 being essentially the introduction of new ideas and obtaining experimental results on their efficiency, we first make what we deem to be sensible assumptions as to what is applicable and what is not without immediately proving their well-foundedness. Closed form formulas then enable us to check the proposed techniques behave as expected.

In chapter 3, we justify a posteriori why our assumptions in chapter 2 are actually correct. Notably, we check the naive technique used to obtain the underlying asset's sensitivities does indeed correspond to a proper discretisation scheme of the stochastic equation describing the evolution of the Greeks and that the discretised sensitivities therefore converge to the exact sensitivities. We then proceed to prove that under reasonably relaxed hypotheses on the coefficients of the evolution SDE (1.2), a Lipschitz payoff means that pathwise sensitivities can be applied (as had been assumed previously).

We then discuss some additional assumptions on the volatility that make the numerical analysis smoother and proceed to present several essential theorems that are used throughout the numerical analysis of multilevel Monte Carlo Greeks (chapters 4, 5, 6, 7 and 8). Those theorems give results on the moments of SDE solutions, on their approximation by continuous interpolations of the Milstein scheme and on the likelihood of so-called “extreme paths”.

In chapter 4, we provide the first proofs for the computational cost of multilevel Greeks. We use pathwise sensitivities to compute the sensitivities of a European option with a smooth Lipschitz payoff. This is a “toy” problem that doesn’t actually correspond to commonly traded contracts but is convenient to introduce some basic ideas (notably this case does not require us to deal with “extreme paths” separately). The efficiency of the multilevel approach is determined by the coefficients α and β of theorem 1.2.1. Our analysis of this case and other cases therefore aims at computing these values.

We then move on to a more realistic payoff type, European options with continuous, yet non-smooth payoffs like the European call whose payoff function is $P(S_T) = (S_T - K)^+$. The payoff’s derivative with respect to the underlying asset’s value S_T being possibly discontinuous, we now have to take special care of “extreme paths”, which can be roughly described as paths with “unusually” large random increments (this is properly formalised in chapter 3) when computing the coefficients α and β.

As explained in chapters 2 and 3, pathwise sensitivities cannot be applied directly to discontinuous payoffs like that of a European digital call whose payoff is $P(S_T) = \mathbf{1}_{[0,T]}$ and we use a variation of pathwise sensitivities, pathwise sensitivities with conditional expectations, which leads to a quite technical analysis of the estimators’ convergence speed. This technique can be seen as a particular form of payoff smoothing and can also be used with non-smooth Lipschitz payoffs like that of the European call. An analysis of this use of pathwise sensitivities with conditional expectations confirms what we had already conjectured from our simulations: the method offers an increased convergence speed compared to simple pathwise sensitivities by compensating for the effects of the discontinuity of the payoff’s first derivative.

We then provide analytical proofs confirming the observed behaviours of “split pathwise sensitivities” and Vibrato Monte Varlo, which both derive from pathwise sensitivities with conditional expectations. This analysis also offers a better insight into the way we should choose the number of final samples for both of these methods.

In chapter 6, we analyse the convergence properties of the multilevel estimators we proposed for Asian options, that is options that depend on the average value of the underlying asset on the time interval $[0, T]$ instead of just its final value $\widehat{S}_T$. The analysis for this sort of payoff is slightly tricky as it is strongly path dependent and estimators involve approximations of the underlying asset's price over the whole time period considered; it therefore requires the use of stronger results than the ones used previously.

Chapter 7 deals with the analysis of multilevel Greeks for barrier options. We consider a down-and-out option where the payoff is $P(S) = (S_T - K)^+ \mathbf{1}_{\min\limits_{t\in[0,T]} S_t > B}$. The discontinuity resulting from the barrier is dealt with by considering the probability of hitting it knowing some intermediate simulated values; in this it is similar to the idea behind pathwise sensitivities with conditional expectations. Nevertheless, as for Asian options, this analysis relies on results that hold on the whole time interval and not just at a fixed point in time.

Finally in chapter 8, we analyse the case of lookback options for which the payoff is $P = S_T - \min\limits_{t\in[0,T]} S_t$. We begin by providing a semi-analytical explanation for the behaviour of its multilevel Greeks' estimators. It highlights the fact that for pricing, the fine and coarse levels of discretisation reaching their respective minima at different times in $[0, T]$ has little impact on the convergence speed of the estimator's variance. Nevertheless, this possibility leads to greatly reduced convergence speed of the Greeks' estimators. A fully analytical approach to this problem is difficult, we therefore analyse another a lookback option with a discretely sampled minimum. When the number of samples is high, its payoff is very similar to that of the continuously sampled lookback option; a more thorough explanation of this last point can be found in [BGK99].

Chapter 2

Simulations

In this chapter, we investigate the joint use of the techniques presented in chapter 1, that is, the ways we can define and implement multilevel Monte Carlo Greeks estimators for various options.

The payoffs considered present various challenges: the digital call's discontinuous payoff prevents a direct use of pathwise sensitivities, Asian and lookback options are path dependent and naive estimators for the computation of prices and sensitivities with the Milstein scheme will generally generate sub-optimal results. Finally, barrier options are path dependent and their payoffs are discontinuous in the underlying asset's extrema.

We present various ideas enabling the techniques introduced in section 1.3 to be exploited in a multilevel setting: building on the relevant multilevel pricing methods found in [BG12], we derive efficient Greeks estimators.

We then use our implementations to get experimental convergence rates for the multilevel estimators, which we then relate to their computational complexity via theorem 1.2.1.

Finally, we present intuitive interpretations of the observed convergence rates. Those give an interesting insight into the ideas underlying the analysis of chapters 4 to 8.

2.1 Setting

In this section, for the sake of generality, we discuss the different techniques in the generic setting where option prices depend on an underlying asset paying no dividends whose value satisfies an Itô diffusion (1.2) under the risk-neutral measure. We still have to pick a specific case for the implementation. For the sake of simplicity and verifiability, we use the well-known Black & Scholes model, in which (1.2) is a geometric Brownian motion and for which the literature provides a number of analytical results.

2.1.1 The Black & Scholes SDE

We perform the simulations in the popular Black & Scholes model for which

$$\begin{aligned} a(S_t, t) &= rS_t \\ b(S_t, t) &= \sigma S_t \end{aligned} \tag{2.1}$$

that is, the evolution equation (1.2) on $[0, T]$ becomes

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t, \tag{2.2}$$

where r is the (constant) riskless interest rate and $\sigma > 0$ the (constant) volatility parameter.

This equation corresponds to a geometric Brownian motion. Several arguments justify its use to model a stock price: we see it is a random process that, when starting from a positive initial value, only takes positive values and that the expected returns are independent of the underlying stock's price. It is clearly not an entirely realistic representation of the stock's evolution, the most obvious shortcomings being the assumption of a constant volatility and the normality of the model's returns (real returns usually have fatter tails than the normal distribution, as shown in [Fam76] for example), nevertheless the simplicity of the above equation and its flexibility (several objections to this model can be addressed in a relatively simple way, see [Wil10]) ensure it is still one of the most common models in use.

SDE (2.2) is convenient in several ways: the coefficients $a(S_t, t)$ and $b(S_t, t)$ are smooth and Lipschitz (we will see in chapter 3 why this is important) and, more importantly, we can get a closed form formula for the density of its solution and therefore for the price and sensitivities of simple options, which will be useful for verification purposes.

2.1.2 Black & Scholes price formulas of common contracts

We briefly recall some well-known analytical results for equation (2.2) and derive the price of some important options.

The SDE corresponds to a lognormal process and can be integrated between 0 and T into

$$S_T = S_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right) T + \sigma\, W_T\right) \tag{2.3}$$

which has a lognormal density

$$p\left(S_T | S_0\right) = \frac{1}{\sigma S_0 \sqrt{2\pi T}} \exp\left(-\frac{\left(\log\left(S_T/S_0\right) - \left(r - \frac{\sigma^2}{2}\right) T\right)^2}{2\sigma^2 T}\right) \tag{2.4}$$

As explained in section 1.1.1, we can get the call's value at $t = 0$ for an initial underlying's price $S_0 = S$ by computing the discounted expectation of call's payoff

$P_{call}(S_T) = (S_T - K)^+$ under the risk neutral measure and we get

$$V_{call}(S,0) = \exp(-rT) \int_0^\infty (S_T - K)^+ \; p(S_T|S_0 = S) \, \mathrm{d}S_T \tag{2.5}$$

Then, defining

$$\begin{aligned} d_1 &= \frac{\log(S/K) + \left(r + \sigma^2/2\right) T}{\sigma\sqrt{T}} \\ d_2 &= \frac{\log(S/K) + \left(r - \sigma^2/2\right) T}{\sigma\sqrt{T}} \end{aligned} \tag{2.6}$$

it can be shown that

$$V_{call}(S,0) = S \; \Phi(d_1) - K \exp(-rT) \; \Phi(d_2) \tag{2.7}$$

and by differentiation,

$$\begin{aligned} \Delta_{call}(S,0) &= \frac{\partial V_{call}}{\partial S} = \Phi(d_1) \\ \nu_{call}(S,0) &= \frac{\partial V_{call}}{\partial \sigma} = S\sqrt{T}\phi(d_1) \end{aligned} \tag{2.8}$$

where Φ is the normal cumulative distribution function and ϕ the normal density function.

Similarly for the digital call, an integration of $P_{digital}(S_T) = \mathbf{1}_{S_T > K}$ gives

$$V_{digital}(S,0) = \exp(-rT) \; \Phi(d_2) \tag{2.9}$$

and

$$\begin{aligned} \Delta_{digital}(S,0) &= \frac{\partial V_{digital}}{\partial S} = \frac{\exp(-rT)\,\phi(d_2)}{\sigma S\sqrt{T}} \\ \nu_{digital}(S,0) &= \frac{\partial V_{digital}}{\partial \sigma} = -\exp(-rT)\,\phi(d_2)\left(\sqrt{T} + \frac{d_2}{\sigma}\right) \end{aligned} \tag{2.10}$$

Note that using Ito's lemma and no arbitrage arguments, we can show that the price $V(S_t, t)$ of a vanilla option on S satisfies the Black & Scholes equation (2.11)

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0 \tag{2.11}$$

The prices V_{call} and $V_{digital}$ can thus also be derived as solutions of this partial differential equation (2.11) using the payoffs $V(S_T, T) = P_{call}$ and $V(S_T, T) = P_{digital}$ as terminal conditions.

For a down-and-out barrier call option with a strike K and a barrier B, we let V_{call} be the Black & Scholes price of a vanilla call with the same strike and expiry as the barrier option considered. We then have

$$V_{barrier}\left(S_t,t\right)=V_{call}\left(S_t,t\right)-\left(\frac{S_t}{B}\right)^{1-(2r/\sigma^2)}V_{call}\left(\frac{B^2}{S_t},t\right)\tag{2.12}$$

We can check (see [Wil07] for example) that this formula still satisfies the Black & Scholes equation (2.11) before the barrier is hit and that the appropriate boundary conditions are respected when $S=B$ or $t=T$. This means that (2.12) is indeed the Black & Scholes price of the barrier option. In particular,

$$V_{barrier}\left(S,0\right)=V_{call}\left(S,0\right)-\left(\frac{S}{B}\right)^{1-(2r/\sigma^2)}V_{call}\left(\frac{B^2}{S},0\right)\tag{2.13}$$

The Asian option satisfies a different partial differential equation, indeed its price V_{asian} is also a function of $I_t=\int\limits_0^t S_t\mathrm{d}t$. Using a reasoning similar to that used to derive (2.11), we get a modified Black & Scholes equation

$$\frac{\partial V_{asian}}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2V_{asian}}{\partial S^2}+rS\frac{\partial V_{asian}}{\partial S}-rV_{asian}+S\frac{\partial V_{asian}}{\partial I}=0\tag{2.14}$$

the solution cannot be written as an analytical formula and has to be found numerically.

The Black & Scholes price for a lookback option also admits a closed form formula which can be found in [MR05] or [Wil07].

$$\begin{aligned}V_{lookback}\left(S,0\right)&=S\Phi\left(d_1\right)-S\exp\left(-rT\right)\Phi\left(d_2\right)\\&\quad+S\exp\left(-rT\right)\frac{\sigma^2}{2r}\Phi\left(-d_1+\frac{2r\sqrt{T}}{\sigma}\right)-S\frac{\sigma^2}{2r}\Phi\left(-d_1\right)\end{aligned}\tag{2.15}$$

2.1.3 Discretisation scheme

As explained in section 2.1.2, the Black & Scholes evolution SDE (2.2) can be integrated exactly into (2.3) and we can perform exact simulations of the underlying asset's price at $t=T$. There is therefore a priori no real need to discretise (2.2) on the interval $[0,T]$ to estimate the price of a vanilla European option via Monte Carlo simulations. It would be sufficient to simulate the value of the Brownian motion at expiry, $W_T\sim\mathcal{N}\left(0,T\right)$ to simulate various values of the payoff $P\left(S_T\right)$ and estimate the option's value via the formula

$$V=\mathbb{E}\left[\exp(-rT)P\left(S_0\,\exp\left(\left(r-\frac{\sigma^2}{2}\right)T+\sigma\,W_T\right)\right)\right]\tag{2.16}$$

For path-dependent options where we need to compute some intermediate values, we could also use an exact integration of the lognormal Black & Scholes SDE on each subdivision $[t_n, t_{n+1}]$ of $[0, T]$,

$$S_{n+1} = S_n \exp\left(\left(r - \frac{\sigma^2}{2}\right)(t_{n+1} - t_n) + \sigma\, \Delta W_n\right) \tag{2.17}$$

where the Brownian increment $\Delta W_n = W_{t_{n+1}} - W_{t_n} \sim \mathcal{N}(0, t_{n+1} - t_n)$. It would also be possible to use the known properties of the geometric Brownian motion on each of these intervals (e.g. distribution of its minimum) to derive more accurate estimators.

Nevertheless, a direct integration of the evolution SDE is not possible in more general cases of (1.2) (e.g. local volatility function) and the use of a proper discretisation then becomes necessary. To illustrate multilevel techniques designed to work in a general setting, we therefore treat the Black & Scholes equation as any other SDE and discretise it on the considered time interval.

The simplest discretisation of (1.2) is the Euler discretisation. Assuming a constant time step $h = t_N - t_{N-1} = \ldots = t_1 - t_0$ between all points of the discretisation, it is written

$$\widehat{S}_{n+1} = \widehat{S}_n + r\,\widehat{S}_n h + \sigma\,\widehat{S}_n \Delta W_n \tag{2.18}$$

As discussed in [Gil08a] and recalled in 1.2.2, the Euler scheme's strong convergence properties are not entirely satisfactory in the context of multilevel Monte Carlo simulations and it is preferable to use the Milstein scheme instead, which is then

$$\widehat{S}_{n+1} = \widehat{S}_n + r\,\widehat{S}_n h + \sigma\,\widehat{S}_n \Delta W_n + \frac{1}{2}\sigma^2\,\widehat{S}_n\left(\Delta W_n^2 - h\right) \tag{2.19}$$

2.1.4 Principle of the numerical simulations

As stated above, the goal is to develop ideas that enable multilevel Monte Carlo to work efficiently in conjunction with the techniques used for the computation of Greeks. We devise and implement multilevel estimators of Greeks based on discretisation (2.19) for the various options considered (European Lipschitz payoffs, European discontinuous payoffs, Asian options, Barrier options, Lookback options).

We recall the superscripts/subscripts f and c denote the values corresponding specifically to the "fine" and "coarse" levels at any given level l of the simulation. When considering the estimator at a given level l, we can drop the index l from our notation for the sake of conciseness as it does not result in any ambiguity. Using the notation of section 1.4, the multilevel estimation of the price sensitivity is written as

$$\frac{\partial V}{\partial \theta} \approx \widehat{Y}_0 + \sum_{l=1}^{L} \widehat{Y}_l \tag{2.20}$$

where $\widehat{Y}_l \approx \mathbb{E}\left(\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\right)$ is defined using fine and coarse discretisations with $N_f(l) = 2^l$ fine time steps and $N_c(l) = 2^{l-1}$ corresponding coarse time steps of size $h := h_f(l) = T/2^l$ and $h_c(l) = T/2^{l-1}$ respectively.

Using the multilevel Monte Carlo complexity theorem, we see that to evaluate the sensivitity $\partial V/\partial\theta$ at a low computational cost equates to constructing estimators $\widehat{Y}_l$ satisfying conditions $A1 - A4$ of theorem 1.2.1 with $\alpha \geq 1/2$ and β as large as possible (ideally $\beta > 1$ to get maximum benefits).

To determine the efficiency of our estimators for each payoff, we consider several levels of discretisation l. Using M_l path simulations, we estimate the the corresponding values $\mathbb{E}\left(\widehat{Y}_l\right)$ and $V_l = M_l\,\mathbb{V}\left(\widehat{Y}_l\right)$. We then plot the results in log-log plots where the x-axis corresponds to the level l (the binary logarithm of the number of fine steps) and the y-axis corrresponds to $\log_2 \mathbb{E}\left(\widehat{Y}_l\right)$ and $\log_2 V_l$. Assuming we have

$$\begin{aligned}\mathbb{E}\left(\widehat{Y}_l\right) &= O\left(h^{\alpha}\right)\\ V_l &= O\left(h^{\beta}\right)\end{aligned} \tag{2.21}$$

we get

$$\begin{aligned}\log_2 \mathbb{E}\left(\widehat{Y}_l\right) &\sim \alpha\,\log_2\left(h\right) \sim -l\,\alpha\\ \log_2 V_l &\sim \beta\,\log_2\left(h\right) \sim -l\,\beta\end{aligned} \tag{2.22}$$

therefore we can estimate the coefficients α and β directly by "measuring" the asymptotic slopes of the graphs resulting from our simulations (i.e. performing a linear regression on the values described above).

The result of section 3.2.4 guarantees that the α measured this way is indeed the same as the one of theorem 1.2.1.

Note that we are interested only in the asymptotic behaviour of the estimators, therefore it may be necessary to exclude low values of l from our slope measurements/linear regressions if the graphs exhibit a non-linear behaviour.

Note also that a certain degree of uncertainty comes with the measurement of these slopes (each point on which the regression is based is a random variable). We obtain reasonable estimates by taking the numbers of samples M_l large enough at each level to keep $\sqrt{\mathbb{V}\left(\widehat{Y}_l\right)} \ll \widehat{Y}_l$.

Unless otherwise stated, the simulations used in this dissertation use the parameters $S_0 = 100$, $K = 100$, $r = 0.05$, $\sigma = 0.20$, $T = 1$.

2.2 Lipschitz payoffs (European call)

We first consider a Lipschitz payoff, that of the European call for which (including the discount in the payoff),

$$P = \exp\left(-rT\right)\left(S_T - K\right)^{+} = \exp\left(-rT\right)\max\left(0, S_T - K\right)$$

We illustrate the techniques by computing delta (Δ) and vega (ν), the sensitivities to the asset's initial value S_0 and to its volatility σ.

2.2.1 Pathwise sensitivities

Since the payoff is Lipschitz, the rule of thumb presented in section 1.3.1 ensures we can use pathwise sensitivities (see section 3.2 for a rigorous justification).

We write the discretisation (2.19) as

$$\widehat{S}_{n+1} = \widehat{S}_n\, D_n \tag{2.23}$$

with

$$D_n := 1 + r\,h + \sigma\,\Delta W_n + \frac{\sigma^2}{2}\,(\Delta W_n^2 - h). \tag{2.24}$$

The differentiation of this scheme gives

$$\frac{\partial \widehat{S}_{n+1}}{\partial \theta} = \frac{\partial \widehat{S}_n}{\partial \theta} \cdot D_n + \widehat{S}_n \frac{\partial D_n}{\partial \theta} \tag{2.25}$$

In particular for Δ,

$$\begin{aligned} \frac{\partial \widehat{S}_0}{\partial S_0} &= 1 \\ \frac{\partial \widehat{S}_{n+1}}{\partial S_0} &= \frac{\partial \widehat{S}_n}{\partial S_0} \cdot D_n \end{aligned} \tag{2.26}$$

and for ν, we have

$$\begin{aligned} \frac{\partial \widehat{S}_0}{\partial \sigma} &= 0 \\ \frac{\partial \widehat{S}_{n+1}}{\partial \sigma} &= \frac{\partial \widehat{S}_n}{\partial \sigma} \cdot D_n + \widehat{S}_n \left(\Delta W_n + \sigma(\Delta W_n^2 - h)\right) \end{aligned} \tag{2.27}$$

Note that as explained later in section 3.1, the differentiation of the discretisation scheme actually corresponds to the discretisation scheme for the sensitivity $\dfrac{\partial S_t}{\partial \theta}$. This means this "naive" technique does indeed provide estimators of the sensitivities and we also get the same weak and strong convergence properties for the discretisation of the asset's value and for its sensitivities.

To compute $\widehat{Y}_l$, we take M_l samples and noticing the call payoff has no direct dependency on θ,

$$\widehat{Y}_l = \frac{1}{M_l} \sum_{i=1}^{M_l} \left[\left(\frac{\partial P}{\partial S}\left(\widehat{S}^f_{N_f}\right) \frac{\partial \widehat{S}^f_{N_f}}{\partial \theta} \right) - \left(\frac{\partial P}{\partial S}\left(\widehat{S}^c_{N_c}\right) \frac{\partial \widehat{S}^c_{N_c}}{\partial \theta} \right) \right]^{(i)} \tag{2.28}$$

As explained for example in [Gil08b], the fine and coarse levels correspond to two discretisations of the same driving Brownian motion. In [ANvST14] it is shown that for the Milstein scheme, defining the fine level as having twice as many steps as the coarse level is near-optimal. This is easily implemented by first generating the fine Brownian increments $\widehat{W}_f$ and then summing them in pairs to get the coarse level's increments $\widehat{W}_c$, that is,

$$
\begin{aligned}
\widehat{W}^f &= \left(\Delta W_0^f, \Delta W_1^f, \ldots, \Delta W_{N_f-1}^f\right) \\
\widehat{W}^c &= \left(\Delta W_0^c, \ldots, \Delta W_{N_c-1}^c\right) = \left(\Delta W_0^f + \Delta W_1^f, \ldots, \Delta W_{N_f-2}^f + \Delta W_{N_f-1}^f\right)
\end{aligned}
\tag{2.29}
$$

Estimated complexity

In figures 2.1, 2.2 we carry out the procedure explained in section 2.1.4: we plot $\mathbb{E}\left(\widehat{Y}_l\right)$ and V_l as a function of the level l, where we take $\widehat{Y}_l$ to be successively the multilevel estimator for the option's value, its delta and its vega.

The slopes of the graphs give numerical estimates of the parameters α and β of theorem 1.2.1, which then gives an estimated complexity of the multilevel algorithm (see table 2.1).

"Intuitive" interpretation

Giles has shown in [Gil08a] that $\alpha = 1$ and $\beta = 2$ for the estimators of the value of a Lipschitz European payoff (and therefore for the estimators of the value of the European call).

Indeed, as explained in section 1.1.2, the Milstein scheme gives $O(h)$ weak and $O(h)$ strong convergence. At expiry, we have

$$
\mathbb{E}\left(P\left(\widehat{S}_N\right) - P\left(S_T\right)\right) = O(h) \tag{2.30}
$$

and

$$
\mathbb{E}\left(\left(\widehat{S}_N - S_T\right)^2\right) = O\left(h^2\right) \tag{2.31}
$$

The payoff P being 1-Lipschitz , we have

$$
\left(P\left(\widehat{S}_N\right) - P\left(S_T\right)\right)^2 < \left(\widehat{S}_N - S_T\right)^2 \tag{2.32}
$$

and equation (2.31) leads to

$$
\mathbb{V}\left[P\left(\widehat{S}_N\right) - P\left(S_T\right)\right] = O\left(h^2\right) \tag{2.33}
$$

Using the convergence properties of the sensitivity at expiry (proved rigorously in section 3.1), we also obtain

$$
\begin{aligned}
\mathbb{E}\left(\frac{\partial \widehat{S}_N}{\partial \theta} - \frac{\partial S_T}{\partial \theta}\right) &= O(h) \\
\mathbb{E}\left(\left|\frac{\partial \widehat{S}_N}{\partial \theta} - \frac{\partial S_T}{\partial \theta}\right|\right) &= O(h)
\end{aligned}
\tag{2.34}
$$

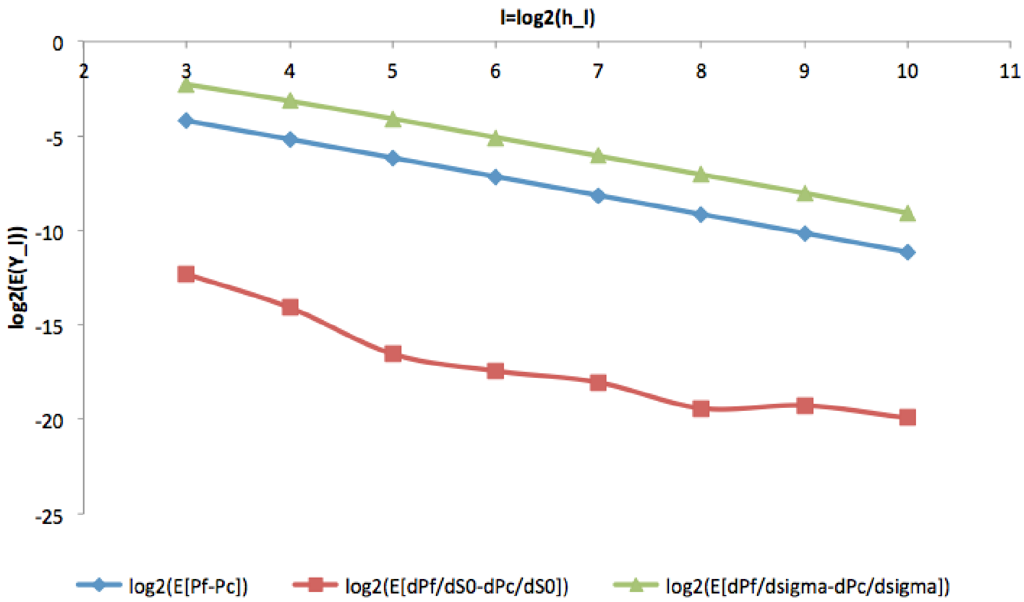

Figure 2.1: Pathwise sensitivities, European call : $\mathbb{E}(\widehat{Y}_l)$

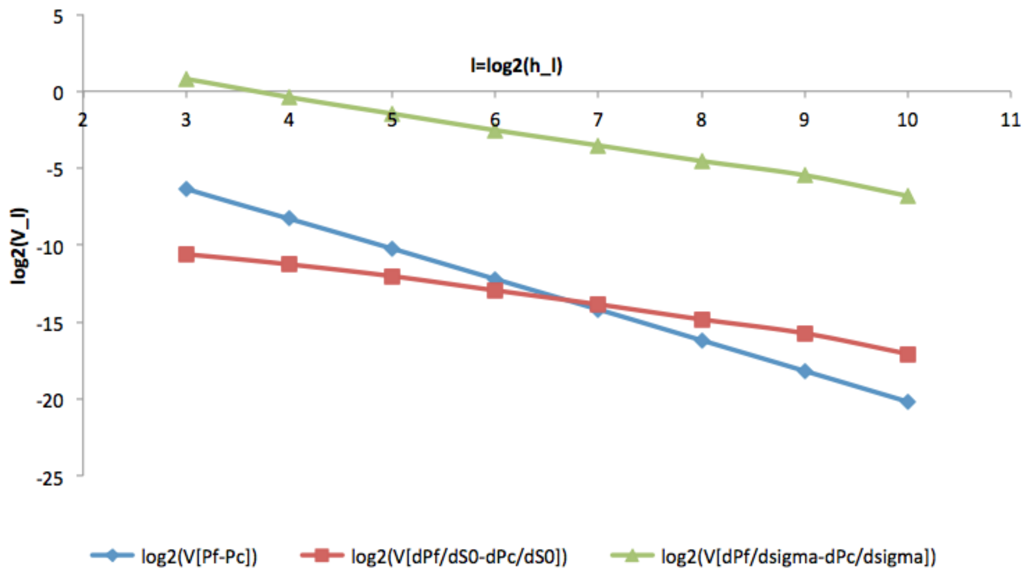

Figure 2.2: Pathwise sensitivities, European call : V_l

Estimator	α	β	MLMC Complexity
Value	≈ 1.0	≈ 2.0	$O(\epsilon^{-2})$
Delta	≈ 1.0	≈ 0.9	$O(\epsilon^{-2.1})$
Vega	≈ 1.0	≈ 1.1	$O(\epsilon^{-2})$

Table 2.1: Pathwise sensitivities, European call : estimated complexity

We explain here how what we observe is similar to what is described for the pricing of discontinous options in [GHM09]: the convergence of Greeks' estimators is degraded by the lack of smoothness of the payoff, i.e. by the discontinuity of $\frac{\partial P}{\partial S}(S_T) = \exp(-rT)\,\mathbf{1}_{S_T>K}$.

Note that this chapter being dedicated to providing an intuitive explanation of the observed convergence rates, we here use the notation $O(.)$ in an informal way to express the characteristic size of various quantities. Rigorous and more formal proofs are provided in later chapters.

We see from equation (2.4) that in the Black & Scholes model, the distribution of S_T is smooth (in the more general setting of an Ito diffusion 1.2 we could assume it is, see section 3.3 for a rigorous justification of this fact). Therefore, a fraction $O(h)$ of all paths are such that the final value S_T is at a distance $O(h)$ from the discontinuity K. The order of strong convergence being $O(h)$, this means that for such paths, there is a $O(1)$ probability that the fine and coarse discretisations $\widehat{S}^f_{N_f}$ and $\hat{S}^c_{N_c}$ are on different sides of the strike K. Then we write

$$\begin{aligned}
&\left(\frac{\partial P}{\partial S}\left(\widehat{S}^f_{N_f}\right)\frac{\partial \widehat{S}^f_{N_f}}{\partial \theta}\right) - \left(\frac{\partial P}{\partial S}\left(\widehat{S}^c_{N_c}\right)\frac{\partial \widehat{S}^c_{N_c}}{\partial \theta}\right) \\
&\qquad = \frac{\partial P}{\partial S}\left(\widehat{S}^f_{N_f}\right)\left(\frac{\partial \widehat{S}^f_{N_f}}{\partial \theta} - \frac{\partial \widehat{S}^c_{N_c}}{\partial \theta}\right) + \left(\frac{\partial P}{\partial S}\left(\widehat{S}^f_{N_f}\right) - \frac{\partial P}{\partial S}\left(\widehat{S}^c_{N_c}\right)\right)\frac{\partial \widehat{S}^c_{N_c}}{\partial \theta} \\
&\qquad = O(1)\,O(h) + O(1)\,O(1) \\
&\qquad = O(1)
\end{aligned} \tag{2.35}$$

The majority of paths (i.e. a fraction $O(1)$) arrive further from K and the strong convergence of the scheme intuitively means their fine and coarse discretisations are bound to be on the same side of the strike where $\frac{\partial P}{\partial S}$ is Lipschitz (actually constant). Therefore for those paths,

$$\left(\frac{\partial P}{\partial S}\left(\widehat{S}^f_{N_f}\right)\frac{\partial \widehat{S}^f_{N_f}}{\partial \theta}\right) - \left(\frac{\partial P}{\partial S}\left(\widehat{S}^c_{N_c}\right)\frac{\partial \widehat{S}^c_{N_c}}{\partial \theta}\right) = O(h) \tag{2.36}$$

Finally we can use the law of total expectation and, with a slight abuse use of the $O(\ldots)$ notation (see notes in 3.5), write that

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial P}{\partial S}\left(\widehat{S}^f_{N_f}\right)\frac{\partial \widehat{S}^f_{N_f}}{\partial \theta}\right) - \left(\frac{\partial P}{\partial S}\left(\widehat{S}^c_{N_c}\right)\frac{\partial \widehat{S}^c_{N_c}}{\partial \theta}\right)\right] &= O(h)\,O(1) + O(1)\,O(h) \\
&= O(h)
\end{aligned} \tag{2.37}$$

and

$$\mathbb{E}\left[\left(\left(\frac{\partial P}{\partial S}\left(\widehat{S}^f_{N_f}\right)\frac{\partial \widehat{S}^f_{N_f}}{\partial \theta}\right)-\left(\frac{\partial P}{\partial S}\left(\widehat{S}^c_{N_c}\right)\frac{\partial \widehat{S}^c_{N_c}}{\partial \theta}\right)\right)^2\right] \tag{2.38}$$
$$= O(h)\, O(1) + O(1)\, O\left(h^2\right)$$
$$= O(h)$$

and thus $V_l = O(h)$. We have thus explained why we expected $\alpha = 1$ and $\beta = 1$ for the Greeks.

2.2.2 Pathwise sensitivities and Conditional Expectations

We have seen that the payoff's lack of smoothness prevents the variance of Greeks' estimators $\widehat{Y}_l$ from decaying quickly and limits the potential benefits of the multilevel approach.

To improve the convergence speed in the case of an option whose payoff only depends on the final value of the underlying asset, we can use conditional expectations to smooth the payoff without introducing any additional bias (see [Gla04] for the basic idea and [Gil08a] for its use in the multilevel context). Let us consider a single level of discretisation with a step $h = T/N$: instead of simulating the whole path from $t_0 = 0$ to $t_N = T$, we simulate it up to the penultimate discretisation time $t_{N-1} = (N-1)\,h$. We get $\widehat{S}_{N-1}$ using the Milstein scheme with the Brownian increments $\widehat{W} = (\Delta W_0, \ldots, \Delta W_{N-2})$. We then consider the full distribution of $\left(\widehat{S}_N | \widehat{S}_{N-1}\right)$ resulting from the use of the Euler scheme based on a Brownian increment ΔW_{N-1} on the last time step $[t_{N-1}, t_N]$.

With

$$\begin{aligned} a_{N-1} &= a\left(\widehat{S}_{N-1}, t_{N-1}\right) \\ b_{N-1} &= b\left(\widehat{S}_{N-1}, t_{N-1}\right) \end{aligned} \tag{2.39}$$

we can write

$$\widehat{S}_N = \widehat{S}_{N-1} + a_{N-1}h + b_{N-1}\Delta W_{N-1} \tag{2.40}$$

this means we get a Normal distribution for $\left(\widehat{S}_N | \widehat{S}_{N-1}\right)$.

Proposition 2.2.1. *Using the Euler scheme for the final step does not degrade the superior strong convergence of the Milstein scheme. Indeed, under the assumptions guaranteeing that the Milstein scheme's order of convergence is* 1 *(see section 3.4 for more details), the strong order of convergence of such a hybrid scheme is also* 1.

Proof. Let $\widehat{S}^{Mil}_N$ be the final simulated value resulting from a "pure" Milstein discretisation of the process on $[0, T]$, the classical convergence results can be applied to $\widehat{S}_N^{Mil}$ and we can write

$$\widehat{S}_N = \widehat{S}_N^{Mil} - \frac{1}{2} b_{N-1} \frac{\partial b_{N-1}}{\partial \widehat{S}_{N-1}} \left(\Delta W_{N-1}^2 - h\right) \tag{2.41}$$

then

$$\begin{aligned}
\mathbb{E}\left[\left(S_T - \widehat{S}_N\right)^2\right] &= \mathbb{E}\left[\left(S_T - \widehat{S}_N^{Mil} + \widehat{S}_N^{Mil} - \widehat{S}_N\right)^2\right] \\
&= \mathbb{E}\left[\left(S_T - \widehat{S}_N^{Mil}\right)^2 + \left(\widehat{S}_N^{Mil} - \widehat{S}_N\right)^2\right. \\
&\qquad \left. +2\left(S_T - \widehat{S}_N^{Mil}\right)\left(\widehat{S}_N^{Mil} - \widehat{S}_N\right)\right]
\end{aligned} \tag{2.42}$$

using the convergence properties of the Milstein scheme, we have

$$\mathbb{E}\left[\left(S_T - \widehat{S}_N^{Mil}\right)^2\right] = O\left(h^2\right) \tag{2.43}$$

also, using Hölder's inequality

$$\begin{aligned}
\mathbb{E}\left[\left(\widehat{S}_N^{Mil} - \widehat{S}_N\right)^2\right] &= \mathbb{E}\left[\frac{1}{4} b_{N-1}^2 \left(\frac{\partial b_{N-1}}{\partial \widehat{S}_{N-1}}\right)^2 \left(\Delta W_{N-1}^2 - h\right)^2\right] \\
&\leq \frac{1}{4} \mathbb{E}\left[{b_{N-1}}^8\right]^{1/4} \mathbb{E}\left[\left(\frac{\partial b_{N-1}}{\partial \widehat{S}_{N-1}}\right)^8\right]^{1/4} \mathbb{E}\left[\left(\Delta W_{N-1}^2 - h\right)^4\right]^{1/2}
\end{aligned} \tag{2.44}$$

Under the assumptions presented in section 3.4, the solution of SDE (1.2) and its discretisation have finite moments. Under these assumptions, $b\left(S,t\right)$ and $\dfrac{\partial b}{\partial S}\left(S,t\right)$ have a linear growth bound. Therefore, $\mathbb{E}\left[{b_{N-1}}^8\right]^{1/4}$ and $\mathbb{E}\left[\left(\dfrac{\partial b_{N-1}}{\partial \widehat{S}_{N-1}}\right)^8\right]^{1/4}$ are finite. $\Delta W_{N-1}^2 - h =: h\ \left(Z_{N-1}^2 - 1\right)$ where Z_{N-1} is a unit normal random variable, therefore $\mathbb{E}\left[\left(\Delta W_{N-1}^2 - h\right)^4\right]^{1/2} = O\left(h^2\right)$. Thus we have

$$\mathbb{E}\left[\left(\widehat{S}_N^{Mil} - \widehat{S}_N\right)^2\right] = O\left(h^2\right) \tag{2.45}$$

Finally,

$$\begin{aligned}
&\mathbb{E}\left[\left(S_T - \widehat{S}_N^{Mil}\right)\left(\widehat{S}_N^{Mil} - \widehat{S}_N\right)\right] = \\
&\qquad \mathbb{E}\left[\left(S_T - \widehat{S}_N^{Mil}\right)\frac{1}{2} b_{N-1}\left(\frac{\partial b_{N-1}}{\partial \widehat{S}_{N-1}}\right)\left(\Delta W_{N-1}^2 - h\right)\right]
\end{aligned} \tag{2.46}$$

and similarly, using Hölder's inequality, the strong convergence property of $\widehat{S}_N^{Mil}$, the finite moments of b_{N-1} and $\dfrac{\partial b_{N-1}}{\partial S}$ and Z_{N-1}, we obtain

$$\mathbb{E}\left[\left(S_T - \widehat{S}_N^{Mil}\right)\left(\widehat{S}_N^{Mil} - \widehat{S}_N\right)\right] = O\left(h^2\right) \tag{2.47}$$

which achieves the proof that

$$\mathbb{E}\left[\left(S_T - \widehat{S}_N\right)^2\right] = O\left(h^2\right) \tag{2.48}$$

i.e. the combination of the Milstein scheme with a final Euler step still results in a strong convergence of order 1. □

Using the final Euler step, the probability density of $\widehat{S}_N$ conditional on $\widehat{S}_{N-1}$ is

$$p(\widehat{S}_N|\widehat{S}_{N-1}) = \frac{1}{\sigma_{N-1}\sqrt{2\pi}}\exp\left(-\frac{\left(\widehat{S}_N - \mu_{N-1}\right)^2}{2\sigma_{N-1}^2}\right) \tag{2.49}$$

$$\text{with}\ \begin{cases} \mu_{N-1} = \widehat{S}_{N-1} + a_{N-1}h \\ \sigma_{N-1} = b_{N-1}\sqrt{h} \end{cases}$$

We can thus compute $\mathbb{E}\left[P\left(\widehat{S}_N\right)|\widehat{S}_{N-1}\right]$. Using the tower property, we get

$$\widehat{V} = \mathbb{E}\left[P(\widehat{S}_N)\right] = \mathbb{E}_{\widehat{S}_{N-1}}\left[\mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right]\right] \tag{2.50}$$

Using $\widehat{S}_{N-1}^{(1)}, \dots, \widehat{S}_{N-1}^{(M)}$, M simulations of $\widehat{S}_{N-1}$, we get

$$\widehat{V} \approx \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\left[P\left(\widehat{S}_N\right)|\widehat{S}_{N-1}^{(m)}\right] \tag{2.51}$$

Let us now compute an analytical expression for $\mathbb{E}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right]$.)

$$\begin{aligned}
\mathbb{E}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right] &= \int_{-\infty}^{\infty} P\left(\widehat{S}_N\right) p\left(\widehat{S}_N\left|\widehat{S}_{N-1}\right.\right) \mathrm{d}\widehat{S}_N \\
&= \int_{K}^{\infty}\left(\widehat{S}_N - K\right)\frac{1}{\sqrt{2\pi}\sigma_{N-1}}\exp\left(-\frac{\left(\widehat{S}_N - \mu_{N-1}\right)^2}{2\sigma_{N-1}^2}\right)\mathrm{d}\widehat{S}_N \\
&= \int_{K}^{\infty}\frac{\widehat{S}_N - \mu_{N-1}}{\sqrt{2\pi}\sigma_{N-1}}\exp\left(-\frac{\left(\widehat{S}_N - \mu_{N-1}\right)^2}{2\sigma_{N-1}^2}\right)\mathrm{d}\widehat{S}_N \\
&\quad + \int_{K}^{\infty}\frac{\mu_{N-1} - K}{\sqrt{2\pi}\sigma_{N-1}}\exp\left(-\frac{\left(\widehat{S}_N - \mu_{N-1}\right)^2}{2\sigma_{N-1}^2}\right)\mathrm{d}\widehat{S}_N
\end{aligned} \tag{2.52}$$

We let ϕ be the normal probability density function and Φ be the normal cumulative distribution functions and obtain

$$\mathbb{E}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right] = \sigma_{N-1}\phi\left(\frac{\mu_{N-1}-K}{\sigma_{N-1}}\right) + (\mu_{N-1}-K)\Phi\left(\frac{\mu_{N-1}-K}{\sigma_{N-1}}\right) \tag{2.53}$$

This expected payoff is smooth with respect to the input parameters μ_{N-1} and σ_{N-1}, which themselves are smooth functions of the input parameters and $\widehat{S}_{N-1}$.

We can apply the pathwise sensitivities technique to this Lipschitz function at time t_{N-1}.

$$\begin{aligned}\frac{\partial \widehat{V}}{\partial \theta} &= \frac{\partial \mathbb{E}_{\widehat{S}_{N-1}}\left[\mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right]\right]}{\partial \theta} \\ &= \mathbb{E}_{\widehat{S}_{N-1}}\left[\frac{\partial \mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right]}{\partial \theta}\right] \\ &= \mathbb{E}_{\widehat{S}_{N-1}}\left[\frac{\partial \mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right]}{\partial \mu_{N-1}}\frac{\partial \mu_{N-1}}{\partial \theta}\right. \\ &\qquad \left. + \frac{\partial \mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right]}{\partial \sigma_{N-1}}\frac{\partial \sigma_{N-1}}{\partial \theta}\right]\end{aligned} \tag{2.54}$$

where $\dfrac{\partial \mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right]}{\partial \mu_{N-1}}$, $\dfrac{\partial \mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)|\widehat{S}_{N-1}\right]}{\partial \sigma_{N-1}}$ result from a direct differentiation of equation (2.53) and $\dfrac{\partial \mu_{N-1}}{\partial \theta}$, $\dfrac{\partial \sigma_{N-1}}{\partial \theta}$ are easily obtained via Pathwise Sensitivities.

In a multilevel setting, the estimator of the Greek is written as

$$\widehat{Y}_l = \frac{1}{M_l}\sum_{i=1}^{M_l}\left[\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\right]^{(i)} \tag{2.55}$$

where $\widehat{P}_f$ and $\widehat{P}_c$ correspond to the smoothed payoff functions based on the conditional expectations of $P\left(\widehat{S}^f_{N_f}\right)$ and $P\left(\widehat{S}^c_{N_c}\right)$ respectively.

At the fine level we simulate $\widehat{S}^f_{N_f-1}$, compute $a_{N_f-1} = a\left(\widehat{S}^f_{N_f-1}, t^f_{N_f-1}\right)$, $b_{N_f-1} = b\left(\widehat{S}^f_{N_f-1}, t^f_{N_f-1}\right)$ and $\mu^f_{N_f-1}$, $\sigma^f_{N_f-1}$ the corresponding values defined in (2.49).

$$\begin{aligned}\mu^f_{N_f-1} &= \widehat{S}_{N-1} + a_{N_f-1}h_f \\ \sigma^f_{N_f-1} &= b_{N_f-1}\sqrt{h_f}\end{aligned} \tag{2.56}$$

In the Black & Scholes model,

$$\begin{aligned}\mu^f_{N_f-1} &= (1+rh_f)\,\widehat{S}^f_{N_f-1} \\ \sigma^f_{N_f-1} &= \sigma S^f_{N_f-1}\sqrt{h_f}\end{aligned} \tag{2.57}$$

Equation (2.53) then gives the analytical expression of $\widehat{P}_f := \mathbb{E}\left[P\left(\widehat{S}^f_{N_f}\right)\right]$,

$$\widehat{P}_f = \sigma^f_{N_f-1}\phi\left(\frac{\mu^f_{N_f-1}-K}{\sigma^f_{N_f-1}}\right) + (\mu^f_{N_f-1}-K)\Phi\left(\frac{\mu^f_{N_f-1}-K}{\sigma^f_{N_f-1}}\right) \tag{2.58}$$

It is a function of $\mu^f_{N_f-1}$ and $\sigma^f_{N_f-1}$. We can then write

$$\frac{\partial\widehat{P}_f}{\partial\theta} = \frac{\partial\mu^f_{N_f-1}}{\partial\theta}\frac{\partial\widehat{P}_f}{\partial\mu^f_{N_f-1}} + \frac{\partial\sigma^f_{N_f-1}}{\partial\theta}\frac{\partial\widehat{P}_f}{\partial\sigma^f_{N_f-1}} \tag{2.59}$$

where $\dfrac{\partial\mu^f_{N_f-1}}{\partial\theta}$ and $\dfrac{\partial\sigma^f_{N_f-1}}{\partial\theta}$ are computed via Pathwise Sensitivities.

Numerical experiments have shown that defining directly $\widehat{P}_c = \mathbb{E}\left(P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1}\right)$ at the coarse level leads to an unsatisfactorily low convergence rate of $\mathbb{V}\left(\widehat{Y}_l\right)$ (intuitively the problem comes from using the diffusion of the underlying process on an interval that is too wide). As explained in (1.17) we can use a modified estimator. To achieve better convergence rates of the variance, we include the knowledge of the final fine Brownian increment in the computation of the conditional expectation over the last coarse Brownian increment. This helps ensure that for a given path there isn't too much discrepancy between the fine and coarse payoff estimators, thereby keeping the variance low.

Using the Euler scheme, $\widehat{S}^c_{N_c}$ is distributed as if it were the value at time T of a simple Brownian motion with constant drift $a\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)$ and volatility $b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)$ on the final coarse step $\left[t^c_{N_c-1}, T\right]$ with value $\widehat{S}^c_{N_c-1}$ at $t^c_{N_c-1}$. Given that the fine Brownian increment on the first half of the final step is $\Delta W^f_{N_f-2}$, we get for such a process the following density:

$$p_c\left(\widehat{S}^c_{N_c}\middle|\widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2}\right) = \frac{1}{\sigma^c_{N_c-1}\sqrt{2\pi}}\exp\left(-\frac{\left(\widehat{S}^c_{N_c}-\mu^c_{N_c-1}\right)^2}{2{\sigma^c_{N_c-1}}^2}\right) \tag{2.60}$$

with

$$\begin{aligned}\mu^c_{N_c-1} &= \widehat{S}^c_{N_c-1} + a\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)h_c + b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)\Delta W^f_{N_f-2} \\ \sigma^c_{N_c-1} &= b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)\sqrt{h_c/2} = b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)\sqrt{h_f}\end{aligned} \tag{2.61}$$

From this distribution we derive that $\mathbb{E}\left[P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2}\right]$ can be expressed via the same payoff formula as before, (2.53) applied to $\mu^c_{N_c-1}$ and $\sigma^c_{N_c-1}$.

$$\begin{aligned}\widehat{P}_c &:= \mathbb{E}\left[P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2}\right] \\ &= \sigma^c_{N_c-1}\phi\left(\frac{\mu^c_{N_c-1}-K}{\sigma^c_{N_c-1}}\right) + (\mu^c_{N_c-1}-K)\Phi\left(\frac{\mu^c_{N_c-1}-K}{\sigma^c_{N_c-1}}\right)\end{aligned} \tag{2.62}$$

In the Black & Scholes model,

$$\begin{aligned} \mu^c_{N_c-1} &= (1+rh_c)\,\widehat{S}^c_{N_c-1} + \sigma \widehat{S}^c_{N_c-1}\Delta W^f_{N_f-2} \\ \sigma^c_{N_c-1} &= \sigma \widehat{S}^c_{N_c-1}\sqrt{h_f} \end{aligned} \tag{2.63}$$

Note that using $\widehat{P}_c$ instead of $\mathbb{E}\left(P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1}\right)$ does not introduce any bias. Indeed, using the tower property on the last coarse time step, we can check that the two expressions have the same expectation.

$$\begin{aligned} &\mathbb{E}_{\Delta W^f_{N_f-2}}\left[\mathbb{E}\left[P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1},\Delta W^f_{N_f-2}\right]|\widehat{S}^c_{N_c-1}\right] \\ &\qquad\qquad = \mathbb{E}_{\Delta W^c_{N_c-1}}\left[P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1}\right] \end{aligned} \tag{2.64}$$

Therefore $\widehat{P}_f$ and $\widehat{P}_c$ satisfy the telescoping sum condition (1.18).

We then apply pathwise sensitivities to $\widehat{P}_c$ and obtain a formula similar to (2.59).

$$\frac{\partial \widehat{P}_c}{\partial\theta} = \frac{\partial \mu^c_{N_c-1}}{\partial\theta}\frac{\partial \widehat{P}_c}{\partial \mu^c_{N_c-1}} + \frac{\partial \sigma^c_{N_c-1}}{\partial\theta}\frac{\partial \widehat{P}_c}{\partial \sigma^c_{N_c-1}} \tag{2.65}$$

where $\dfrac{\partial \mu^c_{N_c-1}}{\partial\theta}$ and $\dfrac{\partial \sigma^c_{N_c-1}}{\partial\theta}$ are computed via Pathwise Sensitivities.

2.2.2.1 Estimated complexity

Our numerical experiments (figures 2.3, 2.4) show the benefits of the conditional expectation technique on the European call: we observe higher convergence rates β which translate into lower complexities (table 2.2).

2.2.2.2 "Intuitive" interpretation

Using the fact that S_T has a smooth density function (see (2.4)), a fraction $O(\sqrt{h})$ of the paths arrive in the area of width $O\left(\sqrt{h}\right)$ around the strike where the first order derivative of the conditional expectation $\dfrac{\partial\mathbb{E}\left(P\left(\widehat{S}_N\right)|\widehat{S}_{N-1}\right)}{\partial \widehat{S}_{N-1}}$ transitions from being almost 0 to almost 1. In this area of width $O\left(\sqrt{h}\right)$, $\dfrac{\partial^2\mathbb{E}\left(P\left(\widehat{S}_N\right)|\widehat{S}_{N-1}\right)}{\partial \widehat{S}^2_{N-1}}$ is therefore intuitively of order $O\left(h^{-1/2}\right)$ and elsewhere it is almost 0. The strong convergence properties of the discretisation scheme imply that the coarse and fine paths differ by $O\left(h\right)$; we thus have $O\left(h\right)O\left(h^{-1/2}\right)$ difference between the fine and coarse values of $\dfrac{\partial\mathbb{E}\left(P\left(\widehat{S}_N\right)|\widehat{S}_{N-1}\right)}{\partial \widehat{S}_{N-1}}$ which results in a $O\left(\sqrt{h}\right)$ difference between the coarse and fine Greeks' estimates.

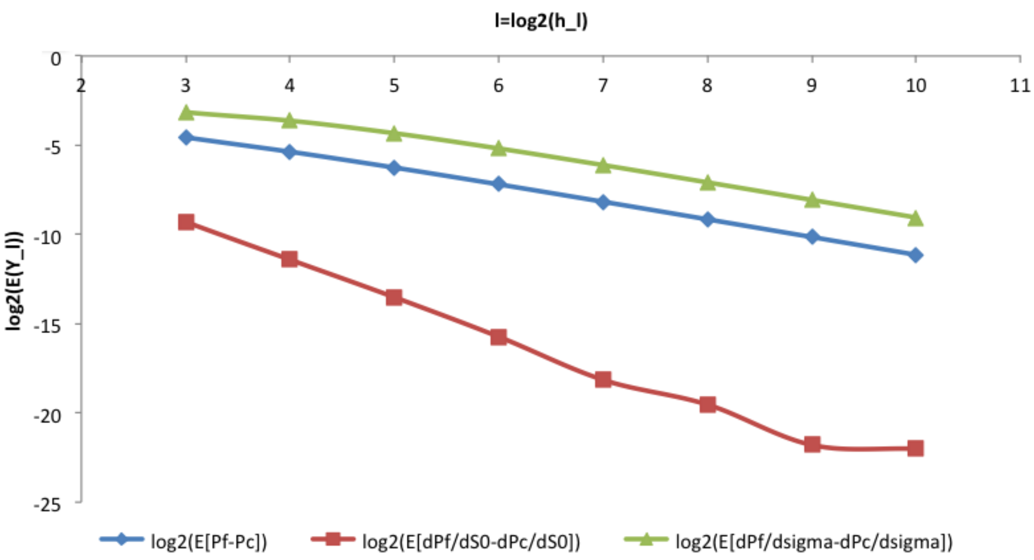

Figure 2.3: Pathwise sensitivities and conditional expectations, European call : $\mathbb{E}(\widehat{Y}_l)$

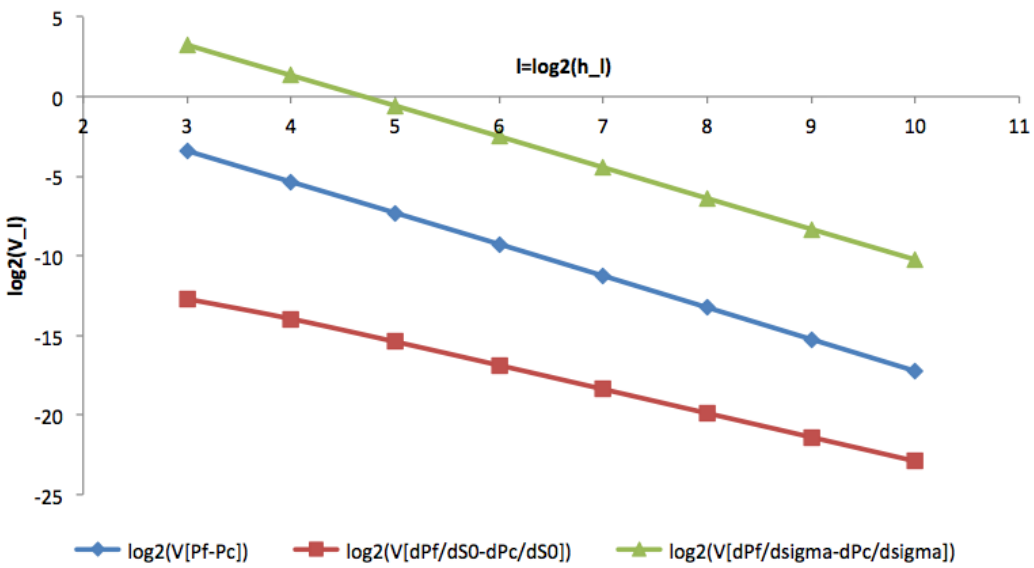

Figure 2.4: Pathwise sensitivities and conditional expectations, European call : V_l

Estimator	α	β	MLMC Complexity
Value	≈ 1.0	≈ 2.0	$O(\epsilon^{-2})$
Delta	≈ 1.6	≈ 1.5	$O(\epsilon^{-2})$
Vega	≈ 1.0	≈ 2.0	$O(\epsilon^{-2})$

Table 2.2: Pathwise sensitivities and conditional expectations, European call : estimated complexity

Reasoning as before, we get

$$\mathbb{E}\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)=O\left(1\right)O\left(h\right)+O\left(\sqrt{h}\right)O\left(\sqrt{h}\right) \\ =O\left(h\right) \tag{2.66}$$

and

$$\mathbb{V}\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)=O\left(1\right)O\left(h\right)^2+O\left(\sqrt{h}\right)O\left(\sqrt{h}\right)^2 \\ =O\left(h^{3/2}\right) \tag{2.67}$$

This means $\alpha = 1$ and $\beta = 3/2$ for the Greeks' estimators, which corresponds to what we observe.

We note that taking the conditional expectation of a non-smooth payoff results in a smooth function, which suggests that this approach is also applicable to discontinuous payoffs. We explore this idea in section 2.3.1.

The main limitation of this method is that in many situations (complicated payoff functions, multidimensional case) it leads to integral computations for which we don't necessarily have analytic solutions. Path splitting, to be discussed next, represents a useful numerical approximation to this technique.

2.2.3 Split pathwise sensitivities

This technique is based on the previous one. The idea is to avoid the potentially tricky computation of $\mathbb{E}\left[P\left(\widehat{S}^f_{N_f}\right)|\widehat{S}^f_{N_f-1}\right]$ and $\mathbb{E}\left[P\left(\widehat{S}^c_{N_c}\right)|\widehat{S}^c_{N_c-1},\Delta W^f_{N_f-2}\right]$. We instead get numerical estimates of these values by "splitting" every path simulation on the final time step.

At the fine level: we use a sequence of Brownian increments $\widehat{W}_f = \left(\Delta W^f_0,\ldots,\Delta W^f_{N_f-2}\right)$ to compute $\widehat{S}^f_{N_f-1}$. For every such path, we then simulate a set of d final increments based on d independent Brownian increments $(\Delta W^f_{N_f-1}{}^{(i)})_{i=1,\ldots,d}$, which we average to get

$$\mathbb{E}\left[P\left(\widehat{S}^f_{N_f}\right)|\widehat{S}^f_{N_f-1}\right]\approx\frac{1}{d}\sum_{i=1}^{d}P\left(\widehat{S}_{N_f}(\widehat{S}^f_{N_f-1},\Delta W^f_{N_f-1}{}^{(i)})\right) \tag{2.68}$$

At the coarse level we use $\widehat{W}_c = \left(\Delta W^f_0+\Delta W^f_1,\ldots,\Delta W^f_{N_f-4}+\Delta W^f_{N_f-3}\right)$ to simulate $\widehat{S}^c_{N_c-1}$. As in the previous section, we improve the convergence rate of $\mathbb{V}\left(\widehat{Y}_l\right)$ by considering estimators of $\mathbb{E}_{\Delta W^f_{N_f-2}}\left[\mathbb{E}\left(P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1},\Delta W^f_{N_f-2}\right)\right]$ instead of $\mathbb{E}\left[P\left(\widehat{S}^c_{N_c}\right)|\widehat{S}^c_{N_c-1}\right]$. We can do so by constructing the final coarse increments as $(\Delta W^c_{N_c-1}{}^{(i)})_{i=1,\ldots,d} = \left(\Delta W^f_{N_f-2}+\Delta W^f_{N_f-1}{}^{(i)}\right)_{i=1,\ldots,d}$ and using them to estimate

$$\mathbb{E}\left[P(\widehat{S}^c_{N_c})|\widehat{S}^c_{N_c-1}\right] \approx \frac{1}{d}\sum_{i=1}^{d} P\left(\widehat{S}^c_{N_c}(\widehat{S}^c_{N_c-1}, \Delta W^c_{N_c-1}{}^{(i)})\right) \tag{2.69}$$

To get the Greeks, we simply compute the corresponding pathwise sensitivities estimators when applicable.

2.2.3.1 Estimated complexity

In figures 2.5, 2.6 we plot the results of our simulations for $d = 400$ (we discuss below the choice of d).

Table 2.3 summarises the observed convergence rates and corresponding multi-level complexities for different values of d.

Note that if we take $d = 1$, then the simulation is identical to the one performed in section 2.2.1 except for the final time step which here uses the Euler scheme instead of the Milstein scheme.

As expected this method yields higher values of β than simple pathwise sensitivities: the convergence rates increase and tend to the rates offered by conditional expectations as d increases and the approximation of the conditional expectation gets more precise. We provide an interpretation for this in section 2.2.3.2.

2.2.3.2 Choice of the number of splittings

We here analyse how the variance of split pathwise sensitivies estimators can be decomposed and show the influence of the number of final samples d.

The following proposition is presented in [Gil09b] and [AG07].

Proposition 2.2.2. *If W and Z are independent random variables and the random variable $f(W,Z)$ is such that $\mathbb{E}_{W,Z}[| f(W,Z) |]$, $\mathbb{E}_W[\mathbb{E}_Z[| f(W,Z) |]]$ and $\mathbb{E}_{W,Z}[| f(W,Z) |]$ are finite, then*

$$\widehat{Y}_{M,d} = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{d}\sum_{n=1}^{d} f\left(W^{(m)}, Z^{(m,n)}\right)\right)$$

with independent samples $W^{(m)}$ and $Z^{(m,n)}$ is an unbiased estimator for

$$\mathbb{E}_{W,Z}[f(W,Z)] = \mathbb{E}_W[\mathbb{E}_Z[f(W,Z)]]$$

and its variance is

$$\mathbb{V}\left[\widehat{Y}_{M,d}\right] = \frac{1}{M}\mathbb{V}_W[\mathbb{E}_Z[f(W,Z)]] + \frac{1}{Md}\mathbb{E}_W[\mathbb{V}_Z[f(W,Z)]]$$

Proof. See [Gil09b]. □

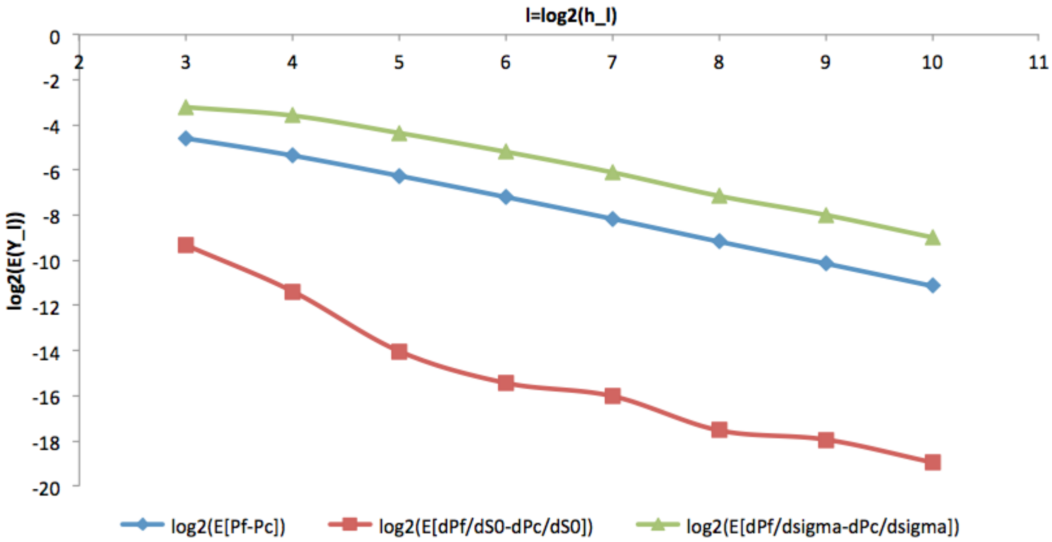

Figure 2.5: Pathwise sensitivities and path splitting, European call : $\mathbb{E}(\widehat{Y}_l)$

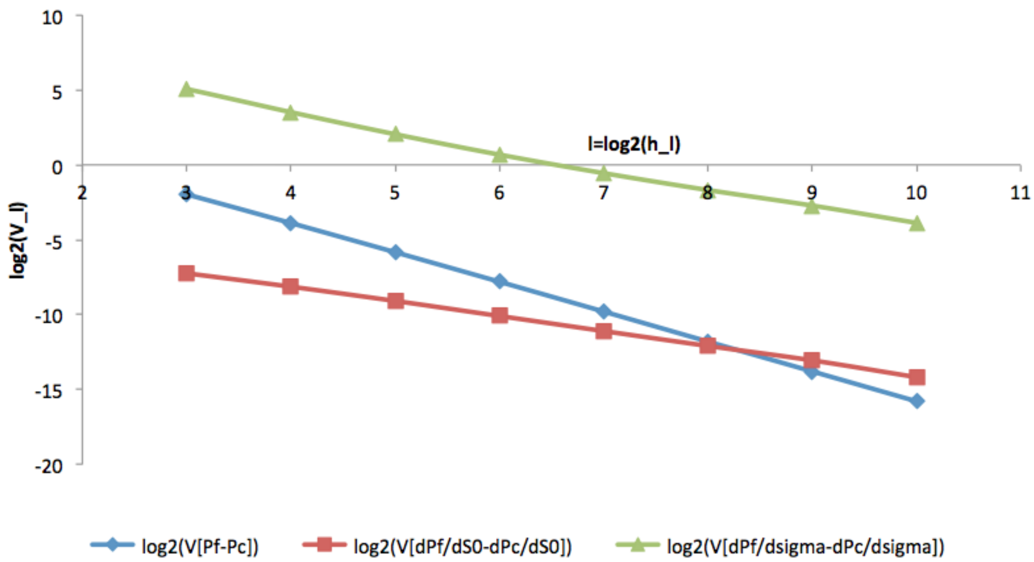

Figure 2.6: Pathwise sensitivities and path splitting, European call : V_l

Estimator	d	α	β	MLMC Complexity
Value	10	≈ 1.0	≈ 2.0	$O(\epsilon^{-2})$
	400	≈ 1.0	≈ 2.0	$O(\epsilon^{-2})$
Delta	10	≈ 1.0	≈ 1.0	$O(\epsilon^{-2}(\log \epsilon)^2)$
	400	≈ 1.3	≈ 1.4	$O(\epsilon^{-2})$
Vega	10	≈ 1.0	≈ 1.6	$O(\epsilon^{-2})$
	400	≈ 1.0	≈ 1.9	$O(\epsilon^{-2})$

Table 2.3: Pathwise sensitivities and path splitting, European call : estimated complexity

Using this decomposition, we can write the variance of the Greeks' estimator as

$$\begin{aligned}\mathbb{V}(\widehat{Y}_l) =& \frac{1}{M_l}\mathbb{V}_{\widehat{W}_f}\left[\mathbb{E}_{\Delta W^f_{N_f-1}}\left(\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\middle|\widehat{W}_f\right)\right] \\ &+ \frac{1}{dM_l}\mathbb{E}_{\widehat{W}_f}\left[\mathbb{V}_{\Delta W^f_{N_f-1}}\left(\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\middle|\widehat{W}_f\right)\right]\end{aligned} \tag{2.70}$$

The first term of this expression is the same as what we obtain with the payoff smoothing of section 2.2.2. As explained before, we can expect

$$\mathbb{V}_{\widehat{W}_f}\left[\mathbb{E}_{\Delta W^f_{N_f-1}}\left(\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\middle|\widehat{W}_f\right)\right] = O\left(h^{3/2}\right)$$

As before, the majority of paths arrive in a region where the payoff is Lipschitz and for which we therefore have

$$\mathbb{V}_{\Delta W^f_{N_f-1}}\left[\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\middle|\widehat{W}_f\right] = O\left(h^2\right)$$

but a proportion $O\left(\sqrt{h}\right)$ of all paths is such that S_T is within $O\left(\sqrt{h}\right)$ of the the payoff's "kink" at K. Then for those paths, there is a $O\left(h/\sqrt{h}\right)$ probability that the fine and coarse discretisations arrive on different sides of K and that therefore $\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta} = O\left(1\right)$. Else we have $\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta} = 0$. Therefore for those paths that are close to K, we have

$$\mathbb{V}_{\Delta W^f_{N_f-1}}\left[\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\middle|\widehat{W}_f\right] = O\left(\sqrt{h}\right)$$

Combining the two contributions via the law of total expectations, we get

$$\mathbb{E}_{\widehat{W}_f}\left[\mathbb{V}_{\Delta W^f_{N_f-1}}\left(\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\middle|\widehat{W}_f\right)\right] = O(h)$$

and finally

$$\mathbb{V}\left(\widehat{Y}_l\right) = \frac{1}{M_l}O\left(h^{3/2}\right) + \frac{1}{dM_l}O\left(h\right) \tag{2.71}$$

We note that taking $d = 1$ gives $\mathbb{V}\left(\widehat{Y}_l\right) = O\left(h\right)$, which matches the results of section 2.2.1.

