Numerical schemes and Monte Carlo techniques for Greeks in stochastic volatility models

A thesis presented for the degree of

Doctor of Philosophy of Imperial College London

by

Ivo Mihaylov

Department of Mathematics

Imperial College London

London SW7 2AZ

NOVEMBER 12, 2015

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed:

Copyright

Abstract

The main objective of this thesis is to propose approximations to option sensitivities in stochastic volatility models. The first part explores sequential Monte Carlo techniques for approximating the latent state in a Hidden Markov Model. These techniques are applied to the computation of Greeks by adapting the likelihood ratio method. Convergence of the Greek estimates is proved and tracking of option prices is performed in a stochastic volatility model.

The second part defines a class of approximate Greek weights and provides high-order approximations and justification for extrapolation techniques. Under certain regularity assumptions on the value function of the problem, Greek approximations are proved for a fully implementable Monte Carlo framework, using weak Taylor discretisation schemes. The variance and bias are studied for the Delta and Gamma, when using such discrete-time approximations.

The final part of the thesis introduces a modified explicit Euler scheme for stochastic differential equations with non-Lipschitz continuous drift or diffusion; a strong rate of convergence is proved. The literature on discretisation techniques for stochastic differential equations has been motivational for the development of techniques preserving the explicitness of the algorithm. Stochastic differential equations in the mathematical finance literature, including the Cox-Ingersoll-Ross, the 3/2 and the Ait-Sahalia models can be discretised, with a strong rate of convergence proved, which is a requirement for multilevel Monte Carlo techniques.

To my family.

Acknowledgement

Foremost, I would like to acknowledge the patience, motivation, enthusiasm and knowledge of Dr Jean-François Chassagneux, and highlight the constant support throughout my research. I equally thank Dr Antoine Jacquier for being my second supervisor—for the guidance, ideas, patience and taking me on as a student. It was great to have the ability to learn from you both, and to spend so many hours discussing the research.

I would like to thank members of the mathematical finance, stochastic analysis and statistics groups in Imperial College London. The environment provided by you has been critical for my development. In addition, I thank my friends for the great moments shared during the thesis and the mathematical (and non-mathematical!) discussions.

I would like to thank my parents, Mihail Mihaylov and Stoia Mihaylova, for their sacrifices and constant support, without which this thesis would have been impossible. My grandparents have been a great inspiration to me, and I am grateful for their serenity, patience and valuable lessons in life. I thank my brother, Dr Preslav Mihaylov, for always being such an inspiration to me, in addition to being such a great friend. Finally, I thank Julia for her understanding and help, and for always believing in me.

Contents

List of Figures

List of Tables

Introduction

This thesis is split in three parts, with links tying numerical methods, stochastic analysis, statistics and finance together. A focus throughout is the use of stochastic differential equations for the modelling of financial instruments and the volatility of a process. A recurring theme is the computation of option price sensitivities, referred to as the Greeks, in various frameworks and models. The Greeks are a practical necessity for trading, hedging and risk-warehousing of financial products. Each chapter concludes with notes on applications to mathematical finance.

The first part of the thesis explores particle filtering techniques, also referred to as Sequential Monte Carlo (SMC). The filtering problem has its origins in signal processing and estimates a hidden state based on observations of a noisy system. A Hidden Markov Model (HMM) set-up is considered in which the asset price is an observable process, and the volatility is the latent driving process of the asset price. Greeks are approximated using a likelihood ratio method, where a smoothing algorithm is applied to approximate the score function (derivative of the log-likelihood of the density function given a set of observations). The method relies on a forward step of the hidden and observed variables to generate a sample path of the observed process; this is followed by a backward pass to compute the score function using particle filtering. A forward-only implementation is considered for applications. It is shown that Greeks in a stochastic volatility framework can be computed using this approach, and convergence results are adapted for such applications.

The second part of the thesis introduces a general technique for approximating option price sensitivities. There are closed-form solutions under some modelling assumptions; Monte Carlo, trees, quadrature (Fourier) and finite-difference methods have been exploited for approximating option prices in full generality. Pricing involves a forward process describing the asset price evolving through time and a backward component describing the option value with appropriate terminal conditions representing the payoff. The option price is the solution to a partial differential equation (PDE), with appropriate boundary conditions. The aim is to compute Greeks alongside option prices by exploring the value function of the PDE and by finding suitable weights. This is achieved by multiplying the payoff by a functional of the increments of the driving Brownian motion. Convergence results are studied with an emphasis on the smoothness requirements of the value function.

The third part of the thesis studies numerical schemes for discretising stochastic differential equations driven by Brownian motion. The focus is to move away from the classical setting where the drift and diffusion functions are assumed to be globally Lipschitz continuous. Such stochastic differential equations are integral to the modelling of financial markets, with the aim of improving the fit of volatility smiles and term-structure exhibited by option prices. A modified explicit Euler scheme is introduced to approximate scalar stochastic processes, for which strong rates of convergence are proved. A family of SDEs considered include those with solutions defined in a domain. For applications inspired by finance this domain is typically the positive half-line (in the case of asset prices, volatility, intensity rates), but can be generalised. Applications include the CIR model, the 3/2 model and the Ait-Sahalia model. A demonstration of multilevel Monte Carlo (MLMC) techniques allows this modified Euler scheme to be used efficiently.

0.1 Preliminaries

The seminal thesis of Bachelier, Einstein's introduction of Brownian motion to physics and the work by Wiener and Lévy provided the foundations of the modern analysis of related topics that followed. An m-dimensional Brownian motion $W = (W_t)_{t\geq 0}$ is an adapted stochastic process on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ such that $W_t(\omega) : [0, \infty) \times \Omega \to \mathbb{R}^m$. It is a Gaussian process, with continuous sample paths. For all times $s < t$, it follows that $W_t - W_s$ is independent from the filtration $\mathcal{F}_s$ (assumed to be right continuous and containing all $\mathbb{P}$-null sets). As this thesis is largely concerned with the simulation of stochastic processes, Brownian motion is an important building block for applications. By having an independent and identically distributed (i.i.d.) sequence of random variables with mean zero and unit variance (readily generated by tossing a coin!), a trajectory which converges in the weak sense to the distribution of a Brownian motion can be constructed.

A process X defined on the probability space equipped with the natural filtration of X, namely $\mathcal{F}_t^X := \sigma(X_u | u \in [0, t])$ (the sigma-algebra generated by the process), is said to be Markovian if and only if for all bounded, measurable functions φ we have that $\mathbb{E}[\varphi(X_t)|\mathcal{F}_s^X] = \mathbb{E}[\varphi(X_t)|X_s]$ for all $s \leq t$; in other words a Markovian process is memoryless and the process in the future only depends on the knowledge at the present time. Later, hidden Markov models for the evolution of the asset prices and their driving volatility processes will be considered. The use of stochastic differential equations is present in a wide range of applications in the natural sciences, economics and finance. A time-homogeneous Itô diffusion in $\mathbb{R}^d$ is a solution to the stochastic differential equation

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + \gamma(X_t)\mathrm{d}W_t\,, \qquad X_0 = x \in \mathbb{R}^d\,, \qquad \forall t \geq 0\,, \tag{0.1.1}$$

for some $f : \mathbb{R}^d \to \mathbb{R}^d$, $\gamma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$. A strong solution of the above SDE is a continuous process X, adapted to the natural filtration of the Brownian motion W, and for all $t \geq 0$ it holds that

$$\int_0^t \left(|f(X_u)| + |\gamma(X_u)|^2\right) \mathrm{d}u \tag{0.1.2}$$

is finite, almost surely. Furthermore, with probability one for all $t \geq 0$, it holds that

$$X_t = x + \int_0^t f(X_u)\mathrm{d}u + \int_0^t \gamma(X_u)\mathrm{d}W_u. \tag{0.1.3}$$

By imposing Lipschitz continuity and linear growth conditions on the drift and diffusion functions, existence and uniqueness of a strong solution are guaranteed.

The notion of a weak solution to the SDE is the triple consisting of the filtered probability space, the $(\mathcal{F}_t)$-Brownian motion W and $\mathcal{F}_t$-progressively measurable process X satisfying the stochastic differential equation with probability one, and being such that (0.1.2) is finite, a.s. for all $t \geq 0$.

0.1.1 Financial option theory

In the mathematical finance literature, an option is a contract between two parties with value based on the future price of an underlying price. A buyer of an option has the right to exercise the contract, but is under no obligation to engage in a transaction. Option specifications are typically described in a term sheet, with characteristics including the exercise type and the payoff. The exercise type describes how the option is exercised (European, American, Bermudan) [Hul14]. The payoff function of the option is based on the price of the underlying instrument throughout its lifetime and on key parameters such as the strike price, K; it can be a combination of path dependence, barriers, Asian/averaging, look-back, digital/binary and many other flavours. Replication of such a derivative, is the formation of a self-financing, hedging strategy. A portfolio is self-financing if there are no external infusions or withdrawals of capital. By the principle of no-arbitrage opportunities, a self-financing portfolio which perfectly replicates the payoff of a derivative has the same value as the derivative. Consider an asset price process $X = (X_t)_{t\geq 0}$. A European call option is a contract that gives the holder the right to purchase one unit of the asset at a fixed strike price at a fixed expiry time, T. The terminal payoff is thus $\max(X_T - K, 0)$. A European put gives the right to sell the stock at a strike price, i.e. the terminal payoff is $\max(K - X_T, 0)$. Options can be exchange traded (typical for vanilla options) or "over-the-counter" transactions for bilateral transactions. The latter tend to be at the more exotic spectrum of products and consist of specialised option transactions.

The Black-Scholes setting is a quoting mechanism and an important modelling framework arising from the seminal paper [BS73]. The model assumes that the underlying asset price is log-normally distributed with a constant drift and volatility. In addition, it makes several assumptions such as the Efficient Market Hypothesis, infinite liquidity of markets, price-continuity, lack of transaction costs and the ability to trade continuously. In the Black-Scholes setting, the underlying asset evolves through the stochastic differential equation (0.1.1), with $f(x) \equiv \mu x$ and $\gamma(x) \equiv \sigma x$, for some constant drift parameter, μ, and some strictly positive volatility, σ. The construction of a self-financing portfolio and the put-call parity are vital concepts in derivatives pricing and structuring, especially given the importance of calls and puts as building blocks for more exotic products.

Itô's Lemma states that for a stochastic process X satisfying (0.1.1) and some functional $F : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ taken to be sufficiently smooth ($F \in \mathcal{C}^{1,2}$), then

$$\mathrm{d}F = \left(\frac{\partial F}{\partial t} + f \frac{\partial F}{\partial X_t} + \frac{1}{2}\gamma^2 \frac{\partial^2 F}{\partial X_t^2} \right) \mathrm{d}t + \gamma \frac{\partial F}{\partial X_t} \mathrm{d}W_t \, ,$$

implying that F is itself an Itô process. The Black-Scholes PDE can be derived using Itô's Lemma and closed-form calculations for the price of vanilla products are well known [Hul14]. In general, let $X = (X_t)_{t\geq 0}$ be the solution to (0.1.1), $g : \mathbb{R}^d \to \mathbb{R}$ be a payoff function at time T and define the option price $V(x)$, as the expectation given the initial condition $X_0 = x$,

$$V(x) := \mathbb{E}\left[g(X_T) | X_0 = x \right] \, . \tag{0.1.4}$$

0.1.2 Stochastic Volatility

The instantaneous volatility of the underlying asset price is an important consideration when modelling the dynamics. A stochastic volatility model is a way of introducing dynamics to the volatility driving the asset price and several stochastic volatility models are reviewed. Options can have extreme sensitivities to future volatility levels; for example cliquet contracts have higher sensitivity to volatility, compared to plain European contracts [Wil01, IO05]. The asset price is typically known, however the volatility of an instrument is not directly observed and later a HMM is adapted to the computation of Greeks. The volatility of a financial asset exhibits variability over time, so it is intuitive to consider a stochastic process for its evolution. There are numerous ways to model the future instantaneous volatility [Cox75, HW87, Sco87, Hes93, Cox96, Wil01]. The simplest suggestion is to treat the volatility σ_t as a Brownian walk or a geometric Brownian motion. An undesirable outcome of treating the volatility process as a random walk is that it could become negative.

The Feller diffusion is a mean-reverting stochastic process [Fel54], defined as the unique strong solution to

$$\mathrm{d}v_t = \kappa(\theta - v_t)\mathrm{d}t + \xi\sqrt{v_t}\mathrm{d}W_t, \qquad v_0 = v > 0, \tag{0.1.5}$$

where W is a Brownian motion and κ, θ, ξ are strictly positive constant parameters (also referred to as the CIR process, named after Cox, Ingersoll and Ross [CIR85]). This process has been widely used in the mathematical finance literature, both for interest rate modelling and as dynamic for the instantaneous variance of a stock price as in the Heston model [CIR85, Hes93, JM11].

In financial markets and options, the skew represents the slope of the implied volatility curve for a given expiration date and the term-structure refers to the implied volatility for different expiration dates. Various types of skew in option prices stem from liquidity constraints, regulatory rules, government intervention and dependence on the asset class [Wil06]. A common feature of stochastic volatility models is the correlation between the Brownian motions of the underlying instrument and of the driving volatility. This correlation is often necessary in calibration in order to fit the skew commonly exhibited in equities, FX and interest rate markets; parameters are fitted to market prices of calls and puts, or other liquid instruments [MN03]. For example, in the Hull-White model both the asset and the volatility follow a geometric Brownian motion with correlated Brownian motion drivers.

The Heston model is a stochastic volatility model where the instantaneous variance follows a Feller diffusion [Hes93]. In this model, option prices admit (semi-)closed form solutions and sensitivities. In practice during market calibration, it is often found that the speed of mean-reversion, κ, is small, as high values of κ reduce the skew exhibited by the model. Additionally, calibration often suggests that the volatility of volatility, ξ, is large. The Feller condition ensures that the variance process in (6.4.1) is positive; if $2\kappa\theta > \xi^2$, then $\mathbb{P}(v_t = 0) = 0$ for all $t \geq 0$ [Fel54]. This makes the Feller condition difficult to satisfy in practice when calibrating to market data [Jac05]. The correlation parameter between the driving Brownian motion of the underlying and variance, ρ, is often negative, because a decrease in the underlying price is often associated with an increase in the variance.

The constant elasticity of variance (CEV) is a stochastic process used for modelling assets using an elasticity factor, $0 \leq \alpha \leq 2$ [Cox75, CR76, Cox96]. The CEV process is the solution to the following SDE for $t \geq 0$:

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t^{\alpha/2} \mathrm{d}W_t \,, \qquad X_0 = x \,,$$

with instantaneous variance of $\mathrm{d}X_t / X_t$ being $\sigma^2 X_t^{\alpha-2}$. Note that when $\alpha = 2$, this is just the Black-Scholes model. Additionally, the instantaneous variance is inversely proportional to the underlying, making the model particularly suitable for fitting empirical data [Bec80]. Using a geometric Brownian motion for the volatility and a CEV process for the underlying asset, the so-called SABR model is widely used in the interest rates industry [HKLW02].

There has been a growing interest in stochastic volatility models in all areas of mathematical finance in recent years. Important considerations when comparing stochastic volatility models are the ability to compute option prices in a closed or semi-closed form, the ability to fit market-observable phenomenon such as skew and finally the ease of calibration to the market. Under the Heston model, European options can be computed efficiently using Fast Fourier Transform algorithms; in addition the model can reproduce a wide range of volatility surfaces implied from the markets [MN03]. The CIR component of the Heston model often breaches the Feller condition when calibrating to the market. Additionally, the original Heston model struggles to create a skew as large as that observed in the market for small T [Gat11, Chapter 5]. It is often necessary to consider time-dependent parameters in order to perform well on calibrating for a

large set of options with different maturities and strikes, over a long period of time [MN03].
Stochastic models for the volatility are required because option prices calculated with simple models are generally not supported by market prices for the whole range of strikes and maturities. As a result, calibration is in practice performed daily, suggesting time-dependent parameters.
In this thesis, a model-agnostic framework is considered for approximating the Greeks using several techniques, focusing on models that have a (semi-)closed form solution to verify the computations; the finite difference methods and Monte Carlo simulations shall also be considered. There are alternative stochastic volatility models that could be studied such as the Scott model, stochastic volatility jump-diffusion processes (SVJD) or SABR alternatives [Sco87, CLS99, RV08].

0.1.3 Discretisation of SDEs

In situations where the solution of an SDE cannot be written in a closed-form, it is important to approximate the solution akin to the numerical integration literature. Let $n \in \mathbb{N}^+$ be a fixed positive integer and $T > 0$ a fixed time horizon. Define the partition of the interval $[0, T]$ by $\pi := \{0 = t_0 < t_1 < \ldots < t_n = T\}$, with $\max_{i=0,\ldots,n-1}(t_{i+1} - t_i) =: h = \mathcal{O}(1/n)$. A first-order approximation is the Euler-Maruyama approximation, when a grid π is used to create an approximation $\hat{X}$ of X, defined via

$$\hat{X}_{t_{i+1}} = \hat{X}_{t_i} + f(\hat{X}_{t_i})h_{i+1} + \gamma(\hat{X}_{t_i})\Delta W_{i+1} \,, \qquad \hat{X}_0 = x \,,$$

where $\Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}$ and $h_{i+1} := t_{i+1} - t_i$, which can be interpolated linearly for all $t \in [0, T]$. The quality of the approximation improves with increasing n, although errors can potentially propagate and explode in certain scenarios. The measures of error are either based on the strong error — how close the process X is tracked by the approximation $\hat{X}$ — or the error in the distributional sense of a particular function. The strong error is referred to as $\mathbb{E}\big[|X_T - \hat{X}_T|\big]$, and the focus of Part III is proving strong rates of convergence for families of SDEs with non-classical assumptions. Sufficient conditions are imposed, so that for a linearly interpolated approximation $\hat{X}$ it holds that for $h > 0$ small enough

$$\mathbb{E}\big[|X_t - \hat{X}_t|^p\big]^{1/p} \leq C_p h^r \,, \qquad \forall t \in [0, T] \,,$$

with some rate of convergence, $r > 0$, for some $p \geq 1$. The weak error for a function g, defined as $|\mathbb{E}[g(X_T)] - \mathbb{E}[g(\hat{X}_T)]\,|$, is a key measure when the focus is evaluating functional driven by diffusion processes. As this is often a requirement in financial derivatives pricing, it will be the focus of Part II.

An explicit Euler discretisation of the instantaneous variance in (6.4.1) on the partition π is:

$$\hat{v}_{t_{i+1}} = \hat{v}_{t_i} + \kappa(\theta - \hat{v}_{t_i})h_{i+1} + \xi\sqrt{\hat{v}_{t_i}}\Delta W_{i+1}\,, \qquad \hat{v}_0 = v\,.$$

Assume that $\hat{v}_{t_i}$ is strictly positive; by conditioning, the probability of the discretised process being negative at time t_{i+1} reads

$$\mathbb{P}(\hat{v}_{t_{i+1}} < 0 | \hat{v}_{t_i} > 0) = \mathbb{P}\left(\Delta W_{i+1} < \frac{\kappa(\hat{v}_{t_i} - \theta)h_{i+1} - \hat{v}_{t_i}}{\xi\sqrt{\hat{v}_{t_i}}}\middle| \hat{v}_{t_i} > 0\right) = \Phi\left(\frac{\kappa(\hat{v}_{t_i} - \theta)h - \hat{v}_{t_i}}{\xi\sqrt{\hat{v}_{t_i}h_{i+1}}}\right)\,,$$

where Φ is the cumulative density function of a standard normal distribution. The probability of a negative variance approximation is positive, even if the Feller condition holds. Upon discretisation, it is possible for approximations to become negative since the continuous-time variance process is approximated with a discrete-time Gaussian process. In an extreme scenario, observe that as ξ gets larger, the probability of a negative approximation for the variance process approaches $1/2$. Enforcing $\max(\hat{v}_t, 0)$ ensures that the instantaneous variance is non-negative and is a possible solution. The emphasis of Part III is to consider a modification of the explicit Euler scheme, for which a strong rate of convergence is proved.

Classical weak and strong convergence results for discretisation schemes of SDEs assume that the drift and the diffusion coefficients are globally Lipschitz continuous (see [KP92]); however many models in the literature violate this assumption e.g. CIR, CEV, Ait-Sahalia models. Typically, in financial derivative pricing weak error is sufficient for applications. Strong convergence rates are important when using multilevel Monte Carlo methods, as the strong rate of convergence can be used to optimise computation of functionals [Gil08b, GHM09].

0.1.4 Greeks

One aim of the thesis is to approximate Greeks for a wide class of stochastic volatility models. A necessity in option trading is the fast and reliable computation of sensitivities of financial derivatives. These sensitivities shown in Table 1 are computed with respect to parameters

	Spot (x)	Volatility	Expiry (T)	Interest rate (r)
Value (V)	Delta (Δ)	Vega ($\mathcal{V}$)	Theta (Θ)	Rho
Delta (Δ)	Gamma (Γ)	Vanna	Charm	
Vega ($\mathcal{V}$)	Vanna	Vomma	Veta	
Gamma (Γ)	Speed	Zomma	Color	

Table 1: Greeks: price and risk (row headings) differentiated with respect to the underlying parameter (column headings).

intrinsic to the option contract, such as the initial underlying price or expiry time of the contract, as well as parameters arising from the modelling assumptions, or parameters from the stochastic volatility model. Greeks are hedging ratios that explain how the profit and loss of a position evolve with changes in the market. Their computation is well studied using different mathematical techniques (for a comprehensive treatment refer to [Gla03, Hul14]). Closed-form Greeks for the Bachelier and Black-Scholes models are known, and Greeks can be computed in semi-closed form for the Heston model [BS73, Hes93]. Monte Carlo methods are commonly used to compute option prices and Greeks through simulation, often making use of classical variance reduction techniques [Cap08, Gla03]. In recent years, Malliavin-inspired techniques have allowed efficient Monte Carlo schemes for Greek computation [Ben01, FLL$^+$99].

0.1.5 Monte Carlo techniques

Monte Carlo techniques approximate solutions of problems that have difficult or intractable analytical solutions. Suppose that $\mathbb{P}$ is a probability measure on some measurable space $(\Omega, \mathcal{F})$, and X is a random variable with support $\mathbb{R}$. Monte Carlo methods are commonly used as a tool for integration, where for example we are interested in the expectation of a random variable with respect to the probability measure, $\mathbb{P}$, or of a functional g. By generating $\{x^{(i)}\}_{i=1,\ldots,N}$, i.i.d. random samples of X according to $\mathbb{P}$, we can approximate the integral

$$I(g) := \mathbb{E}_{\mathbb{P}}[g(X)] = \int_{\mathbb{R}} g(x)\mathbb{P}(x)\mathrm{d}x$$

by $\hat{I}^N(g) := N^{-1}\sum_{i=1}^N g(x^{(i)})$. The Law of Large Numbers, makes convergence of $\hat{I}^N(g)$ to $I(g)$ precise. Provided that the variance, $\mathbb{V}_{\mathbb{P}}[g(X)]$, is finite, the Central Limit Theorem implies that $\sqrt{N}\left(\hat{I}^N(g) - I(g)\right)$ converges in distribution to $\mathrm{N}(0, \mathbb{V}_{\mathbb{P}}[g(X)])$. The convergence rate is $\mathcal{O}(1/\sqrt{N})$, independent of the dimension. In multi-dimensional settings, Monte Carlo is superior to numerical integration [RC05]. Generating i.i.d. samples from $\mathbb{P}$ can be difficult if the probability measure is known only up to a normalising constant. Two methods to handle such a problem are rejection sampling and importance sampling. As rejection sampling is usually only possible in low dimensions, therefore importance sampling will be relied upon for the sequential Monte Carlo methods considered in Chapter 1. The popularity of Monte Carlo has grown due to its versatility; the ability to be used as a method to integrate, optimise and deal with non-linear problems.

One of the main tasks in mathematical finance is the pricing of option derivatives. Typically, the underlying assets are modelled by multi-dimensional SDEs, which rarely admit closed-form solutions and need to be numerically simulated. Therefore, Monte Carlo techniques are used to approximate the prices of options, by simulating sample paths of the underlying assets and estimating functionals to price the financial derivatives of interest (see [Gla03] for a comprehensive overview of such methods with applications to financial engineering). A Monte Carlo approximation of the option price using N simulated trajectories (assuming that the process can be simulated), where path j is denoted by $(X_t^{(j)})_{t\in[0,T]}$, is computed by $V^N(x) := N^{-1}\sum_{j=1,\dots,N} g(X_T^{(j)})$.

0.2 Contributions of this thesis

In this thesis simulation techniques for stochastic differential equations inspired by applications in mathematical finance are developed. The thesis is split into three distinct parts and the contributions are as follows.

The first part of the thesis studies sequential Monte Carlo techniques. Chapter 1 begins with an introduction of the sequential Monte Carlo methodology, with a focus on sampling from a sequence of posterior densities. An observed sequence conditional on a latent process is assumed, to infer the posterior density. Smoothing algorithms are presented to approximate the density in a Hidden Markov Model. The main contribution of the chapter is to consider a novel approach for approximating Greeks using such smoothing algorithms in a setting of unobserved stochastic volatility. This extends the work on SMC methods for option pricing [JDM10], where the use of smoothing algorithms is suggested (but not pursued) for approximating the Greeks. The score vector for a given realisation of an underlying price path is inferred, in order to compute the Greeks under a general stochastic volatility setting. The technique is analytically intractable for most models, therefore SMC is used to perform Bayesian inference. Inspired by likelihood ratio techniques for Greeks, additive functions that appear due to the structure of the models are derived. Upon simulating a volatility path and an underlying path, the volatility is immediately "forgotten". Filtering then recovers a particle approximation of the density, in order to approximate the log-likelihood and the score vector. Such a set-up lends itself to further considerations about hedging of derivatives in general stochastic volatility models. The approach highlights the difficulty in approximating Greeks upon observing the underlying, and moving away from a volatility process behaving as a discrete-space Markov chain. In this set-up, an existing SMC algorithm for smoothing is applied and provides a framework to approximate Greeks [DGA00, DMDS09]. Such techniques are used in the parameter estimation literature [Poy06, Poy11]. Using this approach, convergence results for the Greek estimates are proved in terms of the number of Monte Carlo paths, the number of particles and the number of time steps; the theoretical results are confirmed by numerical examples. The application discusses the tracking error for options using Black-Scholes Greeks and Greeks in a stochastic volatility model. A major drawback of such techniques for Greek approximations is the numerical cost compared with the various alternatives. SMC algorithms are numerically intensive; their inherent propensity to parallelisation has been a well-studied topic in recent years, however there remain challenges in using such techniques.

In Part II, a general technique is proposed to compute option Greeks using Itô-Taylor expansions. The aim is to multiply the payoff by some $\mathcal{F}_h$-measurable weight, for a small time h—this differs from the Malliavin setting, in which the weight is $\mathcal{F}_T$-measurable. The variance of the weights increases as h decreases and for convergence the mean squared error (MSE) is controlled. This technique allows Greeks to be approximated under various stochastic volatility models. In Chapter 3, a numerical approximation is demonstrated for the Delta of a contingent claim. An approximate class of weights are considered to compute high-order approximations of the Delta using weak Taylor schemes. Furthermore, by deriving expansions of these approximations, high-order Greek approximations are extrapolated. In Chapter 4, a family of functions for approximate weights is introduced for the Gamma, with higher-order and extrapolated approximations computed. This part concludes with proposing several directions for future research. Greek approximation under a perturbed model are considered, with an application for the Vega of an option. A brief review of the backward stochastic differential equation (BSDE) literature is provided and a proposed scheme for high-order approximations of the Gamma for non-linear pricing is suggested.

The third part of the thesis is on the discretisation schemes with strong rates of convergence of SDEs with non-Lipschitz continuous coefficients. Upon commencing the research, there were several discretisation schemes for such SDEs, including the implicit families and the various tamed schemes [DNS12, HJ12], which built on the earlier literature of approximations for SDEs admitting a solution in a domain [HMS02, BD04, BBD08]. This approach utilises a projection to ensure that the discretised process stays within a domain of interest. Chapter 6 provides strong convergence rates for a new modified Euler scheme applied to certain SDEs with non-globally Lipschitz continuous coefficients. The scheme introduced uses a projection in the state space, based on the locally Lipschitz continuous coefficients of the drift function of the process. This approach is naturally suited to SDEs with solutions within a domain, as it considers the behaviour at the boundary of the state space. The novelty is to consider the behaviour of the drift function both at zero and at infinity, in order to define the discretisation scheme. This approach relies on first studying the true process of the SDE and then selecting the scheme according to the problem. Examples of SDEs considered include the CIR model, the 3/2-model and the Ait-Sahalia model, all widely used in mathematical finance. A contribution is the extension of the parameter range for which strong rate of convergence holds, compared to the implicit schemes in the literature. Furthermore, for many choices of parameters in the Ait-Sahalia model, an implicit scheme poses significant computational difficulty compared to an explicit scheme. Numerical results supporting the theoretical results are provided. The modified Euler scheme is motivated by an application of multilevel Monte Carlo and an acceleration technique.

Part I

SMC Greeks

1. Sequential Monte Carlo Greeks

Sequential Monte Carlo (SMC) methods are model estimation techniques based on simulation, for approximating expectations with respect to a sequence of densities of increasing dimension. In this chapter, the use of SMC methods in the field of mathematical finance is explored, and the main contribution is the approximation of Greeks for a general stochastic volatility model. SMC methods have gained popularity in the last decade, with applications in engineering and applications to State Space Models. The algorithm is adapted to compute Greeks in a general stochastic volatility setting, with convergence results provided. Tracking option prices is discussed, which has implications in validating stochastic volatility models and their calibrated parameters. A specific application is the tracking of an S&P 500 call option price over a period of a month in the Black-Scholes model and of the Greeks in a stochastic volatility model, where the volatility is a hidden process. An aim is to discuss the practical applications in frameworks with uncertainty, where filtering can be used to compute the Greeks akin to the likelihood ratio method.

1.1 Introduction

Real-world phenomena can produce large time series data, evolving either continuously or discretely in time. Observations are typically discrete in time, with attempts made to describe the processes using models. An increase in computational power has enabled statistical inference for models which aim to describe the dynamics accurately. A common objective is estimating posterior distributions as observations arrive sequentially in time; however, these posterior distributions rarely admit closed-form solutions. The Kalman filter computes a Bayesian estimate for the state of a hidden variable in a linear dynamical system, providing an explicit solution for a linear Markov model perturbed by some Gaussian noise [Kal60]. There have been numerous extensions to this family of methods, such as the extended Kalman Filter for non-linear systems and the Unscented Kalman filter [JU97]. Developments in the 1990s of simulation-based techniques led to approaches consisting of the evolution and the updating of discrete sets of sampled values, with an associated weight [GSS93, Wes93]. In the literature, it has become common to refer to the sampled values as “particles”.

Particle filters are reminiscent of Genetic Algorithm (GA) techniques [SD08]. General steps involve:

- **Initialisation** - drawing from an initial prior distribution;
- **Sampling** (exploration of the space) - "selection" in the GA literature, according to some fitness/likelihood function;
- **Weights** - update weights for the "reproduction step";
- **Resampling** - "mutation-selection" in the GA literature.

Over the past 20 years there has been an explosion in the number of particle filtering techniques [DJ08, CGM07]. Those methods include Sequential Importance Resampling (SIR) and smoothing [GSS93]. There have been various modifications to the original particle filters, such as an auxiliary family of filters which increase the dimension of sampling [PS99]. Other modifications include the Probability Hypothesis Density filter and Approximate Bayesian Computation techniques for particle filtering [WSG10, JMMS12]. Additionally, particle filters can be modified to maintain multi-modality, which is especially desirable for tracking multiple objects [VDP03]; an extension of this is tracking using the Boosted Particle filter [OTdF$^+$04].

A summary of applications specifically in finance is provided by [Cre12]. Other applications include using Kalman filtering to track the state of the "true" order book, assuming the existence of noisy orders [JN11]. A bootstrap particle filter has been applied to estimate spot prices from future tenors in commodity markets [ABT08]. There have been attempts to use particle filtering techniques for inferring the US interest rate, using a monetary model for the economy [LS07]. Other applications include optimal portfolio allocation under stochastic volatility models and estimating default probabilities for collaterised debt obligations [BMV06, Koe11]. In option pricing, SMC methods are used for pricing contingency claims [JDM10]. Inference for stochastic volatility models is considered in [JSDT11]. This work provides inspiration for the smoothing algorithm shall used to approximate option Greeks.

In recent years, there have been huge developments in proving convergence results for SMC algorithms [BC09], including bounds and central limit theorems [CD00, Cho05, HSL08]. The difficulty in analysing convergence comes from the interaction between particles, making them statistically dependent. As a result, classical results from the Monte Carlo literature on convergence cannot be directly applied as the independence condition is not satisfied. Additionally, there is often accumulation of error with time, unless strict mixing conditions are imposed. In real applications, the constants bounding the convergence rate can be very difficult to compute, and can grow exponentially fast with time.

Notations: Capital letters denote random variables and lower case letters denote particular values, particles or realisations. For generic realisations, $(z_k)_{k\in\mathcal{I}}$, use $z_{i:j}$ to denote the vector $\big(z_i, z_{i+1}, \ldots, z_j\big)$. This definition extends naturally for sequences of random variables $(Z_k)_{k\in\mathcal{I}}$ as $Z_{i:j}$. For integration, $\mathrm{d}z_{i:j} \equiv \mathrm{d}z_i\ \mathrm{d}z_{i+1} \ldots\ \mathrm{d}z_j$ is used. Throughout, the convention of n time steps and N as the number of particles is used, denoted by $(X_k^{(i)})_{i=1,\ldots,N}$, at time steps $k = 0, \ldots, n$. Let X, Y be two random variables. Let $\mathcal{X}-$valued variable X with $f(x)$ being its probability density function for all $x \in \mathcal{X}$. For random variable X, $X \ \sim f$ reads as X is distributed according to density f. Write $X \propto Y$ if there exists a finite constant $Z > 0$, such that $X \sim Y/Z$.

Summary: The rest of this chapter is organised as follows. In Section 1.2, existing methodologies within the State Space Model literature and Monte Carlo methods tracing the origins of the particle filtering literature are reviewed [CMR05]. Section 1.3 describes smoothing and motivates the approximation of score vectors. Section 1.4 introduces a general framework for computing Greeks. In Section 1.5, convergence in the SMC literature is reviewed, and convergence results are proved for the proposed Greek approximations. In Section 1.6, numerical results for the tracking of an S&P 500 call option using a Taylor expansion consisting of the Greeks under a stochastic volatility model are presented. The chapter concludes with discussion and possible extensions.

1.2 Inference for State Space Models

State Space Models are a broad family of models describing processes including the Hidden Markov Models. HMMs are a class of models that can be non-linear and non-Gaussian, making them suitable for applications in engineering and finance. Suppose that X and Y are random variables with supports $\mathcal{X}$ and $\mathcal{Y}$. The following model covers a wide range of scenarios and applications: consider an index set $\mathcal{I}$ (typically $\mathbb{N}$) and let $X = (X_k)_{k\in\mathcal{I}}$ be a Markovian unobserved process and $Y = (Y_k)_{k\in\mathcal{I}}$ be an observed process, conditional on X [WH97]. Denote by $\theta \in \Theta \subseteq \mathbb{R}^d$ a set of fixed model parameters:

Definition 1.2.1 (Set of initial parameters, Θ)**.** *Define $\Theta \subseteq \mathbb{R}^d$ to be a family of parameters for a HMM assumed. Let $\theta \in \Theta$ be $\theta = (\theta_1, \ldots, \theta_d)$.*

The following densities are for the evolution of the Markovian process X and the conditional process Y:

1. The Hidden Markov process X is defined by its initial density $X_0 \sim \mu_\theta(.)$—as convention, $f_\theta(\cdot|x_{-1}) := \mu_\theta(\cdot)$—and the transition density $X_k|(X_{k-1} = x_{k-1}) \sim f_\theta(\cdot|x_{k-1})$;

2. The process X is not observed directly, but via the observations of the process Y. For $0 \leq k \leq n$:
$$Y_k|(X_0, \ldots, X_k = x_k, \ldots, X_n) \sim g_\theta(\cdot|x_k) \; .$$

The goal of HMM in a setting where θ is fixed is to filter the density of the unobservable Markovian random variables $X_{0:k}$ given the discrete observations $y_{0:k}$; i.e. to infer the sequences of filtering densities $\pi_\theta(x_k|y_{0:k})$ for $k \geq 0$.

The broad idea of particle filtering is to gradually build up the target distribution using a large set of random particles. The particles' location and likelihood are used to construct an empirical distribution and to perform inference of the hidden state, given the observations $y_{0:n}$. Suppose that θ is a known parameter; then the posterior density is

$$\pi_\theta(x_{0:n}|y_{0:n}) = \frac{\pi_\theta(x_{0:n}, y_{0:n})}{\pi_\theta(y_{0:n})} \; , \qquad \text{where} \qquad \pi_\theta(y_{0:n}) = \int_{\mathcal{X}^{n+1}} \pi_\theta(x_{0:n}, y_{0:n}) \mathrm{d}x_{0:n} \; ,$$

and the joint density is

$$\pi_\theta(x_{0:n}, y_{0:n}) = \pi_\theta(x_{0:n})\pi_\theta(y_{0:n}|x_{0:n}) = \prod_{k=0}^{n} f_\theta(x_k|x_{k-1}) g_\theta(y_k|x_k) \; .$$

SMC methods aim to approximate the posterior distribution, $\pi_\theta(x_{0:n}|y_{0:n})$. An important feature of HMMs is the ability to apply SMC for filtering, smoothing and prediction. These densities are categorised depending on how much information is available:

- **Filtering** - state density given past and present observations, $\pi_\theta(x_n|y_{0:n})$;

- **Smoothing** - state density given past and future observations, $p_\theta(x_k|y_{0:n})$ for $k < n$;

- **Predicting** - state density given past observations, $p_\theta(x_k|y_{0:n})$ for $k > n$.

The following two steps update the filtering densities upon the arrival of new observations, and (1.2.1) and (1.2.2) construct the forward filtering density. After initialisation, Bayes' theorem and marginalisation are used to update this recursion:

$$\pi_\theta(x_k|y_{0:k}) = \frac{g_\theta(y_k|x_k)}{p_\theta(y_k|y_{0:k-1})} p_\theta(x_k|y_{0:k-1}) \ , \tag{1.2.1}$$

and the following predictive density is used to forecast:

$$p_\theta(x_{k+1}|y_{0:k}) = \int_{\mathcal{X}} f_\theta(x_{k+1}|X_k)\pi_\theta(X_k|y_{0:k})\mathrm{d}X_k \ . \tag{1.2.2}$$

Applications of SMC include computing expectations using the approximated densities. Let $\varphi : \mathcal{X}^{n+1} \times \mathcal{Y}^{n+1} \to \mathbb{R}$ be a function of the hidden state and the observations, and suppose that one wishes to compute its expectation recursively in time. $\mathbb{E}_\pi[X]$ and $\mathbb{V}_\pi[X]$ denote the expected value and variance of a random variable X, with respect to the probability measure π. Suppose that φ is integrable with respect to $\pi_\theta(x_{0:n}|y_{0:n})$. By approximating the posterior, the particle approximation can be used to approximate integrals of the form:

$$I(\varphi) = \mathbb{E}_{\pi_\theta(X_{0:n}|y_{0:n})}\left[\varphi(X_{0:n}, y_{0:n})\right] := \int_{\mathcal{X}^{n+1}} \varphi(X_{0:n}, y_{0:n})\pi_\theta(X_{0:n}|y_{0:n})\mathrm{d}X_{0:n} \ . \tag{1.2.3}$$

A possible choice for the function φ is $\varphi(x_{0:n}, y_{0:n}) \equiv x_n$, for the terminal value of the hidden state, so that $I(\varphi)$ approximates the average, final, latent state.

1.2.1 SMC Algorithms

SMC algorithms provide posterior estimation using a series of predicting and updating recursions. The Sequential Importance Sampling technique can be seen as a general framework for particle filtering. Importance sampling is well studied in classical Monte Carlo literature, and can be used as a variance reduction technique. Large variance reduction can be achieved for instance when calculating the Value-at-Risk of large portfolio losses [GHS00]. Importance sampling can also be applied to the efficient calculation of deep out of the money options. In SMC methods, importance sampling is used as a way to associate importance weights to individual particles, to overcome sampling from the "wrong" distribution too often.

Sequential Importance Sampling (SIS) [GSS93]: Suppose that paths $x_{0:n}^{(i)} \sim \pi_\theta(X_{0:n}|y_{0:n})$ can be generated given a set of observations $y_{0:n}$, for $i = 1, \ldots, N$. Thus, the marginal density of the hidden model given some observations can be approximated. Suppose that particle paths $(x_{0:k-1}^{(i)})_{i=1,\ldots,N}$ are available at time $k-1$, weighted equally. An N-particle approximation of the posterior density is

$$\pi_\theta^N(x_{0:k-1}|y_{0:k-1}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_{0:k-1}^{(i)})} ,$$

where δ is the Dirac measure. By sampling $\bar{x}_k^{(i)} \sim f_\theta(.|x_{k-1}^{(i)})$ for $i = 1, \ldots, N$, a prediction for the density at time step k is

$$p_\theta^N(x_{0:k}|y_{0:k-1}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_{0:k-1}^{(i)}, \bar{x}_k^{(i)})} . \tag{1.2.4}$$

The target distribution at time step k is

$$\pi_\theta(x_{0:k}|y_{0:k}) = \frac{g_\theta(y_k|x_k)p_\theta(x_{0:k}|y_{0:k-1})}{\int_{\mathcal{X}} g_\theta(y_k|x_k)p_\theta(x_{0:k}|y_{0:k-1})\mathrm{d}x_k} . \tag{1.2.5}$$

The notation used throughout is $\{(x_k^{(i)}, w_k^{(i)})\}_{i=1}^N$, denoting the set of particle positions and corresponding weights at time step k. A set of particles, weighted according to their likelihood give the following approximations of $\pi_\theta(x_{0:k}|y_{0:k})$, for time steps $k \geq 0$. Substituting the predicted density in (1.2.5) by the approximation (1.2.4) yields

$$\bar{\pi}_\theta^N(x_{0:k}|y_{0:k}) = \sum_{i=1}^{N} w_k^{(i)} \delta_{(x_{0:k}^{(i)})} ,$$

where the weights $(w_k^{(i)})_{i=1,\ldots,N}$ satisfy

$$w_k^{(i)} \propto g_\theta(y_k|\bar{x}_k^{(i)}) \qquad \text{and} \qquad \sum_{i=1}^{N} w_k^{(i)} = 1 .$$

The weighted approximations, $\pi_\theta^N(x_{0:k}|y_{0:k})$, are then propagated through time, up to the terminal time step n. A feature of the SIS algorithm is that the path trajectories $(x_{0:n}^{(i)})_{i=1,\ldots,N}$ are independent and identically distributed. Define

$$\hat{I}^N(\varphi) := \sum_{i=1}^{N} \varphi(x_{0:n}^{(i)}, y_{0:n}) w_n^{(i)}$$

as the SMC estimate of $I(\varphi)$ in (1.2.3). SIS is usually successful for small n, however after several iterations most paths will have a negligible weight [DdFG01, Section 1.3.2]. Eventually one particle will dominate and be used to approximate the expectation, which illustrates the weight degeneracy problem.

Resampling: The variance of the weights increases with the number of time steps, and for a fixed accuracy, the computational cost grows exponentially [KLW94]. To stabilise the variance of weights, resampling methods have been proposed. Resampling consists of choosing a new set of particles based on the original set. The common idea is to increase the number of particles with higher weights, and reduce the number of particles that have low probability. At each time step, k, N particles from the current particle set could be sampled with replacement according to:

$$\mathbb{E}\left[N_k^{(i)}|x_{0:k}^{(i)}\right] = Nw_k^{(i)} .$$

The new particle set consists of $N_k^{(i)}$ realisations of particles $x_{0:k}^{(i)}$, with weights reset to $1/N$ for each resampled particle. Details for resampling schemes and examples of the empirical measures are presented in [Dou05, DMDJ12]. Multinomial resampling draws N new particles from a multinomial distribution according to the normalised weights $(w_k^{(i)})_{i=1,\dots,N}$. Systematic resampling uses a single random uniform draw to generate the new particle set. It is often preferred due to computational simplicity, however the method is sensitive to the ordering of particles [Dou05]. Other methods include residual resampling and stratified resampling [BC09, Dou05]. More complicated schemes have been studied, where the number of particles follow some evolutionary process [CDML99]. Resampling at each discrete time step can be harmful, so metrics such as the effective sample size (ESS) can be used as a trigger for performing a resampling step [LC98].

Definition 1.2.2. *Define the ESS approximation for a set of particles with weights $(w_k^{(i)})_{i=1,\dots,N}$ as:*

$$N_{eff} := \frac{1}{\sum_{i=1}^{N}\left(w_k^{(i)}\right)^2} \in [1, N], k \in \mathcal{I}.$$

N_{eff} approximates the equivalent number of i.i.d. random samples needed for an estimate, such that its Monte Carlo variance is that of the N-particle weighted approximation. A threshold can be set such that when N_{eff} drops below it, a resampling step is performed. In the literature, this threshold is commonly chosen as $N/2$ or $N/3$.

Intuitively, particles with high weights are more likely to be resampled, and particles with low weights will eventually cease to exist upon successive resampling steps. The effect of many successive resampling steps at time n leads to a loss of path diversity at time $n-k$ for some lag $k > 0$, which is referred to as the path degeneracy problem. Attempts have been made to minimise this problem by careful resampling and monitoring of the ESS [LC98, Whi, CDML99]. Path degeneracy is induced from resampling, and eventually approximations of the distribution would be just using one path. The trade-off in resampling can be summarised as controlling the variance of the weights, whilst not dramatically reducing the diversity of particles. Many paths will have the same history when looking through the path of the particles and ultimately all paths will coalesce to a single path [DJ08].

In situations where the consecutive distributions are very different, interpolating distributions have been proposed to reduce the need to resample particles as often [GC00]. Such techniques are often computationally expensive as the number of intermediate distributions could be prohibitive [BLB08].

Particle Filter with Resampling: In Algorithm 1.2.1, the most general particle filter with a resampling step is described. The ESS metric is used as the trigger to resample, according to a user-set resampling scheme. This method is based on the SIS algorithm, with the inclusion of a resampling step.

Algorithm 1.2.1 Particle Filter with Resampling (SIS/R)

Step 0: Initialise

a) For $i = 1 \to N$, sample $x_0^{(i)} \sim \mu(\cdot)$.

b) For $i = 1 \to N$, calculate normalised weights $w_0^{(i)} \propto g_\theta\left(y_0|x_0^{(i)}\right)$.

Step 1: Main recursive step. For $k = 1 \to n$

a) Resample Step

if $N_{eff} < N/2$ **then**

  Resample set $(x_{k-1}^{(i)})_{i=1,\ldots,N}$ according to weights $(\{x_{k-1}^{(i)}, w_{k-1}^{(i)}\})_{i=1,\ldots,N}$.

  For $i = 1 \to N$, set $w_{k-1}^{(i)} := 1/N$.

end if

b) Propagate particles. For $i = 1 \to N$, sample $x_k^{(i)} \sim f_\theta\left(X_k|x_{k-1}^{(i)}\right)$.

c) For $i = 1 \to N$, compute normalised weights, $w_k^{(i)} \propto w_{k-1}^{(i)} g_\theta\left(y_k|x_k^{(i)}\right)$.

The families of SIS algorithms are "online"—the complexity of the algorithm does not increase as the number of time steps increases, and only a fixed memory is required for a fixed number of particles. This is due to the fact that only a forward pass is required. Smoothing algorithms requiring a forward and a backward pass are classified as being "offline".

1.2.2 Convergence Results and Bounds

SMC methods are highly versatile techniques for Bayesian inference. These methods are very useful for dynamic models and are used to estimate a sequence of distributions of growing dimension. A frequently quoted application is the sequential Bayesian inference, which aims to approximate the target distribution $\pi_\theta(x_{0:n}|y_{0:n})$. For convergence results, the number of particles required for a fixed level of precision increases rapidly with the time steps. For $p > 1$, $\mathbb{L}_p$-bounds of the type

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^{N} \varphi(x_{0:n}^{(i)})w_n^{(i)} - \int_{\mathcal{X}^{n+1}} \varphi(X_{0:n})\pi_\theta(X_{0:n}|y_{0:n})\mathrm{d}X_{0:n}\right|^p\right]\right)^{1/p} \leq \frac{C_{p,n}}{\sqrt{N}} ,$$

have been shown, where $C_{p,n}$ is a constant which grows exponentially fast with the number of time steps, n [DM04]. This makes the error increase for fixed number of particles, N. Provided that resampling is used, central limit theorems such as

$$\sqrt{N}\left(\sum_{i=1}^{N} \varphi(x_{0:n}^{(i)})w_n^{(i)} - \int_{\mathcal{X}^{n+1}} \varphi(X_{0:n})\pi_\theta(X_{0:n}|y_{0:n})\mathrm{d}X_{0:n}\right) \xrightarrow{\mathrm{D}} N(0,\sigma_n^2) ,$$

hold as N increases to infinity. The variance, σ_n^2, is a complicated expression, and varies for different SMC algorithms and resampling schemes [DM04, Cho05]. SMC filters and their convergence properties are generally very difficult to study.

1.3 Smoothing

Smoothing is a filtering technique used to approximate the density of a hidden state given past and future observations. The Forward Filtering Backward Smoothing (FFBS) and Forward Smoothing only (FS-SMC) implementations will be summarised. The latter is applied to approximating Greeks (for more details of both algorithms, see [DMDS09]).

Smoothing, in its simplest form, can theoretically be performed alongside a generic particle filter. The Filter-Smoother consists of a standard particle filter which approximates $\pi_\theta(x_{0:n}|y_{0:n})$ using the weighted paths $\{(x_{0:n}^{(i)}, w_n^{(i)})\}_{i=1,\dots,N}$ [Kit96] . The joint smoothing density,

$$\pi_\theta(x_{0:n}|y_{0:n}) \propto g_\theta(y_n|x_n) f_\theta(x_n|x_{n-1}) \pi_\theta(x_{0:n-1}|y_{0:n-1}) ,$$

is marginalised yielding the smoothing density

$$p_\theta(x_k|y_{0:n}) = \int_{\mathcal{X}^n} \pi_\theta(x_{0:n}|y_{0:n}) \mathrm{d}x_{0:k-1} \mathrm{d}x_{k+1:n} . \tag{1.3.1}$$

This method requires storage of the newly sampled particle, $x_k^{(i)}$, in order to construct paths $x_{0:k}^{(i)} := (x_{0:k-1}^{(j_i)}, x_k^{(i)})$, where j_i will be some resampling indices; note that $x_{0:k-1}^{(j_i)}$ are resampled from $(x_{0:k-1}^{(i)})_{i=1,\dots,N}$. The algorithm requires the same $\mathcal{O}(N)$ computational cost of the filter. The particle filter provides accurate approximations for $\pi_\theta(x_k|y_{0:k})$, however resampling reduces the number of distinct paths. This suggests that examining paths of the particles over time, many paths will have coalesced into a single path. Resampling less frequently can reduce this problem, however for increasing $n-k$, the approximation of (1.3.1) will deteriorate as eventually only one path will have any significant weight [FWT10, DJ08].
Let $s_k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a sequence of functions for $k \in \mathbb{N}$, and $S_n : \mathbb{R}^{n+1} \to \mathbb{R}$, for $n \in \mathbb{N}$, be the corresponding sequence of additive functionals, defined as $S_n(x_{0:n}) := \sum_{k=1}^n s_k(x_{k-1}, x_k)$. An objective is computing $\mathcal{S}_n^\theta$, known as the smoothed additive functional, which is the expectation of the functional given the observations $y_{0:n}$:

$$\mathcal{S}_n^\theta := \mathbb{E}[S_n(X_{0:n})|y_{0:n}] , \tag{1.3.2}$$

which is assumed to be finite. The dependency on θ is due to the fixed parameters in the HMM setup. The aim is to construct an SMC estimate of $\mathcal{S}_n^\theta$.

1.3.1 FFBS Recursion

To mitigate the path degeneracy that afflicts the previously mentioned smoothing techniques, SMC approximations of the FFBS algorithm have been developed [DGA00]. This procedure computes the forward filtering densities $(\pi_\theta(x_k|y_{0:k}))_{k=0,\dots,n}$ using Bayes' theorem, followed by a backward pass approximating the marginal smoothed density $(p_\theta(x_{k-1}, x_k|y_{0:n}))_{k=1,\dots,n}$.

The forward filtering step,

$$\pi_\theta(x_{k+1}|y_{0:k+1}) = \frac{g_\theta(y_{k+1}|x_{k+1}) \int_{\mathcal{X}} f_\theta(x_{k+1}|x_k)\pi_\theta(x_k|y_{0:k})\mathrm{d}x_k}{\int_{\mathcal{X}^2} g_\theta(y_{k+1}|x'_{k+1}) f_\theta(x'_{k+1}|x'_k)\pi_\theta(x'_k|y_{0:k})\mathrm{d}x'_{k:k+1}} ,$$

is followed by a backward pass

$$p_\theta(x_{k-1}, x_k|y_{0:n}) = p_\theta(x_k|y_{0:n})p_\theta(x_{k-1}|y_{0:k-1}, x_k) = p_\theta(x_k|y_{0:n})\frac{f_\theta(x_k|x_{k-1})\pi_\theta(x_{k-1}|y_{0:k-1})}{p_\theta(x_k|y_{0:k-1})}; \tag{1.3.3}$$

together these steps approximate the smoothing density. In order to obtain $p_\theta(x_{k-1}|y_{0:n})$, the backward pass is marginalised with respect to x_k. The algorithm is based on the following recursion formula

$$p_\theta(x_{k-1}|y_{0:n}) = \int_{\mathcal{X}} p_\theta(x_k|y_{0:n})\frac{f_\theta(x_k|x_{k-1})\pi_\theta(x_{k-1}|y_{0:k-1})}{p_\theta(x_k|y_{0:k-1})}\mathrm{d}x_k ,$$

which was initially introduced by [Kit96]. The forward pass requires computation and storage of $(\pi_\theta^N(x_k|y_{0:k}))_{k=0,\dots,n}$, which are the approximations of $(\pi_\theta(x_k|y_{0:k}))_{k=0,\dots,n}$. Let the SMC approximation of $p_\theta(x_k|y_{0:n})$ be $p_\theta^N(x_k|y_{0:n}) = \sum_{i=1}^N w_{k|n}^{(i)} \delta_{(x_k^{(i)})}$, for $k \leq n$, with initialisation at the terminal time step $k = n$, by defining $w_{n|n}^{(i)} := w_n^{(i)}$. The rest of the weights are defined recursively using the backward pass (1.3.3):

$$p_\theta^N(x_{k-1}, x_k|y_{0:n}) = \sum_{i=1}^N \sum_{j=1}^N w_{k|n}^{(j)} \frac{w_{k-1}^{(i)} f_\theta(x_k^{(j)}|x_{k-1}^{(i)})}{\sum_{l=1}^N w_{k-1}^{(l)} f_\theta(x_k^{(j)}|x_{k-1}^{(l)})} \delta_{(x_{k-1}^{(i)}, x_k^{(j)})} = \sum_{i=1}^N w_{k-1|n}^{(i)} \delta_{(x_{k-1}^{(i)}, x_k^{(j)})} ,$$

where the weights are defined as

$$w_{k-1|n}^{(i)} := \sum_{j=1}^N w_{k|n}^{(j)} \frac{w_{k-1}^{(i)} f_\theta(x_k^{(j)}|x_{k-1}^{(i)})}{\sum_{l=1}^N w_{k-1}^{(l)} f_\theta(x_k^{(j)}|x_{k-1}^{(l)})} . \tag{1.3.4}$$

Finally, the SMC approximation of $\mathcal{S}_n^\theta$ is defined via

$$\hat{\mathcal{S}}_n^\theta := \sum_{k=1}^n \int_{\mathcal{X}^2} s_k(x_{k-1}, x_k) p_\theta^N(x_{k-1}, x_k|y_{0:n})\mathrm{d}x_{k-1:k} . \tag{1.3.5}$$

Algorithm 1.3.1 is an implementation of the FFBS algorithm to compute the weights of the particles for approximating $\hat{\mathcal{S}}_n^\theta$. This method uses the entire history of each particle, which requires memory proportional to the number of time steps, which makes the algorithm "offline". As a result, the next section considers an online implementation of the FFBS algorithm to circumvent the growing memory requirement.

Algorithm 1.3.1 Forward Filtering Backward Smoothing Algorithm

Step 0: Initialise.

a) For $k = 0 \to n$, $(\{x_k^{(i)}, w_k^{(i)}\})_{i=1,\ldots,N}$ is the SMC approximation of $\pi_\theta(x_k|y_{0:k})$.

b) For $i = 1 \to N$, define $w_{n|n}^{(i)} := w_n^{(i)}$.

Step 1: Main recursive step. For $k = n \to 1$

For $i = 1 \to N$, compute SMC approximation, $(x_{k-1}^{(i)}, w_{k-1|n}^{(i)})_{i=1,\ldots,N}$, for $p_\theta(x_{k-1}|y_{0:n})$ using (1.3.4).

1.3.2 Forward-only version of the FFBS recursion

The forward-only implementation of the FFBS algorithm avoids the backward pass [DMDS09]. The auxiliary function reviewed allows an online implementation [CMR05]. Define the forward smoothing recursion as

$$T_k^\theta(X_k) := \int_{\mathcal{X}^k} S_k(X_{0:k}) p_\theta(X_{0:k-1}|y_{0:k-1}, X_k) \mathrm{d}X_{0:k-1} . \tag{1.3.6}$$

and using this recursion it can be shown that $\mathcal{S}_k^\theta = \int_{\mathcal{X}} T_k^\theta(X_k)\pi_\theta(X_k|y_{0:k})\mathrm{d}X_k$. The approximation $p_\theta^N(x_{k-1}|y_{0:k-1}, x_k)$ of $p_\theta(X_{k-1}|y_{0:k-1}, X_k)$ is substituted in (1.3.6) to compute $\hat{T}_k^\theta(x_k)$, which is the N-particle approximation of $T_k^\theta(X_k)$. For completion, the proposition justifying the updating of the forward smoothing recursions is presented:

Proposition 1.3.1 ([DMDS09, Proposition 2.1])**.** *Define $T_0^\theta(x_0) := 0$. For $k \geq 1$, the smoothing recursion for the auxiliary functions $T_k^\theta(x_k)$ is defined by*

$$T_k^\theta(x_k) := \int_{\mathcal{X}} \left[T_{k-1}^\theta(x_{k-1}) + s_k(x_{k-1}, x_k)\right] p_\theta(x_{k-1}|y_{0:k-1}, x_k)\mathrm{d}x_{k-1} .$$

Algorithm 1.3.2 will be used to approximate the smoothed additive functional, $\mathcal{S}_n^\theta$.

1.3.3 Discussion on Smoothing Methods

The computational cost of FFBS is $\mathcal{O}(N^2)$ for each time step, compared to $\mathcal{O}(N)$ for methods such as the path-space and fixed-lag approximations [OCDM08]. Fixed-lag smoothers require

Algorithm 1.3.2 Forward Smoothing SMC algorithm (FS-SMC) [DMDS09]

Step 0: Initialise

a) Create initial SMC approximation, $(\{x_0^{(i)}, w_0^{(i)}\})_{i=1,\ldots,N}$, for $\pi_\theta(x_0|y_0)$.

b) For $i = 1 \to N$, initialise the forward smoothing recursion, $\hat{T}_0^\theta(x_0^{(i)}) := 0$.

Step 1: Main recursive step. For $k = 1 \to n$

a) Compute new SMC approximation, $(\{x_k^{(i)}, w_k^{(i)}\})_{i=1,\ldots,N}$, for $\pi_\theta(x_k|y_{0:k})$.

b) For $i = 1 \to N$, compute new smoothing recursion:

$$\hat{T}_k^\theta(x_k^{(i)}) = \frac{\sum_{j=1}^N w_{k-1}^{(j)} f_\theta(x_k^{(i)}|x_{k-1}^{(j)}) \left[\hat{T}_{k-1}^\theta(x_{k-1}^{(j)}) + s_k(x_{k-1}^{(j)}, x_k^{(i)})\right]}{\sum_{j=1}^N w_{k-1}^{(j)} f_\theta(x_k^{(i)}|x_{k-1}^{(j)})} .$$

c) Approximate smoothed additive functional, $\mathcal{S}_k^\theta$, as $\hat{\mathcal{S}}_k^\theta = \sum_{i=1}^N \hat{T}_k^\theta(x_k^{(i)}) w_k^{(i)}$.

tuning, rendering them somewhat unattractive, as it is often difficult to gain apriori intuition about the tuning parameters. Other techniques rely on the forgetting properties of the model [KDSM09, (17)], but again custom tuning is required [DdFG01].

The drawback of the FFBS algorithm is that the backward pass gets longer with each consequent time step; the storage requirement of the algorithm is increasing as the number of time steps grow. The use of past data means that the FFBS is an offline algorithm [DMDS09]. The FS-SMC also has computational cost of $\mathcal{O}(N^2)$, however it is online, in the sense that the storage requirements does not increase with the number of time steps. Both the FFBS and FS-SMC can be implemented more efficiently to reduce the cost from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ [KdFD05]. These algorithms has been used to approximate the score vector required for parameter estimation in a sequential Monte Carlo framework [Poy06, Poy11].

The Filter-Smoother provides degenerate results when smoothing, however it is a computationally cheap technique [FWT10]. In terms of convergence, the asymptotic variance for FFBS estimates grows linearly in the time-steps, n [DGMO11]. This suggests better results when compared to the results for the path-space methods, where error increases at least quadratically in time, under favourable mixing conditions [DMD03].

1.4 Greeks and Stochastic Volatility

In this section, an implementation of a smoothing SMC algorithm is used to approximate option Greeks [DMDS09, JDM10]. The setting is that of a HMM where the volatility is a latent process. A cascade of additive functions are defined for Greeks and convergence results are proved for the FS-SMC algorithm. The aim of the section is to suggest a novel approach for approximating Greeks in a general stochastic volatility setting.

Throughout, $(x_k)_{k=0,\dots,n}$ shall denote the latent process path (the unobserved volatility) and $(y_k)_{k=0,\dots,n}$ the observed asset price realisation (recall that $(X_k, Y_k)_{k\geq 0}$ denote the random variables). The density of the asset price path can be written as

$$p_\theta(y_{0:n}) = \int_{\mathcal{X}^{n+1}} \prod_{k=0}^{n} f_\theta(X_k|X_{k-1}) g_\theta(y_k|X_k) \mathrm{d}X_{0:n} \ . \tag{1.4.1}$$

As before, consider n equidistant time steps for the discretisation of the time interval $[0, T]$. A general set-up to approximate Greeks using the likelihood ratio method with respect to parameters $\theta \in \Theta \subseteq \mathbb{R}^d$ is described. Consider options with terminal payoff $\varphi(y_{1:n})$ and value

$$V := \int_{\mathbb{R}^{n+1}} \varphi\left(y_{1:n}\right) p_\theta(y_{0:n}) \mathrm{d}y_{0:n} \ . \tag{1.4.2}$$

Suppose that one can differentiate V through the integral with respect to θ_i to obtain

$$\frac{\partial V}{\partial \theta_i} = \int_{\mathbb{R}^{n+1}} \frac{\partial}{\partial \theta_i} \varphi(y_{1:n}) p_\theta(y_{0:n}) \mathrm{d}y_{0:n} + \int_{\mathbb{R}^{n+1}} \varphi(y_{1:n}) \frac{\partial \log p_\theta(y_{0:n})}{\partial \theta_i} p_\theta(y_{0:n}) \mathrm{d}y_{0:n} \ . \tag{1.4.3}$$

For simplicity, assume that $\frac{\partial}{\partial \theta_i}\varphi(y_{1:n}) = 0$. A Monte Carlo approach for approximating the Greeks is to sample M paths of the underlying, according to $p_\theta(y_{0:n})$ and approximate (1.4.3) by

$$\frac{1}{M}\sum_{j=1}^{M} \varphi(y_{1:n}^{(j)}) \frac{\partial}{\partial \theta_i} \log(p_\theta(y_{0:n}^{(j)})) \ .$$

The difficulty in this strategy lies in the computation of $\frac{\partial}{\partial \theta_i}\log(p_\theta(y_{0:n}^{(j)}))$, which in general is not known. The marginal likelihood can be decomposed using Fisher's identity into an additive function [DMDS09]:

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \log p_\theta(y_{0:n}) = \quad & \mathbb{E}\left[\frac{\partial}{\partial \theta_i} \log \mu_\theta(X_0) | y_{0:n}\right] + \sum_{k=1}^{n} \mathbb{E}\left[\frac{\partial}{\partial \theta_i} \log f_\theta(X_k|X_{k-1}) | y_{0:n}\right] \\ & + \sum_{k=0}^{n} \mathbb{E}\left[\frac{\partial}{\partial \theta_i} \log g_\theta(y_k|X_k) | y_{0:n}\right] \ . \end{aligned} \tag{1.4.4}$$

In (1.4.4), $\frac{\partial}{\partial \theta} \log p_\theta(y_{0:n})$ is the score vector, whose i^{th} component is $\frac{\partial \log p_\theta(y_{0:n})}{\partial \theta_i}$. It is a vector of derivatives of the log-marginal likelihood of the path $y_{0:n}$, with respect to the model parameters, θ_i. The score vector also has direct applications to gradient descent algorithms [CDM09].

By considering the additive form discussed above, it is possible to numerically approximate the score vector for any fixed path $y_{0:n}$ [DMDS09]. This is achieved using Algorithm 1.3.2 for path $y_{0:n}$ with N particles, namely computing the additive expectations in (1.3.2), with the additive functions defined specifically by the structure of the HMM. In fact, this is performed for M Monte Carlo paths $(y_{0:n}^{(j)})_{j=1,\dots,M}$, and the total computational cost is $\mathcal{O}(MN^2)$ for each time step n [DMDS09]. It should also be noted that the techniques is suited to parallelisation across multiple payoffs and strikes, since the bulk of the computational effort is spent on creating and updating the particle approximations and score vector approximations.

1.4.1 Additive Functions for Greeks

From the decomposition in (1.4.4), it is apparent that additive functions appear when approximating Greeks using the SMC approach. In this section, a cascade of these additive functions are derived under a general framework. As before, suppose that $\frac{\partial}{\partial \theta_i}\varphi(y_{1:n}) = 0$, for $i = 1, \dots, d$. Let $p_\theta(y_{0:n})$ be the likelihood of a path of the underlying and for brevity define the following partial derivatives of the log-likelihoods for $i, j, k = 1, \dots, d$:

$$l_i := \frac{\partial \log p_\theta(y_{0:n})}{\partial \theta_i}\,, \qquad l_{i,j} := \frac{\partial^2 \log p_\theta(y_{0:n})}{\partial \theta_i \partial \theta_j}\,, \qquad l_{i,j,k} := \frac{\partial^3 \log p_\theta(y_{0:n})}{\partial \theta_i \partial \theta_j \partial \theta_k}\,.$$

The following proposition defines additive functions for approximating Greeks with respect to different parameters. Recall that φ is the option payoff, and $p_\theta(y_{0:n})$ is defined in (1.4.1). These functions allow Greeks to be expressed as

$$\int_{\mathcal{X}^{n+1}} \varphi(y_{1:n})\phi_\alpha(y_{0:n})p_\theta(y_{0:n})\mathrm{d}y_{0:n}\,,$$

for some additive functions

$$\phi_\alpha := \frac{\partial^\alpha V}{\partial \theta_\alpha} = \frac{\partial^n V}{\partial \theta_1 \dots \partial \theta_n},$$

and a multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \{1, \dots, d\}^n$. The next few results have been adapted from the early discrete-event literature [Rub89, (4) and (7)].

Proposition 1.4.1 (First-order additive functions)**.** *For any $i = 1, \dots, d$, then $\phi_i = l_i$.*

Proof. Observe that

$$p_\theta(y_{0:n})\frac{\partial}{\partial\theta}\log p_\theta(y_{0:n}) = \frac{\partial}{\partial\theta}p_\theta(y_{0:n});$$

for θ_i, differentiation of the initial option value V yields

$$\frac{\partial V}{\partial\theta_i} := \frac{\partial}{\partial\theta_i}\int_{\mathcal{X}^{n+1}}\varphi(y_{1:n})p_\theta(y_{0:n})\mathrm{d}y_{0:n} = \int_{\mathcal{X}^{n+1}}\varphi(y_{1:n})l_i p_\theta(y_{0:n})\mathrm{d}y_{0:n}\ .$$

This follows from (1.4.3), and the assumption that $\frac{\partial}{\partial\theta_i}\varphi(y_{1:n}) = 0$. □

Differentiating V for higher order Greeks, yields further additive functions:

Corollary 1.4.1 (Second-order additive functions)**.** *For any $i, j \in \{1, \ldots, d\}$, then*

$$\phi_{i,j} = \phi_i\phi_j + l_{i,j}\ .$$

Proof. From the proof of Proposition 1.4.1, [Gly89, Section 3] and the product rule

$$\begin{aligned}
\tfrac{\partial^2 V_0}{\partial\theta_i\partial\theta_j} &:= \tfrac{\partial}{\partial\theta_j}\tfrac{\partial}{\partial\theta_i}\int_{\mathbb{R}^{n+1}}\varphi(Y_{1:n})p_\theta(Y_{0:n})\ \mathrm{d}Y_{0:n} \\
&= \tfrac{\partial}{\partial\theta_j}\int_{\mathbb{R}^{n+1}}\varphi(Y_{1:n})l_i p_\theta(Y_{0:n})\mathrm{d}Y_{0:n} \\
&= \int_{\mathbb{R}^{n+1}}\varphi(Y_{1:n})l_i l_j p_\theta(Y_{0:n})\mathrm{d}Y_{0:n} + \int_{\mathbb{R}^{n+1}}\varphi(Y_{1:n})l_{i,j}p_\theta(Y_{0:n})\mathrm{d}Y_{0:n}\ ,
\end{aligned} \tag{1.4.5}$$

which concludes the proof. □

Note that l_i is re-used to calculate the second-order Greeks. Assuming that first-order Greeks were approximated, there is just one new term, $l_{i,j}$, to be calculated.

Corollary 1.4.2 (Third-order additive functions)**.** *For any $i, j, k \in \{1, \ldots, d\}$, then $\phi_{i,j,k} = l_i l_j l_k + l_{i,j}l_k + l_{i,k}l_j + l_{j,k}l_i + l_{i,j,k}$.*

Proof. Differentiating (1.4.5) yields

$$\begin{aligned}
\frac{\partial^3 V_0}{\partial\theta_i\partial\theta_j\partial\theta_k} &:= \frac{\partial}{\partial\theta_k}\int_{\mathbb{R}^{n+1}}\varphi(Y_{1:n})\left[l_i l_j + l_{i,j}\right]p_\theta(Y_{0:n})\mathrm{d}Y_{0:n} \\
&= \int_{\mathbb{R}^{n+1}}\varphi(Y_{1:n})\left(\left[l_i l_j + l_{i,j}\right]\frac{\partial p_\theta(Y_{0:n})}{\partial\theta_k} + \frac{\partial\left[l_i l_j + l_{i,j}\right]}{\partial\theta_k}p_\theta(Y_{0:n})\right)\mathrm{d}Y_{0:n} \\
&= \int_{\mathbb{R}^{n+1}}\varphi(Y_{1:n})\left(\left[l_i l_j + l_{i,j}\right]l_k + \frac{\partial\left[l_i l_j + l_{i,j}\right]}{\partial\theta_k}\right)p_\theta(Y_{0:n})\mathrm{d}Y_{0:n} \\
&= \int_{\mathbb{R}^{n+1}}\varphi(Y_{1:n})\left[l_i l_j l_k + l_{i,j}l_k + l_{i,k}l_j + l_{j,k}l_i + l_{i,j,k}\right]p_\theta(Y_{0:n})\mathrm{d}Y_{0:n}\ .
\end{aligned}$$

□

Again, note that higher-order additive functions contain previously evaluated terms.

Remark 1.4.1. *As a slight abuse of notation, denote l_e as the component of the additive function with respect to e, for all elements $e \in \theta$.*

Remark 1.4.2. *The likelihood ratio method for computing Greeks in the Black-Scholes framework is treated in [Gla03, Chapter 7.3]. The validity of this approach relies on the ability of changing the order of integration and differentiation, which can be justified for smooth probability densities—unlike the pathwise technique for Greeks computation, no smoothness conditions are imposed on the option payoff. To compute Greeks using this approach, suppose that*

$$\frac{\partial}{\partial\theta}\mathbb{E}_{p_\theta}\left[\varphi(X)\right] = \int_{\mathcal{X}} \varphi(X)\frac{\partial}{\partial\theta}p_\theta(X)\mathrm{d}X.$$

The likelihood ratio method usually produces estimates for the Greeks with increasing variance, as the number of time steps increases—this feature is particularly unattractive in a sequential Monte Carlo framework, since the MSE of the algorithm increases with n, as shall be seen in Section 1.5.

1.4.2 Greek Calculations for stochastic volatility model

In the general introduction, several models for stochastic volatility were mentioned. The Black-Scholes model can be thought of as a "HMM", with a constant volatility. This is to motivate the subject of HMMs for the price dynamics of the underlying, conditioned on the volatility which is unobserved. This Black-Scholes formulation is used to test the technique and compare the Greek approximations to the known closed-form values.

Example 1.4.1. *For the Black-Scholes model, sensitivities with respect to parameters $\theta := (x, \sigma, r, T)$ are computed, where x is the initial underlying price, σ is the initial volatility, r is the interest rate (or drift) and T is the expiry time of the option.*

Example 1.4.2 (Random walk for log-volatility)**.** *A Brownian walk is now considered as the dynamic for the* log-volatility *process driving the asset price that follows a geometric Brownian motion. $W^{(1)}$ and $W^{(2)}$ are independent Brownian motions. The model can be written as the solution to the stochastic differential equation*

$$\begin{aligned} \textbf{\textit{Hidden:}}\quad \mathrm{d}\sigma_t &= \tfrac{1}{2}\sigma_t\eta^2\mathrm{d}t + \sigma_t\eta\mathrm{d}W_t^{(1)}, & \sigma_0 &= \sigma, \\ \textbf{\textit{Observed:}}\quad \mathrm{d}s_t &= rs_t\mathrm{d}t + \sigma_t s_t\mathrm{d}W_t^{(2)}, & s_0 &= x. \end{aligned} \tag{1.4.6}$$

This model has parameters $\theta := (x, \sigma, r, T, \eta)$, where η is the "volatility" of the log-volatility *process.*

Assume the following Bayesian set up where $\sigma_{0:n}$ is the volatility path and $s_{0:n}$ is the underlying asset price path. Suppose an equidistant time discretisation with time steps of size $h := T/n$, and define $a_k := r - \sigma_k^2/2$.

Remark 1.4.3. *Consider $(W_k)_{k=1,\dots,n}$ and $(Z_k)_{k=1,\dots,n}$ being i.i.d. $N(0,1)$ distributed random variables, and let*

$$\begin{aligned} \hat{\sigma}_{k+1} &= \hat{\sigma}_k \exp\left(\eta\sqrt{h}W_k\right), & \hat{\sigma}_0 = \sigma, \\ \hat{s}_{k+1} &= \hat{s}_k \exp\left((r - \tfrac{1}{2}\hat{\sigma}_k)h + \hat{\sigma}_k\sqrt{h}Z_k\right), & \hat{s}_0 = x, \end{aligned}$$

where $(\hat{s}_k)_{k\geq 0}$ and $(\hat{\sigma})_{k\geq 0}$ are the discretised processes for the underlying and volatility. The following results on the additive functions is expressed in terms of the random variables $(W_k, Z_k)_{k=1,\dots,n}$.

Once a stochastic volatility model is chosen, the score vector for each sensitivity is approximated according to (1.4.4). This translates to calculating the additive function for the different Greeks. The additive functions are separated by the contribution from the transition densities f_θ and g_θ; denote by l_x^f, the contribution from the latent process transition density $f_\theta(\sigma_k|\sigma_{k-1})$, and l_x^g as the additive function contribution from $g_\theta(s_k|\sigma_k)$, for the sensitivity with respect to parameter x (initial underlying). Define

$$l_x^f := \sum_{k=1}^{n} \mathbb{E}\left[\frac{\partial}{\partial x}\log f_\theta(\sigma_k|\sigma_{k-1})|s_{0:n}\right], \qquad l_x^g := \sum_{k=0}^{n} \mathbb{E}\left[\frac{\partial}{\partial x}\log g_\theta(s_k|\sigma_k)|s_{0:n}\right];$$

the definition extends to l_σ^f, l_σ^g, and other parameters.

Proposition 1.4.2 (Delta). *The additive function for the Delta in Example 1.4.2 is*

$$\phi_x = l_x = l_x^f = \frac{Z_1}{x\sigma\sqrt{h}}.$$

Proof. From Proposition 1.4.1 and one time step,

$$p_\theta(s_{0:1}) = \int_{\mathcal{X}} p_\theta(s_1|\sigma_1, s_0) f_\theta(\sigma_1|\sigma_0) \mathrm{d}\sigma_1.$$

Generalising this for the n steps, and following on from (1.4.4), observe that all terms apart from $\frac{\partial}{\partial x}\log p_\theta(s_1|\sigma_0, s_0)$ are null. Using Fisher's identity (see [DMS14, Appendix D.3, p.495]) yields

$$\begin{aligned}\frac{\partial}{\partial x}\log p_\theta(s_{0:n}) &= \mathbb{E}\left[\frac{\partial}{\partial x}\log f_\theta(\sigma_1|\sigma)|s_{0:n}\right] + \mathbb{E}\left[\frac{\partial}{\partial x}\log p_\theta(s_1|\sigma, s_0)|s_{0:n}\right] \\ &= \mathbb{E}\left[\frac{\partial}{\partial x}\log p_\theta(s_1|\sigma, x)|s_{0:n}\right] ,\end{aligned}$$

which demonstrates that $l_x^f = 0$. It follows from [Gla03, (7.33),(7.34)] that

$$Z_1 = \frac{\log(s_1/x) - a_0 h}{\sigma\sqrt{h}} \sim N(0,1),$$

therefore

$$\frac{\partial}{\partial x}\log p_\theta(s_1|\sigma, x) = \frac{\partial}{\partial x}\left[-\frac{1}{2}\left(\frac{\log(s_1/x) - a_0 h}{\sigma\sqrt{h}}\right)^2\right] = \frac{\partial}{\partial x}\left[-\frac{Z_1^2}{2}\right] = \frac{Z_1}{x\sigma\sqrt{h}}.$$

□

Remark 1.4.4. *It is rather intuitive that l_x^f provides no contribution to the additive function l_x; the density of the latent state differentiated with respect to the initial asset price is zero, since the volatility drives the asset, and not the other way around.*

The additive functions for other sensitivities can be similarly computed:

Proposition 1.4.3 (Vega). *The additive function for the Vega in Example 1.4.2 is $\phi_\sigma = l_\sigma = l_\sigma^f + l_\sigma^g$ where*

$$l_\sigma^f := \frac{W_1}{\sigma\eta\sqrt{h}}, \qquad l_\sigma^g := \sum_{k=1}^{n}\left(\frac{Z_k^2 - 1}{\sigma} - Z_k\sqrt{h}\right).$$

Proof. Recall that for $k = 1, \dots, n$, $W_k := \frac{\log(\sigma_k) - \log(\sigma_{k-1})}{\eta\sqrt{h}} \sim N(0,1)$; upon differentiation of $\log f_\theta(\sigma_k|\sigma_{k-1})$ with respect to σ, observe that analogously to the Delta computation, just one term is left, namely $W_1/(\sigma\eta\sqrt{h})$. For the proof of l_σ^g, see [Gla03, p.405 (7.37)]. □

The sensitivity with respect to the drift parameter, r, and the expiry time of the option, T, are now computed:

Proposition 1.4.4 (Rho). *The additive function for the sensitivity Rho of Example 1.4.2 is*

$$\phi_r = l_r = l_r^g = \sum_{k=1}^{n}\frac{Z_k\sqrt{h}}{\sigma_{k-1}}.$$

Proof. Differentiation for l_r^f clearly yields zero, since the interest rate does not play a role in the transition density for the latent state in the model. To obtain l_r^g, observe that $\partial Z_k/\partial r = -\sqrt{h}/\sigma_{k-1}$, which concludes the proof. □

Proposition 1.4.5 (Theta)**.** *The additive function for the Theta of Example 1.4.2 is $\phi_T = l_T = l_T^f + l_T^g$, where*

$$l_T^f := \sum_{k=1}^{n} \frac{W_k^2}{2T}, \qquad l_T^g := \sum_{k=1}^{n} \left[\frac{Z_k a_{k-1}\sqrt{h}}{\sigma_{k-1}T} + \frac{Z_k^2}{2T} \right].$$

Remark 1.4.5. *There is a particular difference in the forms of the additive functions. For the case of the Delta and Gamma, the additive functions consist only of the first time step contribution, whereas additive functions for σ, r and T have a summation across all time steps. The Gamma additive function, $l_{x,x}^g$ is derived in [Gla03, p.411 (7.45)], and $l_{x,x}^f := 0$ for the stochastic volatility model considered.*

Proposition 1.4.6 (Vanna)**.** *The additive function for the Vanna of Example 1.4.2 is $\phi_{x,\sigma} = l_{x,\sigma} + l_x l_\sigma$, where $l_{x,\sigma} = l_{x,\sigma}^f + l_{x,\sigma}^g$ and*

$$l_{x,\sigma}^f := 0, \qquad l_{x,\sigma}^g := \frac{1}{x\sigma} - \frac{2Z_1}{x\sigma^2\sqrt{h}}.$$

Proof. Combining the results for l_x^f and l_σ^f, the additive function $l_{x,\sigma}^f$ is zero. For $l_{x,\sigma}^f$ observe that

$$l_{x,\sigma}^g = \frac{\partial}{\partial\sigma} l_x^g = \frac{\partial}{\partial\sigma} \frac{Z_1}{x\sigma\sqrt{h}} = \frac{x\sigma\sqrt{h}\frac{\partial}{\partial\sigma}(Z_1) - Z_1 x\sqrt{h}}{x^2\sigma^2 h} = \frac{1}{x\sigma} - \frac{2Z_1}{x\sigma^2\sqrt{h}},$$

and the proof is concluded using Corollary 1.4.1. □

Corollary 1.4.3 (Random walk for the log-volatility)**.** *Suppose the model in* (1.4.6)*, for some fixed θ. Then, the additive function components for the first and second-order Greeks are in Table 1.1.*

The results are derived continuing from the previous five propositions, hence omitted. From the previous conventions it follows that $l_{\sigma,\sigma} := l_{\sigma,\sigma}^f + l_{\sigma,\sigma}^g$, and similarly for $l_{x,T}$ and $l_{\sigma,T}$. Consequently using Proposition 1.4.1 and Corollary 1.4.1 the Greek additive functions can be explicited:

Order	Name	Value
1	l_x^f	0
1	l_x^g	$\frac{Z_1}{x\sigma\sqrt{h}}$
1	l_σ^f	$\frac{W_1}{\sigma\eta\sqrt{h}}$
1	l_σ^g	$\sum_{k=1}^{n}\left(\frac{Z_k^2-1}{\sigma}-Z_k\sqrt{h}\right)$
1	l_r^f	0
1	l_r^g	$\sum_{k=1}^{n}\frac{Z_k\sqrt{h}}{\sigma_{k-1}}$
1	l_T^f	$\sum_{k=1}^{n}\frac{W_k^2}{2T}$
1	l_T^g	$\sum_{k=1}^{n}\left[\frac{Z_k a_{k-1}\sqrt{h}}{\sigma_{k-1}T}+\frac{Z_k^2}{2T}\right]$
2	$l_{x,x}^f$	0
2	$l_{x,x}^g$	$-\frac{1}{x^2\sigma^2 h}-\frac{\log(s_1/x)-a_0h}{x^2\sigma^2h}$
2	$l_{x,\sigma}^f$	0
2	$l_{x,\sigma}^g$	$\frac{1}{x\sigma}-\frac{2Z_1}{x\sigma^2\sqrt{h}}$
2	$l_{x,T}^f$	0
2	$l_{x,T}^g$	$-\frac{a_0}{x\sigma^2T}-\frac{Z_1}{\sigma\sqrt{h}xT}$
2	$l_{\sigma,\sigma}^f$	$\frac{-1}{\sigma^2\eta^2h}-\frac{W_1}{\sigma^2\eta\sqrt{h}}$
2	$l_{\sigma,\sigma}^g$	$\sum_{k=1}^{n}\left(\frac{3\sqrt{h}Z_k\sigma_k-h\sigma_k^2-3Z_k^2+1}{\sigma_k^2}\right)$
2	$l_{\sigma,T}^f$	$\sum_{k=1}^{n}\frac{W_k(\sigma_{k-1}-\sigma_k)}{T\sigma_k\sigma_{k-1}\eta\sqrt{h}}$
2	$l_{\sigma,T}^g$	$\sum_{k=1}^{n}\left[-\frac{Z_k^2}{T\sigma_{k-1}}+\left(\frac{1}{T\sqrt{h}}-\frac{T(a_{k-1}+\sigma_{k-1}^2+a_{k-1}\sqrt{h})}{\sigma_{k-1}^2T^2}\right)Z_k+\frac{a_{k-1}}{\sigma_{k-1}T\sqrt{h}}\right]$

Table 1.1: Components of additive functions for the stochastic volatility model in Example 1.4.2.

Corollary 1.4.4. *The additive functions for the second-order Greeks are:*

$$\begin{aligned}
\textit{Gamma:} \quad \phi_{x,x} &= l_{x,x} + l_x l_x, \\
\textit{Vanna:} \quad \phi_{x,\sigma} &= l_{x,\sigma} + l_x l_\sigma, \\
\textit{Charm:} \quad \phi_{x,T} &= l_{x,T} + l_x l_T, \\
\textit{Vomma:} \quad \phi_{\sigma,\sigma} &= l_{\sigma,\sigma} + l_\sigma l_\sigma, \\
\textit{Veta:} \quad \phi_{\sigma,T} &= l_{\sigma,T} + l_\sigma l_T.
\end{aligned}$$

Remark 1.4.6.

(i) Observe that the variance of the additive functions increases as the step size, h, decreases. This shall be observed in the numerical results section, where the error for Greeks with respect to the latent volatility is considerably higher than that for Greeks with respect to the underlying.

(ii) The additive functions in the likelihood ratio method are agnostic of the option payoff-.

(iii) The choice of the stochastic volatility model is somewhat arbitrary; although tables for other models are not included, it is straightforward to derive the additive functions following the steps prescribed from the previous claims.

1.5 Convergence

This section reviews particle filtering convergence results in discrete time, which are adapted for the proposed approach [BC09, CD00, CD02]. The main results for convergence of the FS-SMC algorithm are shown for bounded payoff functions, however the approach can be extended for a general class of unbounded functions [HSL08].

Recall that the latent process, $X = (X_k)_{k \in \mathbb{N}}$, $X_k \in \mathbb{R}^d$, is a stochastic Markov process defined on the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. For shorthand, denote by $g_k^{y_k}$ the density $g_\theta(y_k|\cdot)$ and f_k for $f_\theta(x_k|\cdot)$. There exists a recurrence formula for the distribution of the random variable, X_k. For $A \subseteq \mathcal{X}$, define the random density $\pi_k^{Y_{0:k}}$ as

$$\pi_k^{Y_{0:k}}(A) := \mathbb{P}(X_k \in A | Y_{0:k})$$

and the expectation with respect to function φ as $\pi_k^{Y_{0:k}}\varphi := \mathbb{E}[\varphi(X_k)|Y_{0:k}]$. Particle filtering approximates the random measure $\pi_k^{Y_{0:k}}$, which can be used to calculate the expectation with respect to any bounded function φ. From now on, fixed paths $Y_{0:n} = y_{0:n}$ are taken. For each individual realisation, $y_{0:n}$, the conditional density and expectations are

$$\pi_k^{y_{0:k}}(A) := \mathbb{P}(X_k \in A | Y_{0:k} = y_{0:k}) , \qquad \pi_k^{y_{0:k}}\varphi := \mathbb{E}[\varphi(X_k)|Y_{0:k} = y_{0:k}] \ .$$

To ease notation, the explicit dependence on the path $y_{0:k}$ is omitted. Define the empirical measure created from the position of the N particles $\{x_k^{(i)}\}_{i=1,\ldots,N}$ as

$$\pi_k^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_k{}^{(i)})} \ ,$$

where the samples are generated from an SMC method (for example, using Algorithm 1.2.1). The weighted measure and the predictive empirical measures are

$$\bar{\pi}_k^N := \sum_{i=1}^{N} w_k^{(i)} \delta_{(x_k{}^{(i)})} \ , \qquad p_k^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{(\bar{x}_k^{(i)})} \ ,$$

where $\bar{x}_k^{(i)} \sim f_\theta\left(.|x_{k-1}^{(i)}\right)$ is the particle position after the predictive step.

Definition 1.5.1. *Let p be a measure (non-null everywhere) and let φ be a non-negative, bounded function. The projective product associated with a function, $\varphi : \mathbb{R} \to \mathbb{R}$, is defined as $p(\varphi) := \int_{\mathbb{R}^d} \varphi(x)p(x)\mathrm{d}x$. Furthermore, suppose that $p(\varphi) > 0$. The projective operator $\star$ is the set function defined by*

$$\varphi \star p(A) := \frac{\int_A \varphi(x)p(x)\mathrm{d}x}{p(\varphi)} \ , \qquad \textit{for all } A \in \mathcal{B}(\mathbb{R}^d) \ .$$

The next result establishes the recurrence formula, consisting of a predictive and updating step:

Proposition 1.5.1 ([BC09, Proposition 10.6])**.** *For a fixed path $y_{0:k}$ the probability measure $\pi_k^{y_{0:k}}$ satisfies the recurrence relation $\pi_k^{y_{0:k}} = g_k^{y_k} \star \left(f_k \pi_{k-1}^{y_{0:k-1}}\right)$, $\mathbb{P}_{Y_{0:k}}$ – almost surely.*

The term $f_k\pi_{k-1}^{y_{0:k-1}}$ is the prediction step, occurring before the new observation y_k becomes available. The second step updates the density taking into account the new information y_k. Intuitively, for the fixed observation case π_k^N converges to $\pi_k^{y_{0:k}}$ and p_k^N converges to $p_k^{y_{0:k-1}}$ almost surely if

- π_0^N tends to the correct initial distribution;
- the limit of the distance between the predictive sequence p_k^N and $f_k\pi_{k-1}^N$ is zero.

Throughout, assume that the following conditions are satisfied:

- p_k^N and π_k^N are random, non-null everywhere measures;
- $p_k^N g_k^{y_k} > 0$ for all $N > 0$ and time steps k.

Convergence for the SMC method presented in Algorithm 1.2.1 is proved inductively, with bounds in terms of the number of particles and a stochastically increasing constant with time:

Lemma 1.5.1 ([BC09, Corollary 10.28]). *Suppose that φ is bounded. Then, for all $k = 1, \ldots, n$, $\mathbb{E}\big[(\pi_k^N \varphi - \pi_k^{y_{0:k}} \varphi)^2\big] \leq C_k / N$, for some constant C_k depending on time.*

1.5.1 Convergence for Greeks

For the application proposed, consider the family of simplified additive functions $S_n(x_{0:n}) = \sum_{k=0}^n s_k(x_k)$, where $s_k : \mathbb{R} \to \mathbb{R}$. Define $\|s_k\| := \sup_{x \in \mathbb{R}} |s_k(x)|$, and denote the oscillation of s_k by $osc(s_k) := \sup_{x,y \in \mathbb{R}} |s_k(x) - s_k(y)|$. The following regularity assumptions are considered:
$(\mathbf{H}b)$: There exist $0 < \rho, \delta < \infty$ such that for all $x, x' \in \mathcal{X}$, $y \in \mathcal{Y}$ and $\theta \in \Theta$,

$$\rho^{-1} \leq f_\theta\left(x'|x\right) \leq \rho\ , \qquad \delta^{-1} \leq g_\theta\left(y|x\right) \leq \delta\ ;$$

furthermore, s_k are bounded and $osc(s_k) \leq 1$ for all $k = 1, \ldots, n$.

Recall (1.3.5). The next lemma provides a bound on the mean squared error of $\mathcal{S}_n^\theta$:

Lemma 1.5.2 ([DMDS09, Theorem 3.1]). *Assume that $(\mathbf{H}b)$ holds and $\theta \in \Theta$. Then,*

$$\mathbb{E}\left(|\hat{\mathcal{S}}_n^\theta - \mathcal{S}_n^\theta|^2\right) \leq \frac{C(n+1)}{N}\left(1 + \sqrt{\frac{n+1}{N}}\right)^2 ,$$

where C is a finite constant, independent of N, θ and the choice of additive functions.

Convergence of Algorithm 1.3.2 relies on $(\mathbf{H}b)$, however the authors suggest that numerical studies do not always require them in order for the algorithm to perform satisfactorily [DMDS09].

The error of the approximation is bounded using the number of particles and the number of simulated paths. Path dependence is introduced in the definition of the smoothed additive functionals, i.e. for path j, define the expectation of the additive functional as

$$\mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)}) := \mathbb{E}\Big[S_n(X_{0:n})|y_{0:n}^{(j)}\Big]\ .$$

Denote $\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)})$ as the SMC approximations of $\mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)})$, using N particles. Using Lemma 1.5.2, the main result of this section on the convergence for approximating Greeks follows:

Theorem 1.5.1. *Assume that* $(\mathbf{H}b)$ *holds and that* φ *is bounded. Consider Algorithm 1.3.2 with N particles and M simulated paths. Then,*

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(y_{0:n}^{(j)})\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)}) - \int_{\mathbb{R}^{n+1}}\varphi(Y_{0:n})\mathcal{S}_n^{\theta}(Y_{0:n})p_\theta(Y_{0:n})\mathrm{d}Y_{0:n}\right)^2\right]\\&\leq C\left[\frac{1}{NM}\left(1+\sqrt{\frac{n+1}{N}}\right)^2+\frac{1}{M}\right],\end{aligned} \tag{1.5.1}$$

where C is a constant independent of N, M, θ depending on the choice of additive functions.

Proof. Applying Minkowski's Lemma to the left-hand side of (1.5.1) yields

$$\mathbb{E}\left[\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(y_{0:n}^{(j)})\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)}) - \int_{\mathbb{R}^{n+1}}\varphi(Y_{0:n})\mathcal{S}_n^{\theta}(Y_{0:n})p_\theta(Y_{0:n})\mathrm{d}Y_{0:n}\right)^2\right] \leq 2\left(\mathbb{E}\left[\Lambda^2\right]+\mathbb{E}\left[\Upsilon^2\right]\right),$$

where

$$\Lambda := \frac{1}{M}\sum_{j=1}^{M}\varphi(y_{0:n}^{(j)})\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)}) - \frac{1}{M}\sum_{j=1}^{M}\varphi(y_{0:n}^{(j)})\mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)}),$$

and

$$\Upsilon := \frac{1}{M}\sum_{j=1}^{M}\varphi(y_{0:n}^{(j)})\mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)}) - \int_{\mathbb{R}^{n+1}}\varphi(Y_{0:n})\mathcal{S}_n^{\theta}(Y_{0:n})p_\theta(Y_{0:n})\mathrm{d}Y_{0:n}.$$

Thus, $\mathbb{E}[\Lambda^2]$ can be bounded using Lemma 1.5.2, the boundedness of φ, the independence of paths $(y_{0:n}^{(j)})_{j=1}^{M}$ and the fact that the approximation $\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)})$ is an unbiased estimator of

$\mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)})$:

$$\begin{aligned}
\mathbb{E}[\Lambda^2] &\leq C\mathbb{E}\left[\left(\frac{1}{M}\sum_{j=1}^{M}\left[\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)}) - \mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)})\right]\right)^2\right] \\
&= \frac{C}{M^2}\sum_{j=1}^{M}\sum_{l=1}^{M}\mathbb{E}\left[\left(\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)}) - \mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)})\right)\left(\hat{\mathcal{S}}_n^{\theta,l}(y_{0:n}^{(l)}) - \mathcal{S}_n^{\theta,l}(y_{0:n}^{(l)})\right)\right] \\
&= \frac{C}{M^2}\sum_{j=1}^{M}\mathbb{E}\left[\left(\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)}) - \mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)})\right)^2\right] \\
&= \frac{C}{M}\mathbb{E}\left[\left(\hat{\mathcal{S}}_n^{\theta,j}(y_{0:n}^{(j)}) - \mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)})\right)^2\right] \leq \frac{C}{NM}\left(1+\sqrt{\frac{n+1}{N}}\right)^2 .
\end{aligned}$$

The expectation of Υ^2 is bounded using the Monte Carlo error by observing that

$$\mathbb{E}\Big[\Upsilon^2\Big] = \mathbb{E}\left[\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(y_{0:n}^{(j)})\mathcal{S}_n^{\theta,j}(y_{0:n}^{(j)}) - \int_{\mathbb{R}^{n+1}}\varphi(Y_{0:n})\mathcal{S}_n^{\theta}(Y_{0:n})p_\theta(Y_{0:n})\mathrm{d}Y_{0:n}\right)^2\right] = \frac{C}{M},$$

where $C := \mathbb{V}[\varphi(Y_{0:n})\mathcal{S}_n^{\theta}(Y_{0:n})] < \infty$, since φ and $\mathcal{S}_n^{\theta}(Y_{0:n})$ are finite. Combining the two bounds proves the claim. □

For general discretisation schemes, assume a weak rate of convergence $q > 0$, i.e. for approximations $(\hat{x}_k, \hat{y}_k)_{k=1,\ldots,n}$ of $(x_k, y_k)_{k=1,\ldots,n}$, assume that

$$\left|\mathbb{E}[\Phi(x_{0:n}, y_{0:n})] - \mathbb{E}[\Phi(\hat{x}_{0:n}, \hat{y}_{0:n})]\right| \leq C/n^q, \tag{1.5.2}$$

for all sufficiently smooth functional $\Phi : \mathcal{X}^{n+1} \times \mathcal{Y}^{n+1} \to \mathbb{R}$. For the next corollary $\Phi(x_{0:n}, y_{0:n}) := \varphi(y_{0:n})\mathcal{S}_n^{\theta}(y_{0:n})$.

Corollary 1.5.1. *Assume that* $(\mathbf{H}b)$ *holds,* φ *is bounded, and that* (1.5.2) *holds for some* $q > 0$. *Consider Algorithm 1.3.2 with N particles and M simulated paths for n time steps. Then,*

$$\begin{aligned}
&\left|\mathbb{E}\left[\frac{1}{M}\sum_{j=1}^{M}\varphi(\hat{y}_{0:n}^{(j)})\hat{\mathcal{S}}_n^{\theta,j}(\hat{y}_{0:n}^{(j)})\right] - \mathbb{E}\left[\int_{\mathbb{R}^{n+1}}\varphi(Y_{0:n})\mathcal{S}_n^{\theta}(Y_{0:n})p_\theta(Y_{0:n})\mathrm{d}Y_{0:n}\right]\right| \\
&\leq C\left(\frac{1}{n^q} + \sqrt{\frac{1}{NM}\left(1+\sqrt{\frac{n+1}{N}}\right)^2 + \frac{1}{M}}\right),
\end{aligned}$$

where C is a constant independent of N, M, θ and depending on the choice of additive functions.

The corollary follows from the previous theorem. Corollary 1.5.1 suggests how parameters N, M should be chosen for a fixed computational effort; the cost of the FS-SMC algorithm is $\mathcal{O}(nN^2M)$. Therefore, choosing N proportional to n^α and M proportional to n^β for some α, β, convergence for discretisation schemes with a weak convergence of rate q are obtained. From the above MSE, $2q \leq \beta$ and $2q \leq 2\alpha + \beta - 1$ imply that it is sufficient to choose $\alpha = 1/2$ and $\beta = 2q$.

Example 1.5.1.

- *Euler scheme with* $q = 1$*: for* $n = 100$ *time steps, selecting* $N = 10$ *and* $M = 10000$ *is a sensible choice of parameters according to the corollary. Observe that the number of particles increases for a large number of paths, M.*

- *For a higher-order scheme with a rate of weak convergence* $q = 2$, $(M, N, n) = (10^8, 10, 100)$ *is an appropriate choice. It seems apparent that considering weak Taylor approximations of higher order for general stochastic volatility models would be beneficial for controlling the cost of the algorithm, given that increasing the number of particles is particularly computationally expensive in this framework.*

1.6 Numerical Results

In this section, numerical results are presented for approximating Greeks. First, the FS-SMC algorithm is applied to the Black-Scholes model in order to validate the technique — the setting is taken to be a "HMM" with fixed volatility. Later, the stochastic volatility model in (1.4.6) is considered. Finally, an indirect method to validate the Greeks for option price replication is implemented. The aim is to track the option price through a Taylor expansion of the option price with respect to first-order and second-order Greeks. It is shown that tracking using a stochastic volatility model greatly outperforms tracking using the Black-Scholes Greeks. An example of tracking an S&P 500 call option over one month is considered using the Delta, Gamma, Vega and Theta.

Denote by (M, N, n, R) the parameters of a particular experiment, where M is the number of simulated underlying paths, N is the number of particles used to approximate the score vector for each path, n is the number of time steps, and R is the number of repeats for the experiment. The number of repeats demonstrates the variability in the Greek approximations across runs.

Remark 1.6.1. *Upon using Algorithm 1.3.2, the bulk of the computational effort is to compute the score vector. The efficiency is greatly improved, by only approximating the score vector for those paths that expire in the money; otherwise the score vector is set to zero. This is particularly important for options with a low probability of expiring in-the-money.*

1.6.1 Black-Scholes

In this section the methodology is applied to the Black-Scholes model. Consider the Delta of a European call option with parameters $(x, \sigma, r, T) = (100, 0.249, 0.03, 30/365)$, and strikes $K = 80, \ldots, 120$, in steps of 1. First, run the SMC algorithm using the parameters $(M, N, n, R) = (10000, 1000, 1, 1)$, which has a runtime of 10345 seconds (2:53 hours), displayed in Figure 1.1 (Left). The step size for the time discretisation is $h = 30/365$. Observe that for a wide range of strikes, the Delta approximation using the SMC approach is of reasonable accuracy by comparison to the closed-form solution.

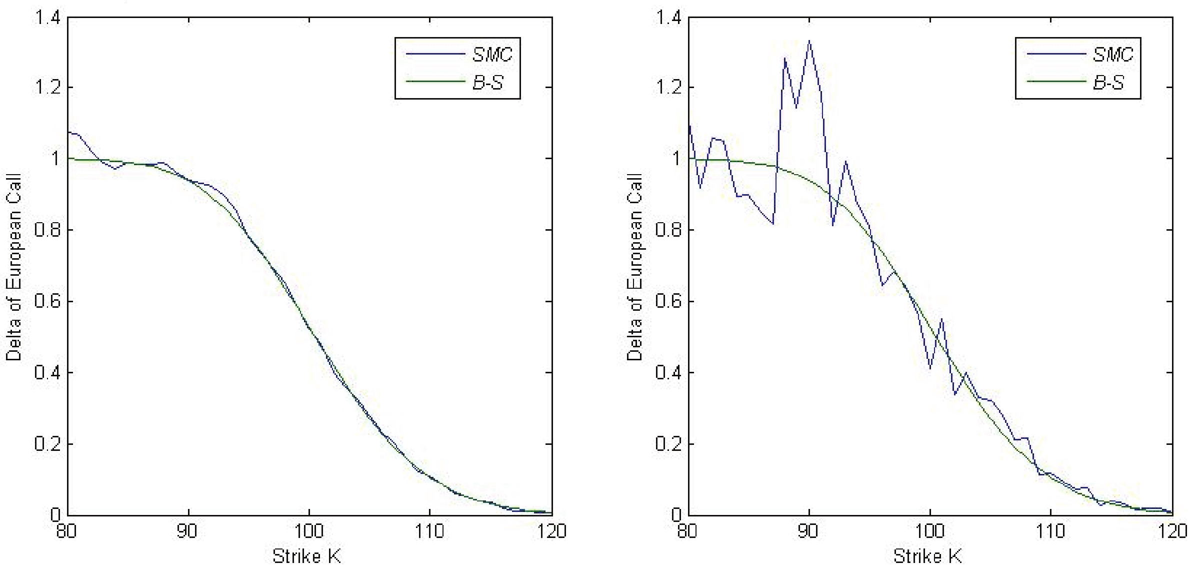

Figure 1.1: (Left): $(M, N, n, R) = (10000, 1000, 1, 1)$. (Right): $(M, N, n, R) = (1000, 1000, 10, 1)$.

Now, the algorithm is repeated for $(M, N, n, R) = (1000, 1000, 10, 1)$, which took 10009 seconds (2:47 hours). The results are displayed in Figure 1.1 (Right). The step size now is $h = 3/365$.

The time steps are increased from 1 to 10, and the number of paths simulated, M, are reduced from 10000 to 1000. From Theorem 1.5.1, the discrepancy between the SMC approximation for the Delta and the true Black-Scholes Delta is expected to increase, as the error propagates as time steps increase and the number of particles decreases. This is demonstrated by the large noise present in the SMC approximation in Figure 1.1 (Right).

Figure 1.1 (Right) shows how a reduction in the number of Monte Carlo paths and an increase in the number of time steps has a detrimental effect to the accuracy of the Delta approximation as expected by the theory. In the Black-Scholes scenario it is unnecessary to have a large number of particles approximating the "hidden" state of the volatility, as it is kept constant—indeed, fixing $N = 1$ yields the same results. Additionally, since the option is not path dependent, $n = 1$ suffices. For each strike, K, the calculations are repeated and a new set of particles approximations are created—the approach used here demonstrates the noise between the different runs.

Example 1.6.1. *Parallelisation across strikes and payoffs is possible so that Greeks for a family of options can be computed using the same SMC approximations. This approach provides "smoother" results as the same realisation is used as demonstrated in Figure 1.2. The run took 12 seconds, for* $(M, N, n, R) = (100000, 1, 1, 1)$*, and the same Black-Scholes parameters.*

Example 1.6.2. *Consider a Black-Scholes model with parameters* $(x, \sigma, r, T) = (100, 0.3, 0.03, 30/365)$*, and a call option with strike* $K = 100$*. The closed-form sensitivities for this option are* $(\Delta, \mathcal{V}, \Theta, \Gamma) = (0.528569, 11.40798, -22.2988, 0.046266)$*. Using the likelihood ratio method in the SMC framework with* $(M, N, n, R) = (1000000, 1, 1, 1)$ *yields* $(\hat{\Delta}, \hat{\mathcal{V}}, \hat{\Theta}, \hat{\Gamma}) = (0.529, 11.4, -22.4, 0.0463)$*, and took* 166 *seconds. Note that due to the absence of stochastic volatility, just one particle is used.*

From Theorem 1.5.1, recall that for fixed time steps n, and fixed number of particles N, the mean squared error of the Greek approximations has rate of convergence $\mathcal{O}(1/M)$, in the number of simulated trajectories. The algorithm is applied with parameters $(N, n, R) = (1, 1, 100)$, to the call option in the previous example, for $M = (2^i)_{i=10,\ldots,18}$, and compute the mean squared error over the $R = 100$ repeats. The results for the four Greeks are presented in Figure 1.3, with the displayed rate matching the predicted rate from Theorem 1.5.1.