Taking a constant number of splittings d for all levels is not optimal. We optimise the variance at a fixed computational cost. As seen in equation (2.71), the variance is of the form

$$\mathbb{V}\left(\widehat{Y}_l\right) = \frac{1}{M_l}\,v_1 + \frac{1}{M_l\,d_l}\,v_2$$

with $v_1 = O\left(h^{3/2}\right)$ and $v_2 = O(h)$ while the cost of the simulation is of the form

$$C\left(\widehat{Y}_l\right) = c_1\, M_l + c_2\, M_l\, d_l$$

where $c_1 = O\left(h^{-1}\right)$ is the cost of a path simulation and $c_2 = O(1)$ the cost of a payoff evaluation. Keeping the computational cost constant, the variance can be minimised by minimising the product

$$\left(\frac{1}{M_l}\, v_1 + \frac{1}{M_l\, d_l}\, v_2\right)(c_1\, M_l + c_2\, M_l\, d_l) = v_1 c_1 + v_1 c_2 d_l + \frac{v_2 c_1}{d_l} + c_2 v_2$$

which gives an optimal value of d as being

$$d_l^{opt} = \sqrt{\frac{v_2 c_1}{v_1 c_2}} = O\left(h^{-3/4}\right) \tag{2.72}$$

Note also that sampling $d_l = O\left(h^{-3/4}\right)$ final samples does not affect the asymptotic cost of the simulation path which is still $C\left(\widehat{Y}_l\right) = M_l O\left(h^{-1}\right)$ and the variance of the estimator is then $\mathbb{V}\left(\widehat{Y}_l\right) = \frac{1}{M_l} O\left(h^{3/2}\right)$. This means that the use of path splitting does not, to leading order, increase the variance or the computational cost compared to the use of exact conditional expectation in the cases where this can be evaluated by a closed form formula.

2.2.4 Vibrato Monte Carlo

Since the path splitting method is still based on the pathwise sensitivity analysis, it is not applicable when payoffs are discontinuous. We also saw in section 2.2.2 that in general, the conditional expectation technique wasn't easily used as it required the computation of analytic formulas. To address these limitations, we use the Vibrato Monte Carlo method introduced by Giles in [Gil09b]. This hybrid method combines pathwise sensitivities and the Likelihood Ratio Method.

We reuse the notations of section 2.2.2 and 2.2.3. Considering again equation (2.50) for a discretisation based on N time steps and noting that, as explained previously, $p\left(\widehat{S}_N \middle| \widehat{S}_{N-1}\right)$ and $\mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N) \middle| \widehat{S}_{N-1}\right]$ are functions of $\mu_{N-1}\left(\widehat{S}_{N-1}, t_{N-1}\right)$ and $\sigma_{N-1}\left(\widehat{S}_{N-1}, t_{N-1}\right)$, we now use the Likelihood Ratio Method on the last time step and we get

$$
\begin{aligned}
\frac{\partial \widehat{V}}{\partial \theta} &= \mathbb{E}_{\widehat{S}_{N-1}}\left[\frac{\partial \mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)\,\middle|\widehat{S}_{N-1}\right]}{\partial \theta}\right] \\
&= \mathbb{E}_{\widehat{S}_{N-1}}\left[\frac{\partial \mu_{N-1}}{\partial \theta}\frac{\partial \mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)\,\middle|\widehat{S}_{N-1}\right]}{\partial \mu_{N-1}}\right. \\
&\qquad\qquad\left.+\frac{\partial \sigma_{N-1}}{\partial \theta}\frac{\partial \mathbb{E}_{\Delta W_{N-1}}\left[P(\widehat{S}_N)\,\middle|\widehat{S}_{N-1}\right]}{\partial \sigma_{N-1}}\right] \\
&= \mathbb{E}_{\widehat{S}_{N-1}}\left[\frac{\partial \mu_{N-1}}{\partial \theta}\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial \log p\left(\widehat{S}_N\,\middle|\widehat{S}_{N-1}\right)}{\partial \mu_{N-1}}\middle|\widehat{S}_{N-1}\right]\right. \\
&\qquad\left.+\frac{\partial \sigma_{N-1}}{\partial \theta}\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial \log p\left(\widehat{S}_N\,\middle|\widehat{S}_{N-1}\right)}{\partial \sigma_{N-1}}\middle|\widehat{S}_{N-1}\right]\right]
\end{aligned}
\tag{2.73}
$$

This leads to the estimator

$$
\begin{aligned}
\frac{\partial \widehat{V}}{\partial \theta} \approx \frac{1}{M_l}\sum_{m=1}^{M_l}&\left(\frac{\partial \mu_{N-1}^{(m)}}{\partial \theta}\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial(\log p)}{\partial \mu_{N-1}}\,\middle|\widehat{S}_{N-1}^{(m)}\right]\right. \\
&\left.+\frac{\partial \sigma_{N-1}^{(m)}}{\partial \theta}\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial(\log p)}{\partial \sigma_{N-1}}\,\middle|\widehat{S}_{N-1}^{(m)}\right]\right)
\end{aligned}
\tag{2.74}
$$

where $\dfrac{\partial \mu_{N-1}^{(m)}}{\partial \theta}$ and $\dfrac{\partial \sigma_{N-1}^{(m)}}{\partial \theta}$ are computed via pathwise sensitivities.

With $\widehat{S}_N^{(m,i)} = \widehat{S}_N(\widehat{S}_{N-1}^{(m)}, \Delta W_{N-1}^{(i)})$ and noting that, as before with a final Euler step, $\widehat{S}_N$ is normally distributed conditionally on $\widehat{S}_{N-1}$, we use the following estimators to evaluate the conditional expectations of (2.74)

$$
\begin{aligned}
&\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial \log p}{\partial \mu_{N-1}}\middle|\widehat{S}_{N-1}^{(m)}\right] \\
&\qquad\approx \frac{1}{d}\sum_{i=1}^{d}\left[P\left(\mu_{N-1}^{(m)}+\sigma_{N-1}^{(m)}\frac{\Delta W_{N-1}^{(i)}}{\sqrt{h}}\right)\right. \\
&\qquad\qquad\left.-P\left(\mu_{N-1}^{(m)}-\sigma_{N-1}^{(m)}\frac{\Delta W_{N-1}^{(i)}}{\sqrt{h}}\right)\right]\frac{\Delta W_{N-1}^{(i)}}{2\sqrt{h}\sigma_{N-1}^{(m)}} \\
&\mathbb{E}_{\Delta W_{N-1}^{(i)}}\left[P\left(\widehat{S}_N\right)\frac{\partial \log p}{\partial \sigma_{N-1}}\middle|\widehat{S}_{N-1}^{(m)}\right] \\
&\qquad\approx \frac{1}{d}\sum_{i=1}^{d}\left[P\left(\mu_{N-1}^{(m)}+\sigma_{N-1}^{(m)}\frac{\Delta W_{N-1}^{(i)}}{\sqrt{h}}\right)-2P\left(\mu_{N-1}^{(m)}\right)\right. \\
&\qquad\qquad\left.+P\left(\mu_{N-1}^{(m)}-\sigma_{N-1}^{(m)}\frac{\Delta W_{N-1}^{(i)}}{\sqrt{h}}\right)\right]\frac{{\Delta W_{N-1}^{(i)}}^{2}-h}{2\sigma_{N-1}^{(m)}h}
\end{aligned}
\tag{2.75}
$$

Note that the estimators of (2.75) use antithetic variables for reducing the variance. See [Gil07] for more details.

At the fine level of a multilevel simulation, we base our simulations on $\widehat{P}_f := \mathbb{E}_{\Delta W^f_{N_f-1}}\left[P\left(\widehat{S}^f_{N_f}\right)\middle|\widehat{S}^f_{N_f-1}\right]$ and the estimation of

$$\mathbb{E}_{\widehat{S}^f_{N_f-1}}\left[\frac{\partial \widehat{P}_f}{\partial\theta}\right] = \mathbb{E}_{\widehat{S}^f_{N_f-1}}\left[\frac{\partial\mu^f_{N_f-1}}{\partial\theta}\frac{\partial\widehat{P}_f}{\partial\mu^f_{N_f-1}} + \frac{\partial\sigma^f_{N_f-1}}{\partial\theta}\frac{\partial\widehat{P}_f}{\partial\sigma^f_{N_f-1}}\right] \tag{2.76}$$

For this, we use a sequence of Brownian increments $\widehat{W}_f$ to simulate $\widehat{S}^f_{N_f-1}$. We then use the set $(\Delta W^f_{N_f-1}{}^{(i)})_{i=1,\ldots,d}$ of final "split" increments to evaluate the sensitivity (2.76) via the Vibrato Monte Carlo estimators (2.74) and (2.75).

At the coarse level, as before, we reuse the fine Brownian increments to construct the coarse ones,

$$\begin{aligned}\widehat{W}_c &= (\Delta W^f_1 + \Delta W^f_2, \ldots, \Delta W^f_{N_f-4} + \Delta W^f_{N_f-3})\\ (\Delta W^{(i)}_{N_c-1})_{i=1,\ldots,d} &= (\Delta W^f_{N_f-2} + \Delta W^f_{N_f-1}{}^{(i)})_{i=1,\ldots,d}\end{aligned} \tag{2.77}$$

Once again, we consider $\widehat{P}_c := \mathbb{E}\left[P\left(\hat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2}\right]$ and the tower property on the last coarse step written in equation (2.64) guarantees the telescoping sum condition (1.18) is verified.

The sensitivity to θ is still given by the estimation of the quantity

$$\mathbb{E}_{\widehat{S}^c_{N_c-1},\Delta W^f_{N_f-2}}\left[\frac{\partial\widehat{P}_c}{\partial\theta}\right] = \mathbb{E}_{\widehat{S}^c_{N_c-1},\Delta W^f_{N_f-2}}\left[\frac{\partial\mu^c_{N_c-1}}{\partial\theta}\frac{\partial\widehat{P}_c}{\partial\mu^c_{N_c-1}} + \frac{\partial\sigma^c_{N_c-1}}{\partial\theta}\frac{\partial\widehat{P}_c}{\partial\sigma^c_{N_c-1}}\right] \tag{2.78}$$

where

$$\begin{aligned}\frac{\partial\widehat{P}_c}{\partial\mu^c_{N_c-1}} &= \mathbb{E}_{\Delta W^f_{N_f-1}}\left[P\left(\widehat{S}^c_{N_c}\right)\frac{\partial\log p_c}{\partial\mu^c_{N_c-1}}\middle|\,\widehat{S}^c_{N_c-1}\,,\Delta W^f_{N_f-2}\right]\\ \frac{\partial\widehat{P}_c}{\partial\sigma^c_{N_c-1}} &= \mathbb{E}_{\Delta W^f_{N_f-1}}\left[P\left(\widehat{S}_{N_c}\right)\frac{\partial\log p_c}{\partial\sigma^c_{N_c-1}}\middle|\,\widehat{S}^c_{N_c-1}\,,\Delta W^f_{N_f-2}\right]\end{aligned} \tag{2.79}$$

which we compute with the following estimators

$$
\begin{aligned}
&\mathbb{E}_{\Delta W^f_{N_f-1}}\left[P\left(\widehat{S}^c_{N_c}\right)\frac{\partial \log p_c}{\partial \mu^c_{N_c-1}}\,\middle|\,\widehat{S}^c_{N_c-1}{}^{(m)},\Delta W^f_{N_f-2}\right]\\
&\qquad\approx \frac{1}{d}\sum_{i=1}^{d}\left(\left[P\left(\mu^c_{N_c-1}+\sigma^c_{N_c-1}\frac{\Delta W^f_{N_f-1}{}^{(i)}}{\sqrt{h_f}}\right)\right.\right.\\
&\qquad\qquad\left.\left.-P\left(\mu^c_{N_c-1}-\sigma^c_{N_c-1}\frac{\Delta W^f_{N_f-1}{}^{(i)}}{\sqrt{h_f}}\right)\right]\frac{\Delta W^f_{N_f-1}{}^{(i)}}{2\sqrt{h_f}\sigma^c_{N_c-1}}\right)^{(m)}\\
&\mathbb{E}_{\Delta W^f_{N_f-1}}\left[P\left(\widehat{S}_{N_c}\right)\frac{\partial \log p_c}{\partial \sigma^c_{N_c-1}}\,\middle|\,\widehat{S}^c_{N_c-1}{}^{(m)},\Delta W^f_{N_f-2}\right]\\
&\qquad\approx \frac{1}{d}\sum_{i=1}^{d}\left(\left[P\left(\mu^c_{N_c-1}+\sigma^c_{N_c-1}\frac{\Delta W^f_{N_f-1}{}^{(i)}}{\sqrt{h_f}}\right)-2P\left(\mu^c_{N_c-1}\right)\right.\right.\\
&\qquad\qquad\left.\left.+P\left(\mu^c_{N_c-1}-\sigma^c_{N_c-1}\frac{\Delta W^f_{N_f-1}{}^{(i)}}{\sqrt{h_f}}\right)\right]\frac{\Delta W^f_{N_f-1}{}^{(i)\,2}-h_f}{2\sigma^c_{N_c-1}h_f}\right)^{(m)}
\end{aligned}
\tag{2.80}
$$

2.2.4.1 Estimated complexity

The results of our numerical experiments are found in figures 2.7 and 2.8. We present the corresponding convergence rates and computational complexities for $d = 10$ in table 2.4.

2.2.4.2 "Intuitive" interpretation

As in section 2.2.3, this is an approximation of the conditional expectation technique, and so getting the same convergence rates as before was expected.

We can also note that increasing the number of samples d does not improve the convergence rate of the algorithm (as was the case for split pathwise sensitivities). The analysis of this fact is beyond the scope of this thesis and will be investigated in future research.

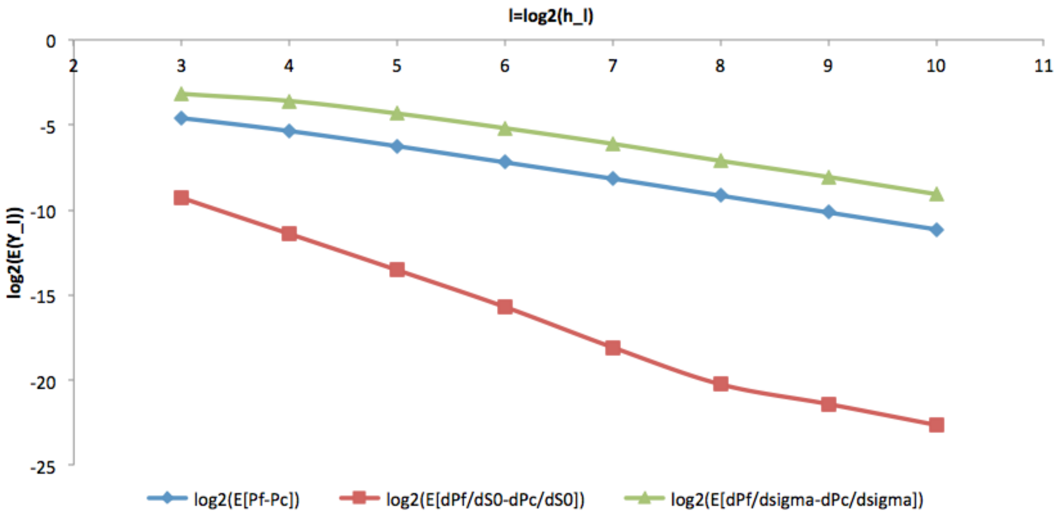

Figure 2.7: Vibrato Monte Carlo, European call : $\mathbb{E}(\widehat{Y}_l)$

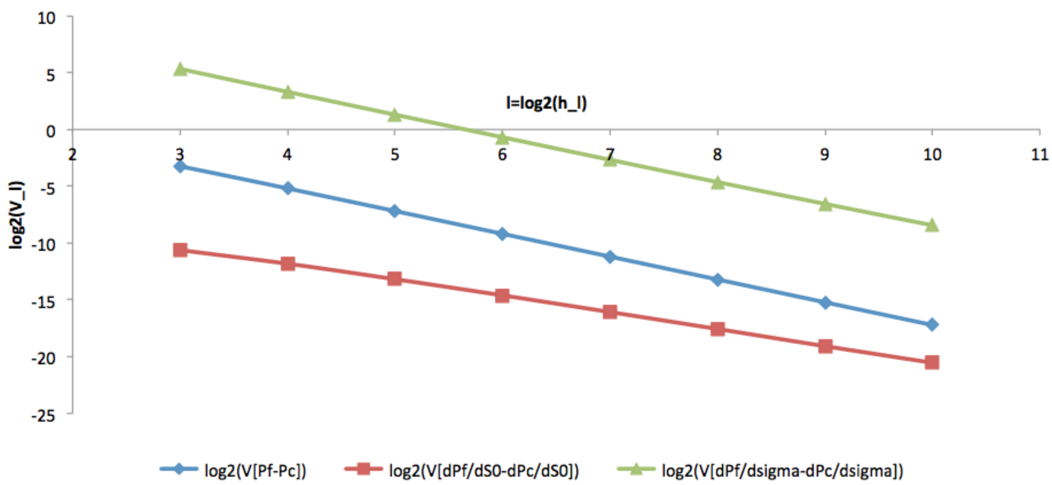

Figure 2.8: Vibrato Monte Carlo, European call : V_l

Estimator	α	β	MLMC Complexity
Value	≈ 1.0	≈ 2.0	$O(\epsilon^{-2})$
Delta	≈ 1.6	≈ 1.5	$O(\epsilon^{-2})$
Vega	≈ 1.0	≈ 2.0	$O(\epsilon^{-2})$

Table 2.4: Vibrato Monte Carlo, European call : estimated complexity

2.3 Discontinuous payoffs (European digital call)

The European digital call's payoff is $P = \mathbf{1}_{S_T > K}$. The discontinuity of the payoff makes the computation of Greeks more challenging. We cannot apply pathwise sensitivities, and so we have to use conditional expectations or Vibrato Monte Carlo.

2.3.1 Pathwise sensitivities and conditional expectations

With the same notation as in section 2.2.2 we compute the conditional expectations of the digital call's payoff at the fine and coarse levels,

$$\widehat{P}_f := \mathbb{E}\left(P\left(\widehat{S}^f_{N_f}\right)\middle|\widehat{S}^f_{N_f-1}\right) = \Phi\left(\frac{\mu^f_{N_f-1} - K}{\sigma^f_{N_f-1}}\right) \tag{2.81}$$

and

$$\widehat{P}_c := \mathbb{E}\left(P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2}\right) = \Phi\left(\frac{\mu^c_{N_c-1} - K}{\sigma^c_{N_c-1}}\right) \tag{2.82}$$

which we use to compute the sensitivities as in (2.59) and (2.65).

2.3.1.1 Estimated complexity

The simulations give figure 2.9, 2.10 and table 2.5. We then obtain the complexities listed in table 2.5.

2.3.1.2 "Intuitive" interpretation

Noting that the first order derivative of the European call's payoff corresponds to the payoff of the digital call, the analysis of the European Call in section 2.2.2 explains why we could expect $\beta = 3/2$ for the value's estimator. Giles has actually proved in [GDR13] that for the digital call we have $\beta = 3/2 - \delta$ for any $\delta > 0$.

S_T having a smooth distribution, a fraction $O\left(\sqrt{h}\right)$ of all paths arrive in the area of width $O\left(\sqrt{h}\right)$ around the strike where $\dfrac{\partial \mathbb{E}\left(P\left(\widehat{S}_N\right)\middle|\widehat{S}_{N-1}\right)}{\partial \widehat{S}_{N-1}}$ is not close to 0. In this area, the second derivative $\dfrac{\partial^2 \mathbb{E}\left(P\left(\widehat{S}_N\right)\middle|\widehat{S}_{N-1}\right)}{\partial \widehat{S}^2_{N-1}} = O\left(h^{-1}\right)$. Using the strong convergence properties of the discretisation scheme, we have $\left|\widehat{S}^f_{N_f} - \widehat{S}^c_{N_c}\right| = O(h)$ which results in a difference of order $O(h)$ between the fine and coarse values of μ_{N-1} and σ_{N-1}. For these paths, we therefore have $O(1)$ difference between the fine and coarse Greeks' estimates.

The majority of paths arrive further from the strike where the payoff is constant and $\dfrac{\partial \mathbb{E}\left(P\left(\widehat{S}_N\right)\middle|\widehat{S}_{N-1}\right)}{\partial \widehat{S}_{N-1}} \approx 0$ and their contribution to the global variance of the

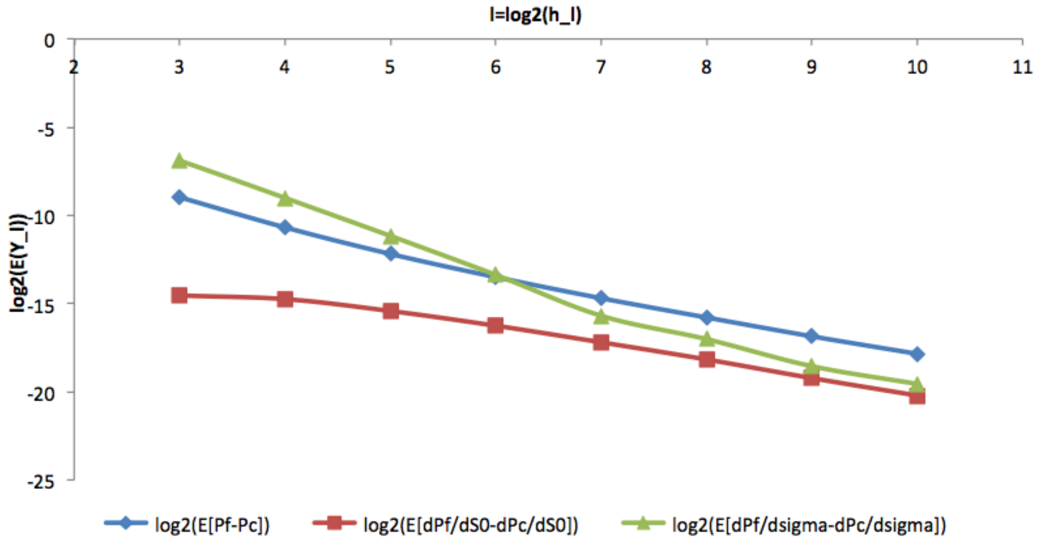

Figure 2.9: Pathwise sensitivities and conditional expectations, digital call: $\mathbb{E}(\widehat{Y}_l)$

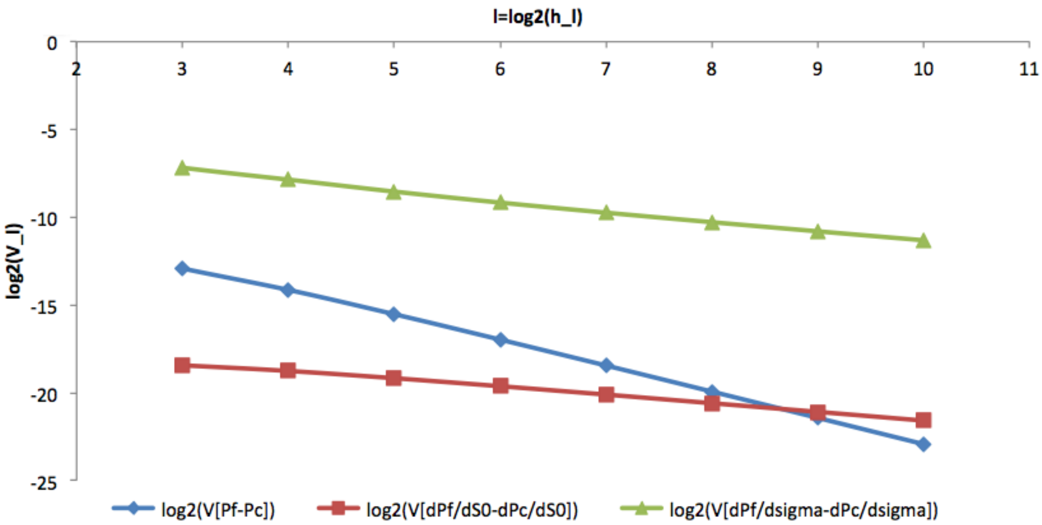

Figure 2.10: Pathwise sensitivities and conditional expectations, digital call: $\mathbb{V}(\widehat{Y}_l)$

Estimator	α	β	MLMC Complexity
Value	≈ 1.1	≈ 1.4	$O(\epsilon^{-2})$
Delta	≈ 0.9	≈ 0.5	$O(\epsilon^{-2.5})$
Vega	≈ 1.6	≈ 0.6	$O(\epsilon^{-2.2})$

Table 2.5: Pathwise sensitivities and conditional expectations, digital call : estimated complexity

multilevel estimator is negligible. Then, the law of total expectations gives

$$\begin{aligned}\mathbb{E}\left(\frac{\partial \widehat{P}_f}{\partial \theta}-\frac{\partial \widehat{P}_c}{\partial \theta}\right) &= O\left(1\right)0+O\left(\sqrt{h}\right)O\left(1\right)\\ &= O\left(\sqrt{h}\right)\end{aligned} \tag{2.83}$$

and

$$\begin{aligned}\mathbb{V}\left(\frac{\partial \widehat{P}_f}{\partial \theta}-\frac{\partial \widehat{P}_c}{\partial \theta}\right) &= O\left(1\right)0^2+O\left(\sqrt{h}\right)O\left(1\right)^2\\ &= O\left(\sqrt{h}\right)\end{aligned} \tag{2.84}$$

This explains the experimental $\alpha \approx 1/2$ and $\beta \approx 1/2$ for the Greeks.

2.3.2 Vibrato Monte Carlo

The Vibrato Monte Carlo technique can be applied to digital options in the same way as for European calls. Writing P the digital call's payoff, the formulas of section 2.2.4 still hold.

2.3.2.1 Estimated complexity

We take $d = 800$ and get figures 2.11, 2.12 and table 2.6.

We observe that unlike in the case of section 2.2.4 (European call), the observed convergence rate depends heavily on the number of samples d taken. A low number of final samples will result in unsatisfactory convergence rates. The study of this behaviour is beyond the scope of this thesis and will be performed in future research.

2.3.2.2 Intuitive interpretation and number of splittings

The reasoning is once again similar to the one presented in section 2.2.3.2: we write again

$$\begin{aligned}\mathbb{V}(\widehat{Y}_l) =& \frac{1}{M_l}\mathbb{V}_{\widehat{W}_f}\left[\mathbb{E}\left(\frac{\partial \widehat{P}_f}{\partial \theta}-\frac{\partial \widehat{P}_c}{\partial \theta}\middle|\widehat{W}_f\right)\right]\\ &+\frac{1}{dM_l}\mathbb{E}_{\widehat{W}_f}\left[\mathbb{V}\left(\frac{\partial \widehat{P}_f}{\partial \theta}-\frac{\partial \widehat{P}_c}{\partial \theta}\middle|\widehat{W}_f\right)\right]\end{aligned} \tag{2.85}$$

The first term of this expression is similar to what we obtain with the payoff smoothing of section 2.3.1. We can expect

$$\mathbb{V}_{\widehat{W}_f}\left[\mathbb{E}\left(\frac{\partial \widehat{P}_f}{\partial \theta}-\frac{\partial \widehat{P}_c}{\partial \theta}\middle|\widehat{W}_f\right)\right]=O\left(\sqrt{h}\right)$$

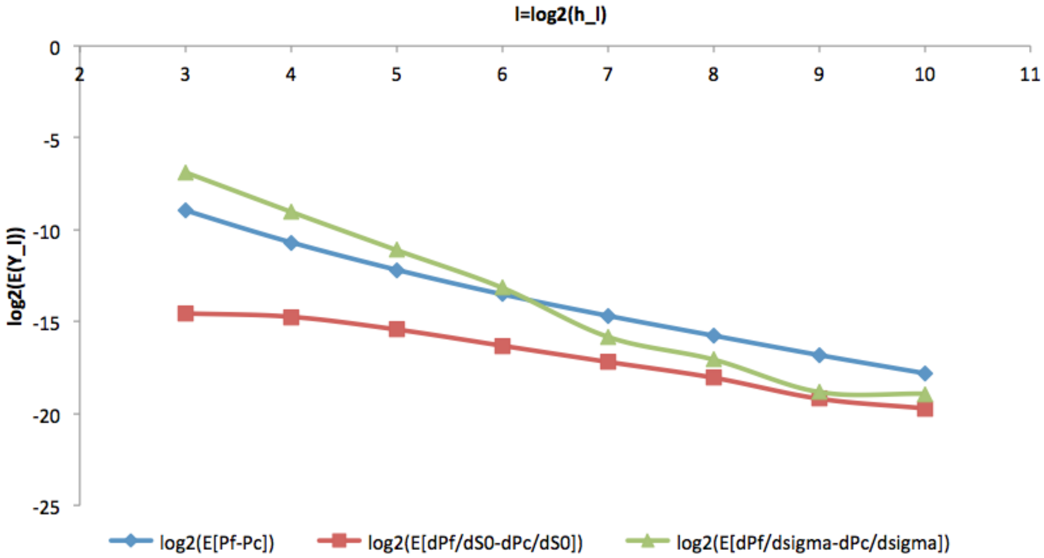

Figure 2.11: Vibrato Monte Carlo, digital call : $\mathbb{E}(\widehat{Y}_l)$

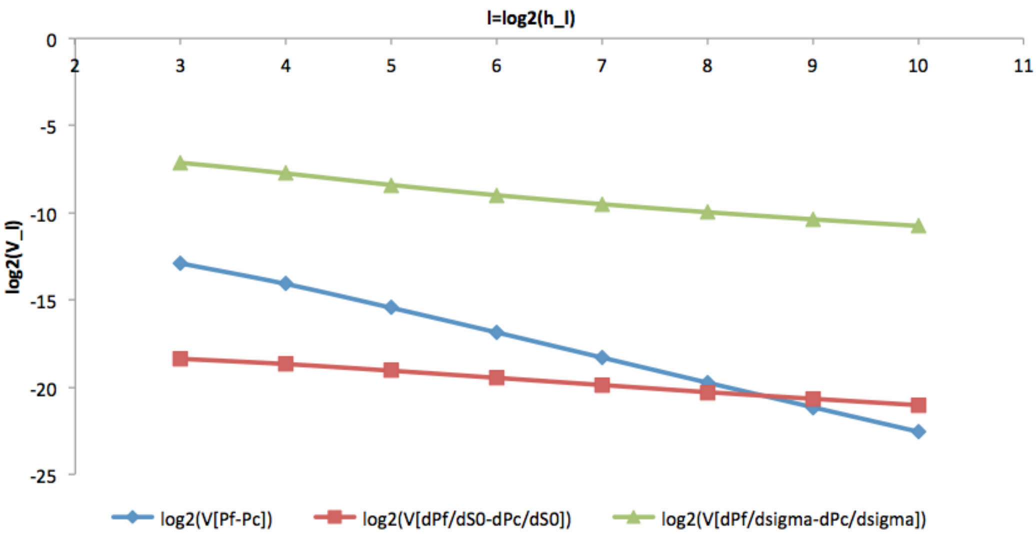

Figure 2.12: Vibrato Monte Carlo, digital call : $\mathbb{V}(\widehat{Y}_l)$

Estimator	α	β	MLMC Complexity
Value	≈ 1.0	≈ 1.4	$O(\epsilon^{-2})$
Delta	≈ 1.1	≈ 0.4	$O(\epsilon^{-2.5})$
Vega	≈ 1.7	≈ 0.5	$O(\epsilon^{-2.3})$

Table 2.6: Vibrato Monte Carlo, digital call : estimated complexity

As explained in equation (2.84), we can also expect

$$\mathbb{E}_{\widehat{W}_f}\left[\mathbb{V}\left(\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\middle| \widehat{W}_f\right)\right] = O\left(\sqrt{h}\right)$$

therefore

$$\mathbb{V}(\widehat{Y}_l) = \frac{1}{M_l}\left(O\left(\sqrt{h}\right) + \frac{1}{d}O\left(\sqrt{h}\right)\right) \tag{2.86}$$

and the two terms have the same order of convergence, it is therefore fine to take d constant.

2.4 Asian call option

The payoff of the Asian call option is of the form

$$P = \left(\overline{S} - K\right)^+ \tag{2.87}$$

where $\overline{S}$ is defined as the average value of the underlying over the considered time interval $[0, T]$.

$$\overline{S} = \frac{1}{T}\int_0^T S_t \mathrm{d}t \tag{2.88}$$

For an Asian call, pathwise sensitivities is the most appropriate technique. Indeed, it is a Lipschitz payoff of the underlying's average price on $[0, T]$ and its approximation based on the discretisation $\widehat{S}_0, \ldots, \widehat{S}_N$ is also Lipschitz in those discretised prices (see below). Also, the fact that this option is path dependent suggests the importance of a fine path discretisation, which, as seen in section 1.3, renders the Likelihood Ratio Method inappropriate.

2.4.1 Payoff estimator

As before, we simulate the underlying asset's price and its sensitivities on a discretisation $t_0, \ldots, t_N$ of the time interval $[0, T]$ using the Milstein scheme. As suggested in [Gil08a], we then use a Brownian bridge construction to define the following continuous extension on each interval $[t_n, t_{n+1}]$.

$$\begin{aligned}
\widehat{S}_{BB}(t) &:= \widehat{S}_n + \frac{t - t_n}{h}\left(\widehat{S}_{n+1} - \widehat{S}_n\right) + b_n\left(W(t) - W_n - \frac{t - t_n}{h}\Delta W_n\right) \\
\frac{\partial \widehat{S}_{BB}(t)}{\partial \theta} &= \frac{\partial \widehat{S}_n}{\partial \theta} + \frac{t - t_n}{h}\left(\frac{\partial \widehat{S}_{n+1}}{\partial \theta} - \frac{\partial \widehat{S}_n}{\partial \theta}\right) \\
&\quad + \left(\frac{\partial \widehat{S}_n}{\partial \theta}\frac{\partial b_n}{\partial S} + \frac{\partial b_n}{\partial \theta}\right)\left(W(t) - W_n - \frac{t - t_n}{h}\Delta W_n\right)
\end{aligned} \tag{2.89}$$

where $a_n = a\left(\widehat{S}_n, t_n\right)$, $b_n = b\left(\widehat{S}_n, t_n\right)$.

We also define a piecewise linear approximation

$$\begin{aligned}\widehat{S}(t) := \widehat{S}_{PL}(t) := \widehat{S}_n + \frac{t-t_n}{h}\left(\widehat{S}_{n+1} - \widehat{S}_n\right) \\ \frac{\partial \widehat{S}(t)}{\partial \theta} = \frac{\partial \widehat{S}_{PL}(t)}{\partial \theta} = \frac{\partial \widehat{S}_n}{\partial \theta} + \frac{t-t_n}{h}\left(\frac{\partial \widehat{S}_{n+1}}{\partial \theta} - \frac{\partial \widehat{S}_n}{\partial \theta}\right)\end{aligned} \tag{2.90}$$

The integral

$$\overline{S} = \frac{1}{T}\int_0^T S(t)\mathrm{d}t \tag{2.91}$$

can be approximated by

$$\overline{\widehat{S}}_{BB} = \frac{1}{T}\int_0^T \widehat{S}_{BB}(t)\mathrm{d}t \tag{2.92}$$

and more easily by the trapezoidal approximation

$$\overline{\widehat{S}} := \overline{\widehat{S}}_{PL} = \frac{1}{T}\int_0^T \widehat{S}_{PL}(t)\mathrm{d}t \tag{2.93}$$

which can be written

$$\begin{aligned}\overline{\widehat{S}} &= \sum_{n=0}^{N-1}\frac{1}{N}\left(\frac{\widehat{S}_n + \widehat{S}_{n+1}}{2}\right) \\ &= \frac{1}{2N}\left(\widehat{S}_0 + \widehat{S}_N\right) + \frac{1}{N}\sum_{n=1}^{N-1}\widehat{S}_n\end{aligned} \tag{2.94}$$

The payoff estimator is then $\widehat{P} := P\left(\overline{\widehat{S}}\right) = \left(\overline{\widehat{S}} - K\right)^+$

2.4.2 Pathwise sensitivities

We note that $\overline{\widehat{S}}$ is defined as a smooth Lipschitz function of all $\left(\widehat{S}_n\right)_{n=1..N}$ and a simple differentiation gives

$$\frac{\partial \overline{\widehat{S}}}{\partial \theta} = \frac{1}{2N}\left(\frac{\partial \widehat{S}_0}{\partial \theta} + \frac{\partial \widehat{S}_N}{\partial \theta}\right) + \frac{1}{N}\sum_{n=1}^{N-1}\frac{\partial \widehat{S}_n}{\partial \theta} \tag{2.95}$$

In the multilevel setting, we use these formulas at the fine level of discretisation by simulating $\widehat{S}_n^f$ and $\dfrac{\partial \widehat{S}_n^f}{\partial \theta}$ for $n = 0, \ldots, N_f$. Equation (2.95) then gives $\overline{\widehat{S}}^f$ and $\dfrac{\partial \overline{\widehat{S}}^f}{\partial \theta}$.

At the coarse level, we simulate $\widehat{S}_n^c$, $\dfrac{\partial \widehat{S}_n^c}{\partial \theta}$ for $n = 0, \ldots, N_c$. We then use these values to compute $\overline{\widehat{S}}^c$ and $\dfrac{\partial \overline{\widehat{S}}^c}{\partial \theta}$ with equation (2.95).

Note that the telescoping property (1.18) needed for the multilevel Monte Carlo to work is obviously respected when we define these estimators which are identical for a discretisation with a given number of time steps, no matter whether they are considered as coarse or fine level estimators.

The payoff $\widehat{P}$ is a Lipschitz function of $\overline{\widehat{S}}$ that is smooth almost everywhere. Therefore $\widehat{P}$ is also a Lipschitz function of all $\left(\widehat{S}_n\right)_{n=1..N}$, which is differentiable everywhere except on the hyperplane defined by $\frac{1}{N}\sum_{n=0}^{N}\widehat{S}_n = K$, i.e. almost everywhere. This confirms that pathwise sensitivities are applicable to this option and we can write

$$\begin{aligned}\frac{\partial \widehat{P}^f}{\partial \theta} &= \frac{\partial P}{\partial \overline{\widehat{S}}}\left(\overline{\widehat{S}}^f\right)\frac{\partial \overline{\widehat{S}}^f}{\partial \theta} \\ \frac{\partial \widehat{P}^c}{\partial \theta} &= \frac{\partial P}{\partial \overline{\widehat{S}}}\left(\overline{\widehat{S}}^c\right)\frac{\partial \overline{\widehat{S}}^c}{\partial \theta}\end{aligned} \tag{2.96}$$

2.4.2.1 Estimated Complexity

The simulations give the results presented in figures 2.13, 2.14 and in table 2.7

2.4.2.2 "Intuitive" interpretation

Under minimal assumptions, we can again assume that the values $\left(\widehat{S}_k\right)_{k=1,\ldots,N}$ have smooth probability density functions and that the average value $\overline{\widehat{S}}$ also has a smooth probability density function. The interpretation of the observed convergence rates is then fairly similar to the one for the European call found in section 2.2.1.

The Milstein scheme's $O(h)$ strong convergence implies that $\mathbb{E}\left(\left|\overline{\widehat{S}}^f - \overline{\widehat{S}}^c\right|\right) = O(h)$. The payoff being Lipschitz, this leads to $\mathbb{V}\left(P\left(\overline{\widehat{S}}^f\right) - P\left(\overline{\widehat{S}}^c\right)\right) = O\left(h^2\right)$, i.e. $\beta = 2$ for the value of the option.

For Greeks, the strong convergence of the Milstein scheme implies there is typically a $O(h)$ difference between $\frac{\partial \overline{\widehat{S}}^f}{\partial \theta}$ and $\frac{\partial \overline{\widehat{S}}^c}{\partial \theta}$. A fraction $O(h)$ of all paths is such that $\overline{\widehat{S}}$ is within $O(h)$ of the strike K. For those, there is a $O(1)$ likelihood that $\overline{\widehat{S}}^f$ and $\overline{\widehat{S}}^c$ are on different sides of the discontinuity in $\frac{\partial P}{\partial \overline{S}}$, implying

$$\left(\frac{\partial P}{\partial \overline{S}}\left(\overline{\widehat{S}}^f\right)\frac{\partial \overline{\widehat{S}}^f}{\partial \theta}\right) - \left(\frac{\partial P}{\partial \overline{S}}\left(\overline{\widehat{S}}^c\right)\frac{\partial \overline{\widehat{S}}^c}{\partial \theta}\right) = O(1) \tag{2.97}$$

For all other paths, $\overline{\widehat{S}}$ is far from K and intuitively, the two values $\overline{\widehat{S}}^f$ and $\overline{\widehat{S}}^c$ are located on the same side of the discontinuity. On each side of the discontinuity, $\frac{\partial P}{\partial \overline{S}}$ is locally Lipschitz (actually constant), therefore for these paths

$$\left(\frac{\partial P}{\partial \overline{S}}\left(\overline{\widehat{S}}^f\right)\frac{\partial \overline{\widehat{S}}^f}{\partial \theta}\right) - \left(\frac{\partial P}{\partial \overline{S}}\left(\overline{\widehat{S}}^c\right)\frac{\partial \overline{\widehat{S}}^c}{\partial \theta}\right) = O(h) \tag{2.98}$$

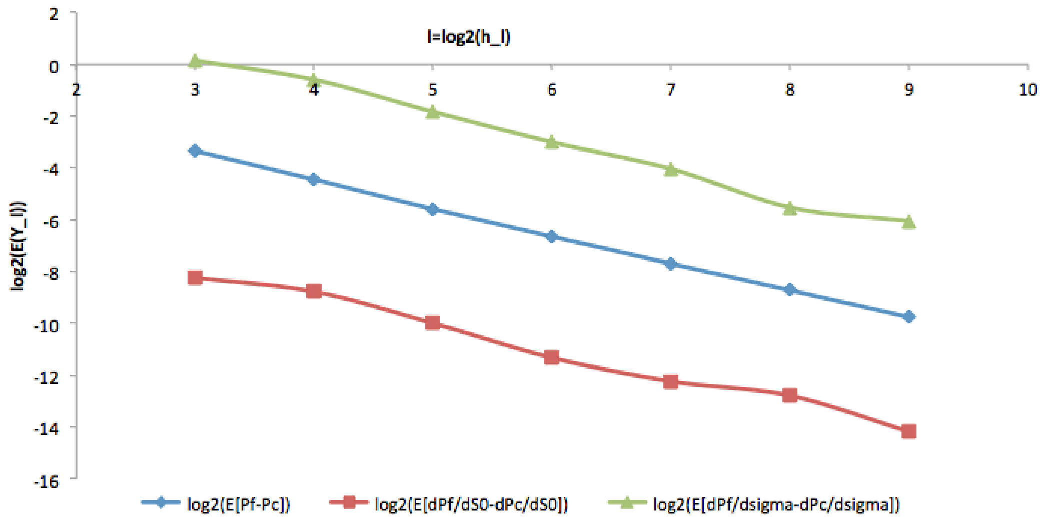

Figure 2.13: Pathwise sensitivities, Asian call: $\mathbb{E}(\widehat{Y}_l)$

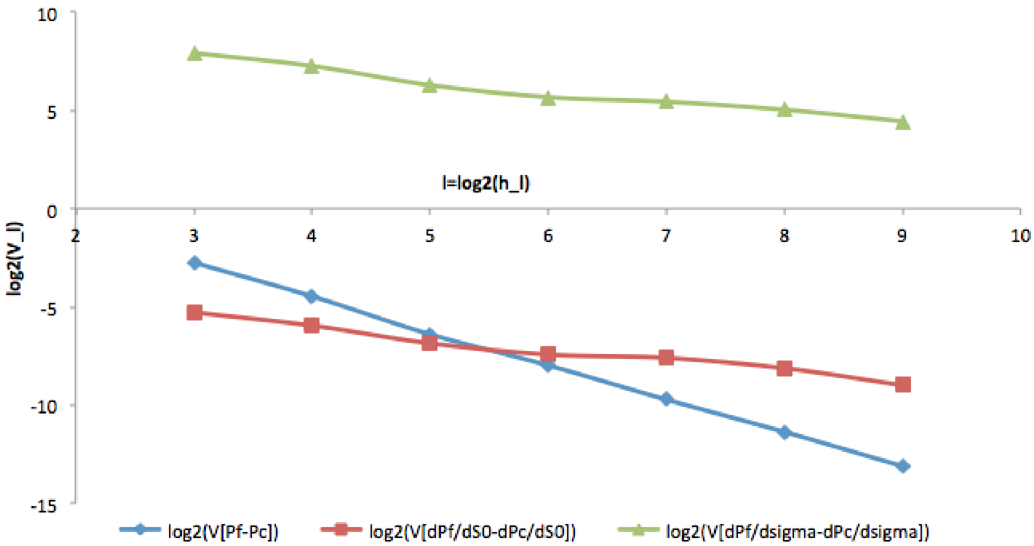

Figure 2.14: Pathwise sensitivities, Asian call: $\mathbb{V}_l$

Estimator	α	β	MLMC Complexity
Value	≈ 1.0	≈ 1.8	$O\left(\epsilon^{-2}\right)$
Delta	≈ 0.9	≈ 1.0	$O\left(\epsilon^{-2}\left(\log \epsilon\right)^{2}\right)$
Vega	≈ 1.0	≈ 1.2	$O\left(\epsilon^{-2}\right)$

Table 2.7: Pathwise sensitivities, Asian call: estimated complexity

and finally,

$$\mathbb{E}\left[\left(\frac{\partial P}{\partial \overline{S}}\left(\overline{\widehat{S}}^f\right)\frac{\partial \overline{\widehat{S}}^f}{\partial \theta}\right)-\left(\frac{\partial P}{\partial \overline{S}}\left(\overline{\widehat{S}}^c\right)\frac{\partial \overline{\widehat{S}}^c}{\partial \theta}\right)\right] = O\left(h\right)O\left(1\right)+O\left(1\right)O\left(h\right) \tag{2.99}$$
$$= O\left(h\right)$$

and

$$\mathbb{E}\left[\left(\left(\frac{\partial P}{\partial \overline{S}}\left(\overline{\widehat{S}}^f\right)\frac{\partial \overline{\widehat{S}}^f}{\partial \theta}\right)-\left(\frac{\partial P}{\partial \overline{S}}\left(\overline{\widehat{S}}^c\right)\frac{\partial \overline{\widehat{S}}^c}{\partial \theta}\right)\right)^2\right] = O\left(h\right)O\left(1\right)+O\left(1\right)O\left(h^2\right)$$
$$= O\left(h\right) \tag{2.100}$$

which suggests $\alpha = 1$ and $\beta = 1$ for Greeks.

2.5 European lookback call

The lookback call payoff is

$$P = S_T - \min_{t\in[0,T]} S_t \tag{2.101}$$

Pathwise sensitivities is once again the most appropriate technique. The payoff is a smooth function of the underlying's value on $[0, T]$. As detailed below, its approximation based on the discretisation $\widehat{S}_0, \ldots, \widehat{S}_N$ remains smooth.

The path dependency once again suggests the importance of a fine discretisation, which makes the Likelihood Ratio Method inappropriate.

2.5.1 Payoff estimator

The naive approximation of the payoff would be $\widehat{P} = (\widehat{S}_N - \min_n \widehat{S}_n)$. Nevertheless, as explained in [Gil08a], it does not result in satisfactory convergence rates with the Milstein scheme. The numerical results of Andersen and Brotherton-Radcliffe [ABR96] and Beaglehole, Dybvig and Zhou [BDZ97] indicate the following approach can be very effective at improving the convergence rates: after simulating $\left(\widehat{S}_n\right)_{n=0,\ldots,N}$, we approximate the behaviour of the process within each fine time step $[t_n, t_{n+1}]$ as a Brownian bridge with constant volatility $b_n = b\left(\widehat{S}_n, t_n\right)$. We write $\widehat{S}(t) := \widehat{S}_{BB}(t)$ the Brownian Bridge extension of the Milstein discretisation of S_t. As shown in section 6.4 of [Gla04] and detailed in section 3.4.2, we can then simulate its local minimum for each time step,

$$\widehat{S}_{n,min} := \frac{1}{2}\left(\widehat{S}_n + \widehat{S}_{n+1} - \sqrt{\left(\widehat{S}_{n+1}-\widehat{S}_n\right)^2 - 2\left(b_n\right)^2 h \log U_n}\right) \tag{2.102}$$

with U_n a uniform random variable on $[0,1]$.

In the multilevel setting, for the sake of clarity, we use indices based on the fine discretisation for both the fine and coarse discretisations, that is,

$$\begin{aligned} \widehat{S}^f_n &:= \widehat{S}^f\left(t^f_n\right) = \widehat{S}^f\left(nh_f\right) \\ \widehat{S}^c_n &:= \widehat{S}^c\left(t^f_n\right) = \widehat{S}^c\left(nh_f\right) \end{aligned} \tag{2.103}$$

We define the fine level's payoff estimator as follows: we simulate the path $\widehat{S}^f_0, \ldots, \widehat{S}^f_{N_f}$ and then with $b^f_n = b(\widehat{S}^f_n, t_n)$, we can simulate the local minimum of each step as

$$\widehat{S}^f_{n,min} := \frac{1}{2}\left(\widehat{S}^f_n + \widehat{S}^f_{n+1} - \sqrt{\left(\widehat{S}^f_{n+1} - \widehat{S}^f_n\right)^2 - 2\left(b^f_n\right)^2 h_f \log U_n}\right) \tag{2.104}$$

The simulated path's minimum is then defined as the minimum of the the local minima over all time steps , i.e.

$$\widehat{P}^f = \widehat{S}^f_{N_f} - \min_{n=0,\ldots,N_f-1} \widehat{S}^f_{n,\min} \tag{2.105}$$

At the coarse level we first simulate the "natural" points of the coarse discretisation (i.e. those with even indices $\widehat{S}^c_{2k}$). To get better convergence rates, the idea is again to consider that the process behaves like a simple Brownian motion on each time step $[t_{2k}, t_{2k+2}]$ with constant volatility b^c_{2k}.

$$\widehat{S}^c(t) = \widehat{S}^c_{2k} + \frac{t - t_{2k}}{h}\left(\widehat{S}^c_{2k+2} - \widehat{S}^c_{2k}\right) + b^c_{2k}\left(W_t - W_{2kh} - \frac{t - t_{2k}}{h}\left(W_{2kh+2} - W_{2k}\right)\right) \tag{2.106}$$

We use the Brownian increments of the fine level to define a midpoint value for each coarse step (thus constructing the values for odd indices $\widehat{S}^c_{2k+1}$).

$$\widehat{S}^c_{2k+1} := \frac{1}{2}\left(\widehat{S}^c_{2k} + \widehat{S}^c_{2k+2} - b^c_{2k}\left(\Delta W^f_{2k+1} - \Delta W^f_{2k}\right)\right) \tag{2.107}$$

We recall $\left(\Delta W^f_{2k+1} - \Delta W^f_{2k}\right)$ is the difference of the fine Brownian increments on the fine time steps $[t_{2k+1}, t_{2k+2}]$ and $[t_{2k}, t_{2k+1}]$. Conditional on this value, we then define the minimum over the whole step $[t_{2k}, t_{2k+2}]$ as the minimum of the minima over each half step, that is

$$\begin{aligned} \widehat{S}^c_{2k,min} := \min\Bigg[&\frac{1}{2}\left(\widehat{S}^c_{2k} + \widehat{S}^c_{2k+1} - \sqrt{\left(\widehat{S}^c_{2k+1} - \widehat{S}^c_{2k}\right)^2 - (b^c_{2k})^2 h_f \log U_{2k}}\right), \\ &\frac{1}{2}\left(\widehat{S}^c_{2k+1} + \widehat{S}^c_{2k+2} - \sqrt{\left(\widehat{S}^c_{2k+2} - \widehat{S}^c_{2k+1}\right)^2 - (b^c_{2k})^2 h_f \log U_{2k+1}}\right)\Bigg] \end{aligned} \tag{2.108}$$

where U_{2k} and U_{2k+1} are the values we sampled to compute the minima of the corresponding time steps $[t_{2k+1}, t_{2k+2}]$ and $[t_{2k}, t_{2k+1}]$ at the fine level. The payoff estimator is then

$$\widehat{P}^c = \widehat{S}^c_{N_f} - \min_{k=0,\ldots,N_f/2-1} \widehat{S}^c_{2k,\min} \tag{2.109}$$

Note that the fine and coarse estimators of two consecutive levels $(l-1), (l)$ are based on the same process $\widehat{S}_t$ defined as a Milstein discretisation of S_t with $N := N_c(l) = N_f(l-1)$ steps and extended as a Brownian bridge with constant volatility within each time step. The only difference in the estimation of the expectation of the minimum is that the coarse estimator of level $(l-1)$ is also based on the simulation of the mid-values on each time step. Using the indices of the fine discretisation at level l, for any function f, we can write the tower property

$$\begin{aligned}\mathbb{E}\left(f\left(\min \widehat{S}_{2k,min}\right)\middle|\widehat{S}_0, \widehat{S}_2, \ldots, \widehat{S}_{N-2}, \widehat{S}_N\right) = \\ \mathbb{E}\left(\mathbb{E}\left(f\left(\min_k \widehat{S}_{2k,min}\right)\middle|\widehat{S}_0, \widehat{S}_1, \widehat{S}_2, \ldots, \widehat{S}_N\right)\middle|\widehat{S}_0, \widehat{S}_2, \ldots, \widehat{S}_{N-2}, \widehat{S}_N\right)\end{aligned} \tag{2.110}$$

Taking $f\left(\min\limits_k \widehat{S}_{2k,min}\right) = \min \widehat{S}_{2k,min}$, the telescoping sum property (1.18) is still respected for the lookback call's payoff, which proves the estimators are valid.

2.5.2 Pathwise sensitivities

Using the treatment described above, we see all the quantities defined in the payoff estimators are almost surely Lipschitz functions of the simulated values $\widehat{S}^f_0, \ldots, \widehat{S}^f_{N_f}$ and $\widehat{S}^c_0, \ldots, \widehat{S}^c_{N_f}$. More precisely, it is locally Lipschitz except when the square root term appearing in equation (2.102) for the computation of the minimum is 0. This happens when for some index n, we have $\widehat{S}_n = \widehat{S}_{n+1} = \min\limits_{t\in[t_n,t_{n+1}]} \widehat{S}_t$, which is clearly a 0-probability event and will be analysed more precisely later.

We can then apply straighforward pathwise sensitivities to compute the multilevel estimator of the sensitivity. Differentiating (2.104) to get the sensitivity of a local minimum yields

$$\begin{aligned}\frac{\partial \widehat{S}^f_{n,min}}{\partial\theta} = \frac{1}{2}\Bigg[&\frac{\partial \widehat{S}^f_n}{\partial\theta} + \frac{\partial \widehat{S}^f_{n+1}}{\partial\theta} \\ &- \frac{\left(\widehat{S}^f_{n+1} - \widehat{S}^f_n\right)\left(\frac{\partial \widehat{S}^f_{n+1}}{\partial\theta} - \frac{\partial \widehat{S}^f_n}{\partial\theta}\right)}{\sqrt{\left(\widehat{S}^f_{n+1} - \widehat{S}^f_n\right)^2 - 2\left(b^f_n\right)^2 h_f \log U_n}} \\ &+ \frac{2b^f_n\left(\frac{\partial b^f_n}{\partial\theta} + \frac{\partial b^f_n}{\partial S_n}\frac{\partial \widehat{S}^f_n}{\partial\theta}\right) h_f \log U_n}{\sqrt{\left(\widehat{S}^f_{n+1} - \widehat{S}^f_n\right)^2 - 2\left(b^f_n\right)^2 h_f \log U_n}}\Bigg]\end{aligned} \tag{2.111}$$

where each term can be easily computed by Pathwise sensitivities.

Under minimal assumptions, for any $0 \leq n_1 < n_2 \leq N_f$, the density functions of the increments $\widehat{S}^f_{n_2,\min} - \widehat{S}^f_{n_1,\min}$ are clearly smooth (see section 3.3 for more details). Therefore, for each path, there is almost surely a unique index $n^f_{\min}$ such that $\widehat{S}^f_{n^f_{\min},\min} = \min_n \widehat{S}^f_{n,\min}$.

Finally we get

$$\frac{\partial \widehat{P}^f}{\partial \theta} = \frac{\partial \widehat{S}^f_{N_f}}{\partial \theta} - \frac{\partial \widehat{S}^f_{n^f_{\min},min}}{\partial \theta} \tag{2.112}$$

At the coarse level, there is also almost surely a unique index $n^c_{\min}$ such that $\widehat{S}^c_{n^c_{\min},\min} = \min_n \widehat{S}^c_{n,\min}$. Note that intuitively the indices $n^c_{\min}$ and $n^f_{\min}$ are likely to be the same but that it is not necessarily the case. The sensitivity of $\widehat{S}^c_{n^c_{\min},\min}$ is computed by differentiating equation (2.108), which results in a formula similar to (2.111) and we then have

$$\frac{\partial \widehat{P}^c}{\partial \theta} = \frac{\partial \widehat{S}^c_{N_f}}{\partial \theta} - \frac{\partial \widehat{S}^c_{n^c_{\min},min}}{\partial \theta} \tag{2.113}$$

2.5.2.1 Estimated complexity

The results of our simulations are presented in figures 2.15, 2.16 and table 2.8.

2.5.2.2 "Intuitive" interpretation

Giles et al have proved in [GDR13] that for the value's estimator, $\beta = 2$. For the lookback option with floating strike with the Black & Scholes model, we have the particular proportionality relationship between the option's value and its delta (Δ). Using equation (2.3), we have

$$\begin{aligned}
\Delta &= \frac{\partial P}{\partial S_0} \\
&= \frac{\partial \left(S_T - S_{\min}\right)}{\partial S_0} \\
&= \frac{1}{S_0}\left(S_T - S_{\min}\right) \\
&= \frac{1}{S_0} P
\end{aligned} \tag{2.114}$$

Therefore $\mathbb{V}\left(\Delta_l - \Delta_{l-1}\right)$ and $\mathbb{V}\left(P_l - P_{l-1}\right)$ are expected to converge at the same speed and $\beta = 2$ for Δ too. Therefore we focus on a more "typical" Greek like the vega (ν).

The convergence speed of vega's estimator cannot be derived as easily as that of the price in [GDR13]. Indeed, while we know from the convergence properties of the Milstein scheme that $\mathbb{E}\left(\frac{\partial \widehat{S}^f_{N_f}}{\partial \theta} - \frac{\partial \widehat{S}^c_{N_f}}{\partial \theta}\right) = O\left(h\right)$ and that for any fixed index n_0, $\mathbb{E}\left(\frac{\partial \widehat{S}^f_{n_0}}{\partial \theta} - \frac{\partial \widehat{S}^c_{n_0}}{\partial \theta}\right) = O\left(h\right)$, it is difficult to predict the behaviour of

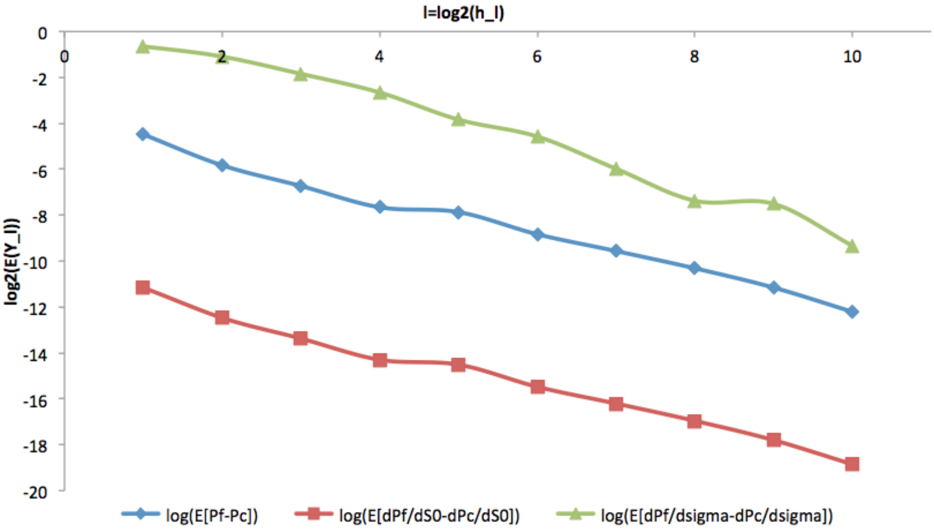

Figure 2.15: Pathwise sensitivities, lookback call : $\mathbb{E}(\widehat{Y}_l)$

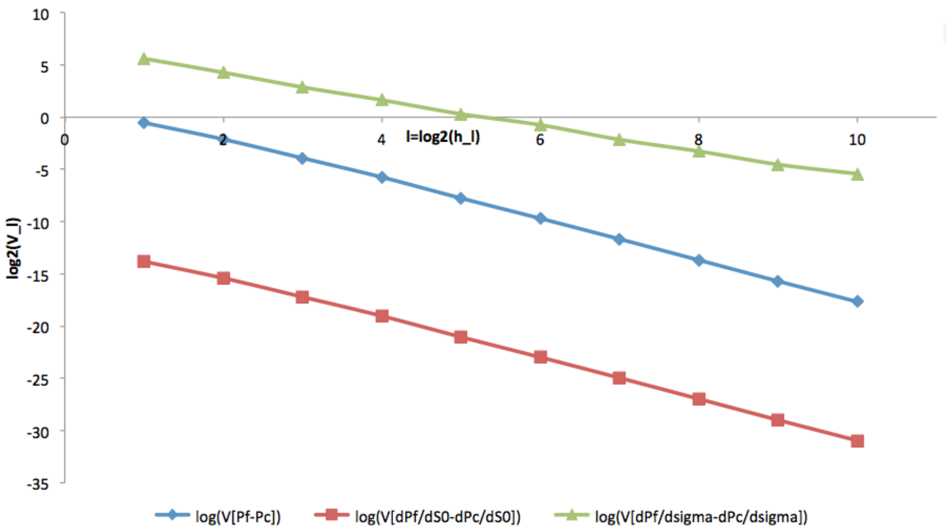

Figure 2.16: Pathwise sensitivities, lookback call : V_l

Estimator	α	β	MLMC Complexity
Value	≈ 0.9	≈ 1.9	$O(\epsilon^{-2})$
Delta	≈ 0.9	≈ 1.9	$O(\epsilon^{-2})$
Vega	≈ 1.0	≈ 1.1	$O(\epsilon^{-2})$

Table 2.8: Pathwise sensitivities, lookback call : estimated complexity

$\mathbb{E}\left(\dfrac{\partial \widehat{S}^f_{n^f_{\min},\min}}{\partial\theta} - \dfrac{\partial \widehat{S}^c_{n^c_{\min},\min}}{\partial\theta}\right)$, which clearly depends on the quantity $t_{n^f_{\min}} - t_{n^c_{\min}} = \left(n^f_{\min} - n^c_{\min}\right) h_f$, the difference between the times at which the fine and coarse paths reach their respective minima. This is discussed in detail in chapter 8.

2.5.3 Conditional Expectations, path splitting or Vibrato Monte Carlo

Unlike the regular call option, the payoff of the lookback call is perfectly smooth and so therefore there is no benefit from using conditional expectations and associated methods.

2.6 European barrier call

Barrier options are contracts which are activated or deactivated when the underlying asset's price S_t reaches a certain barrier value B. We consider here the down-and-out call for which the payoff can be written as

$$P = (S_T - K)^+ \, \mathbf{1}_{\min\limits_{t\in[0,T]}(S_t) > B} \tag{2.115}$$

Once again the path dependency highlights the importance of a fine discretisation and we use pathwise sensitivities based estimators.

2.6.1 Payoff estimator

Both the naive estimators and the approach used with the lookback call (simulating minima on each time step using a Brownian Bridge interpolation) are unsatisfactory: the discontinuity induced by the barrier results in a high variance when pricing and makes pathwise sensitivities inapplicable. Therefore, we use the approach presented in [BCI99] or [GS01] and used in [GDR13]: after simulating the path at times $t_0, \dots, t_N$, we compute for each time step $[t_n, t_{n+1}]$ the probability p_n that the Brownian interpolant $\widehat{S}(t)$ crosses the barrier, i.e. that the local minimum is below B. It can be proved (see [Gla04] for example) that

$$p_n = \exp\left(\frac{-2(\widehat{S}_n - B)^+(\widehat{S}_{n+1} - B)^+}{b_n^2\, h}\right) \tag{2.116}$$

We then define the price estimator

$$\widehat{P} = (\widehat{S}_N - K)^+ \prod_{n=0}^{N-1} (1 - p_n) \tag{2.117}$$

and note that

$$\begin{aligned}
&\mathbb{E}\left(\left(\widehat{S}_N - K\right)^+ \mathbf{1}_{\min \widehat{S}_t > B} \middle| \widehat{S}_0, \ldots, \widehat{S}_N\right) \\
&\qquad = \left(\widehat{S}_N - K\right)^+ \mathbb{E}\left(\mathbf{1}_{\min \widehat{S}_t > B} \middle| \widehat{S}_0, \ldots, \widehat{S}_N\right) \\
&\qquad = \left(\widehat{S}_N - K\right)^+ \mathbb{E}\left(\prod_{n=0}^{N} \mathbf{1}_{\min\limits_{[t_n, t_{n+1}]} \widehat{S}_t > B} \middle| \widehat{S}_0, \ldots, \widehat{S}_N\right) \\
&\qquad = \left(\widehat{S}_N - K\right)^+ \prod_{n=0}^{N} \mathbb{E}\left(\mathbf{1}_{\min\limits_{[t_n, t_{n+1}]} \widehat{S}_t > B} \middle| \widehat{S}_n, \widehat{S}_{n+1}\right) \\
&\qquad = \widehat{P}
\end{aligned} \tag{2.118}$$

As $\widehat{P}$ is the conditional expectation of $\left(\widehat{S}_N - K\right)^+ \mathbf{1}_{\min \widehat{S}_t > B}$, the two have the same expectation (tower property). Just like the Pathwise sensitivities with Conditional Expectations in the case of Vanilla options, this is an instance of what Boyle et al. call the *conditional Monte Carlo method* in [BBG97] and using $\widehat{P}$ instead of sampling the local minimums to estimate the second expression contributes to reducing the variance of the price estimator.

For multilevel simulations we again index both the fine and coarse levels with respect to the fine discretisation grid, as described in equation (2.103).

At the fine level, we simulate $\widehat{S}^f_0, \widehat{S}^f_1, \ldots, \widehat{S}^f_{N_f}$ and using the Brownian Bridge interpolant $\widehat{S}^f(t)$ on each step, we simply apply the previous formula to define

$$\widehat{P}^f = (\widehat{S}^f_{N_f} - K)^+ \prod_{n=0}^{N_f - 1} \left(1 - p^f_n\right) \tag{2.119}$$

with

$$p^f_n = \exp\left(\frac{-2(\widehat{S}^f_n - B)^+ (\widehat{S}^f_{n+1} - B)^+}{(b^f_n)^2\, h_f}\right) \tag{2.120}$$

At the coarse level we proceed as for the lookback option: we first simulate the discretised values on the coarse grid: $\widehat{S}_0, \widehat{S}_2, \ldots, \widehat{S}_{N_f}$. Then,on each coarse time step $[t_{2k}, t_{2k+2}]$, we simulate a midpoint value $\widehat{S}^c_{2k+1}$ corresponding to time $t_{2k+1} = (2k+1)\, h_f$ using the Brownian interpolation formula (2.89). Then for each coarse time step, we consider $p^c_{[2k,2k+2]}$ the probability of not hitting B in $[t_{2k}, t_{2k+2}]$, that is, the conditional probability that the Brownian Bridge interpolant $\widehat{S}^c(t)$ does not hit B in the fine time steps $[t_{2k}, t_{2k+1}]$ and $[t_{2k+1}, t_{2k+2}]$, conditional on its value at t_{2k+1}. Thus we write

$$
\begin{aligned}
\widehat{P}^c &= (\widehat{S}^c_{N_f} - K)^+ \prod_{k=0}^{N_f/2-1} \left(1 - p^c_{[2k,2k+2]}\right) \\
&= (\widehat{S}^c_{N_f} - K)^+ \prod_{k=0}^{N_f/2-1} \left((1 - p^c_{2k})(1 - p^c_{2k+1})\right)
\end{aligned}
\tag{2.121}
$$

where p^c_{2k} and p^c_{2k+1} are the probabilities of the coarse interpolant not hitting B on $[t_{2k}, t_{2k+1}]$ and $[t_{2k+1}, t_{2k+2}]$, conditional on its values $\widehat{S}^c_{2k}, \widehat{S}^c_{2k+1}, \widehat{S}^c_{2k+2}$, i.e.

$$
\begin{aligned}
p^c_{2k} &= \exp\left(\frac{-2(\widehat{S}^c_{2k} - B)^+(\widehat{S}^c_{2k+1} - B)^+}{(b^c_{2k})^2\, h_f}\right) \\
p^c_{2k+1} &= \exp\left(\frac{-2(\widehat{S}^c_{2k+1} - B)^+(\widehat{S}^c_{2k+2} - B)^+}{(b^c_{2k})^2\, h_f}\right)
\end{aligned}
\tag{2.122}
$$

Note that we keep the same volatility on the whole coarse interval, i.e. for $k = 0, \ldots, N_f/2 - 1$, we use $b^c_{2k} = b^c_{2k+1} := b^c_{[2k,2k+2]}$. As in the case of lookback options (see 2.110), the tower property guarantees the estimator satisfies the telescoping sum property (1.18).

2.6.2 Pathwise sensitivities

Note that in the Black & Scholes model we can assume the existence of a strictly positive lower bound of the volatility: if $\widehat{S}_n < B$, we know that $\widehat{P} = 0$ and there is no need for interpolation. Therefore we can always assume $b\,(S,t) = \sigma S_t > \sigma B > 0$.

Assuming the volatility terms b^f_n and b^c_n are non-zero, we see clearly from equations (2.119) and (2.121) that the multilevel estimator $\widehat{Y}_l = \widehat{P}^f - \widehat{P}^c$ is Lipschitz with respect to all $\left(\widehat{S}^f_n\right)_{n=1,\ldots,N_f}$ and $\left(\widehat{S}^c_n\right)_{n=1,\ldots,N_f}$ and we can use pathwise sensitivities.

At the fine level, we have

$$
\begin{aligned}
\frac{\partial \widehat{P}^f}{\partial \theta} &= \mathbf{1}_{\widehat{S}^f_{N_f} > K} \frac{\partial \widehat{S}^f_{N_f}}{\partial \theta} \prod_{n=0}^{N_f-1} \left(1 - p^f_n\right) \\
&\quad - \left(\widehat{S}^f_{N_f} - K\right)^+ \sum_{n=0}^{N_f-1} \left[\prod_{k=0, k\neq n}^{N_f-1} \left(1 - p^f_k\right) \frac{\partial p^f_n}{\partial \theta}\right]
\end{aligned}
\tag{2.123}
$$

with

$$
\begin{aligned}
\frac{\partial p^f_n}{\partial \theta} = \mathbf{1}_{\left(\widehat{S}^f_n, \widehat{S}^f_{n+1} > B\right)} p^f_n &\left[\frac{-2\widehat{\delta}^f_n \left(\widehat{S}^f_{n+1} - B\right)}{{b^f_n}^2\, h_f} + \frac{-2\widehat{\delta}^f_{n+1} \left(\widehat{S}^f_n - B\right)}{{b^f_n}^2\, h_f}\right. \\
&\left. + \frac{4\left(\widehat{S}^f_n - B\right)\left(\widehat{S}^f_{n+1} - B\right) \widetilde{b_n}^f}{{b^f_n}^3\, h_f}\right]
\end{aligned}
\tag{2.124}
$$

where $\widehat{\delta}_n^f := \dfrac{\partial \widehat{S}_n^f}{\partial \theta}$ and $\widetilde{b_n^f} := \left(\dfrac{\partial b_n^f}{\partial \theta} + \dfrac{\partial b_n^f}{\partial S_n} \dfrac{\partial \widehat{S}_n^f}{\partial \theta} \right)$ are computed easily via pathwise sensitivities.

At the coarse level, we obtain similar expressions for $\dfrac{\partial \widehat{P}^c}{\partial \theta}$ by differentiating (2.121), see chapter 7 for more details.

2.6.2.1 Estimated complexity

Our numerical simulations give the results presented in figures 2.17, 2.18 and table 2.9.

2.6.2.2 "Intuitive" interpretation

Giles et al prove in [GDR13] that for the value's estimator, $\beta = \dfrac{3}{2} - \delta$ for any $\delta > 0$. For the Greeks, the intuitive assumption is again that the joint distribution of the discretised values $\widehat{S}_1, \ldots, \widehat{S}_{N_f}$ is continuous. This suggests that a fraction $O\left(\sqrt{h}\right)$ of all paths have their fine discretisation's minimum in an interval of width $O\left(\sqrt{h}\right)$, $I_B = \left[B - O\left(\sqrt{h}\right), B + O\left(\sqrt{h}\right)\right]$.

Also within each time step $[t_n, t_{n+1}]$ of width h, the typical maximum deviation of the Brownian interpolant from the endpoints is $O\left(\sqrt{h}\right)$, this means that the value of the probability $p_n\left(\widehat{S}_n, \widehat{S}_{n+1}\right)$ goes from 1 (when $\widehat{S}_n < B$ or $\widehat{S}_{n+1} < B$) to 0 (when $\min\left(\widehat{S}_n, \widehat{S}_{n+1}\right) - B \gg \sqrt{h}$) in an area of width $O\left(\sqrt{h}\right)$ (when $\widehat{S}_n$ or $\widehat{S}_{n+1}$ move away from the barrier), therefore in this area we intuitively get $\dfrac{\partial p_n}{\partial \widehat{S}_n} = O\left(h^{-1/2}\right)$, $\dfrac{\partial p_n}{\partial \widehat{S}_{n+1}} = O\left(h^{-1/2}\right)$, $\dfrac{\partial^2 p_n}{\partial \widehat{S}_n^2} = O\left(h^{-1}\right)$ and $\dfrac{\partial^2 p_n}{\partial \widehat{S}_{n+1}^2} = O\left(h^{-1}\right)$.

The minima of the fine and coarse discretisations differ by $O(h)$. For paths whose fine discretisation's minimum is not within the interval I_B, they are therefore either both above $B + O\left(\sqrt{h}\right)$, or both below $B - O\left(\sqrt{h}\right)$. Therefore all the probabilities $\left(p_n^f, p_n^c\right)_{n=0,\ldots,N_f-1}$ are going to be almost constant, at either 0 or 1.

In the first case, we never approach the barrier and the payoff is similar to that of the standard European call: $\dfrac{\partial \widehat{P}^f}{\partial \theta} - \dfrac{\partial \widehat{P}^c}{\partial \theta} = O(h)$ for the $O(1)$ fraction of paths such that S_T is far from K and $\dfrac{\partial \widehat{P}^f}{\partial \theta} - \dfrac{\partial \widehat{P}^c}{\partial \theta} = O(1)$ for the fraction $O\left(\sqrt{h}\right)$ of paths that arrive close to K.

In the second case, the barrier is hit by both discretisations and $\dfrac{\partial \widehat{P}^f}{\partial \theta} - \dfrac{\partial \widehat{P}^c}{\partial \theta} = 0$, this happens for a fraction $O(1)$ of all paths.

For paths whose fine discretisation's minimum is in I_B: let us assume the path gets close to B on a given step $[t_n, t_{n+1}]$. Intuitively, the volatility will be similar at the coarse and fine levels and we can approximate $p_n^f(x, y) \approx p_n^c(x, y) := p_n(x, y)$,

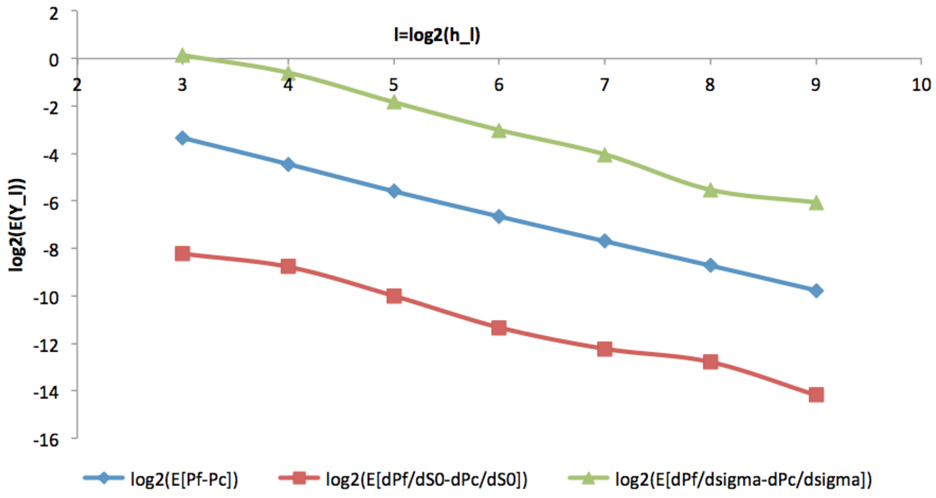

Figure 2.17: Pathwise sensitivities, barrier call : $\mathbb{E}(\widehat{Y}_l)$

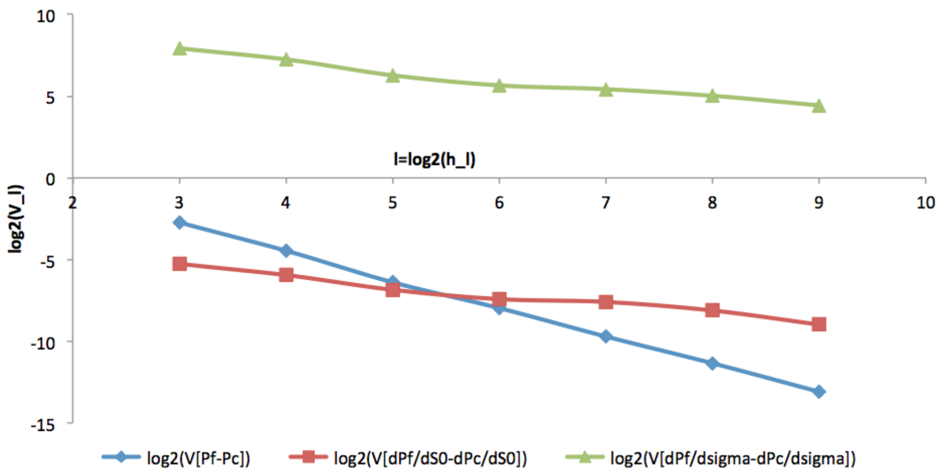

Figure 2.18: Pathwise sensitivities, barrier call : $\mathbb{V}(\widehat{Y}_l)$

Estimator	α	β	MLMC Complexity
Value	≈ 1.0	≈ 1.6	$O(\epsilon^{-2})$
Delta	≈ 1.0	≈ 0.6	$O(\epsilon^{-2.4})$
Vega	≈ 1.0	≈ 0.6	$O(\epsilon^{-2.4})$

Table 2.9: Pathwise sensitivities, barrier call : estimated complexity

we can write the difference

$$
\begin{aligned}
p_n^f - p_n^c &\approx \frac{\partial p_n}{\partial \widehat{S}_n}\left(\widehat{S}_n^f, \widehat{S}_{n+1}^f\right)\left(\widehat{S}_n^f - \widehat{S}_n^c\right) + \frac{\partial p_n}{\partial \widehat{S}_{n+1}}\left(\widehat{S}_n^f, \widehat{S}_{n+1}^f\right)\left(\widehat{S}_n^f - \widehat{S}_n^c\right) \\
&= O\left(h^{-1/2}\right) O\left(h\right) + O\left(h^{-1/2}\right) O\left(h\right) \\
&= O\left(\sqrt{h}\right)
\end{aligned} \tag{2.125}
$$

Now, assuming there is no direct dependency of p_n on θ (which is the case for Δ),

$$
\begin{aligned}
\frac{\partial p_n^f}{\partial \theta} - \frac{\partial p_n^c}{\partial \theta} &= \frac{\partial p_n^f}{\partial \widehat{S}_n}\frac{\partial \widehat{S}_n^f}{\partial \theta} + \frac{\partial p_n^f}{\partial \widehat{S}_{n+1}}\frac{\partial \widehat{S}_{n+1}^f}{\partial \theta} - \frac{\partial p_n^c}{\partial \widehat{S}_n}\frac{\partial \widehat{S}_n^c}{\partial \theta} + \frac{\partial p_n^c}{\partial \widehat{S}_{n+1}}\frac{\partial \widehat{S}_{n+1}^c}{\partial \theta} \\
&= \frac{\partial p_n^f}{\partial \widehat{S}_n}\left(\frac{\partial \widehat{S}_n^f}{\partial \theta} - \frac{\partial \widehat{S}_n^c}{\partial \theta}\right) + \frac{\partial p_n^f}{\partial \widehat{S}_{n+1}}\left(\frac{\partial \widehat{S}_{n+1}^f}{\partial \theta} - \frac{\partial \widehat{S}_{n+1}^c}{\partial \theta}\right) \\
&\quad + \frac{\partial \widehat{S}_n^c}{\partial \theta}\left(\frac{\partial p_n^f}{\partial \widehat{S}_n} - \frac{\partial p_n^c}{\partial \widehat{S}_n}\right) + \frac{\partial \widehat{S}_{n+1}^c}{\partial \theta}\left(\frac{\partial p_{n+1}^f}{\partial \widehat{S}_{n+1}} - \frac{\partial p_{n+1}^c}{\partial \widehat{S}_{n+1}}\right) \\
&\approx O\left(h^{-1/2}\right) O\left(h\right) + O\left(h^{-1/2}\right) O\left(h\right) \\
&\quad + O\left(1\right)\frac{\partial^2 p_n}{\partial \widehat{S}_n^2}\left(\widehat{S}_n^f - \widehat{S}_n^c\right) + O\left(1\right)\frac{\partial^2 p_{n+1}}{\partial \widehat{S}_{n+1}^2}\left(\widehat{S}_{n+1}^f - \widehat{S}_{n+1}^c\right) \\
&= O\left(h^{1/2}\right) + O\left(h^{-1}\right) O\left(h\right) + O\left(h^{-1}\right) O\left(h\right) \\
&= O\left(1\right)
\end{aligned} \tag{2.126}
$$

therefore from formula (2.123) and its coarse equivalent, we expect a difference $\dfrac{\partial \widehat{P}^f}{\partial \theta} - \dfrac{\partial \widehat{P}^c}{\partial \theta} = O\left(1\right)$ for these paths.

Combining the previous results on the contributions of various types of paths using the law of total expectation, we get

$$
\begin{aligned}
\mathbb{E}\left(\left(\frac{\partial \widehat{P}^f}{\partial \theta} - \frac{\partial \widehat{P}^c}{\partial \theta}\right)^2\right) &= O\left(1\right) 0 + O\left(1\right) O\left(h^2\right) + O\left(\sqrt{h}\right) O\left(1\right) + O\left(\sqrt{h}\right) O\left(1\right) \\
&= O\left(\sqrt{h}\right)
\end{aligned} \tag{2.127}
$$

which suggests the rate $\beta \approx 1/2$, which is indeed what we observe experimentally.

Similarly, we can show we expect $\alpha \approx 1/2$.

2.6.3 Conditional Expectations

The low convergence rates observed in the previous section come from from both the discontinuity at the barrier and from the lack of smoothness of the call around K. To address the latter, we could use the techniques described in section 2.2: Conditional Expectations, split pathwise sensitivities and Vibrato Monte Carlo.

Nevertheless, the above analysis reveals that the contributions to the global variance of the two sources of discontinuity have the same amplitude: in both cases, a fraction $O\left(\sqrt{h}\right)$ of all paths (the ones not hitting the barrier and such that S_T is close to K or the ones not hitting the barrier and whose minimum is close to B) result in a difference $O(1)$ between the fine and coarse estimates of the Greek. The techniques mentioned earlier would not reduce the contribution of the barrier and therefore the asymptotic rate of convergence would remain the same. Asymptotically there is therefore no benefit from using them and they would not reduce the computational cost of the multilevel computations.

Chapter 3

Numerical analysis, preliminary notes

In this chapter, we deal with various questions that are essential for a rigorous mathematical analysis of the multilevel techniques presented in chapter 2.

We begin by proving that the method used in chapter 2, the differentiation of the simulated underlying asset's values $\widehat{S}_0, \ldots, \widehat{S}_N$ does result in satisfactory estimators of the underlying asset's sensitivities $\dfrac{\partial S_0}{\partial \theta}, \ldots, \dfrac{\partial S_T}{\partial \theta}$, i.e. the sensitivity analysis of the discrete path approximation is equivalent to a discrete approximation of the sensitivity's SDE. This equivalence seems intuitively natural but needs to be established to prove convergence as the timestep $h \to 0$.

We provide practical conditions ensuring that pathwise sensitivities are applicable. We verify the method is valid in the setting of the Black & Scholes model and explain under which conditions it still is in the more general setting of an Ito process as described by equation (1.2). We then check that the differentiation of the simulated option's value does result in satisfactory estimators of the Greeks.

Certain regularity properties of the underlying asset's density function are necessary or at least desirable to make the analysis valid or to simplify it; we provide practical conditions on the process's volatility ensuring those are satisfied.

We present several important results on the moments of the solution of the evolution equation (1.2), on the properties of its discretisation and its continuous extensions. We introduce the fundamental theorem on which the so-called "extreme paths analysis" used in chapters 4 to 8 is based. Finally we introduce a convenient abuse of notation.

3.1 Estimating the underlying asset's sensitivity

We recall that in chapter 2, we first discretised the asset's evolution equation (2.2): we split the time interval $[0, T]$ into N time steps of width $h = T/N$ and as in equation (2.19) we obtained a discretisation formula of the form

$$\widehat{S}_{n+1} = f\left(\theta, \widehat{S}_n, \Delta W_n\right) \tag{3.1}$$

This enabled us to simulate approximate solutions $\widehat{S}_0, \widehat{S}_1, \ldots, \widehat{S}_N$. Then, to obtain their sensitivities $\frac{\partial \widehat{S}_0}{\partial \theta}, \ldots, \frac{\partial \widehat{S}_N}{\partial \theta}$, we simply differentiated the discretisation scheme (3.1) as in equation (2.25) and obtained a discrete equation of the form

$$\frac{\partial \widehat{S}_{n+1}}{\partial \theta} = \frac{\partial f\left(\theta, \widehat{S}_n, \Delta W_n\right)}{\partial \theta} + \frac{\partial f\left(\theta, \widehat{S}_n, \Delta W_n\right)}{\partial \widehat{S}_n} \frac{\partial \widehat{S}_n}{\partial \theta} \tag{3.2}$$

3.1.1 Order of discretisation and differentiation

Classical convergence results on the Euler and Milstein schemes (see [Mil79], [Tal84], [KP92] or [GS72] and section 1.1.2) guarantee the weak and strong convergence of $\widehat{S}_0, \widehat{S}_1, \ldots, \widehat{S}_N$ towards $S_0, S_h, \ldots, S_{Nh} = S_T$ but they do not guarantee a priori that their sensitivities $\frac{\partial \widehat{S}_0}{\partial \theta}, \ldots, \frac{\partial \widehat{S}_N}{\partial \theta}$ can be used as "naive" estimators that will converge towards $\frac{\partial S_0}{\partial \theta}, \ldots, \frac{\partial S_T}{\partial \theta}$. Here we prove that it is actually the case by showing we can equivalently consider the evolution SDE for the asset's value and its sensitivity, that is, the vector SDE satisfied by $U_t = \left(S_t, \frac{\partial S_t}{\partial \theta}\right)$ and to then discretise it into $\widehat{U}_0, \widehat{U}_1, \ldots, \widehat{U}_N$ using the Euler or Milstein schemes or alternatively consider (as before) the Milstein discretisation for the evolution SDE of the asset S_t and to then differentiate it with respect to the parameter θ. That is, we will prove that for $k = 0, \ldots, N$,

$$\widehat{U}_k = \begin{pmatrix} \widehat{S}_k \\ \frac{\partial \widehat{S}_k}{\partial \theta} \end{pmatrix} \tag{3.3}$$

Note that, more generally, if we want the sensitivities of S_t with respect to n different parameters, we can obviously consider the $n+1$-dimensional process $U_t = \left(S_t, \frac{\partial S_t}{\partial \theta_1}, \ldots, \frac{\partial S_t}{\partial \theta_n}\right)$ still driven by the 1-dimensional Brownian motion W_t and compare its discretisation $\widehat{U}_0, \widehat{U}_1, \ldots, \widehat{U}_N$ to the result of differentiating the discretised asset's value $\widehat{S}_k$ with respect to each of the elements $\theta_1, \ldots, \theta_n$. As the number of sensitivities considered at once does not change the analysis, we focus only on the case $n = 1$ for the sake of readability.

From now on, we also assume for the sake of simplicity that the parameter θ does not have any effect on the size of the time step $h = T/N$. If it did (e.g. considering the sensitivity of the price to the time to expiry T), we would have to add extra terms into our equations to take into account the impact of the parameter on the discretisation itself but the analysis would be essentially the same.

The values $\widehat{U}_0, \widehat{U}_1, \ldots, \widehat{U}_N$ result from a proper Euler or Milstein discretisation of the process $U_t = \left(S_t, \frac{\partial S_t}{\partial \theta}\right)$. The properties of those schemes ensure they converge towards $U_0, U_1, \ldots, U_N$, i.e. for any fixed time $t = k\,h$, we have

$$\widehat{U}_k \underset{h\to 0}{\longrightarrow} \begin{pmatrix} S_t \\ \dfrac{\partial S_t}{\partial \theta} \end{pmatrix} \tag{3.4}$$

The importance of the discretisation-differentiation/differentiation-discretisation equivalence comes from the fact that if we prove it by showing that (3.3) holds, then we have for $k = 0, \ldots, N$,

$$\frac{\partial \widehat{S}_k}{\partial \theta} \underset{h\to 0}{\longrightarrow} \frac{\partial S_t}{\partial \theta} \tag{3.5}$$

that is, the naive estimators $\left(\dfrac{\partial \widehat{S}_k}{\partial \theta}\right)_{k=0,\ldots,N}$ used in chapter 2 do converge (as hoped) towards the sensitivities of the exact solution $\left(\dfrac{\partial S_t}{\partial \theta}\right)_{k=0,\ldots,N}$. The convergence speed is then determined by the weak and strong convergence properties of the schemes used (again see [Mil79], [Tal84], [KP92] or [GS72] and section 1.1.2).

3.1.2 Evolution SDE for the underlying asset's value and its sensitivity

Equation (1.2) means that for $t \in [0, T]$, the solution S_t can be written as

$$S_t = S_0 + \int_0^t a(S_u, u)\,\mathrm{d}u + \int_0^t b(S_u, u)\,\mathrm{d}W_u \tag{3.6}$$

and then, differentiating with respect to θ,

$$\frac{\partial S_t}{\partial \theta} = \frac{\partial S_0}{\partial \theta} + \frac{\partial}{\partial \theta}\int_0^t a(S_u, u)\,\mathrm{d}u + \frac{\partial}{\partial \theta}\int_0^t b(S_u, u)\,\mathrm{d}W_u \tag{3.7}$$

Intuitively, we would like to be able to conclude that we have

$$\begin{aligned} \frac{\partial S_t}{\partial \theta} = \frac{\partial S_0}{\partial \theta} &+ \int_0^t \left(\frac{\partial a}{\partial \theta}(S_u, u) + \frac{\partial a}{\partial S}(S_u, u)\frac{\partial S_u}{\partial \theta}\right)\mathrm{d}u \\ &+ \int_0^t \left(\frac{\partial b}{\partial \theta}(S_u, u) + \frac{\partial b}{\partial S}(S_u, u)\frac{\partial S_u}{\partial \theta}\right)\mathrm{d}W_u \end{aligned} \tag{3.8}$$

that is, in an infinitesimal form,

$$\mathrm{d}\frac{\partial S_t}{\partial \theta} = \left(\frac{\partial a}{\partial \theta}(S_t, t) + \frac{\partial a}{\partial S}(S_t, t)\frac{\partial S_t}{\partial \theta}\right)\mathrm{d}t + \left(\frac{\partial b}{\partial \theta}(S_t, t) + \frac{\partial b}{\partial S}(S_t, t)\frac{\partial S_t}{\partial \theta}\right)\mathrm{d}W_t \tag{3.9}$$

To get (3.8) from (3.7), we need some conditions ensuring it is possible to interchange the order of integration and differentiation in both integrals. Some interesting results enabling this interchange can be found in [Kar83] and [HN84]. We present here slightly more restrictive, yet more convenient and readily applicable conditions on the flows of SDE solutions. These are presented and proved in section V.7 of [Pro90].

We consider an n-dimensional process $X_t = \left(X_t^i\right)_{i=1,\dots,n}$ on the probability space $(\Omega, \mathcal{F}, P)$, solution of a stochastic differential equation of the form

$$X_t = x + \int_0^t F(X_s)\mathrm{d}Z_s \tag{3.10}$$

where $Z_t = (Z_t^\alpha)_{\alpha=1,\dots,m}$ is a continuous m-dimensional semimartingale with $Z_0^\alpha = 0$ for $\alpha = 1, \dots, m$. The vector $x = \left(x^i\right)_{i=1,\dots,n}$ corresponds to the initial values of X_t and $F(X_t)$ is an $n \times m$ matrix of functions $F_\alpha^i(X_t)$ from $\mathbb{R}^n$ to $\mathbb{R}$.

Clearly the solution of the equation depends on the set of initial parameters x and we can study the flow of the equation, that is, the function $\phi : x \to X(t, \omega, x)$ which can be seen as mapping $\mathbb{R}^n \to \mathbb{R}^n$ for (t, ω) fixed or as mapping $\mathbb{R}^n \to \mathcal{D}^n$ for ω fixed, where $\mathcal{D}^n$ is the space of càdlàg functions from $\mathbb{R}^+$ to $\mathbb{R}^n$ with the topology of uniform convergence on compacts (in practice we consider a finite expiry T and this topology is then the topology of uniform convergence on $[0, T]$).

Theorem 3.1.1. (Theorem 38 in [Pro90]) *Assuming that the functions $\left(F_\alpha^i(X_t)\right)_{i=1,\dots,n;\alpha=1,\dots,m}$ are globally Lipschitz on $\mathbb{R}^n$, then there exists a unique càdlàg solution $X(t, \omega, x)$ to equation* (3.10) *on $\mathbb{R}^+ \times \Omega \times \mathbb{R}^n$. For each x, the process $X(t, \omega, x)$ is a solution of the equation and for almost all ω, the flow $x \to X(., \omega, x)$ from $\mathbb{R}^n$ to $\mathcal{D}^n$ is continuous on $\mathbb{R}^n$ in the topology of uniform convergence on compacts.*

The previous theorem gives a result on the continuity of flows. The next one is an extension that deals with their differentiability.

Theorem 3.1.2. (Theorem 39 in [Pro90]) *If in addition to the hypotheses of theorem 3.1.1 we also assume that the functions $\left(F_\alpha^i(X_t)\right)$ have bounded locally Lipschitz first order derivatives, then* (3.10) *has a unique solution $X(t, \omega, x)$ that for almost all ω is continuously differentiable in x.*

For any $k \in [1, n]$, the partial derivative $\dfrac{\partial X}{\partial x_k}(t, \omega, x)$ is then the solution of

$$\frac{\partial X_t^i}{\partial x^k} = \delta_i^k + \sum_{\alpha=1}^m \sum_{j=1}^n \int_0^t \frac{\partial F_\alpha^i}{\partial X^j}(X_s) \frac{\partial X_s^j}{\partial x^k} \mathrm{d}Z_s^\alpha \tag{3.11}$$

where δ_i^k is Kronecker's delta symbol.

Note that as explained in sections V.7 and V.8 of [Pro90], theorem 3.1.2 can be extended to prove that the flow of the solution X_t is N times continuously differentiable in x if we also assume the functions $F_\alpha^i(X_t)$ have locally Lipschitz partial derivatives up to order N (see theorem 40 in [Pro90]).

For our practical application of this result to equation (3.6), we consider X_t, the vector process containing the underlying price S_t as well as different parameters of interest with respect to which we want to differentiate the price (e.g. volatility σ, interest rate r, initial value S_0), that is,

$$x = \begin{pmatrix} S_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}$$

and

$$X_t = \begin{pmatrix} S_t \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}$$

All but the first element of X_t are constant, which corresponds to the case where $\left(F_\alpha^i(X_t)\right)_{i=2,\dots,n} = 0$. Usually $\theta_1, \dots, \theta_n$ are taken to be independent parameters and the sensitivities are then

$$\frac{\partial X^i}{\partial \theta_k}(t) = \begin{pmatrix} \frac{\partial S_t}{\partial \theta_k} \\ \frac{\partial \theta_1}{\partial \theta_k} \\ \frac{\partial \theta_2}{\partial \theta_k} \\ \vdots \\ \frac{\partial \theta_n}{\partial \theta_k} \end{pmatrix} = \begin{pmatrix} \frac{\partial S_t}{\partial \theta_k} \\ \delta_1^k \\ \delta_2^k \\ \dots \\ \delta_n^k \end{pmatrix} \tag{3.12}$$

The continuous semimartingale with respect to which we integrate is

$$Z_t = \begin{pmatrix} W_t \\ t \end{pmatrix}$$

and the matrix F is then

$$F = \begin{pmatrix} b(S,t) & a(S,t) \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix}$$

We can apply theorem 3.1.2 that ensures that if $a(S,t)$ and $b(S,t)$ are Lipschitz and their first order partial derivatives are bounded and locally Lipschitz, then the flow $\theta_k \to S_t$ is differentiable and $\dfrac{\partial S_t}{\partial \theta_k}$ is the solution of

$$\frac{\partial S_t}{\partial \theta_k} = \frac{\partial S_0}{\partial \theta_k} + \sum_{j=1}^{n} \int_0^t \frac{\partial F^i_\alpha}{\partial X^j}(X_s) \frac{\partial X^j_s}{\partial x^k} \mathrm{d}W_s + \sum_{j=1}^{n} \int_0^t \frac{\partial F^i_\alpha}{\partial X^j}(X_s) \frac{\partial X^j_s}{\partial x^k} \mathrm{d}s \tag{3.13}$$

Noting that for $j \geq 2$ the term $\dfrac{\partial X^j_s}{\partial x^k}$ of equation (3.11) is $\dfrac{\partial \theta_j}{\partial \theta_k} = \delta^k_j$, this becomes

$$\begin{aligned} \frac{\partial S_t}{\partial \theta_k} = \frac{\partial S_0}{\partial \theta_k} &+ \left(\int_0^t \frac{\partial b}{\partial S}(X_s, s) \frac{\partial S_s}{\partial \theta^k} + \frac{\partial b}{\partial \theta_k}(X_s, s) \right) \mathrm{d}W_s \\ &+ \left(\int_0^t \frac{\partial a}{\partial S}(X_s, s) \frac{\partial S_s}{\partial \theta^k} + \frac{\partial a}{\partial \theta_k}(X_s, s) \right) \mathrm{d}s \end{aligned} \tag{3.14}$$

This means that assuming $a\,(S, t)$ and $b\,(S, t)$ are Lipschitz and have bounded locally Lipschitz first order derivatives, the solution of (1.2) and its sensitivity with respect to θ follow the joint evolution SDE (3.15)

$$\begin{aligned} \mathrm{d}S(t) &= a(S_t, t)\,\mathrm{d}t + b(S_t, t)\,\mathrm{d}W_t \\ \mathrm{d}\frac{\partial S_t}{\partial \theta} &= \left(\frac{\partial a}{\partial \theta}(S_t, t) + \frac{\partial a}{\partial S}(S_t, t) \frac{\partial S_t}{\partial \theta} \right) \mathrm{d}t \\ &\qquad + \left(\frac{\partial b}{\partial \theta}(S_t, t) + \frac{\partial b}{\partial S}(S_t, t) \frac{\partial S_t}{\partial \theta} \right) \mathrm{d}W_t \end{aligned} \tag{3.15}$$

Note the second term of this joint evolution SDE does indeed correspond to the naively differentiated equation we hoped for in (3.9).

We can rewrite (3.15) as a vector equation describing the evolution of the 2-dimensional process $U_t = \left(S_t, \dfrac{\partial S_t}{\partial \theta} \right)$ driven by a single 1-dimensional Brownian motion W_t.

$$\begin{aligned} \mathrm{d}U_t = &\begin{pmatrix} a(S, t) \\ \\ \frac{\partial a}{\partial \theta}(S_t, t) + \frac{\partial a}{\partial S}(S_t, t) \frac{\partial S_t}{\partial \theta} \end{pmatrix} \mathrm{d}t \\ &\qquad + \begin{pmatrix} b(S, t) \\ \\ \frac{\partial b}{\partial \theta}(S_t, t) + \frac{\partial b}{\partial S}(S_t, t) \frac{\partial S_t}{\partial \theta} \end{pmatrix} \mathrm{d}W_t \end{aligned} \tag{3.16}$$

3.1.3 Differentiation of the discretisation/discretisation of the differentiated SDE

We now compare the formulas resulting from the differentiation of the discretisation of the evolution SDE for S_t and the ones resulting from the discretisation of the evolution SDE for $U_t = \left(S_t, \dfrac{\partial S_t}{\partial \theta}\right)$.

3.1.3.1 Differentiation of the Euler discretisation

We have seen in section 1.1.2 that the Euler discretisation of equation (1.2) is

$$\widehat{S}_{n+1} = \widehat{S}_n + a\left(\widehat{S}_n, t_n\right) h + b\left(\widehat{S}_n, t_n\right) \Delta W_n \tag{3.17}$$

Assuming the coefficients $a(S,t)$ and $b(S,t)$ are differentiable with respect to S and θ, we differentiate (3.17) with respect to θ and obtain

$$\begin{aligned}\frac{\partial \widehat{S}_{n+1}}{\partial \theta} = \frac{\partial \widehat{S}_n}{\partial \theta} &+ \left(\frac{\partial a}{\partial \theta}\left(\widehat{S}_n, t_n\right) + \frac{\partial a}{\partial S}\left(\widehat{S}_n, t_n\right)\frac{\partial \widehat{S}_n}{\partial \theta}\right) h \\ &+ \left(\frac{\partial b}{\partial \theta}\left(\widehat{S}_n, t_n\right) + \frac{\partial b}{\partial S}\left(\widehat{S}_n, t_n\right)\frac{\partial \widehat{S}_n}{\partial \theta}\right) \Delta W_n\end{aligned} \tag{3.18}$$

3.1.3.2 Euler discretisation of the differentiated equation

Now that we have proved that (3.16) is the evolution SDE for $U_t = \left(S_t, \dfrac{\partial S_t}{\partial \theta}\right)$, we can apply the multidimensional Euler discretisation as described in section 10.2 of [KP92]. For $n = 0, \ldots, N-1$, we obtain

$$\begin{aligned}\widehat{U}_{n+1} = \widehat{U}_n &+ \begin{pmatrix} a(S_{t_n}, t_n) \\ \dfrac{\partial a}{\partial \theta}(S_{t_n}, t_n) + \dfrac{\partial a}{\partial S}(S_{t_n}, t_n)\dfrac{\partial S_{t_n}}{\partial \theta} \end{pmatrix} h \\ &+ \begin{pmatrix} b(S_{t_n}, t_n) \\ \dfrac{\partial b}{\partial \theta}(S_{t_n}, t_n) + \dfrac{\partial b}{\partial S}(S_{t_n}, t_n)\dfrac{\partial S_{t_n}}{\partial \theta} \end{pmatrix} \Delta W_n\end{aligned} \tag{3.19}$$

Comparing it to (3.18), this proves that (3.3) holds for the Euler scheme.

3.1.3.3 Differentiation of the Milstein discretisation

As seen in section 1.1.2, for $n = 0, \ldots, N-1$, the Milstein discretisation of equation (1.2) is

$$\begin{aligned}\widehat{S}_{n+1} = \widehat{S}_n &+ a\left(\widehat{S}_n, t_n\right) h + b\left(\widehat{S}_n, t_n\right) \Delta W_n \\ &+ \frac{1}{2} b\left(\widehat{S}_n, t_n\right) \frac{\partial b}{\partial S}\left(\widehat{S}_n, t_n\right) \left(\Delta W_n^2 - h\right)\end{aligned} \tag{3.20}$$

Assuming that $\dfrac{\partial a}{\partial S}, \dfrac{\partial b}{\partial S}, \dfrac{\partial a}{\partial \theta}, \dfrac{\partial b}{\partial \theta}, \dfrac{\partial^2 b}{\partial S^2}, \dfrac{\partial^2 b}{\partial S \partial \theta}$ exist, differentiating (3.20) with respect to θ gives

$$\begin{aligned}
\frac{\partial \widehat{S}_{n+1}}{\partial \theta} = \frac{\partial \widehat{S}_n}{\partial \theta} & \\
&+ \left(\frac{\partial \widehat{S}_n}{\partial \theta} \frac{\partial a}{\partial S} \left(\widehat{S}_n, t_n\right) + \frac{\partial a}{\partial \theta} \left(\widehat{S}_n, t_n\right) \right) h \\
&+ \left(\frac{\partial \widehat{S}_n}{\partial \theta} \frac{\partial b}{\partial S} \left(\widehat{S}_n, t_n\right) + \frac{\partial b}{\partial \theta} \left(\widehat{S}_n, t_n\right) \right) \Delta W_n \\
&+ \frac{1}{2} \left[\frac{\partial \widehat{S}_n}{\partial \theta} \left(\frac{\partial b}{\partial S} \left(\widehat{S}_n, t_n\right) \right)^2 + \frac{\partial \widehat{S}_n}{\partial \theta} b \frac{\partial^2 b}{\partial S^2} \left(\widehat{S}_n, t_n\right) \right. \\
&\qquad \left. + \frac{\partial b}{\partial \theta} \frac{\partial b}{\partial S} \left(\widehat{S}_n, t_n\right) + b \frac{\partial^2 b}{\partial \theta \partial S} \left(\widehat{S}_n, t_n\right) \right] \left(\Delta W_n^2 - h\right)
\end{aligned} \tag{3.21}$$

Writing $\widehat{\delta}_n$ for $\dfrac{\partial \widehat{S}_n}{\partial \theta}$, this is

$$\begin{aligned}
\widehat{\delta}_{n+1} = \widehat{\delta}_n &+ \left(\widehat{\delta}_n \frac{\partial a}{\partial S} + \frac{\partial a}{\partial \theta} \right) \left(\widehat{S}_n, t_n\right) h \\
&+ \left(\widehat{\delta}_n \frac{\partial b}{\partial S} + \frac{\partial b}{\partial \theta} \right) \left(\widehat{S}_n, t_n\right) \Delta W_n \\
&+ \frac{1}{2} \left[\widehat{\delta}_n \left(\frac{\partial b}{\partial S} \right)^2 + \widehat{\delta}_n b \frac{\partial^2 b}{\partial S^2} \right. \\
&\qquad \left. + \frac{\partial b}{\partial \theta} \frac{\partial b}{\partial S} + b \frac{\partial^2 b}{\partial \theta \partial S} \right] \left(\widehat{S}_n, t_n\right) \left(\Delta W_n^2 - h\right)
\end{aligned} \tag{3.22}$$

3.1.3.4 Milstein discretisation of the differentiated equation

We can apply the multidimensional Milstein discretisation to (3.16), as described in section 10.3 of [KP92]. Note that even though U_t is technically a multidimensional SDE, its Milstein scheme is still easily computed as the driving Brownian motion is only 1-dimensional. We get as in p346 of [KP92] that

$$\left\{ \begin{aligned}
\widehat{S}_{n+1} &= \widehat{S}_n + a \left(\widehat{S}_n, t_n\right) h + b \left(\widehat{S}_n, t_n\right) \Delta W_n + \frac{1}{2} b \frac{\partial b}{\partial S} \left(\widehat{S}_n, t_n\right) \left(\Delta W_n^2 - h\right) \\
\widehat{\delta}_{n+1} &= \widehat{\delta}_n + \left(\widehat{\delta}_n \frac{\partial a}{\partial S} + \frac{\partial a}{\partial \theta} \right) \left(\widehat{S}_n, t_n\right) h \\
&\qquad + \left(\widehat{\delta}_n \frac{\partial b}{\partial S} + \frac{\partial b}{\partial \theta} \right) \left(\widehat{S}_n, t_n\right) \Delta W_n \\
&\qquad + \frac{1}{2} \left[\widehat{\delta}_n \left(\frac{\partial b}{\partial S} \right)^2 + \widehat{\delta}_n b \frac{\partial^2 b}{\partial S^2} \right. \\
&\qquad\qquad \left. + \frac{\partial b}{\partial \theta} \frac{\partial b}{\partial S} + b \frac{\partial^2 b}{\partial \theta \partial S} \right] \left(\widehat{S}_n, t_n\right) \left(\Delta W_n^2 - h\right)
\end{aligned} \right. \tag{3.23}$$

Comparing it to (3.22), this proves that (3.3) holds for the Milstein scheme.

3.1.3.5 Conclusion

We have checked that it is equivalent to consider the discretisation of the multi-dimensional SDE describing the evolution of the underlying asset and its sensitivity $U_t = \left(S_t, \frac{\partial S_t}{\partial \theta}\right)$ or to consider the discretisation of the evolution SDE for S_t and to then differentiate it with respect to the parameter θ.

This result confirms that the naive approach adopted in chapter 2 was valid and did indeed yield estimators of $\frac{\partial \widehat{S}_0}{\partial \theta}, \ldots, \frac{\partial \widehat{S}_N}{\partial \theta}$ that converged towards $\frac{\partial S_0}{\partial \theta}, \ldots, \frac{\partial S_T}{\partial \theta}$ with convergence speeds determined by the usual convergence properties of the schemes used.

As already mentioned at the beginning of section 3.1, the result is here written and proved by considering only one sensitivity $\frac{\partial S_t}{\partial \theta}$ but it still holds when considering any number of Greeks.

3.1.4 Note on Brownian Bridge midpoints

Until now we have dealt with points resulting from a standard discretisation scheme (Euler or Milstein) applied to the evolution SDE. As explained in chapter 2, we also use Brownian Bridge interpolants to compute midpoint values within each time step of the coarse level of discretisation (e.g. for the Lookback option in section 2.5). We can also show that the midpoint values constructed this way and their sensitivities also converge quickly towards the exact underlying values and sensitivities.

To prove this result, we consider the Milstein discretisation $\widehat{S}_0, \widehat{S}_2, \ldots, \widehat{S}_{N_f}$ based on the coarse time grid $t_0, t_2, \ldots, t_{N_f-2}, t_{N_f}$ of width $h_c = 2h$. We consider a given coarse step $[t_{2k}, t_{2k+2}]$ and define the point $\widehat{S}^M_{2k+1}$ resulting from a Milstein discretisation on the subinterval $[t_{2k}, t_{2k+1}]$. We show that the Brownian Bridge midpoint $\widehat{S}^{BB}_{2k+1}$ and its sensitivity $\frac{\partial \widehat{S}^{BB}_{2k+1}}{\partial \theta}$ converge quickly towards $\widehat{S}^M_{2k+1}$ and $\frac{\partial \widehat{S}^M_{2k+1}}{\partial \theta}$ and therefore towards S_{2k+1} and $\frac{\partial S_{2k+1}}{\partial \theta}$.

Using the notations $\dot{a}_n = \frac{\partial a_n}{\partial S_n}, \dot{b}_n = \frac{\partial b_n}{\partial S_n}, \widetilde{a}_n = \frac{\partial a_n}{\partial \theta}, \widetilde{b}_n = \frac{\partial b_n}{\partial \theta}, \widehat{\delta}_n = \frac{\partial \widehat{S}_n}{\partial \theta}$, this Milstein midpoint is defined as

$$
\begin{aligned}
\widehat{S}^M_{2k+1} &= \widehat{S}_{2k} + a_{2k}h + b_{2k}\Delta W^f_{2k} + \frac{1}{2} b_n \dot{b}_n \left(\left(\Delta W^f_{2k}\right)^2 - h\right) \\
\widehat{\delta}^M_{2k+1} &= \widehat{\delta}_{2k} + \left(\widetilde{a}_{2k} + \dot{a}_{2k}\widehat{\delta}_{2k}\right) h + \left(\widetilde{b}_{2k} + \dot{b}_{2k}\widehat{\delta}_{2k}\right) \Delta W^f_{2k} \\
&\quad + \frac{1}{2}\left[\left(\widetilde{b}_{2k} + \dot{b}_{2k}\widehat{\delta}_{2k}\right)\dot{b}_{2k}\right. \\
&\qquad\qquad \left. + b_{2k}\left(\frac{\partial^2 b_{2k}}{\partial S_{2k}\partial\theta} + \frac{\partial^2 b_{2k}}{\partial S_{2k}^2}\widehat{\delta}_{2k}\right)\right]\left(\left(\Delta W^f_{2k}\right)^2 - h\right)
\end{aligned}
\tag{3.24}
$$

This point corresponds to a Milstein discretisation with non-uniform time steps: steps of width $2h$ between t_0 and t_{2k}, a step of width h between t_{2k} and t_{2k+1}. The largest time step is $2h$, the convergence properties of the Milstein scheme prove that $\left(\widehat{S}^M_{2k+1}, \dfrac{\partial \widehat{S}^M_{2k+1}}{\partial\theta}\right)$ converges towards $\left(S_{t_{2k+1}}, \dfrac{\partial S_{t_{2k+1}}}{\partial\theta}\right)$ and the order of weak and strong convergence are both 1 as $h \to 0$.

Now, the midpoint value based on the Brownian Bridge interpolant can be written as

$$\begin{aligned}\widehat{S}^{BB}_{2k+1} &= \widehat{S}_{2k} + \frac{1}{2}\left(\widehat{S}_{2k+2} - \widehat{S}_{2k}\right) + b_{2k}\left(W_{2k+1} - W_{2k} - \frac{1}{2}\left(W_{2k+2} - W_{2k}\right)\right)\\ &= \widehat{S}_{2k} + a_{2k}h + b_{2k}\Delta W^f_{2k} + \frac{1}{2} b_{2k}\dot{b}_{2k}\left((\Delta W^c_{2k})^2 - h_c\right)\end{aligned} \tag{3.25}$$

Note that

$$\begin{aligned}(\Delta W^c_{2k})^2 - h_c &= \left(\Delta W^f_{2k} + \Delta W^f_{2k+1}\right)^2 - 2h\\ &= \left(\left(\Delta W^f_{2k}\right)^2 - h\right) + \left(\left(\Delta W^f_{2k+1}\right)^2 + 2\Delta W^f_{2k}\Delta W^f_{2k+1} - h\right)\end{aligned} \tag{3.26}$$

and the difference between the two midpoints can then be written

$$\widehat{S}^{BB}_{2k+1} - \widehat{S}^M_{2k+1} = \frac{1}{2} b_{2k}\dot{b}_{2k}\left(\left(\Delta W^f_{2k+1}\right)^2 + 2\Delta W^f_{2k}\Delta W^f_{2k+1} - h\right) \tag{3.27}$$

From there, assuming $b\left(S,t\right), \dot{b}\left(S,t\right)$ are Lipschitz and using the results of theorems 3.4.1 and 3.4.3 on the finiteness of moments of the solution of the SDE and its Milstein discretisation, we see that

$$\begin{aligned}\mathbb{E}\left[\widehat{S}^{BB}_{2k+1} - \widehat{S}^M_{2k+1}\right] &= 0\\ \mathbb{E}\left[\left(\widehat{S}^{BB}_{2k+1} - \widehat{S}^M_{2k+1}\right)^2\right] &= O\left(h^2\right)\end{aligned} \tag{3.28}$$

The second line of (3.28) means the Brownian Bridge midpoint $\widehat{S}^{BB}_{2k+1}$ converges strongly towards $\widehat{S}^M_{2k+1}$ with order 1. Now we look at the convergence of the sensitivities. We write

$$\begin{aligned}\widehat{\delta}^{BB}_{2k+1} - \widehat{\delta}^M_{2k+1} &= \frac{\partial \widehat{S}^{BB}_{2k+1} - \widehat{S}^M_{2k+1}}{\partial\theta}\\ &= \frac{1}{2}\left(\frac{\partial}{\partial\theta} + \widehat{\delta}_{2k}\frac{\partial}{\partial S_{2k}}\right)\left(b_{2k}\dot{b}_{2k}\right)\\ &\qquad\qquad \left[\left(\Delta W^f_{2k+1}\right)^2 + 2\Delta W^f_{2k}\Delta W^f_{2k+1} - h\right]\end{aligned} \tag{3.29}$$

and as before, this leads to

$$\begin{aligned}\mathbb{E}\left[\widehat{\delta}^{BB}_{2k+1} - \widehat{\delta}^M_{2k+1}\right] &= 0\\ \mathbb{E}\left[\left(\widehat{\delta}^{BB}_{2k+1} - \widehat{\delta}^M_{2k+1}\right)^2\right] &= O\left(h^2\right)\end{aligned} \tag{3.30}$$

The second line means the Brownian Bridge midpoint $\widehat{\delta}_{2k+1}^{BB}$ converges strongly towards $\widehat{\delta}_{2k+1}^{M}$ with order 1.

Putting all these results together, we have a proof that the midpoint values $\left(\widehat{S}_{2k+1}^{BB}, \frac{\partial \widehat{S}_{2k+1}^{BB}}{\partial \theta}\right)$ constructed using a Brownian Bridge interpolation also converge towards the exact values $\left(S_{t_{2k+1}}, \frac{\partial S_{t_{2k+1}}}{\partial \theta}\right)$ with a strong convergence of order 1.

3.2 Applicability of pathwise sensitivity Greeks

As seen in section 1.3.1, for the pathwise sensitivity approach to give unbiased estimators, it is essential that the following re-ordering of the expectation and the partial differentiation holds

$$\frac{\partial \mathbb{E}(P)}{\partial \theta} = \mathbb{E}\left(\frac{\partial P}{\partial \theta}\right) \tag{3.31}$$

We first present a lemma that provides conditions ensuring this interchange is valid. We then proceed to show those conditions are met in the Black & Scholes case and, under certain conditions, in the general case of the exact solution of an Ito evolution equation. Finally we prove that the interchange is also valid when dealing with a discretised version of the underlying Ito process.

3.2.1 Conditions of unbiasedness of pathwise sensitivities

Let θ be a scalar parameter in an open interval Θ and $X(\theta)$ be a vector-valued process. We write $\widetilde{X}(\theta) = (X_1(\theta), \ldots, X_N(\theta))$ its values at various discrete times $t_1, \ldots, t_n$. We consider a real-valued payoff function $P\left(\widetilde{X}(\theta)\right)$. The following lemma is derived from [Gla88] and [BG96].

Lemma 3.2.1. *Assuming the following conditions*

- *A1: For each n, for all $\theta \in \Theta$,*

$$X_n'(\theta) := \lim_{\Delta\theta \to 0} \frac{X_n(\theta + \Delta\theta) - X_n(\theta)}{\Delta\theta} \tag{3.32}$$

  exists with probability 1.

- *A2: Let D_P be the set of points where P is differentiable. For all $\theta \in \Theta$,*

$$\mathbb{P}\left(\widetilde{X}(\theta) \in D_P\right) = 1 \tag{3.33}$$

then the discounted payoff has a pathwise derivative given by

$$\frac{\mathrm{d}P}{\mathrm{d}\theta}\left(\widetilde{X}(\theta)\right) = \sum_{n=1}^{N}\left(\left[\nabla_{x_n} P\left(\widetilde{X}(\theta)\right)\right]^t X_n'(\theta)\right) \tag{3.34}$$

where $\nabla_{x_n} P$ denotes the vector of partial derivatives of P with respect to the components of X_n.

If we also assume that

- *A3: P is k_P-Lipschitz, i.e. there exists a constant k_P such that*

$$| P(X) - P(Y) | \leq k_P \, \|X - Y\| \tag{3.35}$$

  for all X, Y in the domain of P.

- *A4: there exist random variables $(K_n)_{n=1,\ldots,N}$ such that*

$$\|X_n(\theta_2) - X_n(\theta_1)\| \leq K_n | \theta_2 - \theta_1 | \tag{3.36}$$

  for all $\theta_1, \theta_2 \in \Theta$ with $\mathbb{E}(K_n) < \infty$ for all n.

- *$\widetilde{A4}$: Condition A4 is easy to check but slightly restrictive, it can be replaced by the more general condition that for all n the family of functions*

$$\Delta_n(\theta_1, \theta_2) := \frac{X_n(\theta_2) - X_n(\theta_1)}{\theta_2 - \theta_1} \tag{3.37}$$

  be uniformly integrable for all $\theta_1, \theta_2 \in \Theta$.

Then for every $\theta \in \Theta$, $\dfrac{\partial \mathbb{E}(P(X))}{\partial \theta}$ exists and

$$\frac{\partial \mathbb{E}(P(X))}{\partial \theta} = \mathbb{E}\left(\frac{\partial P(X(\theta))}{\partial \theta}\right) \tag{3.38}$$

Proof. With condition $A4$, the validity of (3.38) is a direct consequence of the dominated convergence theorem. The detailed proof can be found in appendix A of [BG96].

With condition $\widetilde{A4}$, the proof carries over verbatim using the Vitali convergence theorem (see for example [Rud86]) instead of the dominated convergence theorem. □

In practice conditions $A1$, $A2$ and $A4$ or $\widetilde{A4}$ are usually satisfied and condition $A3$ is the key requirement for making pathwise sensitivities applicable.

Note that for vanilla options, the payoff P only depends on the price of the underlying asset at expiry S_T. Then, $N = 1$ and $\nabla_x P = \dfrac{\partial P}{\partial S_T}$.

Also note that obviously the case of a direct dependence of P on θ, i.e. the case where we can write $P(X)$ as $P(\theta, X_1(\theta), \ldots, X_N(\theta))$ is encompassed by formula (3.34). It can be seen as a degenerate case where we have an additional variable $X_{N+1} = \theta$. Then we have

$$
\begin{aligned}
\frac{\mathrm{d}P}{\mathrm{d}\theta}\left(X\left(\theta\right)\right) &= \sum_{n=1}^{N+1}\left(\left[\nabla_{x_n} P\left(\widetilde{X}\left(\theta\right)\right)\right]^t X_n'\left(\theta\right)\right) \\
&= \sum_{n=1}^{N}\left(\left[\nabla_{x_n} P\left(\widetilde{X}\left(\theta\right)\right)\right]^t X_n'\left(\theta\right)\right) + \frac{\partial P}{\partial \theta}\left(X\left(\theta\right)\right)
\end{aligned}
\tag{3.39}
$$

3.2.2 Applicability of pathwise sensitivities in the Black & Scholes model

Note that for any parameter θ (e.g. S_0, r, σ), the set of realistic values Θ is clearly bounded. We can legitimately assume there is a closed interval $[\theta_{\min}, \theta_{\max}] \subset \mathbb{R}$ containing Θ.

In the Black & Scholes model, we know (see equation (2.3)) that the asset's price at time t_n is

$$
S_{t_n} = S_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right) t_n + \sigma\, W_{t_n}\right) \tag{3.40}
$$

For any parameter θ we may consider, the function $\theta \to S_{t_n}(\theta)$ is smooth. Condition $A1$ is verified.

The smooth density function of S_{t_n} (see equation (2.4)) ensures that if P is almost everywhere differentiable (e.g. call option, digital call, Asian call, lookback option, barrier option,...), then condition $A2$ is verified.

Condition $A3$ means that the payoff function P must be Lipschitz. This clearly eliminates the digital call or barrier options. Call options or Asian calls satisfy $A3$.

For $\theta = S_0$

$$
\mid S_{t_n}(S_{0,2}) - S_{t_n}(S_{0,1}) \mid = \mid S_{0,2} - S_{0,1} \mid \exp\left(\left(r - \frac{\sigma^2}{2}\right) t_n + \sigma\, W_{t_n}\right) \tag{3.41}
$$

and

$$
\mathbb{E}\left[\exp\left(\left(r - \frac{\sigma^2}{2}\right) t_n + \sigma\, W_{t_n}\right)\right] = \exp\left(r t_n\right) < \infty \tag{3.42}
$$

therefore condition $A4$ is verified.

For $\theta = r$, we have

$$
\begin{aligned}
\mid S_{t_n}(r_2) - S_{t_n}(r_1) \mid = S_0 \exp\left(-\frac{\sigma^2}{2} t_n + \sigma\, W_{t_n}\right) \\
\mid \exp\left(r_2 t_n\right) - \exp\left(r_1 t_n\right) \mid
\end{aligned}
\tag{3.43}
$$

and for a given time t_n, $r \to \exp\left(r t_n\right)$ is Lipschitz on $[r_{\min}, r_{\max}]$, which means there is a constant c_n such that for any r_1, r_2,

$$
\mid S_{t_n}(r_2) - S_{t_n}(r_1) \mid \leq S_0 \exp\left(-\frac{\sigma^2}{2} t_n + \sigma\, W_{t_n}\right) c_n \mid r_2 - r_1 \mid \tag{3.44}
$$

and as before

$$\mathbb{E}\left[S_0 \exp\left(-\frac{\sigma^2}{2}t_n + \sigma\, W_{t_n}\right)c_n\right] < \infty \tag{3.45}$$

Condition $A4$ is satisfied.

For $\theta = \sigma$, we can use the mean value theorem to write that for any σ_1, σ_2 and any W_{t_n} there exists some value $\sigma_3 \in [\sigma_1, \sigma_2]$ such that

$$\left(S_{t_n}(\sigma_2) - S_{t_n}(\sigma_1)\right) = \frac{\partial S_{t_n}}{\partial \sigma}(\sigma_3)\,(\sigma_2 - \sigma_1) \tag{3.46}$$

Therefore

$$|\, S_{t_n}(\sigma_2) - S_{t_n}(\sigma_1)\,| \leq \max_{\sigma \in [\sigma_{\min}, \sigma_{\max}]} \left|\frac{\partial S_{t_n}}{\partial \sigma}(\sigma)\right| |\, \sigma_2 - \sigma_1\,| \tag{3.47}$$

We want the maximum value of

$$\left|\frac{\partial S_{t_n}}{\partial \sigma}\right| = \left| S_0\left(-\sigma t_n + W_{t_n}\right)\exp\left(\left(r - \frac{\sigma^2}{2}\right)t_n + \sigma\, W_{t_n}\right)\right| \tag{3.48}$$

We show that for W_{t_n} given, it is reached for $\sigma_- = \dfrac{W_{t_n} - \sqrt{t_n}}{t_n}$ or $\sigma_+ = \dfrac{W_{t_n} + \sqrt{t_n}}{t_n}$ if σ_- or σ_+ are in $[\sigma_1, \sigma_2]$ or σ_1 or σ_2 if $\sigma_-, \sigma_+ \notin [\sigma_1, \sigma_2]$. Then,

$$\begin{aligned} \left|\frac{\partial S_{t_n}}{\partial \sigma}(\sigma_-)\right| = \left|\frac{\partial S_{t_n}}{\partial \sigma}(\sigma_+)\right| &= S_0\sqrt{t_n}\exp\left(r t_n - \frac{1}{2} + \frac{W_{t_n}^2}{2t_n}\right) \\ \left|\frac{\partial S_{t_n}}{\partial \sigma}(\sigma_i)\right| &= S_0\,|-\sigma_i t_n + W_{t_n}\,|\exp\left(\left(r - \frac{\sigma_i^2}{2}\right)t_n + \sigma_i\, W_{t_n}\right) \end{aligned} \tag{3.49}$$

which means that

$$|\, S_{t_n}(\sigma_2) - S_{t_n}(\sigma_1)\,| \leq K_n\left(W_{t_n}\right)|\, \sigma_2 - \sigma_1\,| \tag{3.50}$$

where K_n is the random variable defined as

$$\begin{aligned} K_n &:= \left|\frac{\partial S_{t_n}}{\partial \sigma}(\sigma_-)\right| & &\text{if } W_{t_n} \in \left[\sigma_1 t_n - \sqrt{t_n}, \sigma_2 t_n + \sqrt{t_n}\right] \\ K_n &:= \left|\frac{\partial S_{t_n}}{\partial \sigma}(\sigma_1)\right| + \left|\frac{\partial S_{t_n}}{\partial \sigma}(\sigma_2)\right| & &\text{if } W_{t_n} \notin \left[\sigma_1 t_n - \sqrt{t_n}, \sigma_2 t_n + \sqrt{t_n}\right] \end{aligned} \tag{3.51}$$

which is clearly integrable, $\mathbb{E}\left(K_n\right) < \infty$, condition $A4$ is satisfied again.