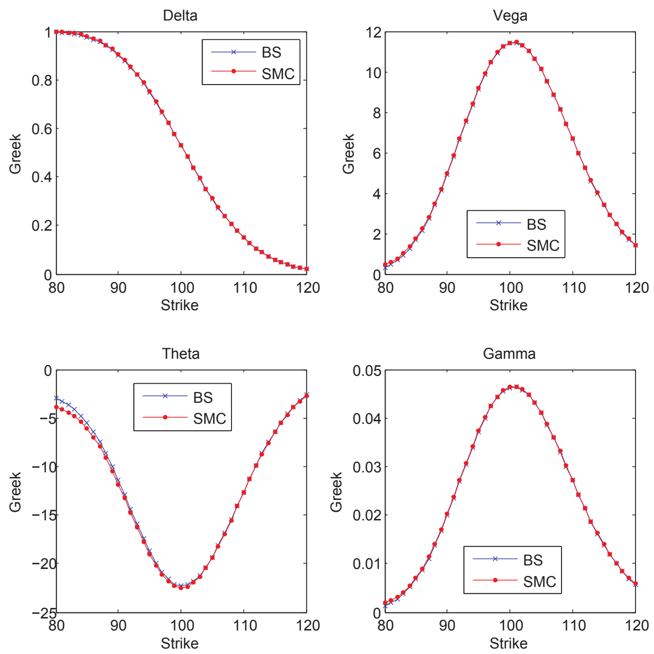

Figure 1.2: BS vs SMC, parallelised across $K = 80, \ldots, 120$. Greeks: Δ, $\mathcal{V}$, Θ and Γ.

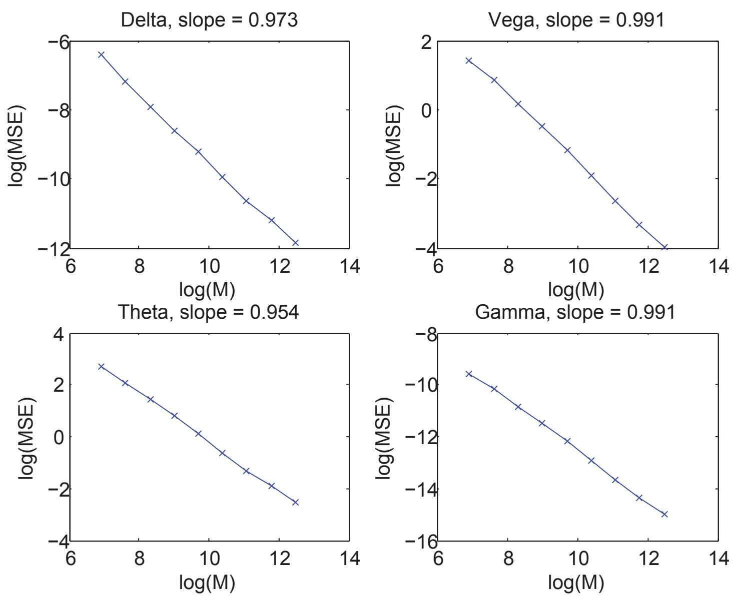

Figure 1.3: Rate of convergence for Δ, $\mathcal{V}$, Θ and Γ ($\log - \log$ scale).

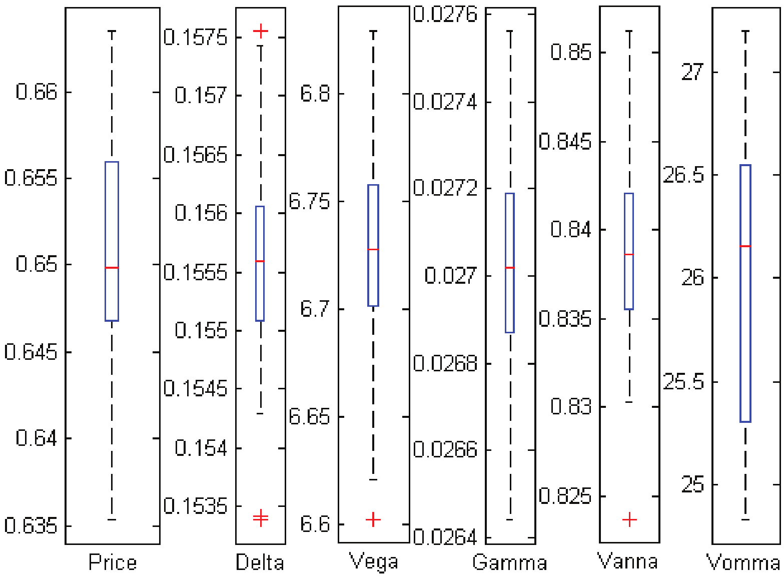

Figure 1.4: Box plots for the Greeks using Monte Carlo bumping.

1.6.2 Stochastic Volatility Greeks

Consider a call option with strike price $K = 110$, and the following parameters: $(x, \sigma, r, T, \eta) = (100, 0.3, 0.03, 30/365, 0.3)$, and compute the Greeks with respect to the underlying and the volatility using Monte Carlo simulation and finite differences. Greek computed using bumping with $(M, n, R) = (1000000, 10, 30)$ are shown in Figure 1.4. Each run took 40 seconds, to give a combined total of 20 minutes.

Remark 1.6.2. *The steps for the finite differences are $h_x := x/\sqrt{M}$ and $h_\sigma := \sigma/\sqrt{M}$. There was some level of tuning for choosing appropriate bump sizes, as picking the offsets too small or too large can produce poor results. Asymptotically, it is well known how to select the optimal h_x and h_σ to control the bias and variance [Gla03, Chapter 7.1.2].*

For the same option parameters, consider the SMC algorithm with parameters $(M, N, n, R) = (100000, 50, 10, 30)$, and the Greeks obtained are summarised in Figure 1.5. Each run took 450 seconds, for a combined total of 225 minutes, which is around ten-fold more computationally expensive compared with the bumping example above. Observe that the approximations using the SMC approach have much wider confidence intervals for the Greeks compared to the

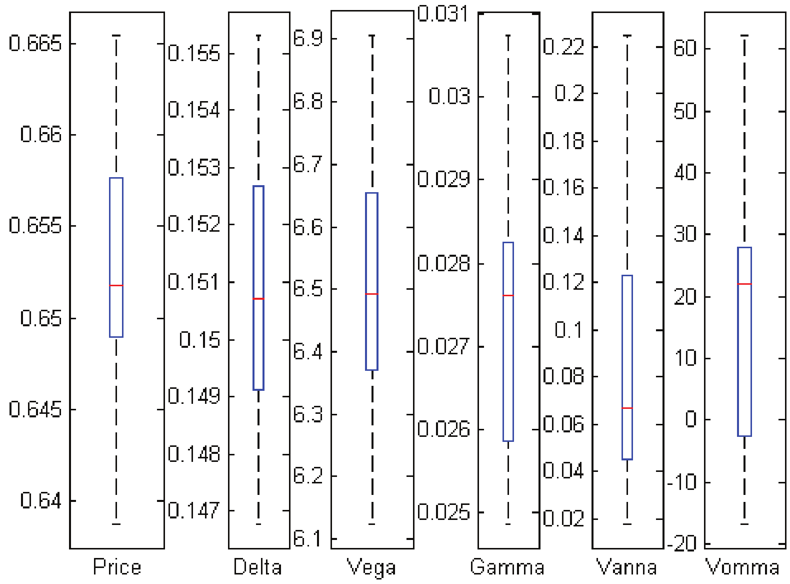

Figure 1.5: Box plots for the Greeks using SMC technique.

bumping approach. In addition, the Vomma approximation is particularly bad, suggesting a need for an increase in particles and paths. Having said that, the mean of the 30 approximations provides a workable Greek approximation, and the bumping approximations are within the interquartile range of the SMC approximations.

Example 1.6.3 (Resampling)**.** *The SMC algorithm included multinomial resampling and a resampling threshold of $N/3$. Repeating the experiment is now slightly more expensive (mainly due to computing the ESS at each time step), and now each run takes* 510 *seconds for the parameters* $(M, N, n, R) = (100000, 50, 10, 30)$. *The results are presented in Figure 1.6. The results are comparable to those without resampling, especially as the rate of resampling is really low (an average of* 0.015 *resampling steps per path). Resampling would have a greater effect for a larger n, when more resampling steps are required.*

Example 1.6.4. *Recall Section 1.5 (recall Corollary 1.5.1). Fix the number of repeats to $R = 250$ and consider $n = 2, 4, 8, \ldots, 64$. The measure of error used is the Mean Absolute Error (MAE). Fix the number of particles N to be proportional to $n^{1/2}$ and the number of paths M to be proportional to n^2, since a weak convergence of order one is supposed for the explicit Euler scheme. In Figure 1.7, convergence for the Greek approximations is observed, albeit with a slow rate for the Vomma. The rate of*

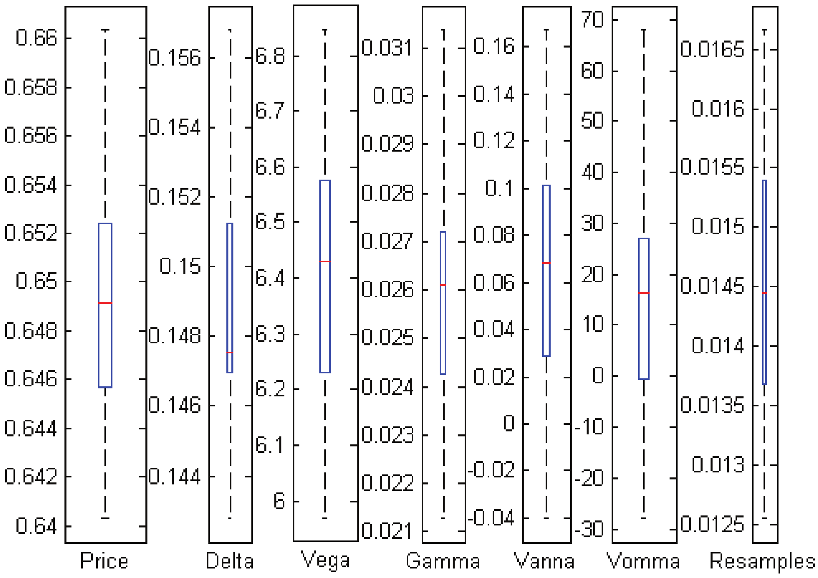

Figure 1.6: Box plots for the Greeks using SMC technique, and multinomial resampling.

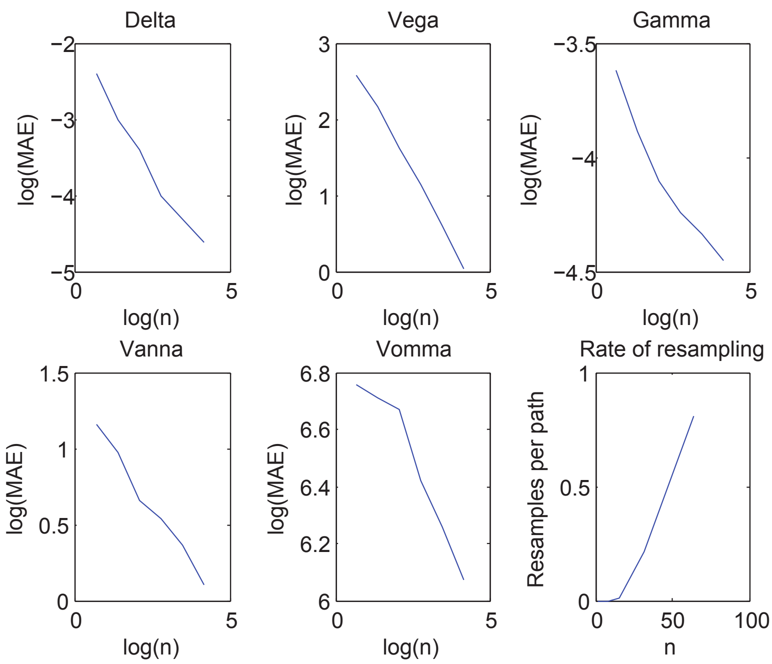

Figure 1.7: Example 1.6.4. Mean absolute error vs n (log − log scale).

resampling increases exponentially for a larger number of time steps, as expected.

1.6.3 Tracking Option prices using Greeks

In this section, an option price is tracked using Greeks, from 02/09/2010 to 28/09/2010. Consider the S&P European call option, with strike $K = 1100$. Taylor expanding the option price, V, using Greeks V_i, $V_{i,j}$ and $V_{i,j,k}$, leads to

$$\mathrm{d}V = \sum_{i=1}^{d} V_i \mathrm{d}\theta_i + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} V_{i,j}\mathrm{d}\theta_i \mathrm{d}\theta_j + \frac{1}{6}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} V_{i,j,k}\mathrm{d}\theta_i \mathrm{d}\theta_j \mathrm{d}\theta_k + \dots \,. \tag{1.6.1}$$

Consider the Taylor expansion

$$\mathrm{d}V = \Delta \mathrm{d}S + \frac{1}{2}\Gamma\left(\mathrm{d}S\right)^2 + \mathcal{V}\mathrm{d}\sigma + \Theta \mathrm{d}T + \epsilon \,. \tag{1.6.2}$$

where ϵ denotes the error, and Δ, $\mathcal{V}$, Γ and Θ are the Delta, Vega, Gamma and Theta.

The SMC algorithm is ran at each time step to calculate the Greeks and track the option price with the Taylor expansion in (1.6.2). For the stochastic volatility model in Example 1.4.2 tracking is superior to that of the Black-Scholes Greeks, as shown in Figure 1.8. The initial volatility is chosen as the implied volatility, computed every day. For the stochastic volatility set-up, a value for η is required; this was approximated to 0.0042 for the option data, by calculating the standard deviation of historical implied volatilities. For each data point, the algorithm was ran using the stochastic volatility model with parameters $(M, N, n, R) = (100, 10, 1, 1000)$ and $(r, \sigma, \eta) = (0.007, 0.22, 0.0042)$, with a total run time of 2:02 hours for all 16 data points. The other input parameters are all market observed. The absolute error terms, $|\epsilon|$, from (1.6.2) are studied. Denote ϵ_{BS} as the error term in the call price after Taylor expansion using the Black-Scholes Greeks, and ϵ_{SV} as the error from Taylor expanding using Greeks from the stochastic volatility model. The mean and variance of the absolute errors compare favourably for the SMC Taylor expansion in comparison to the Taylor expansion using the Black-Scholes Greeks. The errors in tracking using the Black-Scholes Greeks have moments $\mathbb{E}|\epsilon_{BS}| = 6.48$ and $\mathbb{V}|\epsilon_{BS}| = 9.26$. For the SMC tracking, the moments for the errors are $\mathbb{E}|\epsilon_{SV}| = 2.13$ and $\mathbb{V}|\epsilon_{SV}| = 2.21$. There is a considerable amount of improvement in the tracking ability.

Remark 1.6.3. *The tracking of option prices using the Greeks is highly dependent on the movement of*

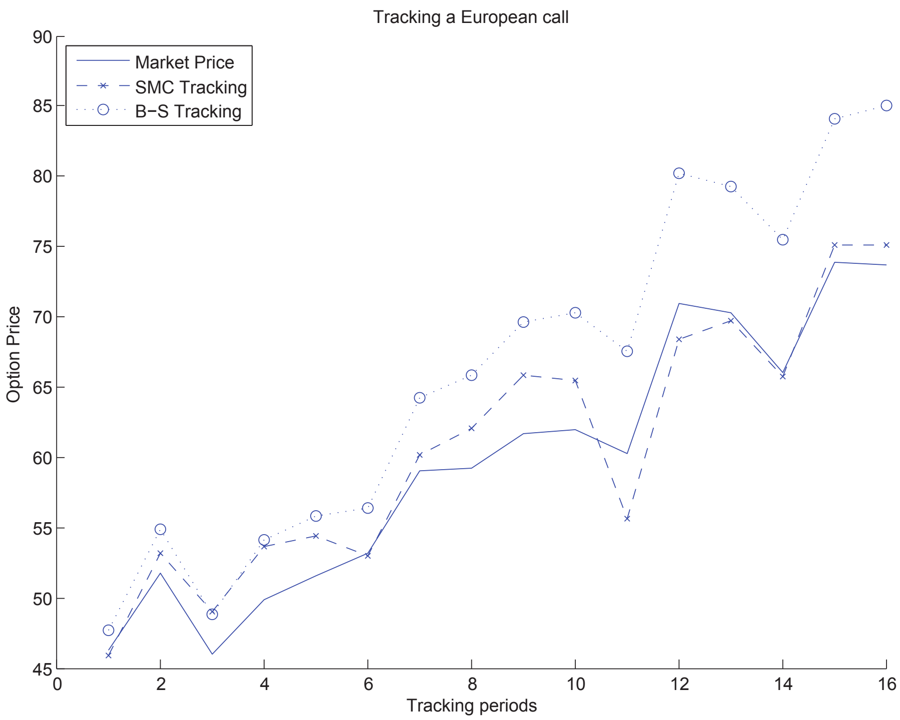

Figure 1.8: Market observed call price, tracked call price using the stochastic volatility model (SMC Tracking) and Black-Scholes tracking.

the underlying every day, $\mathrm{d}S$*, as it affects the tracked price due to the Delta and Gamma. This plays a very important role in the tracking of an option, as it often contributes the biggest change in the value of a derivative. Other components such as the* $\mathrm{d}T$ *(change in time) are obviously predictable due to the arrow of time, and constant expiration of the option. For weekends, and non-working day, it is assumed that several days have gone by; in other words, time is not "stopped" over weekends.*

The results show that robust Greek calculations from a stochastic volatility model can potentially improve the tracking ability of the option price. More Greeks could be used for the Taylor expansion, however they have very little effect on the tracked price.

1.7 Future work

A general framework for calculating option Greeks is introduced and an SMC method is applied to a real financial application. Convergence results have been provided, and optimal choices of the number of time steps, paths and particles have been discussed for simulation schemes of varying weak rates of convergence.

The focus throughout has been on a forward-only smoothing algorithm. This has been with the view that the method is "online" in the sense that the backward filter does not get more expensive with time. As an alternative, the two-filter smoothing method has the potential advantage of putting samples in desirable regions of the state for the particle filter [BDM10]. The trade-off for this method is that it is offline, however it would be interesting to study results from different SMC methodologies.

Example 1.7.1 (Bermudan options)**.** *This SMC framework lends itself to computing Greeks for Bermudan options. The pricing of Bermudan/American-style options can be separated into considering low and high-biased estimates, and there have been important contributions in the pricing of such options by simulation and regression techniques [BDGT00, LS01, Til93, Car96]. The filtering density could be incorporated into the regression functions used for providing low-biased estimates of the option price, i.e. regression in order to decide when the option is exercised [RB10].*

As we have seen in (1.6.1), the option price can be tracked with accurate Greek calculations, with the tracking error providing a measure for different stochastic volatility models. Further investigating is required to determine which model provides Greeks that track the option price best. The motivation behind this approach is that one would be able to compare models using the tracking ability of options.

Part II

A class of approximate Greek weights

2. Theory

There exist several methods for estimating the price sensitivities ("Greeks") for contingent claims: PDE methods, finite-difference approximations through re-simulation, pathwise techniques, likelihood ratio methods, perturbation techniques and Malliavin calculus. The proposed approach uses Itô-Taylor approximations, and $\mathcal{F}_\vartheta$-measurable weights for the option payoff for some small time $0 < \vartheta \leq T$, that produce biased estimates for the Greeks. We derive and analyse Monte Carlo estimators for the Greeks in a general setting, including some families of stochastic volatility models. In certain cases, the Greek weights obtained coincide with those arising in the Malliavin Greeks literature (Bachelier/Black-Scholes model).

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space equipped with an m-dimensional $(\mathcal{F}_t)_{t\in[0,T]}$-adapted Brownian motion $W = (W_t)_{t\in[0,T]}$. Let $X = (X_t)_{t\geq 0}$ be a d-dimensional process, with components $X_t^{(1)}, \ldots, X_t^{(d)}$, $f : [0,T] \times \mathbb{R}^d \times \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}^d$ and $\gamma : [0,T] \times \mathbb{R}^d \times \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}^{d\times m}$ be two functions.

We interpret process X as the stock prices. The payoff of an option depends on X, which is the strong solution to the stochastic differential equation

$$X_t^{x,\theta} = x + \int_0^t f(t, X_t^{x,\theta}, \theta)\mathrm{d}t + \int_0^t \gamma(t, X_t^{x,\theta}, \theta)\mathrm{d}W_t, \qquad X_0^{x,\theta} = x \in \mathbb{R}^d, \tag{2.0.1}$$

where $\theta \in \mathbb{R}^{\mathfrak{d}}$ is a set of given parameters. We shall drop x, θ in the above notation when it is clear from the context. For the i$^{\text{th}}$ component of X, interpret the above as

$$X_t^{(i)} = X_0^{(i)} + \int_0^t f_i(s, X_s, \theta)\mathrm{d}s + \sum_{j=1}^m \int_0^t \gamma_{i,j}(s, X_s, \theta)\mathrm{d}W_s^{(j)}.$$

The use of SDEs to model financial assets is well studied for pricing contingent claims (see [Hul14, Gla03, KN12] and references therein). Fix $T > 0$ as the time horizon of interest. For some $n \in \mathbb{N}$, define the grid $\pi := \{0 = t_0 < t_1 < \ldots < t_n = T\}$. Let $g : \mathbb{R}^{dn} \to \mathbb{R}$ be the payoff function, and in the classical theory of the financial markets, we define the option price $V(x)$ as an expectation given the initial condition, $X_0 = x$, namely

$$V(x) := \mathbb{E}\left[g(X_{t_1}, \ldots, X_{t_n})|X_0 = x\right]. \tag{2.0.2}$$

A Monte Carlo approximation of the option price using N simulated trajectories (assuming that the process can be perfectly simulated), where path j is denoted by $(X_{t_i}^{(j)})_{i=0,\dots,n}$, is

$$V^N(x) := \frac{1}{N}\sum_{j=1}^{N} g(X_{t_1}^{(j)}, \dots, X_{t_n}^{(j)}).$$

Our aim is to provide a representation of the option price sensitivities with respect to the initial condition, $\frac{\partial}{\partial x}V(x)$. Later, we approximate sensitivities with respect to the parameter θ in (2.0.1), namely vector $(\frac{\partial}{\partial \theta_i}V(x))_{i=1,\dots,\mathfrak{d}}$. Common Greeks include the Delta (Δ), defined as the sensitivity of the option price with respect to the initial value, and the Vega ($\mathcal{V}$), i.e. the sensitivity of the option price with respect to the volatility.

For smooth payoff functions g, the Greeks can be approximated through the pathwise approach (see [GM02, Proposition 1.1]). An obvious constraint is the typical non-smoothness exhibited by common payoffs.

An intuitive method to compute Greeks is through re-simulation of the option price $V(x)$, for different values of x, and approximating the option sensitivities through a finite difference. For example, the Delta can be approximated by the forward difference, $\left(V^N(x+\varepsilon) - V^N(x)\right)/\varepsilon$, for some small $\varepsilon > 0$, which has convergence of order $\mathcal{O}(N^{-1/4})$; for a central difference scheme, $\left(V^N(x+\varepsilon) - V^N(x-\varepsilon)\right)/(2\varepsilon)$, this can be improved to $\mathcal{O}(N^{-1/3})$ [Gly89, YK91]. Furthermore, by using common random numbers and a central difference scheme, better convergence results up to the Monte Carlo rate of convergence $\mathcal{O}(N^{-1/2})$ can be achieved. This approach can perform poorly for non-smooth payoff functions and exotic options [GY92].

An alternative method is the likelihood ratio method (LRM), where the computation of Greeks is achieved by $\mathbb{E}[g(X_T)H]$, for random weight H (see [BG96]). Such representation removes the necessity that the payoff is smooth.

The connection between the pathwise and likelihood ratio methods, to Malliavin calculus is explored in [CG07]. Malliavin calculus allows Greeks to be expressed in the form

$$\frac{\partial V(x)}{\partial \theta} = \mathbb{E}\left[g(X_{t_1}, \dots, X_{t_n})\pi_\theta | X_0 = x\right],$$

where π_θ is some weight associated to a sensitivity in the θ direction [FLL+99, Ben01]. Using Malliavin techniques with Monte Carlo simulation allows convergence rates of $\mathcal{O}(N^{-1/2})$. Malliavin weights are known for certain families of jump-diffusion processes [DJ06, EKP04].

In a one-dimensional setting for the underlying asset, the Δ in a Black-Scholes framework has a Malliavin weight of $w_x = W_T/(x\gamma T)$, which is a function of the Brownian motion driving the underlying process and the constant volatility, γ. The Vega of an option, $\mathcal{V}$, can be computed using the Malliavin weight $w_\gamma = W_T^2/(\gamma T) - W_T - 1/\gamma$, and more examples for Greek weights can be found in [FLL$^+$99].

Another family of methods for approximating option prices and sensitivities is the asymptotic expansion schemes. The asymptotic expansion approach introduces a perturbation in the model for approximating an option price. This method perturbs the general SDE in (2.0.1) to

$$\mathrm{d}X_t^{x,\theta,\varepsilon} = f(t, X_t^{x,\theta,\varepsilon}, \theta)\mathrm{d}t + \varepsilon\gamma(t, X_t^{x,\theta,\varepsilon}, \theta)\mathrm{d}W_t, \quad X_0^{x,\theta,\varepsilon} = x \in \mathbb{R}^d, \quad \theta \in \mathbb{R}^{\eth}, \qquad \forall t \geq 0, \tag{2.0.3}$$

for some $\varepsilon \in (0, 1]$. Using small order expansions, up to selected bias $\mathcal{O}(\varepsilon^k)$, for $k \in \mathbb{N}^+$, the option price is approximated under the perturbed model [KT01, MTU04]; there have been extensions for stochastic volatility models in a Markovian setting [KT03]. The validity in the Black-Scholes setting has been demonstrated for an expansion of the option price [KT03, Theorem 3.3]. There has been efforts to apply control variate techniques to reduce the variance when applying asymptotic expansions [MTU04], and more recent results on strong convergence using accelerated schemes [TY12]. An extension to this family of techniques is the asymptotic expansion of the perturbed first variation process. This can be used to derive first-order Greeks [MTU04, Theorem 2, 3].

Motivation: We shall consider a general technique and derive biased estimates for the Greeks, using expansions of the value function u. These approximations will take the form

$$\text{Greek} = \mathbb{E}[g(X_T)w_\vartheta] + \mathcal{O}(\vartheta^l),$$

for some order l depending on the smoothness of the value function and the type of weight considered. In addition, we work under fully implementable schemes and discuss the convergence rate of the Greek approximations using either the strong or the weak rates of convergence of the process X. The idea above will be formalised and higher-order Greek expansions will be derived.

Notations: In the following, denote by C a constant that depends only on T, f, γ, x, θ, but whose value does not depend on the number of steps n (its value may change from line to line). Denote a constant by C_r if it depends on any additional parameter r. Let $\mathcal{C}$ be the set of continuous functions, and denote $\mathcal{C}_b$ the subset of bounded continuous functions. Denote by $\mathcal{C}_b^l$ the set of continuous functions, whose first l derivatives are continuous and bounded. Denote by $\mathcal{C}_p$ the set of continuous functions φ with at most polynomial growth, e.g. for all x it holds that $|\varphi(x)| \leq C(1+|x|^q)$ for some $q > 0$. Let $\mathbb{N}^+$ be the set of strictly positive integers, and $\mathbb{N} := \mathbb{N}^+ \cup \{0\}$. For a matrix M, denote by M^* its transpose. We shall denote by $\mathcal{O}(h^k)$ that the limit for small h is such that

$$\lim_{h\to 0} \frac{\mathcal{O}(h^k)}{h^k} = C,$$

for some constant C that does not depend on n.

2.1 Introduction

Let $(\Omega, \mathcal{F}, (\mathcal{F})_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space equipped with an m-dimensional $(\mathcal{F})_{t\in[0,T]}$-adapted Brownian motion $W = (W_t)_{t\in[0,T]}$. In the setting where f, γ, g are Lipschitz continuous, define $Y_t = \mathbb{E}[g(X_T)|\mathcal{F}_t] = u(t, X_t)$, where $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ is the solution (possibly in the viscosity sense) to the partial differential equation

$$L^{(0)}u(t,\cdot) = 0, \quad t \in [0,T), \text{ and } \quad u(T,\cdot) = g(\cdot), \tag{2.1.1}$$

with the operators $L^{(1)}, \ldots, L^{(m)}, L^{(0)}$ defined as (we use the notation $\partial_x \equiv \frac{\partial}{\partial x}$)

$$L^{(j)} := \sum_{k=1}^{d} \gamma_{k,j}\partial_{x_k} \qquad \text{for } j = 1, \ldots, m, \tag{2.1.2}$$

$$L^{(0)} := \partial_t + \sum_{k=1}^{d} f_k \partial_{x_k} + \frac{1}{2}\sum_{k=1,j=1}^{d} a_{k,j}\partial_{x_k}\partial_{x_j}, \tag{2.1.3}$$

and $a = (a_{k,j}) = \gamma\gamma^*$. Under smoothness assumptions on the coefficients, it is possible to show that $u \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d \to \mathbb{R})$ is a classical solution to the above PDE. In this case application of Itô's formula, yields the representation

$$Y_t = g(X_T) - \int_t^T Z_t^* \mathrm{d}W_t, \qquad Y_T = g(X_T), \tag{2.1.4}$$

where $Y \in \mathbb{R}$, $Z \in \mathbb{R}^d$, $g : \mathbb{R}^d \to \mathbb{R}$ is a measurable function with polynomial growth, and

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

2.2 Multi-indices and Stochastic Taylor expansions

We review the multi-indices notation from [KP92, Chapter 5]. For $l \in \mathbb{N}^+$, define a multi-index $\alpha = (j_1, j_2, \ldots, j_l)$, with $j_i \in \{0, 1, \ldots, m\}$. The number of elements in α is denoted by $l := l(\alpha)$. Let $n(\alpha)$ denote the number of null components of α and $\mathcal{M}$ the set of multi-indices such that

$$\mathcal{M} := \left\{\alpha = (j_1, j_2, \ldots, j_l) : j_i \in \{0, 1, \ldots, m\}, i \in \{1, \ldots, l\} \text{ for } l, m \in \mathbb{N}^+\right\} \cup \{\emptyset\},$$

where $\emptyset$ is the multi-index of length zero (i.e. $l(\emptyset) = 0$). Define the following operations on α, with $l(\alpha) \geq 1$: $-\alpha = (j_2, \ldots, j_l)$, $\alpha- = (j_1, \ldots, j_{l-1})$ and for completeness, $-(j) = (j)- = \emptyset$, for all $j \in \{0, \ldots, m\}$. We define $*$ to be the concatenation operator such that $\alpha * \bar{\alpha} := \left(j_1, \ldots, j_l, \bar{j}_1, \ldots, \bar{j}_l\right)$. For α such that $l(\alpha) \geq 1$, define α^+ to be the multi-index α with null entries removed. Denote by $(j)_l$ the multi-index of length l, with entries all equal to j.
The continuous and adapted process φ belongs to $S^2([0, T])$ if $\mathbb{E}\Big[\sup_{s \in [0,T]} |\varphi_s|^2\Big]$ is finite.

Definition 2.2.1 ([KP92, (5.2.12)]). *Let $\alpha \in \mathcal{M}$ and $\varphi : [0, T] \to \mathbb{R}$, such that $\varphi \in S^2([0, T])$; define the multiple Itô integrals for all $0 \leq s \leq t \leq T$ by*

$$I^\alpha_{s,t}[\varphi(\cdot)] = \begin{cases} \varphi(t), & \text{if } \alpha := \emptyset, \\ \displaystyle\int_s^t I^{\alpha-}_{s,u}[\varphi(\cdot)]\mathrm{d}u, & \text{if } l(\alpha) > 0 \text{ and } j_{l(\alpha)} = 0, \\ \displaystyle\int_s^t I^{\alpha-}_{s,u}[\varphi(\cdot)]\mathrm{d}W^{(j_{l(\alpha)})}_u, & \text{if } l(\alpha) > 0 \text{ and } j_{l(\alpha)} \geq 1, \end{cases} \quad \text{for } s, t \in [0, T]. \tag{2.2.1}$$

For shorthand define $I^\alpha_t[\varphi(\cdot)] := I^\alpha_{0,t}[\varphi(\cdot)]$, for integrals beginning at time zero.

In particular, for $\alpha = (j)$ for all $j \in \{1, \ldots, m\}$, then $I^{(j)}_t[\varphi(\cdot)] = \int_0^t \varphi(s)\mathrm{d}W^{(j)}_s$. For $\alpha \in \mathcal{M} \setminus \{\emptyset\}$, denote by L^α the operator

$$L^\alpha := L^{(j_1)} \circ L^{(j_2)} \circ \ldots \circ L^{(j_l)}, \tag{2.2.2}$$

where $L^{(i)} \circ L^{(j)} u_\cdot \equiv L^{(i)}\left(L^{(j)} u_\cdot\right)$. We shall write $I^{(k)}_t = I^{(k)}_t[1]$, and $u^\alpha_\cdot \equiv L^\alpha u_\cdot$.
For $\alpha \in \mathcal{M}$, define $k_0(\alpha)$ as the number of null components before the first non-zero component, and $k_i(\alpha)$ for $i = 1, \ldots, l(\alpha^+)$ as the number of null components in α between the i^{th} and

$(i+1)^{\text{th}}$ non-zero components. For $\alpha, \beta \in \mathcal{M}$ define

$$w(\alpha, \beta) := l(\alpha^+) + \sum_{i=0}^{l(\alpha^+)} (k_i(\alpha) + k_i(\beta)), \tag{2.2.3}$$

which is the number of non-zero components in α, plus the total number of null components in α and β. We quote a useful result, which is a simplification of [KP92, Lemma 5.7.2]:

Lemma 2.2.1. *Let $\alpha, \beta \in \mathcal{M}$. Then, for any $t \in [0, T]$,*

$$\mathbb{E}\left[I_t^\alpha I_t^\beta\right] = \begin{cases} 0, & \text{if } \alpha^+ \neq \beta^+, \\ \dfrac{t^{w(\alpha,\beta)}}{w(\alpha,\beta)!} \displaystyle\prod_{i=0}^{l(\alpha^+)} C_{k_i(\alpha)+k_i(\beta)}^{k_i(\alpha)}, & \text{if } \alpha^+ = \beta^+, \end{cases} \tag{2.2.4}$$

where $C_i^k := i!/(k!(i-k)!)$ is the usual combinatorial notation.

Itô-Taylor expansions for diffusion processes provide an extension to Itô's formula for a smooth function. These stochastic Taylor expansions can be written concisely using hierarchical sets of multi-indices:

Definition 2.2.2. *A set $\mathcal{A} \subset \mathcal{M}$ is called hierarchical if:*

1. *$\mathcal{A}$ is nonempty;*
2. *$\sup_{\alpha \in \mathcal{A}} l(\alpha)$ is finite;*
3. *for any $\alpha \in \mathcal{A} \setminus \{\emptyset\}$, $-\alpha \in \mathcal{A}$.*

The corresponding remainder set, $\mathcal{B}(\mathcal{A})$, is defined by

$$\mathcal{B}(\mathcal{A}) := \{\alpha \in \mathcal{M} \setminus \mathcal{A} | -\alpha \in \mathcal{A}\}.$$

The following result generalises Itô-Taylor expansions for diffusions:

Theorem 2.2.1 ([KP92, Theorem 5.5.1])**.** *Let ϑ be a stopping time such that $0 \le \vartheta \le T$ almost surely, and let $\mathcal{A} \subset \mathcal{M}$ be a hierarchical set. For X defined in* (2.0.1) *and $u : [0, T] \times \mathbb{R}^d \to \mathbb{R}$, the Itô-Taylor expansion*

$$u(\vartheta, X_\vartheta) = \sum_{\alpha \in \mathcal{A}} I_\vartheta^\alpha [L^\alpha u(0, X_0)] + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_\vartheta^\alpha [L^\alpha u(\cdot, X_\cdot)] \tag{2.2.5}$$

holds, provided that the right-hand side is well defined.

Remark 2.2.1. *The theorem above generalises Itô's Lemma; let $\mathcal{A} = \{\emptyset\}$, since*

$$u(\vartheta, X_\vartheta) = I_\vartheta^\emptyset[u(0, X_0)] + \sum_{\alpha \in \mathcal{B}(\emptyset)} I_\vartheta^\alpha[L^\alpha u(\cdot, X_\cdot)].$$

Clearly, $\mathcal{B}(\emptyset) = \{(0), (1), \ldots, (m)\}$, *and hence the equality* $u(\vartheta, X_\vartheta) = u(0, X_0) + \int_0^\vartheta L^{(0)} u(s, X_s) \mathrm{d}s + \sum_{j=1}^m \int_0^\vartheta L^{(j)} u(s, X_s) \mathrm{d}W_s^{(j)}$ *directly follows from* (2.2.1).

A priori regularity assumptions will be imposed on the value function u (we sometimes abbreviate $u_t := u(t, X_t)$).

Definition 2.2.3. *Let $\alpha \in \mathcal{M} \setminus \{\emptyset\}$, and define $\mathcal{G}_b^\alpha$ as the set of functions $u : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ such that $L^\alpha u_\cdot$ is well defined, continuous and bounded.*

The following assumptions on the value function impose bounds on $L^\alpha u_\cdot$ for different lengths of multi-index α, for some fixed $l \in \mathbb{N}^+$:
($\mathbf{H}u_b^l$): $u \in \mathcal{G}_b^\alpha$ for all $\alpha \in \mathcal{M} \setminus \{\emptyset\}$ such that $l(\alpha) \leq l$.

Example

Recall the process X defined in (2.0.1), and fix $d = m$. The price of an option on X with payoff g is defined via (2.0.2), and we are interested here in computing sensitivities with respect to $X_0^{(1)}, \ldots, X_0^{(d)}$. For $r \in \mathbb{N}$, define the hierarchical set $\mathcal{D}_r := \{\alpha \in \mathcal{M} | l(\alpha) \leq r\}$, with the corresponding remainder set $\mathcal{B}(\mathcal{D}_r) = \mathcal{D}_{r+1} \setminus \mathcal{D}_r = \{\alpha \in \mathcal{M} | l(\alpha) = r+1\}$. Let u be a sufficiently smooth value function. Applying Theorem 2.2.1 with $\mathcal{A} = \{\emptyset\}$ to u, in terms of the operators (2.1.3) and (2.1.2), for any $t \in [0, T]$, we obtain

$$u_t = u_0 + \sum_{j=0}^d I_t^{(j)}[L^{(j)} u_\cdot]. \tag{2.2.6}$$

Similarly, applying Theorem 2.2.1 to $L^{(0)} u, \ldots, L^{(d)} u$ yields

$$L^{(j)} u_t = L^{(j)} u_0 + \sum_{i=0}^d I_t^{(i)}[L^{(i)} \circ L^{(j)} u_\cdot], \qquad j = 0, \ldots, d,$$

and substituting these in (2.2.6), yields the expansion of u_t denoted by

$$u_t = u_0 + \sum_{j=0}^d I_t^{(j)}[L^{(j)} u_0] + \sum_{\alpha \in \mathcal{B}(\mathcal{D}_1)} I_t^\alpha[L^\alpha u_\cdot]. \tag{2.2.7}$$

We motivate the sequel with a result on approximating first-order Greeks with a bias of order $\mathcal{O}(\vartheta)$, for some $\vartheta \in (0, T]$.

Remark 2.2.2. *In a two-dimensional financial setting ($m = d = 2$), suppose that $X^{(1)}$ denotes the instantaneous volatility process and $X^{(2)}$ the underlying asset. In this setting, $\partial_{x_1} u_0 := \frac{\partial}{\partial X_s^{(1)}} u(s, X_s)\Big|_{s=0}$ is the Vega and $\partial x_2 u_0$ is the Delta, so that* (2.2.8) *allows us to solve simultaneously for these first-order Greeks. Throughout, we assume that the diffusion coefficient γ is uniformly positive definite at the initial time.*

The proof of the following result is left to Appendix A.1:

Proposition 2.2.1. *Assume $(\mathbf{H}u_b^2)$ and $d = m$. Then, for $\vartheta \in (0, T]$ and $j = 1, \ldots, d$,*

$$\mathbb{E}\left[g\left(X_T\right) \frac{I_\vartheta^{(j)}}{\vartheta}\right] = \sum_{l=1}^{d} \gamma_{l,j}(x) \partial_{x_l} u_0 + \mathcal{O}(\vartheta). \tag{2.2.8}$$

2.3 Expansion in d dimensions for general order

Fix some $l \in \mathbb{N}^+$ throughout this section. Define the set of multi-indices $\mathcal{M}_{i,j,k}$:

Definition 2.3.1. *For $i, k \in \mathbb{N}^+, i \geq r$, define $\mathcal{M}_{i,r,k}$ as the set of multi-indices of length i, that have r indices equal to k. Formally,*

$$\mathcal{M}_{i,r,k} := \left\{ \alpha = (j_1, \ldots, j_i) \in \mathcal{M} \setminus \{\varnothing\} : l(\alpha) = i, \sum_{p=1}^{i} 1_k(j_p) = r \right\},$$

where $1_k(j) = 1$ if $j = k$ and zero otherwise.

The next proposition generalises Proposition 2.2.1 (proof in Appendix A.1):

Proposition 2.3.1. *Let u be the solution to* (2.1.1) *and suppose that $(\mathbf{H}u_b^{l+1})$ holds. Then, for $j = 1, \ldots, d$ and $\vartheta \in (0, T]$,*

$$\mathbb{E}\left[g(X_T) \tfrac{I_\vartheta^{(j)}}{\vartheta}\right] = u_0^{(j)} + \sum_{i=2}^{l} \sum_{\substack{\alpha \in \mathcal{M}_{i,1,j} \\ j_i \neq 0}} u_0^{\alpha} \frac{\vartheta^{i-1}}{i!} + \mathcal{O}(\vartheta^l). \tag{2.3.1}$$

It is possible to obtain expressions containing higher-order Greeks using this approach. This time u_ϑ is multiplied by I_ϑ^α, for $\alpha = (j_1, j_2)$ with $j_1, j_2 \in \{1, \ldots, d\}$. The results are presented using the notations from Definition 2.3.1 and throughout rely on Lemma 2.2.1.

Proposition 2.3.2 (Second-order expansion)**.** *Let u be the solution to* (2.1.1) *and assume* $(\mathbf{H}u_b^{l+2})$ *holds. Then, for $\vartheta \in (0,T]$ and all $j = 1,\dots,d$,*

$$\mathbb{E}\left[g(X_T)\frac{2I_\vartheta^{(j,j)}}{\vartheta^2}\right] = u_0^{(j,j)} + \sum_{i=3}^{l+1} \sum_{\substack{\alpha\in\mathcal{M}_{i,2,j}\\ j_i\neq 0}} 2u_0^\alpha \frac{\vartheta^{i-2}}{i!} + \mathcal{O}(\vartheta^l). \tag{2.3.2}$$

Proof. Set $\beta := (1,1)$ and use the approach from Proposition 2.3.1. □

Proposition 2.3.3 (Second-order cross terms expansion)**.** *Let* $(\mathbf{H}u_b^3)$ *hold. Then, for $j_1, j_2 \in \{1,\dots,d\}$, it follows that* $\mathbb{E}\Big[2g(X_T)I_\vartheta^{(j_1,j_2)}/\vartheta^2\Big] = u_0^{(j_1,j_2)} + \mathcal{O}(\vartheta)$.

Proof. Expressing $\mathbb{E}\left[u_\vartheta I_\vartheta^{(1,2)}\right]$ for some $\vartheta \in (0,T]$, and Lemma 2.2.1 yield

$$\mathbb{E}\Big[u_\vartheta I_\vartheta^{(1,2)}\Big] = \sum_{\alpha\in\mathcal{D}_2} \mathbb{E}\Big[I_\vartheta^\alpha[u_0^\alpha] I_\vartheta^{(1,2)}\Big] + \sum_{\alpha\in\mathcal{B}(\mathcal{D}_2)} \mathbb{E}\Big[I_\vartheta^\alpha[u_\cdot^\alpha] I_\vartheta^{(1,2)}\Big] = \frac{\vartheta^2}{2}u_0^{(1,2)} + \mathcal{O}(\vartheta^3);$$

since the terms in $\mathcal{B}(\mathcal{D}_2)$ and the boundedness of $L^\alpha u_\cdot$, it holds that

$$\left|\sum_{\alpha\in\mathcal{B}(\mathcal{D}_2)} \mathbb{E}\Big[I_\vartheta^\alpha[u_\cdot^\alpha] I_\vartheta^{(1,2)}\Big]\right| \leq C\mathbb{E}\Big[\Big(I_\vartheta^{(0,1,2)} + I_\vartheta^{(1,0,2)} + I_\vartheta^{(1,2,0)}\Big) I_\vartheta^{(1,2)}\Big] = \mathcal{O}(\vartheta^3).$$

Therefore, $\mathbb{E}\Big[2u_\vartheta I_\vartheta^{(1,2)}/\vartheta^2\Big] = u_0^{(1,2)} + \mathcal{O}(\vartheta)$ and similarly $\mathbb{E}\Big[2u_\vartheta I_\vartheta^{(2,1)}/\vartheta^2\Big] = u_0^{(2,1)} + \mathcal{O}(\vartheta)$. □

Remark 2.3.1. *Terms such as $I_\vartheta^{(j,j)} = \int_0^\vartheta W_s^{(j)} \mathrm{d}W_s^{(j)}$ are easy to compute using Itô's formula. The equality $\int_0^\vartheta W_s^{(j)} \mathrm{d}W_s^{(j)} = (W_\vartheta^{(j)})^2/2 - \vartheta/2$ will be exploited for the MC simulation. Terms such as $I_\vartheta^{(j_1,j_2)} = \int_0^\vartheta W_s^{(j_1)} \mathrm{d}W_s^{(j_2)}$ are difficult to compute directly; for cross terms, with $j_1 \neq j_2$, $\mathbb{E}\left[I_\vartheta^{(j_1,j_2)} + I_\vartheta^{(j_2,j_1)}\right] = \mathbb{E}\left[I_\vartheta^{(j_1)} I_\vartheta^{(j_2)}\right] = \mathbb{E}\left[W_\vartheta^{(j_1)} W_\vartheta^{(j_2)}\right]$ will be used.*

2.4 Convergence and regularity

This section proves convergence for Greek approximations, under certain regularity assumptions. The mean squared error (MSE) of an estimator $\hat{Y}$, with respect to the random variable Y, is defined as $\mathrm{MSE}(\hat{Y}) := \mathbb{E}\big[(\hat{Y} - Y)^2\big]$. As a measure of error, we consider the bias arising from the Itô-Taylor expansion, the Monte Carlo error and the discretisation error, using a partition $\pi := \{0 = t_0 < \ldots < t_n = T\}$, such that $|\pi| := \max_{i=1,\dots,n}(t_i - t_{i-1}) = \mathcal{O}(h)$. These are related by suitable constants $\zeta, \eta > 0$, such that $\vartheta := 1/N^\zeta$ and $h := 1/N^\eta$, where N

is the number of Monte Carlo paths. From Proposition 2.3.1 and Proposition 2.3.2, the bias is based on the order of the expansion. Consider a discretisation scheme and denote by $\hat{X}$ the discretised version of the process X defined in (2.0.1) with an equidistant partition with stepsize $|\pi| = \mathcal{O}(h)$. We shall say that $\hat{X}$ converges strongly with order $k > 0$ at time T if there exist constants $C, h_0 > 0$, such that for all $h \in (0, h_0)$, then $\mathbb{E}\big[|X_T - \hat{X}_T|\big] \le Ch^k$. We shall say that the same approximation converges weakly with order $k > 0$ at time T if for each $g \in \mathcal{C}_p^{2(k+1)}$ there exist constants $C, h_0 > 0$, such that for all $h \in (0, h_0)$, then $|\mathbb{E}[g(X_T)] - \mathbb{E}\big[g(\hat{X}_T)\big] \,| \le Ch^k$. As an example, assume g is a Lipschitz continuous function, and the discretisation scheme has strong rate of convergence k. For $\vartheta \in (0, T]$ and $j \in \{1, \dots, m\}$, the weight W_ϑ/ϑ and the Cauchy-Schwarz inequality yield

$$\begin{aligned}\left|\mathbb{E}\Big[g(X_{t_n})\tfrac{W_\vartheta^{(j)}}{\vartheta}\Big] - \mathbb{E}\Big[g(\hat{X}_{t_n})\tfrac{W_\vartheta^{(j)}}{\vartheta}\Big]\right| &\le \sqrt{\mathbb{E}\left[\big(g(X_{t_n}) - g(\hat{X}_{t_n})\big)^2\right]\mathbb{E}\Big[(W_\vartheta^{(j)}/\vartheta)^2\Big]} \\ &\le C\sqrt{\mathbb{E}\big[(X_{t_n} - \hat{X}_{t_n})^2\big]\,\frac{1}{\vartheta}} \le C\frac{h^k}{\sqrt{\vartheta}}.\end{aligned}$$

This can be extended to polynomials P such that $P(\vartheta) \neq 0$, and multi-indices $\alpha \in \mathcal{M} \setminus \{\emptyset\}$ as

$$\mathbb{E}\left|\big(g(X_{t_n}) - g(\hat{X}_{t_n})\big)\,\frac{I_\vartheta^\alpha}{P(\vartheta)}\right| = \mathcal{O}\left(h^k\sqrt{\mathbb{V}\left[\frac{I_\vartheta^\alpha}{P(\vartheta)}\right]}\right).$$

2.4.1 General convergence result

For $j = 1, \dots, m$, define the approximation for $u_0^{(j)}$ to be

$$\hat{Y}^{(j)} := \frac{1}{N}\sum_{i=1}^{N} g(\hat{X}_{t_n}^i)\frac{I_\vartheta^{(j),i}}{\vartheta}, \tag{2.4.1}$$

where for the i$^{\text{th}}$ Monte Carlo simulation, $\hat{X}_{t_n}^i$ is the approximation of the process X at time T, and $I_\vartheta^{(j),i} := W_\vartheta^{(j),i}$ is the i$^{\text{th}}$ path of the j$^{\text{th}}$ Brownian motion at time ϑ (in total nN Brownian i.i.d. increments).

Theorem 2.4.1. *Assume $(\mathbf{H}u_b^2)$, g is bounded and Lipschitz continuous, and consider a discretisation scheme with a strong convergence rate k. Then, for $\zeta = 1/3$ and $\eta \ge 1/(2k)$, the MSE of the approximation $\hat{Y}^{(j)}$ in (2.4.1) is $\mathcal{O}(N^{-2/3})$, for $j = 1, \dots, m$.*

Proof. Recall the result from Proposition 2.2.1. For $j \in \{1, \dots, m\}$, the bias of the approximation of $u_0^{(j)}$ is

$$\begin{aligned}\left|\mathbb{E}\left[g(\hat{X}_T)\frac{I_\vartheta^{(j)}}{\vartheta}\right] - u_0^{(j)}\right| &= \left|\mathbb{E}\left[g(X_T)\frac{I_\vartheta^{(j)}}{\vartheta}\right] - u_0^{(j)} + \mathbb{E}\left[\left(g(\hat{X}_T) - g(X_T)\right)\frac{I_\vartheta^{(j)}}{\vartheta}\right]\right| \\ &\leq \mathcal{O}(\vartheta) + C\sqrt{\mathbb{E}[(X_T - \hat{X}_T)^2]}\sqrt{\mathbb{E}\left[\left(\frac{I_\vartheta^{(j)}}{\vartheta}\right)^2\right]} = \mathcal{O}(\vartheta) + \mathcal{O}\left(\frac{h^k}{\sqrt{\vartheta}}\right),\end{aligned} \tag{2.4.2}$$

from the Lipschitz continuity of g and the Cauchy-Schwarz inequality. From the boundedness of g, it is clear that $\mathbb{V}\left(g(X_T)\frac{I_\vartheta^{(j)}}{\vartheta}\right) \leq \frac{C_g}{\vartheta}$. This leads to the variance of (2.4.1) being of order $\mathcal{O}(N^{\zeta-1})$. The MSE of approximation (2.4.1) is thus

$$\mathcal{O}(N^{\zeta-1}) + \mathcal{O}(N^{-2\zeta}) + \mathcal{O}(N^{\zeta-2k\eta}) + \mathcal{O}(N^{-\zeta/2-k\eta}),$$

from which it follows that $\zeta = 1/3$, $\eta \geq 1/(2k)$ and that the MSE of (2.4.1) is of order $\mathcal{O}(N^{-2/3})$. This concludes the proof of Theorem 2.4.1. □

Furthermore, the computational cost of the algorithm is $\mathcal{O}(N^{1+1/(2k)})$, therefore a $\log - \log$ plot of the MSE against the computational cost will have a slope of $\frac{-4k}{3(2k+1)}$; as k increases, this quantity approaches $-2/3$.