We conclude that from lemma 3.2.1 that pathwise sensitivities can be applied to the exact solutions of the evolution SDE in the Black & Scholes model with almost-everywhere differentiable Lipschitz payoffs.

3.2.3 Applicability of pathwise sensitivities in the general Ito model

Now that we have proved that pathwise sensitivities were applicable for certain payoffs in the case of the Black & Scholes model, let us consider the more general case of equation (1.2).

Assuming the SDE has Lipschitz coefficients with bounded locally Lipschitz first order derivatives, theorem 3.1.2 ensures that the flow of the solution is differentiable with respect to the parameter θ and conditions $A1$ is satisfied for the asset's prices $S_{t_1}, S_{t_2}, \ldots, S_T$.

Assuming the payoff function P is almost everywhere differentiable and that the prices $S_{t_1}, S_{t_2}, \ldots, S_T$ have smooth density functions (see section 3.3 for conditions ensuring this, e.g. $b_\theta(S_0, t_0) \neq 0$), then $A2$ is also satisfied.

Condition $A3$ is still a condition meaning that P has to be Lipschitz; it is no different from what we had in section 3.2.2.

Condition $A4$ is too restrictive here and we will prove that $\widetilde{A4}$ is satisfied instead. To do this, we now present a stochastic extension of the Gronwall inequality introduced in [Ama05].

Theorem 3.2.2. *We let $M_\omega^2[0,T]$ be the set of real-valued random variables f parametrised by $t \in [0,T]$ such that*

$$\mathbb{E}\left[\int_0^T f^2(\omega,t)\mathrm{d}t\right] < \infty$$

Assume that $\xi(\omega,t)$ and $\eta(\omega,t)$ belong to $M_\omega^2[0,T]$. If there exist functions $a(\omega,t)$ and $b(\omega,t)$ belonging to $M_\omega^2[0,T]$ such that

$$|\,\xi(\omega,t)\,| \leq \left| \int_0^t a(\omega,s)\mathrm{d}s + \int_0^t b(\omega,s)\mathrm{d}W_s \right| \tag{3.52}$$

and there are nonnegative constants $\alpha_0, \alpha_1, \beta_0, \beta_1$ such that

$$\begin{aligned} |\,a(\omega,t)\,| &\leq \alpha_0\,|\,\eta(\omega,t)\,| + \alpha_1\,|\,\xi(\omega,t)\,| \\ |\,b(\omega,t)\,| &\leq \beta_0\,|\,\eta(\omega,t)\,| + \beta_1\,|\,\xi(\omega,t)\,| \end{aligned} \tag{3.53}$$

for $0 \leq t \leq T$, then we have

$$\mathbb{E}\left[\xi^2(\omega,t)\right] \leq 4\left(\alpha_0\sqrt{t}+\beta_0\right)^2 \exp\left(4t\left(\alpha_1\sqrt{t}+\beta_1\right)^2\right)\int_0^t \mathbb{E}\left[\eta^2(\omega,s)\right]\mathrm{d}s \tag{3.54}$$

Proof. See [Ama05] □

Applying it to our problem, we can write that, for a solution of (1.2) with a, b depending on some parameter θ,

$$\begin{aligned} |\,S_t(\theta_2) - S_t(\theta_1)\,| \leq \Bigg| &\int_0^t \left(a_{\theta_2}\left(S_s(\theta_2),s\right) - a_{\theta_1}\left(S_s(\theta_1),s\right)\right)\mathrm{d}s \\ &+ \int_0^t \left(b_{\theta_2}\left(S_s(\theta_2),s\right) - b_{\theta_1}\left(S_s(\theta_1),s\right)\right)\mathrm{d}W_s \Bigg| \end{aligned} \tag{3.55}$$

Assuming a and b are Lipschitz in their parameters, we have

$$\begin{aligned}
|\, a_{\theta_2}\left(S_s(\theta_2), s\right) - a_{\theta_1}\left(S_s(\theta_1), s\right) | &\leq \alpha_1 \,|\, S_s(\theta_2) - S_s(\theta_1) \,| + \alpha_0 \,|\, \theta_2 - \theta_1 \,| \\
|\, b_{\theta_2}\left(S_s(\theta_2), s\right) - b_{\theta_1}\left(S_s(\theta_1), s\right) | &\leq \beta_1 \,|\, S_s(\theta_2) - S_s(\theta_1) \,| + \beta_0 \,|\, \theta_2 - \theta_1 \,|
\end{aligned} \tag{3.56}$$

and applying the previous theorem, we get for any $t \in [0, T]$,

$$\mathbb{E}\left[\left(\frac{S_t(\theta_2) - S_t(\theta_1)}{\theta_2 - \theta_1}\right)^2\right] \leq 4\left(\alpha_0\sqrt{t} + \beta_0\right)^2 \exp\left(4t\left(\alpha_1\sqrt{t} + \beta_1\right)^2\right) t \tag{3.57}$$

Using Jensen's inequality, we can therefore write that for any t fixed and for all $\theta_1, \theta_2 \in \Theta$,

$$\mathbb{E}\left(\left|\frac{\left(S_t\left(\theta_2\right) - S_t\left(\theta_1\right)\right)}{\left(\theta_2 - \theta_1\right)}\right|\right) \leq \sqrt{\mathbb{E}\left(\left(\frac{S_t\left(\theta_2\right) - S_t\left(\theta_1\right)}{\theta_2 - \theta_1}\right)^2\right)} \leq \widetilde{K}_t \tag{3.58}$$

where $\widetilde{K}_t$ is a constant.

Using the Cauchy-Schwarz inequality, we can also write that for any set A such that $\mathbb{P}(A) \leq \delta$, then

$$\begin{aligned}
\mathbb{E}\left(\left|\frac{\left(S_t\left(\theta_2\right) - S_t\left(\theta_1\right)\right)}{\left(\theta_2 - \theta_1\right)}\right| \mathbf{1}_A\right) &\leq \sqrt{\mathbb{E}\left(1_A^2\right)}\sqrt{\mathbb{E}\left(\left(\frac{\left(S_t\left(\theta_2\right) - S_t\left(\theta_1\right)\right)}{\left(\theta_2 - \theta_1\right)}\right)^2\right)} \\
&\leq \sqrt{\delta}\widetilde{K}_t
\end{aligned} \tag{3.59}$$

Therefore for any $\epsilon > 0$, picking $\delta_t = \left(\epsilon / \widetilde{K}_t\right)^2$ ensures that for any set A such that $\mathbb{P}(A) \leq \delta_t$,

$$\mathbb{E}\left(\left|\frac{\left(S_t\left(\theta_2\right) - S_t\left(\theta_1\right)\right)}{\left(\theta_2 - \theta_1\right)}\right| \mathbf{1}_A\right) \leq \epsilon \tag{3.60}$$

Together, (3.58) and (3.60) mean the family of functions $\Delta_t\left(\theta_1, \theta_2\right)$ is uniformly integrable and condition $\widetilde{A4}$ is satisfied.

We have proved that we have all the conditions ensuring that pathwise sensitivities are applicable. We can summarise our result as follows:

Theorem 3.2.3. *Let us consider an option with an almost-everywhere differentiable Lipschitz payoff P depending on the values of an underlying asset S_t. Assuming that S_t follows an Ito process as described by equation* (1.2) *on the interval* $[0, T]$*, that the coefficients $a_\theta(S, t)$, $b_\theta(S, t)$ are Lipschitz, have bounded locally Lipschitz first order derivatives and that $b_\theta(S_0, 0) \neq 0$, then pathwise sensitivities can be applied to compute the option's sensitivities.*

Proof. See above. □

3.2.4 Convergence of the payoff estimators' sensitivities

The results of sections 3.2.2 and 3.2.3 guarantee that we can apply pathwise sensitivities to the exact solution S_t of an Ito evolution SDE. Computing the exact solution is possible in the Black & Scholes model (see section 2.1.2) but in most cases it is impractical and we work with discretised solutions resulting from the Milstein discretisation of the original SDE. We now present results guaranteeing that Pathwise sensitivities are still applicable and that the derivatives of payoff estimators actually result in estimators of the Greeks with a vanishing bias.

3.2.4.1 A simple case

We begin with a simple proof for Lipschitz payoffs P whose first order derivative is Lipschitz in the underlying asset's values and sensitivities at a set of given discretisation times (e.g. a smooth European payoff or a discretely sampled lookback option).

From the results of section 3.2.3, we now know that

$$\frac{\partial}{\partial\theta}\mathbb{E}\left(P\left(S_{t_0},\ldots,S_{t_K}\right)\right)=\mathbb{E}\left(\frac{\partial P\left(S_{t_0},\ldots,S_{t_K}\right)}{\partial\theta}\right) \tag{3.61}$$

and then using the convergence properties of the Milstein discretisation presented in section 3.1, we know that for $k=0,\ldots,K$,

$$\begin{aligned}\mathbb{E}\left(\left|\widehat{S}_k-S_{t_k}\right|\right)&=O\left(h\right)\\ \mathbb{E}\left(\left|\frac{\partial\widehat{S}_k}{\partial\theta}-\frac{\partial S_{t_k}}{\partial\theta}\right|\right)&=O\left(h\right)\end{aligned} \tag{3.62}$$

using the fact that the first order derivative of the payoff is Lipschitz in $\widehat{S}_k$ and $\dfrac{\partial\widehat{S}_k}{\partial\theta}$, we then get

$$\mathbb{E}\left(\frac{\partial P\left(\widehat{S}_0,\ldots,\widehat{S}_{t_K}\right)}{\partial\theta}\right)=\mathbb{E}\left(\frac{\partial P\left(S_{t_0},\ldots,S_{t_K}\right)}{\partial\theta}\right)+O\left(h\right) \tag{3.63}$$

Therefore we indeed have

$$\mathbb{E}\left(\frac{\partial P\left(\widehat{S}_0,\ldots,\widehat{S}_{t_K}\right)}{\partial\theta}\right)=\frac{\partial\mathbb{E}\left(P\left(S_{t_0},\ldots,S_{t_K}\right)\right)}{\partial\theta}+O\left(h\right) \tag{3.64}$$

i.e. the derivative of the payoff's estimator is an estimator that converges towards the exact value of the Greek as the time step h is refined.

3.2.4.2 A more general result

The result of section 3.2.4 does not even apply to simple cases like that of the European call (its derivative is not Lipschitz in S_T). We here present a more general set of conditions ensuring that the payoff estimator's derivative results in an unbiased estimator of the sensitivity.

Lemma 3.2.4. *If the following three conditions are satisfied:*

- *A1: There exist constants $c, \alpha > 0$ such that for all θ in some interval Θ,*

$$\left| \mathbb{E}\left[\widehat{P}_l - P \right] \right| < c\, 2^{-\alpha l}, \tag{3.65}$$

  and

$$\left| \mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta} \right] \right| < c\, 2^{-\alpha l}. \tag{3.66}$$

- *A2:* $$\mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta} \right] \text{ is continuous in } \theta.$$

- *A3:*

$$\frac{\partial}{\partial \theta}\, \mathbb{E}\left[\widehat{P}_l \right] = \mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta} \right] \tag{3.67}$$

then $\mathbb{E}[P]$ is differentiable for all θ in Θ, and there is a second constant c_2 such that

$$\left| \mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta} \right] - \frac{\partial}{\partial \theta}\, \mathbb{E}[P] \right| < c_2\, 2^{-\alpha l}.$$

Proof. Due to (3.66), the sequence $\mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta} \right]$ is a Cauchy sequence and so converges pointwise to some function $Q(\theta)$ as $l \to \infty$.

Furthermore, due to the uniform bound in (3.66), plus condition A2, the uniform convergence theorem proves that $Q(\theta)$ is continuous.

If θ_1, θ_2 lie within the interval Θ, then integrating (3.67) gives

$$\mathbb{E}\left[\widehat{P}_l(\theta_2) \right] - \mathbb{E}\left[\widehat{P}_l(\theta_1) \right] = \int_{\theta_1}^{\theta_2} \mathbb{E}\left[\frac{\partial \widehat{P}_\ell}{\partial \theta} \right] \mathrm{d}\theta.$$

Taking the limit as $l \to \infty$, the dominated convergence theorem gives

$$\mathbb{E}[P(\theta_2)] - \mathbb{E}[P(\theta_1)] = \int_{\theta_1}^{\theta_2} Q(\theta)\, \mathrm{d}\theta.$$

and hence, by the first fundamental theorem of calculus since $Q(\theta)$ is continuous, $\mathbb{E}[P(\theta)]$ is differentiable and its derivative is $Q(\theta)$, that is,

$$\frac{\partial \mathbb{E}\left(P(\theta)\right)}{\partial \theta} = Q(\theta) = \lim_{l \to \infty} \mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta} \right]$$

Finally, defining

$$c_2 \;=\; c\sum_{l=1}^{\infty} 2^{-\alpha l} \;=\; \frac{c\,2^{-\alpha}}{1-2^{-\alpha}},$$

summing (3.66) over l we obtain

$$\left| \mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta}\right] - Q \right| < c_2\, 2^{-\alpha l},$$

which completes the proof. □

Now let us prove that the conditions of theorem 3.2.4 are satisfied in the cases we consider.

The first part of condition $A1$ corresponds to the weak convergence properties of the option value's estimators. The estimators we consider for the European call, the digital call, the lookback option, the Asian option and the barrier option are those presented in [Gil08a] where their experimental weak convergence properties are also established.

In the analysis of the various options (chapters 4 to 8), we establish the convergence properties of $\mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right]$ or $\mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial P}{\partial \theta}\right]$, i.e. we obtain the convergence properties of the sensitivities of the estimators as required by the second part of condition A1.

To check condition $A2$, i.e. that for any fixed l, $\mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta}\right](\theta)$ is continuous on Θ, we use Vitali's convergence theorem: if $\frac{\partial \widehat{P}_l}{\partial \theta}(\theta)$ is almost surely continuous on a neighborhood $N(\Theta)$ of the interval Θ and if the family of functions $\left(\frac{\partial \widehat{P}_l}{\partial \theta}\right)_{\theta\in N(\Theta)}$ is uniformly integrable, then $\mathbb{E}\left[\frac{\partial \widehat{P}_l}{\partial \theta}\right](\theta)$ is indeed continuous on Θ.

We first check that $\frac{\partial \widehat{P}_l}{\partial \theta}$ is continuous in θ: indeed, we can show by iteration that under the usual regularity assumptions on a, b, the discretised values $\widehat{S}_0, \widehat{S}_1, \ldots, \widehat{S}_{N_f(l)}$ and their sensitivities $\frac{\partial \widehat{S}_0}{\partial \theta}, \ldots, \frac{\partial \widehat{S}_{N_f(l)}}{\partial \theta}$ are almost surely continuous in θ. Then, as will be seen in the numerical analysis of chapters 4 to 8, for any given level l, the derivatives of the payoff estimators $\frac{\partial \widehat{P}_l}{\partial \theta}$ are almost surely continuous functions of those values. Therefore $\frac{\partial \widehat{P}_l}{\partial \theta}$ is indeed continuous in θ almost surely.

Then, as in equations (3.58) and (3.60), we see that to prove the uniform integrability of $\left(\frac{\partial \widehat{P}_l}{\partial \theta}\right)_{\theta\in\Theta}$, it is enough to prove that there exists a uniform bound K: for any $\theta \in \Theta$, $\mathbb{E}\left[{\frac{\partial \widehat{P}_l}{\partial \theta}}^2(\theta)\right] < K$. Provided Θ is bounded (which is a reasonable hypothesis for all practical purposes), this will derive from the analysis in chapters 4 to 8.

Condition $A3$ means that at each level l, the interchange between expectation and differentiation is valid, which we can prove using lemma 3.2.1. Indeed, under the same conditions on a, b as before, we can prove by iteration that the simulated values $\widehat{S}_0, \ldots, \widehat{S}_{N_f(l)}$ are almost surely differentiable in θ. At a given level l, the payoff estimators described in chapters 2 and 4 to 8 are Lipschitz functions of those values. In the same way we can also show that for $k = 0, \ldots, N_f(l)$ and for any $(\theta_1, \theta_2) \in \Theta^2$, we can write $\left| \widehat{S}_k(\theta_2) - \widehat{S}_k(\theta_1) \right| \leq K_k \, | \, \theta_2 - \theta_1 \, |$ where K_k is a family of random variables with finite expectations. All the conditions of lemma 3.2.1 are verified and therefore condition $A3$ holds.

Conditions $A1, A2, A3$ of lemma 3.2.4 are verified in the cases we study, therefore the expectation of the derivatives of the payoffs' estimators do converge towards the Greeks. This proves the validity of our approach.

3.3 Assumptions on the volatility

The numerical analysis we perform relies on a few assumptions. One of the most important ones is that the probability density function of the underlying price is regular enough at various times in the inverval $[0, T]$.

Typically we want the density function to be bounded at expiry for European options and also at the various discretisation times $t_1 = T/N, t_2 = 2\,T/N, \ldots, t_N = T$. This hypothesis is crucial as it enables us to link in a simple way the likelihood of a path S_t being at time t_k in a given subinterval of a compact set I and the width of these intervals. Indeed, if we assume that $p\,(S)$, the probability density function of S_{t_k} is continuous on $\mathbb{R}$, then using Heine's theorem (see [Zor04]), it is also uniformly continuous and bounded on the compact set I. We write $p_I^{\max} = \max_{S \in I} p\,(S)$ and then we can write for any subinterval $[a, b] \subset I$,

$$\begin{aligned} \mathbb{P}\,(S_{t_k} \in [a, b]) &= \int_a^b p\,(S_{t_k}) \mathrm{d}S_{t_k} \\ &\leq p_I^{\max} \, | \, b - a \, | \end{aligned} \tag{3.68}$$

We present here a few conditions ensuring that the probability density function for the underlying asset is regular enough.

3.3.1 Black and Scholes density

In the Black & Scholes setting, we recall equation (2.4) and for any times $s < t$, with $s, t \in [0, T]$, we can write

$$p\left(S_t | S_s\right) = \frac{1}{\sigma\sqrt{2\pi(t-s)}} \exp\left(-\frac{\left(\log\left(S_t/S_s\right) - \left(r - \frac{\sigma^2}{2}\right)(t-s)\right)^2}{2\sigma^2(t-s)}\right)$$

which is clearly a smooth function of S_t. This smoothness ensures the analysis of chapters 4 to 8 can be applied to the Black & Scholes model used in chapter 2.

3.3.2 Hörmander's condition

We now present conditions ensuring that the general process described by (1.2) does also have regular density functions. For the sake of simplicity, we only consider the process between the times 0 and t, studying the transition density function between some time $s > 0$ and t is in all points similar.

The following theorem is a 1-dimensional version of a theorem first presented in [BH86].

Theorem 3.3.1. *We write*

$$\tau = \min\left(\inf\left\{t > 0 \,\middle|\, \int_0^t \mathbf{1}_{b(S_s, s) \neq 0} \mathrm{d}s > 0\right\}, T\right)$$

For $t \in [0, T]$, we let S_t be the solution of the stochastic differential equation (1.2) *where the coefficients $a(S, t)$ and $b(S, t)$ safisfy the two following conditions*

- *A1: the coefficients are globally Lipschitz, i.e. there exists a constant $K > 0$ such that for all $t, x, y \in \mathbb{R}^+ \times \mathbb{R}^2$,*

$$|\, a(y, t) - a(x, t)\,| + |\, b(y, t) - b(x, t)\,| \leq K\,|\, y - x\,|$$

- *A2: the functions $t \to a(0, t)$ and $t \to b(0, t)$ are bounded on $[0, T]$.*

Then, for any $t \in [0, T]$, the probability density function of $S(t)$ conditional on $\{t > \tau\}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$.

Note that the previous theorem extends easily to the multidimensional case where S_t is an m-dimensional process driven by a d-dimensional Brownian motion. The multidimensional form of the theorem can be found in section 2.3 of [Nua05].

Note also that S_t is a sample-continuous process (noting a and b are globally Lipschitz and using theorem 3.1.1). If $b(S_0, 0) > 0$, using the sample-continuity of $b(S_t, t)$, for any $t_k > 0$, there is almost surely a non-degenerate time interval $[0, \epsilon]$ (for some $\epsilon \in [0, t_k]$) on which $b(S_t, t) > b(S_0, 0)/2 > 0$ and then $\tau < \epsilon < t_k$, which then means that almost surely, S_{t_k} has an absolutely continuous density functions for all $t_k > 0$.

We now present an alternative condition, which is similar to the intuitive view of theorem 3.3.1 and corresponds to a 1-dimensional version of Hörmander's conditions, as presented in section 2.3 of [Nua05].

Theorem 3.3.2. *We here assume the coefficients $a(S,t)$ and $b(S,t)$ of* (1.2) *do not depend on time, that is,*

$$\mathrm{d}S_t = a(S_t)\,\mathrm{d}t + b(S_t)\,\mathrm{d}W_t$$

If we also assume they are infinitely differentiable with bounded derivatives of all orders and $b(S_0) \neq 0$, then for any $t > 0$, the solution S_t has a probability density function that is absolutely continuous with respect to the Lebesgue measure and is also infinitely differentiable.

The conditions of theorem 3.3.2 are slightly more restrictive than the ones presented in theorem 3.3.1, yet removing the possibility of a time dependency for $a\,(S_t, t)$ and $b\,(S_t, t)$ is not an issue for models like Black & Scholes or Vasicek, in which case this theorem's simplicity is very striking.

3.3.3 A convenient hypothesis

In section 3.3.2 we presented conditions ensuring the solution S_t has sufficiently regular distributions at various times. These conditions can be easily verified and are fairly unrestrictive, yet we will voluntarily consider a slightly more restrictive setting that makes the analysis "cleaner" and lets us focus on essential ideas instead of having to pay too much attention to the detail of particular "ill-behaved" cases.

Several techniques presented in the simulations of chapter 2 rely on the use of the diffusive properties of SDE (1.2). The "conditional expectations" smoothing technique as presented in sections 2.2.2, 2.3.1 and 2.6.3, the vibrato Monte Carlo of sections 2.2.4 and 2.3.2 and finally the treatment of the discontinuity at the barrier in section 2.6.2 all require some diffusion to happen at various discretisation times of $[0, T]$. As is evident in formulas (2.49), (2.60), (2.79), (2.120), 2.122 or (2.124), these methods work "as is" only if the volatility $b\,(S, t)$ is non-zero.

The conditions of theorem 3.3.2 impose that the initial value of the volatility should be strictly positive but they do not exclude situations where the volatility vanishes at some point $t_{k_{vanish}}$ of the discretisation (i.e. $b(\widehat{S}_{k_{vanish}}, t_{k_{vanish}}) = 0$), in which case the formulas mentioned above are not well defined. We can deal with this situation in various ways: in the Black & Scholes model, we notice that if this ever happens, that is if $\sigma \widehat{S}_{k_{vanish}} = 0$, then the discretised solution of the SDE is $\widehat{S}_k = 0$ for all $t_k \geq t_{vanish}$ and conditional expecations are irrelevant. In the general case, if we have $b(\widehat{S}_{k_{vanish}}, t_{k_{vanish}}) = 0$ then the evolution of the discretisation on

$[t_{k_{vanish}}, t_{k_{vanish}+1}]$ is deterministic and there is no use for conditional expectation-related techniques either.

The analysis of the convergence rates of multilevel Monte Carlo techniques involves the quantification of all terms appearing in the aforementioned equations. If we let the volatility be arbitrarily low, then the sensitivities of some of these terms blow up and make the analysis difficult, even though extremely low volatility situations also mean there is little use for smoothing using conditional expectations (the diffusive action is very low). This sort of situation is not a practical issue: in the Black & Scholes setting, low volatility only occurs when $S \approx 0$, that is, far from the discontinuities/kinks in the payoffs where smoothing is needed (around the strike K or the barrier B). Dealing with these ill-behaved cases specifically (e.g. using different estimators depending on the volatility level) would provide very little in terms of benefits and make the analysis unnecessarily intricate.

Therefore for techniques using some form of conditional expectation smoothing, we restrict the analysis to the case of elliptic SDEs, i.e. we make the assumption that the volatility has a lower bound $\epsilon > 0$.

$$\exists \epsilon > 0, \text{ s.t. } \forall (S,t) \in \left(\mathbb{R} \times \mathbb{R}^+\right), b(S,t) > \epsilon \tag{3.69}$$

As explained in [Avi09b] and [Fri64], this ellipticity condition, together with the assumption that the coefficients $a(S,t)$ and $b(S,t)$ are infinitely differentiable with bounded derivatives, also guarantees that the probability distribution function of the solution is bounded (see [CFN98] for slightly less restrictive conditions giving the same result).

Summing up previous remarks, these conditions ensure that the probability densities of the solution S_t at different times $t \in [0,T]$ are smooth (as stated specifically in [DM10] for example) and bounded, that all the aforementioned techniques based on conditional expectations are relevant and that the corresponding formulas are well-defined.

3.4 Essential theorems

In this section, we present a collection of useful lemmas and theorems inspired by the analysis found in [GDR13].

3.4.1 Moments of an SDE's solution

We begin by presenting important results on the moments of the solution of SDE (1.2). They are found on p136 in [KP92].

Theorem 3.4.1. *Let X_t be a scalar process satisfying the following SDE on $[0,T]$,*

$$\begin{aligned}&X(0) \text{ given}\\ &\mathrm{d}X(t) = a(X,t)\,\mathrm{d}t + b(X,t)\,\mathrm{d}W_t\end{aligned} \tag{3.70}$$

Assuming that

- *A1: $a(X,t)$ and $b(X,t)$ are jointly $\mathcal{L}\times\mathcal{L}$ measurable in $\mathbb{R}\times[0,T]$*
- *A2: there exists a constant $C_1 > 0$ such that for all $t,x,y \in \mathbb{R}^+\times\mathbb{R}^2$,* $|a(y,t)-a(x,t)| + |b(y,t)-b(x,t)| \le C_1\,|y-x|$
- *A3: there exists a constant C_2 such that for all $x\in\mathbb{R}$,* $|a(x,t)|^2 + |b(x,t)|^2 \le C_2^2\left(1+|x|^2\right)$

then the SDE (3.70) *has a pathwise unique strong solution X_t on $[0,T]$ with*

$$\sup_{0\le t\le T}\mathbb{E}\left(|X_t|^2\right) < \infty \tag{3.71}$$

$$\mathbb{E}\left(|X_t|^{2n}\right) \le \left(1+|X_0|^{2n}\right)\exp\left(Ct\right) \tag{3.72}$$

$$\mathbb{E}\left(|X_t - X_0|^{2n}\right) \le \widetilde{C}\left(1+|X_0|^{2n}\right)t^n\exp\left(Ct\right) \tag{3.73}$$

$$\mathbb{E}\left(\sup_{0\le t\le T}|X_t|^{2n}\right) \le \widetilde{C}\,\left(|X_0|^{2n} + \left(1+|X_0|^{2n}\right)T^n\exp\left(CT\right)\right) \tag{3.74}$$

$$\mathbb{E}\left(\sup_{0\le t\le T}|X_t - X_0|^{2n}\right) \le \widetilde{C}\,\left(1+|X_0|^{2n}\right)T^n\exp\left(CT\right) \tag{3.75}$$

for $t\in[0,T]$ where $T<\infty$ and $C = 2n\,(2n+1)\max\left(C_1,C_2\right)^2$ and $\widetilde{C}$ is a positive constant depending only on $n, \max\left(C_1,C_2\right)$ and T.

Note that in practice, other important theorems like theorem 3.4.3 add Lipschitz-like conditions for the evolution of the coefficients a and b on $[0,T]$ that make condition $A3$ a direct consequence of condition $A2$.

Also note that the result can be extended to the case where X_0 is a random variable such that $\mathbb{E}\left[|\,X_0\,|^{2n}\right] < \infty$. The inequalities of theorem 3.4.1 still hold if we replace $|\,X_0\,|^{2n}$ by $\mathbb{E}\left[|\,X_0\,|^{2n}\right]$ (see p136 in [KP92]).

Corollary 3.4.2. *Theorem 3.4.1 carries over verbatim to the multidimensional case where $X_t = (X_1(t),\ldots,X_d(t))\in\mathbb{R}^d$ and W_t is still a 1-dimensional Brownian motion, provided the absolute values are replaced by vector and matrix norms such as the Euclidean norms.*

Proof. See [KP92]. □

3.4.2 Milstein scheme and interpolants

We here present different continuous interpolants of the Milstein discretisation of an Ito process. The first one, introduced in [KP92] has interesting convergence properties. The second one, extensively used in [GDR13], is based on a simple Brownian Bridge and is therefore particularly simple and convenient for certain applications.

Theorem 3.4.3. *We consider the case where $X_t \in \mathbb{R}^d$ and W_t is a 1-dimensional Brownian motion. $(a_k)_{(k=1..d)}$ and $(b_k)_{(k=1..d)}$ are the different components of the coefficients a and b in the same vector SDE as before: on $[0,T]$,*

$$\begin{aligned} &X(0) \text{ given} \\ &\mathrm{d}X(t) = a(X,t)\,\mathrm{d}t + b(X,t)\,\mathrm{d}W_t \end{aligned} \tag{3.76}$$

We let

$$L^0 = \frac{\partial}{\partial t} + \sum_{k=1}^{d} a_k \frac{\partial}{\partial X_k} + \frac{1}{2}\sum_{k,l=1}^{d} b_k\, b_l \frac{\partial^2}{\partial X_k \partial X_l}$$

$$L^1 = \sum_{k=1}^{d} b_k \frac{\partial}{\partial X_k}$$

Assuming that

- *A1: $a(x,t)$ is $\mathcal{C}^{(1,1)}\left(\mathbb{R}^d \times \mathbb{R}^+\right)$ and $b(x,t)$ is $\mathcal{C}^{(2,1)}\left(\mathbb{R}^d \times \mathbb{R}^+\right)$*
- *A2: (uniform Lipschitz condition): there exists a constant $C_1 > 0$ such that for all $x, y \in \mathbb{R}^d$,*
  $\|a(y,t) - a(x,t)\| + \|b(y,t) - b(x,t)\| + \left\|L^1 b(y,t) - L^1 b(x,t)\right\| \leq C_1 \|y - x\|$
- *A3: (linear growth bound): There exists a constant C_2 such that for all $x \in \mathbb{R}^d$,*
  $\|a(x,t)\| + \|L_0 a(x,t)\| + \|L_1 a(x,t)\| + \|b(x,t)\| + \|L_0 b(x,t)\| + \|L_1 b(x,t)\| + \|L_0 L_1 b(x,t)\| + \|L_1 L_1 b(x,t)\| \leq C_2\,(1 + |\,x\,|)$
- *A4: (additional Lipschitz-like condition): there exists a constant C_3 such that for all $x \in \mathbb{R}^d$ and $s,t \in \mathbb{R}^+$, $\|b(x,t) - b(x,s)\| \leq C_3\,(1 + |\,x\,|)\sqrt{|\,t - s\,|}$*

Then for each $m > 0$,

$$\mathbb{E}\left(\sup_{0<t<T} \|X_t\|^m\right) < \infty \tag{3.77}$$

Using the Milstein discretisation $\left(\widehat{X}_n\right)_{n=0,\dots,N}$, we define the following continuous time interpolant on each time step $[t_n, t_{n+1}]$ $(n = 0, \dots, N-1)$ and each dimension $k = 1, \dots, d$ as

$$\begin{aligned} \widehat{X}_{KP_k}(t) =& \widehat{X}_{n_k} + a_k\left(\widehat{X}_n, t_n\right)(t - t_n) + b_k\left(\widehat{X}_n, t_n\right)(W_t - W_n) \\ &+ \frac{1}{2}\left(\sum_{l=1}^{k} b_l \frac{\partial b_k}{\partial X_l}\left(\widehat{X}_n, t_n\right)\right)\left((W_t - W_n)^2 - (t - t_n)\right) \end{aligned} \tag{3.78}$$

then for each $m > 0$, there exists a constant C_m such that

$$\mathbb{E}\left(\sup_{0<t<T}\left\|X_t - \widehat{X}_{KP}(t)\right\|^m\right) < C_m\, h^m \tag{3.79}$$

and

$$\mathbb{E}\left(\sup_{0<t<T}\left\|\widehat{X}_{KP}(t)\right\|^m\right) < C_m \tag{3.80}$$

Proof. See theorem 10.6.3 and corollary 10.6.4 in [KP92]. □

Considering the joint evolution equation (3.16) for the asset's price and its sensitivity, we see that this theorem can be applied when the coefficients $a\,(S,t)$, $b\,(S,t)$, $\dfrac{\partial a}{\partial \theta}\,(S,t)$, $\dfrac{\partial b}{\partial \theta}\,(S,t)$ are linear in S and do not depend on t, which corresponds for example to the case of the Black-Scholes model.

Lemma 3.4.4. *Let X_t be the solution of equation* (3.76)*. Given its Milstein discretisation $\left(\widehat{X}_n\right)_{n=0,\dots,N}$, we define on each time step $[t_n, t_{n+1}]$ the continuous extension $\widehat{X}\,(t)$ using the Brownian bridge interpolation*

$$\widehat{X}_t = \widehat{X}_n + \frac{t-t_n}{h}\left(\widehat{X}_{n+1} - \widehat{X}_n\right) + b_n\left(W_t - W_{nh} - \frac{t-t_n}{h}\Delta W_n\right) \tag{3.81}$$

Compared to the Kloeden-Platen interpolant $\widehat{X}_{KP}\,(t)$ defined in theorem 3.4.3, its analytical properties (distribution of the local minima, conditional probability of hitting a barrier) are easier to derive. For example, in the 1-dimensional case we have

- $\displaystyle\int_{t_n}^{t_{n+1}} \widehat{X}_t \mathrm{d}t = \frac{h}{2}\left(\widehat{X}_n + \widehat{X}_{n+1}\right) + b_n I_n$
  where $I_n \sim \mathcal{N}\left(0, \dfrac{h^3}{12}\right)$ is independent of $W_{n+1} - W_n$.
- *Conditional on $\widehat{X}_n, \widehat{X}_{n+1}$, the minimum of $\widehat{X}_t$ on $[t_n, t_{n+1}]$ is given by* $\widehat{X}_{n,\min} = \dfrac{1}{2}\left(\widehat{X}_n + \widehat{X}_{n+1} - \sqrt{\left(\widehat{X}_n - \widehat{X}_{n+1}\right)^2 - 2b_n^2 h \log U_n}\right)$ *where $U_n \sim U\,(0,1)$.*
- *Conditional on $\widehat{X}_n, \widehat{X}_{n+1}$, the probability that the minimum of $\widehat{X}_t$ on $[t_n, t_{n+1}]$ is less than a certain value B is*
  $\mathbb{P}\left(\inf_{t\in[t_n,t_{n+1}]} \widehat{X}_t < B \,\middle|\, \widehat{X}_n, \widehat{X}_{n+1}\right) = \exp\left(\dfrac{-2\left(\widehat{X}_n - B\right)^+\left(\widehat{X}_{n+1} - B\right)^+}{b_n^2 h}\right)$

Under the assumptions of theorem 3.4.3, the relative accuracy of the two estimators is given by

$$\mathbb{E}\left(\sup_{t\in[0,T]}\left\|\widehat{X}\,(t) - \widehat{X}_{KP}\,(t)\right\|^m\right) = O\left((h\log h)^m\right) \tag{3.82}$$

$$\sup_{t\in[0,T]} \mathbb{E}\left(\left\|\widehat{X}\,(t) - \widehat{X}_{KP}\,(t)\right\|^m\right) = O\left(h^m\right) \tag{3.83}$$

$$\mathbb{E}\left(\left\|\int_0^T\left(\widehat{X}(t)-\widehat{X}_{KP}(t)\right)\mathrm{d}t\right\|^2\right)=O\left(h^3\right) \tag{3.84}$$

where we used the norm $\left\|\widehat{X}\left(t\right)-\widehat{X}_{KP}\left(t\right)\right\|=\max_{i=1,\ldots,d}\left|\widehat{X}_k\left(t\right)-\widehat{X}_{KP_k}\left(t\right)\right|$.

Proof. See [GDR13]. □

3.4.3 Extreme paths

The so-called "extreme paths" analysis on which [GHM09] relies and which is detailed in [GDR13] consists in separating events ω into two categories, the ones for which the driving Brownian motion is well-behaved (reasonably small increments between consecutive discretisation times), which are most common, and the "extreme/ill-behaved" ones, which may have large increments but are very rare. Once we have established this dichotomy, the idea underlying the analysis is to first analyse the behaviour of various quantities for non-extreme events for which we have known bounds on Brownian increments or the error resulting from the discretisation.

To prove in lemma 3.4.8 that the contribution of rare extreme paths to the global expectation of the quantities considered is negligible, we begin by presenting lemmas 3.4.5, 3.4.6 and 3.4.7 that are introduced in [GDR13].

Lemma 3.4.5. *If X_l is a scalar random variable defined on level l of the multilevel analysis, and for each positive integer m, $\mathbb{E}\left[\mid X_l\mid^m\right]$ is uniformly bounded, then, for any $\delta>0$,*

$$\mathbb{P}\left(\mid X_l\mid>h_l^{-\delta}\right)=O\left(h_l^p\right),\qquad\qquad \forall p>0 \tag{3.85}$$

Proof. It follows immediately from Markov's inequality

$$\mathbb{P}\left(\mid X_l\mid\geq h_l^{-\delta}\right)=\mathbb{P}\left(\mid X_l\mid^m\geq h_l^{-m\delta}\right)\leq h_l^{m\delta}\mathbb{E}\left[\mid X_l\mid^m\right] \tag{3.86}$$

by choosing $m>p/\delta$ □

Lemma 3.4.6. *If Y_l is a scalar random variable on level l, $\mathbb{E}\left[Y_l^2\right]$ is uniformly bounded, and for each $p>0$, the indicator function $\mathbf{1}_{E_l}$ on level l (which takes value 1 or 0 depending whether or not a path lies within some set E_l) satisfies*

$$\mathbb{E}\left[\mathbf{1}_{E_l}\right]=O\left(h_l^p\right) \tag{3.87}$$

then for each $p>0$,

$$\mathbb{E}\left[\mid Y_l\mid\mathbf{1}_{E_l}\right]=O\left(h_l^{p/2}\right) \tag{3.88}$$

Proof. Immediate consequence of Hölder inequality which gives

$$\mathbb{E}\left[\mid Y_l\mid\mathbf{1}_{E_l}\right]\leq\left(\mathbb{E}\left[Y_l^2\right]\right)^{1/2}\left(\mathbb{E}\left[\mathbf{1}_{E_l}\right]\right)^{1/2} \tag{3.89}$$

□

In proofs, lemma 3.4.5 is used to establish the pre-conditions for lemma 3.4.6 from which it can be concluded by choosing p sufficiently large that the contribution of the paths in E_l is negligible.

Lemma 3.4.7. *If W_t is a Brownian motion with $W_0 = W_1 = 0$, then for $x > 0$,*

$$\mathbb{P}\left(\sup_{[0,1]} W_t > x\right) = \mathbb{P}\left(\inf_{[0,1]} W_t < -x\right) = \exp\left(-2x^2\right) \tag{3.90}$$

and therefore $\mathbb{E}\left[\sup_{[0,1]} |\, W_t\,|^m\right]$ is finite for all integers $m > 0$.

Proof. Equation (3.90) is a corollary of equation (2.116). □

Using this last lemma together with theorem 3.4.3 and lemmas 3.4.4, 3.4.5, 3.4.6, we obtain the following important lemma

Lemma 3.4.8. *Let $\gamma > 0$. We consider a Brownian motion $W(t)$ and $X(t)$, the corresponding solution of SDE (3.76) on $[0, T]$. With the usual conditions of theorem 3.4.3, the probability that the increments $\Delta W_n = W((n+1)h) - W(nh)$, the fine Milstein discretisation $\widehat{X}^f_n$ (step h) and coarse and $\widehat{X}^c_n$ (step $2h$ and midpoints constructed with Brownian bridge) satisfy any of the following extreme conditions*

$$\max_{n=0,\ldots,N}\left(\max\left(\|X(nh)\|, \left\|\widehat{X}^f_n\right\|, \left\|\widehat{X}^c_n\right\|\right)\right) > h^{-\gamma}$$

$$\max_{n=0,\ldots,N}\left(\max\left(\left\|X(nh) - \widehat{X}^c_n\right\|, \left\|X(nh) - \widehat{X}^f_n\right\|, \left\|\widehat{X}^f_n - \widehat{X}^c_n\right\|\right)\right) > h^{1-\gamma}$$

$$\max_{n=0,\ldots,N-1}\left(|\,\Delta W_n\,|\right) > h^{1/2-\gamma}$$

is $o\left(h^p\right)$, for all $p > 0$. Furthermore there exist constants c_1, c_2, c_3, c_4 such that if none of these conditions is satisfied and $\gamma < 1/2$, then

$$\max_{n=1,\ldots,N}\left\|\widehat{X}^f_n - \widehat{X}^f_{n-1}\right\| \le c_1\, h^{1/2-2\gamma}$$

$$\max_{n=1,\ldots,N}\left\|b^f_n - b^f_{n-1}\right\| \le c_2\, h^{1/2-2\gamma}$$

$$\max_{n=0,\ldots,N}\left(\max\left(\left\|b^f_n\right\|, \left\|b^c_n\right\|\right)\right) \le c_3\, h^{-\gamma}$$

$$\max_{n=0,\ldots,N}\left\|b^f_n - b^c_n\right\| \le c_4\, h^{1/2-2\gamma}$$

where b^c_n is defined as b^c_{n-1} if n is odd.

Proof. See [GDR13]. □

3.5 Notes on $O(\ldots)$ notation

For the sake of convenience, we use the following abusive notations in chapters 4 to 8. They correspond to operations on the $o(\ldots)$ and $O(\ldots)$ notations and enable us to describe more clearly the contributions of the different terms when analysing the limiting behaviours of functions in the neighborhood of 0.

Let us consider the functions $f_1, f_2, g_1, g_2 : [0,T] \to \mathbb{R}$, $k \in \mathbb{R}$ and $p \in \mathbb{R}^+$ and assume

$$\begin{aligned} f_1 &= O(g_1) \\ f_2 &= O(g_2) \end{aligned} \tag{3.91}$$

then we write

$$\begin{aligned} k\, f_1 &= k\, O(g_1) = O(g_1) \\ f_1 + f_2 &= O(g_1) + O(g_2) = O(g_1 + g_2) \\ f_1\, f_2 &= O(g_1)\ O(g_2) = O(g_1\ g_2) \\ f_1^p &= O(f_1)^p = O(f_1^p) \end{aligned} \tag{3.92}$$

If we consider $f_3, f_4, g_3, g_4 : [0,T] \to \mathbb{R}$ and assume

$$\begin{aligned} f_3 &= o(g_3) \\ f_4 &= o(g_4) \end{aligned} \tag{3.93}$$

we then write

$$\begin{aligned} k\, f_3 &= k\ o(g_3) = o(g_3) \\ f_3 + f_4 &= o(g_3) + o(g_4) = o(g_3 + g_4) \\ f_3\, f_4 &= o(g_3)\ o(g_4) = o(g_3\ g_4) \\ f_3^p &= o(f_3)^p = o(f_3^p) \end{aligned} \tag{3.94}$$

and also

$$f_1\, f_3 = O(g_1)\, o(g_3) = o(g_1\ g_3) \tag{3.95}$$

In parts of the analysis, we also use the abusive notations $f_1 \leq O(g_1)$ and $f_3 \leq o(g_3)$, for $f_1 = O(g_1)$ and $f_3 = o(g_3)$ when we want to highlight that these result from an inequality. We can actually see it as a shortcut for $f_1 \leq \widetilde{f_1}$ and $f_3 \leq \widetilde{f_3}$ for some quantities $\widetilde{f_1} = O(g_1)$ and $\widetilde{f_3} = o(g_3)$.

Chapter 4

Analysis of Vanilla European options via pathwise sensitivities

We analyse the efficiency of the multilevel Monte Carlo technique for the computation of Greeks of a category of simple options: vanilla European options with almost everywhere differentiable Lipschitz payoffs. To do this, we use the results of chapter 3 to obtain analytical bounds on the coefficients α and β in theorem 1.2.1. We begin with the easiest case, that of smooth Lipschitz payoffs before considering more realistic payoffs whose first order derivative may be discontinuous (e.g. European call).

We always assume that the regularity and growth conditions found in chapter 3 for the coefficients of SDE 1.2 are satisfied, thereby ensuring we can use the results of the corresponding theorems.

4.1 Smooth Lipschitz payoffs

We first consider a European option with a differentiable Lipschitz payoff with a Lipschitz first derivative, i.e. the payoff is of the form $P(S_T)$, where P is a differentiable L_1-Lipschitz function of S_T. $\dfrac{\partial P}{\partial S}$ is also assumed to be L_2-Lipschitz. To simplify the notations, we define $\widetilde{L} = \max(L_1, L_2)$ so that both P and $\dfrac{\partial P}{\partial S}$ are $\widetilde{L}$-Lipschitz.

At level l, with a time step $h_l = T/N_l$, the payoff estimator is $\widehat{P}_l = P(\widehat{S}_{N_l})$ and the multilevel estimator of the Greek as presented in equation (1.35) is written

$$\begin{aligned} \widehat{Y}_l &= M_l^{-1} \sum_{i=1}^{M_l} \left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta} \right)^{(i)} \\ &= M_l^{-1} \sum_{i=1}^{M_l} \left(\frac{\partial \widehat{P}_l}{\partial \widehat{S}_{N_l}} \frac{\partial \widehat{S}_{N_l}}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \widehat{S}_{N_{l-1}}} \frac{\partial \widehat{S}_{N_{l-1}}}{\partial \theta} \right)^{(i)} \end{aligned} \tag{4.1}$$

As explained in section 2.1.4, to determine the efficiency of the multilevel approach, we determine the values of α and β in theorem 1.2.1.

4.1.1 Order of convergence β

To determine β, we analyse the convergence speed of $\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right)$. We can write

$$
\begin{aligned}
\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right) &= \mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial P}{\partial \theta}\right) + \mathbb{V}\left(\frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right) \\
&\quad + 2\,\mathrm{Cov}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial P}{\partial \theta}, \frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right) \\
&\leq \mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial P}{\partial \theta}\right) + \mathbb{V}\left(\frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right) \\
&\quad + 2\sqrt{\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial P}{\partial \theta}\right)\mathbb{V}\left(\frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right)}
\end{aligned} \tag{4.2}
$$

Then, noting that if A and B are two random variables we have

$$
0 \leq \left(\sqrt{\mathbb{V}(A)} - \sqrt{\mathbb{V}(B)}\right)^2 = \mathbb{V}(A) + \mathbb{V}(B) - 2\sqrt{\mathbb{V}(A)\,\mathbb{V}(B)} \tag{4.3}
$$

We get

$$
\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right) \leq 2\left(\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial P}{\partial \theta}\right) + \mathbb{V}\left(\frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right)\right) \tag{4.4}
$$

If we can show that $\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial P}{\partial \theta}\right) \leq C\,h_l^\beta$ for some $C > 0$, then we have

$$
\begin{aligned}
\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right) &\leq 2\,C\,h_l^\beta + 2\,C\,h_{l-1}^\beta \\
&\leq 2\,C\left(1 + \left(\frac{h_{l-1}}{h_l}\right)^\beta\right) h_l^\beta \\
&\leq \widetilde{C}\,h^\beta
\end{aligned} \tag{4.5}
$$

where $\widetilde{C}$ is a positive constant. We therefore study the order of convergence of $\mathbb{V}\left(\frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}_l}{\partial \theta}\right)$ which is the same as that of $\mathbb{V}\left(\widehat{Y}_l\right)$.

To simplify the notation we now write $\widehat{P} = P\left(\widehat{S}_{N_l}\right)$ the payoff resulting from a discretisation with time step h (instead of $\widehat{P}_l$ and h_l respectively), S and $\widehat{S}$ the values of the underlying and its discretised path at T (instead of S_T and $\widehat{S}_{N_l}$ respectively).

We have:

$$\mathbb{V}\left(\frac{\partial P}{\partial \theta}-\frac{\partial \widehat{P}}{\partial \theta}\right) \leq \mathbb{E}\left(\left(\frac{\partial P}{\partial \theta}-\frac{\partial \widehat{P}}{\partial \theta}\right)^{2}\right) \tag{4.6}$$

Then we write

$$\left(\frac{\partial P}{\partial \theta}-\frac{\partial \widehat{P}}{\partial \theta}\right)=\frac{\partial S}{\partial \theta}\left(\frac{\partial P}{\partial S}-\frac{\partial \widehat{P}}{\partial S}\right)+\frac{\partial \widehat{P}}{\partial S}\left(\frac{\partial S}{\partial \theta}-\frac{\partial \widehat{S}}{\partial \theta}\right) \tag{4.7}$$

We let $\mathbb{A}=\dfrac{\partial S}{\partial \theta}\left(\dfrac{\partial P}{\partial S}-\dfrac{\partial \widehat{P}}{\partial S}\right)$, $\mathbb{B}=\dfrac{\partial \widehat{P}}{\partial S}\left(\dfrac{\partial S}{\partial \theta}-\dfrac{\partial \widehat{S}}{\partial \theta}\right)$.

$$\begin{aligned}\mathbb{E}\left(\left(\frac{\partial P}{\partial \theta}-\frac{\partial \widehat{P}}{\partial \theta}\right)^{2}\right) &= \mathbb{E}\left(\mathbb{A}^{2}\right)+\mathbb{E}\left(\mathbb{B}^{2}\right)+2\,\mathbb{E}\left(\mathbb{A}\mathbb{B}\right)\\ &\leq 2\left(\mathbb{E}\left(\mathbb{A}^{2}\right)+\mathbb{E}\left(\mathbb{B}^{2}\right)\right)\end{aligned} \tag{4.8}$$

We can restrict our analysis to that of $\mathbb{E}\left(\mathbb{A}^{2}\right)$ and $\mathbb{E}\left(\mathbb{B}^{2}\right)$.

The payoff being $\widetilde{L}$-Lipschitz, we have

$$\mathbb{B}^{2}=\left(\frac{\partial \widehat{P}}{\partial S}\left(\frac{\partial S}{\partial \theta}-\frac{\partial \widehat{S}}{\partial \theta}\right)\right)^{2} \leq \widetilde{L}^{2}\left(\frac{\partial S}{\partial \theta}-\frac{\partial \widehat{S}}{\partial \theta}\right)^{2} \tag{4.9}$$

As explained in section 3.1, we let $\delta_t := \dfrac{\partial S_t}{\partial \theta}$ and consider the 2-dimensional process $U_t=(S_t,\delta_t)$. The evolution SDE for U_t is given by equation (3.16). We recall it is

$$\mathrm{d}U_t=\begin{pmatrix} a(S_t,t)\\ \delta_t\,\dfrac{\partial a(S_t,t)}{\partial S}+\dfrac{\partial a(S_t,t)}{\partial \theta}\end{pmatrix}\mathrm{d}t+\begin{pmatrix} b(S_t,t)\\ \delta_t\,\dfrac{\partial b(S_t,t)}{\partial S}+\dfrac{\partial b(S_t,t)}{\partial \theta}\end{pmatrix}\mathrm{d}W_t$$

whose Milstein discretisation is, as described in section, 3.1.3.4:

$$\left\{\begin{aligned}\widehat{S}_{n+1} &= \widehat{S}_n + a\left(\widehat{S}_n,t_n\right)h + b\left(\widehat{S}_n,t_n\right)\Delta W_n + \frac{1}{2}\,b\,\frac{\partial b}{\partial S}\left(\widehat{S}_n,t_n\right)\left(\Delta W_n^2-h\right)\\ \widehat{\delta}_{n+1} &= \widehat{\delta}_n + \left(\widehat{\delta}_n\frac{\partial a}{\partial S}+\frac{\partial a}{\partial \theta}\right)\left(\widehat{S}_n,t_n\right)h\\ &\quad + \left(\widehat{\delta}_n\frac{\partial b}{\partial S}+\frac{\partial b}{\partial \theta}\right)\left(\widehat{S}_n,t_n\right)\Delta W_n\\ &\quad + \frac{1}{2}\left[\widehat{\delta}_n\left(\frac{\partial b}{\partial S}\right)^2+\widehat{\delta}_n\,b\,\frac{\partial^2 b}{\partial S^2}\right.\\ &\qquad\qquad \left.+\frac{\partial b}{\partial \theta}\frac{\partial b}{\partial S}+b\,\frac{\partial^2 b}{\partial \theta \partial S}\right]\left(\widehat{S}_n,t_n\right)\left(\Delta W_n^2-h\right)\end{aligned}\right.$$

Theorem 3.4.3 guarantees the multidimensional Milstein scheme applied to U's evolution SDE has a strong convergence of order 1, therefore

$$\mathbb{E}\left[\left(\frac{\partial S}{\partial\theta}-\frac{\partial\widehat{S}}{\partial\theta}\right)^2\right]=O\left(h^2\right) \tag{4.10}$$

This means $\mathbb{E}\left(\mathbb{B}^2\right)=O\left(h^2\right)$.

The derivative $\dfrac{\partial P}{\partial S}$ being $\widetilde{L}$-Lipschitz, we have

$$\mathbb{A}^2=\left(\frac{\partial S}{\partial\theta}\left(\frac{\partial P}{\partial S}-\frac{\partial\widehat{P}}{\partial S}\right)\right)^2\leq\left(\frac{\partial S}{\partial\theta}\right)^2\widetilde{L}^2\left(S-\widehat{S}\right)^2 \tag{4.11}$$

Using Hölder's inequality,

$$\mathbb{E}\left(\mathbb{A}^2\right)\leq K_2\sqrt{\mathbb{E}\left(\left(\frac{\partial S}{\partial\theta}\right)^4\right)}\sqrt{\mathbb{E}\left(\left(S-\widehat{S}\right)^4\right)} \tag{4.12}$$

Corollary 3.4.2 guarantees that there is a constant K_1 such that $\mathbb{E}\left(\left(\dfrac{\partial S}{\partial\theta}\right)^4\right)<K_1$. Therefore,

$$\mathbb{E}\left(\mathbb{A}^2\right)=O\left(\sqrt{\mathbb{E}\left(\left(S-\widehat{S}\right)^4\right)}\right) \tag{4.13}$$

Using theorem 3.4.3,

$$\mathbb{E}\left(\mathbb{A}^2\right)=O\left(h^2\right) \tag{4.14}$$

We thus have $\mathbb{E}\left(\mathbb{A}^2\right)=O\left(h^2\right)$ and $\mathbb{E}\left(\mathbb{B}^2\right)=O\left(h^2\right)$ and finally $\beta=2$

4.1.2 Order of convergence α

With the same notations as before, we now analyse the convergence rate of $\mathbb{E}\left(\dfrac{\partial P}{\partial\theta}-\dfrac{\partial\widehat{P}}{\partial\theta}\right)$.

We have just established that $\mathbb{E}\left[\left(\dfrac{\partial P}{\partial\theta}-\dfrac{\partial\widehat{P}}{\partial\theta}\right)^2\right]=O\left(h^2\right)$. Using Hölder's inequality, we obtain

$$\begin{aligned}\left|\mathbb{E}\left[\left(\frac{\partial P}{\partial\theta}-\frac{\partial\widehat{P}}{\partial\theta}\right)1\right]\right|&\leq\sqrt{\mathbb{E}\left[\left(\frac{\partial P}{\partial\theta}-\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right]\mathbb{E}\left[1\right]}\\&=\sqrt{O\left(h^2\right)O\left(1\right)}\\&=O\left(h\right)\end{aligned} \tag{4.15}$$

We thus have $\mathbb{E}\left[\frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}}{\partial \theta}\right] = O(h)$ and finally $\alpha = 1$.

4.2 Non-smooth Lipschitz payoffs

We can relax the hypothesis that the first derivative of the payoff has to be Lipschitz. We extend the previous result to the case where $\left(\frac{\partial P}{\partial S}\right)$ has a finite number I of discontinuities at points $(K_i)_{i=1..I}$ and is $\widetilde{L}$-Lipschitz between those discontinuities. This broader category of payoffs includes the European call option that has been used in the simulations of section 2.2.

The analysis now has to be a bit more complex than in the previous section. While the analysis of 4.1 mostly relied on ideas already found in the analysis of the multilevel Monte Carlo pricing of call options, we now have to introduce ideas that will be used in all of the following chapters. The main idea is to study the contributions of several categories of paths. We distinguish between "extreme" and "non-extreme" paths. We also analyse separately the contributions of the paths that arrive "far" from the discontinuities, for which the situation is essentially similar to the analysis performed in section 4.1.1 and those that arrive "close" to them for which the effect of the discontinuities is significant ("close" being here defined as "within a distance $h^{1-\delta}$ of a discontinuity", for a certain $\delta > 0$ determined later).

More precisely, we first use the results of section 3.4.3 to define the set of extreme paths E as the set of paths satisfying any of the three conditions of lemma 3.4.8 for a certain $\gamma < 1/2$. We then decompose the set of non-extreme paths E^c into D, the non-extreme paths whose final value is "close" to one of the discontinuities and D^c the non-extreme paths whose final value is "far" from all discontinuities $(K_i)_{i=1..I}$. We can therefore write

$$\Omega = E \sqcup E^c = E \sqcup (D \sqcup D^c) \tag{4.16}$$

4.2.1 Order of convergence β

As before we can write:

$$\begin{aligned}
\mathbb{V}\left(\widehat{Y}_l\right) &= O\left(\mathbb{V}\left(\frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}}{\partial \theta}\right)\right) \\
&= O\left(\mathbb{E}\left[\left(\frac{\partial P}{\partial \theta} - \frac{\partial \widehat{P}}{\partial \theta}\right)^2\right]\right) \\
&= O\left(\mathbb{E}\left(\mathbb{A}^2\right) + \mathbb{E}\left(\mathbb{B}^2\right)\right)
\end{aligned} \tag{4.17}$$

and using the partition of paths of equation (4.16),

$$\mathbb{E}\left(\mathbb{A}^2\right) + \mathbb{E}\left(\mathbb{B}^2\right) = \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right) + \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^2\right) + \mathbb{E}\left(\mathbf{1}_{E}\mathbb{A}^2\right) + \mathbb{E}\left(\mathbf{1}_{E}\mathbb{B}^2\right) \tag{4.18}$$

Contribution of extreme paths

We first show that extreme paths have a negligible contribution to the global expectation: using Hölder's inequality, we have

$$
\begin{aligned}
\mathbb{E}\left(\mathbf{1}_E\mathbb{A}^2\right) &\leq \sqrt{\mathbb{E}\left[\mathbf{1}_E^2\right]}\sqrt{\mathbb{E}\left[\frac{\partial S}{\partial\theta}^4\left(\frac{\partial P}{\partial S}-\frac{\partial\widehat{P}}{\partial S}\right)^4\right]} \\
&\leq \sqrt{\mathbb{E}\left[\mathbf{1}_E\right]}\left(\mathbb{E}\left[\frac{\partial S}{\partial\theta}^8\right]\right)^{1/4}\left(\mathbb{E}\left[\left(\frac{\partial P}{\partial S}-\frac{\partial\widehat{P}}{\partial S}\right)^8\right]\right)^{1/4}
\end{aligned}
\tag{4.19}
$$

Lemma 3.4.8 means that $\sqrt{\mathbb{E}\left(\mathbf{1}_E\right)} = o\left(h^p\right)$ for all $p>0$. Corollary 3.4.2 guarantees that $\sqrt{\mathbb{E}\left(\frac{\partial S}{\partial\theta}^8\right)}$ is finite. $P(S)$ being Lipschitz, we also have

$$
\mathbb{E}\left[\left(\frac{\partial P}{\partial S}-\frac{\partial\widehat{P}}{\partial S}\right)^8\right] \leq \mathbb{E}\left[\left(\left|\frac{\partial P}{\partial S}\right|+\left|\frac{\partial\widehat{P}}{\partial S}\right|\right)^8\right] \leq 2^8\widetilde{L}^8 \tag{4.20}
$$

Therefore equation (4.19) means that for all $p>0$, $\mathbb{E}\left(\mathbf{1}_E\mathbb{A}^2\right) = o\left(h^p\right)$.

Similarly, we write

$$
\begin{aligned}
\mathbb{E}\left(\mathbf{1}_E\mathbb{B}^2\right) &\leq \sqrt{\mathbb{E}\left[\mathbf{1}_E^2\right]}\sqrt{\mathbb{E}\left[\frac{\partial\widehat{P}}{\partial S}^4\left(\frac{\partial S}{\partial\theta}-\frac{\partial\widehat{S}}{\partial\theta}\right)^4\right]} \\
&\leq \sqrt{\mathbb{E}\left[\mathbf{1}_E\right]}\left(\mathbb{E}\left[\frac{\partial\widehat{P}}{\partial S}^8\right]\right)^{1/4}\left(\mathbb{E}\left[\left(\frac{\partial S}{\partial\theta}-\frac{\partial\widehat{S}}{\partial\theta}\right)^8\right]\right)^{1/4}
\end{aligned}
\tag{4.21}
$$

As in (4.20), we use $\left|\frac{\partial\widehat{P}}{\partial S}\right| \leq \widetilde{L}$ to show that $\mathbb{E}\left[\frac{\partial\widehat{P}}{\partial S}^8\right]$ is finite. Applying theorem 3.4.3 proves that $\mathbb{E}\left[\left(\frac{\partial S}{\partial\theta}-\frac{\partial\widehat{S}}{\partial\theta}\right)^8\right]$ is bounded. Therefore, equation (4.21) means that for all $p>0$, $\mathbb{E}\left(\mathbf{1}_E\mathbb{B}^2\right) = o\left(h^p\right)$.

Rewriting equation (4.18), we have

$$
\begin{aligned}
\mathbb{E}\left(\mathbb{A}^2\right)+\mathbb{E}\left(\mathbb{B}^2\right) &= \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right)+\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^2\right)+o\left(h^p\right) \\
&= O\left(\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right)+\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^2\right)\right)
\end{aligned}
\tag{4.22}
$$

This means the contribution of extreme paths is negligible and we can focus only on non-extreme paths.

Contribution of discontinuities

We have now established that

$$\begin{aligned}\mathbb{E}\left(\mathbb{A}^2\right)+\mathbb{E}\left(\mathbb{B}^2\right) &= O\left(\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right)+\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^2\right)\right)\\ &= O\Big(\mathbb{E}\left(\mathbf{1}_{D}\mathbb{A}^2\right)+\mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{A}^2\right)\\ &\qquad +\mathbb{E}\left(\mathbf{1}_{D}\mathbb{B}^2\right)+\mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{B}^2\right)\Big)\end{aligned} \tag{4.23}$$

Let $\mathcal{I}$ be a closed non-degenerate interval such that all discontinuities $(K_i)_{i=1,\ldots,I}$ are contained in its interior. The probability density function of S, $p(S)$ is smooth on $\mathcal{I}$, so it is therefore bounded by some constant $M_{\mathcal{I}}$ on this same interval.

Now let us define D as the set of non-extreme paths for which $\min_i |S-K_i| \leq h^{1-\delta}$, i.e. the set of paths arriving "close" to discontinuities. For $h<h_0$ for some h_0, all paths of D arrive in $\mathcal{I}$. In those conditions, $p<M_{\mathcal{I}}$ for all paths in D and the probability of a path being in D can be written as

$$\begin{aligned}\mathbb{P}(D) &= \int_{S\in D} p(S)\,\mathrm{d}S\\ &= \int_{S\,\middle|\,\min\limits_{i=1,\ldots,I}|S-K_i|<h^{1-\delta}} p(S)\,\mathrm{d}S\\ &\leq 2Ih^{1-\delta}M_{\mathcal{I}}\end{aligned} \tag{4.24}$$

therefore D represents a proportion $O\left(h^{1-\delta}\right)$ of all paths. All other paths are in D^c, which represents a proportion $O(1)$ of all paths. Using equation (4.23), we can write

$$\begin{aligned}\mathbb{E}\left(\mathbb{A}^2\right)+\mathbb{E}\left(\mathbb{B}^2\right) = O\Big(&\mathbb{E}\left(\mathbf{1}_{D^c}\right)\left[\max\left(\mathbb{A}^2 1_{D^c}\right)+\max\left(\mathbb{B}^2 1_{D^c}\right)\right]\\ &+\mathbb{E}\left(\mathbf{1}_{D}\right)\left[\max\left(\mathbb{A}^2 1_{D}\right)+\max\left(\mathbb{B}^2 1_{D}\right)\right]\Big)\end{aligned} \tag{4.25}$$

Therefore

$$\begin{aligned}\mathbb{E}\left(\mathbb{A}^2\right)+\mathbb{E}\left(\mathbb{B}^2\right) =\ & O\left(\max\left(\mathbb{A}^2 1_{D^c}\right)\right)+O\left(\max\left(\mathbb{B}^2 1_{D^c}\right)\right)\\ &+O\left(h^{1-\delta}\right)\left(O\left(\max\left(\mathbb{A}^2 1_{D}\right)\right)+O\left(\max\left(\mathbb{B}^2 1_{D}\right)\right)\right)\end{aligned} \tag{4.26}$$

For paths in D^c, we know by definition of non-extreme paths that $\dfrac{\partial S}{\partial \theta}^2 \leq h^{-2\gamma}$. Also, if we take $\delta = 2\gamma$, the definition of D^c ensures that as h tends to 0,

$$\min_{i=1..I}\left(|S-K_i|\right) \geq h^{1-\delta} \geq h^{1-\gamma} \geq \left|S-\widehat{S}\right| \tag{4.27}$$

Therefore S and $\widehat{S}$ are not separated by a discontinuity and we can use the $\widetilde{L}$-Lipschitz continuity of $\left(\dfrac{\partial P}{\partial S}\right)$ on the interval $\left[\min\left(S,\widehat{S}\right),\max\left(S,\widehat{S}\right)\right]$ and conclude that $\left|\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S}\right| \leq \widetilde{L}\left|S - \widehat{S}\right| \leq \widetilde{L}\, h^{1-\gamma}$. Thus

$$\mathbb{A}^2\, \mathbf{1}_{D^c} = \frac{\partial S}{\partial \theta}^2 \left(\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S}\right)^2 \mathbf{1}_{D^c} \leq \widetilde{L}^2\, h^{2-4\gamma} \tag{4.28}$$

P being differentiable and $\widetilde{L}$-Lipschitz away from the discontinuities, i.e. for all final values S of paths belonging to D^c, we directly get $\frac{\partial \widehat{P}}{\partial S}^2 \leq \widetilde{L}^2$ on D^c. By definition of non-extreme paths, we also get $\left|\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta}\right| \leq h^{1-\gamma}$ on this same set. Therefore

$$\mathbb{B}^2\, \mathbf{1}_{D^c} = \frac{\partial \widehat{P}}{\partial S}^2 \left(\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta}\right)^2 \mathbf{1}_{D^c} \leq \widetilde{L}^2\, h^{2-2\gamma} \tag{4.29}$$

On D, by definition of non-extreme paths, we still have $\left|\frac{\partial S}{\partial \theta}\right| \leq h^{-\gamma}$. Let $J_i := \left|\frac{\partial P}{\partial S}\left(K_i^+\right) - \frac{\partial P}{\partial S}\left(K_i^-\right)\right|$ be the size of the discontinuity of $\frac{\partial P}{\partial S}$ at K_i. The "worst case scenario" for $\left|\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S}\right|$ is when S and $\widehat{S}$ are on different sides of a discontinuity of $\frac{\partial P}{\partial S}$. In this case, $\left|\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S}\right| \leq \max_i\left(J_i\right) + \widetilde{L}\, h^{1-\gamma} = O\left(1\right)$. Therefore

$$\mathbb{A}^2\, \mathbf{1}_{D} = \frac{\partial S}{\partial \theta}^2 \left(\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S}\right)^2 \mathbf{1}_{D} \leq \left(\max_i\left(J_i\right) + \widetilde{L}\, h^{1-\gamma}\right)^2 h^{-2\gamma} \tag{4.30}$$

On D, by definition of non-extreme paths, we also have $\left|\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta}\right| \leq h^{1-\gamma}$ and almost surely (i.e. at all points where the payoff is differentiable) $\frac{\partial \widehat{P}}{\partial S}^2 \leq \widetilde{L}^2$ and

$$\mathbb{B}^2\, \mathbf{1}_{D} = \frac{\partial \widehat{P}}{\partial S}^2 \left(\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta}\right)^2 \mathbf{1}_{D} \leq \widetilde{L}^2\, h^{2-2\gamma} \tag{4.31}$$

Finally, plugging those results into (4.26), we get

$$\begin{aligned}\mathbb{E}\left(\mathbb{A}^2\right) + \mathbb{E}\left(\mathbb{B}^2\right) &= O\left(h^{2-4\gamma}\right) + O\left(h^{2-2\gamma}\right) \\ &\qquad + O\left(h^{1-2\gamma}\right)\left(O\left(h^{2-2\gamma}\right) + O\left(h^{-2\gamma}\right)\right) \\ &= O\left(h^{1-4\gamma}\right)\end{aligned} \tag{4.32}$$

This means that in the case of a Lipschitz payoff whose first derivative is Lipschitz by parts and has a finite number of discontinuities we get $\beta = 1 - 4\gamma$ for any $0 < \gamma < 1/2$.

4.2.2 Order of convergence α

The analysis of the order of weak convergence α is in all points similar to what we have done for the convergence rate β.

Contribution of extreme paths

We show that extreme paths have a negligible contribution to the global expectation using

$$\mathbb{E}\left(\mathbf{1}_E \mathbb{A}\right) \leq \sqrt{\mathbb{E}\left(\mathbf{1}_E^2\right)\, \mathbb{E}\left(\frac{\partial S}{\partial \theta}^2 \left(\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S}\right)^2\right)} \tag{4.33}$$

and

$$\mathbb{E}\left(\mathbf{1}_E \mathbb{B}\right) \leq \sqrt{\mathbb{E}\left(\mathbf{1}_E^2\right)\, \mathbb{E}\left(\frac{\partial \widehat{P}}{\partial S}^2 \left(\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta}\right)^2\right)} \tag{4.34}$$

As before, both of those are $o\left(h^p\right)$ for all $p > 0$, and therefore we can focus on the sole contribution of non-extreme paths

$$\begin{aligned}\mathbb{E}\left(\mathbb{A}\right) + \mathbb{E}\left(\mathbb{B}\right) &= O\left(\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}\right) + \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}\right)\right) \\ &= O\left(\mathbb{E}\left(\mathbf{1}_{D}\mathbb{A}\right) + \mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{A}\right)\right. \\ &\qquad\qquad \left. + \mathbb{E}\left(\mathbf{1}_{D}\mathbb{B}\right) + \mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{B}\right)\right)\end{aligned} \tag{4.35}$$

Contribution of discontinuities

As before, we write

$$\begin{aligned}\mathbb{E}\left(\mathbb{A}\right) + \mathbb{E}\left(\mathbb{B}\right) = &O\left(\max \mathbb{A}\mathbf{1}_{D^c}\right) + O\left(\max \mathbb{B}\mathbf{1}_{D^c}\right) \\ &+ O\left(h^{1-\delta}\right)\, \left(O\left(\max \mathbb{A}\mathbf{1}_{D}\right) + O\left(\max \mathbb{B}\mathbf{1}_{D}\right)\right)\end{aligned} \tag{4.36}$$

Using the same arguments as before,

$$|\,\mathbb{A}\,\mathbf{1}_{D^c}\,| = \left| \frac{\partial S}{\partial \theta}\left(\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S}\right)\mathbf{1}_{D^c}\right| \leq \widetilde{L}\, h^{1-2\gamma} \tag{4.37}$$

$$|\,\mathbb{B}\,\mathbf{1}_{D^c}\,| = \left| \frac{\partial \widehat{P}}{\partial S}\left(\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta}\right)\mathbf{1}_{D^c}\right| \leq \widetilde{L}\, h^{1-\gamma} \tag{4.38}$$

$$|\,\mathbb{A}\,\mathbf{1}_{D}\,| = \left| \frac{\partial S}{\partial \theta}\left(\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S}\right)\mathbf{1}_{D}\right| \leq \left(\max_i\left(J_i\right) + \widetilde{L}\, h^{1-\gamma}\right)\, h^{-\gamma} \tag{4.39}$$

$$|\,\mathbb{B}\,\mathbf{1}_{D}\,| = \left| \frac{\partial \widehat{P}}{\partial S}\left(\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta}\right)\mathbf{1}_{D}\right| \leq \widetilde{L}\, h^{1-\gamma} \tag{4.40}$$

Taking $\delta = 2\gamma$ gives eventually $\mathbb{E}(\mathbb{A}) + \mathbb{E}(\mathbb{B}) = O\left(h^{1-3\gamma}\right)$, i.e. $\alpha = 1 - 3\gamma$ for any $0 < \gamma < 1/2$.

4.2.3 Illustration: Case of the European call

We illustrate this kind of payoff with the case of the European call. We recall the payoff is $P(S) = (S-K)^+$ where K is the strike. It is 1-Lipschitz and differentiable everywhere but at K. Its first derivative $\dfrac{\partial P}{\partial S} = \mathbf{1}_{S>K}$ is discontinuous at K and is piecewise 0-Lipschitz everywhere else.

The analysis presented above applies. $\dfrac{\partial P}{\partial S}$ being piecewise constant actually makes it slightly simpler. We have

$$|\,\mathbb{B}\,| = \left| \frac{\partial \widehat{P}}{\partial S} \left(\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta} \right) \right| \leq \left| \left(\frac{\partial S}{\partial \theta} - \frac{\partial \widehat{S}}{\partial \theta} \right) \right| \tag{4.41}$$

For paths in D^c,

$$\mathbb{A}\mathbf{1}_{D^c} = \frac{\partial S}{\partial \theta} \left(\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S} \right) = 0 \tag{4.42}$$

For paths in D,

$$|\,\mathbb{A}\mathbf{1}_D\,| = \left| \frac{\partial S}{\partial \theta} \left(\frac{\partial P}{\partial S} - \frac{\partial \widehat{P}}{\partial S} \right) \right| \leq \left| \frac{\partial S}{\partial \theta} \right| \tag{4.43}$$

This leads to the same convergence rates as before.

4.3 Conclusion

The numerical analysis gives $\alpha = 1, \beta = 2$ for a Lipschitz payoff whose first derivative is also Lipschitz.

The analysis is extended to payoffs for which these properties hold everywhere except on a finite number of points where we permit the first derivative to be discontinuous. In this case, we derive $\alpha = (1-\delta), \beta = (1-\delta)$ for δ as small as we want.

This result corresponds to the numerical results obtained for a European call in the Black Scholes model. In section 2.2 we observed: $\alpha \approx 1$ and $\beta \approx 1$.

We can summarise the results we have proved in this chapter as follows:

Theorem 4.3.1. *We consider an asset S_t on the time interval $[0, T]$ and a European option with a payoff $P(S_T)$. We assume that S_t follows an Ito process as described by equation* (1.2)*, that the coefficients of the diffusion $a(S,t)$ and $b(S,t)$ satisfy conditions* A1 *to* A4 *of theorem 3.4.3 and that $b(S_0, 0) > 0$.*

If the payoff function P is Lipschitz and differentiable with a Lipschitz first derivative $\dfrac{\partial P}{\partial S}$, then our multilevel estimators of the option's Greeks have an accuracy $O(\epsilon)$ at a cost $O\left(\epsilon^{-2}\right)$.

If now we assume that $\dfrac{\partial P}{\partial S}$ *is discontinuous at a finite number of points (e.g. European call), then our multilevel estimators of the Greeks (see section 2.2.1) have an accuracy* $O\left(\epsilon\right)$ *at a cost* $O\left(\epsilon^{-2}\left(\log\epsilon\right)^{2}\right)$*.*

Proof. See above. $\square$

Chapter 5

Analysis of Vanilla European options via pathwise sensitivities and conditional expectations

We analyse the convergence rates of the multilevel algorithms proposed in sections 2.2.2 and 2.3.1 for the computation of the Greeks of European and digital calls.

Again we assume that the regularity and growth conditions found in chapter 3 for the coefficients of SDE 1.2 are satisfied, thereby ensuring we can use the results of the corresponding theorems to obtain analytical bounds on the coefficients α and β in theorem 1.2.1.

5.1 Discontinuous payoffs: the digital call

5.1.1 Payoff computation

We recall the conditional expectation formula obtained in equation (2.81) for the digital call,

$$\widehat{P}_f := \mathbb{E}\left(P\left(\widehat{S}^f_{N_f}\right)\middle|\widehat{S}^f_{N_f-1}\right) = \Phi\left(\frac{\mu^f_{N_f-1} - K}{\sigma^f_{N_f-1}}\right) \tag{5.1}$$

where

$$\begin{aligned} \mu^f_{N_f-1} &= \widehat{S}^f_{N_f-1} + a\left(\widehat{S}^f_{N_f-1}, t_{N_f-1}\right) h_f := \widehat{S}^f_{N_f-1} + a^f_{N_f-1} h_f \\ \sigma^f_{N_f-1} &= b\left(\widehat{S}^f_{N_f-1}, t_{N_f-1}\right) \sqrt{h_f} := b^f_{N_f-1}\sqrt{h_f} \end{aligned} \tag{5.2}$$

For the sake of brevity, we now write μ_f and σ_f instead of $\mu^f_{N_f-1}$ and $\sigma^f_{N_f-1}$, i.e. $\widehat{P}_f = \Phi\left(\dfrac{\mu_f - K}{\sigma_f}\right)$.

This is infinitely differentiable with respect to the input parameters. The partial derivatives are

$$
\begin{aligned}
\frac{\partial \widehat{P}_f}{\partial \widehat{S}^f_{N_f-1}} &= \left(\frac{\dot{\mu}_f\sigma_f - (\mu_f - K)\dot{\sigma}_f}{\sigma_f^2}\right)\phi\left(\frac{\mu_f - K}{\sigma_f}\right)\\
\frac{\partial \widehat{P}_f}{\partial \theta} &= \left(\frac{\widetilde{\mu}_f\sigma_f - (\mu_f - K)\widetilde{\sigma}_f}{\sigma_f^2}\right)\phi\left(\frac{\mu_f - K}{\sigma_f}\right)
\end{aligned}
\tag{5.3}
$$

where we have used the notations $\dot{x} := \dfrac{\partial x}{\partial \widehat{S}^f_{N_f-1}}$ and $\widetilde{x} := \dfrac{\partial x}{\partial \theta}$, i.e.

$$
\begin{aligned}
\dot{\mu}_f &:= \frac{\partial \mu_f}{\partial \widehat{S}^f_{N_f-1}} = 1 + \frac{\partial a\left(\widehat{S}^f_{N_f-1}, t^f_{N_f-1}\right)}{\partial \widehat{S}^f_{N_f-1}} h_f := 1 + \dot{a}^f_{N_f-1} h_f\\
\dot{\sigma}_f &:= \frac{\partial \sigma_f}{\partial \widehat{S}^f_{N_f-1}} = \frac{\partial b\left(\widehat{S}^f_{N_f-1}, t^f_{N_f-1}\right)}{\partial \widehat{S}^f_{N_f-1}} \sqrt{h_f} := \dot{b}^f_{N_f-1}\sqrt{h_f}
\end{aligned}
$$

and as usual, assuming h_f does not depend on θ,

$$
\begin{aligned}
\widetilde{\mu}_f &:= \frac{\partial \mu_f}{\partial \theta} = \frac{\partial a\left(\widehat{S}^f_{N_f-1}, t^f_{N_f-1}\right)}{\partial \theta}\, h_f := \widetilde{a}^f_{N_f-1}\, h_f\\
\widetilde{\sigma}_f &:= \frac{\partial \sigma_f}{\partial \theta} = \frac{\partial b\left(\widehat{S}^f_{N_f-1}, t^f_{N_f-1}\right)}{\partial \theta} \sqrt{h_f} := \widetilde{b}^f_{N_f-1}\sqrt{h_f}
\end{aligned}
$$

At the coarse level, as seen in equation (2.82), we get an expression similar to the one at the fine level:

$$
\widehat{P}_c := \mathbb{E}\left(P\left(\widehat{S}^c_{N_c}\right) \middle| \widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2}\right) = \Phi\left(\frac{\mu^c_{N_c-1} - K}{\sigma^c_{N_c-1}}\right)
\tag{5.4}
$$

where

$$
\begin{aligned}
\mu^c_{N_c-1} &= \widehat{S}^c_{N_c-1} + a\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)\, h_c + b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)\Delta W^f_{N_f-2}\\
&:= \widehat{S}^c_{N_c-1} + a^c_{N_c-1}\, h_c + b^c_{N_c-1}\Delta W^f_{N_f-2}\\
\sigma^c_{N_c-1} &= b\left(\widehat{S}_{N_c-1}, t^c_{N-1}\right)\sqrt{\frac{h_c}{2}}\\
&:= b^c_{N_c-1}\sqrt{\frac{h_c}{2}}
\end{aligned}
$$

For the sake of brevity, we now write μ_c and σ_c instead of $\mu^c_{N_c-1}$ and $\sigma^c_{N_c-1}$, i.e. $\widehat{P}_c = \Phi\left(\dfrac{\mu_c - K}{\sigma_c}\right)$. We then obtain the partial derivatives

$$
\begin{aligned}
\frac{\partial \widehat{P}_c}{\partial S^c_{N-1}} &= \left(\frac{\dot{\mu}_c\sigma_c - (\mu_c - K)\dot{\sigma}_c}{\sigma_c^2}\right)\phi\left(\frac{\mu_c - K}{\sigma_c}\right)\\
\frac{\partial \widehat{P}_c}{\partial \theta} &= \left(\frac{\widetilde{\mu}_c\sigma_c - (\mu_c - K)\widetilde{\sigma}_c}{\sigma_c^2}\right)\phi\left(\frac{\mu_c - K}{\sigma_c}\right)
\end{aligned}
\tag{5.5}
$$

where

$$
\begin{aligned}
\dot{\mu}_c := \frac{\partial \mu_c}{\partial S^c_{N-1}} &= 1 + \frac{\partial a\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)}{\partial \widehat{S}^c_{N_c-1}} h_c + \frac{\partial b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)}{\partial \widehat{S}^c_{N_c-1}} \Delta W^f_{N_f-2} \\
&:= 1 + \dot{a}^c_{N_c-1} h_c + \dot{b}^c_{N_c-1} \Delta W^f_{N_f-2} \\
\dot{\sigma}_c := \frac{\partial \sigma_c}{\partial S^c_{N-1}} &= \frac{\partial b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)}{\partial \widehat{S}^c_{N_c-1}} \sqrt{\frac{h_c}{2}} \\
&:= \dot{b}^c_{N_c-1} \sqrt{\frac{h_c}{2}}
\end{aligned}
$$

and

$$
\begin{aligned}
\widetilde{\mu}_c = \frac{\partial \mu_c}{\partial \theta} &= \frac{\partial a\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)}{\partial \theta} h_c + \frac{\partial b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)}{\partial \theta} \Delta W^f_{N_f-2} \\
&:= \widetilde{a}^c_{N_c-1} h_c + \widetilde{b}^c_{N_c-1} \Delta W^f_{N_f-2} \\
\widetilde{\sigma}_c = \frac{\partial \sigma_c}{\partial \theta} &= \frac{\partial b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)}{\partial \theta} \sqrt{\frac{h_c}{2}} \\
&:= \widetilde{b}^c_{N_c-1} \sqrt{\frac{h_c}{2}}
\end{aligned}
$$

5.1.2 Order of convergence β

As before, we note that to analyse the convergence speed of $\mathbb{V}\left(\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\right)$, it is sufficient to simply study $\mathbb{E}\left[\left(\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\right)^2\right]$.

Contribution of extreme paths

As in section 4.2, we consider the following partition of all paths:

$$
\Omega = E \sqcup E^c = E \sqcup (D \sqcup D^c) \tag{5.6}
$$

with E of the set of extreme paths (paths not satisfying one of the three conditions of lemma 3.4.8 for a certain $\gamma < 1/2$), D the set of non-extreme paths for which S_T is "close" to the discontinuity at K and D^c the set of non-extreme paths for which S_T is "far" from it. We define later what "close" and "far" precisely mean.

We have

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\right)^2\right] &= \mathbb{E}\left[\mathbf{1}_{E^c}\left(\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\right)^2\right] \\
&\quad + \mathbb{E}\left[\mathbf{1}_{E}\left(\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\right)^2\right]
\end{aligned} \tag{5.7}
$$

We now prove that the influence of extreme paths is negligible. Using lemma 3.4.8, we get that for all $p > 0$,

$$\mathbb{E}\left(\mathbf{1}_E\right) = O\left(h^p\right) \tag{5.8}$$

We then use Hölder's inequality

$$\mathbb{E}\left[\mathbf{1}_E\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^2\right] \leq \sqrt{\mathbb{E}\left(\mathbf{1}_E\right)\mathbb{E}\left[\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^4\right]} \tag{5.9}$$

Then, we write

$$\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^4 = \left(\frac{\partial \widehat{S}^f_{N_f-1}}{\partial\theta}\frac{\partial \widehat{P}_f}{\partial S^f_{N_f-1}} + \frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{S}^c_{N_c-1}}{\partial\theta}\frac{\partial \widehat{P}_c}{\partial S^c_{N_c-1}} - \frac{\partial \widehat{P}_c}{\partial\theta}\right)^4 \tag{5.10}$$

that is,

$$\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^4 = \wp\left(\frac{\partial \widehat{S}_{N_f-1}}{\partial\theta}, \frac{\partial \widehat{P}_f}{\partial S^f_{N_f-1}}, \frac{\partial \widehat{P}_f}{\partial\theta}, \frac{\partial \widehat{S}^c_{N_c-1}}{\partial\theta}, \frac{\partial \widehat{P}_c}{\partial S^c_{N_c-1}}, \frac{\partial \widehat{P}_c}{\partial\theta}\right) \tag{5.11}$$

where $\wp$ is a polynomial of order 4. Hölder's inequality then guarantees that if we can prove that for a finite range of $k = 0, \ldots, k_{\max}$ there is some fixed value $q \geq 0$ such that

$$\begin{aligned}
\mathbb{E}\left(\left(\frac{\partial \widehat{S}_{N_f-1}}{\partial\theta}\right)^k\right) &= O\left(h^{-q}\right)\\
\mathbb{E}\left(\left(\frac{\partial \widehat{P}_f}{\partial S^f_{N_f-1}}\right)^k\right) &= O\left(h^{-q}\right)\\
\mathbb{E}\left(\left(\frac{\partial \widehat{P}_f}{\partial\theta}\right)^k\right) &= O\left(h^{-q}\right)\\
\mathbb{E}\left(\left(\frac{\partial \widehat{S}^c_{N_c-1}}{\partial\theta}\right)^k\right) &= O\left(h^{-q}\right)\\
\mathbb{E}\left(\left(\frac{\partial \widehat{P}_c}{\partial S^c_{N_c-1}}\right)^k\right) &= O\left(h^{-q}\right)\\
\mathbb{E}\left(\left(\frac{\partial \widehat{P}_c}{\partial\theta}\right)^k\right) &= O\left(h^{-q}\right)
\end{aligned} \tag{5.12}$$

then we also have $\mathbb{E}\left(\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^4\right) < \infty$ and eventually using inequality (5.9) that for all $\widetilde{p} > 0$, $\mathbb{E}\left(\mathbf{1}_E\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^2\right) = O\left(h^{\widetilde{p}}\right)$.

We now prove the inequalities of (5.12).

From (3.80) of theorem 3.4.3, we have for $k = 0, \ldots, k_{\max}$ the existence of constants C_k such that

$$\mathbb{E}\left(\left(\frac{\partial \widehat{S}^f_{N_f-1}}{\partial\theta}\right)^k\right) < C_k \tag{5.13}$$

$$\mathbb{E}\left(\left(\frac{\partial \widehat{S}^c_{N_c-1}}{\partial\theta}\right)^k\right) < C_k \tag{5.14}$$

We also have

$$\begin{aligned}
\left|\frac{\partial \widehat{P}_f}{\partial \widehat{S}^f_{N_f-1}}\right| &= \left|\frac{\dot{\mu}_f\sigma_f - (\mu_f - K)\dot{\sigma}_f}{\sigma_f^2}\right| \phi\left(\frac{\mu_f - K}{\sigma_f}\right) \\
&\leq \left|\frac{\dot{\mu}_f\sigma_f - (\mu_f - K)\dot{\sigma}_f}{\sigma_f^2}\right| \\
&\leq \left|\frac{\left(1 + \dot{a}^f_{N_f-1}h_f\right) b^f_{N_f-1} - \left(\widehat{S}^f_{N_f-1} + a^f_{N_f-1}\, h_f - K\right)\dot{b}^f_{N_f-1}}{\left(b^f_{N_f-1}\right)^2 \sqrt{h_f}}\right|
\end{aligned} \tag{5.15}$$

With the assumption we made in section 3.3.3 that $b\,(S,t) > \epsilon > 0$, we get

$$\left|\frac{\partial \widehat{P}_f}{\partial \widehat{S}^f_{N_f-1}}\right| \leq \left|\frac{\left(1 + \dot{a}^f_{N_f-1}h_f\right) b^f_{N_f-1} - \left(\widehat{S}^f_{N_f-1} + a^f_{N_f-1}\, h_f - K\right)\dot{b}^f_{N_f-1}}{\epsilon^2 \sqrt{h_f}}\right| \tag{5.16}$$

Considering the linear growth assumption $A2$ of theorem 3.4.3 applied to the coefficients of the evolution equation for U_t, we get

$$\left|\frac{\partial \widehat{P}_f}{\partial \widehat{S}^f_{N_f-1}}\right|^k \leq \left|\mathbb{Q}^k\left(\widehat{S}^f_{N_f-1}\right)\right| h_f^{-k/2} \tag{5.17}$$

where $\mathbb{Q}$ is a polynomial of order 2.

A similar reasoning gives

$$\left|\frac{\partial \widehat{P}_f}{\partial\theta}\right| \leq \left|\frac{\left(\widetilde{a}^f_{N_f-1}h_f\right) b^f_{N_f-1} - \left(\widehat{S}^f_{N_f-1} + a^f_{N_f-1}\, h_f - K\right)\widetilde{b}^f_{N_f-1}}{\epsilon^2 \sqrt{h_f}}\right| \tag{5.18}$$

$$\left|\frac{\partial \widehat{P}_f}{\partial\theta}\right|^k \leq \left|\widetilde{\mathbb{Q}}^k\left(\widehat{S}^f_{N_f-1}\right)\right| h_f^{-k/2} \tag{5.19}$$

where $\widetilde{\mathbb{Q}}$ is also a polynomial of order 2.

From (3.80) of theorem 3.4.3, we have for $k = 0, \ldots, k_{\max}$ the existence of constants $\widetilde{C}_k$ such that

$$\begin{aligned}\mathbb{E}\left(\left(\widehat{S}^f_{N_f-1}\right)^k\right) &< \widetilde{C}_k \\ \mathbb{E}\left(\left(\widehat{S}^c_{N_c-1}\right)^k\right) &< \widetilde{C}_k\end{aligned} \tag{5.20}$$

Therefore, Hölder's inequality guarantees that there are constants $C_{Q^k}, C_{\widetilde{Q}^k}$ such that

$$\begin{aligned}\mathbb{E}\left(\left|\mathbb{Q}^k\left(\widehat{S}^f_{N_f-1}\right)\right|\right) &< C_{Q^k} \\ \mathbb{E}\left(\left|\widetilde{\mathbb{Q}}^k\left(\widehat{S}^f_{N_f-1}\right)\right|\right) &< C_{\widetilde{Q}^k}\end{aligned} \tag{5.21}$$

and finally

$$\begin{aligned}\mathbb{E}\left(\left|\frac{\partial \widehat{P}_f}{\partial \widehat{S}^f_{N_f-1}}\right|^k\right) &= O\left(h_f^{-k/2}\right) \\ \mathbb{E}\left(\left|\frac{\partial \widehat{P}_f}{\partial \theta}\right|^k\right) &= O\left(h_f^{-k/2}\right)\end{aligned} \tag{5.22}$$

Similarly, we write

$$\begin{aligned}\left|\frac{\partial \widehat{P}_c}{\partial \widehat{S}^c_{N_c-1}}\right| &= \left|\frac{\dot{\mu}_c\sigma_c - (\mu_c - K)\dot{\sigma}_c}{\sigma_c^2}\right| \phi\left(\frac{\mu_c - K}{\sigma_c}\right) \\ &\le \left|\frac{\dot{\mu}_c\sigma_c - (\mu_c - K)\dot{\sigma}_c}{\sigma_c^2}\right| \\ &\le \left|\frac{\left(1 + \dot{a}^c_{N_c-1}h_c + \dot{b}^c_{N_c-1}\Delta W^f_{N_f-2}\right) b^c_{N_c-1}}{\left(b^c_{N_c-1}\right)^2 \sqrt{h_f}}\right. \\ &\qquad \left. - \frac{\left(\widehat{S}^c_{N_c-1} + a^c_{N_c-1}\, h_c + b^c_{N_c-1}\Delta W^f_{N_f-2} - K\right) \dot{b}^c_{N_c-1}}{\left(b^c_{N_c-1}\right)^2 \sqrt{h_f}}\right|\end{aligned} \tag{5.23}$$

With our assumption that $b\,(S,t) > \epsilon > 0$, we get

$$\begin{aligned}\left|\frac{\partial \widehat{P}_c}{\partial \widehat{S}^c_{N_c-1}}\right| &\le \left|\frac{\left(1 + \dot{a}^c_{N_c-1}h_c + \dot{b}^c_{N_c-1}\Delta W^f_{N_f-2}\right) b^c_{N_c-1}}{\epsilon^2 \sqrt{h_f}}\right. \\ &\qquad \left. - \frac{\left(\widehat{S}^c_{N_c-1} + a^c_{N_c-1}\, h_c + b^c_{N_c-1}\Delta W^f_{N_f-2} - K\right) \dot{b}^c_{N_c-1}}{\epsilon^2 \sqrt{h_f}}\right|\end{aligned} \tag{5.24}$$

Considering the linear growth assumption $A2$ of theorem 3.4.3 applied to the coefficients of the evolution equation for U_t, we get

$$\left| \frac{\partial \widehat{P}_c}{\partial \widehat{S}^c_{N_c-1}} \right|^k \leq \left| \mathfrak{R}^k \left(\widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2} \right) \right| h_f^{-k/2} \tag{5.25}$$

where $\mathfrak{R}$ is a polynomial of order 2.

A similar reasoning gives that

$$\begin{aligned} \left| \frac{\partial \widehat{P}_c}{\partial \theta} \right| \leq \Bigg| & \frac{\left(\widetilde{a}^c_{N_c-1} h_c + \widetilde{b}^c_{N_c-1} \Delta W^f_{N_f-2} \right) b^c_{N_c-1}}{\epsilon^2 \sqrt{h_f}} \\ & - \frac{\left(\widehat{S}^c_{N_c-1} + a^c_{N_c-1} h_c + b^c_{N_c-1} \Delta W^f_{N_f-2} - K \right) \widetilde{b}^c_{N_c-1}}{\epsilon^2 \sqrt{h_f}} \Bigg| \end{aligned} \tag{5.26}$$

$$\left| \frac{\partial \widehat{P}_c}{\partial \theta} \right|^k \leq \left| \widetilde{\mathfrak{R}}^k \left(\widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2} \right) \right| h_f^{-k/2} \tag{5.27}$$

where $\widetilde{\mathfrak{R}}$ is also a polynomial of order 2.

We recall that for $k = 0, \dots, k_{\max}$ we have

$$\mathbb{E} \left(\left(\widehat{S}^c_{N_c-1} \right)^k \right) < \widetilde{C}_k \tag{5.28}$$

There is a fixed value M_k for each $k = 0, \dots, k_{\max}$ such that

$$\mathbb{E} \left(\left(\Delta W^f_{N_f-2} \right)^k \right) < M_k \tag{5.29}$$

Therefore, Hölder's inequality guarantees that there are constants $C_{R^k}, C_{\widetilde{R}^k}$ such that

$$\begin{aligned} \mathbb{E} \left(\left| \mathfrak{R}^k \left(\widehat{S}^f_{N_c-1},, \Delta W^f_{N_f-2} \right) \right| \right) &< C_{R^k} \\ \mathbb{E} \left(\left| \widetilde{\mathfrak{R}}^k \left(\widehat{S}^c_{N_c-1}, \Delta W^f_{N_f-2} \right) \right| \right) &< C_{\widetilde{R}^k} \end{aligned} \tag{5.30}$$

and finally

$$\begin{aligned} \mathbb{E} \left(\left| \frac{\partial \widehat{P}_c}{\partial \widehat{S}^c_{N_c-1}} \right|^k \right) &= O\left(h_f^{-k/2} \right) \\ \mathbb{E} \left(\left| \frac{\partial \widehat{P}_c}{\partial \theta} \right|^k \right) &= O\left(h_f^{-k/2} \right) \end{aligned} \tag{5.31}$$

By taking $q = k_{\max}/2$, we have the proof that for all $\widetilde{p} > 0$,

$$\mathbb{E}\left(\mathbf{1}_E\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^2\right) = O\left(h^{\widetilde{p}}\right) \tag{5.32}$$

i.e. the contribution of extreme paths E is negligible. The rest of the analysis therefore focuses on E^c.

Contribution of non-extreme paths

Using again the notation $\widehat{\delta} = \dfrac{\partial \widehat{S}}{\partial \theta}$, we write

$$\begin{aligned}
\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta} &= \frac{\partial \widehat{S}^f_{N_f-1}}{\partial \theta}\frac{\partial \widehat{P}_f}{\partial S^f_{N_f-1}} + \frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{S}^c_{N_c-1}}{\partial \theta}\frac{\partial \widehat{P}_c}{\partial S^c_{N_c-1}} - \frac{\partial \widehat{P}_c}{\partial \theta} \\
&= \left[\frac{\partial \widehat{S}^f_{N_f-1}}{\partial \theta}\left(\frac{\dot{\mu}_f \sigma_f - (\mu_f - K)\dot{\sigma}_f}{\sigma_f^2}\right) + \left(\frac{\widetilde{\mu}_f \sigma_f - (\mu_f - K)\widetilde{\sigma}_f}{\sigma_f^2}\right)\right]\phi\left(\frac{\mu_f - K}{\sigma_f}\right) \\
&\quad - \left[\frac{\partial \widehat{S}^c_{N_c-1}}{\partial \theta}\left(\frac{\dot{\mu}_c \sigma_c - (\mu_c - K)\dot{\sigma}_c}{\sigma_c^2}\right) + \left(\frac{\widetilde{\mu}_c \sigma_c - (\mu_c - K)\widetilde{\sigma}_c}{\sigma_c^2}\right)\right]\phi\left(\frac{\mu_c - K}{\sigma_c}\right) \\
&= \left[\left(\widehat{\delta}^f_{N_f-1}\dot{\mu}_f + \widetilde{\mu}_f\right)\frac{1}{\sigma_f} - (\mu_f - K)\left(\frac{\delta^f_{N_f-1}\dot{\sigma}_f + \widetilde{\sigma}_f}{\sigma_f^2}\right)\right]\phi\left(\frac{\mu_f - K}{\sigma_f}\right) \\
&\quad - \left[\left(\widehat{\delta}^c_{N_c-1}\dot{\mu}_c + \widetilde{\mu}_c\right)\frac{1}{\sigma_c} - (\mu_c - K)\left(\frac{\delta^c_{N_c-1}\dot{\sigma}_c + \widetilde{\sigma}_c}{\sigma_c^2}\right)\right]\phi\left(\frac{\mu_c - K}{\sigma_c}\right)
\end{aligned} \tag{5.33}$$

We then rewrite it as

$$\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta} := \left[\dot{M}_f L_f + M_f \dot{L}_f\right]\phi\left(M_f L_f\right) - \left[\dot{M}_c L_c + M_c \dot{L}_c\right]\phi\left(M_c L_c\right) \tag{5.34}$$

where we define

$$
\begin{aligned}
M_f &:= (\mu_f - K) \\
\dot{M}_f &:= \left(\widehat{\delta}^f_{N_f-1}\dot{\mu}_f + \widetilde{\mu}_f\right) \\
L_f &:= \frac{1}{\sigma_f} \\
\dot{L}_f &:= -\left(\frac{\delta^f_{N_f-1}\dot{\sigma}_f + \widetilde{\sigma}_f}{\sigma_f^2}\right) \\
M_c &:= (\mu_c - K) \\
\dot{M}_c &:= \left(\widehat{\delta}^c_{N_c-1}\dot{\mu}_c + \widetilde{\mu}_c\right) \\
L_c &:= \frac{1}{\sigma_c} \\
\dot{L}_c &:= -\left(\frac{\delta^c_{N_c-1}\dot{\sigma}_c + \widetilde{\sigma}_c}{\sigma_c^2}\right)
\end{aligned}
\tag{5.35}
$$

Note that $\dot{M}_f, \dot{L}_f, \dot{M}_c$ and $\dot{L}_c$ correspond to total derivatives of M_f, L_f, M_c and L_c with respect to θ.

We rewrite the previous equation as

$$
\begin{aligned}
\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta} = &\left(\dot{M}_f L_f \phi\left(M_f L_f\right) - \dot{M}_c L_c \phi\left(M_c L_c\right)\right) \\
&+ \left(M_f \dot{L}_f \phi\left(M_f L_f\right) - M_c \dot{L}_c \phi\left(M_c L_c\right)\right)
\end{aligned}
\tag{5.36}
$$

Defining the notation $\Delta X = X_f - X_c$ (for any value X defined at the fine and coarse level), we use the following decomposition:

$$
\begin{aligned}
\Delta\frac{\mathrm{d}\widehat{P}}{\mathrm{d}\theta} &= \Delta\left(\dot{M} L \phi\left(ML\right)\right) + \Delta\left(M \dot{L} \phi\left(ML\right)\right) \\
&= \Delta\left(\dot{M}\right) L_f \phi\left(M_f L_f\right) + \dot{M}_c \Delta\left(L\right) \phi\left(M_f L_f\right) + \dot{M}_c L_c \Delta\left(\phi\left(ML\right)\right) \\
&\quad + \left(\Delta\left(M\right) \dot{L}_f \phi\left(M_f L_f\right) + M_c \Delta\left(\dot{L}\right) \phi\left(M_f L_f\right) + M_c \dot{L}_c \Delta\left(\phi\left(ML\right)\right)\right)
\end{aligned}
\tag{5.37}
$$

We then note that for any finite family of square integrable random variables X_i, we have

$$
\mathbb{E}\left[\left(\sum_{i=1}^{I} X_i\right)^2\right] \leq I \sum_{i=1}^{I} \mathbb{E}\left[X_i^2\right] \tag{5.38}
$$

We can therefore restrict our analysis to the convergence of the expected squares of $L_f \phi\left(M_f L_f\right) \Delta\left(\dot{M}\right)$, $\dot{M}_c \phi\left(M_f L_f\right) \Delta\left(L\right)$, $\dot{L}_f \phi\left(M_f L_f\right) \Delta\left(M\right)$, $M_c \phi\left(M_f L_f\right) \Delta\left(\dot{L}\right)$, $\dot{M}_c L_c \Delta\left(\phi\left(ML\right)\right)$ and $M_c \dot{L}_c \Delta\left(\phi\left(ML\right)\right)$.

Writing the terms of (5.35) in an expanded form gives

$$
\begin{aligned}
M_f &= \left(\widehat{S}^f_{N_f-1} + a^f_{N_f-1}\, h_f - K\right)\\
\dot{M}_f &= \widehat{\delta}^f_{N_f-1}\left(1 + \dot{a}^f_{N_f-1} h_f\right) + \left(\widetilde{a}^f_{N_f-1}\, h_f\right)\\
L_f &= \frac{1}{b^f_{N_f-1}\sqrt{h_f}}\\
\dot{L}_f &= \left(\frac{\widehat{\delta}^f_{N_f-1}\dot{b}^f_{N_f-1} + \widetilde{b}^f_{N_f-1}}{\left(b^f_{N_f-1}\right)^2\sqrt{h_f}}\right)\\
M_c &= \left(\widehat{S}^c_{N_c-1} + a^c_{N_c-1}\, h_c + b^c_{N_c-1}\Delta W^f_{N_f-2} - K\right)\\
\dot{M}_c &= \widehat{\delta}^c_{N_c-1}\left(1 + \dot{a}^c_{N_c-1}\, h_c + \dot{b}^c_{N_c-1}\Delta W^f_{N_f-2}\right) + \left(\widetilde{a}^c_{N_c-1} h_c + \widetilde{b}^c_{N_c-1}\Delta W^f_{N_f-2}\right)\\
L_c &= \frac{1}{b^c_{N_c-1}\sqrt{\frac{h_c}{2}}}\\
\dot{L}_c &= \left(\frac{\widehat{\delta}^c_{N_c-1}\dot{b}^c_{N_c-1} + \widetilde{b}^c_{N_c-1}}{\left(b^c_{N_c-1}\right)^2\sqrt{\frac{h_c}{2}}}\right)
\end{aligned}
\tag{5.39}
$$

Contribution of the paths "far" from the discontinuity

Here we define D as the set of non-extreme paths "close" to the strike, i.e. for which $|\, S_T - K \,| \leq h^{1/2-3\gamma}$ and D^c the non-extreme paths that do not satisfy this condition, i.e. "far" from the strike.

As shown in section 3.3, the probability density function of S_T is smooth. Using arguments similar to those presented in chapter 4, we can show that a proportion $O\left(h^{1/2-3\gamma}\right)$ of all paths is in D, and a proportion $O\left(1\right)$ of all paths is in D^c, i.e.

$$
\mathbb{P}\left(D\right) = O\left(h^{1/2-3\gamma}\right) \tag{5.40}
$$

Using the law of total expectation,

$$
\begin{aligned}
\mathbb{E}\left(\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^2\right) &= O\left(\mathbb{E}\left[\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^2 \mathbf{1}_{E^c}\right]\right)\\
&= \mathbb{P}\left(D^c\right) O\left(\mathbb{E}\left[\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^2 \middle| D^c\right]\right)\\
&\quad + \mathbb{P}\left(D\right) O\left(\mathbb{E}\left[\left(\frac{\mathrm{d}\widehat{P}_f}{\mathrm{d}\theta} - \frac{\mathrm{d}\widehat{P}_c}{\mathrm{d}\theta}\right)^2 \middle| D\right]\right)
\end{aligned}
\tag{5.41}
$$

On D^c, we have

$$
\begin{aligned}
|\, M_f \,| &= \left| \widehat{S}^f_{N_f-1} + a^f_{N_f-1}\, h_f - K \right| \\
&\geq \left| \left| \widehat{S}^f_{N_f-1} - K \right| - \left| a^f_{N_f-1}\, h_f \right| \right| \\
&\geq \left| \left| \left| \widehat{S}^f_{N_f-1} - S_T \right| - |\, S_T - K \,| \right| - \left| a^f_{N_f-1}\, h_f \right| \right|
\end{aligned}
\tag{5.42}
$$

note that using the linear growth of the coefficient $a(S,t)$ and the bounds imposed on paths of D^c,

$$
\begin{aligned}
|\, S_T - K \,| &\geq h_f^{1/2-3\gamma} \\
\left| \widehat{S}^f_{N_f-1} - S_T \right| &= O\left(h_f^{1/2-2\gamma}\right) = o\left(h_f^{1/2-3\gamma}\right) \\
\left| a^f_{N_f-1}\, h_f \right| &= O\left(h_f^{1-\gamma}\right) = o\left(h_f^{1/2-3\gamma}\right)
\end{aligned}
\tag{5.43}
$$

so that there is a constant $C_{M_f} > 0$ such that

$$
|\, M_f \,| \geq C_{M_f} h_f^{1/2-3\gamma} \tag{5.44}
$$

Using the linear growth of $b(S,t)$, there is a constant $C_{L_f} > 0$ such that

$$
|\, L_f \,| \geq C_{L_f} h_f^{-1/2+\gamma} \tag{5.45}
$$

Therefore

$$
\begin{aligned}
|\, M_f L_f \,| &\geq C_{M_f} h_f^{1/2-3\gamma} C_{L_f} h_f^{-1/2+\gamma} \\
&\geq C_{M_f} C_{L_f} O\left(h_f^{-2\gamma}\right)
\end{aligned}
\tag{5.46}
$$

and finally due to the behaviour of ϕ in its tails, for all $p > 0$ we have

$$
\phi\,(M_f L_f) = O\left(h_f^p\right) \tag{5.47}
$$

Very similarly we can show that on D^c, we also have constants $C_{M_c} > 0$ and $C_{L_c} > 0$ such that

$$
\begin{aligned}
|\, M_c \,| &\geq C_{M_c} h_f^{1/2-3\gamma} \\
|\, L_c \,| &\geq C_{L_c} h_f^{-1/2+\gamma} \\
|\, M_c L_c \,| &\geq C_{M_c} C_{L_c} h_f^{-2\gamma}
\end{aligned}
\tag{5.48}
$$

and finally due to the behaviour of ϕ in its tails, for all $p > 0$ we have

$$
\phi\,(M_c L_c) = O\left(h_f^p\right) \tag{5.49}
$$

This means that unless the path is in D, all the terms we have to analyse are $O\left(h_f^p\right)$, i.e. only the paths of D have a significant contribution to the global variance.

Analysis of $L_f \phi(M_f L_f) \Delta\left(\dot{M}\right)$

We have

$$\begin{aligned}\dot{M}_c = \widehat{\delta}^c_{N_c-1}\left(1+\dot{a}^c_{N_c-1}\,h_c+\dot{b}^c_{N_c-1}\Delta W^f_{N_f-2}\right)\\ +\left(\widetilde{a}^c_{N_c-1}h_c+\widetilde{b}^c_{N_c-1}\Delta W^f_{N_f-2}\right)\end{aligned} \tag{5.50}$$

Noting that at the fine level we used the Milstein scheme on $\left[t^f_{N_f-2}, t^f_{N_f-1}\right]$, we can write that

$$\begin{aligned}\widehat{\delta}^f_{N_f-1} = \widehat{\delta}^f_{N_f-2} &+ \left(\widetilde{a}^f_{N_f-2}+\dot{a}^f_{N_f-2}\widehat{\delta}^f_{N_f-2}\right)h_f+\left(\widetilde{b}^f_{N_f-2}+\dot{b}^f_{N_f-2}\widehat{\delta}^f_{N_f-2}\right)\Delta W^f_{N_f-2}\\ &+\frac{1}{2}\left[\widehat{\delta}^f_{N_f-2}\left(\dot{b}^f_{N_f-2}\right)^2+\widehat{\delta}^f_{N_f-2}b^f_{N^f-2}\ddot{b}^f_{N_f-2}\right.\\ &\left.\qquad+\widetilde{b}^f_{N_f-2}\dot{b}^f_{N_f-2}+\dot{b}^f_{N_f-2}\widetilde{\dot{b}}^f_{N_f-2}\right]\left(\left(\Delta W^f_{N_f-2}\right)^2-h_f\right)\end{aligned} \tag{5.51}$$

We let

$$\begin{aligned}\Delta_{Mil} := \frac{1}{2}&\left[\widehat{\delta}^f_{N_f-2}\left(\dot{b}^f_{N_f-2}\right)^2+\widehat{\delta}^f_{N_f-2}b^f_{N^f-2}\ddot{b}^f_{N_f-2}\right.\\ &\left.+\widetilde{b}^f_{N_f-2}\dot{b}^f_{N_f-2}+\dot{b}^f_{N_f-2}\widetilde{\dot{b}}^f_{N_f-2}\right]\left(\left(\Delta W^f_{N_f-2}\right)^2-h_f\right)\end{aligned} \tag{5.52}$$

Therefore

$$\begin{aligned}\widehat{\delta}^f_{N_f-1} = \widehat{\delta}^f_{N_f-2}\left(1+\dot{a}^f_{N_f-2}h_f+\dot{b}^f_{N_f-2}\Delta W^f_{N_f-2}\right)\\ +\widetilde{a}^f_{N_f-2}h_f+\widetilde{b}^f_{N_f-2}\Delta W^f_{N_f-2}+\Delta_{Mil}\end{aligned} \tag{5.53}$$

and

$$\begin{aligned}\dot{M}_f &= \left[\widehat{\delta}^f_{N_f-2}\left(1+\dot{a}^f_{N_f-2}h_f+\dot{b}^f_{N_f-2}\Delta W^f_{N_f-2}\right)\right.\\ &\qquad\left.+\widetilde{a}^f_{N_f-2}h_f+\widetilde{b}^f_{N_f-2}\Delta W^f_{N_f-2}+\Delta_{Mil}\right]\left(1+\dot{a}^f_{N_f-1}h_f\right)\\ &\qquad+\left(\widetilde{a}^f_{N_f-1}\,h_f\right)\\ &= \widehat{\delta}^f_{N_f-2}\left(1+\frac{\dot{a}^f_{N_f-1}+\dot{a}^f_{N_f-2}}{2}\,h_c+\dot{b}^f_{N_f-2}\Delta W^f_{N_f-2}\right)\\ &\quad+\left(\frac{\widetilde{a}^f_{N_f-2}+\widetilde{a}^f_{N_f-1}}{2}h_c+\widetilde{b}^f_{N_f-2}\Delta W^f_{N_f-2}\right)\\ &\quad+\dot{a}^f_{N_f-1}h_f\left(\dot{a}^f_{N_f-2}h_f+\dot{b}^f_{N_f-2}\Delta W^f_{N_f-2}\right.\\ &\qquad\left.+\widetilde{a}^f_{N_f-2}h_f+\widetilde{b}^f_{N_f-2}\Delta W^f_{N_f-2}\right)\\ &\quad+\Delta_{Mil}\left(1+\dot{a}^f_{N_f-1}h_f\right)\end{aligned} \tag{5.54}$$

As we are only considering the paths in E^c, and recalling that $a(S,t)$ and $b(S,t)$ obey the linear growth and Lipschitz conditions of theorem 3.4.3, we have

$$\begin{aligned}
&\dot{a}^f_{N_f-1}h_f\left[\dot{a}^f_{N_f-2}h_f+\dot{b}^f_{N_f-2}\Delta W^f_{N_f-2}+\widetilde{a}^f_{N_f-2}h_f+\widetilde{b}^f_{N_f-2}\Delta W^f_{N_f-2}\right]\\
&\qquad=O\left(h_f^{-\gamma}\right)O\left(h_f\right)\left[O\left(h_f^{-\gamma}\right)O\left(h_f\right)+O\left(h_f^{-\gamma}\right)O\left(h_f^{1/2-\gamma}\right)\right.\\
&\qquad\qquad\left.+O\left(h_f^{-\gamma}\right)O\left(h_f\right)+O\left(h_f^{-\gamma}\right)O\left(h_f^{1/2-\gamma}\right)\right]\\
&\qquad=O\left(h_f^{1-\gamma}\right)\mathcal{O}\left(h_f^{1/2-\gamma}\right)\\
&\qquad=O\left(h_f^{3/2-2\gamma}\right)
\end{aligned}\tag{5.55}$$

and

$$\begin{aligned}
&\Delta_{Mil}\left(1+\dot{a}^f_{N_f-1}h_f\right)\\
&\quad=\frac{1}{2}\left(\widehat{\delta}^f_{N_f-2}\left(\dot{b}^f_{N_f-2}\right)^2+\widehat{\delta}^f_{N_f-2}b^f_{Nf-2}\ddot{b}^f_{N_f-2}\right.\\
&\qquad\qquad\left.+\widetilde{b}^f_{N_f-2}\dot{b}^f_{N_f-2}+\dot{b}^f_{N_f-2}\dot{\widetilde{b}}^f_{N_f-2}\right)\\
&\qquad\qquad\qquad\left(\left(\Delta W^f_{N_f-2}\right)^2-h_f\right)\left(1+\dot{a}^f_{N_f-1}h_f\right)\\
&\quad=\left(O\left(h_f^{-\gamma}\right)O\left(h_f^{-2\gamma}\right)+O\left(h_f^{-\gamma}\right)O\left(h_f^{-\gamma}\right)O\left(h_f^{-\gamma}\right)\right.\\
&\qquad\qquad\left.+O\left(h_f^{-\gamma}\right)O\left(h_f^{-\gamma}\right)+O\left(h_f^{-\gamma}\right)O\left(h_f^{-\gamma}\right)\right)\\
&\qquad\qquad\qquad\left(O\left(h_f^{1-2\gamma}\right)+O\left(h_f\right)\right)\left(1+O\left(h_f^{-\gamma}\right)h_f\right)\\
&\quad=O\left(h_f^{-3\gamma}\right)O\left(h_f^{1-2\gamma}\right)O\left(1\right)\\
&\quad=O\left(h_f^{1-5\gamma}\right)
\end{aligned}\tag{5.56}$$

We write

$$\begin{aligned}
\Delta\dot{M}&=\dot{M}_f-\dot{M}_c\\
&=\widehat{\delta}^f_{N_f-1}\left(1+\dot{a}^f_{N_f-1}h_f\right)+\left(\widetilde{a}^f_{N_f-1}\,h_f\right)\\
&\qquad-\left[\widehat{\delta}^c_{N_c-1}\left(1+\dot{a}^c_{N_c-1}\,h_c+\dot{b}^c_{N_c-1}\Delta W^f_{N_f-2}\right)+\left(\widetilde{a}^c_{N_c-1}h_c+\widetilde{b}^c_{N_c-1}\Delta W^f_{N_f-2}\right)\right]
\end{aligned}\tag{5.57}$$

Using the previous expansion

$$
\begin{aligned}
\Delta \dot{M} = \widehat{\delta}^{f}_{N_f-2} & \left(1 + \frac{\dot{a}^{f}_{N_f-1} + \dot{a}^{f}_{N_f-2}}{2}\, h_c + \dot{b}^{f}_{N_f-2} \Delta W^{f}_{N_f-2}\right) \\
& + \left(\frac{\widetilde{a}^{f}_{N_f-2} + \widetilde{a}^{f}_{N_f-1}}{2} h_c + \widetilde{b}^{f}_{N_f-2} \Delta W^{f}_{N_f-2}\right) \\
& + O\left(h_f^{3/2-2\gamma}\right) \\
& + O\left(h_f^{1-5\gamma}\right) \\
& - \widehat{\delta}^{c}_{N_c-1} \left(1 + \dot{a}^{c}_{N_c-1}\, h_c + \dot{b}^{c}_{N_c-1} \Delta W^{f}_{N_f-2}\right) \\
& - \left(\widetilde{a}^{c}_{N_c-1} h_c + \widetilde{b}^{c}_{N_c-1} \Delta W^{f}_{N_f-2}\right) \\
= \widehat{\delta}^{f}_{N_f-2} & \left(1 + \frac{\dot{a}^{f}_{N_f-1} + \dot{a}^{f}_{N_f-2}}{2}\, h_c + \dot{b}^{f}_{N_f-2} \Delta W^{f}_{N_f-2}\right) \\
& - \widehat{\delta}^{c}_{N_c-1} \left(1 + \dot{a}^{c}_{N_c-1}\, h_c + \dot{b}^{c}_{N_c-1} \Delta W^{f}_{N_f-2}\right) \\
& + \left(\frac{\widetilde{a}^{f}_{N_f-2} + \widetilde{a}^{f}_{N_f-1}}{2} h_c + \widetilde{b}^{f}_{N_f-2} \Delta W^{f}_{N_f-2}\right) \\
& - \left(\widetilde{a}^{c}_{N_c-1} h_c + \widetilde{b}^{c}_{N_c-1} \Delta W^{f}_{N_f-2}\right) \\
& + O\left(h_f^{1-5\gamma}\right) \\
= \widehat{\delta}^{f}_{N_f-2} & \left[\left(1 + \frac{\dot{a}^{f}_{N_f-1} + \dot{a}^{f}_{N_f-2}}{2}\, h_c + \dot{b}^{f}_{N_f-2} \Delta W^{f}_{N_f-2}\right)\right. \\
& \qquad\qquad \left. - \left(1 + \dot{a}^{c}_{N_c-1}\, h_c + \dot{b}^{c}_{N_c-1} \Delta W^{f}_{N_f-2}\right) \right] \\
& + \left(\widehat{\delta}^{f}_{N_f-2} - \widehat{\delta}^{c}_{N_c-1}\right) \left(1 + \dot{a}^{c}_{N_c-1}\, h_c + \dot{b}^{c}_{N_c-1} \Delta W^{f}_{N_f-2}\right) \\
& + \left(\frac{\widetilde{a}^{f}_{N_f-2} + \widetilde{a}^{f}_{N_f-1}}{2} - \widetilde{a}^{c}_{N_c-1}\right) h_c \\
& + \left(\widetilde{b}^{f}_{N_f-2} - \widetilde{b}^{c}_{N_c-1}\right) \Delta W^{f}_{N_f-2} \\
& + O\left(h_f^{1-5\gamma}\right)
\end{aligned}
\tag{5.58}
$$

$$
\begin{aligned}
\Delta \dot{M} = \widehat{\delta}^f_{N_f-2} & \left(\left(\frac{\dot{a}^f_{N_f-1} + \dot{a}^f_{N_f-2}}{2} - \dot{a}^c_{N_c-1} \right) O\left(h_f\right) + \left(\dot{b}^f_{N_f-2} - \dot{b}^c_{N_c-1} \right) O\left(h_f^{1/2-\gamma} \right) \right) \\
& + \left(\widehat{\delta}^f_{N_f-2} - \widehat{\delta}^c_{N_c-1} \right) \left(1 + O\left(h_f^{-\gamma}\right) O\left(h_f\right) + O\left(h_f^{-\gamma}\right) O\left(h_f^{1/2-\gamma}\right) \right) \\
& + \left(\frac{\widetilde{a}^f_{N_f-2} + \widetilde{a}^f_{N_f-1}}{2} - \widetilde{a}^c_{N_c-1} \right) O\left(h_f\right) \\
& + \left(\widetilde{b}^f_{N_f-2} - \widetilde{b}^c_{N_c-1} \right) O\left(h_f^{1/2-\gamma} \right) \\
& + O\left(h_f^{1-5\gamma} \right)
\end{aligned}
\tag{5.59}
$$

Using once again the fact that $a(S,t)$ and $b(S,t)$ obey the Lipschitz conditions of theorem 3.4.3 and the bounds of lemma (3.4.8) on non-extreme paths, we have

$$
\begin{aligned}
\Delta \dot{M} = O\left(h_f^{-\gamma}\right) & \left(O\left(h_f^{1/2-2\gamma}\right) O\left(h_f\right) + O\left(h_f^{1-\gamma}\right) O\left(h_f^{1/2-\gamma}\right) \right) \\
& + O\left(h_f^{1-\gamma}\right) \left(1 + O\left(h_f^{-\gamma}\right) O\left(h_f\right) + O\left(h_f^{-\gamma}\right) O\left(h_f^{1/2-\gamma}\right) \right) \\
& + O\left(h_f^{1/2-2\gamma}\right) O\left(h_f\right) \\
& + O\left(h_f^{1-\gamma}\right) O\left(h_f^{1/2-\gamma}\right) \\
& + O\left(h_f^{1-5\gamma}\right) \\
= O\left(h_f^{1-5\gamma}\right) &
\end{aligned}
\tag{5.60}
$$

Then as we have assumed $b(S,t) > \epsilon$ in section 3.3.3, we have

$$
L_f = O\left(h_f^{-1/2}\right) \tag{5.61}
$$

We also note that $\phi\left(M_f L_f\right) \leq 1$. Therefore on D we have

$$
L_f \phi\left(M_f L_f\right) \Delta \left(\dot{M}\right) = O\left(h_f^{1/2-5\gamma}\right) \tag{5.62}
$$

Analysis of $\dot{M}_c \phi\left(M_f L_f\right) \Delta\left(L\right)$

Using arguments similar to the ones used before, we have

$$
\begin{aligned}
\dot{M}_c &= \widehat{\delta}^c_{N_c-1} \left(1 + \dot{a}^c_{N_c-1} h_c + \dot{b}^c_{N_c-1} \Delta W^f_{N_f-2} \right) + \left(\widetilde{a}^c_{N_c-1} h_c + \widetilde{b}^c_{N_c-1} \Delta W^f_{N_f-2} \right) \\
&= O\left(h_f^{-\gamma}\right) \left(1 + O\left(h_f^{-\gamma}\right) O\left(h_f\right) + O\left(h_f^{-\gamma}\right) O\left(h_f^{1/2-\gamma}\right) \right) \\
&\qquad + \left(O\left(h_f^{-\gamma}\right) O\left(h_f\right) + O\left(h_f^{-\gamma}\right) O\left(h_f^{1/2-\gamma}\right) \right) \\
&= O\left(h_f^{-\gamma}\right) + O\left(h_f^{1/2-2\gamma}\right) \\
&= O\left(h_f^{-\gamma}\right)
\end{aligned}
\tag{5.63}
$$

Once again on D, $\phi\left(M_f L_f\right) \leq 1$ and

$$\begin{aligned}
\Delta\left(L\right) &= \left(\frac{1}{b^f_{N_f-1}\sqrt{h_f}} - \frac{1}{b^c_{N_c-1}\sqrt{\frac{h_c}{2}}}\right) \\
&= \frac{b^c_{N_c-1} - b^f_{N_f-1}}{b^c_{N_c-1} b^f_{N_f-1}\sqrt{h_f}} \\
&= \frac{O\left(h_f^{1/2-2\gamma}\right)}{\epsilon^2\sqrt{h_f}} \\
&= O\left(h_f^{-2\gamma}\right)
\end{aligned} \tag{5.64}$$

and finally we obtin that on D,

$$\dot{M}_c \phi\left(M_f L_f\right) \Delta\left(L\right) = O\left(h_f^{-3\gamma}\right) \tag{5.65}$$

Analysis of $\dot{L}_f \phi\left(M_f L_f\right) \Delta\left(M\right)$

We have

$$\left(b^f_{N_f-1}\right)^2 \sqrt{h_f} \geq \epsilon^2 \sqrt{h_f} \tag{5.66}$$

and

$$\widehat{\delta}^f_{N_f-1} \dot{b}^f_{N_f-1} + \widetilde{b}^f_{N_f-1} = O\left(h_f^{-\gamma}\right) O\left(h_f^{-\gamma}\right) + O\left(h_f^{-\gamma}\right) \tag{5.67}$$

therefore

$$\begin{aligned}
\dot{L}_f &= \frac{\widehat{\delta}^f_{N_f-1} \dot{b}^f_{N_f-1} + \widetilde{b}^f_{N_f-1}}{\left(b^f_{N_f-1}\right)^2 \sqrt{h_f}} \\
&= O\left(h_f^{-1/2-2\gamma}\right)
\end{aligned} \tag{5.68}$$

and $\phi\left(M_f L_f\right) \leq 1$ and

$$\begin{aligned}
\Delta\left(M\right) &= M_f - M_c \\
&= \left(\widehat{S}^f_{N_f-1} + a^f_{N_f-1}\, h_f - K\right) \\
&\qquad - \left(\widehat{S}^c_{N_c-1} + a^c_{N_c-1}\, h_c + b^c_{N_c-1} \Delta W^f_{N_f-2} - K\right)
\end{aligned} \tag{5.69}$$

We once again write the Milstein scheme on $\left[t^f_{N_f-2}, t^f_{N_f-1}\right]$,

$$\begin{aligned}
\widehat{S}^f_{N_f-1} = \widehat{S}^f_{N_f-2} &+ a^f_{N_f-2} h_f + b^f_{N_f-2} \Delta W^f_{N_f-2} \\
&+ \frac{1}{2} b^f_{N_f-2} \dot{b}^f_{N_f-2} \left(\left(\Delta W^f_{N_f-2}\right)^2 - h_f\right)
\end{aligned} \tag{5.70}$$

and the previous equation becomes

$$
\begin{aligned}
\Delta(M) &= \left(\widehat{S}^f_{N_f-2} + a^f_{N_f-2}h_f + b^f_{N_f-2}\Delta W^f_{N_f-2}\right. \\
&\qquad + \frac{1}{2}b^f_{N_f-2}\dot{b}^f_{N_f-2}\left(\left(\Delta W^f_{N_f-2}\right)^2 - h_f\right) + \left. a^f_{N_f-1}\,h_f - K\right) \\
&\qquad\qquad - \left(\widehat{S}^c_{N_c-1} + a^c_{N_c-1}\,h_c + b^c_{N_c-1}\Delta W^f_{N_f-2} - K\right) \\
&= \widehat{S}^f_{N_f-2} - \widehat{S}^c_{N_c-1} + \left(a^f_{N_f-2} + a^f_{N_f-1} - 2a^c_{N_c-1}\right)h_f \\
&\qquad + \left(b^f_{N_f-2} - b^c_{N_c-1}\right)\Delta W^f_{N_f-2} \\
&\qquad + \frac{1}{2}b^f_{N_f-2}\dot{b}^f_{N_f-2}\left(\left(\Delta W^f_{N_f-2}\right)^2 - h_f\right)
\end{aligned}
\tag{5.71}
$$