2.4.2 Romberg Extrapolation

Recall the process X from (2.0.1), with (x, θ) dependence suppressed. To create a scheme with bias of order $\mathcal{O}(h^l)$ for the first-order Greeks, assume the expansion

$$\mathbb{E}\left[g(X_T)\frac{W_h}{h}\right] = \gamma(x)\Delta + d_1 h + d_2 h^2 + \ldots + d_{l-1}h^{l-1} + \mathcal{O}(h^l) \tag{2.4.3}$$

holds for some constants $d_i \in \mathbb{R}$. In the one-dimensional case, approximating the Δ can be achieved by

$$\mathbb{E}\left[g(X_T)\left(\sum_{i=1}^{l} c_i \frac{W_{ih}}{ih}\right)\right] = \gamma(x)\Delta + \mathcal{O}(h^l), \tag{2.4.4}$$

for some Brownian motion W and $h \in (0, T/l]$, where c_i are scheme-specific constants. Define $Z_k := \int_{(k-1)h}^{kh} \mathrm{d}W_s$ for $k = 1, \ldots, l$, and by independence, it follows that

$$\mathbb{V}\left[\sum_{i=1}^{l} c_i \frac{W_{ih}}{ih}\right] = \mathbb{V}\left[\sum_{i=1}^{l} Z_i \left(\sum_{j=i}^{l} \frac{c_j}{jh}\right)\right] = \frac{1}{h} \sum_{i=1}^{l} \left(\sum_{j=i}^{l} \frac{c_j}{j}\right)^2 =: \frac{C_l}{h}. \tag{2.4.5}$$

The variance for bounded payoffs is controlled by

$$\mathbb{V}\left[g(X_T) \sum_{i=1}^{l} c_i \frac{W_{ih}}{ih}\right] \leq \frac{C_l \|g\|_\infty^2}{h}. \tag{2.4.6}$$

We can solve for c_i such that (2.4.4) is satisfied. For fixed $l \in \mathbb{N}^+$, the system has the following structure,

$$\begin{pmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,l} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,l} \\ \vdots & \vdots & \ddots & \vdots \\ e_{l,1} & e_{l,2} & \cdots & e_{l,l} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_l \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{2.4.7}$$

where $e_{i,j} := j^{i-1}$, for all $i, j \in \{1, \ldots, l\}$. The matrix $[e_{i,j}]_{i,j}$ is invertible, so the system admits a unique solution (inverse of a Vandermonde-matrix, see [Pag07, LP14] and references therein). From a numerical point of view, even though the bias is of higher-order, the variance multiple rapidly increases. For $l = 1, \ldots, 6$, the constants in (2.4.5) read

$$C_1 = 1, \quad C_2 = 2.5, \quad C_3 = 4.83, \quad C_4 = 9.25, \quad C_5 = 18.95, \quad C_6 = 42.68.$$

Define for $j = 1, \ldots, m$, the Monte Carlo approximation for the first-order Greek with bias $\mathcal{O}(\vartheta^l)$ as

$$\hat{Y}^{(j),l} := \frac{1}{N} \sum_{i=1}^{N} g(\hat{X}_{t_n}^i) \left(\sum_{q=1}^{l} c_q \frac{W_{q\vartheta}^{(j),i}}{q\vartheta}\right), \tag{2.4.8}$$

where $(c_q)_{q=1,\ldots,l}$ are the solutions from (2.4.7). The extrapolation weights are step functions. The next theorem shows the MSE for this approximation, using a discretisation scheme with convergence of strong order k:

Theorem 2.4.2. *Assume* $(\mathbf{H}u_b^{l+1})$, *g is bounded and Lipschitz continuous, and consider a discretisation scheme with strong convergence rate k. Then, for $\zeta = 1/(2l+1)$ and $\eta \geq 1/(2k)$ the MSE of $\hat{Y}^{(j),l}$ is* *of order* $\mathcal{O}(N^{-\frac{2l}{2l+1}})$.

Proof. By the approach in Theorem 2.4.1, apply an Itô-Taylor expansion to the value function to obtain an expression as in (2.4.3). For $j \in \{1, \dots, m\}$, $\mathbb{E}\left[g(X_T)\left(\sum_{q=1}^{l} c_q \frac{W_{q\vartheta}^{(j),i}}{q\vartheta}\right)\right]$ is an estimator of $u_0^{(j)}$ with bias of order $\mathcal{O}(\vartheta^l)$, where $(c_q)_{q=1,\dots,l}$ are defined as the solution of (2.4.7). Indeed,

$$\mathbb{E}\left[u_\vartheta\left(\sum_{q=1}^{l} c_q \frac{W_{q\vartheta}^{(j),i}}{q\vartheta}\right)\right] = u_0^{(j)} + \mathcal{O}(\vartheta^l). \tag{2.4.9}$$

For all $\alpha \in \mathcal{M} \setminus \{\emptyset\}$ such that $l(\alpha) = l+1$ and $\alpha^+ = (j)$, applying Lemma 2.2.1 yields

$$\left|\mathbb{E}\left[I_\vartheta^\alpha\left[u_\cdot^\alpha\right]\frac{I_\vartheta^{(j),i}}{\vartheta}\right]\right| \leq \sqrt{\mathbb{E}\left[\left(I_\vartheta^\alpha\left[u_\cdot^\alpha\right]\right)^2\right]}\sqrt{\mathbb{E}\left[\left(\frac{I_\vartheta^{(j),i}}{\vartheta}\right)^2\right]} \leq \frac{C}{\sqrt{\vartheta}}\sqrt{\frac{\vartheta^{w(\alpha,\alpha)}}{w(\alpha,\alpha)}\prod_{q=0}^{l(\alpha+)} C_{2k_q(\alpha)}^{k_q(\alpha)}}. \tag{2.4.10}$$

As for the one-dimensional case, the multi-indices $\alpha \in \mathcal{M}$ of interest are those such that $\alpha^+ = (j)$. From Definition (2.2.3), $w(\alpha,\alpha) = 2(l+1)-1 = 2(l+1/2)$, where $l(\alpha) = l+1$. Simplifying (2.4.10) yields

$$\left|\mathbb{E}\left[I_\vartheta^\alpha\left[u_\cdot^\alpha\right]\frac{I_\vartheta^{(j),i}}{\vartheta}\right]\right| \leq \frac{C}{\sqrt{\vartheta}}\sqrt{\vartheta^{2(l+\frac{1}{2})}} = \mathcal{O}(\vartheta^l).$$

As a result, the Romberg extrapolation technique can be performed, which concludes the claim in (2.4.9). Consider now the bias arising from the discretisation scheme with strong rate of convergence k. Similar to (2.4.2), observe that

$$\begin{aligned}
&\left|\mathbb{E}\left[g(\hat{X}_T)\sum_{q=1}^{l} c_q \frac{W_{q\vartheta}^{(j),i}}{q\vartheta}\right] - u_0^{(j)}\right| \\
&= \left|\mathbb{E}\left[g(X_T)\sum_{q=1}^{l} c_q \frac{W_{q\vartheta}^{(j),i}}{q\vartheta}\right] - u_0^{(j)} + \mathbb{E}\left[\left(g(\hat{X}_T) - g(X_T)\right)\sum_{q=1}^{l} c_q \frac{W_{q\vartheta}^{(j),i}}{q\vartheta}\right]\right| \\
&\leq \mathcal{O}(\vartheta^l) + C\sqrt{\mathbb{E}\left[(X_T - \hat{X}_T)^2\right]}\sqrt{\mathbb{E}\left[\left(\sum_{q=1}^{l} c_q \frac{W_{q\vartheta}^{(j),i}}{q\vartheta}\right)^2\right]} = \mathcal{O}(\vartheta^l) + \mathcal{O}\left(\frac{h^k}{\sqrt{\vartheta}}\right).
\end{aligned}$$

Consider the variance of (2.4.8). By independence and using (2.4.6), it follows that

$$\mathbb{V}(\hat{Y}^{(j),l}) = \frac{1}{N}\mathbb{V}\left(g(\hat{X}^i_{t_n})\sum_{q=1}^{l} c_q \frac{W^{(j),i}_{q\vartheta}}{q\vartheta}\right) \leq \frac{C_{g,l}}{N\vartheta},$$

therefore the variance of the approximation is $\mathcal{O}(N^{\zeta-1})$. The MSE is of order equal to $\mathcal{O}(N^{\zeta-1}) + \mathcal{O}(N^{-2l\zeta}) + \mathcal{O}(N^{-(l-1/2)\zeta-k\eta}) + \mathcal{O}(N^{\zeta-2k\eta})$, and from the first and second terms it follows that $-2l\zeta = \zeta - 1$, hence $\zeta = 1/(2l+1)$. In addition, from the first and last terms it follows that $\eta \geq 1/(2k)$, hence the MSE is of order $\mathcal{O}\left(N^{-2l/(2l+1)}\right)$. □

Remark 2.4.1. *By Theorem 2.4.2, the* log − log *plot of the MSE against the computational cost has a slope of* $-\frac{4lk}{(2l+1)(2k+1)}$, *which approaches* -1 *as k or l tend to infinity.*

2.4.3 Cross sensitivities

For $j_1, j_2 \in \{1, \ldots, m\}$, define the approximation for $u_0^{(j_1,j_2)}$ to be

$$\hat{Y}^{(j_1,j_2)} = \frac{1}{N}\sum_{i=1}^{N} g(\hat{X}^i_{t_n})\frac{2I^{(j_1,j_2),i}_{\vartheta}}{\vartheta^2}, \tag{2.4.11}$$

where for the i^th Monte Carlo simulation, $\hat{X}^i_{t_n}$ is the approximation of the process X at time T, and $I^{(j_1,j_2),i}_{\vartheta}$ is the integral as defined in (2.2.1), at time ϑ.

Theorem 2.4.3. *Let* $(\mathbf{H}u_b^3)$, *g bounded and Lipschitz continuous, and consider a discretisation scheme with strong convergence rate k. Then, for* $\zeta = 1/4$ *and* $\eta \geq 1/(2k)$ *the MSE of* $\hat{Y}^{(j_1,j_2)}$ *is* $\mathcal{O}(N^{-1/2})$, *where* $j_1, j_2 \in \{1, \ldots, m\}$.

Proof. Continuing from Proposition 2.3.3, for $j_1, j_2 \in \{1, \ldots, m\}$, the bias of the approximation of $u_0^{(j_1,j_2)}$ is

$$\left|\mathbb{E}\left[g(\hat{X}_T)\frac{2I^{(j_1,j_2)}_{\vartheta}}{\vartheta^2}\right] - u_0^{(j_1,j_2)}\right| = \left|\mathbb{E}\left[g(X_T)\frac{2I^{(j_1,j_2)}_{\vartheta}}{\vartheta^2}\right] - u_0^{(j_1,j_2)} + \mathbb{E}\left[\left(g(\hat{X}_T) - g(X_T)\right)\frac{2I^{(j_1,j_2)}_{\vartheta}}{\vartheta^2}\right]\right|$$

$$\leq \mathcal{O}(\vartheta) + C\sqrt{\mathbb{E}[(X_T - \hat{X}_T)^2]}\sqrt{\mathbb{E}\left[\left(\frac{2I^{(j_1,j_2)}_{\vartheta}}{\vartheta^2}\right)^2\right]} = \mathcal{O}(\vartheta) + \mathcal{O}\left(\frac{h^k}{\vartheta}\right), \tag{2.4.12}$$

from the Lipschitz continuity of g and the Cauchy-Schwarz inequality. Since g is bounded, then $\mathbb{V}\left(g(X_T)\frac{2I_\vartheta^{(j_1,j_2)}}{\vartheta^2}\right) \leq \frac{C_g}{\vartheta^2}$, which leads to the variance of $\hat{Y}^{(j_1,j_2)}$ in (2.4.11) being $\mathcal{O}(N^{2\zeta-1})$. Therefore, the MSE is of order $\mathcal{O}(N^{2\zeta-1}) + \mathcal{O}(N^{-2\zeta}) + \mathcal{O}(N^{2\zeta-2k\eta}) + \mathcal{O}(N^{-k\eta})$, from which it follows that setting $\zeta = 1/4$, $\eta \geq 1/(2k)$ leads to the MSE of (2.4.11) to be $\mathcal{O}(N^{-1/2})$. □

2.4.4 Black-Scholes: Comparison with Malliavin Greeks

Consider the Black-Scholes model with zero drift, under which the asset price process X is the solution to

$$\mathrm{d}X_t = \gamma X_t \mathrm{d}W_t, \qquad X_0 = x > 0, \tag{2.4.13}$$

for some constant volatility parameter $\gamma > 0$ [BS73].

Lemma 2.4.1. *Let* $(\mathbf{H}u_b^3)$, *g bounded and Lipschitz continuous, and consider a discretisation scheme with strong convergence rate k. Then, the weights for the Delta and Gamma of the driftless Black-Scholes model in* (2.4.13), *with corresponding MSE rates in terms of the number of Monte Carlo paths are summarised in Table 2.1.*

Greek	Weight	Value	Bias	ζ	η	MSE
Delta	$\frac{W_\vartheta}{\vartheta x \gamma}$	$\partial_x u_0$	$\mathcal{O}(\vartheta)$	$1/3$	$\geq 1/(2k)$	$\mathcal{O}(N^{-2/3})$
Gamma	$\frac{W_\vartheta^2}{\vartheta^2 x^2 \gamma^2} - \frac{1}{\vartheta x^2 \gamma^2} - \frac{W_\vartheta}{\vartheta x^2 \gamma}$	$\partial_{xx} u_0$	$\mathcal{O}(\vartheta)$	$1/4$	$\geq 1/(2k)$	$\mathcal{O}(N^{-1/2})$

Table 2.1: Black-Scholes Delta and Gamma.

Proof. Follows as a corollary of Theorems 2.4.1 and 2.4.3, and recalling Remark 2.3.1. □

Observe that the weight for computing the Delta is $W_\vartheta/(\vartheta x \gamma)$, whilst the Malliavin weight is $W_T/(Tx\gamma)$ (see [Ben01]). Replacing ϑ by T in the expression for the Gamma in Table 2.1 yields the Malliavin weight (see [Ben01, Chapter 2.3]). In the Black-Scholes model, the Vega ($\mathcal{V}$) and the Gamma (Γ) are related by $\mathcal{V} = T\gamma x^2 \Gamma$. As a result, the Vega can be approximated using the Gamma approximation.

2.4.5 Stochastic Volatility

We consider the couple $\mathfrak{X} = (X, \gamma)$, that is the solution to the following stochastic differential equations

$$\begin{aligned} \mathrm{d}X_t &= \gamma_t X_t \mathrm{d}W_t^{(1)}, & X_0 &= x > 0, \\ \mathrm{d}\gamma_t &= \theta\gamma_t \mathrm{d}W_t^{(2)}, & \gamma_0 &= \gamma > 0, \end{aligned} \tag{2.4.14}$$

where $\theta \in \mathbb{R}^+$ is a fixed constant parameter. We study approximating the Delta (sensitivity with respect to x), the Vega (sensitivity with respect to γ) and the Gamma. This is a specific example of the SABR model, with skewness parameter and correlation set to zero (i.e. $\beta = 1$ and $\rho = 0$ in the notation of [HKLW02]). Since, $\mathbb{E}\Big[g(X_T) I_\vartheta^{(1)}/\vartheta\Big] = u_0^{(1)} + \mathcal{O}(\vartheta) = \gamma x \Delta + \mathcal{O}(\vartheta)$, it follows that $\Delta = \mathbb{E}\Big[g(X_T)\frac{I_\vartheta^{(1)}}{\gamma x \vartheta}\Big] + \mathcal{O}(\vartheta)$. Similarly, the Vega can be approximated as $\mathcal{V} = \mathbb{E}\Big[g(X_T)\frac{I_\vartheta^{(2)}}{\theta\gamma\vartheta}\Big] + \mathcal{O}(\vartheta)$. For the Gamma, consider

$$\mathbb{E}\left[g(X_T)\frac{2I_\vartheta^{(1,1)}}{\vartheta^2}\right] = u_0^{(1,1)} + \mathcal{O}(\vartheta) = \gamma^2 x(\Delta + x\Gamma) + \mathcal{O}(\vartheta),$$

so that the Gamma can be expressed by $\Gamma = \mathbb{E}\Big[g(X_T)\frac{2I_\vartheta^{(1,1)}}{\gamma^2 x^2 \vartheta^2}\Big] - \frac{\Delta}{x} + \mathcal{O}(\vartheta)$.

Corollary 2.4.1. *Let* $(\mathbf{H}u_b^3)$*, g bounded and Lipschitz continuous, and consider a discretisation scheme with a strong convergence rate k. Then, the weights for the Delta, Vega, Gamma, Vanna and Vomma are presented in Table 2.2.*

Greek	Weight	Value	Bias	ζ	η	MSE
Delta	$\frac{W_\vartheta^{(1)}}{\vartheta x \gamma}$	$\partial_x u_0$	$\mathcal{O}(\vartheta)$	1/3	$\geq 1/(2k)$	$\mathcal{O}(N^{-2/3})$
Vega	$\frac{W_\vartheta^{(2)}}{\vartheta \gamma \theta}$	$\partial_\gamma u_0$	$\mathcal{O}(\vartheta)$	1/3	$\geq 1/(2k)$	$\mathcal{O}(N^{-2/3})$
Gamma	$\frac{(W_\vartheta^{(1)})^2}{\vartheta^2 x^2 \gamma^2} - \frac{1}{\vartheta x^2 \gamma^2} - \frac{W_\vartheta^{(1)}}{\vartheta x^2 \gamma}$	$\partial_{xx} u_0$	$\mathcal{O}(\vartheta)$	1/4	$\geq 1/(2k)$	$\mathcal{O}(N^{-1/2})$
Vanna	$\frac{W_\vartheta^{(1)} W_\vartheta^{(2)}}{\vartheta^2 x \gamma^2 \theta} - \frac{W_\vartheta^{(1)}}{2\vartheta x \gamma^2}$	$\partial_{x\gamma} u_0$	$\mathcal{O}(\vartheta)$	1/4	$\geq 1/(2k)$	$\mathcal{O}(N^{-1/2})$
Vomma	$\frac{(W_\vartheta^{(2)})^2}{\vartheta^2 \gamma^2 \theta^2} - \frac{1}{\vartheta \gamma^2 \theta^2} - \frac{W_\vartheta^{(2)}}{\vartheta \gamma^2 \theta}$	$\partial_{\gamma\gamma} u_0$	$\mathcal{O}(\vartheta)$	1/4	$\geq 1/(2k)$	$\mathcal{O}(N^{-1/2})$

Table 2.2: First and second-order Greeks for SV model (2.4.14).

3. Numerical Approximation of the Delta

The focus in this chapter is the approximation of the first-order sensitivity of the value function u, solution to the Cauchy problem in (2.1.1), with respect to the space variables. We demonstrate two approaches to approximate the Delta (Δ) of an option with high-order of convergence.

The first technique is inspired from the BSDE literature on numerical methods and consists of multiplying the option payoff by weights based on the driving Brownian motion. We describe ψ-functions, that characterise such weights, and discuss the variance properties of the Delta approximations, using weights characterised by polynomials $\psi_{p,l}$ and step functions $\psi_{s,l}$, within the ψ-family of functions. We state approximation results when using a discrete-time approximation for process X. By studying such fully implementable algorithms, we obtain order 1 approximations for the Delta, improving the rate $1/2$ proved in the backward stochastic differential equation (BSDE) literature [Cha14].

The second approach follows from the ideas of the seminal work by [TT90] and builds on [Cha14]. The aim is to justify an expansion of the Delta allowing intuitive extrapolation techniques to be applied in order to obtain higher-order approximations. We improve the results from the previous chapter on convergence of the Greek approximations, by considering the weak order of convergence upon discretisation using weak Taylor schemes.

3.1 A class of Δ weights

We review a class of functions which are used to define weights to approximate the Δ of a contingent claim and they are inspired from the work done on the numerical approximations of BSDEs.

Definition 3.1.1 (ψ-functions [CC14, Definition 1.5 (i)]). *For $l \in \mathbb{N}$, define $\mathcal{B}^l_{[0,1]}$ as the set of bounded, measurable functions $\psi : [0,1] \to \mathbb{R}$ such that*

$$\int_0^1 \psi(s)\mathrm{d}s = 1, \qquad \textit{and if } l \in \mathbb{N}^+,\ \int_0^1 \psi(s)s^k\mathrm{d}s = 0 \qquad \textit{for all } 1 \leq k \leq l.$$

The solution (X, Y, Z) of (2.0.1) and (2.1.4) is a special case of a BSDE with a zero-driver. In a financial context, the process Z is related to the first-order sensitivity, and $Z_0 = \gamma(x)\Delta$ at the initial time. We define the weight $H^{\psi}_{\cdot}$, which is used to approximate the Z process in (2.1.4):

Definition 3.1.2 (H^{ψ}_h-functionals [CC14, Definition 1.5 (ii)]). *Let $\psi \in \mathcal{B}^l_{[0,1]}$, and for $0 < h \leq T$, define the row vector $H^{\psi}_{t,h}$ with entries $j = 1, \ldots, m$ by*

$$(H^{\psi}_{t,h})_j := \frac{1}{h} \int_{s=t}^{t+h} \psi^j \left(\frac{s-t}{h} \right) \mathrm{d}W^{(j)}_s,$$

and for shorthand $H^{\psi}_h := H^{\psi}_{0,h}$.

Recall the smoothness assumptions of the value function from Definition 2.2.3, $(\mathbf{H}u^l_b)$, and the operators defined in (2.1.2)-(2.1.3). The value function u throughout this chapter will be the solution to (2.1.1).

Proposition 3.1.1 ([CC14, Proposition 2.3]). *Fix $l \in \mathbb{N}$. Let $(\mathbf{H}u^{l+2}_b)$ hold, $\psi \in \mathcal{B}^l_{[0,1]}$. Then, for $h \in (0, T]$,*

$$\mathbb{E}\Big[(H^{\psi}_h)_j g(X_T)\Big] = u^{(j)}_0 + \mathcal{O}(h^{l+1}), \qquad j = 1, \ldots, m. \tag{3.1.1}$$

Proof. By an application of the conditional expectation, $\mathbb{E}\Big[(H^{\psi}_h)_j g(X_T)\Big] = \mathbb{E}\Big[(H^{\psi}_h)_j u(h, X_h)\Big]$. Consider the class of theoretical coefficients $H^{\psi}_{\cdot}$; from [CC14, Proposition 2.3 (i)], we can expand any sufficiently smooth function using a weak Taylor expansion. For $u \in \mathcal{G}^{l+2}_b$, $\psi \in \mathcal{B}^l_{[0,1]}$, for all $1 \leq j \leq m$, we then obtain

$$\mathbb{E}\Big[(H^{\psi}_{t,h})_j u(t+h, X^{t,x}_{t+h})\Big] = u^{(j)}(t,x) + h u^{(j,0)}(t,x) + \cdots + \frac{h^l}{l!} u^{(j)*(0)_l}(t,x) + \mathcal{O}(h^{l+1}), \tag{3.1.2}$$

where $(X^{t,x}_s)$ is the process at time $s \geq t$, with initial conditions $(t,x) \in [0,T] \times \mathbb{R}^d$, i.e. $X^{t,x}_t = x$. The result immediately follows by considering the initial time $t = 0$, since all terms $u^{(j)*(0)_l}_0$ for $l \in \mathbb{N}^+$ are equal to zero from (2.1.1). $\square$

Remark 3.1.1. *The expansion* (3.1.1) *from Proposition 3.1.1 holds true with $\psi \equiv 1$ in the special case where $L^{(1)} \circ L^{(0)} = L^{(0)} \circ L^{(1)}$ for partial differential equations such that $u^{(0)}_0 = 0$.*

Corollary 3.1.1. *For all $j = 1, \ldots, m$, the following statements hold:*

1. *Let $(\mathbf{H}u^2_b)$ and $\psi_0 \equiv 1$, belonging to $\mathcal{B}^0_{[0,1]}$. Then, $\mathbb{E}\Big[(H^{\psi_0}_h)_j u_h\Big] = u^{(j)}_0 + \mathcal{O}(h)$.*
2. (a) *$(\mathbf{H}u^3_b)$, $\psi_{p,1}(u) \equiv 4 - 6u \in \mathcal{B}^1_{[0,1]}$ implies that $\mathbb{E}\Big[(H^{\psi_{p,1}}_h)_j u_h\Big] = u^{(j)}_0 + \mathcal{O}(h^2)$.*

(b) $(\mathbf{H}u_b^3)$, *for* $c \in (0,1)$ *the function* $\psi_{s,1}(u) \equiv \frac{1}{c(c-1)}1_{[1-c,1]}(u) + \frac{c-2}{c-1}$ *is also in* $\mathcal{B}^1_{[0,1]}$. *Then,* $\mathbb{E}\Big[(H_h^{\psi_{s,1}})_j u_h\Big] = u_0^{(j)} + \mathcal{O}(h^2)$.

3. *Let* $(\mathbf{H}u_b^4)$ *and fix distinct* $c, c' \in (0,1)$. *Define*

$$\psi_{s,2}(u) \equiv \frac{1-c'}{c(1-c)(c'-c)}1_{[1-c,1]}(u) + \frac{c-1}{c'(1-c')(c'-c)}1_{[1-c',1]}(u) + \left(1 + \frac{1}{1-c} + \frac{1}{1-c'}\right),$$

which belongs to $\mathcal{B}^2_{[0,1]}$. *Then,* $\mathbb{E}\Big[(H_h^{\psi_{s,2}})_j u_h\Big] = u_0^{(j)} + \mathcal{O}(h^3)$.

Proof. This corollary is a by-product of Proposition 3.1.1, [CC14, Example 2.1] and [CC14, Proposition 2.4]. □

3.1.1 Variance properties

The previous corollary explicited several $\psi_{\cdot,l}$ functions and the corresponding bias of the approximation of $u_0^{(j)}$. For $\psi \equiv 1$, by direct calculation $\mathbb{V}[H_h^{\psi}] = 1/h$; we study the variance of higher-order weights using functions belonging to $\mathcal{B}^1_{[0,1]}$ and $\mathcal{B}^2_{[0,1]}$. The variance of these weights, coupled with the associated bias allows the MSE of the Greek approximations to be studied.

Example 3.1.1 (Step function $\psi_{s,1} \in \mathcal{B}^1_{[0,1]}$). *In order to simulate the weight* $H_h^{\psi_{s,1}}$ *using the step function* $\psi_{s,1}$, *we fix* $c \in (0,1)$. *From the definition of* $H_h^{\psi_{s,1}}$, *it follows that*

$$\begin{aligned} H_h^{\psi_{s,1}} &= \tfrac{1}{h}\int_0^h \psi_{s,1}(s/h)\mathrm{d}W_s = \tfrac{1}{h}\int_0^h \left(\tfrac{1}{c(c-1)}1_{[1-c,1]}(s/h) + \tfrac{c-2}{c-1}1_{[0,1]}(s/h)\right)\mathrm{d}W_s \\ &= \tfrac{1}{hc(c-1)}(W_h - W_{h(1-c)}) + \tfrac{c-2}{h(c-1)}W_h = \tfrac{c-1}{c}\tfrac{W_h}{h} + \tfrac{1}{c}\tfrac{W_{h(1-c)}}{h(1-c)}. \end{aligned}$$

This weight is simulated using the Brownian motion at times h *and* $h(1-c)$. *The variance of* $H_h^{\psi_{s,1}}$, *given a fixed* c, *is*

$$\mathbb{V}(H_h^{\psi_{s,1}}) = \frac{(c-1)^2 + 2(c-1) + 1/(1-c)}{c^2 h} = \frac{c^2 - c - 1}{ch(c-1)}, \tag{3.1.3}$$

and the minimum variance of $5/h$ *is attained independently of* h *with* $c = 1/2$.

Example 3.1.2 (Polynomial $\psi_{p,1} \in \mathcal{B}^1_{[0,1]}$). *We now consider* $\psi_{p,1}(u) \equiv 4 - 6u$ *and the variance of* *the weight* $H_h^{\psi_{p,1}}$. *From Definition 3.1.1, it follows that*

$$H_h^{\psi_{p,1}} = \frac{1}{h}\int_0^h \psi_{p,1}(s/h)\mathrm{d}W_s = \frac{4}{h}W_h - \frac{6}{h^2}\int_0^h s\mathrm{d}W_s.$$

To compute this weight, we sample from the vector

$$\begin{pmatrix} W_h \\ \int_0^h s\mathrm{d}W_s \end{pmatrix} \sim N(0,\Sigma), \qquad \text{where } \Sigma := \begin{pmatrix} h & h^2/2 \\ h^2/2 & h^3/3 \end{pmatrix}.$$

By performing a Cholesky decomposition, $\Sigma = LL^*$ *where*

$$L = \begin{pmatrix} \sqrt{h} & 0 \\ \frac{1}{2}h^{3/2} & \frac{1}{2\sqrt{3}}h^{3/2} \end{pmatrix},$$

and using independent $Z_1, Z_2 \sim N(0,1)$, *the vector can be sampled by setting*

$$\begin{pmatrix} W_h \\ \int_0^h s\mathrm{d}W_s \end{pmatrix} = \begin{pmatrix} \sqrt{h} & 0 \\ \frac{1}{2}h^{3/2} & \pm\frac{1}{2\sqrt{3}}h^{3/2} \end{pmatrix}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}. \tag{3.1.4}$$

The variance of $H_h^{\psi_{p,1}}$ *is easily computed using Itô's isometry and evaluates to*

$$\mathbb{V}(H_h^{\psi_{p,1}}) = \frac{16}{h^2}\int_0^h \mathrm{d}s - \frac{48}{h^3}\int_0^h s\mathrm{d}s + \frac{36}{h^4}\int_0^h s^2\mathrm{d}s = \frac{4}{h},$$

noting that it is lower than that of the weight defined using $\psi_{s,1}$ *in Example 3.1.1.*

Example 3.1.3 (Step function $\psi_{s,2} \in \mathcal{B}^2_{[0,1]}$). *For distinct* $(c,c') \in (0,1)^2$ *fixed, the weight* $H_h^{\psi_{s,2}}$ *is given by*

$$H_h^{\psi_{s,2}} = \frac{\left(W_h - W_{(1-c)h}\right)(1-c')}{hc(1-c)(c'-c)} + \frac{\left(W_h - W_{(1-c')h}\right)(c-1)}{hc'(1-c')(c'-c)} + \frac{W_h}{h}\left(1 + \frac{1}{1-c} + \frac{1}{1-c'}\right).$$

The minimal variance $\mathbb{V}[H_h^{\psi_{s,2}}] = 11.1/h$ *is achieved at* $(c,c') = (0.775, 0.126)$, *independently of* h.

Example 3.1.4 (Polynomial $\psi_{p,2} \in \mathcal{B}^2_{[0,1]}$). *The unique quadratic belonging to* $\mathcal{B}^2_{[0,1]}$ *is* $\psi_{p,2}(u) \equiv 9 - 36u + 30u^2$, *with a corresponding weight of*

$$H_h^{\psi_{p,2}} = \frac{9}{h}\int_0^h \mathrm{d}W_s - \frac{36}{h^2}\int_0^h s\mathrm{d}W_s + \frac{30}{h^3}\int_0^h s^2\mathrm{d}W_s.$$

The vector has the following distribution:

$$\begin{pmatrix} W_h \\ \int_0^h s\mathrm{d}W_s \\ \int_0^h s^2\mathrm{d}W_s \end{pmatrix} \sim N(0,\Sigma), \qquad \textit{where } \Sigma := \begin{pmatrix} h & h^2/2 & h^3/3 \\ h^2/2 & h^3/3 & h^4/4 \\ h^3/3 & h^4/4 & h^5/5 \end{pmatrix},$$

and by Cholesky's decomposition, $\Sigma = LL^*$ *where*

$$L = \begin{pmatrix} \sqrt{h} & 0 & 0 \\ \frac{1}{2}h^{3/2} & \frac{1}{2\sqrt{3}}h^{3/2} & 0 \\ \frac{1}{3}h^{5/2} & \frac{1}{2\sqrt{3}}h^{5/2} & \frac{1}{6\sqrt{5}}h^{5/2} \end{pmatrix}.$$

Using independent $Z_1, Z_2, Z_3 \sim N(0,1)$, *the desired vector can be sampled by* LZ *where* Z *is the column vector consisting of* Z_1, Z_2, Z_3. *The variance of the weight is*

$$\mathbb{V}[H_h^{\psi_{p,2}}] = \frac{81}{h} + \frac{1296}{3h} + \frac{900}{5h} - \frac{648}{2h} + \frac{540}{3h} - \frac{2160}{4h} = \frac{9}{h},$$

which is slightly less than the variance of $\psi_{s,2}$ *in Example 3.1.3.*

3.1.2 Optimal function ψ

We consider the MSE of the approximations using $h := 1/N^{\zeta}$, where N is the number of Monte Carlo realisations. Assume a setting where we can perfectly simulate the process without any discretisation error. Consider the MSE bounds denoted by $\mathfrak{M}_l$ for some fixed l, of the Greek approximations using $\psi \in \mathcal{B}^l_{[0,1]}$, given the optimal values of ζ which yields the same order of convergence for the bias and variance components of the mean squared error. The MSE of an approximation using $\psi \in \mathcal{B}^0_{[0,1]}$ (i.e. $\psi \equiv 1$) for optimal $\zeta = 1/3$ can be expressed as

$$\mathfrak{M}_0 := \frac{C_1^2}{N^{2/3}} + \frac{C_4}{N^{1-1/3}} = \frac{C_1^2 + C_4}{N^{2/3}},$$

where C_1 is the bias constant in the expansion using $\psi \equiv 1$, and C_4 is a bound on the variance term. For weights defined by functions $\psi \in \mathcal{B}^1_{[0,1]}$ and using the optimal $\zeta = 1/5$, the MSE can be expressed as

$$\mathfrak{M}_1 := \frac{C_2^2}{N^{4/5}} + \frac{C_3C_4}{N^{1-1/5}} = \frac{C_2^2 + C_3C_4}{N^{4/5}},$$

where C_2 is the bias constant in the expansion using $\psi \in \mathcal{B}^1_{[0,1]}$, $C_3 = 5$ for the step function $\psi_{s,1}$ and $C_3 = 4$ for $\psi_{p,1} \in \mathcal{B}^1_{[0,1]}$ (recalling the computations in Examples 3.1.1-3.1.2). The constants C_1 and C_2 above are bounds depending on the higher-order sensitivities of the value function. Therefore, by comparing the mean squared errors bounds, $\mathfrak{M}_1 \leq \mathfrak{M}_0$ holds only when

$$N \geq \left(\frac{C_2^2 + C_3 C_4}{C_1^2 + C_4} \right)^{15/2}.$$

From this, we observe that depending on the above constants, the critical value of N for which a higher-order scheme produces a smaller MSE can be quite large. For example, when $C_1 = C_2 = C_4 = 1$, a lower MSE occurs for $\psi_{p,1}$ when $N \approx 10^3$. A slight change of the constants to $C_1 = C_4 = 1$ and $C_2 = 6$, implies $\mathfrak{M}_1 \leq \mathfrak{M}_0$ only when $N \approx 10^{10}$. This shows how the optimal choice of scheme is dependent on the constants arising from the expansions and the MSE computation.

Remark 3.1.2.

(i) *The MSE considered for comparison purposes above is the upper bound, as opposed to the actual value.*

(ii) *This section highlights a practical consideration which can be observed when performing the numerical simulations: it is imperative to consider the variance increase upon the selection of higher-order weights.*

3.2 Weak Taylor schemes

We now combine the above with higher-order approximations of the process X, recalling the iterated Itô integrals from Definition 2.2.1:

Definition 3.2.1 (Weak Taylor scheme of order r [KP92, (14.5.4)]). *Consider a discretised process $\hat{X} = (\hat{X}_t)_{t\in[0,T]}$ using a weak Taylor scheme of order r of the process X in* (0.1.3). *For a grid $\pi := \{0 := t_0 < t_1 < \ldots < t_n := T\}$, we define $\hat{X}$ using the hierarchical set $\mathcal{D}_r$ (recall definition on page 77) for $t \in [t_i, t_{i+1}]$ as*

$$\hat{X}_t := \hat{X}_{t_i} + \sum_{\alpha \in \mathcal{D}_r \setminus \{\emptyset\}} \mathfrak{f}^\alpha(t_i, \hat{X}_{t_i}) I^\alpha_{t_i,t}, \qquad \hat{X}_{t_0} := X_{t_0},$$

$\mathfrak{f}(t,x) \equiv x$, with $\mathfrak{f}^{(0)} = f$, $\mathfrak{f}^{(1)} = \gamma$, and for $l \geq 1$ such that $\alpha = (j_1, \ldots, j_l)$, then $\mathfrak{f}^{\alpha} = L^{(j_1)}\mathfrak{f}^{-\alpha}$ (for the $d = m = 1$ case).

Remark 3.2.1.

(i) We have extended the usual definition of the weak Taylor scheme for all $t \in [0, T]$ as opposed to just defining the discretisation at the grid points of π.

(ii) The (continuous) Euler scheme on $[0, T]$ is the weak Taylor scheme of order $r = 1$. Define $(\hat{X})_{t\in[0,T]}$ on an equidistant grid π such that $|\pi| = T/n$, for $t \in [t_i, t_{i+1}]$ by

$$\hat{X}_t := \hat{X}_{t_i} + f(\hat{X}_{t_i})(t - t_i) + \gamma(\hat{X}_{t_i})(W_t - W_{t_i}), \qquad \hat{X}_0 := X_0.$$

(iii) We sometimes highlight the number of time steps n in π by referring to the approximation as $\hat{X}^n := (\hat{X}^n_t)_{t\in[0,T]}$.

3.2.1 Approximation using Euler scheme

Let us observe that on $[0, h]$, the Euler scheme is a Brownian motion with constant drift $f(y)$ and volatility $\gamma(y)$, if the process X start at y at $t = 0$. We denote by $\hat{L}^{(j)}_y$, $j = 0, \ldots, m$ the operators associated to this process:

Definition 3.2.2. *Define the fixed space operators $\hat{L}^{(j)}_y$ for some $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ acting on $\mathcal{C}^{1,2}(\mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R})$ functions φ by:*

$$\{\hat{L}^{(j)}_y \varphi\}(t,x) := \sum_{k=1}^{d} \gamma_{k,j}(y)\partial_{x_k}\varphi(t,x) \qquad \text{for } j = 1, \ldots, m, \tag{3.2.1}$$

$$\{\hat{L}^{(0)}_y \varphi\}(t,x) := \left(\partial_t + \sum_{k=1}^{d} f_k(y)\partial_{x_k} + \frac{1}{2}\sum_{j=1}^{m} \hat{L}^{(j)}_y \circ \hat{L}^{(j)}_y\right)\varphi(t,x). \tag{3.2.2}$$

Remark 3.2.2. *Consider the Euler scheme and fix $y = \hat{X}_{t_i}$: then $\hat{L}^{(0)}_y$ is the operator associated to the diffusion process $(\hat{X}_t)_{t\in[t_i,t_{i+1}]}$. Recall the operators defined in (2.1.1); note that $L^{(0)}\varphi(t, X_t) = \hat{L}^{(0)}_{X_t}\varphi(t, X_t)$ and $L^{(1)}\varphi(t, X_t) = \hat{L}^{(1)}_{X_t}\varphi(t, X_t)$ for this example.*

Example 3.2.1. *In the one-dimensional case $(d = m = 1)$, we consider several examples, to distinguish between $L^{\alpha}u(t,x)$ and $\hat{L}^{\alpha}_x u(t,x)$ (supposing that $f \equiv 0$).*

(i) Observe that $L^{(1,1)}u(t,x) = \gamma^2(x)\partial_{xx}u(t,x) + \gamma(x)\gamma'(x)\partial_x u(t,x)$*. Now,* $\hat{L}_y^{(1)} \circ (\gamma(y)\partial_x)u(t,x) = \gamma^2(y)\partial_{xx}u(t,x)$*, therefore substituting* $y = x$ *yields* $\hat{L}_x^{(1,1)}u(t,x) = \gamma^2(x)\partial_{xx}u(t,x)$*. Combining the two expressions yields*

$$L^{(1,1)}u(t,x) = \hat{L}_x^{(1,1)}u(t,x) + \gamma(x)\gamma'(x)\partial_x u(t,x). \tag{3.2.3}$$

(ii) $L^{(0,0)}u(t,x)$ *can be similarly expanded. Consider* $\hat{L}_y^{(0,0)}u(t,x) = \partial_{tt}u(t,x) + \gamma^2(y)\partial_{txx}u(t,x) + \frac{1}{4}\gamma^4(y)\partial_{xxxx}u(t,x)$*, and setting* $y = x$ *yields*

$$L^{(0,0)}u(t,x) = \hat{L}_x^{(0,0)}u(t,x) + \frac{1}{2}\gamma^2(x)\partial_x\left(\gamma(x)\gamma'(x)\right)\partial_{xx}u(t,x) + \gamma^3\gamma'(x)\partial_{xxx}u(t,x). \tag{3.2.4}$$

Recall the Euler scheme from Remark 3.2.1(ii): define $(\hat{X}_u^{s,y})_{u\geq t}$ for $(s,y) \in [0,T) \times \mathbb{R}^d$ as the process such that $\hat{X}_t^{s,y} = y + \int_s^t f(\hat{X}_r^{s,y})\mathrm{d}r + \int_s^t \gamma(\hat{X}_r^{s,y})\mathrm{d}W_r$, and $\hat{X}_s^{s,y} = y$. For the Euler scheme, we are able to write the operators $\hat{L}_y^\alpha$ and for multi-indices such that $l(\alpha) \leq 1$, we have that $\hat{L}_y^\alpha u(s,y) = L^\alpha u(s,y)$.

We now state the following result using the function $\psi \equiv 1$, for the expansion of any smooth value function using (3.1.2):

Lemma 3.2.1. *Fix* $l \in \mathbb{N}$ *and consider an Euler scheme. For any* $v \in \mathcal{G}_b^{l+2}$ *and* $x \in \mathbb{R}^d$*, then*

$$\begin{aligned}\mathbb{E}\Big[v(h,\hat{X}_h^{0,x})\Big] &= v(0,x) + \hat{L}_x^{(0)}v(0,x)h + \cdots + \hat{L}_x^{(0)_{l+1}}v(0,x)\frac{h^{l+1}}{(l+1)!} + \mathcal{O}(h^{l+2}),\\ \mathbb{E}\Big[(H_h^1)_j v(h,\hat{X}_h^{0,x})\Big] &= \hat{L}_x^{(j)}v(0,x) + \hat{L}_x^{(j,0)}v(0,x)h + \cdots + \hat{L}_x^{(j)*(0)_l}v(0,x)\frac{h^l}{l!} + \mathcal{O}(h^{l+1}),\end{aligned}$$

with $(H_h^1)_j = W_h^{(j)}/h$*, for all* $j = 1,\ldots,m$*.*

Proof. The first part is [CC14, Proposition 2.2]. The second part simply follows since $\hat{L}_x^{(0)} \circ \hat{L}_x^{(1)} = \hat{L}_x^{(1)} \circ \hat{L}_x^{(0)}$, so we quote [CC14, Proposition 2.3 (iii)]. □

Remark 3.2.3.

(i) Recall the multi-variate version of Taylor's theorem. For a multi-index $\alpha = (\alpha_1,\ldots,\alpha_n)$*, and* $x \in \mathbb{R}^n$*, define* $x^\alpha := \prod_{i=1}^n x_i^{\alpha_i}$*. Furthermore, define* $\alpha! := \prod_{i=1}^n(\alpha_i!)$ *and*

$$\partial_\alpha f := \frac{\partial^{l(\alpha)}f}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}}.$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is k times differentiable at point $b \in \mathbb{R}^n$, then there exists some remainder R, such that

$$f(x) = \sum_{l(\alpha)\leq k} \frac{\partial_\alpha f(b)}{\alpha!}(x-b)^\alpha + R,$$

with R approaching zero as x approaches b.

(ii) *For the value function $u : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$, $l(\alpha) = d+1$. We shall restrict the proofs to $d = 1$, i.e. two-dimensional Taylor's theorem shall be used with respect to time and space.*

(iii) *This multi-index α should not be confused with the multi-indices used for the Itô-Taylor expansions.*

The next result is a fully implementable technique for computing the Delta of an option, and is the primary contribution of this chapter. We consider the Euler scheme discretisation, set $\psi \equiv 1 \in \mathcal{B}^0_{[0,1]}$ and state the main result for approximating the Δ using an Euler scheme:

Theorem 3.2.1. *Suppose that $(\mathbf{H}u^3_b)$ holds for a value function u, $\psi \in \mathcal{B}^0_{[0,1]}$, and suppose an Euler scheme on an equidistant mesh π, such that $|\pi| = h$. Then,*

$$\mathbb{E}\Big[(H^\psi_h)_j g(\hat{X}_T)\Big] = L^{(j)}u(0,x) + \mathcal{O}(h).$$

Proof. We begin by fixing an equidistant time grid π with n points of size h.

i) By a telescoping sum it follows that

$$\mathbb{E}\Big[(H^\psi_h)_j g(\hat{X}_T)\Big] = \mathbb{E}\left[(H^\psi_h)_j \sum_{i=1}^{n-1}\Big\{u(t_{i+1},\hat{X}_{t_{i+1}}) - u(t_i,\hat{X}_{t_i})\Big\}\right] + \mathbb{E}\Big[(H^\psi_h)_j u(t_1,\hat{X}_{t_1})\Big], \quad (3.2.5)$$

and from Lemma 3.2.3 we note that $\mathbb{E}\Big[(H^\psi_h)_j u(h,\hat{X}_h)\Big] = L^{(j)}u(0,x) + \mathcal{O}(h)$, where $h = t_1$.

ii) It is left to deal with the telescoping series; consider

$$\begin{aligned} u(t_{i+1},\hat{X}_{t_{i+1}}) - u(t_i,\hat{X}_{t_i}) &= \int_{t_i}^{t_{i+1}} \hat{L}^{(0)}_{\hat{X}_{t_i}} u(s,\hat{X}_s)\mathrm{d}s + \sum_{j=1}^m \int_{t_i}^{t_{i+1}} \hat{L}^{(j)}_{\hat{X}_{t_i}} u(s,\hat{X}_s)\mathrm{d}W^{(j)}_s \\ &= h\hat{L}^{(0)}_{\hat{X}_{t_i}} u(t_i,\hat{X}_{t_i}) + \int_{t_i}^{t_{i+1}}\int_{t_i}^{s} \hat{L}^{(0,0)}_{\hat{X}_{t_i}} u(r,\hat{X}_r)\mathrm{d}r\mathrm{d}s + R_2 + R_1, \end{aligned} \quad (3.2.6)$$

where $R_1 := \sum_{j=1}^m \int_{t_i}^{t_{i+1}} \hat{L}^{(j)}_{\hat{X}_{t_i}} u(s,\hat{X}_s)\mathrm{d}W^{(j)}_s$ and $R_2 := \sum_{j=1}^m \int_{t_i}^{t_{i+1}}\int_{t_i}^{s} \hat{L}^{(j,0)}_{\hat{X}_{t_i}} u(r,\hat{X}_r)\mathrm{d}W^{(j)}_r\mathrm{d}s$. The term $\hat{L}^{(0)}_{\hat{X}_{t_i}} u(t_i,\hat{X}_{t_i})$ is zero directly from the partial differential equation, since $\hat{L}^{(0)}_{\hat{X}_{t_i}} u(t_i,\hat{X}_{t_i}) =$

$L^{(0)}u(t_i, \hat{X}_{t_i}) = 0$. We now consider the second term, which can be rewritten as

$$\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \hat{L}^{(0,0)}_{\hat{X}_{t_i}} u(r, \hat{X}_r) \mathrm{d}r\mathrm{d}s = \frac{h^2}{2}\varphi(t_i, \hat{X}_{t_i}) + R_3,$$

where $\varphi(t, x_t) := \hat{L}^{(0,0)}_{x_t} u(t, x_t)$ and $R_3 := \int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \{\varphi(r, \hat{X}_r) - \varphi(t_i, \hat{X}_{t_i})\}\mathrm{d}r\mathrm{d}s$. We now combine

$$\mathbb{E}\Big[(H^{\psi}_h)_j\{u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i})\}\Big] = \mathbb{E}\Big[(H^{\psi}_h)_j\{\tfrac{h^2}{2}\varphi(t_i, \hat{X}_{t_i}) + R_3 + R_2 + R_1\}\Big]. \qquad (3.2.7)$$

It is apparent that for all $i = 1, \ldots, n-1$ it holds that $\mathbb{E}\Big[(H^{\psi}_h)_j \mathbb{E}_{t_i}[R_1 + R_2]\Big] = 0$, by taking a conditional expectation and noting that the Brownian increments are independent. An application of the Cauchy-Schwarz inequality, yields $\mathbb{E}\Big[(H^{\psi}_h)_j R_3\Big] \leq \|(H^{\psi}_h)_j\|_2 \|R_3\|_2 \leq Ch^2$ for $\psi \equiv 1$ since $\psi \in \mathcal{B}^0_{[0,1]}$, and φ is sufficiently smooth. Furthermore, $\mathbb{E}\Big[(H^{\psi}_h)_j \varphi(t_i, \hat{X}_{t_i})\Big] = \mathbb{E}\Big[(H^{\psi}_h)_j \tilde{\varphi}_i(h, \hat{X}_h)\Big]$ by the Markov property of $(\hat{X}_{t_i})_{i=1,\ldots,n}$, where $\tilde{\varphi}_i(h, x) = \mathbb{E}\big[\varphi(t_i, \hat{X}_{t_i}) | \hat{X}_{t_i} = x\big]$. From this, we obtain $\mathbb{E}\Big[(H^{\psi}_h)_j \varphi(t_i, \hat{X}_{t_i})\Big] = \hat{L}^{(j)}_x \varphi_i(0, x) + \mathcal{O}(h)$, since we can perfectly simulate $(H^{\psi}_h)_j$ and

$$\mathbb{E}_{t_i}\Big[(H^{\psi}_{t_i,h})_j v(t_{i+1}, \hat{X}_{i+1})\Big] = v^{(j)}(t_i, \hat{X}_i) + \mathcal{O}(h), \qquad \text{for } v \in \mathcal{G}^1_b.$$

Therefore, $\frac{h^2}{2}\sum_{i=1}^{n} \mathbb{E}\Big[(H^{\psi}_h)_j \varphi(t_i, \hat{X}_{t_i})\Big] = \mathcal{O}(h)$. To conclude, summation over $i = 1, \ldots, n-1$ for (3.2.7) yields $\mathbb{E}\Big[(H^{\psi}_h)_j g(\hat{X}_T)\Big] = u^{(j)}_0 + \mathcal{O}(h)$.

□

In the next section, we study the case for higher order schemes, using weights defined by functions $\psi \in \mathcal{B}^l_{[0,1]}$, for $l \geq 1$.

3.2.2 Approximation using higher-order weak Taylor scheme

We now consider a weak Taylor scheme of order r, and introduce the operators $\hat{L}^{\alpha,r}_y$ where $l(\alpha) \leq r$, and argue that these operators are such that for all $\alpha \in \mathcal{D}_r$, $\hat{L}^{\alpha,r}_y u(s, y) = L^{\alpha} u(s, y)$; we do not attempt to explicit these operators for weak Taylor schemes of higher orders.

Definition 3.2.3. *Consider a weak Taylor scheme of order $r \geq 2$, a multi-index α such that $\alpha = (0)_l$ for $l \geq 1$, and a smooth function u. Itô-Taylor expanding $u(h, \hat{X}^{0,x}_h)$ yields the smooth function*

$$\mathbb{E}\Big[u(h, \hat{X}^{0,x}_h)\Big] = u(0, x) + C_1 u(0, x)h + C_2 u(0, x)h^2 + \ldots + C_{l+1} u(0, x)h^{l+1} + \mathcal{O}(h^{l+2})$$

for some constants C_i, and the operators $\hat{L}_x^{\alpha,r}$ are defined implicitly by

$$\mathbb{E}\Big[u(h, \hat{X}_h^{0,x})\Big] = u(0,x) + \hat{L}_x^{(0),r}u(0,x)h + \hat{L}_x^{(0,0),r}u(0,x)\frac{h^2}{2} + \ldots + \hat{L}_x^{(0)_{l+1},r}u(0,x)\frac{h^{l+1}}{(l+1)!} + \mathcal{O}(h^{l+2}).$$

Lemma 3.2.2. *Suppose that $(\mathbf{H}u_b^{l+2})$ holds for the value function u and consider a weak Taylor scheme of order $r = l+1$. For all $\alpha \in \mathcal{D}_r$ such that α is a multi-index with all entries being equal to zero, and $(0,x) \in [0,T] \times \mathbb{R}^d$, it holds that $\hat{L}_x^{\alpha,r}u(0,x) \equiv L^\alpha u(0,x)$.*

Proof. Consider a multi-index of the form $\alpha = (0)_k$ for all $k = 1, \ldots, l+1$. We use the properties of the value function to recall that

$$\mathbb{E}\big[u(h, \hat{X}_h)\big] = \mathbb{E}[u(h, X_h)] + \mathcal{O}(h^{r+1}), \tag{3.2.8}$$

given the order of the weak Taylor scheme; this follows from [KP92, Theorem 14.5.2], as we are just considering one time step (see [KP92, p.474, (14.5.12)]). Furthermore, by extending [CC14, Proposition 2.2], we can write the expansion of the weak Taylor scheme of order r, using the (unspecified) operators $\hat{L}_x^{\alpha,r}$:

$$\mathbb{E}\Big[u(h, \hat{X}_h^{0,x})\Big] = u(0,x) + \hat{L}_x^{(0),r}u(0,x)h + \ldots + \hat{L}_x^{(0)_{l+1},r}u(0,x)\frac{h^{l+1}}{(l+1)!} + \mathcal{O}(h^{l+2}).$$

For the true process X, the value function can be expanded as

$$\mathbb{E}[u(h, X_h)] = u(0,x) + L^{(0)}u(0,x)h + \ldots + L^{(0)_{l+1}}u(0,x)\frac{h^{l+1}}{(l+1)!} + \mathcal{O}(h^{l+2}).$$

Observe that from (3.2.8), and the previous two equalities, we can compare coefficients of h to establish that

$$\mathbb{E}\big[u(h, \hat{X}_h)\big] - \mathbb{E}[u(h, X_h)] = \sum_{k=1}^{l+1}\Big(\hat{L}_x^{(0)_k,r}u(0,x) - L^{(0)_k}u(0,x)\Big)\frac{h^k}{k!} + \mathcal{O}(h^{l+2})$$

and

$$\mathbb{E}\big[u(h, \hat{X}_h)\big] - \mathbb{E}[u(h, X_h)] = \mathcal{O}(h^{r+1})$$

therefore by division by h^k for $k = 0, \ldots, l+1$, $\hat{L}_x^{(0)_k,r}u(0,x) = L^{(0)_k}u(0,x)$ holds. □

The above lemma is required so that we can use the notation $\hat{L}_x^{\alpha,r}u(0,x) \equiv L^\alpha u(0,x)$ for sufficiently smooth value functions with $\alpha = (0)_k$ and $k \leq r$. We have included a few examples throughout the following sections that explicit the Taylor expansions of the value function using weak Taylor schemes.