As before, being in E^c and the coefficients having a linear growth, we get

$$
\begin{aligned}
\Delta(M) &= O\left(h_f^{1-\gamma}\right) + O\left(h_f^{1/2-2\gamma}\right)O(h_f) + O\left(h_f^{1-\gamma}\right)O\left(h_f^{1/2-\gamma}\right) \\
&\qquad + O\left(h_f^{-\gamma}\right)O\left(h_f^{-\gamma}\right)\left(O\left(h_f^{1-2\gamma}\right) + O(h_f)\right) \\
&= O\left(h_f^{1-\gamma}\right) + O\left(h_f^{3/2-2\gamma}\right) + O\left(h_f^{3/2-2\gamma}\right) + O\left(h_f^{1-4\gamma}\right) \\
&= O\left(h_f^{1-4\gamma}\right)
\end{aligned}
\tag{5.72}
$$

Finally we get on D

$$
\dot{L}_f\phi(M_fL_f)\,\Delta(M) = O\left(h_f^{1/2-6\gamma}\right) \tag{5.73}
$$

Analysis of $M_c\phi(M_fL_f)\,\Delta\left(\dot{L}\right)$

In general on E^c,

$$
M_c = \left(\widehat{S}^c_{N_c-1} + a^c_{N_c-1}\,h_c + b^c_{N_c-1}\Delta W^f_{N_f-2} - K\right) = O\left(h_f^{-\gamma}\right) \tag{5.74}
$$

and in particular on D,

$$
M_c = O\left(h_f^{1/2-3\gamma}\right) \tag{5.75}
$$

and $\phi(M_fL_f) \leq 1$. Then,

$$
\begin{aligned}
\Delta\left(\dot{L}\right) &= \left(\frac{\widehat{\delta}^f_{N_f-1}\dot{b}^f_{N_f-1} + \widetilde{b}^f_{N_f-1}}{\left(b^f_{N_f-1}\right)^2\sqrt{h_f}}\right) - \left(\frac{\widehat{\delta}^c_{N_c-1}\dot{b}^c_{N_c-1} + \widetilde{b}^c_{N_c-1}}{\left(b^c_{N_c-1}\right)^2\sqrt{\frac{h_c}{2}}}\right) \\
&= \Delta\left(\widehat{\delta}_{N-1}\dot{b}_{N-1} + \widetilde{b}_{N_f-1}\right)\frac{1}{\left(b^f_{N_f-1}\right)^2\sqrt{h_f}} \\
&\qquad + \left(\widehat{\delta}^c_{N_c-1}\dot{b}^c_{N_c-1} + \widetilde{b}^c_{N_c-1}\right)\Delta\left(\frac{1}{\left(b_{N-1}\right)^2\sqrt{h_f}}\right)
\end{aligned}
\tag{5.76}
$$

$$\left|\Delta\left(\dot{L}\right)\right| \leq \Delta\left(\widehat{\delta}_{N-1}\dot{b}_{N-1}+\widetilde{b}_{N_f-1}\right)\frac{1}{\epsilon^2\sqrt{h_f}} + \left(h_f^{-\gamma}h_f^{-\gamma}+h_f^{-\gamma}\right)\Delta\left(\frac{1}{\left(b_{N-1}\right)^2\sqrt{h_f}}\right) \tag{5.77}$$

that is

$$\Delta\left(\dot{L}\right) = \Delta\left(\widehat{\delta}_{N-1}\dot{b}_{N-1}+\widetilde{b}_{N_f-1}\right)O\left(h_f^{-1/2+\gamma}\right) + O\left(h_f^{-2\gamma}\right)\Delta\left(\frac{1}{\left(b_{N-1}\right)^2\sqrt{h_f}}\right) \tag{5.78}$$

we then have

$$\begin{aligned}\Delta\left(\frac{1}{\left(b_{N-1}\right)^2\sqrt{h_f}}\right) &= \frac{\left(b^c_{N_c-1}\right)^2-\left(b^f_{N_f-1}\right)^2}{\left(b^c_{N_c-1}\right)^2\left(b^f_{N_f-1}\right)^2\sqrt{h_f}} \\ &= \frac{\left(b^c_{N_c-1}-b^f_{N_f-1}\right)\left(b^c_{N_c-1}+b^f_{N_f-1}\right)}{\left(b^c_{N_c-1}\right)^2\left(b^f_{N_f-1}\right)^2\sqrt{h_f}}\end{aligned} \tag{5.79}$$

and

$$\begin{aligned}\Delta\left(\frac{1}{\left(b_{N-1}\right)^2\sqrt{h_f}}\right) &\leq \frac{O\left(h_f^{1/2-2\gamma}\right)O\left(h_f^{-\gamma}\right)}{\epsilon^4\sqrt{h_f}} \\ &= O\left(h_f^{-3\gamma}\right)\end{aligned} \tag{5.80}$$

and

$$\begin{aligned}\Delta\left(\widehat{\delta}_{N-1}\dot{b}_{N-1}+\widetilde{b}_{N_f-1}\right) &= \widehat{\delta}^f_{N_f-1}\dot{b}^f_{N_f-1}+\widetilde{b}^f_{N_f-1}-\left(\widehat{\delta}^c_{N_c-1}\dot{b}^c_{N_c-1}+\widetilde{b}^c_{N_c-1}\right) \\ &= \left(\widehat{\delta}^f_{N_f-1}-\widehat{\delta}^c_{N_c-1}\right)\dot{b}^f_{N_f-1}+\widehat{\delta}^c_{N_c-1}\left(\dot{b}^f_{N_f-1}-\dot{b}^c_{N_c-1}\right) \\ &\quad +\widetilde{b}^f_{N_f-1}-\widetilde{b}^c_{N_c-1}\end{aligned} \tag{5.81}$$

therefore

$$\begin{aligned}\Delta\left(\widehat{\delta}_{N-1}\dot{b}_{N-1}+\widetilde{b}_{N_f-1}\right) &= O\left(h_f^{1/2-2\gamma}\right)O\left(h_f^{-\gamma}\right)+O\left(h_f^{-\gamma}\right)O\left(h_f^{1/2-2\gamma}\right) \\ &\quad +O\left(h_f^{1/2-2\gamma}\right) \\ &= O\left(h_f^{1/2-3\gamma}\right)\end{aligned} \tag{5.82}$$

and finally

$$
\begin{aligned}
\Delta\left(\dot{L}\right) &= O\left(h_f^{1/2-3\gamma}\right) O\left(h_f^{-1/2+\gamma}\right) \\
&\qquad + O\left(h_f^{-2\gamma}\right) O\left(h_f^{-3\gamma}\right) \\
&= O\left(h_f^{-5\gamma}\right)
\end{aligned}
\tag{5.83}
$$

Those results finally yield on D

$$
M_c \phi\left(M_f L_f\right) \Delta\left(\dot{L}\right) = O\left(h_f^{1/2-8\gamma}\right) \tag{5.84}
$$

Analysis of $\dot{M}_c L_c \Delta\phi\left(ML\right)$

We have from above

$$
\dot{M}_c = O\left(h_f^{-\gamma}\right) \tag{5.85}
$$

and as before

$$
\begin{aligned}
\left| L_c \right| &= \left| \frac{1}{b^c_{N_c-1}\sqrt{\frac{h_c}{2}}} \right| \\
&\leq \frac{1}{\epsilon\sqrt{h_f}}
\end{aligned}
\tag{5.86}
$$

Thus

$$
L_c = O\left(h_f^{-1/2}\right) \tag{5.87}
$$

Let us now analyse $\Delta\phi\left(ML\right)$.

$$
\begin{aligned}
\Delta\phi\left(ML\right) &= \phi\left(M_f L_f\right) - \phi\left(M_c L_c\right) \\
&= \phi\left(\left(\widehat{S}^f_{N_f-1} + a^f_{N_f-1}\, h_f - K\right) \frac{1}{b^f_{N_f-1}\sqrt{h_f}}\right) \\
&\qquad - \phi\left(\left(\widehat{S}^c_{N_c-1} + a^c_{N_c-1}\, h_c + b^c_{N_c-1}\Delta W^f_{N_f-2} - K\right) \frac{1}{b^c_{N_c-1}\sqrt{\frac{h_c}{2}}}\right)
\end{aligned}
\tag{5.88}
$$

We first study $\Delta\left(ML\right)$. This will be useful in conjunction with the mean value theorem. We write

$$
\Delta\left(ML\right) = L_f \Delta M + M_c \Delta L \tag{5.89}
$$

From above, we get

$$\begin{aligned} \Delta M &= O\left(h_f^{1-4\gamma}\right) \\ \Delta L &= O\left(h_f^{-2\gamma}\right) \\ L_f &= O\left(h_f^{-1/2}\right) \end{aligned} \tag{5.90}$$

and

$$\begin{aligned} M_c &= \left(\widehat{S}^c_{N_c-1} - K + a^c_{N_c-1}\, h_c + b^c_{N_c-1} \Delta W^f_{N_f-2}\right) \\ &= O\left(h_f^{1/2-3\gamma}\right) + O\left(h_f^{1-\gamma}\right) + O\left(h_f^{1/2-2\gamma}\right) \\ &= O\left(h_f^{1/2-3\gamma}\right) \end{aligned} \tag{5.91}$$

therefore

$$\begin{aligned} \Delta\left(ML\right) &= O\left(h_f^{1-4\gamma}\right) O\left(h_f^{-1/2}\right) + O\left(h_f^{1/2-3\gamma}\right) O\left(h_f^{-2\gamma}\right) \\ &= O\left(h_f^{1/2-5\gamma}\right) \end{aligned} \tag{5.92}$$

Then the mean value theorem guarantees there is a certain value $C \in \left(M_c L_c, M_f L_f\right)$ such that

$$\Delta\phi\left(ML\right) = \Delta\left(ML\right) \phi'\left(C\right) \tag{5.93}$$

and ϕ' is bounded, therefore we finally get

$$\Delta\phi\left(ML\right) = O\left(h_f^{1/2-5\gamma}\right) \tag{5.94}$$

on D. And hence

$$\dot{M}_c L_c \Delta\phi\left(ML\right) = O\left(h_f^{1/2-5\gamma}\right) \tag{5.95}$$

Analysis of $M_c \dot{L}_c \Delta\left(\phi\left(ML\right)\right)$

As before on D^c,

$$\begin{aligned} M_c &= O\left(h_f^{1/2-3\gamma}\right) &\quad \text{on } D \\ M_c &= O\left(h_f^{-\gamma}\right) &\quad \text{on } D^c \end{aligned} \tag{5.96}$$

and

$$
\begin{aligned}
\dot{L}_c &= \frac{\widehat{\delta}^c_{N_c-1}\dot{b}^c_{N_c-1} + \widetilde{b}^c_{N_c-1}}{\left(b^c_{N_c-1}\right)^2\sqrt{\frac{h_c}{2}}} \\
&= \frac{O\left(h_f^{-\gamma}\right)O\left(h_f^{-\gamma}\right) + O\left(h_f^{-\gamma}\right)}{\epsilon^2 O\left(h_f^{1/2}\right)} \\
&= O\left(h_f^{-1/2-2\gamma}\right)
\end{aligned} \tag{5.97}
$$

and as before on D we have

$$
\Delta\phi\left(ML\right) = O\left(h_f^{1/2-5\gamma}\right) \tag{5.98}
$$

Therefore

$$
M_c\dot{L}_c\Delta\left(\phi\left(ML\right)\right) = O\left(h_f^{1/2-10\gamma}\right) \tag{5.99}
$$

Putting things together

Putting these results together using equation (5.41) and the analysis of the terms listed in (5.39), we have

$$
\begin{aligned}
\mathbb{V}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right) &= O\left(\mathbb{E}\left[\Delta\frac{\partial\widehat{P}}{\partial\theta}^2\right]\right) \\
&= \mathbb{P}\left(D^c\right)O\left(\mathbb{E}\left[\left.\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right| D^c\right]\right) \\
&\qquad + \mathbb{P}\left(D\right)\,O\left(\mathbb{E}\left[\left.\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right| D\right]\right)
\end{aligned} \tag{5.100}
$$

and from equation (5.37) and the pertaining remarks,

$$
\mathbb{E}\left[\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right] = O\left(\sum_{i=1}^{6}\mathbb{E}\left(A_i^2\right)\right) \tag{5.101}
$$

with

$$
\begin{aligned}
A_1 &= L_f\phi\left(M_fL_f\right)\Delta\left(\dot{M}\right) \\
A_2 &= \dot{M}_c\phi\left(M_fL_f\right)\Delta\left(L\right) \\
A_3 &= \dot{L}_f\phi\left(M_fL_f\right)\Delta\left(M\right) \\
A_4 &= M_c\phi\left(M_fL_f\right)\Delta\left(\dot{L}\right) \\
A_5 &= \dot{M}_cL_c\Delta\left(\phi\left(ML\right)\right) \\
A_6 &= M_c\dot{L}_c\Delta\left(\phi\left(ML\right)\right)
\end{aligned} \tag{5.102}
$$

The analysis above reveals that A_2 is the largest term and using (5.100),

$$\begin{aligned}\mathbb{V}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right) &= O(1)\,O\left(h_f^{2p}\right)+O\left(h_f^{1/2-3\gamma}\right)\,O\left(h_f^{-6\gamma}\right)\\ &= O\left(h_f^{1/2-9\gamma}\right)\end{aligned} \tag{5.103}$$

This means that the order of convergence for the Digital Call using Conditional Expectations is $\beta=\frac{1}{2}-9\gamma$ for any $\gamma>0$ (and $\gamma<1/2$).

5.1.3 Order of convergence α

In lemma 3.2.4, we have established that to study the rate of weak convergence of the Greeks' estimators $\mathbb{E}\left(\frac{\partial\widehat{P}}{\partial\theta}-\frac{\partial P}{\partial\theta}\right)$, we can equivalently study the rate of convergence of $\mathbb{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)$. The analysis of the weak convergence rate α is then similar to the analysis of β.

$$\begin{aligned}\mathbb{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right) &= \mathbb{P}\left(D^c\right)O\left(\mathbb{E}\left[\Delta\frac{\partial\widehat{P}}{\partial\theta}\middle| D^c\right]\right)\\ &\qquad+\mathbb{P}\left(D\right)\,O\left(\mathbb{E}\left[\Delta\frac{\partial\widehat{P}}{\partial\theta}\middle| D\right]\right)\end{aligned} \tag{5.104}$$

and using the results of section 5.1.2, we have

$$\mathbb{E}\left[\Delta\frac{\partial\widehat{P}}{\partial\theta}\right]=\sum_{i=1}^{6}\mathbb{E}\left(A_i\right) \tag{5.105}$$

We see from our previous results that A_2 is again the limiting factor and we have finally

$$\mathbb{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)=O\left(h^{1/2-6\gamma}\right) \tag{5.106}$$

This means that the order of weak convergence for the Digital Call using Conditional Expectations is $\alpha=\frac{1}{2}-6\gamma$ for any $\gamma>0$ (and $\gamma<1/2$).

5.2 Continuous payoffs with discontinuous first derivative: the European call

We now analyse the computation of Greeks for the European call using pathwise sensitivities and conditional expectations as presented in section 2.2.2.

5.2.1 Payoff computation

We recall that for the European call, with the notations of section 5.1.1, the conditional expectation technique gives the fine payoff estimator

$$\widehat{P}_f = \mathbb{E}\left[P\left(\widehat{S}^f_{N_f}\right)\middle|\widehat{S}^f_{N_f-1}\right] = \sigma_f\,\phi\left(\frac{\mu_f-K}{\sigma_f}\right)+(\mu_f-K)\,\Phi\left(\frac{\mu_f-K}{\sigma_f}\right) \tag{5.107}$$

where we reuse the notations of sections 2.2.2 and 5.1.1

$$\begin{aligned}
\mu_f &:= \mu^f_{N_f-1} = \widehat{S}^f_{N_f-1} + a\left(\widehat{S}^f_{N_f-1}, t_{N_f-1}\right)h_f := \widehat{S}^f_{N_f-1} + a^f_{N_f-1}h_f\\
\sigma_f &:= \sigma^f_{N_f-1} = b\left(\widehat{S}^f_{N_f-1}, t_{N_f-1}\right)\sqrt{h_f} = b^f_{N_f-1}\sqrt{h_f}
\end{aligned} \tag{5.108}$$

When differentiating this expression, some terms cancel out and we get

$$\begin{aligned}
\frac{\partial \widehat{P}_f}{\partial \widehat{S}^f_{N_f-1}} &= \dot{\mu}_f\Phi\left(\frac{\mu_f-K}{\sigma_f}\right)+\dot{\sigma}_f\phi\left(\frac{\mu_f-K}{\sigma_f}\right)\\
\frac{\partial \widehat{P}_f}{\partial \theta} &= \widetilde{\mu}_f\Phi\left(\frac{\mu_f-K}{\sigma_f}\right)+\widetilde{\sigma}_f\phi\left(\frac{\mu_f-K}{\sigma_f}\right)
\end{aligned} \tag{5.109}$$

with

$$\begin{aligned}
\dot{\mu}_f &:= \frac{\partial \mu_f}{\partial \widehat{S}^f_{N_f-1}} = 1+\dot{a}^f_{N_f-1}h_f\\
\dot{\sigma}_f &:= \frac{\partial \sigma_f}{\partial \widehat{S}^f_{N_f-1}} = \dot{b}^f_{N_f-1}\sqrt{h_f}\\
\widetilde{\mu}_f &:= \frac{\partial \mu_f}{\partial \theta} = \widetilde{a}^f_{N_f-1}\,h_f\\
\widetilde{\sigma}_f &:= \frac{\partial \sigma_f}{\partial \theta} = \widetilde{b}^f_{N_f-1}\sqrt{h_f}
\end{aligned} \tag{5.110}$$

At the coarse level,

$$\widehat{P}_c = \mathbb{E}\left[P\left(\widehat{S}^c_{N_c}\right)\middle|\widehat{S}^c_{N_c-1},\Delta W^f_{N_f-2}\right] = \sigma_c\,\phi\left(\frac{\mu_c-K}{\sigma_c}\right)+(\mu_c-K)\,\Phi\left(\frac{\mu_c-K}{\sigma_c}\right) \tag{5.111}$$

with the usual notations

$$\begin{aligned}
\mu_c &:= \mu^c_{N_c-1} = \widehat{S}^c_{N_c-1} + a\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)\,h_c + b\left(\widehat{S}^c_{N_c-1}, t^c_{N_c-1}\right)\Delta W^f_{N_f-2}\\
&:= \widehat{S}^c_{N_c-1} + a^c_{N_c-1}\,h_c + b^c_{N_c-1}\Delta W^f_{N_f-2}\\
\sigma_c &:= \sigma^c_{N_c-1} = b\left(\widehat{S}_{N_c-1}, t^c_{N-1}\right)\,\sqrt{\frac{h_c}{2}}\\
&:= b^c_{N_c-1}\,\sqrt{\frac{h_c}{2}}
\end{aligned}$$

and differentiation gives

$$
\begin{aligned}
\frac{\partial \widehat{P}_c}{\partial \widehat{S}^c_{N_c-1}} &= \dot{\mu}_c \Phi\left(\frac{\mu_c - K}{\sigma_c}\right) + \dot{\sigma}_c \phi\left(\frac{\mu_c - K}{\sigma_c}\right) \\
\frac{\partial \widehat{P}_c}{\partial \theta} &= \widetilde{\mu}_c \Phi\left(\frac{\mu_c - K}{\sigma_c}\right) + \widetilde{\sigma}_c \phi\left(\frac{\mu_c - K}{\sigma_c}\right)
\end{aligned}
\tag{5.112}
$$

with

$$
\begin{aligned}
\dot{\mu}_c &= \frac{\partial \mu_c}{\partial \widehat{S}^c_{N_c-1}} = 1 + \dot{a}^f_{N_f-1} h_f = 1 + \dot{a}^c_{N_c-1} h_c + \dot{b}^c_{N_c-1} \Delta W^f_{N_f-2} \\
\dot{\sigma}_c &= \frac{\partial \sigma_c}{\partial \widehat{S}^c_{N_c-1}} = \dot{b}^f_{N_f-1} \sqrt{h_f} = \dot{b}^c_{N_c-1} \sqrt{\frac{h_c}{2}} \\
\widetilde{\mu}_c &= \frac{\partial \mu_c}{\partial \theta} = \widetilde{a}^f_{N_f-1}\, h_f = \widetilde{a}^c_{N_c-1} h_c + \widetilde{b}^c_{N_c-1} \Delta W^f_{N_f-2} \\
\widetilde{\sigma}_c &= \frac{\partial \sigma_c}{\partial \theta} = \widetilde{b}^f_{N_f-1} \sqrt{h_f} = \widetilde{b}^c_{N_c-1} \sqrt{\frac{h_c}{2}}
\end{aligned}
\tag{5.113}
$$

5.2.2 Order of convergence β

As in section 5.1.2, to analyse the variance of the multilevel estimator, we study the expectation of its square, that is, we study

$$
\mathbb{E}\left[\left(\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta}\right)^2\right] \tag{5.114}
$$

with

$$
\frac{\partial \widehat{P}_f}{\partial \theta} - \frac{\partial \widehat{P}_c}{\partial \theta} = \Delta\left(\frac{\partial \widehat{S}_{N-1}}{\partial \theta} \frac{\partial \widehat{P}}{\partial S_{N-1}} + \frac{\partial \widehat{P}}{\partial \theta}\right) \tag{5.115}
$$

and a reasoning similar to the one found in section 5.1.2 proves that extreme paths have a negligible contribution to the total variance.

From the results of section 5.2.1, we derive

$$
\begin{aligned}
\Delta\left(\frac{\partial \widehat{S}_{N-1}}{\partial \theta} \frac{\partial \widehat{P}}{\partial S_{N-1}} + \frac{\partial \widehat{P}}{\partial \theta}\right) &= \Delta\left(\left(\delta_{N-1}\dot{\sigma} + \widetilde{\sigma}\right) \phi\left(\frac{\mu - K}{\sigma}\right)\right) \\
&\quad + \Delta\left(\left(\delta_{N-1}\dot{\mu} + \widetilde{\mu}\right) \Phi\left(\frac{\mu - K}{\sigma}\right)\right)
\end{aligned}
\tag{5.116}
$$

where as before $\delta_{N-1} = \dfrac{\partial \widehat{S}_{N-1}}{\partial \theta}$.

Reusing the notations introduced before,

$$
\begin{aligned}
\Delta\left(\frac{\partial \widehat{S}_{N-1}}{\partial \theta}\frac{\partial \widehat{P}}{\partial S_{N-1}}+\frac{\partial \widehat{P}}{\partial \theta}\right) &= \Delta\left(\sigma^2 \dot{L}\phi\left(ML\right)\right)+\Delta\left(\dot{M}\Phi\left(ML\right)\right) \\
&= \Delta\sigma^2\ \dot{L}_f\ \phi\left(M_f L_f\right) \\
&\quad + \sigma_c^2\ \Delta\dot{L}\ \phi\left(M_f L_f\right) \\
&\quad + \sigma_c^2\ \dot{L}_c\ \Delta\phi\left(ML\right) \\
&\quad + \Delta\dot{M}\ \Phi\left(M_f L_f\right) \\
&\quad + \dot{M}_c\ \Delta\Phi\left(ML\right)
\end{aligned}
\tag{5.117}
$$

We study the expectation of the square of this difference, whose order of convergence is known thanks to Hölder's inequality to be the same as that of the sum of the squares of the terms appearing in equation (5.117). As before we decompose non-extreme paths E^c into D and D^c, the paths whose final value are "close to" and "far from" the strike K. From the computations of section 5.1.2, for all $p > 0$ and any $\gamma > 0$, we get

$$
\begin{aligned}
\sigma_c^2 &= O\left(h_f^{1-2\gamma}\right) \\
\Delta\sigma^2 &= O\left(h_f^{1-3\gamma}\right) \\
\dot{L}_f &= O\left(h_f^{-1/2-2\gamma}\right) \\
\dot{L}_c &= O\left(h_f^{-1/2-2\gamma}\right) \\
\Delta\dot{L} &= O\left(h_f^{-5\gamma}\right) \\
\Delta\dot{M} &= O\left(h_f^{1-5\gamma}\right) \\
\dot{M}_c &= O\left(h_f^{-\gamma}\right) \\
\phi\left(M_f L_f\right) &= O\left(h_f^{p}\right) && \text{on } D^c \\
\phi\left(M_f L_f\right) &= O\left(1\right) && \text{on } D \\
\Delta\phi\left(ML\right) &= O\left(h_f^{p}\right) && \text{on } D^c \\
\Delta\phi\left(ML\right) &= O\left(h_f^{1/2-5\gamma}\right) && \text{on } D \\
\Phi\left(M_f L_f\right) &= O\left(1\right) \\
\Delta\Phi\left(ML\right) &= O\left(h_f^{p}\right) && \text{on } D^c \\
\Delta\Phi\left(ML\right) &= O\left(h_f^{1/2-5\gamma}\right) && \text{on } D
\end{aligned}
\tag{5.118}
$$

Hence

$$
\begin{aligned}
\Delta\dot{M}\Phi\left(M_f L_f\right) &= O\left(h_f^{1-5\gamma}\right) & & \\
\Delta\sigma^2\dot{L}_f\phi\left(M_f L_f\right) &= O\left(h_f^{p}\right) & &\text{on } D^c \\
\Delta\sigma^2\dot{L}_f\phi\left(M_f L_f\right) &= O\left(h_f^{1/2-5\gamma}\right) & &\text{on } D \\
\sigma_c^2\Delta\dot{L}\phi\left(M_f L_f\right) &= O\left(h_f^{p}\right) & &\text{on } D^c \\
\sigma_c^2\Delta\dot{L}\phi\left(M_f L_f\right) &= O\left(h_f^{1-7\gamma}\right) & &\text{on } D \\
\sigma_c^2\dot{L}_c\Delta\phi\left(ML\right) &= O\left(h_f^{p}\right) & &\text{on } D^c \\
\sigma_c^2\dot{L}_c\Delta\phi\left(ML\right) &= O\left(h_f^{1-9\gamma}\right) & &\text{on } D \\
\dot{M}_c\Delta\Phi\left(ML\right) &= O\left(h_f^{p}\right) & &\text{on } D^c \\
\dot{M}_c\Delta\Phi\left(ML\right) &= O\left(h_f^{1/2-6\gamma}\right) & &\text{on } D
\end{aligned}
\tag{5.119}
$$

The third and last terms are clearly the limiting ones.

The law of total expectation finally gives

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)^2\right] &= \mathbb{P}\left(D\right)\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)^2\middle| D\right]+\mathbb{P}\left(D^c\right)\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)^2\middle| D^c\right] \\
&= O\left(1\right)O\left(h_f^{2-10\gamma}\right) \\
&\quad + O\left(h_f^{1/2-3\gamma}\right)\left[O\left(h_f^{2-10\gamma}\right)+O\left(h_f^{1-10\gamma}\right)\right. \\
&\qquad\qquad \left.+O\left(h_f^{2-14\gamma}\right)+O\left(h_f^{2-18\gamma}\right)+O\left(h_f^{1-12\gamma}\right)\right] \\
&= O\left(h_f^{3/2-15\gamma}\right)
\end{aligned}
\tag{5.120}
$$

This means that the order of convergence for the European Call using Conditional Expectations is $\beta = \dfrac{3}{2} - 15\gamma$ for any $\gamma > 0$ (and $\gamma < 1/2$).

5.2.3 Order of convergence α

Very similarly, we write

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right] &= \mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)\middle| D\right]\mathbb{P}\left(D\right) \\
&\quad +\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)\middle| D^c\right]\mathbb{P}\left(D^c\right) \\
&\quad +\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)\middle| E\right]\mathbb{P}\left(E\right)
\end{aligned}
\tag{5.121}
$$

We once again show that extreme paths have no significant contribution to this value. Therefore,

$$\mathbb{E}\left[\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right]=O\left(\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)\middle| D\right]\mathbb{P}\left(D\right)\right.$$
$$\left.+\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)\middle| D^c\right]\mathbb{P}\left(D^c\right)\right) \tag{5.122}$$

and finally, using previous results,

$$\begin{aligned}\mathbb{E}\left[\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right]&=O\left(\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)\middle| D\right]\mathbb{P}\left(D\right)\right.\\&\qquad\left.+\mathbb{E}\left[\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)\middle| D^c\right]\mathbb{P}\left(D^c\right)\right)\\&=O\left(1\right)O\left(h_f^{1-5\gamma}\right)\\&\quad+O\left(h_f^{1/2-3\gamma}\right)\left[O\left(h_f^{1-5\gamma}\right)+O\left(h_f^{1/2-5\gamma}\right)\right.\\&\qquad\left.+O\left(h_f^{1-7\gamma}\right)+O\left(h_f^{1-9\gamma}\right)+O\left(h_f^{1/2-6\gamma}\right)\right]\\&=O\left(h_f^{1-9\gamma}\right)\end{aligned} \tag{5.123}$$

This means that the order of weak convergence for the European Call using Conditional Expectations is $\alpha = 1 - 9\gamma$ for any $\gamma > 0$ (and $\gamma < 1/2$).

5.3 Conclusion

The Conditional Expectations technique enables us to perform pathwise sensitivities computations with non-Lipschitz payoffs. For the digital call it yields the following convergence rates: $\beta = \frac{1}{2} - 9\gamma$ and $\alpha = \frac{1}{2} - 6\gamma$ for any $\gamma > 0$ (and $\gamma < 1/2$).

For Lipschitz yet non-smooth payoffs it is also useful: it offers improved convergence rates over simple Pathwise Sensitivities as described in chapter 4. For the European call those are: $\beta = \frac{3}{2} - 15\gamma$ and $\alpha = 1 - 9\gamma$ for any $\gamma > 0$ (and $\gamma < 1/2$).

We can summarise the results of this chapter as follows:

Theorem 5.3.1. *We consider an asset S_t on the time interval $[0, T]$ and a European option with a payoff $P(S_T)$. We assume that S_t follows an Ito process as described by equation (1.2), that the coefficients of the diffusion $a(S, t)$ and $b(S, t)$ satisfy conditions $A1$ to $A4$ of theorem 3.4.3 and that there exists a constant $\epsilon > 0$ such that $b(S, t) \geq \epsilon$.*

Multilevel pathwise sensitivities can be used jointly with the so-called "conditional expectation technique" to construct the estimators of digital options' Greeks described in section 2.3.1. Those have an accuracy $O(\epsilon)$ at a cost $O(\epsilon^{-3})$.

This multilevel technique can also be applied to construct the estimators of Greeks of options with Lipschitz yet non-smooth payoffs. In the case of the European call, our estimators (see section 2.2.2) have an accuracy $O(\epsilon)$ at a cost $O(\epsilon^{-2})$.

Proof. See above. □

Chapter 6

Analysis of Asian call options

Here we analyse the efficiency of the multilevel Monte Carlo technique for the computation of Greeks of Asian options. As in previous chapters, we do this by using the results of chapter 3 to obtain analytical bounds on the coefficients α and β of theorem 1.2.1 in this setting.

We refer to section 2.4 for the definition of the Asian call's payoff, its multilevel estimators and the use of pathwise sensitivities to compute its Greeks.

6.1 Analysis

For the analysis of Asian options, we use index convention (2.103), i.e. we use indexes based on the fine discretisation for both the fine and coarse discretisations, that is for any quantity $\widehat{X}$ defined at the fine and coarse level,

$$\begin{aligned}\widehat{X}^f_n &:= \widehat{X}^f\left(t^f_n\right) = \widehat{X}^f\left(nh_f\right)\\ \widehat{X}^c_n &:= \widehat{X}^c\left(t^f_n\right) = \widehat{X}^c\left(nh_f\right)\end{aligned} \tag{6.1}$$

We also reuse the notation of equation (2.88), i.e. we define

$$\overline{X} = \frac{1}{T}\int_0^T X_t \mathrm{d}t \tag{6.2}$$

for any quantity X_t.

6.1.1 Order of convergence β

We analyse the convergence speed of $\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial\theta} - \frac{\partial \widehat{P}_{l-1}}{\partial\theta}\right)$ and write as before that

$$\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial\theta} - \frac{\partial \widehat{P}_{l-1}}{\partial\theta}\right) \leq 2\left(\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial\theta} - \frac{\partial P}{\partial\theta}\right) + \mathbb{V}\left(\frac{\partial P}{\partial\theta} - \frac{\partial \widehat{P}_{l-1}}{\partial\theta}\right)\right) \tag{6.3}$$

which shows it is sufficient to study $\mathbb{E}\left[\left(\frac{\partial P}{\partial \theta}-\frac{\partial \widehat{P}}{\partial \theta}\right)^2\right]$ for $\frac{\partial \widehat{P}}{\partial \theta}$ resulting either from the fine or from the coarse discretisation of the path. We note that the fine and coarse estimators corresponding to a given level of discretisation are here identical and do not requisite a separate analysis.

Redefinition of extreme paths

Reusing the notation of chapters 2 and 3, we write $U_t=\left(S_t, \frac{\partial S_t}{\partial \theta}\right)$ and $\widehat{U}_{KP,t}, \widehat{U}_{BB,t}, \widehat{U}_t$ the approximations based on the Kloeden-Platen, the Brownian Bridge and the piecewise interpolants respectively.

We first prove that $\mathbb{E}\left[\left\|\overline{U}-\overline{\widehat{U}}\right\|^2\right]=O\left(h^2\right)$. We write

$$U_t-\widehat{U}_t=U_t-\widehat{U}_{KP,t}+\widehat{U}_{KP,t}-\widehat{U}_{BB,t}+\widehat{U}_{BB,t}-\widehat{U}_t \tag{6.4}$$

and therefore

$$\overline{U}-\overline{\widehat{U}}=\overline{U}-\overline{\widehat{U}}_{KP}+\overline{\widehat{U}}_{KP}-\overline{\widehat{U}}_{BB}+\overline{\widehat{U}}_{BB}-\overline{\widehat{U}} \tag{6.5}$$

Then,

$$\left\|\overline{U}-\overline{\widehat{U}}\right\|^2 \leq 3\left[\left\|\overline{U}-\overline{\widehat{U}}_{KP}\right\|^2+\left\|\overline{\widehat{U}}_{KP}-\overline{\widehat{U}}_{BB}\right\|^2+\left\|\overline{\widehat{U}}_{BB}-\overline{\widehat{U}}\right\|^2\right] \tag{6.6}$$

and

$$\begin{aligned}\mathbb{E}\left[\left\|\overline{U}-\overline{\widehat{U}}\right\|^2\right] \leq 3\Bigg[&\mathbb{E}\left[\left\|\overline{U}-\overline{\widehat{U}}_{KP}\right\|^2\right]\\ &+\mathbb{E}\left[\left\|\overline{\widehat{U}}_{KP}-\overline{\widehat{U}}_{BB}\right\|^2\right]+\mathbb{E}\left[\left\|\overline{\widehat{U}}_{BB}-\overline{\widehat{U}}\right\|^2\right]\Bigg]\end{aligned} \tag{6.7}$$

From theorem 3.4.3, we get that for each $m>0$, there exists a constant C_m such that

$$\mathbb{E}\left(\sup_{0<t<T}\left\|U_t-\widehat{U}_{KP}(t)\right\|^m\right)<C_m\, h^m \tag{6.8}$$

and writing

$$\left\|\overline{U}-\overline{\widehat{U}}_{KP}\right\|^2=\left\|\frac{1}{T}\int_0^T\left(U_t-\widehat{U}_{KP,t}\right)\mathrm{d}t\right\|^2 \leq \sup_{0<t<T}\left\|U_t-\widehat{U}_{KP}(t)\right\|^2 \tag{6.9}$$

we get $\mathbb{E}\left[\left\|\overline{U}-\overline{\widehat{U}}_{KP}\right\|^2\right]=O\left(h^2\right)$.

Lemma (3.4.4) gives

$$\mathbb{E}\left(\left\|\int_0^T \left(\widehat{U}_{KP}(t) - \widehat{U}_{BB}(t)\right) \mathrm{d}t\right\|^2\right) = O\left(h^3\right) \tag{6.10}$$

that is, $\mathbb{E}\left[\left\|\overline{\widehat{U}}_{KP} - \overline{\widehat{U}}_{BB}\right\|^2\right] = O\left(h^3\right)$.

For the analysis of $\mathbb{E}\left[\left\|\overline{\widehat{U}}_{BB} - \overline{\widehat{U}}\right\|^2\right]$, we let $N = T/h$ and note that using lemma (3.4.4) and its notations, we have

$$\frac{1}{T}\int_0^T \left(\widehat{U}_{BB}(t) - \widehat{U}(t)\right) \mathrm{d}t = \frac{1}{T}\sum_{n=0}^{N-1} b_n I_n \tag{6.11}$$

where I_n are i.i.d. $\mathcal{N}\left(0, h^3/12\right)$ variables independent from the variables b_n, which leads to

$$\begin{aligned}\mathbb{E}\left[\left\|\overline{\widehat{U}}_{BB} - \overline{\widehat{U}}\right\|^2\right] &= \frac{h^3}{12\,T^2}\sum_{n=0}^{N-1}\mathbb{E}\left(\|b_n\|^2\right)\\ &= O\left(h^2\right)\end{aligned} \tag{6.12}$$

where the second equality comes from the fact that the results of section 3.4 together with the linear growth properties of $b\,(U,t)$ mean that $\mathbb{E}\left[\max_n \|b_n\|^2\right]$ is bounded.

Putting back those results into equation (6.6), we finally obtain $\mathbb{E}\left[\left\|\overline{U} - \overline{\widehat{U}}\right\|^2\right] = O\left(h^2\right)$.

We can extend this result to prove that $\mathbb{E}\left[\left\|\overline{U} - \overline{\widehat{U}}\right\|^p\right] = O\left(h^{p-\gamma}\right)$ for any $p > 2$ and for $\gamma > 0$ as small as we want.

We again use decomposition (6.5) and study the behaviours of $\mathbb{E}\left[\left\|\overline{U} - \overline{\widehat{U}}_{KP}\right\|^p\right]$, $\mathbb{E}\left[\left\|\overline{\widehat{U}}_{KP} - \overline{\widehat{U}}_{BB}\right\|^p\right]$ and $\mathbb{E}\left[\left\|\overline{\widehat{U}}_{BB} - \overline{\widehat{U}}\right\|^p\right]$.

As before we use theorem 3.4.3 and immediately obtain $\mathbb{E}\left[\left\|\overline{U} - \overline{\widehat{U}}_{KP}\right\|^p\right] = O\left(h^p\right)$.

Lemma (3.4.4) gives

$$\mathbb{E}\left(\sup_{t\in[0,T]}\left\|\widehat{U}_{KP}\left(t\right) - \widehat{U}_{BB}\left(t\right)\right\|^p\right) = O\left(\left(h\log h\right)^p\right) \tag{6.13}$$

We write

$$\begin{aligned}\left\|\overline{\widehat{U}}_{KP} - \overline{\widehat{U}}_{BB}\right\|^p &\leq \left\|\frac{1}{T}\int_0^T \left(\widehat{U}_{KP}(t) - \widehat{U}_{BB}(t)\right) \mathrm{d}t\right\|^p\\ &\leq \sup_{t\in[0,T]}\left\|\widehat{U}_{KP}\left(t\right) - \widehat{U}_{BB}\left(t\right)\right\|^p\end{aligned} \tag{6.14}$$

and conclude that $\mathbb{E}\left[\left\|\overline{\widehat{U}}_{KP}-\overline{\widehat{U}}_{BB}\right\|^{p}\right]=O\left((h\log h)^{p}\right)=O\left(h^{p-\gamma}\right)$ for any $\gamma>0$.

To study $\mathbb{E}\left[\left\|\overline{\widehat{U}}_{BB}-\overline{\widehat{U}}\right\|^{p}\right]$, we consider the discrete martingale

$$M_n=\frac{1}{T}\int\limits_0^{t_n}\left(\widehat{U}_{BB}(t)-\widehat{U}(t)\right)\mathrm{d}t \tag{6.15}$$

and note that $\left(\overline{\widehat{U}}_{BB}-\overline{\widehat{U}}\right)=M_N$. As in (6.11), we write

$$\begin{aligned} M_n&=\frac{1}{T}\sum_{k=0}^{n-1}\int\limits_{t_k}^{t_{k+1}}\left(\widehat{U}_{BB}(t)-\widehat{U}(t)\right)\mathrm{d}t\\ &=\frac{1}{T}\sum_{k=0}^{n-1}b_kI_k \end{aligned} \tag{6.16}$$

where for $k=0,\dots,N-1$, $I_k\sim\mathcal{N}\left(0,\dfrac{h^3}{12}\right)$ is independent of b_k. We apply the discrete Burkholder-Davis-Gundy inequality (see [Kal02]) and obtain

$$\begin{aligned} \mathbb{E}\left[\left\|\overline{\widehat{U}}_{BB}-\overline{\widehat{U}}\right\|^{p}\right]&\leq\mathbb{E}\left[\sup_{n=0,\dots,N}\|M_n\|^p\right]\\ &=O\left(\mathbb{E}\left[\left(\sum_{n=0}^{N-1}\|M_{n+1}-M_n\|^2\right)^{p/2}\right]\right)\\ &=O\left(\mathbb{E}\left[\left(\sum_{n=0}^{N-1}\|b_nI_n\|^2\right)^{p/2}\right]\right)\\ &=O\left(\mathbb{E}\left[\max_{n=0,\dots,N-1}\|b_n\|^p\left(\sum_{n=0}^{N-1}I_n^2\right)^{p/2}\right]\right)\\ &=O\left(\mathbb{E}\left[\max_{n=0,\dots,N-1}\|b_n\|^p\right]\mathbb{E}\left[\left(\sum_{n=0}^{N-1}I_n^2\right)^{p/2}\right]\right) \end{aligned} \tag{6.17}$$

using as before the boundedness of the moments of b_n, then Jensen's inequality and finally writing $I_n=\left(h^3/12\right)^{1/2}J_n$ (where $J_n\sim\mathcal{N}(0,1)$), we obtain

$$\begin{aligned} \mathbb{E}\left[\left\|\overline{\widehat{U}}_{BB}-\overline{\widehat{U}}\right\|^{p}\right]&=O\left(\mathbb{E}\left[\left(\sum_{n=0}^{N-1}I_n^2\right)^{p/2}\right]\right)\\ &=O\left(N^{p/2-1}\mathbb{E}\left[\sum_{n=0}^{N-1}I_n^p\right]\right)\\ &=O\left(N^{p/2}\mathbb{E}\left[I_0^p\right]\right)\\ &=O\left(N^{p/2}\left(h^3/12\right)^{p/2}\mathbb{E}\left[J_0^p\right]\right) \end{aligned} \tag{6.18}$$

and using the boundedness of moments of the unit standard distribution,

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\overline{U}}_{BB}-\widehat{\overline{U}}\right\|^p\right] &= O\left(N^{p/2}\left(h^3/12\right)^{p/2}\right)\\ &= O\left(h^p\right)\end{aligned} \tag{6.19}$$

We have thus proved that $\mathbb{E}\left[\left\|\overline{U}-\widehat{\overline{U}}\right\|^p\right]=O\left(h^{p-\gamma}\right)$ for any $p>2$ and for $\gamma>0$ as small as we want. Therefore we have $\mathbb{E}\left[\left\|\frac{\overline{U}-\widehat{\overline{U}}}{h}\right\|^p\right]=O\left(h^{-\gamma}\right)$.

As in lemma 3.4.5, we can now write that for all $\delta>0$, $m>0$,

$$\begin{aligned}\mathbb{P}\left(\left\|\overline{U}-\widehat{\overline{U}}\right\|\geq h^{1-\delta}\right) &= \mathbb{P}\left(\left\|\frac{\overline{U}-\widehat{\overline{U}}}{h}\right\|\geq h^{-\delta}\right)\\ &= \mathbb{P}\left(\left\|\frac{\overline{U}-\widehat{\overline{U}}}{h}\right\|^m\geq h^{-m\delta}\right)\\ &\leq h^{m\delta}\mathbb{E}\left[\left\|\frac{\overline{U}-\widehat{\overline{U}}}{h}\right\|^m\right]\\ &= O\left(h^{m\delta-\gamma}\right)\end{aligned} \tag{6.20}$$

and by picking m sufficiently large this proves that for any $\delta>0$, $\mathbb{P}\left(\left\|\overline{U}-\widehat{\overline{U}}\right\|\geq h^{1-\delta}\right)=o\left(h^p\right)$ for all $p>0$.

We can therefore extend lemma 3.4.8 by redefining extreme paths E as the paths satsifying any of the extreme conditions of lemma 3.4.8 or satisfying $\left\|\overline{U}-\widehat{\overline{U}}\right\|>h^{1-\gamma}$. Those still have a likelihood $o\left(h^p\right)$ for all $p>0$. From there, the analysis of Asian options is similar to the analysis of European options in chapter 4.

Contribution of extreme paths

Letting $\mathbb{A}=\frac{\partial\overline{S}}{\partial\theta}\left(\frac{\partial P}{\partial\overline{S}}-\frac{\partial\widehat{P}}{\partial\widehat{\overline{S}}}\right)$ and $\mathbb{B}=\frac{\partial\widehat{P}}{\partial\widehat{\overline{S}}}\left(\frac{\partial\overline{S}}{\partial\theta}-\frac{\partial\widehat{\overline{S}}}{\partial\theta}\right)$, we can write that

$$\mathbb{V}\left(\frac{\partial P}{\partial\theta}-\frac{\partial\widehat{P}}{\partial\theta}\right)\leq\mathbb{E}\left[\left(\frac{\partial P}{\partial\theta}-\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right]=O\left(\mathbb{E}\left(\mathbb{A}^2\right)+\mathbb{E}\left(\mathbb{B}^2\right)\right) \tag{6.21}$$

and using the usual decomposition between extreme and non-extreme paths (with the above redefinition of E and E^c), we write

$$\mathbb{E}\left[\left(\frac{\partial P}{\partial\theta}-\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right]=O\left(\mathbb{E}\left(\mathbb{A}^2\mathbf{1}_E\right)+\mathbb{E}\left(\mathbb{A}^2\mathbf{1}_{E^c}\right)+\mathbb{E}\left(\mathbb{B}^2\mathbf{1}_E\right)+\mathbb{E}\left(\mathbb{B}^2\mathbf{1}_{E^c}\right)\right) \tag{6.22}$$

As before we show that extreme paths have a negligible contribution to the global variance. Indeed, using Hölder's inequality, we have

$$\mathbb{E}\left(\mathbb{A}^2\mathbf{1}_E\right) \leq \sqrt{\mathbb{E}\left(\mathbf{1}_E\right)}\left(\mathbb{E}\left[\frac{\partial\overline{S}}{\partial\theta}^8\right]\right)^{1/4}\left(\mathbb{E}\left[\left(\frac{\partial P}{\partial\overline{S}}-\frac{\partial\widehat{P}}{\partial\widehat{\overline{S}}}\right)^8\right]\right)^{1/4} \tag{6.23}$$

Lemma 3.4.8 shows that $\sqrt{\mathbb{E}\left(\mathbf{1}_E\right)} = o\left(h^p\right)$ for all $p > 0$.

To show that $\mathbb{E}\left[\frac{\partial\overline{S}}{\partial\theta}^8\right]$ is finite, we note that

$$\left|\frac{\partial\overline{S}}{\partial\theta}\right| \leq \sup_t\left|\frac{\partial S_t}{\partial\theta}\right| \tag{6.24}$$

Then applying theorem 3.4.3 to $U_t = \left(S_t, \frac{\partial S_t}{\partial\theta}\right)$, we get that for all $k \geq 0$,

$$\mathbb{E}\left[\sup_t\left|\frac{\partial S_t}{\partial\theta}\right|^k\right] < \infty \tag{6.25}$$

therefore we get the result we want: for any $k \geq 0$ (in particular $k = 8$),

$$\mathbb{E}\left[\frac{\partial\overline{S}}{\partial\theta}^k\right] < \infty \tag{6.26}$$

Exploiting the fact that $0 \leq \frac{\partial P}{\partial\overline{S}} \leq 1$ and $0 \leq \frac{\partial\widehat{P}}{\partial\widehat{\overline{S}}} \leq 1$ for the Asian call, we get that

$$\mathbb{E}\left[\left(\frac{\partial P}{\partial\overline{S}}-\frac{\partial\widehat{P}}{\partial\widehat{\overline{S}}}\right)^8\right] \leq 1 \tag{6.27}$$

Plugging the previous results back into the equation gives that for all $p > 0$, $\mathbb{E}\left(\mathbf{1}_E\mathbb{A}^2\right) = o\left(h^p\right)O\left(1\right)O\left(1\right) = o\left(h^p\right)$.

Similarly, we can prove that

$$\begin{aligned}\mathbb{E}\left(\mathbb{B}^2\mathbf{1}_E\right) &\leq \sqrt{\mathbb{E}\left(\mathbf{1}_E\right)}\left(\mathbb{E}\left[\frac{\partial\widehat{P}}{\partial\widehat{\overline{S}}}^8\right]\right)^{1/4}\left(\mathbb{E}\left[\left(\frac{\partial\overline{S}}{\partial\theta}-\frac{\partial\widehat{\overline{S}}}{\partial\theta}\right)^8\right]\right)^{1/4}\\ &= o\left(h^p\right)O\left(1\right)O\left(1\right)\end{aligned} \tag{6.28}$$

Indeed, lemma 3.4.8 gives that $\sqrt{\mathbb{E}\left(\mathbf{1}_E\right)} = o\left(h^p\right)$ for all $p > 0$. Using the fact that $\frac{\partial\widehat{P}}{\partial\widehat{\overline{S}}} \leq 1$, we obtain that $\mathbb{E}\left[\frac{\partial\widehat{P}}{\partial\widehat{\overline{S}}}^8\right]$ is finite. Finally, to prove that $\mathbb{E}\left[\left(\frac{\partial\overline{S}}{\partial\theta}-\frac{\partial\widehat{\overline{S}}}{\partial\theta}\right)^8\right]$

is finite, we write

$$\mathbb{E}\left[\left(\frac{\partial\overline{S}}{\partial\theta}-\frac{\partial\overline{\widehat{S}}}{\partial\theta}\right)^{8}\right]=\mathbb{E}\left[\mathscr{P}\left(\frac{\partial\overline{S}}{\partial\theta},\frac{\partial\overline{\widehat{S}}}{\partial\theta}\right)\right] \tag{6.29}$$

where $\mathscr{P}$ is a polynomial of degree 8. Then, using Hölder's inequality, it can be bounded by a function of $\mathbb{E}\left[\frac{\partial\overline{S}}{\partial\theta}^{k}\right]$, $\mathbb{E}\left[\frac{\partial\overline{\widehat{S}}}{\partial\theta}^{k}\right]$, which, as explained before, are finite. We then conclude that $\mathbb{E}\left[\left(\frac{\partial\overline{S}}{\partial\theta}-\frac{\partial\overline{\widehat{S}}}{\partial\theta}\right)^{8}\right]<\infty$.

Plugging the previous results back into the previous equation gives that for all $p>0$,

$$\mathbb{E}\left(\mathbf{1}_{E}\mathbb{B}^{2}\right)=o\left(h^{p}\right) \tag{6.30}$$

This means that the contributions of extreme paths are negligible. Therefore we have

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^{2}\right)&=O\left(\mathbb{E}\left(\mathbb{A}^{2}\right)\right)\\ \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^{2}\right)&=O\left(\mathbb{E}\left(\mathbb{B}^{2}\right)\right)\end{aligned} \tag{6.31}$$

$$\mathbb{V}\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)\leq\mathbb{E}\left(\mathbb{A}^{2}\right)+\mathbb{E}\left(\mathbb{B}^{2}\right)=O\left(\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^{2}\right)+\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^{2}\right)\right) \tag{6.32}$$

Contribution of discontinuities

We have now established that only paths of E^c contribute to the variance. Therefore we now restrict our analysis to non-extreme paths.

We define D the set of non-extreme paths such that $\left|\overline{\widehat{S}}-K\right|<h^{1-\delta}$ for some $\delta>0$ (i.e. close to the strike). D^c is then the set of non-extreme paths that do not verify this condition. We now write

$$\begin{aligned}\mathbb{V}\left(\frac{\partial\widehat{P}_f}{\partial\theta}-\frac{\partial\widehat{P}_c}{\partial\theta}\right)&=O\left(\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^{2}\right)+\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^{2}\right)\right)\\ &=O\left(\mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{A}^{2}\right)+\mathbb{E}\left(\mathbf{1}_{D}\mathbb{A}^{2}\right)\right)+O\left(\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^{2}\right)\right)\end{aligned} \tag{6.33}$$

Under the assumptions of section 3.3.3, the SDE is elliptic and the increments $\left(\widehat{S}_{n+1}-\widehat{S}_n\right)_{n=0,\ldots,N-1}$ have smooth density functions $p_{\widehat{S}_{n+1}-\widehat{S}_n}$. We can then prove that the probability density function of $\overline{\widehat{S}}$ is also smooth. Indeeed $\overline{\widehat{S}}$ can be seen as a weighted sum of the independent increments $\left(\widehat{S}_{n+1} - \widehat{S}_n\right)$.

$$\begin{aligned}\overline{\widehat{S}} &= \frac{1}{2N}\left(\widehat{S}_0 + \widehat{S}_N\right) + \frac{1}{N}\sum_{n=1}^{N-1}\widehat{S}_n \\ &= S_0 + \frac{1}{N}\sum_{n=1}^{N}\left(\left(N - n + \frac{1}{2}\right)\left(\widehat{S}_n - \widehat{S}_{n-1}\right)\right)\end{aligned} \tag{6.34}$$

The probability density function for $\overline{\widehat{S}}$ is then derived from the convolution product

$$p_{\overline{\widehat{S}}}(x) = \left(p_{\left(N-N+\frac{1}{2}\right)\left(\widehat{S}_N - \widehat{S}_{N-1}\right)} * \cdots * p_{\left(N-1+\frac{1}{2}\right)\left(\widehat{S}_1 - \widehat{S}_0\right)}\right)(x) \tag{6.35}$$

therefore $\overline{\widehat{S}}$ also has a smooth probability density function. As in chapter 4, we can therefore conclude that the paths of D represent a proportion $O\left(h^{1-\delta}\right)$ of all paths.

Using (6.33), we write

$$\begin{aligned}\mathbb{V}\left(\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\right) &= O\left(\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^2\right)\right) \\ &\quad + O\left(h^{1-\delta}\right) O\left(\max\left(\mathbb{A}^2\mathbf{1}_D\right)\right) + O\left(\max\left(\mathbb{A}^2\mathbf{1}_{D^c}\right)\right)\end{aligned} \tag{6.36}$$

Paths in D and D^c are non extreme and therefore $\left(\frac{\partial\overline{S}}{\partial\theta}\right)^2 \leq h^{-2\gamma}$. Taking $\delta = 2\gamma$, we can write

$$\left|\overline{S} - \overline{\widehat{S}}\right| \leq h^{1-\gamma} \leq h^{1-\delta} \leq \left|\overline{\widehat{S}} - K\right| \tag{6.37}$$

For paths in D^c, there can therefore be no payoff discontinuity between $\overline{\widehat{S}}$ and $\overline{S}$. The payoff being piecewise linear away from the discontinuity, we thus have $\mathbb{A}^2\mathbf{1}_{D^c} = 0$.

For paths in D, we use the fact that $\left|\frac{\partial P}{\partial\overline{S}} - \frac{\partial\widehat{P}}{\partial\overline{\widehat{S}}}\right| \leq 1$ to conclude that $\max\left(\mathbb{A}^2\mathbf{1}_D\right) = O\left(h^{-2\gamma}\right)$.

Then, noting that $\left|\frac{\partial\widehat{P}}{\partial\overline{\widehat{S}}}\right| \leq 1$, we can write

$$\mathbb{E}\left[\frac{\partial\widehat{P}}{\partial\overline{\widehat{S}}}^2\left(\frac{\partial\overline{S}}{\partial\theta} - \frac{\partial\overline{\widehat{S}}}{\partial\theta}\right)^2\mathbf{1}_{E^c}\right] \leq \mathbb{E}\left[\left(\frac{\partial\overline{S}}{\partial\theta} - \frac{\partial\overline{\widehat{S}}}{\partial\theta}\right)^2\right] = O\left(h^2\right) \tag{6.38}$$

where the last equality corresponds to what we proved earlier in this chapter.

Putting all those results together, we finally obtain

$$\begin{aligned}\mathbb{V}\left(\frac{\partial \widehat{P}_f}{\partial \theta}-\frac{\partial \widehat{P}_c}{\partial \theta}\right) &= O\left(\mathbb{E}\left[\mathbb{A}^2+\mathbb{B}^2\right]\right)\\ &= O\left(h^{1-\delta-2\gamma}\right)+O\left(h^2\right)\\ &= O\left(h^{1-\delta-2\gamma}\right)\\ &= O\left(h^{1-4\gamma}\right)\end{aligned} \tag{6.39}$$

that is, $\beta = 1-4\gamma$ for any $\gamma > 0$.

6.1.2 Order of convergence α

Similarly, we write

$$\mathbb{E}\left[\frac{\partial P}{\partial \theta}-\frac{\partial \widehat{P}}{\partial \theta}\right] = \mathbb{E}\left[\mathbb{A}\right]+\mathbb{E}\left[\mathbb{B}\right] \tag{6.40}$$

and using the results of section 6.1.1, we obtain

$$\begin{aligned}\mathbb{E}\left[\frac{\partial P}{\partial \theta}-\frac{\partial \widehat{P}}{\partial \theta}\right] &= \mathbb{E}\left[\mathbb{A}\right]+\mathbb{E}\left[\mathbb{B}\right]\\ &= O\left(\mathbb{E}\left[\mathbb{A}\mathbf{1}_{E^c}\right]\right)+O\left(\mathbb{E}\left[\mathbb{B}\mathbf{1}_{E^c}\right]\right)\\ &= O\left(\mathbb{E}\left[\mathbb{A}\mathbf{1}_{D}\right]\right)+O\left(\mathbb{E}\left[\mathbb{A}\mathbf{1}_{D^c}\right]\right)+O\left(\mathbb{E}\left[\mathbb{B}\mathbf{1}_{E^c}\right]\right)\\ &= O\left(h^{1-\delta-\gamma}\right)+0+O\left(h\right)\\ &= O\left(h^{1-3\gamma}\right)\end{aligned} \tag{6.41}$$

This means that $\alpha = 1-3\gamma$ for any $\gamma > 0$.

6.2 Conclusion

We have proved that for the convergence rates for Asian call options are $\alpha = 1-3\gamma$ and $\beta = 1-4\gamma$ for $\gamma \in]0, 1/2[$ as small as we want.

This corresponds to the convergence rates observed experimentally in section 2.4.2: $\alpha \approx 1$ and $\beta \approx 1$.

We can summarise the results we have proved in this chapter as follows:

Theorem 6.2.1. *Let us consider an asset S_t on the time interval $[0, T]$. We assume that the underlying asset's price S_t follows an Ito process as described by equation* (1.2)*, that the coefficients of the diffusion $a(S, t)$ and $b(S, t)$ satisfy conditions A1 to A4 of theorem 3.4.3 and that there exists a constant $\epsilon > 0$ such that $b(S, t) \geq \epsilon$.*

Our multilevel estimators of the Greeks of Asian call options (see section 2.4) have an accuracy $O(\epsilon)$ at a cost $O\left(\epsilon^{-2} (\log \epsilon)^2\right)$.

Proof. See above. $\square$

Chapter 7

Analysis of Barrier options

In this chapter we use the results of chapter 3 to obtain analytical bounds on the coefficients α and β of theorem 1.2.1 for barrier options. This allows us to determine the efficiency of the multilevel Monte Carlo technique in this setting.

7.1 Setting

We once again use index convention (2.103), i.e. we use indexes based on the fine discretisation for both the fine and coarse discretisations, that is for any quantity $\widehat{X}$ defined at the fine and coarse level,

$$\begin{aligned} \widehat{X}_n^f &:= \widehat{X}^f\left(t_n^f\right) = \widehat{X}^f\left(nh_f\right) \\ \widehat{X}_n^c &:= \widehat{X}^c\left(t_n^f\right) = \widehat{X}^c\left(nh_f\right) \end{aligned} \tag{7.1}$$

We recall from section 2.6.1 that the payoff of the down-and-out barrier call option is of the form

$$P = (S_T - K)^+ \, \mathbf{1} \min_{t\in[0,T]} (S_t) > B \tag{7.2}$$

As explained before, the payoff estimator at the fine level is

$$\widehat{P}^f = (\widehat{S}^f_{N_f} - K)^+ \prod_{n=0}^{N_f-1} \left(1 - p_n^f\right) \tag{7.3}$$

with

$$p_n^f = \exp\left(\frac{-2(\widehat{S}_n^f - B)^+(\widehat{S}_{n+1}^f - B)^+}{(b_n^f)^2 \, h_f}\right)$$

At the coarse level, the payoff estimator is

$$\widehat{P}^c = (\widehat{S}^c_{N_f} - K)^+ \prod_{k=0}^{N_f/2-1} \left((1 - p_{2k}^c)(1 - p_{2k+1}^c)\right) \tag{7.4}$$

with

$$\begin{aligned} p_{2k}^c &= \exp\left(\frac{-2(\widehat{S}_{2k}^c - B)^+ (\widehat{S}_{2k+1}^c - B)^+}{(b_{2k}^c)^2\, h_f}\right) \\ p_{2k+1}^c &= \exp\left(\frac{-2(\widehat{S}_{2k+1}^c - B)^+ (\widehat{S}_{2k+2}^c - B)^+}{(b_{2k}^c)^2\, h_f}\right) \end{aligned} \tag{7.5}$$

Note that we keep the same volatility on the whole coarse interval $[t_{2k}, t_{2k+2}]$: for $k = 0 \dots N_f/2 - 1$, we define $b_{2k+1}^c := b_{2k}^c$ to write (7.5) as

$$p_n^c = \exp\left(\frac{-2(\widehat{S}_n^c - B)^+ (\widehat{S}_{n+1}^c - B)^+}{(b_n^c)^2\, h_f}\right) \tag{7.6}$$

7.2 Analysis

7.2.1 Order of convergence β

As usual, we let $\Delta X = \left(X^f - X^c\right)$ for corresponding quantities X^f and X^c defined at the fine and coarse levels.

We now analyse the convergence speed of $\mathbb{V}\left(\frac{\partial \widehat{P}_l}{\partial \theta} - \frac{\partial \widehat{P}_{l-1}}{\partial \theta}\right) = \mathbb{V}\left(\Delta \frac{\partial \widehat{P}}{\partial \theta}\right)$.

We have

$$\frac{\partial \widehat{P}^f}{\partial \theta} = \mathbf{1}_{\widehat{S}_{N_f}^f > K} \frac{\partial \widehat{S}_{N_f}^f}{\partial \theta} \prod_{n=0}^{N_f-1} \left(1 - p_n^f\right) - \left(\widehat{S}_{N_f}^f - K\right)^+ \sum_{n=0}^{N_f-1} \left[\prod_{k=0, k\neq n}^{N_f-1} \left(1 - p_k^f\right) \frac{\partial p_n^f}{\partial \theta}\right] \tag{7.7}$$

with (writing $\dot{b} := \frac{\partial b}{\partial S}$ and $\widetilde{b} := \frac{\partial b}{\partial \theta}$)

$$\begin{aligned} \frac{\partial p_n^f}{\partial \theta} = \mathbf{1}_{\left(\widehat{S}_n^f, \widehat{S}_{n+1}^f > B\right)} p_n^f \Bigg[\widehat{\delta}_n^f \left(\frac{-2\left(\widehat{S}_{n+1}^f - B\right)}{{b_n^f}^2 h_f} + \frac{4\left(\widehat{S}_n^f - B\right)\left(\widehat{S}_{n+1}^f - B\right) \dot{b_n}^f}{{b_n^f}^3 h_f} \right) \\ + \widehat{\delta}_{n+1}^f \frac{-2\left(\widehat{S}_n^f - B\right)}{{b_n^f}^2 h_f} + \frac{4\left(\widehat{S}_n^f - B\right)\left(\widehat{S}_{n+1}^f - B\right) \widetilde{b_n}^f}{{b_n^f}^3 h_f} \Bigg] \end{aligned} \tag{7.8}$$

and

$$\frac{\partial \widehat{P}^c}{\partial \theta} = \mathbf{1}_{\widehat{S}_{N_f}^c > K} \frac{\partial \widehat{S}_{N_f}^c}{\partial \theta} \prod_{n=0}^{N_f-1} (1 - p_n^c) - \left(\widehat{S}_{N_f}^c - K\right)^+ \sum_{n=0}^{N_f-1} \left[\prod_{k=0, k\neq n}^{N_f-1} (1 - p_k^c) \frac{\partial p_n^c}{\partial \theta}\right] \tag{7.9}$$

with

$$\frac{\partial p_n^c}{\partial \theta} = \mathbf{1}_{\left(\widehat{S}_n^c, \widehat{S}_{n+1}^c > B\right)} p_n^c \left[\widehat{\delta}_n^c \left(\frac{-2\left(\widehat{S}_{n+1}^c - B\right)}{{b_n^c}^2 h_f} + \frac{4\left(\widehat{S}_n^c - B\right)\left(\widehat{S}_{n+1}^c - B\right) \dot{b_n}^c}{{b_n^c}^3 h_f} \right) \right.$$
$$\left. + \widehat{\delta}_{(n+1)}^c \frac{-2\left(\widehat{S}_n^c - B\right)}{{b_n^c}^2 h_f} + \frac{4\left(\widehat{S}_n^c - B\right)\left(\widehat{S}_{n+1}^c - B\right) \widetilde{b_n}^c}{{b_n^c}^3 h_f} \right] \tag{7.10}$$

Then we write

$$\Delta \frac{\partial \widehat{P}}{\partial \theta} = \Delta \left[\mathbf{1}_{\widehat{S}_N > K} \frac{\partial \widehat{S}_N}{\partial \theta} \prod_{n=0}^{N_f - 1} (1 - p_n) \right] - \Delta \left[\left(\widehat{S}_N - K\right)^+ \sum_{n=0}^{N_f-1} \left[\prod_{k=0, k \neq n}^{N_f - 1} (1 - p_k) \frac{\partial p_n}{\partial \theta} \right] \right] \tag{7.11}$$

and define

$$\begin{aligned} \mathbb{A} &= \Delta \left[\mathbf{1}_{\widehat{S}_N > K} \frac{\partial \widehat{S}_N}{\partial \theta} \prod_{n=0}^{N_f - 1} (1 - p_n) \right] \\ \mathbb{B} &= \Delta \left[\left(\widehat{S}_N - K\right)^+ \sum_{n=0}^{N_f-1} \left[\prod_{k=0, k \neq n}^{N_f - 1} (1 - p_k) \frac{\partial p_n}{\partial \theta} \right] \right] \end{aligned} \tag{7.12}$$

that is,

$$\Delta \frac{\partial \widehat{P}}{\partial \theta} = \mathbb{A} + \mathbb{B} \tag{7.13}$$

and as usual we have

$$\mathbb{V}\left(\Delta \frac{\partial \widehat{P}}{\partial \theta}\right) \leq \mathbb{E}\left[\left(\Delta \frac{\partial \widehat{P}}{\partial \theta}\right)^2\right] \sim \mathbb{E}\left(\mathbb{A}^2\right) + \mathbb{E}\left(\mathbb{B}^2\right) \tag{7.14}$$

We define the set E of extreme paths as in section 4.2. It is the set of paths satisfying any of the three conditions of lemma 3.4.8 for a certain $\gamma < 1/2$. As before, the idea is to first show those have a negligible contribution to the variances/expectations considered. This will then enable us to focus only the set E^c of "non-extreme" paths, which we split between the set of paths D for which $\left(\min_n \widehat{S}_n\right)$ is "close" to the barrier and D^c the set of paths for which $\left(\min_n \widehat{S}_n\right)$ is "far" from it (a notion we will more precisely define later). We therefore write as before

$$\Omega = E \sqcup E^c = E \sqcup (D \sqcup D^c) \tag{7.15}$$

Contribution of extreme paths

We first write

$$\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right)=\mathbb{E}\left(\mathbf{1}_{E^c}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right)+\mathbb{E}\left(\mathbf{1}_{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right)\tag{7.16}$$

We now prove that the influence of extreme paths is negligible. Using lemma 3.4.8, we get that for all $p>0$,

$$\mathbb{E}\left(\mathbf{1}_E\right)=O\left(h^p\right)\tag{7.17}$$

We then use Hölder's inequality

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right)&\leq\sqrt{\mathbb{E}\left(\mathbf{1}_E\right)\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^4\right)}\\&=O\left(h^p\right)\sqrt{\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^4\right)}\end{aligned}\tag{7.18}$$

Therefore it is sufficient to prove that $\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^4\right)$ does not explode "too quickly" as $h\to 0$. We write

$$\left|\Delta\frac{\partial\widehat{P}}{\partial\theta}\right|\leq|\,\mathbb{A}\,|+|\,\mathbb{B}\,|\tag{7.19}$$

thus,

$$\begin{aligned}\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^4\right)&\leq\mathbb{E}\left((|\,\mathbb{A}\,|+|\,\mathbb{B}\,|)^4\right)\\&\sim\mathbb{E}\left(|\,\mathbb{A}\,|^4\right)+\mathbb{E}\left(|\,\mathbb{B}\,|^4\right)\end{aligned}\tag{7.20}$$

We thus study $\mathbb{E}\left(|\,\mathbb{A}\,|^4\right)$ and $\mathbb{E}\left(|\,\mathbb{B}\,|^4\right)$.

We have

$$\begin{aligned}|\,\mathbb{A}\,|&\leq\mathbf{1}_{\widehat{S}^f_{N_f}>K}\left|\frac{\partial\widehat{S}^f_{N_f}}{\partial\theta}\right|\prod_{n=0}^{N_f-1}\left(1-p^f_n\right)+\mathbf{1}_{\widehat{S}^c_{N_c}>K}\left|\frac{\partial\widehat{S}^c_{N_c}}{\partial\theta}\right|\prod_{n=0}^{N_f-1}\left(1-p^c_n\right)\\&\leq\left|\frac{\partial\widehat{S}^f_{N_f}}{\partial\theta}\right|+\left|\frac{\partial\widehat{S}^c_{N_c}}{\partial\theta}\right|\end{aligned}\tag{7.21}$$

As in previous chapters, from (3.80) of theorem 3.4.3 we know that for all $k>0$

there is some constant $(C_{\mathbb{A},k})$ s.t.

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{N_f-1}}{\partial\theta}\right)^k\right] < C_{\mathbb{A},k} < \infty \\ \mathbb{E}\left[\left(\frac{\partial \widehat{S}^c_{N_c-1}}{\partial\theta}\right)^k\right] < C_{\mathbb{A},k} < \infty\end{aligned} \tag{7.22}$$

Therefore Hölder's inequality guarantees there exists some constant $C_{\mathbb{A}} < \infty$ such that

$$\mathbb{E}\left(|\,\mathbb{A}\,|^4\right) < C_{\mathbb{A}} \tag{7.23}$$

We also have

$$\begin{aligned}|\,\mathbb{B}\,| \leq \left|\,\widehat{S}^f_{N_f} - K\,\right| \sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}\left(1-p^f_k\right)\left|\,\frac{\partial p^f_n}{\partial\theta}\,\right|\right] \\ + \left|\,\widehat{S}^c_{N_c} - K\,\right| \sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}(1-p^c_k)\left|\,\frac{\partial p^c_n}{\partial\theta}\,\right|\right]\end{aligned} \tag{7.24}$$

Therefore

$$\begin{aligned}|\,\mathbb{B}\,|^4 \leq \mathcal{P}\Bigg(\left(\widehat{S}^f_{N_f} - K\right), \left(\widehat{S}^c_{N_c} - K\right), \\ \sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}\left(1-p^f_k\right)\left|\,\frac{\partial p^f_n}{\partial\theta}\,\right|\right], \\ \sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}(1-p^c_k)\left|\,\frac{\partial p^c_n}{\partial\theta}\,\right|\right]\Bigg)\end{aligned} \tag{7.25}$$

where $\mathcal{P}$ is a polynomial of order 4.

From our previous computations,

$$\begin{aligned}\left|\,\frac{\partial p^f_n}{\partial\theta}\,\right| \leq p^f_n \frac{2}{{b^f_n}^2 h_f}\Bigg|\widehat{\delta}^f_n\left(-\left(\widehat{S}^f_{n+1} - B\right) + \frac{2\left(\widehat{S}^f_n - B\right)\left(\widehat{S}^f_{n+1} - B\right)\dot{b_n}^f}{b^f_n}\right) \\ - \widehat{\delta}^f_{n+1}\left(\widehat{S}^f_n - B\right) + \frac{2\left(\widehat{S}^f_n - B\right)\left(\widehat{S}^f_{n+1} - B\right)\widetilde{b_n}^f}{b^f_n}\Bigg| \\ \left|\,\frac{\partial p^c_n}{\partial\theta}\,\right| \leq p^c_n \frac{2}{{b^c_n}^2 h_f}\Bigg|\widehat{\delta}^c_n\left(-\left(\widehat{S}^c_{n+1} - B\right) + \frac{2\left(\widehat{S}^c_n - B\right)\left(\widehat{S}^c_{(n+1)} - B\right)\dot{b_n}^c}{b^c_n}\right) \\ - \widehat{\delta}^c_{n+1}\left(\widehat{S}^f_n - B\right) + \frac{2\left(\widehat{S}^c_n - B\right)\left(\widehat{S}^c_{n+1} - B\right)\widetilde{b}^c_n}{b^c_n}\Bigg|\end{aligned} \tag{7.26}$$

As before, from (3.80) of theorem 3.4.3: for all $k > 0$, there are constants $(C_{\mathbb{B},k})$, s.t. for $n = 0, \ldots, N_f$,

$$\begin{aligned}
\mathbb{E}\left(\left(\widehat{S}_n^f - K\right)^k\right) &< C_{\mathbb{B},k} < \infty \\
\mathbb{E}\left(\left(\widehat{S}_n^c - K\right)^k\right) &< C_{\mathbb{B},k} < \infty \\
\mathbb{E}\left(\left(\widehat{\delta}_n^f\right)^k\right) = \mathbb{E}\left(\left(\frac{\partial \widehat{S}_n^f}{\partial \theta}\right)^k\right) &< C_{\mathbb{B},k} < \infty \\
\mathbb{E}\left(\left(\widehat{\delta}_n^c\right)^k\right) = \mathbb{E}\left(\left(\frac{\partial \widehat{S}_n^c}{\partial \theta}\right)^k\right) &< C_{\mathbb{B},k} < \infty
\end{aligned} \tag{7.27}$$

Assumption $A2$ of theorem 3.4.3 applied to U_t as before implies that $\dot{b}$ and $\widetilde{b}$ satisfy linear growth conditions. Therefore, for any $k >= 0$ and for $n = 0, \ldots, N_f - 1$,

$$\begin{aligned}
\mathbb{E}\left[\left(\dot{b}_n^f\right)^k\right] &< \infty \\
\mathbb{E}\left[\left(\widetilde{b}_n^f\right)^k\right] &< \infty \\
\mathbb{E}\left[\left(\dot{b}_n^c\right)^k\right] &< \infty \\
\mathbb{E}\left[\left(\widetilde{b}_n^c\right)^k\right] &< \infty
\end{aligned} \tag{7.28}$$

We also use the hypothesis that our SDE is elliptic (see section 3.3.3)

$$\exists \epsilon > 0 \text{ s.t. } \forall (S,t) \in \mathbb{R}^{+2}, b(S,t) > \epsilon \tag{7.29}$$

We apply Hölder's inequality repeatedly and get the existence of constants $\left(\widetilde{C}_{\mathbb{B},k}\right)$ such that for $n = 0, \ldots, N_f$

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial p_n^f}{\partial \theta}\right)^k\right] &\leq \widetilde{C}_{\mathbb{B},k} h_f^{-k} \\
\mathbb{E}\left[\left(\frac{\partial p_{n/2}^c}{\partial \theta}\right)^k\right] &\leq \widetilde{C}_{\mathbb{B},k} h_f^{-k}
\end{aligned} \tag{7.30}$$

Then

$$
\begin{aligned}
\mathbb{E}\left(\left(\sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)\left|\frac{\partial p_n^f}{\partial\theta}\right|\right]\right)^k\right) &\leq \mathbb{E}\left(\left(\sum_{n=0}^{N_f-1}\left|\frac{\partial p_n^f}{\partial\theta}\right|\right)^k\right)\\
&\leq \mathbb{E}\left(\left(N_f\max_{n=0..N_f-1}\left|\frac{\partial p_n^f}{\partial\theta}\right|\right)^k\right)\\
&\leq N_f^k\,\mathbb{E}\left(\left(\max_{n=0..N_f-1}\left|\frac{\partial p_n^f}{\partial\theta}\right|\right)^k\right)\\
&= O\left(h_f^{-k}\right)O\left(h_f^{-k}\right)\\
&= O\left(h_f^{-2k}\right)
\end{aligned}
\tag{7.31}
$$

Similarly

$$
\mathbb{E}\left(\sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}(1-p_k^c)\left|\frac{\partial p_n^c}{\partial\theta}\right|\right]\right) = O\left(h_f^{-2k}\right) \tag{7.32}
$$

Finally, we recall that from equation (7.25),

$$
\begin{aligned}
\mathbb{E}\left(\mathbb{B}^4\right) \leq \mathbb{E}\Bigg[\wp\Bigg(&\left(\widehat{S}_{N_f}^f-K\right),\left(\widehat{S}_{N_c}^c-K\right),\\
&\sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)\left|\frac{\partial p_n^f}{\partial\theta}\right|\right],\\
&\sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}(1-p_k^c)\left|\frac{\partial p_n^c}{\partial\theta}\right|\right]\Bigg)\Bigg]
\end{aligned}
\tag{7.33}
$$

and we have proved that for all $k \geq 0$,

$$
\begin{aligned}
\mathbb{E}\left(\left(\widehat{S}_{N_f}^f-K\right)^k\right) &< C_k\\
\mathbb{E}\left(\left(\widehat{S}_{N_c}^c-K\right)^k\right) &< C_k\\
\mathbb{E}\left(\left(\sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)\left|\frac{\partial p_n^f}{\partial\theta}\right|\right]\right)^k\right) &= O\left(h_f^{-2k}\right)\\
\mathbb{E}\left(\left(\sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}(1-p_k^c)\left|\frac{\partial p_n^c}{\partial\theta}\right|\right]\right)^k\right) &= O\left(h_f^{-2k}\right)
\end{aligned}
\tag{7.34}
$$

Hölder's inequality applied to $\mathbb{E}\left[\wp(\ldots)\right]$ then guarantees the existence of some value $K_{\mathbb{B}} < \infty$ such that

$$\mathbb{E}\left[\wp\left(\ldots\right)\right] = O\left(h^{-K_{\mathbb{B}}}\right) \tag{7.35}$$

and thus

$$\mathbb{E}\left(\mathbb{B}^4\right) = O\left(h^{-K_{\mathbb{B}}}\right) \tag{7.36}$$

Combining all the results since equation (7.20),

$$\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^4\right) = O\left(h^{-K_{\mathbb{B}}}\right) \tag{7.37}$$

and from (7.16) and (7.18), we finally get for all $p > 0$

$$\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right) = \mathbb{E}\left(\mathbf{1}_{E^c}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right) + O\left(h^p\right) \tag{7.38}$$

This means extreme paths only have a negligible contribution to the global variance and we can now focus solely on the analysis of the variance for non-extreme paths.

Contribution of non-extreme paths

Using the previous results, we write as in (7.20) that for all $p > 0$,

$$\begin{aligned}\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right) &\leq \mathbb{E}\left(\left(|\,\mathbb{A}\,| + |\,\mathbb{B}\,|\right)^2\right)\\ &\sim \mathbb{E}\left(\mathbf{1}_{E^c}\left(|\,\mathbb{A}\,| + |\,\mathbb{B}\,|\right)^2\right) + O\left(h^p\right)\\ &\sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right) + \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^2\right) + O\left(h^p\right)\end{aligned} \tag{7.39}$$

We have established that only paths of E^c contribute significantly to the variance: we now study $\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right)$ and $\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^2\right)$. In the following computations, we restrict our study to non-extreme paths as defined by lemma 3.4.8 for some $\gamma > 0$.

$\mathbb{E}\left(\mathbb{A}^2\mathbf{1}_{E^c}\right)$: We begin with $\mathbb{E}\left(\mathbb{A}^2\right)$ on E^c and write

$$\begin{aligned}\mathbb{A} &= \Delta\left[\mathbf{1}_{\widehat{S}_N > K}\frac{\partial\widehat{S}_N}{\partial\theta}\prod_{n=0}^{N_f-1}\left(1 - p_n\right)\right]\\ &= \Delta\left[\mathbf{1}_{\widehat{S}_N > K}\frac{\partial\widehat{S}_N}{\partial\theta}\right]\prod_{n=0}^{N_f-1}\left(1 - p_n^f\right)\\ &\qquad + \mathbf{1}_{\widehat{S}_N^c > K}\frac{\partial\widehat{S}_N^c}{\partial\theta}\Delta\left[\prod_{n=0}^{N_f-1}\left(1 - p_n\right)\right]\\ &:= \mathbb{A}1 + \mathbb{A}2\end{aligned} \tag{7.40}$$

with

$$
\begin{aligned}
\mathbb{A}1 &:= \Delta\left[\mathbf{1}_{\widehat{S}_N>K}\frac{\partial \widehat{S}_N}{\partial\theta}\right]\prod_{n=0}^{N_f-1}\left(1-p_n^f\right)\\
\mathbb{A}2 &:= \mathbf{1}_{\widehat{S}_N^c>K}\frac{\partial \widehat{S}_N^c}{\partial\theta}\Delta\left[\prod_{n=0}^{N_f-1}(1-p_n)\right]
\end{aligned}
\tag{7.41}
$$

Note that

$$
\mathbb{E}\left(\mathbb{A}^2\mathbf{1}_{E^c}\right)\sim\mathbb{E}\left((\mathbb{A}1)^2\,\mathbf{1}_{E^c}\right)+\mathbb{E}\left((\mathbb{A}2)^2\,\mathbf{1}_{E^c}\right)
\tag{7.42}
$$

We have $\prod_{n=0}^{N_f-1}(1-p_n)\leq 1$, therefore

$$
\mathbb{E}\left[(\mathbb{A}1)^2\,\mathbf{1}_{E^c}\right]=O\left(\mathbb{E}\left[\left(\Delta\left[\mathbf{1}_{\widehat{S}_N>K}\frac{\partial \widehat{S}_N}{\partial\theta}\right]\right)^2\mathbf{1}_{E^c}\right]\right)
\tag{7.43}
$$

This actually corresponds to the case of the European call, which was dealt with in chapter 4. Therefore we can conclude that there is a certain $K_{\mathbb{A}1}$ such that

$$
\mathbb{E}\left((\mathbb{A}1)^2\,\mathbf{1}_{E^c}\right)=O\left(h_f^{2-K_{\mathbb{A}1}\gamma}\right)
\tag{7.44}
$$

For the study of $\mathbb{E}\left((\mathbb{A}2)^2\,\mathbf{1}_{E^c}\right)$, we note that

$$
\mathbb{A}2\leq\frac{\partial \widehat{S}_N^c}{\partial\theta}\Delta\left[\prod_{n=0}^{N_f-1}(1-p_n)\right]
\tag{7.45}
$$

By definition of non-extreme paths,

$$
\frac{\partial \widehat{S}_N^c}{\partial\theta}\leq h_f^{-\gamma}
\tag{7.46}
$$

From the analysis of the pricing of barrier options in [GDR13], we get

$$
\Delta\left[\prod_{n=0}^{N_f-1}(1-p_n)\right]=O\left(h_f^{1/2-5\gamma}\right)
\tag{7.47}
$$

Therefore there is some $K_{\mathbb{A}2}<\infty$ such that

$$
\mathbb{E}\left((\mathbb{A}2)^2\,\mathbf{1}_{E^c}\right)=O\left(h_f^{1-K_{\mathbb{A}2}\gamma}\right)
\tag{7.48}
$$

Finally putting things together, we get from (7.42)

$$
\begin{aligned}
\mathbb{E}\left(\mathbb{A}^2\mathbf{1}_{E^c}\right)&=O\left(h_f^{2-K_{\mathbb{A}1}\gamma}\right)+O\left(h_f^{1-K_{\mathbb{A}2}\gamma}\right)\\
&=O\left(h_f^{1-K_{\mathbb{A}2}\gamma}\right)
\end{aligned}
\tag{7.49}
$$

$\mathbb{E}\left(\mathbb{B}^2\mathbf{1}_{E^c}\right)$ **:** We then study $\mathbb{E}\left(\mathbb{B}^2\right)$ on E^c. We write

$$
\begin{aligned}
\mathbb{B} &= \Delta\left[\left(\widehat{S}_N - K\right)^+ \sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}(1-p_k)\frac{\partial p_n}{\partial\theta}\right]\right] \\
&= \Delta\left[\left(\widehat{S}_N - K\right)^+\right] \sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)\frac{\partial p_n^f}{\partial\theta}\right] \\
&\qquad + \left(\widehat{S}_N^c - K\right)^+ \sum_{n=0}^{N_f-1}\Delta\left[\prod_{k=0,k\neq n}^{N_f-1}(1-p_k)\frac{\partial p_n}{\partial\theta}\right] \\
&:= \mathbb{B}1 + \mathbb{B}2
\end{aligned} \tag{7.50}
$$

with

$$
\begin{aligned}
\mathbb{B}1 &:= \Delta\left[\left(\widehat{S}_N - K\right)^+\right] \sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)\frac{\partial p_n^f}{\partial\theta}\right] \\
\mathbb{B}2 &:= \left(\widehat{S}_N^c - K\right)^+ \sum_{n=0}^{N_f-1}\Delta\left[\prod_{k=0,k\neq n}^{N_f-1}(1-p_k)\frac{\partial p_n}{\partial\theta}\right]
\end{aligned} \tag{7.51}
$$

We define D the set of paths for which the minimum is "close" to B as the set of paths for which $\left|\min_{[0,T]} S_t - B\right| \leq 2h_f^{1/2-\delta}$ for some $\delta > 3\gamma$. Using (7.15) and (7.39), we get for all $p > 0$

$$
\begin{aligned}
\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right) &\sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right) + \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}^2\right) + O\left(h_f^p\right) \\
&\sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right) + \mathbb{E}\left(\mathbf{1}_{D}\mathbb{B}^2\right) + \mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{B}^2\right) + O\left(h_f^p\right)
\end{aligned} \tag{7.52}
$$

We now prove that paths in D^c have a negligible contribution to the global variance.

On D^c, we have two possible cases:

- If $S_{t_{\min}} := \min_{[0,T]} S_t \leq B - 2h_f^{1/2-\delta}$ and $t_{\min} \in [t_{n_{\min}}, t_{n_{\min}+1}]$, we write

$$
\left|\widehat{S}_{n_{\min}}^f - S_{t_{\min}}\right| \leq \left|\widehat{S}_{n_{\min}}^f - \widehat{S}^f\left(t_{\min}\right)\right| + \left|\widehat{S}^f\left(t_{\min}\right) - S_{t_{\min}}\right| \tag{7.53}
$$

  where the first term corresponds to the evolution of the interpolant (as defined in lemma 3.4.4) and the second one to the error between the exact solution and the interpolant at $t_{\min}$. We have

$$
\widehat{S}_{n_{\min}}^f - \widehat{S}^f\left(t_{\min}\right) = \frac{t_{\min} - t_{n_{\min}}}{h_f}\left(\widehat{S}_{n_{\min}+1}^f - \widehat{S}_{n_{\min}}^f\right) + b_{n_{\min}}^f\left(W_{t_{\min}} - W_{n_{\min}}\right) \tag{7.54}
$$

We apply the bounds of lemma 3.4.8 and get for non-extreme paths

$$\left| \widehat{S}^f_{n_{\min}} - S_{t_{\min}} \right| = O\left(h_f^{1/2-2\gamma}\right) = o\left(h_f^{1/2-\delta}\right) \tag{7.55}$$

and therefore $\widehat{S}^f_{n_{\min}} < B$, the barrier is hit by the fine path and $p^f_{n_{\min}} = 1$, $\dfrac{\partial p^f_{n_{\min}}}{\partial \theta} = 0$.

We also note that $\max\limits_n \left(\widehat{S}^f_n - \widehat{S}^c_n\right) \le h_f^{1-\gamma}$ which leads to $\widehat{S}^c_{n_{\min}} < B$, the barrier is hit by the coarse path.

Therefore in this first case, $\mathbb{B} = 0$.

- If $S_{t_{\min}} := \min\limits_{[0,T]} S_t \ge B + 2h_f^{1/2-\delta}$

$$\min_k \left(\widehat{S}^f_k, \widehat{S}^c_k\right) \ge B + 2h_f^{1/2-\delta} - h_f^{1-\gamma} \ge B + h_f^{1/2-\delta} \tag{7.56}$$

Then we note that with the inequalities of lemma 3.4.8 for non-extreme paths,

$$\frac{\left(\widehat{S}^f_n - B\right)^+ \left(\widehat{S}^f_{n+1} - B\right)^+}{{b^f_n}^2 h_f} \ge O\left(\frac{h_f^{1-2\delta}}{h_f^{1-2\gamma}}\right) \ge O\left(h_f^{-2(\delta-\gamma)}\right) \tag{7.57}$$

and as $\delta - \gamma > \gamma > 0$ we therefore have for all $p > 0$

$$p^f_n = O\left(\exp\left(-h_f^{-2(\delta-\gamma)}\right)\right) = O\left(h^p\right) \tag{7.58}$$

which we use again with (7.26) and the bounds of lemma 3.4.8 to write

$$\begin{aligned}
\left| \frac{\partial p^f_n}{\partial \theta} \right| &\le p^f_n \frac{2}{{b^f_n}^2 h_f} \Bigg| \widehat{\delta}^f_n \left(-\left(\widehat{S}^f_{n+1} - B\right) + \frac{2\left(\widehat{S}^f_n - B\right)\left(\widehat{S}^f_{n+1} - B\right) \dot{b}^f_n}{b^f_n} \right) \\
&\qquad + \widehat{\delta}^f_{n+1}\left(\widehat{S}^f_n - B\right) + \frac{2\left(\widehat{S}^f_n - B\right)\left(\widehat{S}^f_{n+1} - B\right) \widetilde{b}^f_n}{b^f_n} \Bigg| \\
&\le p^f_n \frac{2}{\epsilon^2 h_f} \left[h_f^{-\gamma} \left(h_f^{-\gamma} + \frac{2h_f^{-\gamma} h_f^{-\gamma} C_1 h_f^{-\gamma}}{\epsilon} \right) \right. \\
&\qquad \left. + h_f^{-\gamma} h_f^{-\gamma} + \frac{2h_f^{-\gamma} h_f^{-\gamma} C_1 h_f^{-\gamma}}{\epsilon} \right] \\
&= O\left(h^{p-1-4\gamma}\right)
\end{aligned} \tag{7.59}$$

Similarly we get $p^c_n = O\left(h^p\right)$ and $\dfrac{\partial p^c_n}{\partial \theta} = O\left(h^{p-1-4\gamma}\right)$. Noting that

$N_f = O\left(h_f^{-1}\right)$, we can then write for all $p > 0$

$$
\begin{aligned}
|\,\mathbb{B}1\,| &= \left| \Delta\left[\left(\widehat{S}_N - K\right)^+\right] \sum_{n=0}^{N_f-1} \left[\prod_{k=0,k\neq n}^{N_f-1} \left(1-p_k^f\right) \frac{\partial p_n^f}{\partial\theta}\right]\right| \\
&\leq \left|\,\Delta\widehat{S}_N\,\right| \sum_{n=0}^{N_f-1} \left|\,\frac{\partial p_n^f}{\partial\theta}\,\right| \\
&\leq O\left(h_f^{1-\gamma}\right) O\left(h_f^{-1}\right) o\left(h_f^{p-1-4\gamma}\right) \\
&= O\left(h_f^{p-1-5\gamma}\right)
\end{aligned}
\tag{7.60}
$$

and

$$
\begin{aligned}
|\,\mathbb{B}2\,| &\leq \left|\,\widehat{S}_N\,\right| \sum_{n=0}^{N_f-1} \left[\left|\prod_{k=0,k\neq n}^{N_f-1} \left(1-p_k^f\right) \frac{\partial p_n^f}{\partial\theta}\right| + \left|\prod_{k=0,k\neq n}^{N_f-1} \left(1-p_k^c\right) \frac{\partial p_n^c}{\partial\theta}\right|\right] \\
&\leq \left|\,\widehat{S}_N\,\right| \sum_{n=0}^{N_f-1} \left[\left|\,\frac{\partial p_n^f}{\partial\theta}\,\right| + \left|\,\frac{\partial p_n^c}{\partial\theta}\,\right|\right] \\
&\leq O\left(h_f^{-\gamma}\right) O\left(h_f^{-1}\right) o\left(h_f^{p-1-4\gamma}\right) \\
&= O\left(h_f^{p-2-5\gamma}\right)
\end{aligned}
\tag{7.61}
$$

Thus in this second case $\mathbb{B} = \mathbb{B}1 + \mathbb{B}2 = O\left(h_f^{\widetilde{p}}\right)$ for all $\widetilde{p} > 0$.

The analysis of those two cases finally gives for all $p > 0$,

$$
\mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{B}^2\right) = O\left(h_f^p\right) \tag{7.62}
$$

and from (7.52),

$$
\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right) \sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right) + \mathbb{E}\left(\mathbf{1}_D\mathbb{B}^2\right) + O\left(h_f^p\right) \tag{7.63}
$$

$\mathbb{E}\left(\mathbf{1}_D\mathbb{B}^2\right)$ **:** We now analyse $\mathbb{E}\left(\mathbb{B}^2\right)$ for paths of D, i.e. non-extreme paths for which $\left|\min_{[0,T]} S_t - B\right| \leq 2h_f^{1/2-\delta}$ for some $\delta > 3\gamma$. Assuming $\min_{[0,T]}(S_t)$ has a bounded density around B means a proportion $O\left(h_f^{1/2-\delta}\right)$ of all paths are in D.

As in (7.55), we have $\left|\,\widehat{S}^f_{n_{\min}} - S_{t_{\min}}\,\right| = O\left(h_f^{1/2-2\gamma}\right) = o\left(h_f^{1/2-\delta}\right)$ and $\max_n\left(\widehat{S}^f_n - \widehat{S}^c_n\right) \leq h_f^{1-\gamma}$ therefore on D,

$$
\begin{aligned}
\left|\min_n \widehat{S}^f_n - B\right| &= O\left(h_f^{1/2-2\gamma}\right) \\
\left|\min_n \widehat{S}^c_n - B\right| &= O\left(h_f^{1/2-2\gamma}\right)
\end{aligned}
\tag{7.64}
$$

and we define the set R of indices n such that $B < \widehat{S}_n^f, \widehat{S}_n^c, \widehat{S}_{n+1}^f, \widehat{S}_{n+1}^c < B + h_f^{1/2-3\gamma}$
.

$$\mathbb{E}\left(\mathbf{1}_D \mathbb{B}^2\right) \sim \mathbb{E}\left(\mathbf{1}_D \left(\mathbb{B}1\right)^2\right) + \mathbb{E}\left(\mathbf{1}_D \left(\mathbb{B}2\right)^2\right) \tag{7.65}$$

We first study the contribution of D's paths to $\mathbb{B}1$. We let

$$|\,\mathbb{B}1\,| = |\,\mathbb{B}11\,|\,|\,\mathbb{B}12\,| \tag{7.66}$$

with

$$\begin{aligned} \mathbb{B}11 &= \Delta\left[\left(\widehat{S}_N - K\right)^+\right] \\ \mathbb{B}12 &= \sum_{n=0}^{N_f-1}\left[\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)\frac{\partial p_n^f}{\partial\theta}\right] \end{aligned} \tag{7.67}$$

We have

$$|\,\mathbb{B}11\,| = \left|\,\Delta\left[\left(\widehat{S}_N - K\right)^+\right]\,\right| \leq \left|\,\Delta\widehat{S}_N\,\right| = O\left(h_f^{1-\gamma}\right) \tag{7.68}$$

then, we note that (7.59) still holds in general and that for indices in R, we have

$$\begin{aligned} \left|\,\frac{\partial p_n^f}{\partial\theta}\,\right| &\leq p_n^f \frac{2}{{b_n^f}^2 h_f}\left|\widehat{\delta}_n^f\left(-\left(\widehat{S}_{n+1}^f - B\right) + \frac{2\left(\widehat{S}_n^f - B\right)\left(\widehat{S}_{n+1}^f - B\right)\dot{b_n}^f}{b_n^f}\right)\right. \\ &\qquad\left. + \,\widehat{\delta}_{n+1}^f\left(\widehat{S}_n^f - B\right) + \frac{2\left(\widehat{S}_n^f - B\right)\left(\widehat{S}_{n+1}^f - B\right)\widetilde{b_n}^f}{b_n^f}\right| \\ &\leq p_n^f \frac{2}{\epsilon^2 h_f}\left[h_f^{-\gamma}\left(h_f^{1/2-3\gamma} + \frac{2h_f^{1/2-3\gamma}h_f^{1/2-3\gamma}C_1 h_f^{-\gamma}}{\epsilon}\right)\right. \\ &\qquad\left. + \,h_f^{-\gamma}h_f^{1/2-3\gamma} + \frac{2h_f^{1/2-3\gamma}h_f^{1/2-3\gamma}C_1 h_f^{-\gamma}}{\epsilon}\right] \\ &\leq p_n^f K_7 h_f^{-1/2-4\gamma} \end{aligned} \tag{7.69}$$

for some constant K_7. Then, using a reasoning similar to (7.60) on indices in R^c, for all $p > 0$,

$$
\begin{aligned}
|\,\mathbb{B}12\,| &\leq \sum_{n\in R}\left|\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)\frac{\partial p_n^f}{\partial\theta}\right| + \sum_{n\in R^c}\left|\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)\frac{\partial p_n^f}{\partial\theta}\right| \\
&\leq \sum_{n\in R}\prod_{k=0,k\neq n}^{N_f-1}\left|\left(1-p_k^f\right)\frac{\partial p_n^f}{\partial\theta}\right| + O\left(h_f^p\right) \\
&\leq \sum_{n\in R}\prod_{k=0,k\neq n}^{N_f-1}\left|\left(1-p_k^f\right)p_n^f K_7 h_f^{-1/2-4\gamma}\right| + O\left(h_f^p\right)
\end{aligned}
\tag{7.70}
$$

where the last inequality comes directly from (7.69). Then we use

$$
\sum_{n}\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)p_n^f \leq \prod_{k=0}^{N_f-1}\left((1-p_k^f)+p_k^f\right) = 1 \tag{7.71}
$$

to derive

$$
\begin{aligned}
|\,\mathbb{B}12\,| &\leq K_7 h_f^{-1/2-4\gamma}\sum_{n\in R}\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k^f\right)p_n^f + O\left(h_f^p\right) \\
&= O\left(h_f^{-1/2-4\gamma}\right)
\end{aligned}
\tag{7.72}
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}\left(\mathbf{1}_D\mathbb{B}1^2\right) &= \mathbb{E}\left(\mathbf{1}_D\mathbb{B}11^2\mathbb{B}12^2\right) \\
&= \mathbb{E}\left(\mathbb{B}11^2\mathbb{B}12^2|D\right)\mathbb{P}\left(D\right) \\
&\leq O\left(h_f^{2-2\gamma}\right)O\left(h_f^{-1-8\gamma}\right)O\left(h_f^{1/2-\delta}\right) \\
&= O\left(h_f^{3/2-10\gamma-\delta}\right)
\end{aligned}
\tag{7.73}
$$

We then study the contribution of D's paths to $\mathbb{B}2$. We let

$$
|\,\mathbb{B}2\,| = |\,\mathbb{B}21\,|\,|\,\mathbb{B}22\,| \tag{7.74}
$$

with

$$
\begin{aligned}
\mathbb{B}21 &= \left(\widehat{S}_N^c - K\right)^+ \\
\mathbb{B}22 &= \Delta\left[\sum_{n=0}^{N_f-1}\prod_{k=0,k\neq n}^{N_f-1}\left(1-p_k\right)\frac{\partial p_n}{\partial\theta}\right]
\end{aligned}
\tag{7.75}
$$

We are always dealing with non-extreme paths, thus we directly get

$$
|\,\mathbb{B}21\,| = O\left(h_f^{-\gamma}\right) \tag{7.76}
$$

The analysis of $|\,\mathbb{B}22\,|$ is slightly more intricate. Reasoning as before, we have for all $p > 0$

$$
\begin{aligned}
|\,\mathbb{B}22\,| &\leq \left| \Delta \left[\sum_{\substack{n=0\\ n\in R}}^{N_f-1} \prod_{k=0,k\neq n}^{N_f-1} (1-p_k) \frac{\partial p_n}{\partial \theta} \right] \right| + \left| \Delta \left[\sum_{\substack{n=0\\ n\in R^c}}^{N_f-1} \prod_{k=0,k\neq n}^{N_f-1} (1-p_k) \frac{\partial p_n}{\partial \theta} \right] \right| \\
&\leq \left| \Delta \left[\sum_{\substack{n=0\\ n\in R}}^{N_f-1} \prod_{k=0,k\neq n}^{N_f-1} (1-p_k) \frac{\partial p_n}{\partial \theta} \right] \right| + O\left(h_f^p\right)
\end{aligned}
\tag{7.77}
$$

We let $p_n(z) = p_n^c + z\left(p_n^f - p_n^c\right)$, $\dfrac{\partial p_n(z)}{\partial \theta} = \dfrac{\partial p_n^c}{\partial \theta} + z\left(\dfrac{\partial p_n^f}{\partial \theta} - \dfrac{\partial p_n^c}{\partial \theta}\right)$ and $f(z) = \displaystyle\sum_{\substack{n=0\\ n\in R}}^{N_f-1} \left[\prod_{\substack{k=0\\ k\neq n}}^{N_f-1} (1-p_k(z)) \frac{\partial p_n(z)}{\partial \theta} \right]$. Then

$$
\begin{aligned}
\mathbb{B}22 &= \Delta \left[\sum_{\substack{n=0\\ n\in R}}^{N_f-1} \prod_{\substack{k=0\\ k\neq n}}^{N_f-1} (1-p_k) \frac{\partial p_n}{\partial \theta} \right] + O\left(h_f^p\right) \\
&= f(1) - f(0) + O\left(h_f^p\right) \\
&\leq 1 \cdot \sup_{z\in[0,1]} \left| \frac{\partial f}{\partial z} \right| + O\left(h_f^p\right)
\end{aligned}
\tag{7.78}
$$

where for $z \in [0,1]$

$$
\begin{aligned}
\frac{\partial f}{\partial z} &= \sum_{\substack{n=0\\ n\in R}} \left[\sum_{l\neq n} \left(\prod_{k\neq n,l} -(1-p_k) \frac{\partial p_l}{\partial z} \frac{\partial p_n}{\partial \theta} \right) + \prod_{k\neq n} (1-p_k) \frac{\partial^2 p_n}{\partial z \partial \theta} \right] \\
&= \sum_{\substack{n=0\\ n\in R}} \left[\sum_{l\neq n} \left(\prod_{k\neq n,l} -(1-p_k) \Delta p_l \frac{\partial p_n}{\partial \theta} \right) + \prod_{k\neq n} (1-p_k) \Delta \frac{\partial p_n}{\partial \theta} \right]
\end{aligned}
\tag{7.79}
$$

We recall that on R, $\left| \dfrac{\partial p_n}{\partial \theta} \right| \leq K_6 h_f^{-1/2} p_n$ for some constant K_6.

Therefore,

$$
\left| \frac{\partial f}{\partial z} \right| \leq K_6 h_f^{-1/2} \sum_{\substack{n=0\\ n\in R}} \left[\sum_{l\neq n} \left(\prod_{k\neq n,l} (1-p_k)\,|\,\Delta p_l\,|\; p_n \right) + \prod_{k\neq n} (1-p_k) \left| \Delta \frac{\partial p_n}{\partial \theta} \right| \right]
\tag{7.80}
$$

Then if we let

$$p_n = \exp\left(X_n\right) \tag{7.81}$$

with

$$\begin{aligned} X_n^f &= \frac{-2(\widehat{S}_n^f - B)^+(\widehat{S}_{n+1}^f - B)^+}{(b_n^f)^2\, h_f} \\ X_n^c &= \frac{-2(\widehat{S}_n^c - B)^+(\widehat{S}_{n+1/2}^c - B)^+}{(b_n^c)^2\, h_f} \end{aligned} \tag{7.82}$$

it can be shown as in the analysis of the pricing of barrier options in [GDR13] that

$$\left| X_n^f - X_n^c \right| \leq K_8 h_f^{1/2-4\gamma} \tag{7.83}$$

for some constant K_8. We can then write

$$\begin{aligned} |\, \Delta p_n \,| &= \left| p_n^f - p_n^c \right| = p_n^c \left| \exp\left(X_n^c - X_n^f\right) - 1 \right| \leq p_n^c \Delta_h \\ |\, \Delta p_n \,| &= \left| p_n^f - p_n^c \right| = p_n^f \left| 1 - \exp\left(X_n^f - X_n^c\right) \right| \leq p_n^f \widetilde{\Delta}_h \end{aligned} \tag{7.84}$$

where

$$\begin{aligned} \Delta_h &= \exp\left(K_8 h_f^{1/2-4\gamma} - 1\right) \\ \widetilde{\Delta}_h &= 1 - \exp\left(-K_8 h_f^{1/2-4\gamma}\right) \end{aligned} \tag{7.85}$$

and as $h_f \to 0$, we have via a Taylor expansion

$$\begin{aligned} \Delta_h &\sim K_8 h_f^{1/2-4\gamma} \\ \widetilde{\Delta}_h &\sim K_8 h_f^{1/2-4\gamma} \end{aligned} \tag{7.86}$$

Therefore there is some constant K_9 such that

$$\begin{aligned} |\, \Delta p_n \,| &\leq p_n^f K_9 h_f^{1/2-4\gamma} \\ |\, \Delta p_n \,| &\leq p_n^c K_9 h_f^{1/2-4\gamma} \end{aligned} \tag{7.87}$$

and for $z \in [0,1]$,

$$|\, \Delta p_n(z) \,| \leq p_n(z) K_9 h_f^{1/2-4\gamma} \tag{7.88}$$

Using the same notation,

$$\begin{aligned} \left| \Delta \frac{\partial p_n}{\partial \theta} \right| &\leq p_n^f \left| \Delta \frac{\partial X_n}{\partial \theta} \right| + \left| \frac{\partial X_n^c}{\partial \theta} \right| |\, \Delta p_n \,| \\ \left| \Delta \frac{\partial p_n}{\partial \theta} \right| &\leq p_n^c \left| \Delta \frac{\partial X_n}{\partial \theta} \right| + \left| \frac{\partial X_n^f}{\partial \theta} \right| |\, \Delta p_n \,| \end{aligned} \tag{7.89}$$

as before, we can show that on R, $\left|\Delta\dfrac{\partial X_n}{\partial\theta}\right| = O\left(h_f^{-7\gamma}\right)$ and $\dfrac{\partial X_n}{\partial\theta} = O\left(h_f^{-1/2-4\gamma}\right)$. Then, using equation (7.88), we get

$$\left|\Delta\frac{\partial p_n}{\partial\theta}(z)\right| \le p_n(z)K_{10}h_f^{-8\gamma} \tag{7.90}$$

We use this in (7.80) and letting $K_{11} = \max\left(K_9, K_{10}\right)$, we get

$$\begin{aligned}\left|\frac{\partial f}{\partial z}\right| &\le K_6 h_f^{-1/2} K_{11} h_f^{1/2-4\gamma} \sum_{\substack{n=0\\ n\in R}} \left[\sum_{l\neq n}\left(\prod_{k\neq n,l}\left(1-p_k\right)p_l\,p_n\right) + \prod_{k\neq n}\left(1-p_k\right)p_n\right] \\ &\le K_6\,K_{11} h_f^{-4\gamma} \sum_{\substack{n=0\\ n\in R}} \left[\sum_{l\neq n}\left(\prod_{k\neq n,l}\left(1-p_k\right)p_l\,p_n\right) + \prod_{k\neq n}\left(1-p_k\right)p_n\right]\end{aligned} \tag{7.91}$$

and using the same idea as in (7.71),

$$\begin{aligned}&\sum_{n}\left[\sum_{l\neq n}\left(\prod_{k\neq n,l}\left(1-p_k\right)p_l\,p_n\right) + \prod_{k\neq n}\left(1-p_k\right)p_n\right] \\ &\qquad\qquad\le \prod_{k=0}^{N_f-1}\left(\left(1-p_k^f\right)+p_k^f\right) = 1\end{aligned} \tag{7.92}$$

and

$$\left|\frac{\partial f}{\partial z}\right| \le K_6\,K_{11} h_f^{-4\gamma} \tag{7.93}$$

from (7.78), we thus get

$$\left|\,\mathbb{B}22\,\right| \le \sup_{z\in[0,1]}\left|\frac{\partial f}{\partial z}\right| + O\left(h_f^p\right) = O\left(h_f^{-4\gamma}\right) \tag{7.94}$$

Putting things together, we get from (7.74)

$$\left|\,\mathbb{B}2\,\right| = O\left(h_f^{-5\gamma}\right) \tag{7.95}$$

and we can write

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_D\mathbb{B}2^2\right) &= \mathbb{E}\left(\mathbb{B}2^2|D\right)\mathbb{P}\left(D\right) \\ &\le O\left(h_f^{-10\gamma}\right)O\left(h_f^{1/2-\delta}\right) \\ &= O\left(h_f^{1/2-10\gamma-\delta}\right)\end{aligned} \tag{7.96}$$

Finally from the previous equation, (7.65) and (7.73), we get as in (7.65)

$$\mathbb{E}\left(\mathbf{1}_D\mathbb{B}^2\right) \sim O\left(h_f^{3/2-10\gamma-\delta}\right) + O\left(h_f^{1/2-10\gamma-\delta}\right) \sim O\left(h_f^{1/2-10\gamma-\delta}\right) \tag{7.97}$$

and using it with (7.63) and (7.49)

$$\begin{aligned}\mathbb{E}\left(\left(\Delta\frac{\partial \widehat{P}}{\partial\theta}\right)^2\right) &\sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}^2\right) + \mathbb{E}\left(\mathbf{1}_D\mathbb{B}^2\right) + O\left(h_f^p\right) \\ &= O\left(h_f^{1-K_{\mathbb{A}2}\gamma}\right) + O\left(h_f^{1/2-10\gamma-\delta}\right) + O\left(h_f^p\right) \\ &= O\left(h_f^{1/2-10\gamma-\delta}\right)\end{aligned} \tag{7.98}$$

This means we have $\beta = 1/2 - \widetilde{\gamma}$ for $\widetilde{\gamma} > 0$ as small as we want.

7.2.2 Order of convergence α

The analysis of α is very similar to the analysis of β. As in previous chapters, we analyse the convergence speed of $\mathbb{E}\left(\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\right)$. With the same notation as before,

$$\mathbb{E}\left(\frac{\partial \widehat{P}_f}{\partial\theta} - \frac{\partial \widehat{P}_c}{\partial\theta}\right) \sim \mathbb{E}\left(\mathbb{A}\right) + \mathbb{E}\left(\mathbb{B}\right) \tag{7.99}$$

Contribution of extreme paths

As before, we can write

$$\mathbb{E}\left(\Delta\frac{\partial \widehat{P}}{\partial\theta}\right) = \mathbb{E}\left(\mathbf{1}_{E^c}\Delta\frac{\partial \widehat{P}}{\partial\theta}\right) + \mathbb{E}\left(\mathbf{1}_{E}\Delta\frac{\partial \widehat{P}}{\partial\theta}\right) \tag{7.100}$$

As before

$$\mathbb{E}\left(\mathbf{1}_E\right) = O\left(h^p\right) \tag{7.101}$$

and using Hölder's inequality we show it is sufficient to prove that $\mathbb{E}\left(\left(\Delta\frac{\partial \widehat{P}}{\partial\theta}\right)^2\right)$ does not explode "too quickly" as $h \to 0$ to conclude that extreme paths have a negligible contribution to $\mathbb{E}\left(\Delta\frac{\partial \widehat{P}}{\partial\theta}\right)$.

We note that $\mathbb{E}\left(\left(\Delta\frac{\partial \widehat{P}}{\partial\theta}\right)^2\right) \sim \mathbb{E}\left(|\,\mathbb{A}\,|^2\right) + \mathbb{E}\left(|\,\mathbb{B}\,|^2\right)$ and study $\mathbb{E}\left(|\,\mathbb{A}\,|^2\right)$ and $\mathbb{E}\left(|\,\mathbb{B}\,|^2\right)$.

As before we can prove there exists some constant $\widetilde{C}_{\mathbb{A}} < \infty$ such that

$$\mathbb{E}\left(|\,\mathbb{A}\,|^2\right) < \widetilde{C}_{\mathbb{A}} \tag{7.102}$$

As before we can also prove the existence of some $\widetilde{K}_{\mathbb{B}} < \infty$ such that

$$\mathbb{E}\left(\mathbb{B}^2\right) = O\left(h^{-\widetilde{K}_{\mathbb{B}}}\right) \tag{7.103}$$

Combining the previous results,

$$\mathbb{E}\left(\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right)^2\right) = O\left(h^{-\widetilde{K}_{\mathbb{B}}}\right) \tag{7.104}$$

and as before we finally get for all $p > 0$

$$\mathbb{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right) = \mathbb{E}\left(\mathbf{1}_{E^c}\,\Delta\frac{\partial\widehat{P}}{\partial\theta}\right) + O\left(h^p\right) \tag{7.105}$$

This means extreme paths only have a negligible contribution and we can now focus solely on the analysis of the variance for non-extreme paths.

Contribution of non-extreme paths

Using the previous results, we write as before that for all $p > 0$,

$$\mathbb{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right) = \mathbb{E}\left(\mathbf{1}_{E^c}\,\mathbb{A}\right) + \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}\right) + O\left(h^p\right) \tag{7.106}$$

We now study $\mathbb{E}\left(\mathbf{1}_{E^c}\,\mathbb{A}\right)$ and $\mathbb{E}\left(\mathbf{1}_{E^c}\,\mathbb{B}\right)$, that is we now restrict our study to non-extreme paths only. With a slight abuse of notation, we write $\mathbb{E}\left(...\right)$ instead of $\mathbb{E}\left(...|E^c\right)$ when no ambiguity arises.

$\mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}\right)$: We study $\mathbb{E}\left(\mathbb{A}\right)$ on E^c. We recall

$$\begin{aligned}\mathbb{A}1 &:= \Delta\left[\mathbf{1}_{\widehat{S}_N>K}\frac{\partial\widehat{S}_N}{\partial\theta}\right]\prod_{n=0}^{N_f-1}\left(1-p_n^f\right)\\ \mathbb{A}2 &:= \mathbf{1}_{\widehat{S}_N^c>K}\frac{\partial\widehat{S}_N^c}{\partial\theta}\Delta\left[\prod_{n=0}^{N_f-1}\left(1-p_n\right)\right]\end{aligned} \tag{7.107}$$

and

$$\mathbb{E}\left(\mathbf{1}_{E^c}\,\mathbb{A}\right) = \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}1\right) + \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}2\right) \tag{7.108}$$

As before, we can show there is a certain $\widetilde{K}_{\mathbb{A}1}$ such that

$$\mathbb{E}\left(\mathbf{1}_{E^c}\,\mathbb{A}1\right) = O\left(h_f^{1-\widetilde{K}_{\mathbb{A}1}\gamma}\right) \tag{7.109}$$

For $\mathbb{E}\left(\mathbf{1}_{E^c}\,\mathbb{A}2\right)$, note that

$$\mathbb{A}2 \leq \frac{\partial \widehat{S}^c_N}{\partial \theta} \Delta \left[\prod_{n=0}^{N_f-1} (1-p_n) \right] \tag{7.110}$$

and there is some $\widetilde{K}_{\mathbb{A}2} < \infty$ such that

$$\mathbb{E}\left(\mathbf{1}_{E^c}\,\mathbb{A}2\right) = O\left(h_f^{1/2-\widetilde{K}_{\mathbb{A}2}\gamma}\right) \tag{7.111}$$

Finally putting things together, we get

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_{E^c}\,\mathbb{A}\right) &= O\left(h_f^{1-\widetilde{K}_{\mathbb{A}1}\gamma}\right) + O\left(h_f^{1/2-\widetilde{K}_{\mathbb{A}2}\gamma}\right) \\ &= O\left(h_f^{1/2-\widetilde{K}_{\mathbb{A}2}\gamma}\right)\end{aligned} \tag{7.112}$$

$\mathbb{E}\left(\mathbb{B}\,\mathbf{1}_{E^c}\right)$ **:** We then study $\mathbb{E}\left(\mathbb{B}\right)$ on E^c. We recall

$$\mathbb{B} := \mathbb{B}1 + \mathbb{B}2 \tag{7.113}$$

with

$$\begin{aligned}\mathbb{B}1 &:= \Delta\left[\left(\widehat{S}_N - K\right)^+\right] \sum_{n=0}^{N_f-1} \left[\prod_{k=0,k\neq n}^{N_f-1} \left(1-p^f_k\right) \frac{\partial p^f_n}{\partial \theta} \right] \\ \mathbb{B}2 &:= \left(\widehat{S}^c_N - K\right)^+ \sum_{n=0}^{N_f-1} \Delta \left[\prod_{k=0,k\neq n}^{N_f-1} (1-p_k) \frac{\partial p_n}{\partial \theta} \right]\end{aligned} \tag{7.114}$$

We use again D, the set of paths for which the minimum is "close" to B as the set of paths for which $\left| \min_{[0,T]} S_t - B \right| \leq 2h_f^{1/2-\delta}$ for some $\delta > 3\gamma$. Using (7.15) and (7.39), as in (7.52), we get for all $p > 0$

$$\begin{aligned}\mathbb{E}\left(\Delta \frac{\partial \widehat{P}}{\partial \theta}\right) &\sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}\right) + \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{B}\right) + O\left(h_f^p\right) \\ &\sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}\right) + \mathbb{E}\left(\mathbf{1}_{D}\mathbb{B}\right) + \mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{B}\right) + O\left(h_f^p\right)\end{aligned} \tag{7.115}$$

As before, we prove that paths in D^c have a negligible contribution to the global expectation.

On D^c, we have two possibile cases:

- If $S_{t_{\min}} := \min_{[0,T]} S_t \leq B - 2h_f^{1/2-\delta}$ and $t_{\min} \in [t_{n_{\min}}, t_{n_{\min}+1}]$, we have as before that $\widehat{S}^f_{n_{\min}} < B$, the barrier is hit by the fine path and $p^f_{n_{\min}} = 1$, $\dfrac{\partial p^f_{n_{\min}}}{\partial \theta} = 0$. Similarly we also get that $\widehat{S}^c_{n_{\min}} < B$, the barrier is hit by the coarse path. Therefore $\mathbb{B} = 0$.

- If $S_{t_{\min}} := \min_{[0,T]} S_t \geq B + 2h_f^{1/2-\delta}$ we have as before that for all $p > 0$,

$$\begin{aligned} p_n^f &= O\left(h^p\right) \\ \frac{\partial p_n^f}{\partial \theta} &= O\left(h^{p-1-4\gamma}\right) \end{aligned} \tag{7.116}$$

and

$$\begin{aligned} p_n^c &= O\left(h^p\right) \\ \frac{\partial p_n^c}{\partial \theta} &= O\left(h^{p-1-4\gamma}\right) \end{aligned} \tag{7.117}$$

Then we conclude that

$$\begin{aligned} |\,\mathbb{B}1\,| &= O\left(h_f^{p-1-5\gamma}\right) \\ |\,\mathbb{B}2\,| &= O\left(h_f^{p-2-5\gamma}\right) \end{aligned} \tag{7.118}$$

Thus in this case $\mathbb{B} = \mathbb{B}1 + \mathbb{B}2 = O\left(h_f^{\widetilde{p}}\right)$ for all $\widetilde{p} > 0$.

The analysis of those two cases finally gives for all $p > 0$,

$$\mathbb{E}\left(\mathbf{1}_{D^c}\mathbb{B}\right) = O\left(h_f^p\right) \tag{7.119}$$

and from (7.52),

$$\mathbb{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right) \sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}\right) + \mathbb{E}\left(\mathbf{1}_D\mathbb{B}\right) + O\left(h_f^p\right) \tag{7.120}$$

$\mathbb{E}\left(\mathbf{1}_D\,\mathbb{B}\right)$ **:** We now analyse $\mathbb{E}\left(\mathbb{B}\right)$ for paths of D, i.e. non-extreme paths for which $\left|\min_{[0,T]} S_t - B\right| \leq 2h_f^{1/2-\delta}$ for some $\delta > 3\gamma$.

As in (7.65),

$$\mathbb{E}\left(\mathbf{1}_D\,\mathbb{B}\right) \sim \mathbb{E}\left(\mathbf{1}_D\,\mathbb{B}1\right) + \mathbb{E}\left(\mathbf{1}_D\,\mathbb{B}2\right) \tag{7.121}$$

We first study the contribution of D's paths to $\mathbb{B}1$. As before we write

$$|\,\mathbb{B}1\,| = |\,\mathbb{B}11\,|\,|\,\mathbb{B}12\,| \tag{7.122}$$

and use the fact established earlier that on D,

$$|\,\mathbb{B}11\,| = \left|\,\Delta\left[\left(\widehat{S}_N - K\right)^+\right]\,\right| \leq \left|\,\Delta\widehat{S}_N\,\right| = O\left(h_f^{1-\gamma}\right) \tag{7.123}$$

and

$$|\,\mathbb{B}12\,| = O\left(h_f^{-1/2-4\gamma}\right) \tag{7.124}$$

thus

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_D\,\mathbb{B}1\right) &= \mathbb{E}\left(\mathbf{1}_D\,\mathbb{B}11\,\mathbb{B}12\right)\\ &= \mathbb{E}\left(\mathbb{B}11\mathbb{B}12|D\right)\mathbb{P}\left(D\right)\\ &\leq O\left(h_f^{1-\gamma}\right)O\left(h_f^{-1/2-4\gamma}\right)O\left(h_f^{1/2-\delta}\right)\\ &= O\left(h_f^{1-5\gamma-\delta}\right)\end{aligned} \tag{7.125}$$

We then study the contribution of D's paths to $\mathbb{B}2$. We write as before

$$|\,\mathbb{B}2\,| = |\,\mathbb{B}21\,|\,|\,\mathbb{B}22\,| \tag{7.126}$$

and as before we have on D that

$$|\,\mathbb{B}21\,| = O\left(h_f^{-\gamma}\right) \tag{7.127}$$

$$|\,\mathbb{B}22\,| = O\left(h_f^{-4\gamma}\right) \tag{7.128}$$

Putting things together, we get as before

$$|\,\mathbb{B}2\,| = O\left(h_f^{-5\gamma}\right) \tag{7.129}$$

and we can write

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_D\,\mathbb{B}2\right) &= \mathbb{E}\left(\mathbb{B}2|D\right)\mathbb{P}\left(D\right)\\ &\leq O\left(h_f^{-5\gamma}\right)O\left(h_f^{1/2-\delta}\right)\\ &= O\left(h_f^{1/2-5\gamma-\delta}\right)\end{aligned} \tag{7.130}$$

Finally we get

$$\mathbb{E}\left(\mathbf{1}_D\,\mathbb{B}\right) = O\left(h_f^{1-5\gamma-\delta}\right) + O\left(h_f^{1/2-5\gamma-\delta}\right) \;= O\left(h_f^{1/2-5\gamma-\delta}\right) \tag{7.131}$$

and plugging all the results into (7.120),

$$\begin{aligned}\mathbb{E}\left(\Delta\frac{\partial\widehat{P}}{\partial\theta}\right) &\sim \mathbb{E}\left(\mathbf{1}_{E^c}\mathbb{A}\right) + \mathbb{E}\left(\mathbf{1}_D\mathbb{B}\right) + O\left(h_f^p\right)\\ &= O\left(h_f^{1/2-\widetilde{K}_{\mathbb{A}2}\gamma}\right) + O\left(h_f^{1/2-5\gamma-\delta}\right) + O\left(h_f^p\right)\end{aligned} \tag{7.132}$$

This means there is a constant K_γ such that we have $\alpha = 1/2 - K_\gamma\,\widetilde{\gamma}$ for any $\widetilde{\gamma} > 0$.

7.3 Conclusion

We have proved $\beta = (1/2 - \widetilde{\gamma})$, $\alpha = (1/2 - K_\gamma\,\widetilde{\gamma})$ for $\widetilde{\gamma}$ as small as we want.

We can summarise the results of this chapter as follows:

Theorem 7.3.1. *Let us consider a barrier option on an underlying asset S_t. We assume that the underlying asset's price S_t follows an Ito process as described by equation* (1.2)*, that the coefficients of the diffusion $a(S,t)$ and $b(S,t)$ satisfy conditions A1 to A4 of theorem 3.4.3 and that there exists a constant $\epsilon > 0$ such that $b(S,t) \geq \epsilon$.*

Our multilevel estimators of the Greeks of Barrier call options (see section 2.6) have an accuracy $O(\epsilon)$ at a cost $O\left(\epsilon^{-3}\right)$.

Proof. See above. $\square$

Chapter 8

Lookback options

We analyse the efficiency of multilevel Monte Carlo for the Greeks of lookback options. For doing so, we determine analytical bounds for the coefficients α and β of theorem 1.2.1 in this setting.

We begin by examining the case of continuously sampled lookback options. We highlight what makes the corresponding analysis surprisingly difficult compared to that of other payoffs' sensitivities and to that of the pricing of the option itself.

While most closed-form expressions available for pricing were historically based on continuous-time models (see [GSG79], [CV91], [HK95]), it is important to note that many (if not most) traded options are based on discrete price fixings. As explained in [BGK97], [GKB98] or [Kou07] the prices of all those options are actually closely related.

Thus, instead of considering the original problem (continuously sampled minimum) as described in chapter 2, we then analyse the case of lookback options with a discretely sampled minimum. We begin with the the case where the minimum is sampled at only two points of the time interval $[0,T]$ and use this analysis to generalise to the case where the minimum is sampled at any finite number of points.

8.1 Lookback options with continuously sampled minimum

We reuse the notations of section 2.5 and write $\widehat{S}^f_{t^f_{\min}} := \min_{k=0,\dots,N_f-1} \widehat{S}^f_{k,\min}$ with $t^f_{\min}$ corresponding to the time at which $\widehat{S}^f_t$ reaches its simulated minimum and $\widehat{S}^c_{t^c_{\min}} := \min_{k=0,\dots,N_f/2-1} \widehat{S}^c_{2k,\min}$ with $t^c_{\min}$ corresponding to the time at which $\widehat{S}^c_t$ reaches its simulated minimum.

We recall that the analysis of the multilevel pricing of continuously sampled lookback options (see [GDR13]) involves the analysis of $\mathbb{E}\left[\left(\widehat{S}^f_T - \widehat{S}^c_T\right)^2\right]$ and $\mathbb{E}\left[\left(\widehat{S}^f_{t^f_{\min}} - \widehat{S}^c_{t^c_{\min}}\right)^2\right]$. Under the usual assumptions of chapter 3, the strong convergence of the Milstein scheme gives directly $\mathbb{E}\left[\left(\widehat{S}^f_T - \widehat{S}^c_T\right)^2\right] = O\left(h^2\right)$ and in section 3.5 of [GDR13], it is shown that $\mathbb{E}\left[\left(\widehat{S}^f_{t^f_{\min}} - \widehat{S}^c_{t^c_{\min}}\right)^2\right] = O\left(h^2 \log h^2\right)$, which concludes the analysis.

Now let us consider the Greeks and see how their analysis differs from that of the pricing.

8.1.1 Analysis of the convergence rate β

As usual we start by using

$$\mathbb{V}\left(\frac{\partial \widehat{P}^f}{\partial\theta} - \frac{\partial \widehat{P}^c}{\partial\theta}\right) \leq \mathbb{E}\left[\left(\frac{\partial \widehat{P}^f}{\partial\theta} - \frac{\partial \widehat{P}^c}{\partial\theta}\right)^2\right] \tag{8.1}$$

With the notation introduced above, we have

$$\mathbb{V}\left(\frac{\partial \widehat{P}^f}{\partial\theta} - \frac{\partial \widehat{P}^c}{\partial\theta}\right) \leq \mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_T}{\partial\theta} - \frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_T}{\partial\theta} + \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] \tag{8.2}$$

and using the fact that cross-terms cannot determine the order of convergence of the variance (see section 4.1.1),

$$\begin{aligned}\mathbb{V}\left(\frac{\partial \widehat{P}^f}{\partial\theta} - \frac{\partial \widehat{P}^c}{\partial\theta}\right) = O&\left(\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_T}{\partial\theta} - \frac{\partial \widehat{S}^c_T}{\partial\theta}\right)^2\right]\right) \\ &+ O\left(\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right)\end{aligned} \tag{8.3}$$

which is reminiscent of the terms involved in the analysis of the pricing.