We now similarly define the operators $\hat{L}_x^{\alpha,r}$ for $\alpha = (j) * (0)_l$ for $l \geq 0$, using the weight $(H_h^\psi)_j$:

Definition 3.2.4. *For $\psi \in \mathcal{B}^l_{[0,1]}$, a weak Taylor scheme of order r and a smooth value function u, an Itô-Taylor expansion yields*

$$\mathbb{E}\Big[(H_h^\psi)_j u(h, \hat{X}_h^{0,x})\Big] = L^{(j)}u(0,x) + \tilde{C}_1 h + \tilde{C}_2 h^2 + \ldots + \tilde{C}_l h^l + \mathcal{O}(h^{l+1}),$$

for some constants $\tilde{C}_i$, and we define implicitly the operators $\hat{L}_x^{\alpha,r}$

$$\mathbb{E}\Big[(H_h^\psi)_j u(h, \hat{X}_h^{0,x})\Big] = \hat{L}_x^{(j),r}u(0,x) + \hat{L}_x^{(1,0),r}u(0,x)h + \ldots + \hat{L}_x^{(1)*(0)_l,r}u(0,x)\frac{h^l}{l!} + \mathcal{O}(h^{l+1}),$$

for multi-indices $\alpha = (1) * (0)_k$ for $k = 0, \ldots, l$.

Lemma 3.2.3. *Fix $l \in \mathbb{N}$. Suppose $(\mathbf{H}u_b^{l+2})$ holds, a weak Taylor scheme of order $l+1$ and $\psi \in \mathcal{B}^l_{[0,1]}$. Then,*

$$\mathbb{E}_{t_i}\left[H^\psi_{t_i,h} u(t_{i+1}, X_{t_{i+1}}^{t_i,\hat{X}_{t_i}})\right] = \mathbb{E}_{t_i}\Big[H^\psi_{t_i,h} u(t_{i+1}, \hat{X}_{t_{i+1}})\Big] + \mathcal{O}(h^{l+1}).$$

Proof. We prove only for the weak Taylor scheme of order 2, using a function $\psi \in \mathcal{B}^1_{[0,1]}$. We consider the first time step, i.e. $i = 0$, and $h := t_1$ (equidistant grid). Fix $\psi \in \mathcal{B}^1_{[0,1]}$ and $(\mathbf{H}u_b^3)$ holds; the weak Taylor 2 scheme for one step (with $f \equiv 0$) is

$$\hat{X}_h = x + \gamma\sqrt{h}\Delta W + \frac{1}{2}\gamma\gamma' h\left((\Delta W)^2 - h\right) + \frac{1}{2}\gamma^2\gamma''\left(h\Delta W - \Delta Z\right), \tag{3.2.9}$$

where $\Delta W := I_h^{(j)} = \int_0^h \mathrm{d}W_s^{(j)}$, and $\Delta Z := I_h^{(j,0)} = \int_0^h W_s^{(j)}\mathrm{d}s$. Apply a Taylor expansion to $u(h, \hat{X}_h)$ around $(0, x)$ (recalling Remark 3.2.3(i)), multiply by the weight $(H_h^\psi)_j := \frac{1}{h}\int_0^h \psi(s/h)\mathrm{d}W_s^{(j)}$, and take the expectation to obtain

$$\begin{aligned}
\mathbb{E}\Big[(H_h^\psi)_j u(h,\hat{X}_h)\Big] &= \mathbb{E}\Big[(\tfrac{1}{h}\textstyle\int_0^h \psi(s/h)\mathrm{d}W_s^{(j)})u(0,x)\Big] \\
&+ \mathbb{E}\Big[(\tfrac{1}{h}\textstyle\int_0^h \psi(s/h)\mathrm{d}W_s^{(j)})\left((\hat{X}_h - x)\partial_x u(0,x) + h\partial_t u(0,x)\right)\Big] \\
&+ \mathbb{E}\Big[(\tfrac{1}{h}\textstyle\int_0^h \psi(s/h)\mathrm{d}W_s^{(j)})\tfrac{(\hat{X}_h-x)^2\partial_{xx}u(0,x)+2h(\hat{X}_h-x)\partial_{tx}u(0,x)+h^2\partial_{tt}u(0,x)}{2!}\Big] \\
&+ \mathbb{E}\Big[(H_h^\psi)_j\tfrac{(\hat{X}_h-x)^3\partial_{xxx}u(0,x)+3h(\hat{X}_h-x)^2\partial_{txx}u(0,x)+3h^2(\hat{X}_h-x)\partial_{ttx}u(0,x)+h^3\partial_{ttt}u(0,x)}{3!}\Big] \\
&+ \mathbb{E}\Big[(H_h^\psi)_j\left(\tfrac{(\hat{X}_h-x)^4\partial_{xxxx}+4h(\hat{X}_h-x)^3\partial_{txxx}+6h^2(\hat{X}_h-x)^2\partial_{ttxx}+4h^3(\hat{X}_h-x)\partial_{tttx}+h^4\partial_{tttt}}{4!}\right)u(0,x)\Big] \\
&+ \ldots;
\end{aligned} \tag{3.2.10}$$

now consider the individual terms. The first term on the RHS of (3.2.10) is zero since ψ is bounded. For the second term, observe that by Itô's isometry, and the properties of $\psi \in \mathcal{B}^l_{[0,1]}$ in Definition 3.1.1, then

$$\begin{aligned}
&\mathbb{E}\Big[\big(\tfrac{1}{h}\int_0^h \psi(s/h)\mathrm{d}W_s^{(j)}\big)\left((\hat{X}_h - x)\partial_x u(0,x) + h\partial_t u(0,x)\right)\Big] \\
&= \mathbb{E}\Big[\tfrac{1}{h}\int_0^h \psi(s/h)\mathrm{d}s\gamma(x) + \tfrac{1}{h}\tfrac{1}{2}\gamma''(x)\gamma^2(x)\int_0^h \psi(s/h)s\mathrm{d}s\Big]\partial_x u(0,x) + h\mathbb{E}\Big[\int_0^h \psi(s/h)\mathrm{d}W_s^{(j)}\Big]\partial_t u(0,x) \\
&= \gamma(x)\partial_x u(0,x) = \gamma(x)\Delta.
\end{aligned} \tag{3.2.11}$$

From the third line of (3.2.10), consider

$$\mathbb{E}\left[\left(\frac{1}{h}\int_0^h \psi(s/h)\mathrm{d}W_s^{(j)}\right)\frac{(\hat{X}_h - x)^2\partial_{xx}u(0,x) + 2h(\hat{X}_h - x)\partial_{tx}u(0,x) + h^2\partial_{tt}u(0,x)}{2!}\right];$$

the first term evaluates to ($\gamma \to \gamma(x)$, $\gamma' \to \gamma'(x)$, $\gamma'' \to \gamma''(x)$)

$$\mathbb{E}\left[(H_h^\psi)_j\frac{(\hat{X}_h - x)^2}{2!}\partial_{xx}u(0,x)\right] = \left(\gamma'\gamma^2 h + \frac{1}{4}\gamma''\gamma^3\gamma' h^2\right)\partial_{xx}u(0,x);$$

since from one of the cross terms

$$\mathbb{E}\left[(H_h^\psi)_j\frac{1}{2}\gamma'\gamma^2((\Delta W)^3 - \Delta W h)\right] = \frac{3}{2}\gamma^2\gamma' h - \frac{1}{2}\gamma'\gamma^2 h = \gamma'\gamma^2 h.$$

and other terms such as

$$\mathbb{E}\left[\frac{1}{4}(H_h^\psi)_j\gamma'\gamma''\gamma^3\left((\Delta W)^2 - h\right)(h\Delta W - \Delta Z)\right] = \mathcal{O}(h^2).$$

Other terms evaluate to zero using $\int_0^1 \psi(s)s\mathrm{d}s = 0$ and $h\Delta W - \Delta Z = \int_0^h s\mathrm{d}W_s^{(j)}$. We also consider

$$\mathbb{E}\left[(H_h^\psi)_j\frac{(\hat{X}_h - x)^3}{3!}\partial_{xxx}u(0,x)\right] = \frac{1}{2}\gamma^3 h\partial_{xxx}u(0,x) + \mathcal{O}(h^2),$$

which can be observed using the same properties.

For the cross-term $\partial_{tx}u(0,x)$, observe that

$$\begin{aligned}
\mathbb{E}\left[(H_h^\psi)_j\frac{2(\hat{X}_h - x)h}{2!}\right] &= \mathbb{E}\left[\left(\int_0^h \psi(s/h)\mathrm{d}W_s^{(j)}\right)\left(\gamma\Delta W + \frac{\gamma'\gamma}{2}\left((\Delta W)^2 - h\right) + \frac{\gamma''\gamma^2}{2}\int_0^h s\mathrm{d}W_s^{(j)}\right)\right] \\
&= \gamma(x)\int_0^h \psi(s/h)\mathrm{d}s + \frac{1}{2}\gamma''\gamma^2\int_0^h \psi(s/h)s\mathrm{d}s = \gamma(x)h,
\end{aligned}$$

by the properties in Definition 3.1.1 and a change of variables.

Similarly, for higher order Taylor expansions, the expectation of the product with the weight H_h^ψ is controlled by $\mathcal{O}(h^2)$. By combining the above equalities, we can conclude that

$$\mathbb{E}\Big[(H_h^\psi)_j u(h,\hat{X}_h)\Big] = \gamma(x)\partial_x u(0,x) + \mathcal{O}(h^2),$$

since $L^{(1,0)}u(0,x) = \gamma(x)\Big\{\partial_{tx} + \gamma(x)\gamma'(x)\partial_{xx} + \frac{1}{2}\gamma^2(x)\partial_{xxx}\Big\}u(0,x) = 0$. The proof similarly follows for higher order weak Taylor schemes and the corresponding functions $\psi \in \mathcal{B}^l_{[0,1]}$. □

Corollary 3.2.1. *Consider the value function solving* (2.1.1), *where* $(\mathbf{H}u_b^{l+2})$ *holds. For* $\psi \in \mathcal{B}^l_{[0,1]}$, *and a weak Taylor scheme of order* r, *then the operators* $\hat{L}_x^{\alpha,r}$ *are defined as*

$$\hat{L}_x^{(1),r}u(0,x) = L^{(1)}u(0,x), \qquad \cdots \qquad , \qquad \hat{L}_x^{(1)*(0)_l,r}u(0,x) = L^{(1)*(0)_l}u(0,x).$$

Proof. Recall, that for $\psi \in \mathcal{B}^l_{[0,1]}$, and a weak Taylor scheme of order $r = l+1$, we have that for value functions such that $(\mathbf{H}u_b^{l+2})$,

$$\mathbb{E}\Big[H_h^\psi u(h,X_h)\Big] = L^{(1)}u(0,x) + L^{(1,0)}u(0,x)h + \ldots + L^{(1)*(0)_l}u(0,x)\frac{h^l}{l!} + \mathcal{O}(h^{l+1});$$

since the value function is such that $L^{(0)}u(0,x) = 0$, then $\mathbb{E}\Big[H_h^\psi u(h,X_h)\Big] = L^{(1)}u(0,x) + \mathcal{O}(h^{l+1})$. We conclude by applying Lemma 3.2.3, since $\mathbb{E}\Big[H_h^\psi u(h,X_h)\Big] = \mathbb{E}\Big[H_h^\psi u(h,\hat{X}_h)\Big] + \mathcal{O}(h^{l+1})$. □

Remark 3.2.4.

(i) We can similarly obtain the operators $\hat{L}_x^{\alpha,r}$ *for other multi-indices* α, *by appropriately selecting a weight. We shall not require them, so we do not explicit them here.*

(ii) We see that by imposing smoothness and boundedness assumptions on $L^\alpha u(0,x)$, *these properties can be passed on to* $\hat{L}_x^{\alpha,r}u(0,x)$ *for weak Taylor schemes of sufficiently high order.*

Theorem 3.2.1 paves the way for a general result for higher-order weak Taylor schemes, which can be justified by [TT90, Theorem 1 (iv)], where the results are shown in the $d = m = 1$ case:

Corollary 3.2.2. *Fix* $l \in \mathbb{N}$. *Consider a weak Taylor scheme of order* $l+1$, *on an equidistant mesh* π, *such that* $|\pi| = h$, *suppose* $(\mathbf{H}u_b^{l+3})$ *holds for a value function* u, *and let* $\psi \in \mathcal{B}^l_{[0,1]}$. *Then,*

$$\mathbb{E}\Big[H_h^\psi g(\hat{X}_T)\Big] = L^{(1)}u(0,x) + \mathcal{O}(h^{l+1}).$$

Proof. From Lemma 3.2.3 it follows that $\mathbb{E}\Big[H_h^\psi u(h,\hat{X}_h)\Big] = L^{(1)}u(0,x) + \mathcal{O}(h^{l+1})$, where $h = t_1$. Define $\varphi(t, x_t) \equiv \hat{L}_{x_t}^{(0)_{l+2},r}u(t,x_t)$, where the operator is associated to a weak Taylor scheme of order r. From an extension of (3.2.7), then

$$\mathbb{E}\left[H_h^\psi\left\{u(t_{i+1},\hat{X}_{t_{i+1}}) - u(t_i,\hat{X}_{t_i})\right\}\right] = \mathbb{E}\left[H_h^\psi\left\{\frac{h^{l+2}}{(l+2)!}\varphi(t_i,\hat{X}_{t_i}) + R\right\}\right],$$

where $R := I_{t_i,t_{i+1}}^{(0)_{l+2}}\left[\varphi(\cdot,\hat{X}_\cdot) - \varphi(t_i,\hat{X}_{t_i})\right]$. An application of the Cauchy-Schwarz inequality, yields $\mathbb{E}\Big[H_h^\psi R\Big] \le \|H_h^\psi\|_2\|R\|_2 \le Ch^{l+2}$ for $\psi \in \mathcal{B}_{[0,1]}^l$ and φ sufficiently smooth. We conclude with a first-order expansion of the sum of $\mathbb{E}\Big[(H_h^\psi)_j\varphi(t_i,\hat{X}_{t_i})\Big]$ which is treated similarly as in the proof of Theorem 3.2.1. □

The above corollary is constructive for selecting a discretisation scheme on $[0,T]$, and an appropriate function $\psi \in \mathcal{B}_{[0,1]}^l$ in order to have approximations of Δ to a higher order of bias.

3.3 Extrapolation method

We now recall Section 2.4.2 and the assumed expansion. We shall prove that the expansion in (2.4.3) holds for the $\hat{X}$ discretisation using weak Taylor schemes.

3.3.1 Euler scheme

The next lemma provides an approximation of an integral, using summations for a general function $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, which is characterised by the smoothness of the function:

Lemma 3.3.1. *For $v \in \mathcal{G}_b^2$, then*

$$\sum_{i=0}^{n-1} h\mathbb{E}[v(t_i,X_{t_i})] = \int_0^T \mathbb{E}[v(s,X_s)]\,\mathrm{d}s + h\int_0^T \frac{1}{2}\mathbb{E}\Big[L^{(0)}v(s,X_s)\Big]\,\mathrm{d}s + \mathcal{O}(h^2)\,.$$

Proof. 1. First, recall that if $f : [0,T] \to \mathbb{R}$ is $\mathcal{C}^1$, then we recognise an order 1 approximation of an integral, which allows us to write

$$\sum_{i=0}^{n-1} hf(t_i) = \int_0^T f(s)\mathrm{d}s + \mathcal{O}(h). \tag{3.3.1}$$

Indeed, observe that

$$\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\{f(s)-f(t_i)\}\mathrm{d}s=\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\int_0^1 f'(t_i+\lambda(s-t_i))(s-t_i)\mathrm{d}\lambda\mathrm{d}s.$$

The results follow from the continuity of f' on $[0,T]$.

2. We now compute

$$\int_0^T \mathbb{E}[v(s,X_s)]\,\mathrm{d}s-\sum_{i=0}^{n-1}h\mathbb{E}[v(t_i,X_{t_i})]=\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}[v(s,X_s)-v(t_i,X_{t_i})]\,\mathrm{d}s.$$

From the weak expansion in Lemma 3.2.1, we have

$$\begin{aligned}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}[v(s,X_s)-v(t_i,X_{t_i})]\,\mathrm{d}s&=\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\mathbb{E}\Big[L^{(0)}v(t_i,X_{t_i})\Big]\,(s-t_i)\mathrm{d}s+\mathcal{O}(h^2)\\&=h\left(h\sum_{i=0}^{n-1}\frac{1}{2}\mathbb{E}\Big[L^{(0)}v(t_i,X_{t_i})\Big]\right)+\mathcal{O}(h^2).\end{aligned}$$

The proof of the lemma is concluded by using the first step. □

We expand $\mathbb{E}\Big[(H_h^{\psi})_j g(\hat{X}_T)\Big]$ in the step size h with $\psi\equiv 1$, to justify an extrapolation method:

Proposition 3.3.1. *Suppose that $u\in\mathcal{G}_b^4$, $\psi\equiv 1$ and assume an Euler scheme for the discretisation of the process X. Then,*

$$\mathbb{E}\Big[(H_h^1)_j g(\hat{X}_T)\Big]=u^{(j)}(0,x)+hC_{1,j,x,T}+\mathcal{O}(h^2).$$

Proof. Note $t_1=h$, and $H_h:=(H_h^1)_j$.

1. Applying Ito's Formula, we compute successively (similarly to (3.2.6))

$$\begin{aligned}\mathbb{E}_{t_i}[u(t_{i+1},\hat{X}_{t_{i+1}})-u(t_i,\hat{X}_{t_i})]&=\mathbb{E}_{t_i}\Big[\int_{t_i}^{t_{i+1}}\hat{L}^{(0)}_{\hat{X}_{t_i}}u(s,\hat{X}_s)\mathrm{d}s\Big]\\&=h\hat{L}^{(0)}_{\hat{X}_{t_i}}u(t_i,\hat{X}_{t_i})+\mathbb{E}_{t_i}\Big[\int_{t_i}^{t_{i+1}}\int_{t_i}^{s}\hat{L}^{(0,0)}_{\hat{X}_{t_i}}u(r,\hat{X}_r)\mathrm{d}r\mathrm{d}s\Big]\\&=\frac{h^2}{2}\hat{L}^{(0,0)}_{\hat{X}_{t_i}}u(t_i,\hat{X}_{t_i})+\frac{h^3}{6}\hat{L}^{(0,0,0)}_{\hat{X}_{t_i}}u(t_i,\hat{X}_{t_i})+\mathcal{O}(h^4),\end{aligned}\tag{3.3.2}$$

where to get the last equality we used also the fact that $\hat{L}^{(0)}_{\hat{X}_{t_i}}u(t_i,\hat{X}_{t_i})=0$, and the boundedness of the derivatives of the value function. We define

$$\phi_e^1(s,x) := \frac{1}{2}\hat{L}_x^{(0,0)}u(s,x)\ , \qquad \phi_e^2(s,x) := \frac{1}{6}\hat{L}_x^{(0,0,0)}u(s,x).$$

With this notation, we obtain,

$$\mathbb{E}\big[H_h\{g(\hat{X}_T)-u(h,\hat{X}_h)\}\big] = \mathbb{E}\left[H_h\left(\sum_{i=1}^{n-1}\mathbb{E}_{t_1}\Big[h^2\phi_e^1(t_i,\hat{X}_{t_i})+h^3\phi_e^2(t_i,\hat{X}_{t_i})\Big]+\mathcal{O}(h^4)\right)\right]. \quad (3.3.3)$$

From [TT90, Theorem 1], we know that

$$\mathbb{E}_{t_1}\Big[\phi_e^1(t_i,\hat{X}_{t_i})\Big] = \mathbb{E}_{t_1}\Big[\phi_e^1(t_i,X_{t_i}^{t_1,\hat{X}_{t_1}})\Big]+h\tilde{\phi}_{e,i}^1(t_1,\hat{X}_{t_1})+\mathcal{O}(h^2)\ ,$$

for some bounded function $\tilde{\phi}_{e,i}^1$, and

$$\mathbb{E}_{t_1}\Big[\phi_e^2(t_i,\hat{X}_{t_i})\Big] = \mathbb{E}_{t_1}\Big[\phi_e^2(t_i,X_{t_i}^{t_1,\hat{X}_{t_1}})\Big]+\mathcal{O}(h).$$

Combining these equalities with (3.3.3), we obtain

$$\begin{aligned}\mathbb{E}\big[H_h\{g(\hat{X}_T)-u(h,\hat{X}_h)\}\big] &= \mathcal{O}(h^{\frac{5}{2}})+h^3\mathbb{E}\left[H_h\sum_{i=1}^{n-1}\tilde{\phi}_{e,i}^1(t_1,\hat{X}_{t_1})\right]\\ &+\mathbb{E}\left[H_h\left(\sum_{i=1}^{n-1}\mathbb{E}_{t_1}\Big[h^2\phi_e^1(t_i,X_{t_i}^{t_1,\hat{X}_{t_1}})+h^3\phi_e^2(t_i,X_{t_i}^{t_1,\hat{X}_{t_1}})\Big]\right)\right],\end{aligned} \quad (3.3.4)$$

using the Cauchy-Schwarz inequality and the variance of H_h. Using Lemma 3.2.1, we observe that

$$\mathbb{E}\left[H_h\sum_{i=1}^{n-1}\tilde{\phi}_{e,i}^1(t_1,\hat{X}_{t_1})\right] = \sum_{i=1}^{n-1}\Big[L^{(j)}\tilde{\phi}_{e,i}^1(0,x)+\mathcal{O}(h)\Big] = \mathcal{O}\left(\frac{1}{h}\right). \quad (3.3.5)$$

We also compute

$$\sum_{i=1}^{n-1}h\mathbb{E}_{t_1}\Big[\phi_e^2(t_i,X_{t_i}^{t_1,\hat{X}_{t_1}})\Big] = \mathbb{E}_{t_1}\Big[\int_{t_1}^T\phi_e^2(s,X_s^{t_1,\hat{X}_{t_1}})\mathrm{d}s\Big]+\mathcal{O}(h), \quad (3.3.6)$$

leading to

$$h^2\mathbb{E}\left[H_hh\sum_{i=1}^{n-1}\phi_e^2(t_i,X_{t_i}^{t_1,\hat{X}_{t_1}})\right] = h^2\left(L^{(j)}\varphi_e^2(0,x)+\mathcal{O}(h^{\frac{1}{2}})\right) = \mathcal{O}(h^2), \quad (3.3.7)$$

where $\varphi_e^2(t,x) := \mathbb{E}\Big[\int_t^T \phi_e^2(s,X_s)\mathrm{d}s\Big]$ and using Lemma 3.3.1. Similarly,

$$\begin{aligned}\sum_{i=1}^{n-1} h\mathbb{E}_{t_1}\Big[\phi_e^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\Big] &= \mathbb{E}_{t_1}\Big[\int_{t_1}^T \phi_e^1(s, X_s^{t_1,\hat{X}_{t_1}})\mathrm{d}s\Big] + \frac{h}{2}\mathbb{E}_{t_1}\Big[\int_{t_1}^T L^{(0)}\phi_e^1(s, X_s^{t_1,\hat{X}_{t_1}})\mathrm{d}s\Big] + \mathcal{O}(h^2)\\ &= \varphi_e^1(t_1,\hat{X}_{t_1}) + \tilde{\varphi}_e^1(t_1,\hat{X}_{t_1})h + \mathcal{O}(h^2),\end{aligned}$$

where for any $(t,x) \in [0,T]\times\mathbb{R}^d$, $\varphi_e^1(t,x) := \mathbb{E}_t\Big[\int_t^T \phi_e^1(s,X_s^{t,x})\mathrm{d}s\Big]$ and $\tilde{\varphi}_e^1(t,x) := \mathbb{E}_t\Big[\int_t^T L^{(0)}\phi_e^1(s,X_s^{t,x})\mathrm{d}s\Big]/2$. Note that $\varphi_e^1 \in \mathcal{G}_b^2$ and $\tilde{\varphi}_e^1 \in \mathcal{G}_b^1$.
We compute

$$h\mathbb{E}\left[H_h h\sum_{i=1}^{n-1}\phi_e^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\right] = hL^{(j)}\varphi_e^1(0,x) + \mathcal{O}(h^2); \tag{3.3.8}$$

combining (4.1.6), (3.3.5), (3.3.7) and (3.3.8) we get

$$\mathbb{E}\big[H_h\{g(\hat{X}_T) - u(h,\hat{X}_h)\}\big] = hL^{(j)}\varphi_e^1(0,x) + \mathcal{O}(h^2). \tag{3.3.9}$$

2. We now observe that

$$\begin{aligned}\mathbb{E}\big[H_h g(\hat{X}_T)\big] &= \mathbb{E}\big[H_h\{g(\hat{X}_T) - u(h,\hat{X}_h)\}\big] + \mathbb{E}\big[H_h u(h,\hat{X}_h)\big]\\ &= hL^{(j)}\varphi_e^1(0,x) + \mathcal{O}(h^2) + \mathbb{E}\big[H_h u(h,\hat{X}_h)\big].\end{aligned}$$

Using Lemma 3.2.1, we have

$$\mathbb{E}\big[H_h u(h,\hat{X}_h)\big] = L^{(j)}u(0,x) + h\hat{L}_x^{(j,0)}u(0,x) + \mathcal{O}(h^2).$$

Combining the above expansion with (3.3.9), we finally obtain

$$\mathbb{E}\big[H_h g(\hat{X}_T)\big] = L^{(j)}u(0,x) + h\left(L^{(j)}\varphi_e^1(0,x) + \hat{L}_x^{(j,0)}u(0,x)\right) + \mathcal{O}(h^2),$$

which completes the proof. □

We consider $\hat{X}^{n/2} = (\hat{X}_t^{n/2})_{t\in[0,T]}$, the Euler scheme associated with a grid of stepsize of $2h$, recalling the notation from Remark 3.2.1(iii). The following result yields a second order approximation of the Δ using the Romberg method:

Theorem 3.3.1. *Suppose that $u \in \mathcal{G}_b^4$ and $\psi \equiv 1$. Using an Euler scheme we have*

$$2\mathbb{E}\Big[(H_h^{\psi})_j g(\hat{X}_T^n)\Big] - \mathbb{E}\Big[(H_{2h}^{\psi})_j g(\hat{X}_T^{n/2})\Big] = L^{(j)}u(0,x) + \mathcal{O}(h^2). \tag{3.3.10}$$

Proof. The proof is a direct consequence of the previous proposition, and noting that $\hat{L}_x^{(j)}u(0,x) = L^{(j)}u(0,x)$. □

Remark 3.3.1.

(i) By the above arguments, we can present a third-order scheme using the Euler scheme; following the same steps, we can show that for $(\mathbf{H}u_b^5)$ *and an Euler scheme with $\psi \equiv 1$, then*

$$\mathbb{E}\Big[H_h^{\psi}u(h,\hat{X}_h)\Big] = \gamma(x)\Delta + d_1 h + d_2\frac{h^2}{2} + \mathcal{O}(h^3);$$

straightforward extrapolation suggests

$$3\mathbb{E}\Big[H_h^{\psi}g(\hat{X}_T^n)\Big] - \frac{5}{2}\mathbb{E}\Big[H_{2h}^{\psi}g(\hat{X}_T^{n/2})\Big] + \mathbb{E}\Big[H_{3h}^{\psi}g(\hat{X}_T^{n/3})\Big] = \gamma(x)\Delta + \mathcal{O}(h^3).$$

(ii) The step functions defined using (2.4.4) *are similar to the step functions defined using the families $\mathcal{B}_{[0,1]}^l$. Setting $c = 1/2$ for $\psi_{s,1}$, coincides with the scheme using* (2.4.4) *yielding the same step function and weight variance. For higher-order schemes, we saw in Example 3.1.3 that the optimal $(c,c') \neq (1/3,2/3)$ which are the suggested parameters for the scheme using equidistant step functions; therefore weights defined using $\psi \in \mathcal{B}_{[0,1]}^l$ achieve a better variance bound compared to the equidistant step functions defined using* (2.4.4).

3.3.2 Weak Taylor scheme of order 2

We extend the previous result for the Euler scheme to a weak Taylor scheme of order 2 to perform extrapolation. Recall the scheme (3.2.9), with the drift set to zero. For extrapolation, we proceed from Lemma 3.2.3 with an additional level of Taylor expansions ($d = m = 1$):

Lemma 3.3.2. *Suppose a weak Taylor 2 scheme, $\psi \in \mathcal{B}_{[0,1]}^1$ and* $(\mathbf{H}u_b^4)$*. Then,* $\mathbb{E}\Big[H_h^{\psi}u(h,\hat{X}_h)\Big] = u_0^{(1)} + C_{2,x,T}h^2 + \mathcal{O}(h^3)$.

Proof. This proof extends Lemma 3.2.3 to an additional order of Taylor expansions, up to $\mathcal{O}(h^3)$ terms. For a general $\psi \in \mathcal{B}^1_{[0,1]}$:

$$\begin{aligned}\mathbb{E}\Big[H_h^{\psi} u(h,\hat{X}_h)\Big] =&\gamma(x)\partial_x u(0,x) + \Big\{\frac{1}{2}\gamma\partial_{ttx} + \frac{1}{4}\gamma^2\gamma''\partial_{tx} + \gamma^2\gamma'\partial_{txx} + \frac{1}{2}\gamma^3\partial_{txxx} \\ &+ \frac{1}{4}\gamma^3\gamma'\gamma''\partial_{xx} + \left(\frac{3}{2}\gamma^3(\gamma')^2 + \frac{1}{2}\gamma\left[-\frac{1}{2}\gamma^2(\gamma')^2 + \frac{1}{2}\gamma''\gamma^3\right] + \frac{1}{8}\gamma''\gamma^4\right)\partial_{xxx} \\ &+ \gamma^4\gamma'\partial_{xxxx} + \frac{1}{8}\gamma^5\partial_{xxxxx}\Big\} u(0,x)h^2 + \mathcal{O}(h^3),\end{aligned}$$

therefore we can conclude. □

Theorem 3.3.2. *Suppose that $u \in \mathcal{G}^5_b$. Then, for $\psi \in \mathcal{B}^1_{[0,1]}$, and a weak Taylor scheme of order 2, with an equidistant stepsize $|\pi| = h$*

$$\frac{4}{3}\mathbb{E}\Big[(H_h^{\psi})_j g(\hat{X}_T^n)\Big] - \frac{1}{3}\mathbb{E}\Big[(H_{2h}^{\psi})_j g(\hat{X}_T^{n/2})\Big] = L^{(j)}u(0,x) + \mathcal{O}(h^3).$$

Proof. Observe that for $j = 1, \ldots, m$

$$\begin{aligned}\mathbb{E}\Big[(H_h^{\psi})_j g(\hat{X}_T)\Big] &= \mathbb{E}\Big[(H_h^{\psi})_j\{g(\hat{X}_T) - u(h,\hat{X}_h)\}\Big] + \mathbb{E}\big[(H_h)_j u(h,\hat{X}_h)\big] \\ &= L^{(j)}u(0,x) + \mathbb{E}\Big[(H_h^{\psi})_j\{g(\hat{X}_T) - u(h,\hat{X}_h)\}\Big] + C_{2,j,x,T}h^2 + \mathcal{O}(h^3),\end{aligned}$$

using Lemma 3.3.2 for Taylor expanding the value function $u(h, \hat{X}_h)$ using a weight $\psi \in \mathcal{B}^1_{[0,1]}$. Now recall the telescoping term (3.3.2): since we are using a weak Taylor scheme of order 2, from Lemma 3.2.2 we know that $\hat{L}_x^{(0,0),2}u(s,x) = L^{(0,0)}u(s,x) = 0$, therefore

$$\mathbb{E}_{t_i}\big[u(t_{i+1},\hat{X}_{t_{i+1}}) - u(t_i,\hat{X}_{t_i})\big] = \frac{h^3}{6}\hat{L}_{\hat{X}_{t_i}}^{(0,0,0),2}u(t_i,\hat{X}_{t_i}) + \frac{h^4}{24}\hat{L}_{\hat{X}_{t_i}}^{(0,0,0,0),2}u(t_i,\hat{X}_{t_i}) + \mathcal{O}(h^5).$$

The proof follows from the same argument as the proof for the Euler scheme extrapolation (Proposition 3.3.1). For $(s,x) \in [0,T] \times \mathbb{R}^d$, we denote for our second order scheme

$$\phi_s^1(s,x) := \frac{1}{6}\hat{L}_x^{(0,0,0),2}u(s,x)\ ,\ \phi_s^2(s,x) := \frac{1}{24}\hat{L}_x^{(0,0,0,0),2}u(s,x),$$

where $\hat{L}_x^{\alpha,2}$ is the operator of the weak Taylor scheme of order 2 for multi-index α. With this notation, we obtain

$$\mathbb{E}\Big[(H_h^{\psi})_j\{g(\hat{X}_T) - u(h,\hat{X}_h)\}\Big] = \mathbb{E}\left[(H_h^{\psi})_j\sum_{i=1}^{n-1}\left(\mathbb{E}_{t_1}\Big[h^3\phi_s^1(t_i,\hat{X}_{t_i}) + h^4\phi_s^2(t_i,\hat{X}_{t_i})\Big] + \mathcal{O}(h^5)\right)\right].$$

Observe that we can approximate as an integral using Lemma 3.3.1,

$$\begin{aligned}\sum_{i=1}^{n-1} h\mathbb{E}_{t_1}\Big[\phi_s^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\Big] &= \mathbb{E}_{t_1}\Big[\int_{t_1}^T \phi_s^1(s, X_s^{t_1,\hat{X}_{t_1}})\mathrm{d}s\Big] + \frac{h}{2}\mathbb{E}_{t_1}\Big[\int_{t_1}^T L^{(0)}\phi_s^1(s, X_s^{t_1,\hat{X}_{t_1}})\mathrm{d}s\Big] + \mathcal{O}(h^2)\\ &= \varphi_s^1(t_1, \hat{X}_{t_1}) + \tilde{\varphi}_s^1(t_1, \hat{X}_{t_1})h + \mathcal{O}(h^2),\end{aligned}$$

where for any $(t,x) \in [0,T] \times \mathbb{R}^d$, $\varphi_s^1(t,x) := \mathbb{E}_t\Big[\int_t^T \phi_s^1(s, X_s^{t,x})\mathrm{d}s\Big]$ and $\tilde{\varphi}_s^1(t,x) := \mathbb{E}_t\Big[\int_t^T L^{(0)}\phi_s^1(s, X_s^{t,x})\mathrm{d}s\Big]/2$. Note that with $\varphi_s^1 \in \mathcal{G}_b^2$ and $\tilde{\varphi}_s^1 \in \mathcal{G}_b^1$, we compute as before

$$h\mathbb{E}\left[(H_h)_j h \sum_{i=1}^{n-1} \phi_s^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\right] = hL^{(j)}\varphi_s^1(0,x) + \mathcal{O}(h^2),$$

and similarly to (3.3.6) and (3.3.7), we have

$$h^3\mathbb{E}\left[(H_h^\psi)_j \sum_{i=1}^{n-1} h\phi_s^2(t_i, \hat{X}_{t_i})\right] = h^3\left(L^{(j)}\varphi_s^2(0,x) + \mathcal{O}(h^{\frac{1}{2}})\right) = \mathcal{O}(h^3),$$

where $\varphi_s^2(t,x) := \mathbb{E}\Big[\int_t^T \phi_e^2(s, X_s^{t,x})\mathrm{d}s\Big]$. To conclude, observe that using the Cauchy-Schwarz inequality, $\mathbb{E}\Big[(H_h^\psi)_j \sum_{i=1}^{n-1}\mathcal{O}(h^5)\Big] = \mathbb{E}\Big[(H_h^\psi)_j\mathcal{O}(h^4)\Big] = \mathcal{O}(h^{7/2})$, which enables us to conclude that

$$\mathbb{E}\Big[(H_h^\psi)_j g(\hat{X}_T^n)\Big] = L^{(j)}u(0,x) + C_{2,j,x,T}h^2 + \mathcal{O}(h^3);$$

by extrapolation we obtain a scheme for Δ approximations with a bias of $\mathcal{O}(h^3)$. □

3.4 Simulation results

We now price contingent claims and approximate the Greeks using finite difference methods. It is often the preferred technique for small-dimensional problems. Finite difference methods replace the partial derivatives by their approximations on a grid, in order to reduce the problem to a finite set of algebraic equations (for details, see [Duf06]).

We implement the explicit finite difference method, which marches-back in time from the terminal payoff at expiry time T, to the initial time $t = 0$. The benefit of this technique is the quick implementation, whilst the drawback is that in order to guarantee stability of the algorithm, doubling the number of space steps increases the number of required time steps four-fold.

We consider the following smooth, Lipschitz continuous diffusion and payoff function for our numerical examples:

Example 3.4.1 (Smooth diffusion and payoff)**.** *Suppose zero drift, diffusion* $\gamma(u) \equiv 1 + \sin^2(u)$*, and payoff* $g(u) \equiv \arctan(u)$*. Consider initial condition* $x = 0.3$ *and* $T = 1$ *as the parameters. The true price,* Δ *and* Γ *of this option are* $(0.155, 0.503, -0.086)$*, computed using the finite difference method,* 1000 *spatial steps and* 4000000 *time steps.*

Remark 3.4.1. *We compute the slope of the straight line of the mean squared error against the computational cost (*$\log - \log$ *scale), which is proxied by the runtime (measured in seconds) of the algorithms. Throughout this numerical section, we include the slope and constant of the straight line in the legend for the various plots, which are an indication of the complexity of the various techniques. We use the weak Taylor schemes (see [KP92, Chapter 14] for more details). The parameter* ζ *determines the size of* $h := 1/N^{\zeta}$*, for which the first step is simulated, and also defines the equidistant step size. We summarise the parameters in tables for the different techniques, explaining the scheme, weight and convergence properties.*

3.4.1 High-order Δ approximation

Consider N simulations, and fix the step size $|\pi|$ to equal to the h-increment of the weight defined; i.e. $|\pi| := h$. For this example, to approximate the Δ, the approximation $\mathbb{E}\Big[H_h^{\psi} g(\hat{X}_T^n)\Big]$ has a bias of $\mathcal{O}(h^r)$, where r is the order of the scheme used ($r = 1$ corresponds to the Euler scheme, $r = 2$ corresponds to the second order weak Taylor scheme, etc). In Table 3.1, we explicit the implementation of the Δ using the different schemes, and weight requirements, where $h = |\pi| = 1/N^{\zeta}$.

r (Scheme)	Weight	ζ	MSE	Cost	Slope
1 (Euler)	$\psi \equiv 1$	$1/3$	$\mathcal{O}(N^{-2/3})$	$\mathcal{O}(N^{4/3})$	$-1/2$
2 (WT2)	$\psi_{s,1}$, $\psi_{p,1}$	$1/5$	$\mathcal{O}(N^{-4/5})$	$\mathcal{O}(N^{6/5})$	$-2/3$
3 (WT3)	$\psi_{s,2}$, $\psi_{p,2}$	$1/7$	$\mathcal{O}(N^{-6/7})$	$\mathcal{O}(N^{8/7})$	$-3/4$

Table 3.1: Implementation and MSE for the Delta.

In Figure 3.1, we consider high-order approximations for the Δ, for Example 3.4.1. Observe that the slope of the approximation improves dramatically from the Euler scheme ($\psi \equiv 1$), to the weak Taylor scheme 2 (with $\psi_{s,1}$) and consequently the weak Taylor order 3 scheme with $\psi_{s,2}$.

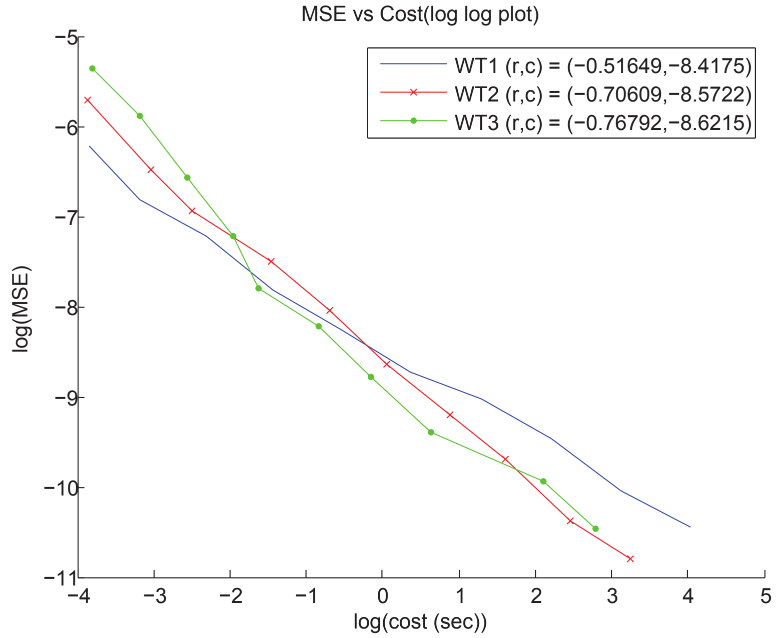

Figure 3.1: MSE vs Cost (log − log) in seconds for the Delta, each with 500 repeats. Parameters as in Table 3.1.

For this example, for an MSE of approximately $\exp(-10)$, the weak Taylor order 3 scheme takes 20 seconds, whilst the Euler scheme takes approximately 60 seconds; even though the higher weak Taylor scheme is more computationally demanding, the fact that ζ is much lower means that the step sizes are considerably bigger, translating to a faster runtime. In addition, recall the discussion from Section 3.1.2 on the size of N for which high-order schemes are preferred to higher-order schemes for a fixed computational effort available: from Figure 3.1, we observe that for runtimes of more than approximately 0.3 seconds, high-order Δ approximations are preferred to the Euler scheme and $\psi \equiv 1$ (basic approximation).

Remark 3.4.2 (Different schemes on $[0, h]$ and $[h, T]$).

(i) We could consider using different schemes on $[0, h]$ and $[h, T]$, where $[0, h]$ is discretised using one time step.

(ii) The computational cost of each method is determined by the step size and scheme for the discretisation of $[h, T]$.

3.4.2 Extrapolation Delta

We now consider Romberg-Richardson style extrapolation from Theorems 3.3.1 and 3.3.2.

ψ	A	B	Scheme	ζ	MSE	Cost	Slope
$\psi \equiv 1$	2	1	Euler	$1/5$	$\mathcal{O}(N^{-4/5})$	$\mathcal{O}(N^{6/5})$	$-2/3$
$\psi_{s,1}$	$4/3$	$1/3$	WT2	$1/7$	$\mathcal{O}(N^{-6/7})$	$\mathcal{O}(N^{8/7})$	$-3/4$

Table 3.2: Parameters for extrapolating the Δ using Euler and the weak Taylor scheme of order 2, using $h := 1/N^{\zeta}$. MSE, Computational cost and $\log - \log$ slope. For numerics, see Figure 3.3.

Example 3.4.2. *We consider extrapolation with independent Brownian paths and the same Brownian paths, for Example 3.4.1:*

(i) ***Independent Brownian paths:*** *the two extrapolation terms in* (3.3.10) *are calculated independently. This achieves the expected strong slope of MSE vs Cost of $-2/3$ as expected from Table 3.2 for the Euler scheme with $\zeta = 1/5$. In Figure 3.2, we show the actual average values of Δ obtained, showing the superior performance of the extrapolation.*

(ii) ***Same Brownian paths:*** *the two extrapolation terms in* (3.3.10) *are calculated using the same Brownian path. This achieves the same strong rate of convergence, however the constant is lower, since the variance of* (3.3.10) *has a smaller constant.*

In Figure 3.3, we consider parameters from Table 3.2. The rate of convergence increases as expected for the higher order extrapolation; we observe that extrapolation using a weak Taylor scheme of order 2 is an improvement on higher-order Δ using a weak Taylor scheme of order 3 and $\psi_{s,2}$.

Remark 3.4.3. *Recall Remark 3.3.1(i). The optimal $\zeta = 1/7$ for the extrapolated Δ yields an MSE of $\mathcal{O}(N^{-6/7})$, with computational cost $\mathcal{O}(N^{8/7})$ and a theoretical slope of $-3/4$ for the* $\log - \log$ *plot of the MSE against the computational cost.*

It is natural to compare high-order approximations from the previous section to the Greeks using extrapolation: we compare Figure 3.1 and Figure 3.3. Comparing a weak Taylor 2 scheme with $\psi_{s,1}$, and an extrapolated Euler scheme with $\psi = 1$, the performance is similar. Upon comparing the weak Taylor 3 scheme with $\psi_{s,2}$ and an extrapolated weak Taylor 2 scheme with

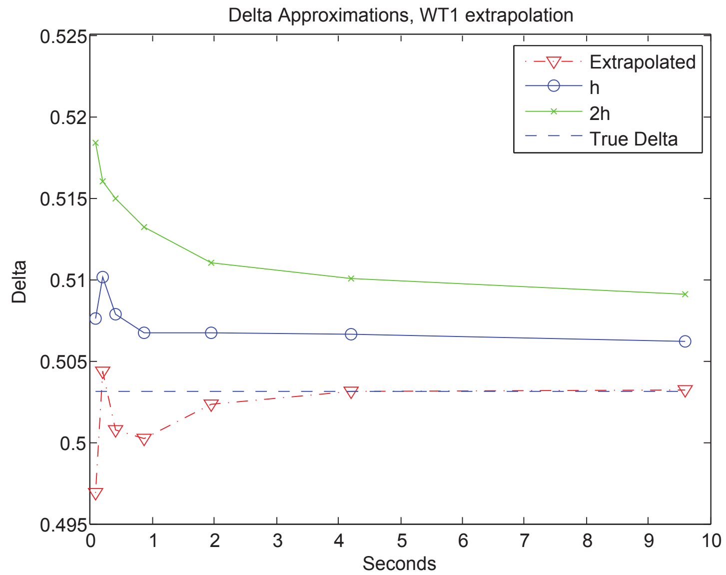

Figure 3.2: See Example 3.4.2(i). Δ values obtained against time in seconds for the extrapolated Δ, the value with stepsize h, $2h$ and the true Δ. Each run is repeated 100 times, with the number of Monte Carlo paths $N = 2^{14}, \ldots, 2^{20}$. Euler scheme extrapolation with $\zeta = 1/5$.

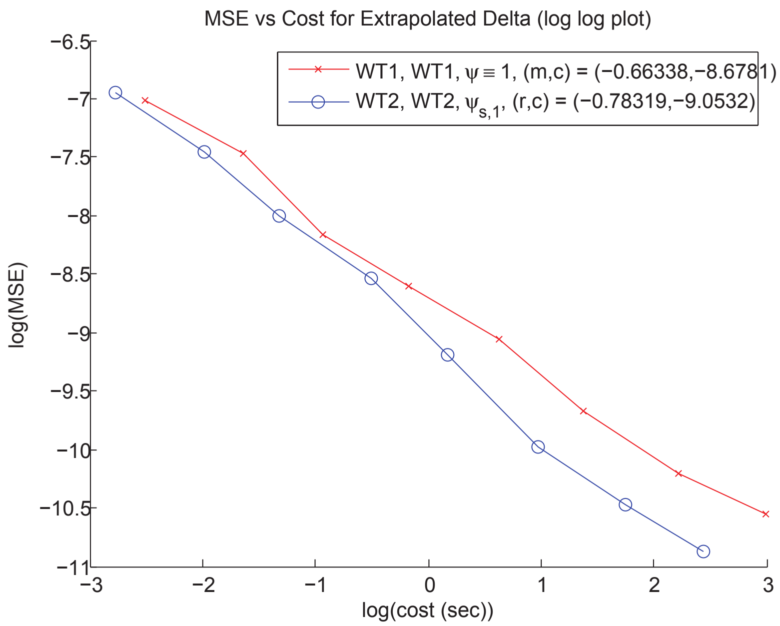

Figure 3.3: MSE vs Cost ($\log - \log$) for the extrapolated Δ using WT1 and WT2 schemes with $\psi \equiv 1$ and $\psi_{s,1}$, each run is repeated 100 times. Parameters as in Table 3.2.

$\psi_{s,1}$, then the extrapolated scheme achieves an improved constant, since the variance of the weight is much smaller. Furthermore, it is worth highlighting that the extrapolated scheme is easier to implement, and can be further parallelised.

3.4.3 Heston Greeks

We now apply the results of the previous sections even if the assumptions required for the proofs are not satisfied to perform numerics. Consider an asset price process $S = (S_t)_{t\geq 0}$, with $S_0 = x > 0$, defined on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P})$, assuming some constant interest rate $r \in \mathbb{R}$. In the Heston model, the variance process is modelled as a mean-reverting square-root diffusion stochastic process. The tuple (S_t, X_t) is the unique strong solution to

$$\begin{aligned} \mathrm{d}S_t &= rS_t\mathrm{d}t + \sqrt{X_t}S_t\mathrm{d}B_t^{(1)}\,, && S_0 = x > 0\,,\\ \mathrm{d}X_t &= \kappa(\theta - X_t)\mathrm{d}t + \xi\sqrt{X_t}\mathrm{d}B_t^{(2)}\,, && X_0 = v > 0\\ \mathrm{d}\langle B^{(1)}, B^{(2)}\rangle_t &= \rho\mathrm{d}t\,, && |\rho| \leq 1\,, \end{aligned} \tag{3.4.1}$$

with $\kappa, \theta, \xi > 0$, $B^{(1)} = (B^{(1)})_{t\geq 0}$ and $B^{(2)} = (B^{(2)})_{t\geq 0}$ being two correlated Brownian motions. If $2\kappa\theta \geq \xi^2$, then $\mathbb{P}(X_t = 0) = 0$, for all $t \geq 0$. This is referred to as the Feller condition, and when satisfied ensures that the origin is unattainable for the variance process (see [RW00]). We also mention that moments of the Heston model can explode depending on the parameters. For exponents $p > 1$, $\mathbb{E}[S_t^p]$ is finite for all $t > 0$ if and only if $\rho \leq \kappa/(\xi p) - \sqrt{(p-1)/p}$ [AP07]. The Heston model can be represented with independent Brownian motions $W^{(1)} = (W_t^{(1)})_{t\geq 0}$ and $W^{(2)} = (W_t^{(2)})_{t\geq 0}$ as

$$\mathrm{d}\begin{pmatrix} S_t \\ X_t \end{pmatrix} = \begin{pmatrix} rS_t \\ \kappa(\theta - X_t) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} \sqrt{1-\rho^2}\sqrt{X_t}S_t & \rho\sqrt{X_t}S_t \\ 0 & \xi\sqrt{X_t} \end{pmatrix}\begin{pmatrix} \mathrm{d}W_t^{(1)} \\ \mathrm{d}W_t^{(2)} \end{pmatrix}, \qquad \begin{aligned} S_0 &= x, \\ X_0 &= v. \end{aligned} \tag{3.4.2}$$

Consider now approximating the Delta under the Heston model. The difficulty in simulating the Heston model is the CIR volatility process as it can become negative using the Euler-Maruyama scheme; as a result, we consider several techniques for approximating the process. We consider an explicit Euler scheme, and a drift-implicit scheme [Alf13a]. Future work could be to consider a second-order discretisation scheme for the CIR process [Alf08]. By a suitable

Lamperti transform, we can obtain the log-Heston model [DNS12]

$$\begin{aligned} \mathrm{d}\log(S_t) &= (r - \tfrac{1}{2}Y_t^2)\mathrm{d}t + Y_t\left(\sqrt{1-\rho^2}\mathrm{d}W_t^{(1)} + \rho\mathrm{d}W_t^{(2)}\right), & S_0 = x \\ \mathrm{d}Y_t &= \left(\tfrac{4\kappa\theta - \xi^2}{8}\tfrac{1}{Y_t} - \tfrac{\kappa}{2}Y_t\right)\mathrm{d}t + \tfrac{\xi}{2}\mathrm{d}W_t^{(2)}, & Y_0 = \sqrt{v}, \end{aligned} \tag{3.4.3}$$

where $Y := \sqrt{X}$. Apply the Euler scheme for the log-price equation and the drift-implicit square root scheme for the volatility process to obtain $\bar{H}_{t_k}$, which is an approximation of $\log(S_{t_k})$ given by

$$\bar{H}_{t_k} = \log(x) + \sum_{l=0}^{k-1}\left(r - \frac{1}{2}\bar{Y}_{t_l}^2\right)\Delta t_{l+1} + \sum_{l=0}^{k-1}\bar{Y}_{t_l}\left(\sqrt{1-\rho^2}\Delta W_{l+1}^{(1)} + \rho\Delta W_{l+1}^{(2)}\right), \tag{3.4.4}$$

and $\bar{S}_{t_k} := \exp(\bar{H}_{t_k})$ is an approximation of the asset price at time t_k. We now consider several schemes for approximating the CIR process in the Heston model, in order to use the above discretisation scheme for the log-price.

Consider a terminal payoff function g, of the asset price, and suppose that the correlation parameter is set to zero, in a zero interest rate environment (i.e. $\rho \equiv 0$, $r \equiv 0$). Applying our previous results, a suggested scheme for the Δ in the Heston model is

$$\Delta = \mathbb{E}\left[g(X_T)\frac{\Delta W_h^{(1)}}{hx\sqrt{v}}\right] + \mathcal{O}(h).$$

Example 3.4.3 (Modified explicit Euler scheme). *One approach is to apply the modified explicit scheme, which is the next part of the thesis (see Part III). If the Feller condition $2\kappa\theta/\xi^2 > 1$ holds, then the transformed process $Y = \sqrt{X}$ is the unique strong solution to*

$$\mathrm{d}Y_t = f(Y_t)\mathrm{d}t + c\mathrm{d}W_t^{(2)}, \qquad Y_0 = \sqrt{v},$$

with drift function $f(x) \equiv a/x + bx$, $a := (4\kappa\theta - \xi^2)/8 > 0$, $b := -\kappa/2$ and $c := \xi/2$ from (3.4.3).

Example 3.4.4 (Drift Implicit scheme). *The drift-implicit Euler method can be written as*

$$\bar{Y}_{t_{k+1}} = \bar{Y}_{t_k} + f(\bar{Y}_{t_{k+1}})\Delta t_{k+1} + c\Delta W_{k+1}^{(2)}, \qquad \bar{Y}_0 = \sqrt{v}.$$

We can take the positive root of the quadratic equation, solving for $\bar{Y}_{t_{k+1}}$, to obtain the explicit solution

$$\bar{Y}_{t_{k+1}} = \frac{\bar{Y}_{t_k} + c\Delta W_{k+1}^{(2)}}{2(1-b\Delta t_{k+1})} + \sqrt{\frac{(\bar{Y}_{t_k} + c\Delta W_{k+1}^{(2)})^2}{4(1-b\Delta t_{k+1})^2} + \frac{a\Delta t_{k+1}}{1-b\Delta t_{k+1}}}, \qquad \bar{Y}_{t_0} = \sqrt{v}.$$

We can approximate the CIR process using $\bar{X}_{t_k} = \bar{Y}^2_{t_k}$*. For convergence of the modified Euler approximation to the* log-Heston price *in* (3.4.4)*, refer to [KN12, Corollary 5.5].*

Example 3.4.5 (Heston Call option)**.** *Consider the following parameters for the Heston model:* $(\kappa, \theta, \xi, r, \rho, x, v) = (1.15, 0.04, 0.2, 0, 0, 100, 0.04)$*. We consider a European Call option, with strike* $K = 100$*, and terminal time* $T = 1$*. The true price and semi-analytic* Δ *are computed to be* $(11.03, 0.555)$.

In Figure 3.4, we consider the explicit and drift-implicit approximations, with the parameters $\zeta = 1/3$, with $\psi \in \mathcal{B}^0_{[0,1]}$. Note that the schemes are quite similar, with comparable performance as expected; in fact the drift-implicit Euler has a slightly lower constant as it is more computationally intensive.

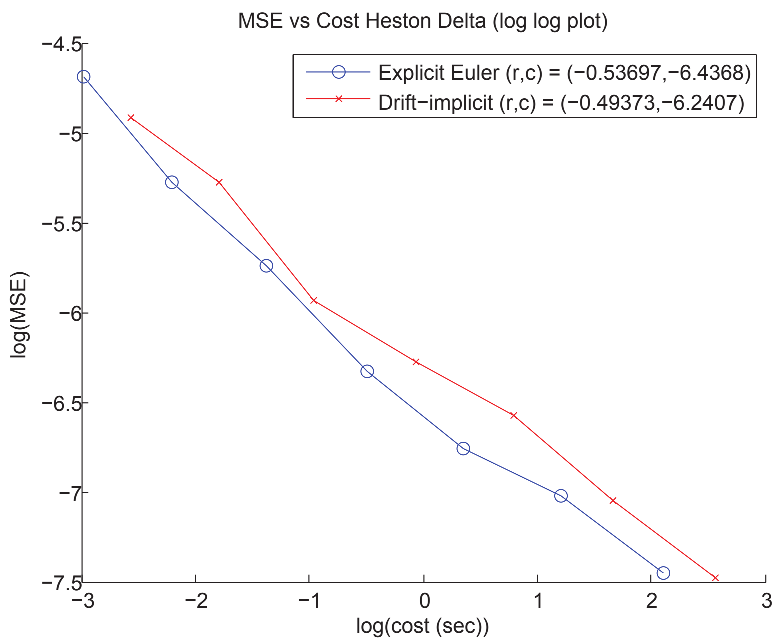

Figure 3.4: Heston model: MSE vs Cost $(\log - \log)$ in seconds for the Δ of the option in Example 3.4.5, 100 repeats, $\psi \equiv 1$, $\zeta = 1/3$. Explicit Euler scheme is from Example 3.4.3, Drift-implicit is from Example 3.4.4.

By following the same techniques, we can approximate the Heston Vega ($\mathcal{V} := \partial_v u(0, x)$) by

$$\mathcal{V} = \mathbb{E}\left[g(X_T) \frac{(H^\psi_h)_2}{\xi\sqrt{v}} \right] + \mathcal{O}(h),$$

using the second Brownian motion to define the weight—performance is similar hence omitted.

3.5 Discussion

In this chapter, we have shown two valid approaches for computing the Δ. The first uses specific weights which improve the order of bias; the second technique is based on the Romberg extrapolation technique. We have shown that $\mathbb{E}\Big[H_h^{\psi} g(\hat{X}_T)\Big]$ admits an expansion in terms of the equidistant step size h for weak Taylor schemes of varying order. Combined with a particular choice of ψ functions, we can use the ideas from the theoretical expansions from Chapter 2 to create higher order approximations, or further improve the Δ procedure by Romberg extrapolation.

The main measure of error used is the MSE compared to the runtime; it is seen that the extrapolation techniques obtain superior slope compared to just high-order techniques. Extrapolation of the Euler scheme is particularly appealing due to the fact that it is not necessary to compute potentially difficult derivatives of the drift and diffusion functions that are required for weak Taylor schemes of high-order. Furthermore, it is hard to justify schemes of extremely high orders, due to the implementation and the increasing variance constant of the weights. Another advantage of extrapolation is that it lends itself to natural parallelisation, so in a production environment one would divide the work effort across the two runs. Expansion methods for the Euler scheme allow Greek computation for general models without having to differentiate the drift and diffusion coefficients; this can really be a “black-box” in real-life applications.

It is important to consider the function ψ in tandem with the weak Taylor scheme used; using an inappropriate combination can increase the variance unnecessarily. There is the subtlety of the smoothness required of the value function; it would be interesting to consider examples of value functions for which $(\mathbf{H}u_b^l)$ holds, but $(\mathbf{H}u_b^{l+1})$ doesn’t, and then perform extrapolation or higher order schemes for which convergence cannot be justified theoretically.

We have also considered an example for the Heston model Δ using a modified explicit Euler scheme, and the more numerically demanding drift-implicit scheme, for which proving the results is more challenging.

4. Numerical Approximation of the Gamma

In this chapter, we extend the themes from Chapter 3 to approximate the Gamma of an option. This quantity in a financial setting is $\Gamma := \partial_{xx}u(0,x)$ at the initial time, the second derivative of the value function with respect to the initial spot price. In this chapter, we use numerical schemes to approximate expressions containing the Γ within them. A new set of functions are introduced to define suitable weights for these approximations.

4.1 High-order approximations

The aim now is to generalise the high order approximations for the Δ to approximate the Γ. We proceed with considering families of functions, which will be used to define weights Γ_h^ϕ (vector of length m) to approximate $u_0^{(j,j)}$ by

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j g(X_T)\Big] = \mathbb{E}\Big[(\Gamma_h^\phi)_j u(h, X_h)\Big] , \tag{4.1.1}$$

where we recall that $u_0^{(j,j)} = L^{(j,j)}u(0,x)$.

Definition 4.1.1 (ϕ-functions). *For $l \in \mathbb{N}^+$, define $\mathcal{K}_{[0,1]}^l$ as the set of bounded, measurable functions $\phi : [0,1] \to \mathbb{R}$ such that*

$$\int_0^1 \phi(s)s\mathrm{d}s = 1, \tag{4.1.2}$$

and if $l \geq 2$, then for all $k = 2, \ldots, l$,

$$\int_0^1 \phi(s)s^k\mathrm{d}s = 0. \tag{4.1.3}$$

We now define the general family of weights Γ_h^ϕ for $\phi \in \mathcal{K}_{[0,1]}^l$:

Definition 4.1.2 (Γ_h^ϕ-weights). *Let $\phi \in \mathcal{K}_{[0,1]}^l$, and for $0 < h \leq T$, define the row vector $\Gamma_{t,h}^\phi$ as*

$$(\Gamma_{t,h}^\phi)_j := \frac{1}{h^2}\int_t^{t+h} \phi\left(\frac{s-t}{h}\right) W_s^{(j)}\mathrm{d}W_s^{(j)} \text{ for } j = 1, \ldots, m,$$

and for shorthand $\Gamma_h^\phi := \Gamma_{0,h}^\phi$.

Example 4.1.1. *We motivate the family of functions defined with an example using $\phi \in \mathcal{K}^1_{[0,1]}$. Suppose a weak Taylor scheme of order 2 ($r = 2$) and that ($\mathbf{H}u^3_b$) holds. Following from* (4.1.1), $u_0^{(1,1)}$ *can be approximated using Itô-Taylor expansions using the hierarchical set $\mathcal{D}_2$ and the remainder set $\mathcal{B}(\mathcal{D}_2)$; the terms from the remainder set are bounded by $\mathcal{O}(h)$ from the smooth, bounded derivatives, and* $\mathbb{E}\Big[(\Gamma_h^{\phi})_j I_h^{(j,j)}\Big] = 1$ *from* (4.1.2). *This concludes that* $\mathbb{E}\Big[(\Gamma_h^{\phi})_j g(X_T)\Big] = u_0^{(j,j)} + \mathcal{O}(h)$.

Example 4.1.2. *Suppose $\phi \in \mathcal{K}^2_{[0,1]}$. By considering an Itô-Taylor expansion of $u(h, X_h)$ ($d = m = 1$ case) with a hierarchical set $\mathcal{D}_3$, observe that*

$$\mathbb{E}\Big[\Gamma_h^{\phi} u(h, X_h)\Big] = \mathbb{E}\Big[\Gamma_h^{\phi}\left(u_0^{(1,1)} I_h^{(1,1)} + u_0^{(0,1,1)} I_h^{(0,1,1)} + u_0^{(1,0,1)} I_h^{(1,0,1)}\right)\Big] + \sum_{\alpha \in \mathcal{B}(\mathcal{D}_3)} \mathbb{E}\Big[\Gamma_h^{\phi} I_h^{\alpha}[u^{\alpha}_{\cdot}]\Big].$$

Let ($\mathbf{H}u^4_b$), *and consider the various terms individually:*

(i) The first term is evaluated using the definition of Γ_h^{ϕ}, Itô's isometry, a change of variables and (4.1.2)*:*

$$\mathbb{E}\Big[\Gamma_h^{\phi} u_0^{(1,1)} I_h^{(1,1)}\Big] = u_0^{(1,1)} \frac{1}{h^2} \int_0^h \phi\left(\frac{s}{h}\right) s \mathrm{d}s = u_0^{(1,1)}. \tag{4.1.4}$$

(ii) The second term can be explicited using Itô's isometry and (4.1.3)*:*

$$\mathbb{E}\Big[\Gamma_h^{\phi} I_h^{(0,1,1)}\Big] = \frac{1}{h^2} \int_{s=0}^h \phi\left(\frac{s}{h}\right) \frac{s^2}{2} \mathrm{d}s = \frac{h}{2} \int_0^1 \phi(s) s^2 \mathrm{d}s = 0.$$

(iii) For the third term observe that $\int_{u=0}^s W_u \mathrm{d}u = \int_{u=0}^s (s-u) \mathrm{d}W_u$ by an integration by parts argument, therefore using (4.1.3) *we obtain*

$$\mathbb{E}\Big[\Gamma_h^{\phi} I_h^{(1,0,1)}\Big] = h \int_{s=0}^1 \phi(s) \int_{u=0}^s (s-u) \mathrm{d}u \mathrm{d}s = \frac{h}{2} \int_0^1 \phi(s) s^2 \mathrm{d}s = 0.$$

The term $u_0^{(1,1,0)}$ is equal to zero from $L^{(0)} u(\cdot, X_\cdot) = 0$. Combining (i)-(iii) and noting that $\sum_{\alpha \in \mathcal{B}(\mathcal{D}_3)} \mathbb{E}\Big[\Gamma_h^{\phi} I_h^{\alpha}[u^{\alpha}_{\cdot}]\Big] = \mathcal{O}(h^2)$, *it follows that* $\mathbb{E}\Big[(\Gamma_h^{\phi})_j g(X_T)\Big] = u_0^{(1,1)} + \mathcal{O}(h^2)$.

4.1.1 Weights and variance properties

We now consider a general result to show that the weights Γ_h^{ϕ} are suitable for approximating $u_0^{(j,j)}$ by considering higher-order terms in the Itô-Taylor expansion of $u(h, X_h)$:

Theorem 4.1.1. *Fix $l \in \mathbb{N}^+$. Suppose $(\mathbf{H}u_b^{l+2})$ holds and $\phi \in \mathcal{K}^l_{[0,1]}$. Then, for $h \in (0,T]$ and $j = 1, \dots, m$,*

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j g(X_T)\Big] = u_0^{(j,j)} + \mathcal{O}(h^l).$$

Proof. 1. We compute $\mathbb{E}\Big[\Gamma_h^\phi I_h^\alpha\Big]$ recalling $L^{(0)}u_\cdot = 0$, then it is sufficient to consider multi-indices such that $l(\alpha) = q$, $\alpha^+ = (j,j)$ and $2 \le q \le l+1$. Then, for every such multi-index, there exists $a \in \mathbb{N}^+$, such that $2 \le a \le q$ and the multi-index can be expressed as $\alpha = (0)_{a-2} * (j) * (0)_{q-a} * (j)$. For such multi-index, we have

$$\begin{aligned}\mathbb{E}\Big[(\Gamma_h^\phi)_j I_h^\alpha\Big] &= \frac{1}{h^2}\mathbb{E}\Bigg[\left(\int_{s=0}^h \phi(s/h) W_s^{(j)} \mathrm{d}W_s^{(j)}\right) I_h^{(0)_{a-2}*(j)*(0)_{q-a}*(j)}\Bigg] \\ &= \frac{1}{h^2}\int_0^h \phi(s/h)\mathbb{E}\Big[I_s^{(j)} I_s^{(0)_{a-2}*(j)*(0)_{q-a}}\Big]\,\mathrm{d}s.\end{aligned}$$

Since $k_0((j)) = k_1((j)) = 0$, and $k_0(\alpha) = a-2$ and $k_1(\alpha) = q-a$, it follows from Lemma 2.2.1 that

$$\mathbb{E}\Big[I_s^{(j)} I_s^{(0)_{a-2}*(j)*(0)_{q-a}}\Big] = \frac{s^{q-1}}{(q-1)!}.$$

Therefore, for $\alpha = (0)_{a-2} * (j) * (0)_{q-a} * (j)$,

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j I_h^\alpha\Big] = \frac{1}{(q-1)!h^2}\int_0^h \phi(s/h)s^{q-1}\mathrm{d}s = \frac{h^{q-2}}{(q-1)!}\int_0^1 \phi(s)s^{q-1}\mathrm{d}s = 0,$$

unless $q = a = 2$, which yields 1 as seen in (4.1.4).

2. Consider an Itô-Taylor expansion for $u(h, X_h)$ using the hierarchical set $\mathcal{D}_{l+1}$, and remainder set $\mathcal{B}(\mathcal{D}_{l+1})$. The only non-zero expectation terms are those with multi-indices α such that $\alpha^+ = (j,j)$; therefore, α is again of the form $\alpha = (0)_{a-2} * (j) * (0)_{l-a} * (j)$, for $q = l+2$. Recalling (2.2.3), observe that $k_0(\alpha) = a-2$, $k_1(\alpha) = l+2-a$, $k_2(\alpha) = 0$, and $k_0((j,j)) = k_1((j,j)) = k_2((j,j)) = 0$, leading to $w((j,j),\alpha) = l+2$ for all $\alpha \in \mathcal{B}(\mathcal{D}_{l+1})$. From the regularity $(\mathbf{H}u_b^{l+2})$, it follows that $\sum_{\alpha\in\mathcal{B}(\mathcal{D}_{l+1})} \mathbb{E}\Big[(\Gamma_h^\phi)_j I_h^\alpha[u^\alpha_\cdot]\Big] = \mathcal{O}(h^l)$. □

We now consider various functions $\phi \in \mathcal{K}^l_{[0,1]}$, and again categorise them in polynomials $\phi_{p,l}$ and step functions $\phi_{s,l}$.

Polynomial functions $\phi_{p,l} \in \mathcal{K}^l_{[0,1]}$

We now derive the polynomials that belong to $\mathcal{K}^l_{[0,1]}$ by simultaneously solving equations from the conditions imposed on $\phi_{p,l}$ by (4.1.2) and (4.1.3):

Lemma 4.1.1. *Suppose $m = 1$.*

1. *$\phi_{p,1} \equiv 2 \equiv \psi_{s,1}$, belongs to $\mathcal{K}^1_{[0,1]}$. The weight defined using $\phi_{p,1}$ has variance $\mathbb{V}[\Gamma_h^{\phi_{p,1}}] = \frac{2}{h^2}$.*

2. *$\phi_{p,2}(s) \equiv 18 - 24s$ belongs to $\mathcal{K}^2_{[0,1]}$.*

3. *$\phi_{p,3}(s) \equiv 72 - 240s + 180s^2$ belongs to $\mathcal{K}^3_{[0,1]}$.*

Remark 4.1.1. *We can easily simulate $\Gamma_h^{\phi_{p,1}}$; however for $\phi_{p,l}$ for $l \geq 2$, we require terms $\int_0^h s^k W_s \mathrm{d}W_s$ for each $k = 2, \ldots, l$, which are generally difficult to simulate.*

Step functions $\phi_{s,l} \in \mathcal{K}^l_{[0,1]}$

Weights defined using step functions are easier to simulate. Therefore, we explicit $\Gamma_h^{\phi_{s,2}}$, for some fixed $c \in (0,1)$:

Lemma 4.1.2. *Function $\phi_{s,2}(u) \equiv \frac{-2}{c(c-1)^2} 1_{[1-c,1]}(u) + (2 - \frac{2}{c-1} + \frac{2}{(c-1)^2})$ is a bounded, measurable step function for any $c \in (0,1)$ and belongs to $\mathcal{K}^2_{[0,1]}$. Furthermore, the minimum variance for the weight $\Gamma_h^{\phi_{s,2}}$ is attained when $c = 3/2 - \sqrt{5}/2$, independently of h.*

Proof. We choose a step function $\phi_{s,2} : [0,1] \to \mathbb{R}$, with one step at point $c \in (0,1)$. From the properties of $\phi_{s,2}$, we require $\int_0^1 \phi_{s,2}(u) u \mathrm{d}u = 1$ and $\int_0^1 \phi_{s,2}(u) u^2 \mathrm{d}u = 0$. By forming the simultaneous equations

$$A \int_{(1-c)}^1 u \mathrm{d}u + B \int_0^1 u \mathrm{d}u = 1, \qquad A \int_{(1-c)}^1 u^2 \mathrm{d}u + B \int_0^1 u^2 \mathrm{d}u = 0,$$

it follows that

$$\phi_{s,2}(u) \equiv \frac{-2}{c(c-1)^2} 1_{[1-c,1]}(u) + \left(2 - \frac{2}{c-1} + \frac{2}{(c-1)^2}\right).$$

We explicit the weight $\Gamma_h^{\phi_{s,2}}$ using Definition 4.1.2 as

$$\begin{aligned}
\Gamma_h^{\phi_{s,2}} &= \tfrac{1}{h^2} \int_0^h \phi_{s,2}(s/h) W_s \mathrm{d}W_s \\
&= \tfrac{1}{h^2} \int_0^h \left(\tfrac{-2}{c(c-1)^2} 1_{[1-c,1]}(s/h) W_s + (2 - \tfrac{2}{c-1} + \tfrac{2}{(c-1)^2}) 1_{[0,1]}(s/h) W_s \right) \mathrm{d}W_s \\
&= \tfrac{-1}{h^2 c(c-1)^2} \left[(W_h^2 - h) - \left(W_{h(1-c)}^2 - h(1-c) \right) \right] + \left(1 - \tfrac{1}{c-1} + \tfrac{1}{(c-1)^2} \right) \tfrac{(W_h^2 - h)}{h^2} \\
&= \tfrac{(c-1)}{ch^2} \left(W_h^2 - h \right) + \tfrac{1}{h^2 c(c-1)^2} \left(W_{h(1-c)}^2 - h(1-c) \right).
\end{aligned}$$

The weight has mean zero, so to compute the variance, we square and take expectations; simplifying the resulting expression using

$$\mathbb{E}\Big[W_h^4\Big] = 3h^2 \qquad \text{and} \qquad \mathbb{E}\Big[W_h^2 W_{(1-c)h}^2\Big] = h^2(1-c)(3-2c) \text{ for } c \in (0,1).$$

Differentiating the variance of the weight with respect to c, yields a minimum variance attained at $c = 3/2 - \sqrt{5}/2 \approx 0.382$, independently of h. The minimum variance of $\Gamma_h^{\phi_{s,2}}$ is hence given by

$$\mathbb{V}(\Gamma_h^{\phi_{s,2}}) = \frac{9404\sqrt{5} - 21028}{h^2(3\sqrt{5}-7)(21\sqrt{5}-47)(4\sqrt{5}-9)} \approx \frac{24.2}{h^2}.$$

□

For the step function $\phi_{s,3}$ with steps at distinct points $c, c' \in (0,1)$, we consider three simultaneous equations from the definition of the function $\phi \in \mathcal{K}^3_{[0,1]}$. Their solution yields

$$\phi_{s,3}(u) \equiv s_1 1_{[1-c,1]}(u) + s_2 1_{[1-c',1]}(u) + s_3,$$

where

$$s_1 := -2\,\frac{1}{c\,(c-1)^2} - 2\frac{1}{c\,(c-1)\,(c'-c)},$$

$$s_2 := \frac{-2c+2}{cc'} + 2\frac{1}{c\,(c-1)\,(c'-c)} + \frac{2c-4}{(c-1)\,(c'-1)} - 2\left(c'-1\right)^{-2}$$

and

$$s_3 := 2 + 2\,(c-1)^{-2} - 2\,(c-1)^{-1} + \frac{-2c+4}{(c-1)\,(c'-1)} + 2\left(c'-1\right)^{-2}.$$

Lemma 4.1.3. *The weight defined using the step function $\phi_{s,3}$ attains its minimal variance at $c = 0.676$, $c' = 0.104$ independently of h, and $\mathbb{V}[\Gamma_h^{\phi_{s,3}}] = 95.7/h^2$.*

The proof is omitted as it follows the same argument as in the proof of Lemma 4.1.2.

4.1.2 Approximating the Γ using the Euler scheme

We now discretise the process using an Euler scheme, and consider approximating the Γ. The next lemma will be required for the main result in this section:

Lemma 4.1.4. *Suppose an Euler scheme, $\phi \equiv 2$, and $(\mathbf{H}u_b^2)$ holds. Then,*

$$\mathbb{E}\Big[(\Gamma_h^{\phi})_j u(h, \hat{X}_h)\Big] = \hat{L}_x^{(j,j)} u(0,x) + \mathcal{O}(h) = \gamma^2 \partial_{xx} u(0,x) + \mathcal{O}(h);$$

assuming $(\mathbf{H}u_b^3)$, *we have*

$$\mathbb{E}\Big[(\Gamma_h^{\phi})_j u(h,\hat{X}_h)\Big] = \hat{L}_x^{(j,j)}u(0,x) + \hat{L}_x^{(j,j,0)}u(0,x)h + \mathcal{O}(h^2).$$

Proof. We show the proof for $d=m=1$, which extends naturally. Using $\hat{X}_h = x + f(x)h + \gamma(x)\Delta W$, perform a Taylor expansion on $u(h,\hat{X}_h)$ using the multivariate Taylor theorem around $(0,x)$, observing that when $\phi \equiv 2$, then $\Gamma_h^{\phi} = \left((I_h^{(1)})^2 - h\right)/h^2$:

$$\begin{aligned}\mathbb{E}\Big[\Gamma_h^{\phi}u(h,\hat{X}_h)\Big] =& \mathbb{E}\Big[\Gamma_h^{\phi}u(0,x)\Big] + \mathbb{E}\Bigg[\left(\frac{(\Delta W)^2-h}{h^2}\right)\left((\hat{X}_h-x)\partial_x u(0,x) + h\partial_t u(0,x)\right)\Bigg]\\ &+\mathbb{E}\Bigg[\left(\frac{(\Delta W)^2-h}{h^2}\right)\frac{(\hat{X}_h-x)^2\partial_{xx}u(0,x) + 2h(\hat{X}_h-x)\partial_{tx}u(0,x) + h^2\partial_{tt}u(0,x)}{2!}\Bigg]\\ &+\ldots,\end{aligned}$$

where $\Delta W := \int_0^h \mathrm{d}W_s = I_h^{(1)}$, and $\Delta Z := \int_0^h W_s \mathrm{d}s$. We now consider the terms individually, starting from $\mathbb{E}\Big[\Gamma_h^{\phi}u(0,x)\Big] = 0$.

1. For the first part, we expand up to $\mathcal{O}(h)$ terms. Furthermore, observe that the $\partial_{xx}u(0,x)$ terms are

$$\begin{aligned}\mathbb{E}\Bigg[\Gamma_h^{\phi}\frac{(\hat{X}_h-x)^2}{2!}\Bigg] =& \mathbb{E}\Bigg[\frac{\left(I_h^{(1)}\right)^2-h}{h^2}\left(f(x)h+\gamma(x)I_h^{(1)}\right)^2\Bigg]\\ =&\mathbb{E}\Bigg[\frac{1}{2}\gamma^2\left(I_h^{(1)}\right)^4 - \frac{1}{2}\gamma^2\left(I_h^{(1)}\right)^2 + \left\{f\gamma\left(I_h^{(1)}\right)^3 - f\gamma I_h^{(1)}\right\}\sqrt{h}\Bigg]\\ &+\mathbb{E}\Bigg[\left\{\frac{1}{2}f^2\left(I_h^{(1)}\right)^2 - \frac{1}{2}f^2\right\}h\Bigg] = \frac{3}{2}\gamma(x)^2 - \frac{1}{2}\gamma(x)^2 = \gamma^2(x),\end{aligned}$$

using the properties of the Brownian motion. We now consider the term containing $\partial_{xxx}u(0,x)$:

$$\begin{aligned}\mathbb{E}\Bigg[\Gamma_h^{\phi}\frac{(\hat{X}_h-x)^3}{3!}\Bigg] =& \mathbb{E}\Bigg[\frac{1}{6}\gamma^3\left(\left(I_h^{(1)}\right)^4 - \frac{1}{2}\gamma^2\left(I_h^{(1)}\right)^5 - \left(I_h^{(1)}\right)^3\right) + \left\{f\gamma\left(I_h^{(1)}\right)^3 - f\gamma I_h^{(1)}\right\}\sqrt{h}\Bigg]\\ &+\mathbb{E}\Bigg[\frac{1}{2}f\gamma^2\left\{\left(I_h^{(1)}\right)^4 - \left(I_h^{(1)}\right)^2\right\}h\Bigg] + \mathbb{E}\Bigg[\frac{1}{2}f\gamma^2\left\{\left(I_h^{(1)}\right)^4 - \left(I_h^{(1)}\right)^2\right\}h^{3/2}\Bigg] + \mathcal{O}(h^2)\\ =& f(x)\gamma(x)^2 h + \mathcal{O}(h^2).\end{aligned}$$

For the $\partial_{xxxx}u(0,x)$ terms, we have

$$\begin{aligned}\mathbb{E}\left[\Gamma_h^{\phi}\frac{(\hat{X}_h-x)^4}{4!}\right] =&\frac{15}{24}\gamma^4 h+3\left(\frac{1}{4}f^2\gamma^2h^2-\frac{1}{24}\gamma^4h\right)+\left(\frac{1}{24}f^4h^3-\frac{1}{4}f^2\gamma^2h^2\right)-\frac{1}{24}f^4h^3\\ =&\frac{1}{2}\gamma(x)^4h+\frac{1}{2}f(x)^2\gamma(x)^2h^2.\end{aligned}$$

Now, we consider several of the cross terms containing $\partial_{tx}u(0,x)$:

$$\mathbb{E}\left[\Gamma_h^{\phi}\frac{2h(\hat{X}_h-x)}{2!}\right]=\mathbb{E}\left[f(x)\left(I_h^{(1)}\right)^2h-f(x)h+\gamma(x)\sqrt{h}\left(I_h^{(1)}\right)^3-\gamma(x)\sqrt{h}I_h^{(1)}\right]=0.$$

For terms relating to $\partial_{txx}u(0,x)$,

$$\begin{aligned}\mathbb{E}\left[\Gamma_h^{\phi}\frac{3h(\hat{X}_h-x)^2}{3!}\right]=&\mathbb{E}\left[\frac{1}{2}\gamma^2\left\{\left(I_h^{(1)}\right)^4-\left(I_h^{(1)}\right)^2\right\}h+f(x)\gamma(x)\left\{\left(I_h^{(1)}\right)^3-I_h^{(1)}\right\}h^{3/2}+\mathcal{O}(h^2)\right]\\ =&\gamma(x)^2h+\mathcal{O}(h^2),\end{aligned}$$

and similarly we can check that the $\partial_{ttx}u(0,x)$ terms are $\mathcal{O}(h^2)$. For higher-order terms, we have that terms such as $\hat{L}_x^{\alpha}u(0,x)$ are continuous and bounded by assumption. Collecting the terms up to $\mathcal{O}(h)$ proves the first part.

2. For the second result, collect the terms until $\mathcal{O}(h^2)$, with the additional smoothness in the value function. We continue to Taylor expand the value function, and observe that the $\partial_{xxxxx}u(0,x)$ terms are

$$\mathbb{E}\left[\Gamma_h^{\phi}\frac{(\hat{X}_h-x)^5}{5!}\right]=\mathbb{E}\left[\frac{1}{120}\gamma^5\left\{\left(I_h^{(1)}\right)^7-\left(I_h^{(1)}\right)^5\right\}h^{3/2}\right]+\mathcal{O}(h^2)=\mathcal{O}(h^2);$$

and similarly for higher order terms, we can check that $\mathbb{E}\left[\Gamma_h^{\phi}\frac{(\hat{X}_h-x)^k}{k!}\right]=\mathcal{O}(h^2)$ for $k\geq 5$.

$$\begin{aligned}\mathbb{E}\left[\Gamma_h^{\phi}u(h,\hat{X}_h)\right]&=\left(\gamma(x)^2\partial_{xx}+\{\gamma^2\partial_{txx}+f\gamma^2\partial_{xxx}+\frac{1}{2}\gamma^4\partial_{xxxx}\}h+\mathcal{O}(h^2)\right)u(0,x)\\ &=\hat{L}_x^{(1,1)}u(0,x)+\hat{L}_x^{(1,1,0)}u(0,x)h+\mathcal{O}(h^2).\end{aligned}$$

□

Remark 4.1.2.

(i) An alternative proof to Lemma 4.1.4 is to use the Euler scheme and an extension to Lemma 3.2.1.

(ii) The above proof can be shown for general $\phi\in\mathcal{K}^1_{[0,1]}$—we pick $\phi\equiv 2$ as it defines a weight with

the smallest variance.

We now conclude the result for approximating the Γ using an Euler scheme:

Theorem 4.1.2. *Suppose* $(\mathbf{H}u_b^4)$ *and* $\phi \equiv 2$. *Using an Euler scheme, then*

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j g(\hat{X}_T)\Big] = \hat{L}_x^{(j,j)} u(0,x) + \mathcal{O}(h).$$

Proof. This proof follows essentially the same steps as in the proof of Theorem 3.2.1. By a telescoping sum and the first part of the previous lemma,

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j g(\hat{X}_T)\Big] = \hat{L}_x^{(j,j)} u(0,x) + \mathcal{O}(h) + \mathbb{E}\left[(\Gamma_h^\phi)_j \sum_{i=1}^{n-1} \left(u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i})\right)\right].$$

Applying Ito's Formula, we compute

$$\begin{aligned} \mathbb{E}_{t_i}\big[u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i})\big] &= \mathbb{E}_{t_i}\left[\int_{t_i}^{t_{i+1}} \hat{L}_{\hat{X}_{t_i}}^{(0)} u(s, \hat{X}_s)\mathrm{d}s\right] \\ &= h\hat{L}_{\hat{X}_{t_i}}^{(0)} u(t_i, \hat{X}_{t_i}) + \mathbb{E}_{t_i}\left[\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \hat{L}_{\hat{X}_{t_i}}^{(0,0)} u(r, \hat{X}_r)\mathrm{d}r\mathrm{d}s\right] \\ &= \frac{h^2}{2}\hat{L}_{\hat{X}_{t_i}}^{(0,0)} u(t_i, \hat{X}_{t_i}) + \mathcal{O}(h^3), \end{aligned} \tag{4.1.5}$$

where to get the last equality we used also the fact that $\hat{L}_{\hat{X}_{t_i}}^{(0)} u(t_i, \hat{X}_{t_i}) = 0$, and the boundedness of the derivatives of the value function. For $(s,y) \in [0,T] \times \mathbb{R}^d$, define $\phi_e^1(s,y) := \frac{1}{2}\hat{L}_y^{(0,0)} u(s,y)$. With this notation, we obtain,

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j \{g(\hat{X}_T) - u(h, \hat{X}_h)\}\Big] = \mathbb{E}\left[(\Gamma_h^\phi)_j \sum_{i=1}^{n-1} \left(\mathbb{E}_{t_1}\Big[h^2\phi_e^1(t_i, \hat{X}_{t_i})\Big] + \mathcal{O}(h^3)\right)\right].$$

From the smoothness of ϕ_e^1, then $\mathbb{E}_{t_1}[\phi_e^1(t_i, \hat{X}_{t_i})] = \mathbb{E}_{t_1}\left[\phi_e^1(t_i, X_{t_i}^{t_1, \hat{X}_{t_1}})\right] + \mathcal{O}(h)$. Combining these equalities, we obtain

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j \{g(\hat{X}_T) - u(h, \hat{X}_h)\}\Big] = \mathcal{O}(h) + \mathbb{E}\left[(\Gamma_h^\phi)_j \left(\textstyle\sum_{i=1}^{n-1} \mathbb{E}_{t_1}\left[h^2\phi_e^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\right]\right)\right], \tag{4.1.6}$$

using the Cauchy-Schwarz inequality and the variance of weight $(\Gamma_h^\phi)_j$. We observe that,

$$\sum_{i=1}^{n-1} h\mathbb{E}_{t_1}\left[\phi_e^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\right] = \varphi_e^1(t_1, \hat{X}_{t_1}) + \mathcal{O}(h),$$

where for any $(t,x) \in [0,T] \times \mathbb{R}^d$, $\varphi_e^1(t,x) := \mathbb{E}_t\Big[\int_t^T \phi_e^1(s, X_s^{t,x})\mathrm{d}s\Big]$, noting that $\varphi_e^1 \in \mathcal{G}_b^2$.
We compute

$$h\mathbb{E}\left[(\Gamma_h^{\phi})_j h \sum_{i=1}^{n-1} \phi_e^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\right] = h\hat{L}_x^{(j,j)}\varphi_e^1(0,x) + \mathcal{O}(h^2);$$

therefore

$$\mathbb{E}\Big[(\Gamma_h^{\phi})_j g(\hat{X}_T)\Big] = \hat{L}_x^{(j,j)} u(0,x) + \mathcal{O}(h).$$

□

4.1.3 High-order expansion for Γ

We now pursue high-order approximations, which shall lead to approximations of $u_0^{(1,1)}$, as opposed to $\hat{L}_x^{(1,1)} u(0,x)$ using the approximations from the previous section. In Chapter 3, we applied weak Taylor schemes of order r used for approximating the Δ. The aim is to extend these results for the Γ—we begin with the analogues result to Lemma 3.2.3:

Lemma 4.1.5. *Fix $l \in \mathbb{N}^+$. Suppose that $(\mathbf{H}u_b^{l+2})$ holds for a value function and $L^{(0)}u_. = 0$, $\phi \in \mathcal{K}_{[0,1]}^l$, and suppose that the order of the weak Taylor scheme is $l+1$. Then,*

$$\mathbb{E}_{t_i}\left[\Gamma_{t_i,h}^{\phi} u(t_{i+1}, X_{t_{i+1}}^{t_i,\hat{X}_{t_i}})\right] = \mathbb{E}_{t_i}\Big[\Gamma_{t_i,h}^{\phi} u(t_{i+1}, \hat{X}_{t_{i+1}})\Big] + \mathcal{O}(h^l);$$

in particular, $\mathbb{E}\Big[\Gamma_h^{\phi} u(h, X_h)\Big] = \mathbb{E}\Big[\Gamma_h^{\phi} u(h, \hat{X}_h)\Big] + \mathcal{O}(h)$, when $\phi \equiv 2$, $(\mathbf{H}u_b^3)$ and a weak Taylor scheme of order 2 is used.

Proof. We show the proof in the $d = m = 1$ case, and only in the case of a weak Taylor 2 scheme.
i) We first show the result for a weak Taylor scheme of order 2 (without drift), for $\phi \in \mathcal{K}_{[0,1]}^1$. The proof is analogues to Lemma 3.3.2, but using weights for the Γ. Recall that

$$\hat{X}_h := x + \gamma(x) I_h^{(1)} + \frac{1}{2}\gamma\gamma'\left\{\left(I_h^{(1)}\right)^2 - h\right\} + \frac{1}{2}\gamma^2\gamma'' I_h^{(0,1)},$$

where

$$I_h^{(1)} = \int_0^h \mathrm{d}W_s, \qquad I_h^{(0,1)} = \int_0^h \int_0^s \mathrm{d}v \mathrm{d}W_s.$$

Now, multiply the Taylor expansion of $u(h, \hat{X}_h)$ by Γ_h^ϕ and take expectations:

$$\begin{aligned}\mathbb{E}\Big[\Gamma_h^\phi u(h,\hat{X}_h)\Big] = \;& \mathbb{E}\Big[\Gamma_h^\phi u(0,x)\Big] + \mathbb{E}\Big[\Gamma_h^\phi \left((\hat{X}_h - x)\partial_x u(0,x) + h\partial_t u(0,x)\right)\Big] \\ &+\mathbb{E}\Big[\Gamma_h^\phi \tfrac{(\hat{X}_h-x)^2\partial_{xx}u(0,x)+2h(\hat{X}_h-x)\partial_{tx}u(0,x)+h^2\partial_{tt}u(0,x)}{2!}\Big] \\ &+\ldots,\end{aligned} \tag{4.1.7}$$

and we explicit the individual terms, in this case up to $\mathcal{O}(h)$, with

$$\Gamma_h^\phi = \frac{1}{h^2}\int_0^h \phi(s/h)W_s\mathrm{d}W_s.$$

For the first term, $\mathbb{E}\Big[\Gamma_h^\phi u(0,x)\Big] = 0$. The terms $\partial_x u(0,x)$ are

$$\begin{aligned}\mathbb{E}\Big[\Gamma_h^\phi(\hat{X}_h - x)\Big] &= \mathbb{E}\Big[\Gamma_h^\phi\left(\gamma I_h^{(1)} + \frac{1}{2}\gamma\gamma'\Big\{\left(I_h^{(1)}\right)^2 - h\Big\} + \frac{1}{2}\gamma''\gamma^2\int_0^h s\mathrm{d}W_s\right)\Big] \\ &= \frac{1}{2}\gamma\gamma'\mathbb{E}\Big[\frac{2}{h^2}\int_0^h \phi(s/h)s\mathrm{d}s\Big] = \gamma\gamma',\end{aligned}$$

using a change of variables and (4.1.2). For the $\partial_t u(0,x)$ and $\partial_{tt}u(0,x)$ terms, we have $\mathbb{E}\Big[\Gamma_h^\phi h\Big] = \mathbb{E}\Big[\Gamma_h^\phi h^2\Big] = 0$. For the terms containing $\partial_{xx}u(0,x)$, we obtain

$$\begin{aligned}\mathbb{E}\Big[\Gamma_h^\phi\frac{(\hat{X}_h - x)^2}{2}\Big] =&\mathbb{E}\Big[\Gamma_h^\phi\left(\frac{1}{8}\gamma^2\gamma''\left(I_h^{(1)}\right)^4 + \frac{1}{2}\gamma^2\gamma'\left(I_h^{(1)}\right)^3\right)\Big] \\ &+\mathbb{E}\Big[\Gamma_h^\phi\left(\frac{1}{2}\gamma^2 - \frac{1}{4}\gamma^2(\gamma')^2 h + \frac{1}{4}\gamma^3\gamma'\gamma'' I_h^{(0,1)}\right)\left(I_h^{(1)}\right)^2\Big] \\ &+\mathbb{E}\Big[\Gamma_h^\phi\left(-\frac{1}{2}\gamma^2\gamma' h + \frac{1}{2}\gamma^3\gamma'' I_h^{(0,1)}\right)I_h^{(1)}\Big] \\ &+\mathbb{E}\Big[\Gamma_h^\phi\Big\{\frac{1}{8}\gamma^2(\gamma')^2h^2 - \frac{1}{4}\gamma^3\gamma'\gamma'' h I_h^{(0,1)} + \frac{1}{8}\gamma^4(\gamma'')^2\left(I_h^{(0,1)}\right)^2\Big\}\Big] \\ =&\mathbb{E}\Big[\frac{\gamma^2}{2h^2}\left(3\int_0^h\phi(s/h)s\mathrm{d}s - \int_0^h \phi(s/h)s\mathrm{d}s\right)\Big] + \mathcal{O}(h) \\ =&\gamma^2\int_0^1\phi(s)s\mathrm{d}s + \mathcal{O}(h) = \gamma^2 + \mathcal{O}(h).\end{aligned}$$

Higher order and cross terms can be dealt with in a similar manner. For the remainder terms, observe that since $(\mathbf{H}u_b^3)$, then

$$\mathbb{E}\Big[\Gamma_h^\phi u(h,\hat{X}_h)\Big] = L^{(1,1)}u(0,x) + \mathcal{O}(h).$$

ii) For the second result, consider a weak Taylor order 3 scheme, with $\phi \in \mathcal{K}^2_{[0,1]}$; using the property $\int_0^1 \phi(s)s^2 ds = 0$, the remainder terms are of $\mathcal{O}(h^2)$. For this we require an additional level of smoothness in the value function; i.e. $(\mathbf{H}u_b^4)$. □

We now prove convergence for the weak Taylor scheme with $r = 2$:

Theorem 4.1.3. *Suppose that* $(\mathbf{H}u_b^3)$ *holds,* $\phi \in \mathcal{K}^1_{[0,1]}$*, and suppose a weak Taylor scheme of order* 2*, on an equidistant time grid* π *such that* $|\pi| = h$*. Then,*

$$u_0^{(j,j)} = \mathbb{E}\Big[(\Gamma_h^\phi)_j g(\hat{X}_T)\Big] + \mathcal{O}(h).$$

Proof. We begin by fixing the equidistant time grid π with n points of size h.
i) By a telescoping sum it follows that $\mathbb{E}\Big[(\Gamma_h^\phi)_j g(\hat{X}_T)\Big]$ can be expressed as

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j u(t_n, \hat{X}_{t_n})\Big] = \mathbb{E}\left[(\Gamma_h^\phi)_j \sum_{i=1}^{n-1} \{u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i})\}\right] + \mathbb{E}\Big[(\Gamma_h^\phi)_j u(t_1, \hat{X}_{t_1})\Big], \tag{4.1.8}$$

and from Lemma 4.1.5 we note that $\mathbb{E}\Big[(\Gamma_h^\phi)_j u(h, \hat{X}_h)\Big] = u_0^{(j,j)} + \mathcal{O}(h)$, where $h := t_1$.
ii) It is left to deal with the telescoping series. Consider

$$\begin{aligned} u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i}) &= \textstyle\int_{t_i}^{t_{i+1}} \hat{L}^{(0),2}_{\hat{X}_{t_i}} u(s, \hat{X}_s) ds + \sum_{j=1}^m \int_{t_i}^{t_{i+1}} \hat{L}^{(j),2}_{\hat{X}_{t_i}} u(s, \hat{X}_s) dW_s^{(j)} \\ &= h\hat{L}^{(0),2}_{\hat{X}_{t_i}} u(t_i, \hat{X}_{t_i}) + \textstyle\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \hat{L}^{(0,0),2}_{\hat{X}_{t_i}} u(r, \hat{X}_r) dr ds + R, \end{aligned}$$

and observe that since we have used a weak Taylor 2 scheme,

$$u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i}) = \frac{h^2}{2} \hat{L}^{(0,0),2}_{\hat{X}_{t_i}} u(t_i, \hat{X}_{t_i}) + \mathcal{O}(h^3) + R; \tag{4.1.9}$$

R are terms that have conditional expectation equal to zero given the filtration $\mathcal{F}_{t_i}$, and for a weak Taylor scheme of order 2, $\hat{L}^{(0,0),2}_{\hat{X}_{t_i}} u(t_i, \hat{X}_{t_i}) = L^{(0,0)} u(t_i, \hat{X}_{t_i}) = 0$ from Lemma 3.2.2. From this, we can conclude by summation that

$$\mathbb{E}\left[(\Gamma_h^\phi)_j \sum_{i=1}^{n-1} \{u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i})\}\right] \le C\sqrt{\frac{h^4}{h^2}} = \mathcal{O}(h),$$

from the Cauchy-Schwarz inequality, $(\mathbf{H}u_b^3)$, and observing that $\|\Gamma_h^\phi\|_2 = \sqrt{2}/h$. Therefore,

$$\mathbb{E}\Big[(\Gamma_h^\phi)_j u(t_n, \hat{X}_{t_n})\Big] = \mathbb{E}\left[(\Gamma_h^\phi)_j \sum_{i=1}^{n-1} \{u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i})\}\right] + \mathbb{E}\Big[(\Gamma_h^\phi)_j u(h, \hat{X}_h)\Big] = u_0^{(j,j)} + \mathcal{O}(h),$$

which concludes the proof. $\square$

Remark 4.1.3 (Simplified weak Taylor schemes)**.** *For higher-order schemes, it could be advantageous to consider simplified weak Taylor schemes; for an Euler scheme we replace the* $\int_{t_i}^{t_{i+1}} \mathrm{d}W_s$ *components by simple expressions such as the random variable* $\Delta\hat{W}_{i+1}$*, where* $h_{i+1} = t_{i+1} - t_i$ *and* $\mathbb{P}(\Delta\hat{W}_{i+1} = \pm\sqrt{h_{i+1}}) = 1/2$*. For a weak Taylor scheme of order* 2*, also replace* $\int_{t_i}^{t_{i+1}} W_s \mathrm{d}s$ *by* $\frac{1}{2}\Delta\hat{W}_{i+1}h_{i+1}$*, with*

$$\mathbb{P}(\Delta\hat{W}_{i+1} = \pm\sqrt{3h_{i+1}}) = 1/6 \qquad \textit{and} \qquad \mathbb{P}(\Delta\hat{W}_{i+1} = 0) = 2/3.$$

The use of such simplified schemes enables the techniques to be implemented in a deterministic manner as a binomial/trinomial lattice; if there is no "recombination" of the tree, then the computational cost grows exponentially.

We now state a more general result for higher-order approximations of $u_0^{(j,j)}$:

Theorem 4.1.4. *Fix* $l \in \mathbb{N}^+$*. Suppose that* $(\mathbf{H}u_b^{l+2})$ *holds for a value function* u*,* $\phi \in \mathcal{K}^l_{[0,1]}$*, and suppose a weak Taylor scheme of order* $l+1$*, on an equidistant time grid* π*, such that* $|\pi| = h$*. Then,*

$$u_0^{(j,j)} = \mathbb{E}\Big[(\Gamma_h^\phi)_j g(\hat{X}_T)\Big] + \mathcal{O}(h^l).$$

4.2 Combination of weak Taylor schemes

We can see from Lemma 4.1.4 that for an Euler scheme and $h := t_1$, we can write

$$\mathbb{E}\Big[\Gamma_h^\phi u(h, \hat{X}_h)\Big] = \hat{L}_x^{(1,1)} u(0,x) + \mathcal{O}(h) = \gamma^2(x)\partial_{xx}u(0,x) + \mathcal{O}(h), \tag{4.2.1}$$

where $\hat{L}_x^{(1,1)} u(0,x) = \gamma^2(x)\partial_{xx}u(0,x)$, which includes the Γ. Furthermore, from the previous section, using a weak Taylor scheme of order 2, then

$$\mathbb{E}\Big[\Gamma_h^\phi u(h, \hat{X}_h)\Big] = L^{(1,1)} u(0,x) + \mathcal{O}(h) = \gamma(x)^2\partial_{xx}u(0,x) + \gamma(x)\gamma'(x)\partial_x u(0,x) + \mathcal{O}(h). \tag{4.2.2}$$

As a result, we have several alternatives for approximating the Γ. We could set $\phi \equiv 2$ and:

(a) Use an Euler scheme for the first time step, and the weak Taylor order 2 scheme for the remainder of the time steps, yielding (4.2.1).

(b) Use the weak Taylor order 2 scheme throughout for all time steps, and approximate the Γ by rearranging (4.2.2).

(c) Euler scheme throughout, with Γ approximated as in part a).

(d) For completeness, one could use a weak Taylor 2 scheme for the first step, followed by an Euler scheme for the remainder of the steps. The Γ is approximated as in part b) above.

Remark 4.2.1. *Observe that using the Euler scheme has the apparent advantage of not requiring the* Δ*, since* $\hat{L}_x^{(1,1)}u(0,x)$ *contains the* Γ*.*

4.3 Extrapolation

We begin by performing extrapolation for the Γ using the Euler scheme, to obtain a result similar to Theorem 3.3.1. The proof of the next theorem is very similar to that for the Delta, so for completeness is included in the appendix.