With the same assumptions as before, the strong convergence properties of the Milstein scheme guarantee that

$$\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_T}{\partial\theta} - \frac{\partial \widehat{S}^c_T}{\partial\theta}\right)^2\right] = O\left(h_l^2\right) \tag{8.4}$$

Now studying $\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right]$ is less obvious as there is a priori nothing linking the values of $\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta}$ and $\frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}$ directly when $t^c_{\min} \neq t^f_{\min}$. Indeed we can write

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] &= \mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^f_{\min}}}{\partial\theta} + \frac{\partial S_{t^f_{\min}}}{\partial\theta}\right.\right. \\
&\qquad\qquad \left.\left. - \frac{\partial S_{t^c_{\min}}}{\partial\theta} + \frac{\partial S_{t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] \\
&= O\left(\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2\right]\right) + \\
&\qquad O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right) + \\
&\qquad\qquad O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right)
\end{aligned} \tag{8.5}
$$

We can write

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2\right] &= \mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^f_{KP,t^f_{\min}}}{\partial\theta} + \frac{\partial \widehat{S}^f_{KP,t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2\right] \\
\mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] &= \mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{KP,t^c_{\min}}}{\partial\theta} + \frac{\partial \widehat{S}^c_{KP,t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right]
\end{aligned} \tag{8.6}
$$

and thus

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2\right] &= O\left(\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^f_{KP,t^f_{\min}}}{\partial\theta}\right)^2\right]\right) \\
&\qquad + O\left(\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{KP,t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2\right]\right) \\
\mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] &= O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{KP,t^c_{\min}}}{\partial\theta}\right)^2\right]\right) \\
&\qquad + O\left(\mathbb{E}\left[\left(\frac{\partial \widehat{S}^c_{KP,t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right)
\end{aligned} \tag{8.7}
$$

Equation (3.82) in lemma 3.4.4 and equation (3.79) in theorem 3.4.3 then give that for any $\left(t^f_{\min},\, t^c_{\min}\right) \in [0,T]^2$

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2\right] &= O\left((h_l \log h_l)^2\right) \\
\mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] &= O\left((h_l \log h_l)^2\right)
\end{aligned} \tag{8.8}$$

Therefore

$$\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] = O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right) + O\left((h_l \log h_l)^2\right) \tag{8.9}$$

To study $\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]$, we assume wlog that $t^f_{\min} < t^c_{\min}$. We let $\Delta t_{\min} = t^f_{\min} - t^c_{\min}$. Under the usual assumptions, theorem 3.4.1 then yields

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta} - \frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2 \middle| t^f_{\min}, t^c_{\min} \text{ fixed}\right] \\
&\qquad \leq D\left(1 + \mathbb{E}\left[\left|\frac{\partial S_{t^f_{\min}}}{\partial\theta}\right|^2 \middle| t^f_{\min}, t^c_{\min} \text{ fixed}\right]\right) |\,\Delta t_{\min}\,| \exp\left(C\,|\,\Delta t_{\min}\,|\right)
\end{aligned} \tag{8.10}$$

where C and D are constants that depend on T, not on $\left(t^c_{\min}, t^f_{\min}\right)$. As we consider a finite time interval $[0,T]$, this same theorem also guarantees that $\mathbb{E}\left(\left|\frac{\partial S_{t^f_{\min}}}{\partial\theta}\right|^2 \middle| t^f_{\min}, t^c_{\min} \text{ fixed}\right)$ is bounded uniformly. More precisely for any $t^f_{\min}$, we have

$$\mathbb{E}\left(\left|\frac{\partial S_{t^f_{\min}}}{\partial\theta}\right|^2 \middle| t^f_{\min}, t^c_{\min} \text{ fixed}\right) \leq \left(1 + \frac{\partial S_0}{\partial\theta}^2\right) \exp\left(CT\right) \tag{8.11}$$

and

$$1 \leq \exp\left(C\,|\,\Delta t_{\min}\,|\right) \leq \exp\left(CT\right) \tag{8.12}$$

therefore for T fixed, there exists a constant $\widetilde{D}$ such that

$$\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2 \middle| t^f_{\min}, t^c_{\min} \text{ fixed}\right] \leq \widetilde{D}\left|t^c_{\min} - t^f_{\min}\right| \tag{8.13}$$

And finally, using the tower property,

$$\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]=\mathbb{E}_{t^f_{\min},t^c_{\min}}\left(\mathbb{E}\left[\left(\frac{\partial S^c_{t^f_{\min}}}{\partial\theta}-\frac{\partial S^c_{t^c_{\min}}}{\partial\theta}\right)^2\middle| t^f_{\min},t^c_{\min}\right]\right)$$
$$\leq \widetilde{D}\,\mathbb{E}\left(\left|\,t^c_{\min}-t^f_{\min}\right|\right) \tag{8.14}$$

and combining it with (8.9) and (8.2), we get

$$\mathbb{V}\left(\frac{\partial\widehat{P}^f}{\partial\theta}-\frac{\partial\widehat{P}^c}{\partial\theta}\right)=O\left(\mathbb{E}\left[|\,\Delta t_{\min}\,|\right]\right)+O\left((h_l\log h_l)^2\right) \tag{8.15}$$

This highlights the fact that the values of $\dfrac{\partial\widehat{S}^f_{t^f_{\min}}}{\partial\theta}$ and $\dfrac{\partial\widehat{S}^c_{t^c_{\min}}}{\partial\theta}$ may diverge when $t^c_{\min}\neq t^f_{\min}$ and may contribute significantly to the global variance. The order of convergence $\beta\approx 1$ observed in section 2.5 actually suggests it is the limiting factor.

8.1.2 Experimental behaviour of $\Delta t_{\min}$

Unfortunately the analysis of the order of convergence of $\mathbb{E}\left[|\,\Delta t_{\min}\,|\right]$ is tricky and prevents us from concluding the analysis we started. To have a general idea of how $\mathbb{E}\left[|\,\Delta t_{\min}\,|\right]$ behaves, we perform the following simulation: we simulate $\widehat{S}^f_n$ for $n=0,\ldots,N_f$ and compute the local minimum on each fine time step $[t_n,t_{n+1}]$ as described in section 2.5.1. Similarly we compute $\widehat{S}^c_n$ for $n=0,\ldots,N_f$ and compute the local minima on the coarse time steps $[t_{2k},t_{2k+2}]$ for $k=0,\ldots,N_f/2-1$.

Our method does not enable us to know precisely where $t^f_{\min}$ and $t^c_{\min}$ would be. Nevertheless, to have a rough idea of the behaviour of $\mathbb{E}\left[|\,\Delta t_{\min}\,|\right]$, we arbitrarily place $t^f_{\min}$ in the middle of the fine interval where the global minimum of the fine path is reached and, similarly, we place $t^c_{\min}$ in the middle of the coarse interval where the global minimum of the coarse path is reached.

Simulations in the Black & Scholes setting as in section 2.5 suggest the following convergence rates (see figure 8.1)

$$
\begin{aligned}
\mathbb{E}\left(\Delta t_{\min}\right) &= O\left(h_l^{1.2}\right) \\
\mathbb{E}\left(\mid \Delta t_{\min} \mid\right) &= O\left(h_l^{1.0}\right) \\
\mathbb{E}\left[\left(\Delta t_{\min}\right)^2\right] &= O\left(h_l^{1.0}\right) \\
\mathbb{E}\left[\left(\Delta t_{\min}\right)^3\right] &= O\left(h_l^{1.2}\right) \\
\mathbb{E}\left[\left(\Delta t_{\min}\right)^4\right] &= O\left(h_l^{1.0}\right) \\
\mathbb{E}\left[\left(\Delta t_{\min}\right)^5\right] &= O\left(h_l^{1.0}\right) \\
\mathbb{E}\left[\left(\Delta t_{\min}\right)^6\right] &= O\left(h_l^{1.0}\right) \\
\mathbb{E}\left[\left(\Delta t_{\min}\right)^7\right] &= O\left(h_l^{0.9}\right) \\
&\cdots
\end{aligned}
\tag{8.16}
$$

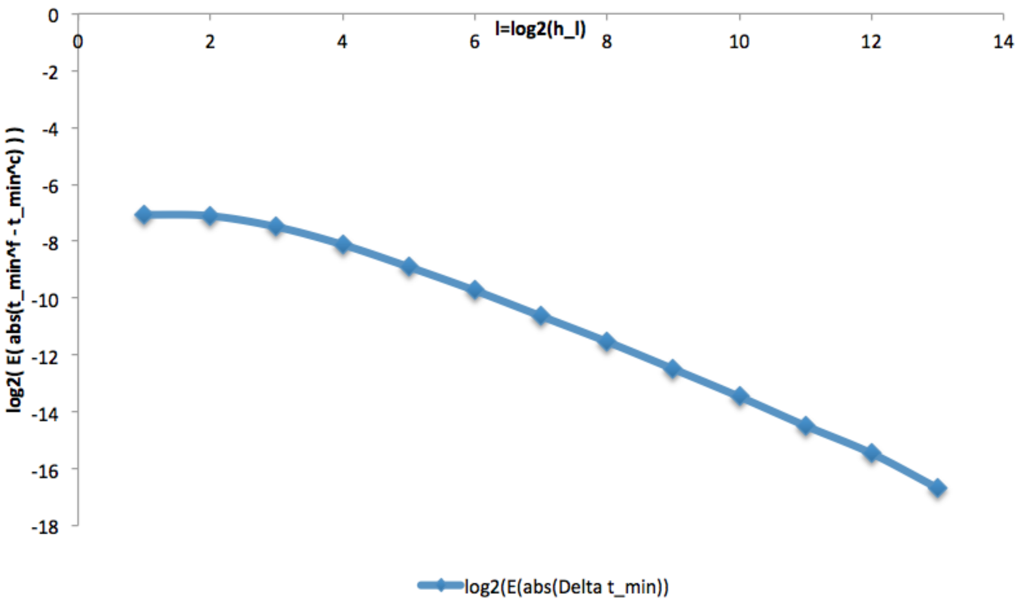

Figure 8.1: $\mathbb{E}\left(\mid \Delta t_{\min} \mid\right)$, $l = 2..10$

Finally, using these numerical results in (8.15),

$$
\begin{aligned}
\mathbb{V}\left(\frac{\partial \widehat{P}^f}{\partial \theta} - \frac{\partial \widehat{P}^c}{\partial \theta}\right) &= O\left(\mathbb{E}\left(\mid \Delta t_{\min} \mid\right)\right) + O\left(\left(h_l \log h_l\right)^2\right) \\
&= O\left(h_l^{1.0}\right)
\end{aligned}
\tag{8.17}
$$

This corresponds to what we observed in chapter 2 and suggests the simulation gives us a reasonable insight into the behaviour of the terms that contribute to the estimators' variance.

Our attempted analysis still ends up relying on a numerical simulation and it is arguable that it does not represent a major improvement over a purely numerical estimation of β as in section 2.5. Nevertheless, as explained earlier, we can legitimately focus on the related case of discretely sampled lookback options for which we now provide a complete analysis.

8.2 Analysis of a similar payoff: discretely sampled lookback option

Discretely sampled lookback options are lookback options where the minimum of the path is tracked on a fixed set of points $\mathcal{T} = (T_0, T_1, \ldots, T_{M-1} = T)$ and the payoff is defined as

$$P = S_T - \min_{k \in 0,\ldots,M-1} (S_{T_k}) \tag{8.18}$$

For the sake of simplicity, we assume there is a certain level l_0 such that the points of $\mathcal{T}$ are all located on the multilevel discretisation grids $(t_0 = 0, t_1 = h_l, \ldots, t_{N_l} = T)$ for $l > l_0$, e.g. taking $M = N_{l_0} + 1$ and the uniformly spaced sampling points $\mathcal{T} = (0, h_{l_0}, 2h_{l_0}, \ldots, (N_{l_0} - 1)\, h_{l_0}, N_{l_0} h_{l_0})$.

Results in [GKB98] and [Kou07] show that in the Black & Scholes case, the price of the discretely sampled lookback call converges towards the price of a continuously sampled lookback call with a small continuity correction term that depends on M. The price of the discretely sampled option converges towards that of the continuous one as we increase the sampling frequency, i.e. $M \to \infty$.

We now analyse the multilevel Greeks of such options.

8.2.1 Case where $M = 2$

To introduce the methods used for the analysis, we start with the simple case where $M = 2$, i.e. the minimum of the path is sampled at only two points in time: expiry T and some intermediate time $U \in [0, T]$. The payoff can thus be written as

$$P = S_T - \min (S_U, S_T) \tag{8.19}$$

We write with a slight abuse of notation $t^f_{\min}$ the point of $\{U, T\}$ where the fine path $\widehat{S}^f$ is minimum (respectively $t^c_{\min}$ the point of $\{U, T\}$ where the coarse path $\widehat{S}^c$ is minimum).

8.2.1.1 Order of convergence β

As in section 8.1.1, we write for some parameter θ

$$
\begin{aligned}
\mathbb{V}\left(\frac{\partial P^f}{\partial\theta}-\frac{\partial P^c}{\partial\theta}\right) &\leq \mathbb{E}\left[\left(\frac{\partial P^f}{\partial\theta}-\frac{\partial P^c}{\partial\theta}\right)^2\right] \\
&\leq \mathbb{E}\left[\left(\frac{\partial\widehat{S}^f_T}{\partial\theta}-\frac{\partial\widehat{S}^f_{t^f_{\min}}}{\partial\theta}-\frac{\partial\widehat{S}^c_T}{\partial\theta}+\frac{\partial\widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] \\
&= O\left(\mathbb{E}\left[\left(\frac{\partial\widehat{S}^f_T}{\partial\theta}-\frac{\partial\widehat{S}^c_T}{\partial\theta}\right)^2\right]\right)+O\left(\mathbb{E}\left[\left(\frac{\partial\widehat{S}^f_{t^f_{\min}}}{\partial\theta}-\frac{\partial\widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right)
\end{aligned}
\tag{8.20}
$$

As before, the convergence properties of the Milstein scheme give

$$
\mathbb{E}\left[\left(\frac{\partial\widehat{S}^f_T}{\partial\theta}-\frac{\partial\widehat{S}^c_T}{\partial\theta}\right)^2\right] = O\left(h_l^2\right)
\tag{8.21}
$$

Again, we use the decomposition

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial\widehat{S}^f_{t^f_{\min}}}{\partial\theta}-\frac{\partial\widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] &= \mathbb{E}\left[\left(\frac{\partial\widehat{S}^f_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^f_{\min}}}{\partial\theta}+\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}+\frac{\partial S_{t^c_{\min}}}{\partial\theta}-\frac{\partial\widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] \\
&= O\left(\mathbb{E}\left[\left(\frac{\partial\widehat{S}^f_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2\right]\right)+ \\
&\qquad O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right)+ \\
&\qquad\quad O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta}-\frac{\partial\widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right)
\end{aligned}
\tag{8.22}
$$

Using theorem 3.4.3 for the properties of the Milstein scheme and lemma 3.4.4 for the properties of the Brownian Bridge used for the coarse path interpolation,

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial\widehat{S}^f_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2\right] &= O\left(h_l^2\right) \\
\mathbb{E}\left[\left(\frac{\partial S_{t^c_{\min}}}{\partial\theta}-\frac{\partial\widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right)^2\right] &= O\left(\left(h_l\log h_l\right)^2\right)
\end{aligned}
\tag{8.23}
$$

and thus, as in (8.9),

$$\begin{aligned}\mathbb{V}\left(\frac{\partial P^f}{\partial \theta}-\frac{\partial P^c}{\partial \theta}\right) &= O\left(h_l^2\right)+O\left(\mathbb{E}\left[\left(\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial \theta}-\frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial \theta}\right)^2\right]\right)\\ &= O\left(\left(h_l \log h_l\right)^2\right)+O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^2\right]\right)\end{aligned} \tag{8.24}$$

We now study

$$\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^2\right] \tag{8.25}$$

Contribution of extreme paths As before, we define the set of extreme paths E satisfying any of the three conditions of lemma 3.4.8 for a certain $\gamma < 1/2$. We show those have a negligible contribution to the variance.

$$\Omega = E \sqcup E^c \tag{8.26}$$

Therefore

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^2\right] &= \mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^2 \mathbf{1}_{E^c}\right]\\ &\quad + \mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^2 \mathbf{1}_{E}\right]\end{aligned} \tag{8.27}$$

Using Hölder's inequality,

$$\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^2 \mathbf{1}_{E}\right] \leq \sqrt{\mathbb{E}\left[\mathbf{1}_E\right]}\sqrt{\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^4\right]} \tag{8.28}$$

lemma 3.4.8 and theorem 3.4.1 ensure that for all $p > 0$,

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_E\right] &= O\left(h_l^p\right)\\ \mathbb{E}\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^4 &= O\left(1\right)\end{aligned} \tag{8.29}$$

and thus for all $\tilde{p} > 0$,

$$\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial \theta}-\frac{\partial S_{t^c_{\min}}}{\partial \theta}\right)^2 \mathbf{1}_{E}\right] = O\left(h_l^{\tilde{p}}\right) O\left(1\right) = O\left(h_l^{\tilde{p}}\right) \tag{8.30}$$

that is, extreme paths have a negligible contribution and we can focus on the contribution of non-extreme paths:

$$\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]=\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\mathbf{1}_{E^c}\right]+O\left(h_l^{\tilde{p}}\right) \tag{8.31}$$

In the coming sections, the analysis is therefore restricted to non-extreme paths. Except when specified otherwised, it is always assumed that we are working exclusively within E^c.

Contribution of paths for which $t^f_{\min}=t^c_{\min}$ Now we define D, the set of non-extreme paths such that $t^f_{\min}=t^c_{\min}$, and D^c its complementary. On D, we have

$$\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2=\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^f_{\min}}}{\partial\theta}\right)^2=0 \tag{8.32}$$

We can thus write

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\mathbf{1}_{E^c}\right]&=\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\mathbf{1}_{D^c}\right]\\&\quad+\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\mathbf{1}_{D}\right]\\&=\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\mathbf{1}_{D^c}\right]\end{aligned} \tag{8.33}$$

This means the only significant contribution comes from non-extreme paths for which $t^f_{\min}\neq t^c_{\min}$. Using Hölder's inequality again, we get

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\mathbf{1}_{D^c}\right]\\&\qquad\leq\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^{2p}\mathbf{1}_{E^c}\right]^{1/p}\mathbb{E}\left[\mathbf{1}_{D^c}\right]^{1/q}\end{aligned} \tag{8.34}$$

where $p,q>1$ are such that $\dfrac{1}{p}+\dfrac{1}{q}=1$. Note we can let $p\to\infty$ and $q\to 1$.

We note that $\mathbb{E}\left[\mathbf{1}_{D^c}\right]=\mathbb{P}\left(\left(t^f_{\min}\neq t^c_{\min}\right)\cap E^c\right)$. In the case $M=2$, $\left(t^f_{\min}\neq t^c_{\min}\right)$ corresponds to

$$\left(t^f_{\min}=U\cap t^c_{\min}=T\right)\cup\left(t^f_{\min}=T\cap t^c_{\min}=U\right) \tag{8.35}$$

that is

$$\left(\left(\widehat{S}^f_U<\widehat{S}^f_T\right)\cap\left(\widehat{S}^c_T<\widehat{S}^c_U\right)\right)\cup\left(\left(\widehat{S}^f_T<\widehat{S}^f_U\right)\cap\left(\widehat{S}^c_U<\widehat{S}^c_T\right)\right) \tag{8.36}$$

thus

$$\mathbb{E}\left[\mathbf{1}_{D^c}\right] = \mathbb{P}\left(\left(\widehat{S}^f_U < \widehat{S}^f_T\right) \cap \left(\widehat{S}^c_T < \widehat{S}^c_U\right) \cap E^c\right) + \mathbb{P}\left(\left(\widehat{S}^f_T < \widehat{S}^f_U\right) \cap \left(\widehat{S}^c_U < \widehat{S}^c_T\right) \cap E^c\right) \tag{8.37}$$

We now study $\mathbb{P}\left(\left(\widehat{S}^f_U < \widehat{S}^f_T\right) \cap \left(\widehat{S}^c_T < \widehat{S}^c_U\right) \cap E^c\right)$, i.e. the case $t^f_{\min} = U$. We use the properties of non-extreme paths:

$$\begin{aligned} \widehat{S}^f_T &\leq \widehat{S}^c_T + h_l^{1-\gamma} \\ \widehat{S}^c_U &\leq \widehat{S}^f_U + h_l^{1-\gamma} \end{aligned} \tag{8.38}$$

and get

$$\widehat{S}^f_U \leq \widehat{S}^f_T \leq \widehat{S}^c_T + h_l^{1-\gamma} < \widehat{S}^c_U + h_l^{1-\gamma} \leq \widehat{S}^f_U + 2h_l^{1-\gamma} \tag{8.39}$$

This means that for non-extreme paths, $\widehat{S}^f_T \in \left[\widehat{S}^f_U, \widehat{S}^f_U + 2h_l^{1-\gamma}\right]$ and thus $S_T \in \left[S_U - 2h_l^{1-\gamma}, S_U + 4h_l^{1-\gamma}\right]$. Summing up previous results, this means that for paths in E^c,

$$\left(t^f_{\min} = U \cap t^c_{\min} = T\right) \Rightarrow S_T \in \left[S_U - 2h^{1-\gamma}, S_U + 4h_l^{1-\gamma}\right] \tag{8.40}$$

Note that for some event X,

$$\begin{aligned} \mathbb{P}\left(X \cap E^c\right) &= \mathbb{P}\left(X | E^c\right) \mathbb{P}\left(E^c\right) \\ &= \mathbb{P}\left(X | E^c\right)\left(1 - O\left(h_l^p\right)\right) \\ &\sim \mathbb{P}\left(X | E^c\right) \end{aligned} \tag{8.41}$$

Therefore, we can write

$$\begin{aligned} \mathbb{P}\left(t^f_{\min} = U \cap t^c_{\min} = T \cap E^c\right) &\leq \mathbb{P}\left(t^f_{\min} = U \cap t^c_{\min} = T | E^c\right) \\ &\leq \mathbb{P}\left(S_T \in \left[S_U - 2h^{1-\gamma}, S_U + 4h_l^{1-\gamma}\right] | E^c\right) \\ &\sim \mathbb{P}\left(S_T \in \left[S_U - 2h^{1-\gamma}, S_U + 4h_l^{1-\gamma}\right] \cap E^c\right) \\ &\leq \mathbb{P}\left(S_T \in \left[S_U - 2h^{1-\gamma}, S_U + 4h_l^{1-\gamma}\right]\right) \end{aligned} \tag{8.42}$$

As before, the ellipticity condition (see section 3.3.3) ensures that the SDE's solution has a smooth and bounded transition probability density function $p_{S_T - S_U}$ between U and T. Therefore it is uniformly bounded on the intervals $\left[-2\,h_l^{1-\gamma}, 4\,h_l^{1-\gamma}\right]$

as $h_l \to 0$ and

$$\begin{aligned}\mathbb{P}\left(S_T \in \left[S_U - 2\,h_l^{1-\gamma}, S_U + 4\,h_l^{1-\gamma}\right]\right) &= \int\limits_{-2\,h_l^{1-\gamma}}^{4\,h_l^{1-\gamma}} p_{S_T - S_U}(S)\mathrm{d}S \\ &\leq 6\,h_l^{1-\gamma} \max_{\left[-2\,h_l^{1-\gamma}, 4\,h_l^{1-\gamma}\right]} \left(p_{S_T - S_U}(S)\right) \\ &= O\left(h_l^{1-\gamma}\right)\end{aligned} \tag{8.43}$$

Therefore

$$\mathbb{P}\left(\left(\widehat{S}_U^f < \widehat{S}_T^f\right) \cap \left(\widehat{S}_T^c < \widehat{S}_U^c\right) \cap E^c\right) = O\left(h_l^{1-\gamma}\right) \tag{8.44}$$

Similarly, in the case $t_{\min}^f = T$, we show that

$$\mathbb{P}\left(\left(\widehat{S}_T^f < \widehat{S}_U^f\right) \cap \left(\widehat{S}_U^c < \widehat{S}_T^c\right) \cap E^c\right) = O\left(h_l^{1-\gamma}\right) \tag{8.45}$$

and

$$\mathbb{E}\left[\mathbf{1}_{D^c}\right] = O\left(h_l^{1-\gamma}\right) \tag{8.46}$$

Equation (8.34) becomes

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^2 \mathbf{1}_{D^c}\right] \\ &\qquad \leq \mathbb{E}\left[\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^{2p} \mathbf{1}_{E^c}\right]^{1/p} O\left(h_l^{\frac{1-\gamma}{q}}\right)\end{aligned} \tag{8.47}$$

We now analyse $\mathbb{E}\left[\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^{2p} \mathbf{1}_{E^c}\right]^{1/p}$. If the path is extreme, then

$$\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^{2p} \mathbf{1}_{E^c} = 0 \tag{8.48}$$

If the path is non-extreme, then

$$\begin{aligned}\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^{2p} \mathbf{1}_{E^c} &\leq \left(\left|\frac{\partial S_{t_{\min}^f}}{\partial \theta}\right| + \left|\frac{\partial S_{t_{\min}^c}}{\partial \theta}\right|\right)^{2p} \\ &\leq 4^p h_l^{-2p\gamma}\end{aligned} \tag{8.49}$$

finally

$$\mathbb{E}\left[\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^{2p} \mathbf{1}_{E^c}\right]^{1/p} \leq 4h_l^{-2\gamma} \tag{8.50}$$

Equation (8.34) is then

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2 \mathbf{1}_{D^c}\right] &= O\left(h_l^{-2\gamma}\right) O\left(h_l^{\frac{1-\gamma}{q}}\right) \\
&= O\left(h_l^{\frac{1-\gamma}{q}-2\gamma}\right)
\end{aligned}
\tag{8.51}
$$

From (8.33), (8.31) and (8.24),

$$
\begin{aligned}
\mathbb{V}\left(\frac{\partial P^f}{\partial\theta}-\frac{\partial P^c}{\partial\theta}\right) &= O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right) \\
&= O\left(h_l^{\frac{1-\gamma}{q}-2\gamma}\right)
\end{aligned}
\tag{8.52}
$$

As $q \to 1^+$, $\dfrac{1-\gamma}{q} \to 1-\gamma$, therefore for any $\gamma > 0$, we can pick $q(\gamma) > 1$ sufficiently close to 1 such that $1-4\gamma \leq \dfrac{1-\gamma}{q(\gamma)} - 2\gamma$. Then we take the corresponding $p(\gamma)$ such that $\dfrac{1}{p(\gamma)} + \dfrac{1}{q(\gamma)} = 1$. With this pair $p(\gamma), q(\gamma)$, equation (8.52) becomes

$$
\mathbb{V}\left(\frac{\partial P^f}{\partial\theta}-\frac{\partial P^c}{\partial\theta}\right) = O\left(h_l^{1-4\gamma}\right)
\tag{8.53}
$$

We have proved that $\beta = (1-\widetilde{\gamma})$ for $\widetilde{\gamma}$ as small as we want.

8.2.1.2 Order of convergence α

The analysis of the weak convergence is very similar to that of the strong convergence. We write for some parameter θ

$$
\begin{aligned}
\mathbb{E}\left(\frac{\partial P^f}{\partial\theta}-\frac{\partial P^c}{\partial\theta}\right) &= \mathbb{E}\left[\frac{\partial \widehat{S}^f_T}{\partial\theta} - \frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_T}{\partial\theta} + \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right] \\
&= \mathbb{E}\left[\frac{\partial \widehat{S}^f_T}{\partial\theta} - \frac{\partial \widehat{S}^c_T}{\partial\theta}\right] + \mathbb{E}\left[\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial\theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial\theta}\right]
\end{aligned}
\tag{8.54}
$$

Under the usual assumptions, the weak convergence properties of the Milstein scheme give

$$
\mathbb{E}\left[\frac{\partial \widehat{S}^f_T}{\partial\theta} - \frac{\partial \widehat{S}^c_T}{\partial\theta}\right] = O\left(h_l\right)
\tag{8.55}
$$

We now study

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial \theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial \theta}\right] &= \mathbb{E}\left[\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial \theta} - \frac{\partial S_{t^f_{\min}}}{\partial \theta} + \frac{\partial S_{t^f_{\min}}}{\partial \theta} - \frac{\partial S_{t^c_{\min}}}{\partial \theta} + \frac{\partial S_{t^c_{\min}}}{\partial \theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial \theta}\right] \\
&= \mathbb{E}\left[\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial \theta} - \frac{\partial S_{t^f_{\min}}}{\partial \theta}\right] + \\
&\quad \mathbb{E}\left[\frac{\partial S_{t^f_{\min}}}{\partial \theta} - \frac{\partial S_{t^c_{\min}}}{\partial \theta}\right] + \\
&\quad \mathbb{E}\left[\frac{\partial S_{t^c_{\min}}}{\partial \theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial \theta}\right]
\end{aligned} \tag{8.56}$$

Under the usual conditions,

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial \theta} - \frac{\partial S_{t^f_{\min}}}{\partial \theta}\right] &= O\left(h_l\right) \\
\mathbb{E}\left[\frac{\partial S_{t^c_{\min}}}{\partial \theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial \theta}\right] &= O\left(h_l \log h_l\right)
\end{aligned} \tag{8.57}$$

and thus

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial P^f}{\partial \theta} - \frac{\partial P^c}{\partial \theta}\right) &= O\left(h_l\right) + O\left(\mathbb{E}\left[\frac{\partial \widehat{S}^f_{t^f_{\min}}}{\partial \theta} - \frac{\partial \widehat{S}^c_{t^c_{\min}}}{\partial \theta}\right]\right) \\
&= O\left(h_l \log h_l\right) + O\left(\mathbb{E}\left[\frac{\partial S_{t^f_{\min}}}{\partial \theta} - \frac{\partial S_{t^c_{\min}}}{\partial \theta}\right]\right)
\end{aligned} \tag{8.58}$$

We now study

$$\mathbb{E}\left[\frac{\partial S_{t^f_{\min}}}{\partial \theta} - \frac{\partial S_{t^c_{\min}}}{\partial \theta}\right] \tag{8.59}$$

Contribution of extreme paths Again we can check extreme paths have only a negligible contribution to the global expectation: for all $p > 0$,

$$\begin{aligned}
\left|\mathbb{E}\left[\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right]\right| &\leq \left|\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{E^c}\right]\right| \\
&\quad + \left|\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{E}\right]\right| \\
&\leq \left|\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{E^c}\right]\right| \\
&\quad + \sqrt{\mathbb{E}\left[\mathbf{1}_E\right]}\sqrt{\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]} \\
&\leq \left|\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{E^c}\right]\right| + O\left(h_l^p\right)O\left(1\right)
\end{aligned} \tag{8.60}$$

Contribution of paths for which $t^f_{\min} = t^c_{\min}$ As before we can write

$$\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{E^c}\right] = \mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{D^c}\right] \tag{8.61}$$

We thus focus on non-extreme paths for which $t^f_{\min} \neq t^c_{\min}$. Using Hölder's inequality,

$$\begin{aligned}
&\left|\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{D^c}\right]\right| \\
&\qquad \leq \mathbb{E}\left[\left|\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right|^p \mathbf{1}_{E^c}\right]^{1/p} \mathbb{E}\left[\mathbf{1}_{D^c}\right]^{1/q}
\end{aligned} \tag{8.62}$$

where $p, q > 1$ are such that $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

From (8.46), $\mathbb{E}\left[\mathbf{1}_{D^c}\right] = O\left(h_l^{1-\gamma}\right)$ and the previous equation becomes

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{E^c}\mathbf{1}_{D^c}\right] \\
&\qquad \leq \mathbb{E}\left[\left|\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right|^p \mathbf{1}_{E^c}\right]^{1/p} O\left(h_l^{\frac{1-\gamma}{q}}\right)
\end{aligned} \tag{8.63}$$

As before, we note that for both extreme and non-extreme paths,

$$\left|\frac{\partial S_{t^f_{\min}}}{\partial\theta} - \frac{\partial S_{t^c_{\min}}}{\partial\theta}\right|^p \mathbf{1}_{E^c} \leq 2^p h_l^{-p\gamma} \tag{8.64}$$

finally

$$\mathbb{E}\left[\left|\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right|^p \mathbf{1}_{E^c}\right]^{1/p} \leq 2h_l^{-\gamma} \tag{8.65}$$

Equation (8.63) is then

$$\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)\mathbf{1}_{E^c}\mathbf{1}_{D^c}\right] = O\left(h_l^{\frac{1-\gamma}{q}-\gamma}\right) \tag{8.66}$$

From (8.58), (8.60) and (8.61),

$$\mathbb{E}\left(\frac{\partial P^f}{\partial\theta}-\frac{\partial P^c}{\partial\theta}\right) = O\left(h_l^{\frac{1-\gamma}{q}-\gamma}\right) \tag{8.67}$$

As before, we can pick $q(\gamma)>1$ such that $1-3\gamma \leq \dfrac{1-\gamma}{q(\gamma)}-\gamma$. and then $p(\gamma)$ such that $\dfrac{1}{p(\gamma)}+\dfrac{1}{q(\gamma)}=1$. With this pair $p(\gamma),q(\gamma)$, equation (8.67) becomes

$$\mathbb{E}\left(\frac{\partial P^f}{\partial\theta}-\frac{\partial P^c}{\partial\theta}\right) = O\left(h_l^{1-3\gamma}\right) \tag{8.68}$$

We have proved that $\alpha=(1-\widetilde{\gamma})$ for $\widetilde{\gamma}$ as small as we want.

8.2.2 Case where $M>2$

The analysis of the more realistic case where the number of sample points M is larger ($M>2$) is extremely similar to the case $M=2$.

8.2.2.1 Order of convergence β

As before we obtain

$$\mathbb{V}\left(\frac{\partial P^f}{\partial\theta}-\frac{\partial P^c}{\partial\theta}\right) = O\left((h_l\log h_l)^2\right)+O\left(\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right]\right) \tag{8.69}$$

As in (8.33) and (8.34), we can write

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\right] &\sim \mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^2\mathbf{1}_{D^c}\right]\\ &\leq \mathbb{E}\left[\left(\frac{\partial S_{t^f_{\min}}}{\partial\theta}-\frac{\partial S_{t^c_{\min}}}{\partial\theta}\right)^{2p}\mathbf{1}_{E^c}\right]^{1/p}\mathbb{E}\left[\mathbf{1}_{D^c}\right]^{1/q}\end{aligned} \tag{8.70}$$

where $p,q>1$ are such that $\dfrac{1}{p}+\dfrac{1}{q}=1$.

Again we compute

$$\mathbb{E}\left[\mathbf{1}_{D^c}\right] = \mathbb{P}\left(\left(t_{\min}^f \neq t_{\min}^c\right) \cap E^c\right) \tag{8.71}$$

Now $\mathcal{T}^2$ has M^2 elements and there are M couples $\left(t_{\min}^f, t_{\min}^c\right)$ such that $t_{\min}^f = t_{\min}^c$, therefore there are $M^2 - M$ possible couples $\left(t_{\min}^f, t_{\min}^c\right)$ to consider. Let us consider one of them: we write $t_{\min}^f = U$ and $t_{\min}^c = V$. As before, we get

$$\begin{aligned}\mathbb{P}\left(\left(t_{\min}^f \neq t_{\min}^c\right) \cap E^c\right) &= \sum_{(U \neq V) \in \mathcal{T}^2} \mathbb{P}\left(t_{\min}^f = U \cap t_{\min}^c = V \cap E^c\right) \\ &= O\left(h_l^{1-\gamma}\right)\end{aligned} \tag{8.72}$$

That is, $\mathbb{E}\left[\mathbf{1}_{D^c}\right]^{1/q} = O\left(h_l^{\frac{1-\gamma}{q}}\right)$ and as before we can construct $q(\gamma)$ sufficiently small such that

$$\mathbb{E}\left[\mathbf{1}_{D^c}\right]^{1/q} = O\left(h_l^{1-2\gamma}\right) \tag{8.73}$$

As in (8.48) and (8.49), for both extreme and non-extreme paths

$$\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^{2p} \mathbf{1}_{E^c} \leq 4^p h_l^{-2p\gamma} \tag{8.74}$$

thus

$$\mathbb{E}\left[\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^{2p} \mathbf{1}_{E^c}\right]^{1/p} \leq 4 h_l^{-\gamma} \tag{8.75}$$

and finally

$$\mathbb{V}\left(\frac{\partial P^f}{\partial \theta} - \frac{\partial P^c}{\partial \theta}\right) = O\left(h_l^{-2\gamma}\right) O\left(h_l^{1-2\gamma}\right) = O\left(h_l^{1-4\gamma}\right) \tag{8.76}$$

We have proved that $\beta = (1 - \widetilde{\gamma})$ for $\widetilde{\gamma}$ as small as we want.

8.2.2.2 Order of convergence α

As before we prove

$$\mathbb{E}\left(\frac{\partial P^f}{\partial \theta} - \frac{\partial P^c}{\partial \theta}\right) = O\left(h \log h_l\right) + O\left(\mathbb{E}\left[\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)\right]\right) \tag{8.77}$$

and

$$\mathbb{E}\left[\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)\right] \leq \mathbb{E}\left[\left(\frac{\partial S_{t_{\min}^f}}{\partial \theta} - \frac{\partial S_{t_{\min}^c}}{\partial \theta}\right)^p \mathbf{1}_{E^c}\right]^{1/p} \mathbb{E}\left[\mathbf{1}_{D^c}\right]^{1/q} \tag{8.78}$$

where $p, q > 1$ are such that $\frac{1}{p} + \frac{1}{q} = 1$. Finally, we obtain

$$\mathbb{E}\left(\frac{\partial P^f}{\partial \theta} - \frac{\partial P^c}{\partial \theta}\right) = O\left(h_l^{-\gamma}\right) O\left(h_l^{1-2\gamma}\right) = O\left(h_l^{1-3\gamma}\right) \tag{8.79}$$

We have proved that $\alpha = (1 - \widetilde{\gamma})$ for $\widetilde{\gamma}$ as small as we want.

8.3 Conclusion

The analysis of sections 8.2.1 and 8.2.2 reveals that in the discrete case the order of convergence is limited by the fact that $\frac{\partial S}{\partial \theta}$, the sensitivity of the underlying asset, can differ significantly between the times at which the fine and the coarse paths reach their respective minimums. This also corresponds to what the analysis of section 8.1.1 together with the simulation of section 8.1.2 suggest for the continuous case.

We have proved that in the case of discretely sampled lookback options, $\alpha = 1 - \widetilde{\gamma}$ and $\beta = 1 - \widetilde{\gamma}$ for $\widetilde{\gamma} > 0$ as small as we want. Those rates of convergence are also the ones observed in chapter 2 for continuously sampled lookback options. Intuitively this is explained by the fact that the price of a discretely sampled lookback option converges towards the price of its continuously sampled equivalent when the number of sampling points M is increased.

We can summarise the results we have proved in this chapter as follows:

Theorem 8.3.1. *Let us consider a lookback option on an asset S_t. We assume that this underlying asset's price follows an Ito process as described by equation (1.2), that the coefficients of the diffusion $a(S,t)$ and $b(S,t)$ satisfy conditions A1 to A4 of theorem 3.4.3 and that there exists a constant $\epsilon > 0$ such that $b(S,t) \geq \epsilon$.*

Multilevel pathwise sensitivities estimators of the Greeks of discretely sampled lookback options have an accuracy $O(\epsilon)$ at a cost $O\left(\epsilon^{-2} (\log \epsilon)^2\right)$.

Proof. See above. □

Chapter 9

Conclusion and future work

Option prices' sensitivities, the Greeks, are indicators of financial risk. Their estimation is an important problem in mathematical finance. A commonly used method is Monte Carlo simulations. However, obtaining accurate estimates of those sensitivities can be computationally costly, much more so than simply estimating option prices. The computational cost of obtaining a sensitivity with an accuracy of ϵ is at best of order $O\left(\epsilon^{-3}\right)$ and for most options it is even higher. Efficient algorithms enabling faster computations therefore become crucial in this setting.

Multilevel techniques provide computational savings in various settings, notably for option pricing. Nevertheless such methods have never been used for the computation of Greeks. In this thesis we tried to answer the following questions: can multilevel Monte Carlo techniques be applied to this problem? If so, how should we use them to obtain computationally efficient estimators of the sensitivities? Finally what complexity improvements can be achieved using these methods?

Here we summarise the results we obtained and later we will describe new questions arising from this work, new directions for future research.

9.1 Summary

Introduction - chapter 1

We first described the setting in which we are working. We then presented existing Monte Carlo techniques used for pricing . We also recalled other Monte Carlo techniques (pathwise sensitivities, Likelihood Ratio Method, Vibrato Monte Carlo) for estimating the Greeks, i.e. their sensitivities and the basic principles of multilevel Monte Carlo pricing.

Then we explained how we would combine those ideas in the following chapters: we would use them as building blocks for techniques intended to efficiently evaluate Greeks.

Algorithms and simulations - chapter 2

Here we first explained in detail the implementation of the ideas mentioned in chapter 1 for various option types. We then ran simulations to verify the validity of our algorithms. This would also provide a better insight into the potential computational savings offered by those new multilevel estimators.

As a preliminary step we gave details about how to read the experimental convergence rates from our simulations. We also explained how to use those readings with the complexity theorem of chapter 1 to numerically estimate the computational cost of our algorithms.

We first dealt with simple European options with regular payoffs for which we could use pathwise sensitivities.

However, using multilevel techniques efficiently for other option types proved more tricky: digital options required a special treatment to take into account the payoff's discontinuity and permit the use of pathwise sensitivities. This lead to the so-called conditional expectation technique which can also help achieve higher convergence rates for non-smooth Lipschitz options; we also showed how Vibrato Monte Carlo or the path splitting technique could be used as an alternative to the conditional expectation technique in a multilevel setting.

While Asian options can be easily dealt with using multilevel pathwise sensitivities, other exotic options like barrier or lookback options require more attention. In those cases, obtaining maximum computational savings from multilevel ideas could prove particularly tricky. We needed to take into account the behaviour of the path between discretisation times to simulate local minima or to estimate their likelihood of crossing the barrier within each time step.

Preliminary theoretical results - chapter 3

In chapter 2 we obtained experimental evidence of the validity of our approach. We now needed to rigorously prove that it resulted in asymptotically unbiased estimators of the Greeks. We also had to confirm the computational benefits they offered. This would involve a detailed analysis of each estimator, which we would perform in chapters 4 to 8. Beforehand in chapter 3 we established several important theoretical results.

We first proved that under certain assumptions, the stochastic differential equation followed by an asset and its sensitivities did indeed correspond to the equation obtained by naively differentiating the asset value's evolution equation. This enabled us to prove the validity of the discretisation scheme used in chapter 2.

We then came up with practical conditions ensuring that the pathwise sensitivities technique was applicable when working with the exact solution of an Itô process. We proved that this method was also applicable when working with discretised solutions of the process and that the sensitivities of the payoff estimators did result in estimators of the Greeks with a vanishing bias.

We then provided conditions on the process's volatility ensuring that its transition probabilities are well-behaved (smooth, bounded). Those conditions may be slightly restrictive but contribute to having a cleaner analysis.

We finally introduced various path approximations that are used in following chapters and established their properties by recalling essential theorems about the moments of SDE solutions and about the so-called extreme-path analysis.

Detailed analysis of multilevel Greeks estimators - chapters 4 to 8

As explained earlier, chapters 4 to 8 consist of a detailed analysis of the convergence and computational complexity of the multilevel algorithms suggested in chapter 2.

Chapter 4 - Multilevel pathwise sensitivities for European options. In chapter 4 we provided an analytical proof for the convergence rates/computational complexities observed in chapter 2 in the case of European options with Lipschitz payoffs. To introduce essential ideas of the analysis and its basic outline, we began with the simple case of options with smooth payoffs and proved that the computational complexity using multilevel Monte Carlo was reduced to $O\left(\epsilon^{-2}\right)$ instead of $O\left(\epsilon^{-3}\right)$ with standard (single level) Monte Carlo.

We then adapted this analysis to the case of non-smooth payoffs; this analysis revealed that we had to consider the respective contributions of paths arriving near the payoff's kinks and of all other paths, which involved the use of the extreme-path analysis. The kinks reduce the convergence speed of multilevel estimators and therefore limit the complexity to $O\left(\epsilon^{-2}\left(\log\epsilon\right)^2\right)$.

Chapter 5 - Multilevel pathwise sensitivities and conditional expectations technique. Chapter 5 dealt with the joint use of pathwise sensitivities and the conditional expectation technique. While the technique was initially introduced to extend the use of pathwise sensitivities to discontinuous payoffs (it enables the computation of the Greeks of options with discontinuous payoffs with a complexity $O\left(\epsilon^{-3}\right)$), it is also useful in the case of non-smooth Lipschitz payoffs and improves the computational savings obtained with multilevel Monte Carlo. We indeed obtained complexities of $O\left(\epsilon^{-2}\right)$ for the European call instead of $O\left(\epsilon^{-2}\left(\log\epsilon\right)^2\right)$ for simple pathwise sensitivities as explained in chapter 4.

Chapter 6 - Asian options. We analysed Asian options in chapter 6. While this case a priori required the analysis of the paths and their discretisations on the whole time interval considered and therefore required the use of stronger results than previous chapters, a minor change of the fundamental "extreme path theorem" of chapter 3 made the analysis fairly similar to that of European options with non-smooth Lipschitz payoffs. We obtained a complexity $O\left(\epsilon^{-2}\left(\log\epsilon\right)^2\right)$.

Chapter 7 - Barrier options. Chapter 7 provided an analysis of our multilevel estimator for barrier options' sensitivities. As the payoff depends on the path's entire trajectory and more specifically on its minimum, only considering its values at the discretisation times was insufficient and we computed the likelihood that the path's approximation (Milstein discretisation with Brownian bridge interpolation within each time step) reached the barrier. This resulted in a complexity $O\left(\epsilon^{-3}\right)$.

Chapter 8 - Lookback options. Finally, we devoted chapter 8 to lookback options. We showed that the analysis of the (mostly academic) case of continuously sampled lookback options was made particularly difficult by the fact that the fine and coarse paths used by our estimators didn't necessarily reach their minima at the same time. Noting that the prices of discretely sampled lookback options converge towards the prices of equivalent continuously sampled lookback options when the sampling frequency increases, we could get around the problem by studying the former. The results obtained for discretely sampled options were consistent with what we expected from continuously sampled options and we proved that the complexity of the multilevel Greeks estimator was $O\left(\epsilon^{-2}\left(\log\epsilon\right)^2\right)$.

9.2 Future work

We answered our initial questions. We have shown that multilevel Monte Carlo techniques could be used for the estimation of Greeks. We have proved it did provide significant complexity improvements over standard Monte Carlo.

Our analysis also left some questions open for future research. How can we rigorously analyse the choice of the optimal number of final samples for split pathwise sensitivities? How can we analyse the multilevel Vibrato Monte Carlo method? Once again how would we choose the optimal number of samples for the final time step?

This research also raised new questions which would have to be investigated in the future. In [Xia11], Giles and Xia investigate techniques for option pricing using multilevel Monte Carlo with jump processes. How would we adapt the methods we presented to work in the context of jump processes? As explained for example in [CG07], [FLL$^+$99b], [Ben03] or [GKH03] , the use of Malliavin calculus provides interesting hybrid estimators of Greeks. How could we use our ideas in conjunction with Malliavin techniques for an efficient computation of Greeks?

Appendix A

Numerical verification methods

In this appendix, we present some useful techniques that can be used to validate our code and check the correctness of our computations of multilevel Greeks. The techniques related to the numerical evaluation of the derivatives of functions defined by a computer program are collectively known as algorithmic differentiation or automatic differentiation techniques.

To illustrate the ideas, let's consider some function f specified by some computer code `function` that takes some real inputs: $input = (i_1, ..., i_M)$ and returns the real outputs: $output = (o_1, ..., o_N)$. Using MATLAB's notation,

```
input=[i1,...,iM];
output=[o1,..., oN];
output=function(i1,...,iM);
```

A.1 Finite difference

A naive technique to evaluate the sensitivity of some code's outputs with respect to its inputs is to use finite differences, also known as variable "bumping".

A.1.1 Principle

The idea is simply to "bump" by a small amount ϵ the k-th input. Indeed, assuming f is sufficiently differentiable, we can write the Taylor expansion:

$$f(..., i_k + \epsilon, ...) = f(..., i_k, ...) + \epsilon \, \frac{\partial f(..., i_k, ...)}{\partial i_k} + \frac{1}{2} \, \epsilon^2 \, \frac{\partial^2 f(..., i_k, ...)}{\partial i_k^2} + \dots \tag{A.1}$$

The k-th sensitivity can then be simply evaluated with an accuracy $O\left(\epsilon\right)$ via the forward difference scheme

$$\frac{\partial output}{\partial i_k} \approx \frac{f(..., i_k + \epsilon, ...) - f(..., i_k, ...)}{\epsilon} \tag{A.2}$$

that is, we compute

```
output=function(i1,...,ik,..., iM);
output_p=function(i1,...,ik+eps,...,iM);
```

and then,

$$\frac{\partial output}{\partial i_k} \approx \texttt{(output\_p-output)/eps}$$

Alternatively, a more accurate stencil known as a "central difference" uses

```
output_m=function(i1,...,ik-eps/2,..., iM);
output_p=function(i1,...,ik+eps/2,...,iM);
```

The k-th sensitivity is then evaluated as

$$\frac{\partial output}{\partial i_k} \approx \texttt{(output\_p-output\_m)/eps}$$

There is a rich literature on finite difference schemes. More details on the basic techniques presented here and higher order schemes can be found in [CB80] or [Duf06].

A.1.2 Complexity and limitations

We see that the computation of the Jacobian of f with the forward difference requires $M+1$ simulations and central difference requires $2M$ simulations. The cost is proportional to the number of inputs of the function.

The limitations of this method are also well-known and detailed in the literature (e.g. [CB80], [BSPSM81]). Let's recall the major issues associated with this technique.

Floating point arithmetics First, the accuracy of the scheme is clearly determined by the size of the "bump" ϵ; we should *a priori* take ϵ as small as possible to decrease the bias of our estimate. Nevertheless, the limitations of simple and double precision arithmetic prevent us from doing so. In practice, taking $\epsilon \approx 10^{-4} i_k$ or $\epsilon \approx 10^{-6} i_k$ is safe for single and double precision respectively and results in a small bias for continuous payoffs.

Discontinuous payoffs and variance explosion When pricing discontinuous payoffs, the variance of the finite difference estimator of the sensitivity explodes as $\epsilon \to 0$. Considering the example of the digital call with strike K (with a discontinuity around K), we can write that most samples are away from K and then, $f(S+\epsilon) - f(S) = O(\epsilon)$ and for a fraction $O(\epsilon)$ of all samples, $f(S+\epsilon) - f(S) = O(1)$, which leads to

$$\begin{aligned}
\mathbb{E}\left[\frac{f(S+\epsilon) - f(S)}{\epsilon}\right] &= O(1) \\
\mathbb{E}\left[\left(\frac{f(S+\epsilon) - f(S)}{\epsilon}\right)^2\right] &= O\left(\frac{1}{\epsilon}\right)
\end{aligned} \tag{A.3}$$

and therefore the variance of the estimator is $O\left(\epsilon^{-1}\right)$. The choice of ϵ is then the result of a tradeoff between bias and variance to minimise the estimator's mean square error.

A.2 Complex variable "trick"

An alternative to finite differences is the so-called "complex variable trick", as presented in [ST98].

A.2.1 Principle

To study the sensitivity with respect to some input, the idea is again to "bump" it, this time by a pure imaginary number $\mathrm{i}\epsilon$ (where $\epsilon \in \mathbb{R}$ and $\mathrm{i} = \sqrt{-1}$). For analytic functions with real values on the real line, we can write the beginning of the Taylor expansion

$$f(i_1, ..., i_k + \mathrm{i}\epsilon, ..., i_M) = f(i_1, ..., i_k + \mathrm{i}\epsilon, ..., i_M) + \mathrm{i}\epsilon \frac{\partial f}{\partial i_k} - \epsilon^2 \frac{1}{2} \frac{\partial^2 f}{\partial i_k^2} + O\left(\epsilon^3\right) \quad \text{(A.4)}$$

Therefore

$$\frac{\partial f}{\partial i_k} \approx \mathrm{Imag}\left[\frac{f(i_1, ..., i_k + \mathrm{i}\epsilon, ..., i_M)}{\epsilon}\right] + O\left(\epsilon^2\right) \quad \text{(A.5)}$$

The benefit of this approximation of the sensitivity is that its accuracy is of order $O\left(\epsilon^2\right)$, and also it doesn't involve a subtraction which would make the estimator vulnerable to floating point arithmetic errors.

A.2.2 Complexity and limitations

As before, this technique only gives one sensitivity at a time and we need to run it M times (bumping each input once) to get all sensitivities. The cost of running what is initially a real function on complex numbers is likely to be close to two times the cost of a "normal" run with real variables.

Language Another limitation of this technique is that it is obviously better adapted to programming languages with a native support of complex numbers (e.g. MATLAB). It is still possible to use this idea in other languages, provided we can define a complex type and the corresponding operations. This approach would actually be very similar to the idea of algorithmic differentiation via operator overloading (see [Gri89] or [BHN02]).

Definition of analytical extensions Some operations are not mathematically defined for complex numbers or not analytic (e.g. $\max, \min, |\,,\,| \ldots$). Some languages (e.g. MATLAB) sometimes define extensions but they rarely behave in the way we require. We therefore need to define our own analytic extensions.

Payoff discontinuities The technique gives accurate estimates of pathwise sensitivities, yet it does not necessarily mean it gives an accurate estimate of the Greeks. It does so only when pathwise sensitivities are applicable (see section 1.3.1).

Bibliography

[ABR96] L. Andersen and R. Brotherton-Ratcliffe. Exact exotics. *Risk*, 9:85–89, 1996.

[AG07] A. Asmussen and P. Glynn. *Stochastic Simulation*. Springer, New York, 2007.

[Ama05] K. Amano. A stochastic Gronwall inequality and its applications. *Journal of Inequalities in Pure and Applied Mathematics*, 6, 2005.

[ANvST14] Abdul Lateef Haji Ali, Fabio Nobile, Erik von Schwerin, and Raul Tempone. Optimization of mesh hierarchies in multilevel monte carlo samplers, 2014.

[Avi09a] R. Avikainen. Convergence rates for approximations of functionals of SDEs. *Finance and Stochastics*, 13(3):381–401, 2009.

[Avi09b] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13(3):381–401, 2009.

[BBG97] P. Boyle, M. Broadie, and P. Glasserman. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21(8-9):1267–1321, June 1997.

[BCI99] P. Baldi, L. Caramellino, and M.G. Iovino. Pricing complex barrier options with general features using sharp large deviation estimate. *Proceedings of the MCQMC Conference, Claremont (LA), USA*, 1999.

[BDZ97] D.R. Beaglehole, P.H. Dybvig, and G. Zhou. Going to extremes: Correcting simulation bias in exotic option valuation. *Fin. Anal. Jour.*, 53:62–68, 1997.

[Ben03] E. Benhamou. Optimal Malliavin weighting function for the computation of the Greeks. *Mathematical Finance*, 13(1):37–53, 2003.

[BG96] M. Broadie and P. Glasserman. Estimating security price derivatives using simulation. *Management Science*, 42(2):269–285, 1996.

[BG12] S. Burgos and M. Giles. Computing Greeks using multilevel path simulation. *Monte Carlo and Quasi Monte Carlo Methods 2010*, pages 281–296, 2012.

[BGK97] M. Broadie, P. Glasserman, and S. Kou. A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4):325–349, 1997.

[BGK99] M. Broadie, P. Glasserman, and S.G. Kou. Connecting discrete and continuous path-dependent options. *Finance and Stochastics*, 3(1):55–82, 1999.

[BH86] N. Bouleau and F. Hirsch. Formes de Dirichlet générales et densité des variables aléatoires réelles sur l'espace de Wiener. (General Dirichlet forms and density of real random variables on Wiener space). *J. Funct. Anal.*, 69:229–259, 1986.

[BHN02] Christian H. Bischof, Paul D. Hovland, and Boyana Norris. Implementation of automatic differentiation tools. In Peter Thiemann, editor, *PEPM*, pages 98–107. ACM, 2002.

[BHR12] K. Bujok, B. Hambly, and C. Reisinger. Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives. *ArXiv preprint: 1211.0707*, 2012.

[BSD13] D. Belomestny, J. Schoenmakers, and F. Dickmann. Multilevel dual approach for pricing American style derivatives. *Finance and Stochastics*, 2013.

[BSPSM81] P. Brezillon, J-F. Staub, A-M. Perault-Staub, and G. Milhaud. Numerical estimation of the first order derivative: approximate evaluation of an optimal step. *Computers and Mathematics with Applications*, 7(4):333 – 347, 1981.

[BT95] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations, I: convergence rate of the distribution function. *Probability Theory and Related Fields*, 104(1):43–60, 1995.

[CB80] S.D. Conte and C.W. De Boor. *Elementary Numerical Analysis: An Algorithmic Approach.* McGraw-Hill Higher Education, 3rd edition, 1980.

[CFN98] M-E. Caballero, B. Fernández, and D. Nualart. Estimation of densities and applications. *J. Theoret. Probab.*, 11(3):831–851, 1998.

[CG07] N. Chen and P. Glasserman. Malliavin Greeks without Malliavin calculus. *Stochastic Processes and their Applications*, 117:1689–1723, 2007.

[CGST11] K.A. Cliffe, M.B. Giles, R. Scheichl, and A.L. Teckentrup. Multilevel monte carlo methods and applications to elliptic pdes with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.

[CV91] A. Conze and Viswanathan. Path dependent options: The case of lookback options. *Journal of Finance*, 46(5):1893–907, 1991.

[DG95] D. Duffie and P. Glynn. Efficient Monte Carlo simulation of security prices. *The Annals of Applied Probability*, 5(4):897–905, 1995.

[DH11] S. Dereich and F. Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Processes and their Applications*, 121(7):1565–1587, 2011.

[DM10] S. De Marco. *On probability distributions of diffusions and financial models with non-globally smooth coefficients.* PhD thesis, Université Paris-Est, 2010.

[Duf06] D. Duffy. *Finite Difference Methods in Financial Engineering: A Partial Differential Equation Approach.* John Wiley & Sons, 2006.

[Fam76] Eugene F. Fama. *Foundations of finance.* Basic Books, New York, NY, 1976.

[FCKSS13] A. Ferreiro-Castilla, A.E. Kyprianou, R. Scheichl, and G. Suryanarayana. Multilevel monte carlo simulation for lévy processes based on the wiener-hopf factorisation. *ArXiv*, 2013.

[FCvS13] A. Ferreiro-Castilla and K. van Schaik. Applying the wiener-hopf monte carlo simulation technique for levy processes to path functionals such as first passage times, undershoots and overshoots. *ArXiv*, 2013.

[FLL+99a] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3:391–412, 1999.

[FLL+99b] E. Fournie, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of malliavin calculus to monte carlo methods in finance. *Finance and Stochastics 3*, pages 391–412, 1999.

[FLL+01] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance, II. *Finance and Stochastics*, 5:201–236, 2001.

[Fri64] A. Friedman. Partial differential equations of parabolic type, 1964.

[GDR13] M.B. Giles, K. Debrabant, and A. Roessler. Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation. *ArXiv preprint: 1302.4676*, 2013.

[GG06] M.B. Giles and P. Glasserman. Smoking adjoints: fast Monte Carlo Greeks. *RISK*, January 2006.

[GH13] T. Gerstner and S. Heinz. Dimension- and time-adaptive multilevel Monte Carlo methods. In *Sparse Grids and Applications*, volume 88 of *Lecture Notes in Computational Science and Engineering*, pages 107–120. Springer, 2013.

[GHM09] M. Giles, D. Higham, and X. Mao. Analysing multi-level monte carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3):403–413, 2009.

[Gil07] M.B. Giles. Monte Carlo evaluation of sensitivities in computational finance. Technical Report NA07/12, 2007.

[Gil08a] M.B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer-Verlag, 2008.

[Gil08b] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[Gil09a] M.B. Giles. Multilevel Monte Carlo for basket options. In M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference*, pages 1283–1290. IEEE, 2009.

[Gil09b] M.B. Giles. Vibrato Monte Carlo sensitivities. In P. L'Ecuyer and A. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pages 369–382. Springer, 2009.

[GKB98] P. Glasserman, S.G. Kou, and M. Broadie. Connecting discrete and continuous path-dependent options. *Finance and Stochastics*, 3(1):55–82, 1998.

[GKH03] E. Gobet and A. Kohatsu-Higa. Computation of Greeks for barrier and lookback options using Malliavin calculus. *Electron. Commun. Probab.*, 8, 2003.

[GL94] J.G. Gaines and T.J. Lyons. Random generation of stochastic integrals. *SIAM Journal of Applied Mathematics*, 54(4):1132–1146, 1994.

[Gla88] P. Glasserman. Performance continuity and differentiability in Monte Carlo optimization. In *Proceedings of the 20th conference on Winter simulation*, WSC '88, pages 518–524, New York, NY, USA, 1988. ACM.

[Gla04] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.

[GN13] T. Gerstner and M. Noll. *Randomized multilevel quasi-Monte Carlo path simulation*, pages 349–372. World Scientific, 2013.

[GR12] M.B. Giles and C. Reisinger. Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance. *SIAM Journal of Financial Mathematics*, 3(1):572–592, 2012.

[Gri89] Andreas Griewank. On automatic differentiation. In *IN MATHEMATICAL PROGRAMMING: RECENT DEVELOPMENTS AND APPLICATIONS*, pages 83–108. Kluwer Academic Publishers, 1989.

[GS72] I.I. Gikhman and A.V. Skorokhod. *Stochastic differential equations*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag, 1972.

[GS01] P. Glasserman and J. Staum. Conditioning on one-step survival for barrier option simulations. *Operations Research*, 49:2001, 2001.

[GS13] M.B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Annals of Applied Probability*, 2013.

[GSG79] M.B. Goldman, H.B. Sosin, and M.A. Gatto. Path dependent options: "buy at the low, sell at the high". *Journal of Finance*, 34(5):1111–27, 1979.

[GW09] M.B. Giles and B.J. Waterhouse. Multilevel quasi-Monte Carlo path simulation. In *Advanced Financial Modelling*, Radon Series on Computational and Applied Mathematics, pages 165–181. De Gruyter, 2009.

[HK95] R.C. Heynen and H.M. Kat. Lookback options with discrete and partial monitoring of the underlying price. *Applied Mathematical Finance*, 2(4):273–284, 1995.

[HK04] P. Hunt and J. Kennedy. *Financial Derivatives in Theory and Practice*. Wiley Series in Probability and Statistics. Wiley, 2004.

[HN84] J. Hutton and P. Nelson. Interchanging the order of differentiation and stochastic integration. *Stochastic Processes and their Applications 18*, pages 371–377, 1984.

[HSST12] H. Hoel, E. von Schwerin, A. Szepessy, and R. Tempone. Adaptive multilevel Monte Carlo simulation. In B. Engquist, O. Runborg, and Y.-H.R. Tsai, editors, *Numerical Analysis of Multiscale Computations*, number 82 in Lecture Notes in Computational Science and Engineering, pages 217–234. Springer-Verlag, 2012.

[Kal02] Olav Kallenberg. *Foundations of modern probability.* Probability and its applications. Springer, New York, Berlin,, Paris, 2002. Sur la 4e de couv. : This new edition contains four new chapters as well as numerous improvements throughout the text.

[Kar83] R. Karandikar. Interchanging the order of stochastic integration and ordinary differentiation. *Sankhya: The Indian Journal of Statistics, Series A (1961-2002), Vol. 45, No. 1*, pages 120–124, 1983.

[KKPS11] A. Kuznetsov, A.E. Kyprianou, J-C. Pardo, and K. Van Schaik. A Wiener-Hopf Monte Carlo simulation technique for Lévy processes. *Annals of Applied Probability*, 21(6):2171–2190, December 2011.

[Kou07] S.G. Kou. *Handbooks in Operations Research and Management Science*, volume Volume 15, chapter Chapter 8. Discrete Barrier and Lookback Options, pages 343–373. Elsevier, 2007.

[KP92] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1992.

[L'E90] P. L'Ecuyer. A unified view of the IPA, SF and LR gradient estimation techniques. *Management Science*, 36(11):1364–1383, 1990.

[Mil79] G. Milstein. A method of second-order accuracy integration of stochastic differential equations. *Theory of Probability and Its Applications*, 23(2):396–401, 1979.

[Mil95] G.N. Milstein. *Numerical Integration of Stochastic Differential Equations.* Environmental Science and Technology. Springer, 1995.

[MR05] M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling.* Stochastic Modelling and Applied Probability. Springer, 2005.

[Nua05] D. Nualart. *The Malliavin Calculus and Related Topics*, chapter 2. Springer, 2nd edition, 2005.

[Pro90] P. Protter. *Stochastic integration and differential equations.* Springer, 1990.

[Rud86] W. Rudin. *Real and Complex Analysis.* McGraw-Hill Science/Engineering/Math, 1986.

[RW01] T. Rydén and M. Wiktorsson. On the simulation of iterated Itô integrals. *Stochastic Processes and their Applications*, 91(1):151–168, January 2001.

[ST98] W. Squire and G. Trapp. Using complex variables to estimate derivatives of real functions. *SIAM Rev.*, 40(1):110–112, mar 1998.

[Tal84] D. Talay. Efficient numerical schemes for the approximation of expectations of functionals of the solution of an SDE, and applications. In H. Korezlioglu, G. Mazziotto, and J. Szpirglas, editors, *Filtering and Control of Random Processes*, volume 61 of *Lecture Notes in Control and Information Sciences*, pages 294–313. Springer Berlin Heidelberg, 1984.

[WHD95] P. Wilmott, S. Howison, and J. Dewynne. *The mathematics of financial derivatives: a student introduction*. Cambridge University Press, 1995.

[Wik01] M. Wiktorsson. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Annals of Applied Probability*, 11(2):470–487, 2001.

[Wil07] P. Wilmott. *Paul Wilmott Introduces Quantitative Finance*. Wiley Desktop Editions. Wiley, 2007.

[Wil10] P. Wilmott. *Frequently Asked Questions in Quantitative Finance*. Wiley, 2010.

[XG12] Y. Xia and M.B. Giles. Multilevel path simulation for jump-diffusion SDEs. In L. Plaskota and H. Woźniakowski, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2010*, pages 695–708. Springer-Verlag, 2012.

[Xia11] Y. Xia. Multilevel Monte Carlo method for jump-diffusion SDEs. *ArXiv e-prints*, June 2011.

[Zor04] V. A. Zorich. *Mathematical analysis. I*. Universitext. Springer-Verlag, Berlin, 2004. Translated from the 2002 fourth Russian edition by Roger Cooke.