Theorem 4.3.1. *Consider an Euler scheme throughout. Suppose that* $u \in \mathcal{G}_b^4$*. Then, for* $\phi \equiv 2 \in \mathcal{K}^1_{[0,1]}$*,*

$$2\mathbb{E}\Big[(\Gamma_h^{\phi})_j g(\hat{X}_T^n)\Big] - \mathbb{E}\Big[(\Gamma_{2h}^{\phi})_j g(\hat{X}_T^{n/2})\Big] = \hat{L}_x^{(j,j)}u(0,x) + \mathcal{O}(h^2).$$

We now show an expansion using a weak Taylor 2 scheme and $\phi \in \mathcal{K}^1_{[0,1]}$ in order to justify the extrapolation technique for the Γ:

Lemma 4.3.1. *Consider a weak Taylor scheme of order* 2*, and* $\phi \equiv 2 \in \mathcal{K}^1_{[0,1]}$*. Suppose that* $(\mathbf{H}u_b^4)$ *holds. Then,*

$$\mathbb{E}\Big[\Gamma_h^{\phi} g(\hat{X}_T)\Big] = u_0^{(1,1)} + Ch + \mathcal{O}(h^2).$$

Proof. i) We apply the weak Taylor 2 scheme, and consider terms in the Taylor expansion of $u(h,\hat{X}_h)$ as in (4.1.7), expanding up to $\mathcal{O}(h^2)$. We start with $\mathbb{E}\Big[\Gamma_h^{\phi}u(0,x)\Big] = 0$. We now take expectation of $\mathbb{E}\Big[\Gamma_h^{\phi}(\hat{X}_h - x)\partial_x u(0,x)\Big]$ to obtain

$$\begin{aligned}
&\Big[\tfrac{3}{2}\gamma\gamma' + (f - \gamma\gamma' + \tfrac{1}{2}ff'h + \tfrac{1}{4}hf''\gamma^2) + (-f + \tfrac{1}{2}\gamma\gamma' - \tfrac{1}{2}hff' - \tfrac{1}{4}hf''\gamma^2)\Big]\,\partial_x u(0,x)\\
&= \Big[\gamma\gamma' + h\left(\tfrac{1}{2}ff' + \tfrac{1}{2}f''\gamma^2 - \tfrac{1}{2}ff' - \tfrac{1}{2}f''\gamma^2\right)\Big]\,\partial_x u(0,x) = \gamma\gamma'\partial_x u(0,x)
\end{aligned}$$

where $f := f(x)$, $f' := \frac{\mathrm{d}f(x)}{\mathrm{d}x}$ and likewise for γ and higher order derivatives. We now consider the second term in our expansion, namely $\mathbb{E}\Big[\Gamma_h^{\phi}\frac{(\hat{X}_h - x)^2}{2}\partial_{xx}u(0,x)\Big]$ and expand in powers of h

to obtain

$$\begin{aligned}
&\left[\tfrac{15}{8}h\gamma^2(\gamma')^2+3\{-\tfrac{3}{8}h\gamma^2(\gamma')^2+\tfrac{1}{2}h\gamma^2 f'+\tfrac{1}{4}h\gamma^3\gamma'+\tfrac{1}{2}\gamma^2+hf\gamma\gamma'\}\right]\partial_{xx}u(0,x)\\
&+\left[-\tfrac{1}{2}h\gamma^2 f'-\tfrac{1}{4}h\gamma^3\gamma''+\tfrac{3}{8}h\gamma^2(\gamma')^2+\tfrac{1}{2}hf^2-\tfrac{1}{2}\gamma^2-\tfrac{3}{2}hf\gamma\gamma'\right]\partial_{xx}u(0,x)\\
&+\left[\tfrac{1}{2}hf\gamma\gamma'-\tfrac{1}{8}h\gamma^2(\gamma')^2-\tfrac{1}{2}hf^2\right]\partial_{xx}u(0,x)+\mathcal{O}(h^2)\\
&=\left(\gamma^2+h\{\gamma^2(\gamma')^2+\gamma^2 f'+\tfrac{1}{2}\gamma^3\gamma''+2f\gamma\gamma'\}\right)\partial_{xx}u(0,x)+\mathcal{O}(h^2).
\end{aligned}$$

We now consider $\mathbb{E}\Big[\Gamma_h^{\phi}\frac{(\hat{X}_h-x)^3}{3!}\partial_{xxx}u(0,x)\Big]$ and repeat the prescribed steps to obtain

$$\begin{aligned}
&\left[\tfrac{15}{4}h\gamma^3+3\{-\tfrac{1}{2}h\gamma^3\gamma'+\tfrac{1}{2}hf\gamma^2\}+\tfrac{1}{4}h\gamma^3\gamma'-\tfrac{1}{2}hf\gamma^2\right]\partial_{xxx}u(0,x)+\mathcal{O}(h^2)\\
&=h(\tfrac{5}{2}\gamma^3\gamma'+f\gamma^2)\partial_{xxx}u(0,x)+\mathcal{O}(h^2).
\end{aligned}$$

Studying $\mathbb{E}\Big[\Gamma_h^{\phi}\frac{(\hat{X}_h-x)^4}{4!}\partial_{xxxx}u(0,x)\Big]$ yields

$$\tfrac{1}{2}h\gamma^4\partial_{xxxx}u(0,x)+\mathcal{O}(h^2),$$

and with further effort

$$\mathbb{E}\Big[\Gamma_h^{\phi}\frac{(\hat{X}_h-x)^5}{5!}\partial_{xxxxx}u(0,x)\Big]=\mathcal{O}(h^2).$$

We now consider the cross terms

$$\mathbb{E}\Big[\Gamma_h^{\phi}2\frac{(\hat{X}_h-x)h}{2!}\partial_{tx}u(0,x)\Big]=\gamma\gamma' h\partial_{tx}u(0,x)+\mathcal{O}(h^2),$$

$$\mathbb{E}\Big[\Gamma_h^{\phi}3\frac{(\hat{X}_h-x)^2h}{3!}\partial_{txx}u(0,x)\Big]=\gamma^2 h\partial_{txx}u(0,x)+\mathcal{O}(h^2)$$

and $\mathbb{E}\Big[\Gamma_h^{\phi}3\frac{(\hat{X}_h-x)h^2}{3!}\partial_{ttx}u(0,x)\Big]=\mathcal{O}(h^2)$, $\mathbb{E}\Big[\Gamma_h^{\phi}4\frac{(\hat{X}_h-x)^3h}{4!}\partial_{txxx}u(0,x)\Big]=\mathcal{O}(h^2)$,

$$\mathbb{E}\Big[\Gamma_h^{\phi}6\frac{(\hat{X}_h-x)^2h^2}{4!}\partial_{ttxx}u(0,x)\Big]=\mathcal{O}(h^2),$$

and $\mathbb{E}\Big[\Gamma_h^{\phi}6\frac{(\hat{X}_h-x)h^3}{4!}\partial_{tttx}u(0,x)\Big]=\mathcal{O}(h^3)$. This is sufficient to show that for this example, we have $\mathbb{E}\Big[\Gamma_h^{\phi}u(h,\hat{X}_h)\Big]=u_0^{(1,1)}+hC_1+\mathcal{O}(h^2)$.

ii) We now consider the telescoping terms: using the second part of the proof of Theorem 4.3.1, we can conclude that

$$\mathbb{E}\left[\Gamma_h^\phi \sum_{i=1}^{n-1}\left\{u(t_{i+1},\hat{X}_{t_{i+1}}) - u(t_i,\hat{X}_{t_i})\right\}\right] = C_2 h + \mathcal{O}(h^2).$$

Therefore, $\mathbb{E}\left[\Gamma_h^\phi g(\hat{X}_T)\right] = u_0^{(1,1)} + Ch + \mathcal{O}(h^2)$ with $C := C_1 + C_2$. □

The next theorem extrapolates for the Γ analogously to Theorem 3.3.1 for the Δ:

Theorem 4.3.2. *Consider a weak Taylor scheme order 2 throughout. Suppose that $u \in \mathcal{G}_b^4$. Then, for $\phi \equiv 2 \in \mathcal{K}^1_{[0,1]}$,*

$$2\mathbb{E}\left[(\Gamma_h^\phi)_j g(\hat{X}_T^n)\right] - \mathbb{E}\left[(\Gamma_{2h}^\phi)_j g(\hat{X}_T^{n/2})\right] = L^{(j,j)}u(0,x) + \mathcal{O}(h^2).$$

Proof. Application of Lemma 4.3.1. □

Theorem 4.3.3. *Consider a weak Taylor scheme of order 3. Suppose that $u \in \mathcal{G}_b^5$. Then, for $\phi \in \mathcal{K}^2_{[0,1]}$,*

$$\frac{4}{3}\mathbb{E}\left[(\Gamma_h^\phi)_j g(\hat{X}_T^n)\right] - \frac{1}{3}\mathbb{E}\left[(\Gamma_{2h}^\phi)_j g(\hat{X}_T^{n/2})\right] = L^{(j,j)}u(0,x) + \mathcal{O}(h^3).$$

4.4 Simulation results

We consider higher-order approximations of the Γ and extrapolation results. We study Example 3.4.1 throughout.

4.4.1 High-order Γ

We now summarise parameter configurations in Table 4.4.1 for high-order Γ approximations.

ϕ	Expression	Scheme	ζ	MSE	Cost	Slope
$\phi \equiv 2 \in \mathcal{K}^1_{[0,1]}$	$\hat{L}_x^{(1,1)}u_0$	Euler	1/4	$\mathcal{O}(N^{-1/2})$	$\mathcal{O}(N^{5/4})$	-2/5
$\phi \equiv 2 \in \mathcal{K}^1_{[0,1]}$	$L^{(1,1)}u_0$	WT2	1/4	$\mathcal{O}(N^{-1/2})$	$\mathcal{O}(N^{5/4})$	-2/5
$\phi_{s,2} \in \mathcal{K}^2_{[0,1]}$	$L^{(1,1)}u_0$	WT3	1/6	$\mathcal{O}(N^{-2/3})$	$\mathcal{O}(N^{7/6})$	-4/7

Table 4.1: Approximating Γ in different ways, using different ζ and schemes.

In this example, we consider $\zeta = 1/4$. The computational cost is $\mathcal{O}(N^{5/4})$ and the MSE is $\mathcal{O}(N^{-1/2})$, which suggests a gradient of $-2/5$. This slope is confirmed by Figure 4.1.

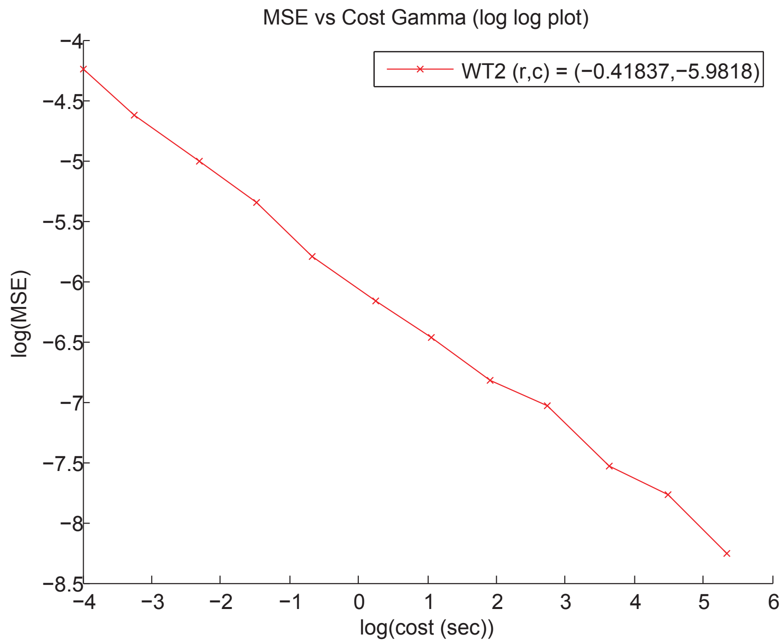

Figure 4.1: MSE vs Cost (log − log) in seconds for the Γ, 250 repeats. Parameters as in Table 4.4.1 (i.e. $\zeta = 1/4$).

Example 4.4.1. *We consider Example 3.4.1. For the scheme configurations $a), b), c)$ from Section 4.2, in Figure 4.2 we use $\zeta = 1/4$ and the true Δ (we require Δ for examples $b)$ and $d)$ above). We consider $N = 2^{18}, \ldots, 2^{23}$, with 30 repeats. We see that when $\zeta = 1/4$, all schemes convergence with the same rate; the predicted value is $-2/5$ since the computational cost is $\mathcal{O}(N^{5/4})$ and the MSE is $\mathcal{O}(N^{-1/2})$.*

4.4.2 Extrapolation for Γ

We consider the three different examples of extrapolations in Figure 4.3, with the parameters summarised in Table 4.2. The first example uses an Euler scheme, with $\zeta = 1/6$. Extrapolating for the Γ using an Euler scheme approximates $\hat{L}_x^{(1,1)} = \gamma^2(x)\Gamma$, which does not include the Δ term. In this example, this is highly attractive, as we do not require an approximation of the Δ to obtain the Γ. The second example, uses a weak Taylor scheme of order 2 throughout, with $\zeta = 1/6$. The approximation now is of $L^{(1,1)}u(0,x)$, which is an expression containing the Δ. We see that this is slightly worse compared to the Γ using an Euler scheme, as we have now used the approximation of the Δ at each step, as opposed to the true value. For both of these examples, the extrapolation is performed using $(A, B) = (2, 1)$, yielding a bias of $\mathcal{O}(h^2)$.

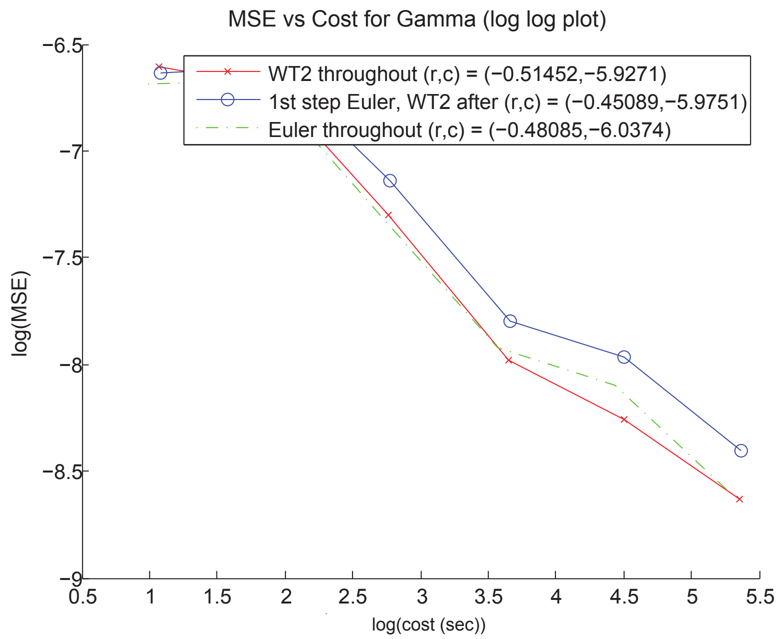

Figure 4.2: See Example 4.4.1. Gamma approximated using various schemes, $\zeta = 1/4$.

ϕ	Expression	Scheme	A	B	ζ	MSE	Cost	Slope
$\phi \equiv 2 \in \mathcal{K}^1_{[0,1]}$	$\hat{L}^{(1,1)}_x u_0$	Euler	2	1	1/6	$\mathcal{O}(N^{-2/3})$	$\mathcal{O}(N^{7/6})$	-4/7
$\phi \equiv 2 \in \mathcal{K}^1_{[0,1]}$	$L^{(1,1)} u_0$	WT2	2	1	1/6	$\mathcal{O}(N^{-2/3})$	$\mathcal{O}(N^{7/6})$	-4/7
$\phi_{s,2} \in \mathcal{K}^2_{[0,1]}$	$L^{(1,1)} u_0$	WT3	4/3	1/3	1/8	$\mathcal{O}(N^{-3/4})$	$\mathcal{O}(N^{9/8})$	-2/3

Table 4.2: Approximating Γ using extrapolation, using different ζ and schemes.

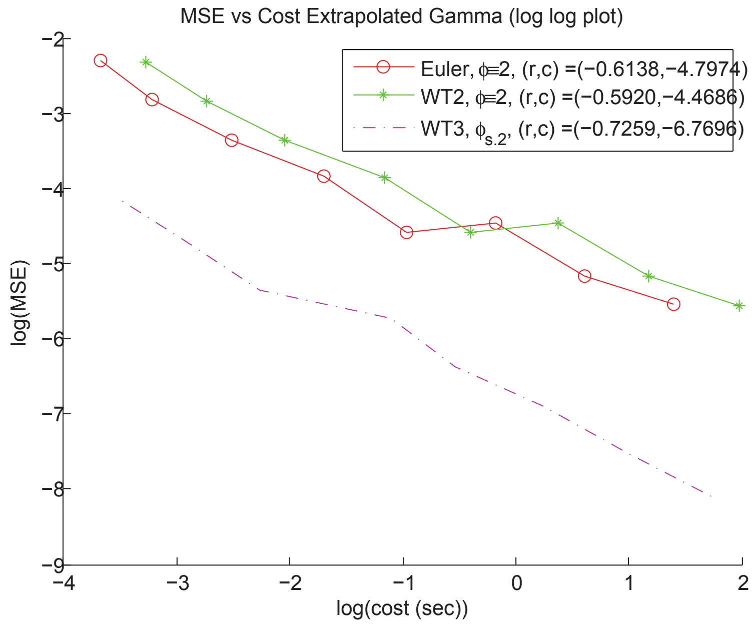

Figure 4.3: MSE vs Cost ($\log - \log$) in seconds for the Γ, 100 repeats, using extrapolation. Euler scheme and WT2 with $\phi \equiv 2$, and $(A, B) = (2, 1)$. Third plot is WT3, using $\psi_{s,2}$ and $(A, B) = (4/3, 1/3)$. See Table 4.2.

Note that the weak Taylor 2 scheme is slightly more computationally tasking. For these two examples, the computational cost is $\mathcal{O}(N^{7/6})$ and the MSE is $\mathcal{O}(N^{-2/3})$, which suggests a slope of $-4/7$, confirmed by the numerics. The third example uses a weak Taylor scheme of order 3, with $\zeta = 1/8$ and $\phi_{s,2}$. The performance is superior and even though though the scheme is very computationally expensive. The computational cost is now $\mathcal{O}(N^{9/8})$, and with an MSE of $\mathcal{O}(N^{-3/4})$ the theoretical slope is $-2/3$, which is observed.

5. Possible extensions

Thus far, we have considered stochastic differential equations of the form (2.0.1), and approximated option sensitivities with respect to the initial state variable, $x \in \mathbb{R}^d$. In this section, we approximate the sensitivity with respect to a parameter, other than a state variable. For example, the Vega of an option in a Black-Scholes model is defined as the sensitivity of the option price to a change in the fixed, initial volatility; the idea here is to make the constant volatility parameter stochastic by introducing a perturbation. It will be seen that the technique relies on the expansion of the value function with a perturbation parameter ε.

Let $X^\varepsilon = (X_t^\varepsilon)_{t\geq 0}$ be a perturbed version of process X, and let $u_\cdot^\varepsilon$ be the value function of the perturbed Cauchy problem. Informally, the aim is to be able to make statements such as $X_t^\varepsilon = X_t + \mathcal{O}(\varepsilon)$ in some probabilistic sense, and similarly for the solution of the PDE. In [FSW12, Theorem 1.2], it is shown that assuming $(\mathbf{H}f1)$ then for all $t, \delta > 0$ it holds that $\mathbb{E}|X_t^\varepsilon - X_t|^2 \leq C_t \varepsilon^2$ and $\lim_{\varepsilon\downarrow 0} \mathbb{P}(\max_{0\leq s\leq t} |X_t^\varepsilon - X_t| > \delta) = 0$.

5.1 General perturbation

We consider a perturbation with an independent Brownian motion. Recalling (2.0.1), consider a driftless, time-homogeneous n-dimensional stochastic process $X = (X_t)_{t\leq 0}$ satisfying

$$\mathrm{d}X_t = \gamma(X_t, \theta)\mathrm{d}W_t^{(1)}, \quad X_0 = x \in \mathbb{R}^d, \tag{5.1.1}$$

where $\theta \in \mathbb{R}^{\mathfrak{d}}$ is fixed, and $W^{(1)}$ is an m-dimensional Brownian motion. Introduce a small perturbation $\varepsilon > 0$, an independent (from $W^{(1)}$) $\mathfrak{d}$-dimensional Brownian motion $W^{(2)} = (W_t^{(2)})_{t\geq 0}$ and consider the perturbed couple $\mathfrak{X}^\varepsilon = (X^\varepsilon, \theta^\varepsilon)$, solution to the following SDEs:

$$\begin{aligned} \mathrm{d}X_t^\varepsilon &= \gamma(X_t^\varepsilon, \theta_t^\varepsilon)\mathrm{d}W_t^{(1)}, & X_0^\varepsilon = x \in \mathbb{R}^d, \\ \mathrm{d}\theta_t^\varepsilon &= \varepsilon \mathrm{d}W_t^{(2)}, & \theta_0^\varepsilon = \theta \in \mathbb{R}^{\mathfrak{d}}. \end{aligned} \tag{5.1.2}$$

Essentially, the dimension of the system has increased from $\mathbb{R}^d$ to $\mathbb{R}^{d+\mathfrak{d}}$, whilst the number of parameters has gone from the $\mathbb{R}^{\mathfrak{d}}$ (the dimensions of θ) to $\mathbb{R}$ (that of ε). Supposing Lipschitz continuity and linear growth on the driving coefficients of (5.1.2) guarantees existence and uniqueness of the solution. We choose to continue with the expansion approach described in the previous section, and fix $d = \mathfrak{d} = 1$ and $m = 1$. The new Brownian motion, $W^{(2)}$, allows us to compute the sensitivity with respect to θ, using a suitable $\mathcal{F}_\vartheta$-measurable weight multiplied by the payoff. In the perturbed model, consider a terminal payoff function $g \in \mathcal{C}_p$. Suppose that $(u_t^\varepsilon)_{t\in[0,T]}$ is the value function, where $u_t^\varepsilon := u_t^\varepsilon(t, X_t^\varepsilon, \theta_t^\varepsilon, \varepsilon)$. Formally, the option pricing paradigm can now be represented as

$$u_0^\varepsilon = \mathbb{E}[g\left((\mathfrak{X}_T^\varepsilon)_1\right)] = \mathbb{E}[g\left(X_T^\varepsilon\right)],$$

where $(a)_i$ is notation for the i^{th} entry of a. It is clear that $u_0^0(0, x, \theta, 0) = u_0(0, x, \theta)$, i.e. the value function of (5.1.1) coincides with that of (5.1.2) when the perturbation parameter ε is zero. However, it is not obvious how the limit of $\lim_{\varepsilon\downarrow 0} u_0^\varepsilon(0, x, \theta, \varepsilon)$ behaves, whether it exists, and whether it equals to $u_0(0, x, \theta)$.

Throughout, we make the following assumptions on the ability to expand the value function in terms of the perturbation ε:

$(\mathbf{H}u_\varepsilon^{r,l})$: There exists a rate $r > 0$ such that for all $\varepsilon > 0$, $u_\cdot^\varepsilon = u_\cdot^0 + C\varepsilon^r + o(\varepsilon^r)$ holds pointwise and for all multi-indices α such that $l(\alpha) \le l$ also $L^\alpha u_\cdot^\varepsilon = L^\alpha u_\cdot^0 + C\varepsilon^r + o(\varepsilon^r)$ holds pointwise.

From now on assume $(\mathbf{H}u_\varepsilon^{1,l})$ holds throughout for some $l \ge 1$, although we do not attempt to impose conditions for this strong condition to hold. The analysis will be performed assuming the rate $r = 1$, and it could be repeated in the same manner for a general $r > 0$.

The aim is to approximate sensitivities, such as

$$\left.\frac{\partial u_0^\varepsilon(0, x, \theta, \varepsilon)}{\partial x}\right|_{\varepsilon\downarrow 0} \qquad \text{and} \qquad \left.\frac{\partial u_0^\varepsilon(0, x, \theta, \varepsilon)}{\partial \theta}\right|_{\varepsilon\downarrow 0},$$

in addition to higher-order derivatives.

Remark 5.1.1. *We do not require Delta approximations, however insist on including them in the analysis for the included perturbation. The reason behind that is so that all Greeks can be computed with one forward pass using the perturbed stochastic differential equation.*

Remark 5.1.2. *Expansions for the volatility of volatility in a similar framework have been considered in [Lew00, Chapter 3]. The author presents two expansions: in terms of the option price, and in terms of the implied volatility, and is able to present some asymptotics for them (see [Lew00, Table 3.2], with better performance for the second type of expansion).*

In the next proposition, the aim is to approximate sensitivities with respect to x and θ, and compute the Greeks using $(X_t^\varepsilon)_{t\in[0,T]}$, whilst ε tends to zero. Again, the operators $L^{(0)}, L^{(1)}, L^{(2)}$ defined in (2.1.3) and (2.1.2) will be used. As a convention (and slight abuse of notation), γ denotes $\gamma(x,\theta)$, where (x,θ) are the initial values of the driving SDEs.

Proposition 5.1.1. *Consider the model in* (5.1.2)*, and assume* $(\mathbf{H}u_b^3)$ *and* $(\mathbf{H}u_\varepsilon^{1,2})$*. Let* $\gamma : \mathbb{R}^2 \to \mathbb{R}$ *be three times continuously differentiable, with bounded derivatives. Then, Table 5.1 shows expressions containing the first and second-order sensitivities with respect to x and* θ*, by multiplying the payoff by the given weights:*

Weight	Value	Bias
$\frac{I_\vartheta^{(1)}}{\vartheta\gamma}$	$\partial_x u_0^0$	$\mathcal{O}(\vartheta)+\mathcal{O}(\varepsilon)$
$\frac{I_\vartheta^{(2)}}{\vartheta\varepsilon}$	$\partial_\theta u_0^0$	$\mathcal{O}(\vartheta)+\mathcal{O}(\varepsilon)$
$\frac{2I_\vartheta^{(1,1)}}{\vartheta^2}$	$L^{(1,1)}u_0^0$	$\mathcal{O}(\vartheta)+\mathcal{O}(\varepsilon)$
$\frac{2I_\vartheta^{(1)}I_\vartheta^{(2)}}{\vartheta^2}$	$L^{(1,2)}u_0^0+L^{(2,1)}u_0^0$	$\mathcal{O}(\vartheta\varepsilon)+\mathcal{O}(\varepsilon^2)$
$\frac{2I_\vartheta^{(2,2)}}{\vartheta^2\varepsilon^2}$	$\partial_{\theta\theta}u_0^0$	$\mathcal{O}(\vartheta)+\mathcal{O}(\varepsilon)$

Table 5.1: Two dimensional sensitivities and weights for the general perturbed model.

Proof. The Itô-Taylor expansion in (2.2.7) can be recalled for u_ϑ^ε; multiplying it by $I_\vartheta^{(1)}$ yields

$$\begin{aligned}
\mathbb{E}\left[u_\vartheta^\varepsilon I_\vartheta^{(1)}\right] &= \mathbb{E}\left[I_\vartheta^{(1)}\left[L^{(1)}u_0^\varepsilon\right]I_\vartheta^{(1)}\right]+\mathbb{E}\left[I_\vartheta^{(1,0)}\left[L^{(1,0)}u_\cdot^\varepsilon\right]I_\vartheta^{(1)}\right]+\mathbb{E}\left[I_\vartheta^{(0,1)}\left[L^{(0,1)}u_\cdot^\varepsilon\right]I_\vartheta^{(1)}\right]\\
&= \vartheta L^{(1)}u_0^\varepsilon+\mathcal{O}(\vartheta^2)\\
&= \vartheta L^{(1)}u_0^0+\mathcal{O}(\vartheta^2)+\mathcal{O}(\vartheta\varepsilon)\\
&= \vartheta\gamma\partial_x u_0^0+\mathcal{O}(\vartheta^2)+\mathcal{O}(\vartheta\varepsilon),
\end{aligned}$$

thus obtaining the weight $I_\vartheta^{(1)}/(\vartheta\gamma)$ for $\partial_x u_0^0$ and the corresponding bias. The proof for $\partial_\theta u_0^0$ is similar, thus omitted. In summary, the perturbed solution X^ε is multiplied by the following weights to approximate the first-order Greeks:

$$\begin{aligned}
\partial_x u_0^0 &= \mathbb{E}\left[g(X_T^\varepsilon)\frac{I_\vartheta^{(1)}}{\vartheta\gamma}\right]+\mathcal{O}(\varepsilon)+\mathcal{O}(\vartheta),\\
\partial_\theta u_0^0 &= \mathbb{E}\left[g(X_T^\varepsilon)\frac{I_\vartheta^{(2)}}{\vartheta\varepsilon}\right]+\mathcal{O}(\varepsilon)+\mathcal{O}(\vartheta).
\end{aligned} \tag{5.1.3}$$

For the second-order sensitivities, recall the hierarchical set $\mathcal{D}_2$ and consider the Itô-Taylor expansion of u^ε_ϑ which has the following Itô-Taylor expansion:

$$u^\varepsilon_\vartheta = \sum_{\alpha\in\mathcal{D}_2} I^\alpha_\vartheta \left[L^\alpha u^\varepsilon_0\right] + \sum_{\alpha\in\mathcal{B}(\mathcal{D}_2)} I^\alpha_\vartheta \left[L^\alpha u^\varepsilon_\cdot\right].$$

Multiply the expansion by $I^{(1,1)}_\vartheta$, and by

$$\mathbb{E}\left[u^\varepsilon_\vartheta \frac{2I^{(1,1)}_\vartheta}{\vartheta^2}\right] = L^{(1,1)}u^\varepsilon_0 + \mathcal{O}(\vartheta),$$

we obtain

$$\mathbb{E}\Big[g(X^\varepsilon_T)2I^{(1,1)}_\vartheta/\vartheta^2\Big] = L^{(1,1)}u^\varepsilon_0 + \mathcal{O}(\vartheta) = L^{(1,1)}u^0_0 + \mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon). \tag{5.1.4}$$

Similar analysis can be performed for the weight $2I^{(2,2)}_\vartheta/\vartheta^2$ to obtain an expression containing $L^{(2,2)}u^0_0$; since $L^{(2)} = \varepsilon\partial_\theta$, hence it holds that

$$\mathbb{E}\left[g(X^\varepsilon_T)\frac{2I^{(2,2)}_\vartheta}{\vartheta^2\varepsilon^2}\right] = \partial_{\theta\theta}u^0_0 + \mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon). \tag{5.1.5}$$

For the cross-terms, repeat the same computations, and recall that $\mathbb{E}\Big[I^{(1,2)}_\vartheta + I^{(2,1)}_\vartheta\Big] = \mathbb{E}\Big[I^{(1)}_\vartheta I^{(2)}_\vartheta\Big]$ from Remark 2.3.1, so that the expression

$$\mathbb{E}\left[g(X^\varepsilon_T)\frac{2I^{(1)}_\vartheta I^{(2)}_\vartheta}{\vartheta^2}\right] = L^{(1,2)}u^0_0 + L^{(2,1)}u^0_0 + \mathcal{O}(\vartheta\varepsilon) + \mathcal{O}(\varepsilon^2) \tag{5.1.6}$$

holds, assuming smoothness for the functions γ and u^ε_ϑ. Combining expressions (5.1.3), (5.1.4), (5.1.5) and (5.1.6) proves the results in Table 5.1. □

5.1.1 Example: perturbed Bachelier model

Consider the perturbed Bachelier model:

Definition 5.1.1. *The driftless perturbed Bachelier model is* (5.1.2) *with* $\gamma(x,\theta) \equiv \theta$.

The strategy is to approximate sensitivities with respect to x and θ, and compute the Greeks using the perturbed process $(X^\varepsilon_t)_{t\in[0,T]}$, whilst the perturbation ε tends to zero. We proceed with a corollary (proof in Appendix A.2), which is a consequence of Proposition 5.1.1.

Corollary 5.1.1 (Perturbed Bachelier Greeks). *Suppose that* $(\mathbf{H}u_b^3)$ *and* $(\mathbf{H}u_\varepsilon^{1,2})$ *hold. Then, the following expressions for the first and second-order Greeks from Table 5.2 hold:*

Greek	Weight	Value	Bias
Delta	$\frac{I_\vartheta^{(1)}}{\vartheta\theta}$	$\partial_x u_0^0$	$\mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon)$
Vega	$\frac{I_\vartheta^{(2)}}{\vartheta\varepsilon}$	$\partial_\theta u_0^0$	$\mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon)$
Gamma	$\frac{2I_\vartheta^{(1,1)}}{\vartheta^2\theta^2}$	$\partial_{xx} u_0^0$	$\mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon)$
Vanna	$\frac{I_\vartheta^{(1)} I_\vartheta^{(2)}}{\vartheta^2\theta\varepsilon} - \frac{I_\vartheta^{(1)}}{2\vartheta\theta^2}$	$\partial_{\theta x} u_0^0$	$\mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon)$
Vomma	$\frac{2I_\vartheta^{(2,2)}}{\vartheta^2\varepsilon^2}$	$\partial_{\theta\theta} u_0^0$	$\mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon)$

Table 5.2: Bachelier Greeks using two-dimensional uncorrelated Brownian motion for the underlying and the volatility.

We now prove convergence results for the Bachelier Greek approximations in Table 5.2, when using a discretisation scheme with a strong rate of convergence of order k. We proceed as in Section 2.4 to compute the MSE of the Greek approximations, including the error introduced from the perturbed SDE using the parameter ε. The parameters ζ, η are chosen as before (recall definitions on page 80), and the perturbation is set to $\varepsilon := 1/N^\nu$.

Consider an approximation for the Delta with N paths under the perturbed model to be $\hat{\Delta}_N^\varepsilon := \frac{1}{N}\sum_{i=1}^N g(\hat{X}_T^{\varepsilon,i}) I_\vartheta^{(1),i}/(\vartheta\theta)$, where $\hat{X}_T^{\varepsilon,i}$ is the i^{th} simulated path.

Proposition 5.1.2 (Delta). *Assume* $(\mathbf{H}u_b^2)$ *and* $(\mathbf{H}u_\varepsilon^{1,2})$*, g is bounded and Lipschitz continuous, and consider a discretisation scheme with strong convergence rate k. Then, for $\zeta = 1/3$ and $\nu, k\eta \geq 1/3$ the MSE of $\hat{\Delta}_N^\varepsilon$ is $\mathcal{O}(N^{-2/3})$.*

Proof. The bias of the approximation of the Delta is $\mathbb{E}[\Delta_N^\varepsilon] - \Delta = \mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon) + \mathcal{O}(\vartheta\varepsilon)$. By using a discretisation scheme with strong rate of convergence k, it follows that $\mathbb{E}[\hat{\Delta}_N^\varepsilon] = \Delta + \mathcal{O}(h^k) + \mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon) + \mathcal{O}(\vartheta\varepsilon)$. The variance of $\hat{\Delta}_N^\varepsilon$ is $\mathcal{O}(1/(N\vartheta))$. The proof then follows from observing that the MSE of $\hat{\Delta}_N^\varepsilon$ is

$$\mathcal{O}(N^{\zeta-1}) + \mathcal{O}(N^{-2k\eta}) + \mathcal{O}(N^{-2\nu}) + \mathcal{O}(N^{-2\zeta}) + \mathcal{O}(N^{-\nu-k\eta}) + \mathcal{O}(N^{-k\eta-\zeta}) + \mathcal{O}(N^{-\zeta-\nu}).$$

From the first and the fourth terms it follows that $\zeta = 1/3$. To match the other errors, it is necessary to have $\nu, k\eta \geq 1/3$. With such choice of parameters, the MSE is of order $\mathcal{O}(N^{-2/3})$.

This concludes the proof of Proposition 5.1.2. □

The next Greek of interest is the $\mathcal{V}$ and its approximation is $\hat{\mathcal{V}}^{\varepsilon}_N := \frac{1}{N}\sum_{i=1}^N g(\hat{X}^{\varepsilon,i}_T)\frac{I^{(2),i}_{\vartheta}}{\vartheta\varepsilon}$. Similarly, the bias of this approximation, when using a strong-order scheme with rate k is $\mathbb{E}[\hat{\mathcal{V}}^{\varepsilon}_N] - \mathcal{V} = \mathcal{O}(h^k) + \mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon) + \mathcal{O}(\vartheta\varepsilon)$.

Proposition 5.1.3 (Vega). *Assume* $(\mathbf{H}u^2_b)$ *and* $(\mathbf{H}u^{1,2}_{\varepsilon})$, *$g$ is bounded and Lipschitz continuous, and consider a discretisation scheme with strong convergence rate k. Then, for $\zeta = \nu = 1/5$, $k\eta \geq 1/5$ the MSE of $\hat{\mathcal{V}}^{\varepsilon}_N$ is $\mathcal{O}(N^{-2/5})$.*

Proof. For the Vega, the variance of $\mathcal{V}^{\varepsilon}_N$ is

$$\mathbb{V}(\hat{\mathcal{V}}^{\varepsilon}_N) = \frac{\mathbb{V}\left(g(\hat{X}^{\varepsilon}_T)I^{(2)}_{\vartheta}\right)}{N\vartheta^2\varepsilon^2} = \mathcal{O}\left(\frac{1}{N\vartheta\varepsilon^2}\right).$$

Combining this with the bias, the MSE is proportional to

$$\mathcal{O}(N^{\zeta+2\nu-1}) + \mathcal{O}(N^{-2k\eta}) + \mathcal{O}(N^{-2\nu}) + \mathcal{O}(N^{-2\zeta}) + \mathcal{O}(N^{-\nu-k\eta}) + \mathcal{O}(N^{-k\eta-\zeta}) + \mathcal{O}(N^{-\zeta-\nu}).$$

From this, it follows that $1-\zeta-2\nu = 2\nu = 2\zeta$, therefore $\zeta = \nu = 1/5$. In addition, if $k\eta \geq 1/5$ is chosen, the MSE is of order $\mathcal{O}(N^{-2/5})$. □

Consider now the second-order Greeks, namely Gamma, Vanna and Vomma, referring to them as Γ, $\mathcal{V}_a$ and $\mathcal{V}_o$. Their corresponding approximations in the perturbed model are defined by

$$\begin{aligned}
\hat{\Gamma}^{\varepsilon}_N &:= \frac{1}{N}\sum_{i=1}^N g(\hat{X}^{\varepsilon,i}_T)\frac{2I^{(1,1),i}_{\vartheta}}{\vartheta^2\theta^2},\\
\hat{\mathcal{V}}^{\varepsilon}_{a,N} &:= \frac{1}{N}\sum_{i=1}^N g(\hat{X}^{\varepsilon,i}_T)\left[\frac{I^{(1),i}_{\vartheta}I^{(2),i}_{\vartheta}}{\vartheta^2\theta\varepsilon} - \frac{I^{(1),i}_{\vartheta}}{2\vartheta\theta^2}\right],\\
\hat{\mathcal{V}}^{\varepsilon}_{o,N} &:= \frac{1}{N}\sum_{i=1}^N g(\hat{X}^{\varepsilon,i}_T)\frac{2I^{(2,2),i}_{\vartheta}}{\vartheta^2\varepsilon^2}.
\end{aligned} \tag{5.1.7}$$

Proposition 5.1.4 (Gamma, Vanna, Vomma). *Assume* $(\mathbf{H}u^3_b)$ *and* $(\mathbf{H}u^{1,2}_{\varepsilon})$, *$g$ bounded and Lipschitz continuous, and consider a discretisation scheme with strong convergence rate k. Then,*

1. *for $\zeta = 1/4$ and $\nu, k\eta \geq 1/4$ the MSE of $\hat{\Gamma}^{\varepsilon}_N$ is of order $\mathcal{O}(N^{-1/2})$;*

2. *for $\zeta = \nu = 1/6$ and $k\eta \geq 1/6$ the MSE of $\hat{\mathcal{V}}^{\varepsilon}_{a,N}$ is of order $\mathcal{O}(N^{-1/3})$;*

3. *for $\zeta = \nu = 1/8$ and $k\eta \geq 1/8$ the MSE of $\hat{\mathcal{V}}^{\varepsilon}_{o,N}$ is of order $\mathcal{O}(N^{-1/4})$.*

Proof. All three proofs follow in the spirit of Proposition 5.1.2 and 5.1.3. The different parameter constraints on $\zeta, \nu, k\eta$ arise from the variance term, so that the MSE can be controlled.
1. The variance of $\hat{\Gamma}^{\varepsilon}_N$ is $\mathcal{O}(1/(N\vartheta^2))$. The MSE for the approximation of the Gamma is $\mathcal{O}(N^{2\zeta-1}) + \mathcal{O}(N^{-2k\eta}) + \mathcal{O}(N^{-2\nu}) + \mathcal{O}(N^{-2\zeta}) + \mathcal{O}(N^{-\nu-k\eta}) + \mathcal{O}(N^{-k\eta-\zeta}) + \mathcal{O}(N^{-\zeta-\nu})$. Therefore, to balance the error from the first and the fourth terms, $1 - 2\zeta = 2\zeta$ yields $\zeta = 1/4$ and the MSE is $\mathcal{O}(N^{-1/2})$ when $k\eta, \nu \geq 1/4$.
2. The variance of $\hat{\mathcal{V}}^{\varepsilon}_{a,N}$ is $\mathcal{O}(1/(N\vartheta^2\varepsilon^2))$. From the MSE, $1 - 2\zeta - 2\nu = 2\nu = 2\zeta$ yields $\zeta = \nu = 1/6$. By choosing $k\eta \geq 1/6$, it follows that the MSE is $\mathcal{O}(N^{-1/3})$.
3. The variance of $\hat{\mathcal{V}}^{\varepsilon}_{o,N}$ is $\mathcal{O}(1/(N\vartheta^2\varepsilon^4))$. From the MSE of the Vomma, it follows that $1 - 2\zeta - 4\nu = 2\nu = 2\zeta$ therefore $\zeta = \nu = 1/8$, and for $k\eta \geq 1/8$ it follows that the MSE is $\mathcal{O}(N^{-1/4})$. □

A summary of Proposition 5.1.2-5.1.4 is included in Figure 5.3.

Greek	ζ	ν	$k\eta$	MSE
Delta	$1/3$	$\geq 1/3$	$\geq 1/3$	$\mathcal{O}(N^{-2/3})$
Vega	$1/5$	$1/5$	$\geq 1/5$	$\mathcal{O}(N^{-2/5})$
Gamma	$1/4$	$\geq 1/4$	$\geq 1/4$	$\mathcal{O}(N^{-1/2})$
Vanna	$1/6$	$1/6$	$\geq 1/6$	$\mathcal{O}(N^{-1/3})$
Vomma	$1/8$	$1/8$	$\geq 1/8$	$\mathcal{O}(N^{-1/4})$

Table 5.3: Parameters and constraints, with MSE.

5.1.2 Numerical results: Vega with perturbation

We now consider the Bachelier model with parameters $(x, \theta, T) = (100, 20, 1)$, and consider a European call option with strike $K = 105$. Our focus is to compute the Vega, and from Table 5.3, we use an Euler scheme and fix $(\zeta, \nu, \eta) = (1/5, 1/5, 1/5)$. In Figure 5.1, we show the improved approximations for these Greeks against N.

Remark 5.1.3. *For this example, it is not necessary to include a perturbation to compute the Delta and Gamma; the plots simply demonstrate that even with the ε perturbation, we can compute these Greeks using the perturbed model.*

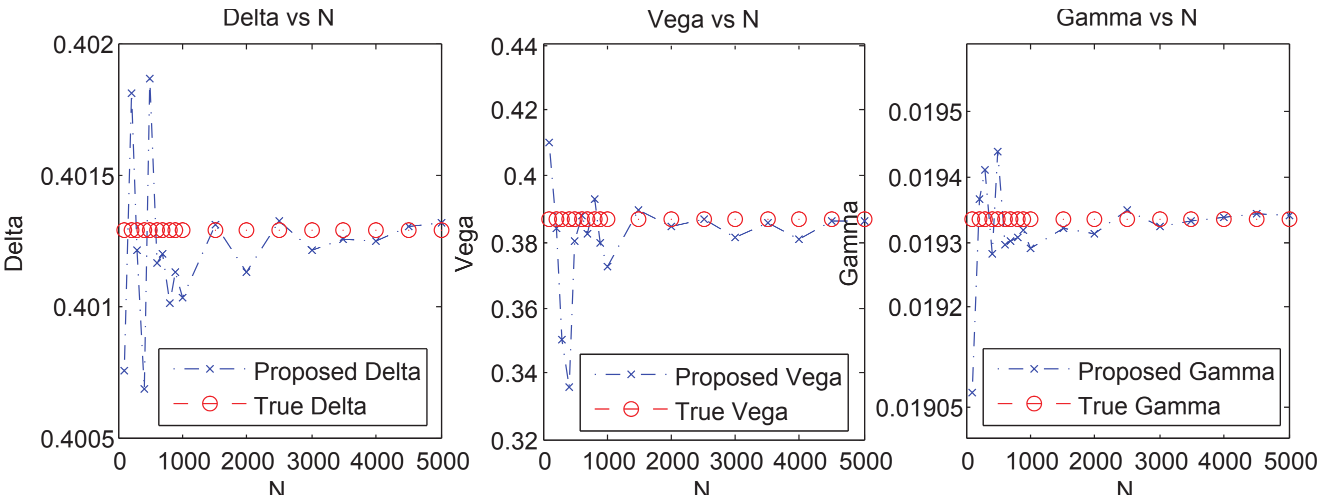

Figure 5.1: Bachelier model Delta, Vega, and Gamma, using a perturbation. MSE vs N using antithetic variables and an Euler scheme.

In Figure 5.2, we plot the mean squared error against the computational cost (measured in seconds), in a $\log - \log$ scale. The slope of this is -0.73, which is an improvement on the predicted value of $-1/3$ (since the computational cost is $\mathcal{O}(N^{6/5})$ for $\zeta = \nu = \eta = 1/5$).

5.2 High-order Greeks for non-linear pricing

Backward stochastic differential equations (BSDEs) have been widely used in stochastic control, and in mathematical finance for pricing problems, see e.g. [EKPQ97, MY99, PP92, EKHM08] and references therein. The solution of a (decoupled) forward-backward stochastic differential equation consisting of the adapted processes (Y, Z) satisfying

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + \gamma(X_t)\mathrm{d}W_t, \qquad X_0 = x, \tag{5.2.1}$$

$$-\mathrm{d}Y_t = h(X_t, Y_t, Z_t)\mathrm{d}t - Z_t^* \mathrm{d}W_t, \qquad Y_T = g(X_T), \tag{5.2.2}$$

where $h : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is the driver, and $f : \mathbb{R}^d \to \mathbb{R}^d$, $\gamma : \mathbb{R}^d \to \mathbb{R}^d$, $g : \mathbb{R}^d \to \mathbb{R}$ and h are some Lipschitz continuous functions.

The approximation of the forward X process is well studied, and in the BSDE literature the focus is on approximating (Y, Z) in a backward, recursive manner. Recent methods extend the setting to a broader class of BSDEs (with drivers) based on Euler approximations [BT04]; other

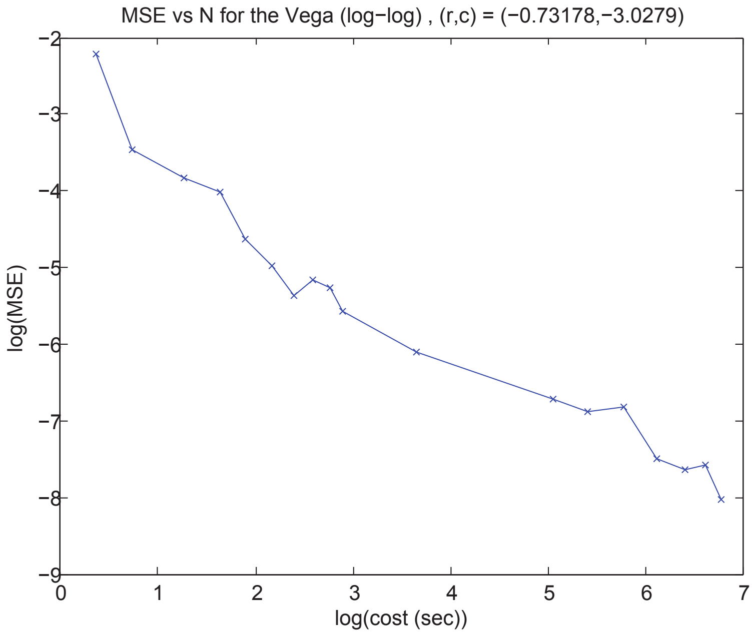

Figure 5.2: Bachelier model Vega: MSE vs Cost ($\log - \log$) using antithetic variates and an Euler scheme.

work considers multi-step and Runge-Kutta schemes [Cha14, CC14]. A variance reduction technique is considered in [AA13]. A possible extension could be extending the solution of a BSDE, to include process $G = (G_t)_{t \in [0,T]}$. Formally, define $G_t := L^{(1,1)} u(t, X_t)$, for $t \in [0, T]$, and consider a solution (X, Y, Z, G).

Consider an equidistant mesh π with n time points $0 = t_0 < t_1 < \ldots < t_n = T$ and denote by (Y_i, Z_i, G_i) the approximation of $(Y_{t_i}, Z_{t_i}, G_{t_i})$ for $i = 0, \ldots, n$, where $h := T/n$.

On this time grid, we suggest a one-step fully-implementable approximation:

(i) Initialize the terminal conditions, (Y_n, Z_n, G_n), which are $\mathcal{F}_{t_n}$-measurable, square-integrable random variables.

(ii) Let approximations (Y_i, Z_i, G_i) be given by

$$Y_i = \mathbb{E}_{t_i}[Y_{i+1} + (t_{i+1} - t_i) h(t_i, X_i, Y_i, Z_i)], \qquad Z_i = \mathbb{E}_{t_i}[H^{\psi}_{t_i,h} Y_{i+1}], \qquad G_i = \mathbb{E}_{t_i}[\Gamma^{\phi}_{t_i,h} Y_{i+1}],$$

where the coefficients $H^{\psi}_{t_i,h}$ and $\Gamma^{\phi}_{t_i,h}$ are $\mathcal{F}_{t_{i+1}}$-measurable random variables, such that for some positive Λ

$$h\mathbb{E}\Big[|H^{\psi}_{t_i,h}|^2\Big] \leq \Lambda, \qquad \mathbb{E}_{t_i}[H^{\psi}_{t_i,h}] = 0, \text{ and } \qquad h^2\mathbb{E}\Big[|\Gamma^{\phi}_{t_i,h}|^2\Big] \leq \Lambda, \qquad \mathbb{E}_{t_i}[\Gamma^{\phi}_{t_i,h}] = 0.$$

Such schemes have been considered in [FTW11, CSTV07]. The value of (Y_n, Z_n, G_n) is given by $(g(X_T), \gamma(X_T)\partial_x g(X_T), \gamma(X_T)\partial_x(\gamma(X_T)\partial_x g(X_T)))$ extending the notion of a solution in [PP92]. Convergence properties for (Y, Z) are well studied for one-step, multi-step and Runge-Kutta schemes [Zha04, Cha14, CC14].

Future work in this general direction would be to adapt what has been done in the previous sections to the non-linear setting (extrapolation method, Gamma approximations).

Part III

Explicit Euler scheme for SDEs

6. An Explicit Euler scheme for financial SDEs with non-Lipschitz coefficients

We propose here a modified explicit Euler-Maruyama discretisation scheme for a class of stochastic differential equations with non-Lipschitz drift or diffusion coefficients. This scheme yields strong convergence, with a rate, which, under some regularity and integrability conditions on the coefficients of the SDE, is actually optimal. We then apply it to some widely used diffusion models in the mathematical finance literature, including the Cox-Ingersoll-Ross, the CEV, the 3/2 and the Ait-Sahalia models, as well as to a family of mean-reverting processes with locally smooth coefficients.

6.1 Introduction

One of the main tasks in mathematical finance is to evaluate complex derivative products, where the underlying assets are modelled by multi-dimensional SDEs which rarely admit closed-form solutions. Monte Carlo techniques are therefore needed to approximate these prices, and Glasserman's book [Gla03] has become the main reference for a comprehensive overview of such methods with applications to financial engineering.

Classical weak and strong convergence results for discretisation schemes of SDEs assume that the drift and the diffusion coefficients are globally Lipschitz continuous [KP92]; however many models used in the literature, such as the CIR, CEV, Ait-Sahalia models, violate this assumption. For pricing purposes, weak error is usually sufficient, but strong convergence rates are needed when using multilevel Monte Carlo methods (MLMC), in order to optimise the computational complexity [Gil08b, GHM09].

In traditional Euler-Maruyama discretisation schemes, the constructed approximation can potentially escape the domain of the true solution of the SDE. In recent years, a lot of effort has been focused on deriving schemes staying in restricted domains for SDEs with non-Lipschitz continuous coefficients [Alf13a, BBD08, BD04, HMS02, HJK12, NS12]. Several modifications have been introduced such as the drift-implicit [DNS12] and the increment-tamed explicit Euler schemes [HJ12, Theorem 3.15]; in the context of mathematical finance, a thorough overview of these can be found in [KN12].

A now classical trick is to apply a suitable Lamperti transform in order to obtain an SDE with constant diffusion coefficient, thereby translating all the non-smoothness to the drift. In the context of non-globally Lipschitz coefficients, this idea, introduced by Alfonsi [Alf05], was further exploited in [Alf13b, NS12] to obtain strong L^p-convergence rates for implicit "Lamperti-Euler" schemes, in particular for the CIR and the Ait-Sahalia models, and for scalar SDEs with one-sided Lipschitz continuous drift and constant diffusion [NS12].

Under sufficient differentiability conditions, modified Itô-Taylor schemes [JKN09] of order $\psi > 0$ provide pathwise convergence results of order $\psi - \varepsilon$ (for arbitrarily small $\varepsilon > 0$). This approach relies on a localisation argument similar to that in [Gyö98], with an auxiliary drift and diffusion function chosen upon the discretised process exiting a sub-domain. For irregular coefficients, some strong rates of convergence have been obtained under more restrictive conditions in [Gyö98, GR11, Yan02, NT13].

Motivated by these different approaches, our main contribution is to provide an efficient numerical approximation of SDEs with non-globally Lipschitz coefficients.

We first present an explicit Euler scheme with a projection for SDEs with locally Lipschitz and globally one-sided Lipschitz drift coefficient, which has a computational cost of the same order as the explicit Euler-Maruyama scheme. We prove strong rates of convergence for a wide family of SDEs, often exceeding the parameter range of the implicit schemes available in the literature. Under suitable assumptions, we are able to obtain fast convergence reaching the optimal rates of convergence. The scheme shares some of the features of the tamed-scheme family. Its analysis however does not require heavy technical tools. Having in mind application to mathematical finance, the analysis is made for SDEs whose support is included in $(0, \infty)$. Nevertheless, the techniques used here can be extended to the multi-dimensional cases under some suitable assumptions. An important contribution is the choice of the scheme in relation to considering the rate of explosion of the drift function at the boundaries of the domain through a locally Lipschitz continuous condition. To the best of our knowledge, thus far in the literature of tamed schemes, only the exploding behaviour at infinity has been considered.

We then turn our attention on SDEs with non-globally Lipschitz diffusion coefficients, as often encountered in finance. We apply a Lamperti transformation to the process of interest in order to shift the non-Lipschitz behaviour from the diffusion to the drift function, before using the modified scheme. This allows us to prove rate of convergence for the original process in the $L^{1+\varepsilon}$-norm for $\varepsilon \geq 0$. The rate of convergence for the value $\varepsilon = 1$ can then be used for MLMC applications.

The remainder of the chapter is structured as follows. In Section 6.2, the modified Euler-Maruyama scheme is introduced. In Section 6.3, the main convergence result is proven for the scheme. In Section 6.4, the scheme is applied to families of SDEs, such as the CIR, the 3/2 and the Ait-Sahalia models, widely used in mathematical finance, and the Ginzburg-Landau. In Section 6.5, numerical results for the rates of convergence obtained are shown and discussed.

Notations: In the sequel, D is the interval $(0, \infty)$. We denote by $\tilde{D}_\eta$ the domain $[\eta, \infty)$, and $\bar{D} := \tilde{D}_0$. Furthermore, we define the interval $\check{D}_\zeta := (-\infty, \zeta]$ and $\check{D}_{\eta,\zeta} = \tilde{D}_\eta \cap \check{D}_\zeta$, for $\eta \leq \zeta$. We denote by $\mathcal{C}^2(D)$ the space of twice differentiable functions with continuous derivatives on D, and by $\mathcal{C}_b^2(D)$ the space of functions in $\mathcal{C}^2(D)$ with first and second bounded derivatives. We shall denote by $\mathbb{N}^+$ the set of strictly positive integers. For $m > 0$, we denote L^m the set of random variable Z such that $\|Z\|_m := \mathbb{E}[|Z|^m]^{1/m} < +\infty$.

6.2 Definitions and assumptions

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space, and $W = (W_t)_{t\geq 0}$ a standard $(\mathcal{F}_t)$-adapted Brownian motion. Consider a stochastic differential equation of the form

$$dY_t = f(Y_t)dt + \gamma(Y_t)dW_t, \qquad Y_0 = y_0. \tag{6.2.1}$$

Throughout this article, we shall assume the following:

$(\mathbf{H}y0)$: the SDE (6.2.1) admits a unique strong solution in $D = (0, \infty)$; the drift f is locally Lipschitz continuous and globally one-sided Lipschitz continuous on D, namely there exist $\alpha, \beta \geq 0$, $K > 0$, such that for all $(x, y) \in D^2$:

$$|f(x) - f(y)| \leq K\left(1 + |x|^\alpha + |y|^\alpha + \frac{1}{|x|^\beta} + \frac{1}{|y|^\beta}\right)|x - y|, \tag{6.2.2}$$

$$(x - y)\left(f(x) - f(y)\right) \leq K|x - y|^2; \tag{6.2.3}$$

furthermore, the diffusion function γ is K-Lipschitz continuous on $\bar{D}$ for some $K > 0$: for all $(x, y) \in \bar{D}^2$, the inequality $|\gamma(x) - \gamma(y)| \leq K|x - y|$ holds.

Remark 6.2.1. *The function γ could as well be defined on D. However, assuming the Lipschitz* *continuity of γ on D would lead to a natural extension of γ on $\bar{D}$.*

Remark 6.2.2. *In many models used in practice (in particular the Feller/CIR diffusion in mathematical finance, see Section 6.4.1), these assumptions are not met. A suitable change of variables, however, allows us to bypass this: consider an SDE of the form*

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \qquad X_0 = x_0, \tag{6.2.4}$$

where the process X takes values in some domain $D_X \subseteq \mathbb{R}$. If $\sigma(x) > 0$ for all $x \in D_X$, the Lamperti transformation of X is defined as $F(x) \equiv \int^x \sigma(z)^{-1}\mathrm{d}z$, and Itô's Lemma implies that the process defined pathwise by $Y := F(X)$ satisfies (6.2.1) *with $f \equiv F'\mu + \frac{1}{2}F''\sigma^2$ and $\gamma \equiv F'\sigma$ is constant.*

Let $n \in \mathbb{N}^+$ be a fixed positive integer and $T > 0$ a fixed time horizon. Define the partition of the interval $[0, T]$ by $\pi := \{0 = t_0 < t_1 < \ldots < t_n = T\}$, with $\max_{i=0,\ldots,n-1}(t_{i+1} - t_i) =: h = \mathcal{O}(1/n)$.

For a closed interval $\mathcal{C} \subset \mathbb{R}$, we define $p_{\mathcal{C}} : \mathbb{R} \to \mathcal{C}$ as the projection operator onto $\mathcal{C}$. For ease of notation, we define also $p_n = p_{D_n}$, for $x \in \mathbb{R}$,

$$p_n(x) = \begin{cases} n^{-k} \vee x \wedge n^{k'} & , \ D_n = \check{D}_{n^{-k},n^{k'}} \quad \text{if } \alpha > 0, \beta > 0 \\ n^{-k} \vee x & , \ D_n = \tilde{D}_{n^{-k}} \quad \text{if } \alpha = 0, \beta > 0 \\ x \wedge n^{k'} & , \ D_n = \check{D}_{n^{k'}} \quad \text{if } \alpha > 0, \beta = 0 \\ x & , \ D_n = \bar{D} \quad \text{if } \alpha = \beta = 0 \end{cases}. \tag{6.2.5}$$

In the following, we denote by C a constant that depends only on K, T, α, β, y_0, but whose value may change from line to line. We denote it by C_p if it depends on an extra parameter p. We now introduce our explicit scheme for the discretisation process $\hat{Y}$:

Definition 6.2.1. *Set $\hat{Y}_0 := Y_0$ and for $i = 0, \ldots, n-1$,*

$$\hat{Y}_{t_{i+1}} := \hat{Y}_{t_i} + f_n(\hat{Y}_{t_i})h_{i+1} + \bar{\gamma}(\hat{Y}_{t_i})\Delta W_{i+1},$$

with $h_{i+1} := t_{i+1} - t_i$, $\Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}$, $f_n := f \circ p_n$ and $\bar{\gamma} := \gamma \circ p_{\bar{D}}$.

Remark 6.2.3.

(i) For some applications, it may be interesting to force the scheme to take values in a domain, e.g. intervals $\bar{D}$, $\bar{D}_\eta$, $\check{D}_\zeta$ or even $\check{D}_{\eta,\zeta}$. To this end, we introduce some extensions of the previous

scheme. For all $i \le n$, we define $\bar{Y}_{t_i} := p_{\bar{D}}(\hat{Y}_{t_i})$, $\tilde{Y}_{t_i} := p_{\bar{D}_\eta}(\hat{Y}_{t_i})$, $\check{Y}_{t_i} := p_{\check{D}_\zeta}(\hat{Y}_{t_i})$ and $\check{Y}_{t_i} := p_{\check{D}_{\eta,\zeta}}(\hat{Y}_{t_i})$, for some $\eta, \zeta > 0$ to be determined later on, see Corollary 6.3.1 for details. In Proposition 6.3.3, we prove finite moments and finite inverse moments for these modifications.

(ii) *Observe that for $\alpha = \beta = 0$, $\hat{Y}$ is the usual Euler-Maruyama scheme, up to a projection onto $\bar{D}$.*

The following lemma shows how the properties of the initial drift f translate into the new projected drift f_n:

Lemma 6.2.1. *For any $n \in \mathbb{N}^+$, the composition $f_n \equiv f \circ p_n$ is Lipschitz continuous with Lipschitz constant $L(n) = 2K(1 + n^{k\beta}\mathbf{1}_{\{\beta>0\}} + n^{k'\alpha}\mathbf{1}_{\{\alpha>0\}})$, and one-sided Lipschitz continuous with the same constant K as the one-sided Lipschitz continuous constant of f.*

Proof. The fact that f_n is $L(n)$-Lipschitz continuous is straightforward. We prove the one-sided Lipschitz property in two steps below.
Step 1. Let $r > l > 0$ such that $D_n \subset (l, r)$. Assume that f is $C^1(l, r)$. From (6.2.2), we have, for $z, z' \in D_n$, $z > z'$, $\frac{f(z)-f(z')}{z-z'} \le K$, and letting $z' \to z$, we retrieve that $f'(z) \le K$. This shows that $f = g + \ell$, where g is a non-increasing function and ℓ is K-Lipschitz continuous, setting e.g. $g(x) \equiv \int_{\frac{l+r}{2}}^x f'(u)\mathbf{1}_{\{f'(u)\le 0\}}\mathrm{d}u$ and $\ell(x) \equiv \int_{\frac{l+r}{2}}^x f'(u)\mathbf{1}_{\{f'(u)>0\}}\mathrm{d}u$. Since p_n is non-decreasing and 1-Lipschitz on $\mathbb{R}$, we have $f_n = g \circ p_n + \ell \circ p_n$, with $g \circ p_n$ non-increasing and $\ell \circ p_n$ K-Lipschitz continuous on $\mathbb{R}$. This shows that f_n satisfies (6.2.3) as well on $\mathbb{R}$.
Step 2. We now deal with the general case using a smoothing argument. Let $l, r \in D$, $r > l$, such that for all $D_n \subset (l, r)$. We consider a sequence $(\varphi_m)_{m\ge1}$ of mollifiers whose supports are included in $[-\frac{l}{2}, \frac{l}{2}]$ and define $f^m \equiv \varphi_m \star f \equiv \int_{[-\frac{l}{2},\frac{l}{2}]} \varphi_m(u) f(x-u)\mathrm{d}u$ as the convolution of φ_m and f. We observe that, for all $x, y \in (l, r)$,

$$
\begin{aligned}
(x-y)(f^m(x) - f^m(y)) &= \int_{[-\frac{l}{2},\frac{l}{2}]} \varphi_m(u)\{(x-y)(f(x-u) - f(y-u))\}\mathrm{d}u \\
&\le K|x-y|^2 \int_{[-\frac{l}{2},\frac{l}{2}]} \varphi_m(u)\mathrm{d}u \ \le K|x-y|^2 ,
\end{aligned}
$$

where we used (6.2.3) and the fact that $\int_D \varphi_m(u)\mathrm{d}u = 1$. Since f^m is smooth, we can apply Step 1 to obtain, for all $(x, y) \in \mathbb{R}^2$,

$$
(x-y)\left(f^m(p_n(x)) - f^m(p_n(y))\right) \le K|x-y|^2 .
$$

Letting m go to infinity, we then obtain

$$(x-y)\left(f(p_n(x)) - f(p_n(y))\right) \leq K|x-y|^2 ,$$

for all $x, y \in \mathbb{R}$, which concludes the proof. □

Remark 6.2.4. *For any $n \in \mathbb{N}^+$, since f_n and γ are Lipschitz continuous, an easy induction shows that the scheme in Definition 6.2.1 satisfies $\max_{i=0,\dots,n} \|\hat{Y}_{t_i}\|_2 < \infty$. The bound is a priori non-uniform in n, since the Lipschitz constant of f_n depends on n.*

We now introduce the following assumption, which implies that $L(n)^2 h \leq C$, for all $n \in \mathbb{N}^+$, and which relates the locally Lipschitz exponents α and β to the size of the truncated domain D_n:

$(\mathbf{H}p)$: the strictly positive constants k, k' satisfy $2\beta k \leq 1$ and $2\alpha k' \leq 1$.

We require additional assumptions to prove the strong convergence rate of our scheme: below $(\mathbf{H}y1)$ imposes a condition on the moments of the process Y in terms of the locally Lipschitz exponents α and β, to obtain a minimal convergence rate. We shall further impose regularity conditions on f and γ to obtain a better rate of convergence.

$(\mathbf{H}y1)$: $(\mathbf{H}p)$ holds and there exist $q' > 2(\alpha+1)$ and $q > 2\beta$ such that $\mathbb{E}\left[|Y_t|^{q'}\right]$ and $\mathbb{E}[|Y_t|^{-q}]$ are finite for all $t \in [0, T]$.

$(\mathbf{H}y2)$: $(\mathbf{H}y1)$ holds, the drift function f is of class $\mathcal{C}^2(D)$, and

$$\sup_{t\in[0,T]} \mathbb{E}\left[|\gamma(Y_t) f'(Y_t)|^2 + \left| f'(Y_t) f(Y_t) + \frac{\gamma^2(Y_t)}{2} f''(Y_t) \right|^2\right] < \infty. \tag{6.2.6}$$

For an implicit scheme, strong rates of convergence have been derived in [NS12] assuming $(\mathbf{H}y2)$; inspired by this paper, our motivation is to recover strong rates of convergence for the explicit scheme in Definition 6.2.1.

6.3 Convergence results

In this section we prove strong rate of convergence for the scheme in Definition 6.2.1 under some of the assumptions stated above; this result follows from estimates for the regularity of the processes Y and $f(Y)$, and the discretisation error of the scheme. Below, we give the results for the general case $\alpha, \beta \geq 0$, but in the proof we restrict to the most complicated case $\alpha > 0, \beta > 0$.

6.3.1 Preliminary estimates

Our first two results concern the error due to projecting the true solution Y on D_n.

Lemma 6.3.1. *Assume that* $(\mathbf{H}y0)$ *and* $(\mathbf{H}y1)$ *hold. Then, for any* $t \in [0, T]$,

$$\mathbb{E}\Big[|Y_t - p_n(Y_t)|^2\Big] \leq C_{q,q'} \left(\frac{1}{n^{k(q+2)}} \mathbf{1}_{\{\beta>0\}} + \frac{1}{n^{k'(q'-2)}} \mathbf{1}_{\{\alpha>0\}} \right) =: K_1(n, q, q'),$$

where q, q' *are given by* $(\mathbf{H}y1)$.

Proof. For any $t \in [0, T]$, we can write

$$\mathbb{E}\Big[|Y_t - p_n(Y_t)|^2\Big] \leq \frac{1}{n^{2k}} \mathbb{P}\left(Y_t < \frac{1}{n^k}\right) + \mathbb{E}\Big[|Y_t|^2 \mathbf{1}_{\{Y_t > n^{k'}\}}\Big] .$$

Set $\eta = q'/2$ and $\theta = q'/(q'-2)$, its conjugate exponent. Hölder's inequality yields

$$\mathbb{E}\Big[|Y_t|^2 \mathbf{1}_{\{Y_t > n^{k'}\}}\Big] \leq \mathbb{E}\Big[|Y_t|^{q'}\Big]^{1/\eta} \mathbb{P}\{Y_t > n^{k'}\}^{1/\theta}.$$

Using $(\mathbf{H}y1)$ and the set equality $\{Y_t > n^{k'}\} = \{Y_t^{q'} > n^{k'q'}\}$, Markov's inequality implies $\mathbb{E}\Big[|Y_t|^2 \mathbf{1}_{\{Y_t > n^{k'}\}}\Big] \leq C_{q'} n^{-k'(q'-2)}$. Likewise, since $\{Y_t < n^{-k}\} = \{Y_t^{-q} > n^{kq}\}$, Markov's inequality yields $\mathbb{P}(Y_t < n^{-k}) \leq C_q n^{-kq}$, and the lemma follows. □

Lemma 6.3.2. *Assume that* $(\mathbf{H}y0)$ *and* $(\mathbf{H}y1)$ *hold. Then, for any* $t \in [0, T]$,

$$\mathbb{E}\Big[|f(Y_t) - f_n(Y_t)|^2\Big] \leq C_{q,q'} \left(\frac{1}{n^{k(q-2(\beta-1))}} \mathbf{1}_{\{\beta>0\}} + \frac{1}{n^{k'(q'-2(\alpha+1))}} \mathbf{1}_{\{\alpha>0\}} \right) =: K_2(n, q, q'),$$

where q, q' *are given by* $(\mathbf{H}y1)$.

Proof. Using (6.2.2), we observe that

$$\begin{aligned}
|f(Y_t) - f_n(Y_t)|^2 &\leq C\left(1 + |Y_t|^{-2\beta} + |Y_t|^{2\alpha}\right) |Y_t - p_n(Y_t)|^2 \\
&\leq C\left(1 + |Y_t|^{-2\beta}\right) \frac{1}{n^{2k}} \mathbf{1}_{\{Y_t < n^{-k}\}} + C\left(1 + |Y_t|^{2\alpha}\right) |Y_t|^2 \mathbf{1}_{\{Y_t > n^{k'}\}} \\
&:= A_1 + A_2.
\end{aligned}$$

Set $\eta := q/(2\beta)$ and $\theta := q/(q-2\beta)$. Hölder's inequality then yields

$$\mathbb{E}[A_1] \leq \frac{C_q}{n^{2k}} \mathbb{E}\big[|Y_t|^{-q}\big]^{1/\eta} \, \mathbb{P}\{Y_t < n^{-k}\}^{1/\theta},$$

and $(\mathbf{H}y1)$ together with Markov's inequality imply $\mathbb{E}[A_1] \leq C_q n^{-k(q-2(\beta-1))}$. Setting $\eta' := \frac{q'}{2(\alpha+1)}$ and $\theta' := \frac{q'}{q'-2(\alpha+1)}$, a similar computation gives $\mathbb{E}[A_2] \leq C_{q'} n^{-k'(q'-2(\alpha+1))}$. □

The following lemma provides a regularity result for the process Y and will be required for the main convergence result. For a given stochastic process X on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and the partition π, we define its "regularity" by

$$\mathcal{R}_\pi[X] := \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\Big[|X_t - X_{t_i}|^2\Big] \, \mathrm{d}t \, . \tag{6.3.1}$$

Lemma 6.3.3. *Assume that* $(\mathbf{H}y0)$ *and* $(\mathbf{H}y1)$ *hold. The regularity of* Y *satisfies* $\mathcal{R}_\pi[Y] \leq C_{q,q'}h$, *where* q, q' *are given by* $(\mathbf{H}y1)$.

Proof. For $t \in (t_i, t_{i+1}]$, since γ is K-Lipschitz, $(\mathbf{H}y1)$ implies

$$\mathbb{E}\Big[|Y_t - Y_{t_i}|^2\Big] \leq C\mathbb{E}\left[\left(\int_{t_i}^t f(Y_s)\mathrm{d}s\right)^2 + \int_{t_i}^t (|Y_s|^2+1)\mathrm{d}s\right] \leq Ch\left(1 + \frac{1}{h}\mathbb{E}\left[\left(\int_{t_i}^t f(Y_s)\mathrm{d}s\right)^2\right]\right).$$

For $t \in (t_i, t_{i+1}]$, we now compute

$$\begin{aligned}
\frac{1}{h}\mathbb{E}\left[\left(\int_{t_i}^t f(Y_s)\mathrm{d}s\right)^2\right] &\leq \mathbb{E}\left[\int_{t_i}^{t_{i+1}} |f(Y_s)|^2 \mathrm{d}s\right] \\
&\leq 2\left[\int_{t_i}^{t_{i+1}} \mathbb{E}\Big[|f(Y_s) - f_n(Y_s)|^2\Big] \, \mathrm{d}s + \int_{t_i}^{t_{i+1}} \mathbb{E}\Big[|f_n(Y_s)|^2\Big] \, \mathrm{d}s\right] \\
&\leq Ch\left(K_2(n,q,q') + L(n)^2 \sup_{t\in[t_i,t_{i+1}]} \mathbb{E}\Big[1 + |Y_t|^2\Big]\right).
\end{aligned}$$

Using $(\mathbf{H}y1)$ and the inequality $L(n)^2 h \leq C$, which holds under $(\mathbf{H}p)$, we obtain $\mathbb{E}\big[|Y_t - Y_{t_i}|^2\big] \leq C_{q,q'}h$ for $t \in (t_i, t_{i+1}]$, and the lemma follows from the upper bound

$$\mathcal{R}_\pi[Y] = \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}\Big[|Y_t - Y_{t_i}|^2\Big] \, \mathrm{d}t \leq C \max_{i=0,\ldots,n-1} \sup_{t\in[t_i,t_{i+1}]} \mathbb{E}\Big[|Y_t - Y_{t_i}|^2\Big] \leq C_{q,q'}h \, .$$

□

We now compute upper bounds for the regularity of $f(Y)$.

Lemma 6.3.4. *Assume that* $(\mathbf{H}y0)$ *and* $(\mathbf{H}y1)$ *hold.*

(i) Then $\mathcal{R}_\pi[f(Y)] \leq C\left(K_2(n,q,q') + L(n)^2 h\right)$, *where* q, q' *are given by* $(\mathbf{H}y1)$.

(ii) If moreover $(\mathbf{H}y2)$ *holds, then* $\mathcal{R}_\pi[f(Y)] \leq Ch$.

Proof. The inequality in (i) is a direct consequence of the following computation:

$$\begin{aligned}\int_{t_i}^{t_{i+1}} \mathbb{E}\Big[|f(Y_t) - f(Y_{t_i})|^2\Big]\,\mathrm{d}t \leq C\Big(& \int_{t_i}^{t_{i+1}} \mathbb{E}\Big[|f(Y_t) - f_n(Y_t)|^2\Big]\,\mathrm{d}t \\ & + \int_{t_i}^{t_{i+1}} \mathbb{E}\Big[|f_n(Y_t) - f_n(Y_{t_i})|^2\Big]\,\mathrm{d}t \\ & + h\mathbb{E}\Big[|f_n(Y_{t_i}) - f(Y_{t_i})|^2\Big]\Big) \leq Ch\left(K_2(n,q,q') + L(n)^2 h\right),\end{aligned}$$

where we used Lemma 6.3.2, Lemma 6.3.3, and $(\mathbf{H}p)$. Let us now prove (ii). The drift function f is of class $\mathcal{C}^2(D)$ by $(\mathbf{H}y2)$, and Itô's Formula on the interval $[t_i, t_{i+1}]$ implies

$$f(Y_{t_{i+1}}) - f(Y_{t_i}) = \int_{t_i}^{t_{i+1}} \left(f'(Y_t)f(Y_t) + \frac{1}{2}f''(Y_t)\gamma(Y_t)^2\right)\mathrm{d}t + \int_{t_i}^{t_{i+1}} f'(Y_t)\gamma(Y_t)\mathrm{d}W_t.$$

Squaring and applying the Cauchy-Schwarz inequality the yields

$$\mathbb{E}\Big[|f(Y_{t_{i+1}}) - f(Y_{t_i})|^2\Big] \leq \int_{t_i}^{t_{i+1}} \mathbb{E}\left[|\gamma(Y_t)f'(Y_t)|^2 + h\left|f'(Y_t)f(Y_t) + \frac{\gamma^2(Y_t)}{2}f''(Y_t)\right|^2\right]\mathrm{d}t,$$

and (ii) follows from (6.2.6), direct integration on $[t_i, t_{i+1}]$ and summation. □

6.3.2 Convergence result

We consider here the discretisation error between the true process Y and the discretised process $\hat{Y}$. Let us introduce the following notations:

$$\delta Y_i := Y_{t_i} - \hat{Y}_{t_i}, \qquad \delta_n f_i := f_n(Y_{t_i}) - f_n(\hat{Y}_{t_i}), \qquad \delta\gamma_i := \gamma(Y_{t_i}) - \bar{\gamma}(\hat{Y}_{t_i}). \tag{6.3.2}$$

The following key proposition provides a bound on the squared differences $|\delta Y_i|^2$, which depends on both the partition size and the regularity (in the sense of (6.3.1)), and which will be refined further below in Theorem 6.3.1.

Proposition 6.3.1. *Assume that* $(\mathbf{H}y0)$ *and* $(\mathbf{H}y1)$ *hold, then*

$$\max_{i=0,\dots,n} \mathbb{E}\Big[|\delta Y_i|^2\Big] \le C\left(K_2(n,q,q') + \mathcal{R}_\pi[f(Y)] + \mathcal{R}_\pi[Y]\right), \tag{6.3.3}$$

where q, q' *are given by* $(\mathbf{H}y1)$.

Proof. 1. We first show that the global error between the scheme and the solution is controlled by the sum of local truncation errors defined below. Indeed, observe that

$$Y_{t_{i+1}} = Y_{t_i} + f_n(Y_{t_i})h_{i+1} + \bar{\gamma}(Y_{t_i})\Delta W_{i+1} + \zeta^d_{i+1} + \zeta^w_{i+1},$$

for $i \le n-1$, where

$$\zeta^d_{i+1} := \int_{t_i}^{t_{i+1}} \left(f(Y_t) - f_n(Y_{t_i})\right) \mathrm{d}t,$$
$$\zeta^w_{i+1} := \int_{t_i}^{t_{i+1}} \left(\gamma(Y_t) - \bar{\gamma}(Y_{t_i})\right) \mathrm{d}W_t = \int_{t_i}^{t_{i+1}} \left(\gamma(Y_t) - \gamma(Y_{t_i})\right) \mathrm{d}W_t.$$

The last equality comes from the fact that Y takes values in D and $\bar{\gamma}(Y_{t_i}) = \gamma(Y_{t_i})$, for all $i \le n$. Therefore, squaring the difference δY_{i+1} gives

$$|\delta Y_{i+1}|^2 = |\delta Y_i|^2 + 2\delta Y_i \delta_n f_i h_{i+1} + 2\delta Y_i \delta\gamma_i \Delta W_{i+1} + 2\delta Y_i \zeta^d_{i+1} + 2\delta Y_i \zeta^w_{i+1}$$
$$+ |\delta_n f_i h_{i+1} + \delta\gamma_i \Delta W_{i+1} + \zeta^d_{i+1} + \zeta^w_{i+1}|^2 .$$

Using the simple identity $\mathbb{E}_{t_i}[2\delta Y_i \delta\gamma_i \Delta W_{i+1} + 2\delta Y_i \zeta^w_{i+1}] = 0$ and an application of Young's inequality yields

$$\mathbb{E}\Big[|\delta Y_{i+1}|^2\Big] \le (1+Ch)\mathbb{E}\Big[|\delta Y_i|^2\Big] + C\mathbb{E}\left[|\delta_n f_i h_{i+1}|^2 + |\delta\gamma_i|^2 h_{i+1} + \frac{|\mathbb{E}_{t_i}[\zeta^d_{i+1}]\,|^2}{h} + |\zeta^d_{i+1}|^2 + |\zeta^w_{i+1}|^2\right]$$
$$\le \left(1 + Ch + CL(n)^2h^2\right)\mathbb{E}\Big[|\delta Y_i|^2\Big] + C\mathbb{E}\left[\frac{\left(\mathbb{E}_{t_i}[\zeta^d_{i+1}]\right)^2}{h} + |\zeta^d_{i+1}|^2 + |\zeta^w_{i+1}|^2\right],$$

since f_n is one-sided Lipschitz continuous (Lemma 6.2.1), locally Lipschitz continuous with Lipschitz constant $L(n)$ and γ is Lipschitz continuous. Under $(\mathbf{H}p)$, $L(n)^2 h \le C$ and an iteration yields

$$\max_{i=0,\dots,n} \mathbb{E}\Big[|\delta Y_i|^2\Big] \le C \sum_{j=1}^{n} \mathbb{E}\left[\frac{\left(\mathbb{E}_{t_j}\big[\zeta_j^d\big]\right)^2}{h} + |\zeta_j^d|^2 + |\zeta_j^w|^2\right] \tag{6.3.4}$$

$$\le C \sum_{j=1}^{n} \mathbb{E}\left[\frac{|\zeta_j^d|^2}{h} + |\zeta_j^w|^2\right]. \tag{6.3.5}$$

2. We now provide explicit errors for the global truncation. As γ is K-Lipschitz, we have $\mathbb{E}\big[|\zeta_{i+1}^w|^2\big] \le C\int_{t_i}^{t_{i+1}} \mathbb{E}\big[|Y_t - Y_{t_i}|^2\big]\,\mathrm{d}t$, and hence

$$\sum_{i=1}^{n} \mathbb{E}\Big[|\zeta_i^w|^2\Big] \le C\mathcal{R}_\pi[Y]. \tag{6.3.6}$$

We now compute an upper bound for $\mathbb{E}\big[|\zeta_{i+1}^d|^2\big]$. Since

$$\zeta_{i+1}^d := \int_{t_i}^{t_{i+1}} (f(Y_t) - f_n(Y_{t_i}))\mathrm{d}t = \int_{t_i}^{t_{i+1}} (f(Y_t) - f(Y_{t_i}))\mathrm{d}t + \int_{t_i}^{t_{i+1}} (f(Y_{t_i}) - f_n(Y_{t_i}))\mathrm{d}t, \tag{6.3.7}$$

The Cauchy-Schwarz inequality yields

$$\mathbb{E}\Big[|\zeta_{i+1}^d|^2\Big] \le Ch\left(\int_{t_i}^{t_{i+1}} \mathbb{E}\Big[|f(Y_t) - f(Y_{t_i})|^2\Big]\,\mathrm{d}t + h\mathbb{E}\Big[|f(Y_{t_i}) - f_n(Y_{t_i})|^2\Big]\right),$$

and Lemma 6.3.2 implies $\mathbb{E}\big[|\zeta_{i+1}^d|^2\big] \le Ch(\int_{t_i}^{t_{i+1}} \mathbb{E}\big[|f(Y_t) - f(Y_{t_i})|^2\big]\,\mathrm{d}t + hK_2(n,q,q'))$ and $\frac{1}{h}\sum_{i=1}^{n} \mathbb{E}\big[|\zeta_i^d|^2\big] \le C\left(K_2(n,q,q') + \mathcal{R}_\pi[f(Y)]\right)$. Combining this with (6.3.5) and (6.3.6) concludes the proof. □

We have kept the above result general, without *a priori* assuming that the drift function belongs to $C^2(D)$. If we consider a constant diffusion and $(\mathbf{H}y2)$, we can recover a better upper bound using (6.3.4) instead of (6.3.5) in the first part of the previous proof and prove a first-order strong rate of convergence. This will be illustrated in Proposition 6.3.2 below.
We now state the main result of our paper, namely a strong rate for δY_i defined in (6.3.2).

Theorem 6.3.1. *Assume that* $(\mathbf{H}y0)$ *holds, then the inequality*

$$\max_{i=0,\dots,n} \|\delta Y_i\|_2 \le C_{q,q'} h^r \tag{6.3.8}$$

holds with $r = \min(\frac{1}{2} - \frac{\beta}{q+2}, \frac{1}{2} - \frac{\alpha}{q'-2}) > 0$ *under* $(\mathbf{H}y1)$ *by setting* $(k,k') = (\frac{1}{q+2}, \frac{1}{q'-2})$ *and*

$r = \min(\frac{1}{2}, \frac{q+2}{4\beta} - \frac{1}{2}, \frac{q'-2}{4\alpha} - \frac{1}{2}) > 0$ *under* $(\mathbf{H}y2)$ *by setting* $(k,k') = (\frac{1}{2\beta}, \frac{1}{2\alpha})$.

Proof. 1. Assume $(\mathbf{H}y1)$. Combining Lemma 6.3.3 and Lemma 6.3.4(i) with (6.3.3) yields

$$\begin{aligned}\max_{i=0,\dots,n} \mathbb{E}\Big[|\delta Y_i|^2\Big] &\le C(K_2(n,q,q') + L(n)^2 h + h);\\ &\le C_{q,q'}(h^{1-2\beta k} + h^{k(q+2)-2\beta k} + h^{1-2\alpha k'} + h^{k'(q'-2)-2\alpha k'} + h)\ .\end{aligned}$$

To balance the error terms, set $k = \frac{1}{q+2}$ and $k' = \frac{1}{q'-2}$, observing that under $(\mathbf{H}y1)$, $(\mathbf{H}p)$ holds for this choice of parameters. Thus, we obtain $\max_{i=0,\dots,n} \|\delta Y_i\|_2 \le C_{q,q'}h^r$, with $r = \min(\frac{1}{2} - \frac{\beta}{q+2}, \frac{1}{2} - \frac{\alpha}{q'-2})$, with $r > 0$.

2. Assume $(\mathbf{H}y2)$. Lemma 6.3.3 and Lemma 6.3.4(ii) with (6.3.3) imply

$$\max_{i=0,\dots,n} \mathbb{E}\Big[|\delta Y_i|^2\Big] \le C(K_2(n,q,q') + h)\ .$$

Setting $k = \frac{1}{2\beta}, k' = \frac{1}{2\alpha}$ yields $\max_{i=0,\dots,n} \|\delta Y_i\|_2 \le C_{q,q'}h^r$, where $r = \min(1/2, \frac{q+2}{4\beta} - 1/2, \frac{q'-2}{4\alpha} - 1/2)$. Since $(\mathbf{H}y2)$ implies $(\mathbf{H}y1)$, we observe that $r > 0$. $\square$

We now state the convergence results associated to the extensions of the scheme defined in Remark 6.2.3.

Corollary 6.3.1. *Assume that* $(\mathbf{H}y0)$ *holds. Then the approximations* $(\tilde{Y}_{t_i})_{i\le n}$ *and* $(\check{Y}_{t_i})_{i\le n}$ *defined in Remark 6.2.3 satisfy*

$$\max_{i=0,\dots,n} \big(\|Y_{t_i} - \bar{Y}_{t_i}\|_2 + \|Y_{t_i} - \tilde{Y}_{t_i}\|_2 + \|Y_{t_i} - \check{Y}_{t_i}\|_2\big) \le C_{q,q'}h^r,$$

holds with $r = \min(\frac{1}{2} - \frac{\beta}{q+2}, \frac{1}{2} - \frac{\alpha}{q'-2}) > 0$ *under* $(\mathbf{H}y1)$ *by setting* $(k,k') = (\frac{1}{q+2}, \frac{1}{q'-2})$ *and* $r = \min(\frac{1}{2}, \frac{q+2}{4\beta} - \frac{1}{2}, \frac{q'-2}{4\alpha} - \frac{1}{2}) > 0$ *under* $(\mathbf{H}y2)$ *by setting* $(k,k') = (\frac{1}{2\beta}, \frac{1}{2\alpha})$, *where* $\eta := h^{2r/q}$ *and* $\zeta := h^{-2r/(q'-2)}$.

Proof. The proof follows by computing upper bounds for each of the three quantities on the left-hand side. For all $i \le n$, since $p_{\bar{D}}$ is 1-Lipschitz continuous, we can write

$$\mathbb{E}\Big[|Y_{t_i} - \bar{Y}_{t_i}|^2\Big] = \mathbb{E}\Big[|p_{\bar{D}}(Y_{t_i}) - p_{\bar{D}}(\hat{Y}_{t_i})|^2\Big] \le \mathbb{E}\Big[|Y_{t_i} - \hat{Y}_{t_i}|^2\Big] = \mathbb{E}\,|\delta Y_i|^2\,,$$

and the upper bound for $\|Y_{t_i} - \bar{Y}_{t_i}\|_2$ follows from Theorem 6.3.1.

Set now $\eta = h^{2r/q}$. For $i \le n$,

$$\begin{aligned}\mathbb{E}\Big[|Y_{t_i} - \tilde{Y}_{t_i}|^2\Big] &\le 2\left(\mathbb{E}\Big[|Y_{t_i} - p_{\bar{D}_\eta}(Y_{t_i})|^2\Big] + \mathbb{E}\Big[|p_{\bar{D}_\eta}(Y_{t_i}) - p_{\bar{D}_\eta}(\hat{Y}_{t_i})|^2\Big]\right)\\ &\le 2\left(\mathbb{E}\Big[|Y_{t_i} - p_{\bar{D}_\eta}(Y_{t_i})|^2\Big] + \mathbb{E}\Big[|Y_{t_i} - \hat{Y}_{t_i}|^2\Big]\right)\\ &\le C_{q,q'}\left(\mathbb{E}\Big[|Y_{t_i} - p_{\bar{D}_\eta}(Y_{t_i})|^2\Big] + h^{2r}\right), \end{aligned} \tag{6.3.9}$$

where the last inequality follows from Theorem 6.3.1. A straightforward adaptation of the proof of Lemma 6.3.1 yields $\mathbb{E}\Big[|Y_{t_i} - p_{\bar{D}_\eta}(Y_{t_i})|^2\Big] \le C_q\eta^q$, which gives the second bound. Similarly, for $i \le n$, the equality $\mathbb{E}[|Y_{t_i} - p_{\check{D}_\zeta}(Y_{t_i})|^2] = \mathbb{E}[|Y_{t_i} - \zeta|^2 1_{\{Y_{t_i} > \zeta\}}]$ holds, and an application of Hölder's inequality gives $\mathbb{E}[|Y_{t_i} - p_{\check{D}_\zeta}(Y_{t_i})|^2] \le C_{q'}\zeta^{-(q'-2)}$. Choosing $\zeta = h^{-2r/(q'-2)}$ concludes the proof. □

Remark 6.3.1. *For SDEs defined on the whole real line, strong convergence rates have been proved using tamed explicit schemes [HJK12, Sab13]. The authors assumed that the drift satisfies* (6.2.2) *and* (6.2.3) *with locally Lipschitz exponents $\alpha \in (0, \infty)$, $\beta = 0$, $D = \mathbb{R}$ and that the diffusion is K-Lipschitz. Under these assumptions,* (6.2.1) *has a unique strong solution [Kry90]. Our modified scheme and a slight modification of the projection, namely, $p_n(x) \equiv -n^{k'} \vee x \wedge n^{k'}$ can be applied to cover this case.*

We now show that, as for the classical Euler scheme, our modified scheme may have a first-order strong rate of convergence if the diffusion coefficient is constant. This can be observed in practice, as shown in Section 6.5.1. This also suggests that a similarly modified Milstein scheme, when the diffusion coefficient is not constant, will have a first-order strong rate of convergence.

Proposition 6.3.2. *Assume that $\gamma(x) \equiv \gamma > 0$ for all $x \in D$, and that $(\mathbf{H}y0)$ and $(\mathbf{H}y2)$ hold, with $q > 6\beta - 2$ and $q' > 6\alpha + 2$. Then,*

$$\max_{i=0,\dots,n}\left(\|\delta Y_i\|_2 + \|Y_{t_i} - \bar{Y}_{t_i}\|_2 + \|Y_{t_i} - \tilde{Y}_{t_i}\|_2 + \|Y_{t_i} - \check{Y}_{t_i}\|_2\right) \le C_{q,q'}h\,,$$

where we set $\eta := h^{2/q}$ and $\zeta := h^{-2/(q'-2)}$ in the definition of $\tilde{Y}$ and $\check{Y}$.

Proof. The proof is similar to Step 2 in the proof of Proposition 6.3.1, but uses the sharper upper bound (6.3.4). Since the diffusion function is constant, $\sum_{i=1}^n \mathbb{E}\Big[|\zeta_i^w|^2\Big]$ is null, and using (6.3.7), we can write

$$\max_i \mathbb{E}\Big[|\delta Y_i|^2\Big] \le \sum_{i=0}^{n-1} \mathbb{E}\left[|\zeta_{i+1}^d|^2 + \frac{(\mathbb{E}_{t_i}[\zeta_{i+1}^d])^2}{h}\right] \tag{6.3.10}$$

$$\le K_2(n,q,q') + \sum_{i=0}^{n-1} \mathbb{E}\left[\left|\int_{t_i}^{t_{i+1}} (f(Y_t) - f(Y_{t_i}))\mathrm{d}t\right|^2 + \frac{1}{h}\left(\mathbb{E}_{t_i}\Big[\int_{t_i}^{t_{i+1}} (f(Y_t) - f(Y_{t_i}))\mathrm{d}t\Big]\right)^2\right].$$

Moreover, Itô's Lemma implies

$$\int_{t_i}^{t_{i+1}} (f(Y_t) - f(Y_{t_i}))\mathrm{d}t = \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^{t} f'(Y_u)f(Y_u) + \frac{1}{2}f''(Y_u)\gamma^2\mathrm{d}u + \int_{t_i}^{t} f'(Y_u)\gamma \mathrm{d}W_u\right)\mathrm{d}t$$

which we can rewrite as

$$\int_{t_i}^{t_{i+1}} \left(\int_{t_i}^{t} f'(Y_u)f(Y_u) + \frac{1}{2}f''(Y_u)\gamma^2\mathrm{d}u\right)\mathrm{d}t + \int_{t_i}^{t_{i+1}} (t_{i+1} - t)f'(Y_t)\gamma \mathrm{d}W_t.$$

Under $(\mathbf{H}y2)$, we then obtain easily, recalling (6.3.10), that

$$\max_i \mathbb{E}\Big[|\delta Y_i|^2\Big] \le C(K_2(n,q,q') + h^2)\,.$$

The proposition then follows by setting $(k,k') = (\frac{1}{2\beta}, \frac{1}{2\alpha})$ and using the fact that $q > 6\beta - 2$ and $q' > 6\alpha + 2$, from Lemma 6.3.2.

The statement for $\|Y_{t_i} - \bar{Y}_{t_i}\|_2$, $\|Y_{t_i} - \tilde{Y}_{t_i}\|_2$, $\|Y_{t_i} - \check{Y}_{t_i}\|_2$, follows from the same arguments as in Corollary 6.3.1. □

6.3.3 Moment properties of the schemes

For later use, we show that our approximations have uniformly bounded second moments, which completes the result of Remark 6.2.4.

Lemma 6.3.5. *Assume that* $(\mathbf{H}y0)$ *and* $(\mathbf{H}y1)$ *hold. Then, for* q, q' *given by* $(\mathbf{H}y1)$*,*

$$\max_{i=0,\dots,n} \mathbb{E}\Big[|\hat{Y}_{t_i}|^2 + |\bar{Y}_{t_i}|^2 + |\check{Y}_{t_i}|^2 + |\tilde{Y}_{t_i}|^2\Big] \le C_{q,q'}$$

with for $\check{Y}$, $\zeta := h^{-2r/(q'-2)}$ *and for* $\tilde{Y}$, $\eta := h^{2r/q}$, *recall Remark 6.2.3, and with* $r = \min(\frac{1}{2} - \frac{\beta}{q+2}, \frac{1}{2} - \frac{\alpha}{q'-2}) > 0$, *under* $(\mathbf{H}y2)$ $r = \min(\frac{1}{2}, \frac{q+2}{4\beta} - \frac{1}{2}, \frac{q'-2}{4\alpha} - \frac{1}{2}) > 0$, *and if moreover,* $q > 6\beta - 2$, $q' > 6\alpha + 2$ *and* $\gamma(\cdot) \equiv \gamma > 0$, $r = 1$.

Proof. Since $|\hat{Y}_i|^2 \le 2(|Y_{t_i} - \hat{Y}_{t_i}|^2 + |Y_{t_i}|^2)$, $(\mathbf{H}y1)$ and Theorem 6.3.1 imply that

$$\mathbb{E}\Big[|\hat{Y}_{t_i}|^2\Big] \le 2\left(\mathbb{E}\Big[|Y_{t_i} - \hat{Y}_{t_i}|^2\Big] + \mathbb{E}\Big[|Y_{t_i}|^2\Big]\right) \le C_{q,q'}(h^{2r}+1) \le C_{q,q'}$$

holds for any $i \le n$, which proves the claim.
The statement for $\bar{Y}$, $\check{Y}$ and $\tilde{Y}$ follows from Corollary 6.3.1 or Proposition 6.3.2. □

We now consider the modifications $\tilde{Y}$ and $\check{Y}$ defined in Remark 6.2.3 and prove some finite moments or inverse moments for them, extending the previous result.

Proposition 6.3.3. *Assume that* $(\mathbf{H}y0)$ *hold and let* $\zeta := h^{-2r/(q'-2)}$ *and* $\eta := h^{2r/q}$, *where* q *and* q' *are given by* $(\mathbf{H}y1)$.

(i) if $(\mathbf{H}y1)$ *holds, then* $\max_{i=0,\dots,n} \mathbb{E}\Big[\check{Y}_{t_i}^p\Big] \le C_{p,q,q'}$ *for all* $p \in [1,(q'-1)\vee 2]$;

(ii) if $(\mathbf{H}y1)$ *holds with* $q \ge 4$, *then* $\max_{i=0,\dots,n} \mathbb{E}\Big[\tilde{Y}_{t_i}^{-p}\Big] \le C_{p,q,q'}$ *for all* $p \in [1, q-3]$.

Proof. 1. We first prove (i). We remark that the result for $p \in [1,2]$ follows directly from Lemma 6.3.5. We now assume that $1 < p \le q-1$ and we introduce the sets $A := \{Y_{t_i} \le \zeta\}$ and $B := \{|\delta Y_{t_i}| > 1\}$, where $\delta Y := \check{Y} - Y$. We then observe that

$$\check{Y}_{t_i}^p = \check{Y}_{t_i}^p \mathbf{1}_{A^c} + \check{Y}_{t_i}^p \mathbf{1}_{A\cap B^c} + \check{Y}_{t_i}^p \mathbf{1}_{A\cap B}$$

and deal which each terms in the right hand side separately.
Since $\check{Y} \le \zeta$ by definition, we compute, for the first term,

$$\mathbb{E}\Big[\check{Y}_{t_i}^p \mathbf{1}_{A^c}\Big] \le \mathbb{E}\Big[Y_{t_i}^p\Big] \le C_p\ . \tag{6.3.11}$$

For the second term, as $|\delta Y_{t_i}| \le 1$ on B^c, we obtain

$$\mathbb{E}\Big[\check{Y}_{t_i}^p \mathbf{1}_{A\cap B^c}\Big] \le C_p(1 + \mathbb{E}\Big[Y_{t_i}^p\Big]) \le C_p. \tag{6.3.12}$$

For the last term, we first observe that for non negative y, y' and $\theta \ne 1$,

$$(y')^\theta - y^\theta = \theta \int_0^1 \big((1-\lambda)y + \lambda y'\big)^{\theta-1} \mathrm{d}\lambda (y'-y). \tag{6.3.13}$$

Using the above equality for $y' = \check{Y}_{t_i}$, $y' = Y_{t_i}$ and $\theta = p$ we compute that

$$|\check{Y}_{t_i}^p - Y_{t_i}^p| \le C_p(\check{Y}_{t_i}^{p-1} + Y_{t_i}^{p-1})|\delta Y_{t_i}|\ .$$

Then since, $\check{Y}_{t_i}^p \mathbf{1}_{A\cap B} \le Y_{t_i}^p + |\check{Y}_{t_i}^p - Y_{t_i}^p| \mathbf{1}_{A\cap B}$, we observe that

$$\mathbb{E}\Big[\check{Y}_{t_i}^p \mathbf{1}_{A\cap B}\Big] \le C_p(1+\zeta^{p-1})\mathbb{E}\Big[|\delta Y_{t_i}| \mathbf{1}_{|\delta Y_{t_i}|>1}\Big] \le C_p(1+\zeta^{p-1})\mathbb{E}\Big[|\delta Y_{t_i}|^2\Big]$$

Applying Corollary 6.3.1, we thus obtain

$$\mathbb{E}\Big[\check{Y}_{t_i}^p \mathbf{1}_{A\cap B}\Big] \le C_p(1+\zeta^{p-1}h^{2r}) \le C_p \,. \tag{6.3.14}$$

The proof of the first statement is concluded by combining the previous inequality with (6.3.11) and (6.3.12).

2. We now prove (ii). We assume that $p \in [1, q-3]$ and that $q \ge 4$. We introduce the set $A = \{Y_{t_i} \ge \eta\}$ and $B = \{|\delta Y| > \eta^2\}$, where $\delta Y := \tilde{Y} - Y$. We observe that

$$\tilde{Y}_{t_i}^{-p} = \tilde{Y}_{t_i}^{-p} \mathbf{1}_{A^c} + \tilde{Y}_{t_i}^{-p} \mathbf{1}_{A\cap B^c} + \tilde{Y}_{t_i}^{-p} \mathbf{1}_{A\cap B} \,.$$

We are going to upper bound separately the expectation of each terms appearing in the right hand side of the above equality.

For the first term, since on A^c, $Y_{t_i} \le \tilde{Y}_{t_i}$ holds by definition, we get

$$\mathbb{E}\Big[\tilde{Y}_{t_i}^{-p} \mathbf{1}_{A^c}\Big] \le \mathbb{E}\Big[Y_{t_i}^{-p} \mathbf{1}_{A^c}\Big] \le C_p \,.$$

For the second term, observing that $\frac{1}{Y_{t_i}} - \frac{1}{\tilde{Y}_{t_i}} = \frac{\delta Y_{t_i}}{Y_{t_i}\tilde{Y}_{t_i}}$, we compute

$$\mathbb{E}\Big[\tilde{Y}_{t_i}^{-p} \mathbf{1}_{A\cap B^c}\Big] \le C_p \mathbb{E}\left[Y_{t_i}^{-p} + \left|\frac{\delta Y_{t_i}}{Y_{t_i}\tilde{Y}_{t_i}}\right|^p \mathbf{1}_{A\cap B^c}\right] \le C_p,$$

since on $A \cap B^c$, $|\delta Y_{t_i}| \le \eta^2$ and $\frac{1}{Y_{t_i}} \le \frac{1}{\eta}$. For the last term, we compute that

$$\mathbb{E}\Big[\tilde{Y}_{t_i}^{-p} \mathbf{1}_{A\cap B}\Big] \le C_p \mathbb{E}\Big[Y_{t_i}^{-p} + |\tilde{Y}_{t_i}^{-p} - Y_{t_i}^{-p}| \mathbf{1}_{A\cap B}\Big]$$

and using (6.3.13), we get

$$\begin{aligned}\mathbb{E}\Big[\tilde{Y}_{t_i}^{-p} \mathbf{1}_{A\cap B}\Big] &\le C_p(1+\mathbb{E}\Big[(\tilde{Y}_{t_i}^{-p-1} + Y_{t_i}^{-p-1})|\delta Y_{t_i}| \mathbf{1}_{A\cap B}\Big] \\ &\le C_p(1+\eta^{-(p+1)})\mathbb{E}\Big[|\delta Y_{t_i}| \mathbf{1}_{\{|\delta Y_{t_i}|>\eta^2\}}\Big] \,.\end{aligned}$$

Using the Cauchy-Schwarz inequality and then applying Chebyshev's inequality, we obtain

$$\mathbb{E}\Big[\tilde{Y}_{t_i}^{-p}\mathbf{1}_{A\cap B}\Big] \leq C_p(1+\eta^{-(p+3)})h^{2r} \leq C_p\ ,$$

which concludes the proof for this step. □

6.4 Applications to financial SDEs

We now apply our results to various stochastic differential equations widely used in the literature.

6.4.1 CIR model

We consider the Feller diffusion [Fel54], defined as the unique strong solution to

$$\mathrm{d}X_t = \kappa(\theta - X_t)\mathrm{d}t + \xi\sqrt{X_t}\mathrm{d}W_t, \qquad X_0 = x_0 > 0, \tag{6.4.1}$$

where W is a Brownian motion, and κ, θ, ξ are strictly positive constant parameters. This process has been widely used in the mathematical finance literature, both for interest rate modelling [CIR85] and for the instantaneous variance of a stock price process [Hes93]. Under the Feller condition $\omega := 2\kappa\theta/\xi^2 > 1$, X remains strictly positive almost surely, and Itô's Lemma implies that the Lamperti transform $Y = \sqrt{X}$ satisfies

$$\mathrm{d}Y_t = f(Y_t)\mathrm{d}t + c\,\mathrm{d}W_t, \qquad Y_0 = \sqrt{x_0} > 0, \tag{6.4.2}$$

where

$$f(x) \equiv a/x + bx, \qquad a := (4\kappa\theta - \xi^2)/8, \qquad b := -\kappa/2, \qquad c := \xi/2; \tag{6.4.3}$$

furthermore, $a > 0$ when the Feller condition holds. Since $X = Y^2$, proving a rate of convergence for a discretisation scheme for the process Y will allow us to obtain a rate of convergence for the process X. In the following corollary, we apply Theorem 6.3.1 to provide bounds for $\|\delta Y_i\|_2$ and $\|\delta X_i\|_1$, where $\delta X_i := X_{t_i} - \hat{X}_{t_i} = Y_{t_i}^2 - \hat{Y}_{t_i}^2$.

Corollary 6.4.1. *For $\omega > 2$, $\max_{i=0,\dots,n}\left(\|\delta Y_i\|_2 + \|\delta X_i\|_1\right) \le C_r h^r$ holds, where*

$$\begin{cases} r \in \left(\dfrac{1}{6}, \dfrac{1}{2} - \dfrac{1}{\omega+1}\right), & \text{if } 2 < \omega \le 3, \\ r = 1/2, & \text{if } 3 < \omega \le 5, \\ r = 1, & \text{if } \omega > 5. \end{cases} \tag{6.4.4}$$

Proof. Consider first the bound for $\|\delta Y_i\|_2$. The drift of Y is one-sided Lipschitz continuous and locally Lipschitz continuous with exponents $\alpha = 0$ and $\beta = 2$, and the diffusion is constant, hence Lipschitz continuous. From [DNS12, page 5], we know that $\sup_{t\in[0,T]} \mathbb{E}(|X_t|^p) < +\infty$ for all $p > -2\kappa\theta/\xi^2$, and therefore

$$\sup_{t\in[0,T]} \mathbb{E}(|Y_t|^{-\ell}) < +\infty \ \text{ for all } \ \ell < 4\kappa\theta/\xi^2 = 2\omega. \tag{6.4.5}$$

In the case $2 < \omega \le 3$, we choose $q \in (4, 2\omega)$ and fix $k = 1/(q+2)$, so that $(\mathbf{H}p)$ holds (no condition on k' is required since $\alpha = 0$) and $(\mathbf{H}y1)$ holds as well. From Theorem 6.3.1 it follows that the convergence rate is given by $r := 1/2 - \beta/(q+2)$. We compute easily, since $\beta = 2$, that $r \in (\frac{1}{6}, \frac{1}{2} - \frac{1}{\omega+1})$, depending on the choice of $q \in (4, 2\omega)$.
Consider now the case $3 < \omega$. We compute that $\mathbb{E}(|f(Y_t)f'(Y_t) + \frac{1}{2}c^2 f''(Y_t)|^2) \le C\mathbb{E}(|Y_t|^2 + |Y_t|^{-6}) \le C$ hold. Combining the previous inequality with (6.4.5), we obtain that $(\mathbf{H}y2)$ holds. Fix $q \in (6, 2\omega)$ and set $k = 1/4$, it follows that $r = \min(1/2, (q+2)/8 - 1/2) = 1/2$ from Theorem 6.3.1. The case $\omega > 5$ follows directly from Proposition 6.3.2.
We now prove the corollary for the difference δX_i. The Cauchy-Schwarz inequality and the result above imply

$$\begin{aligned} \mathbb{E}[|\delta X_i|] = \mathbb{E}\left[|(Y_{t_i} - \hat{Y}_{t_i})(Y_{t_i} + \hat{Y}_{t_i})|\right] &\le \sqrt{\mathbb{E}(|\delta Y_i|^2)\mathbb{E}\left[|Y_{t_i} + \hat{Y}_{t_i}|^2\right]} \\ &\le C_r h^r \sqrt{\mathbb{E}(|Y_{t_i}|^2) + \mathbb{E}(|\hat{Y}_{t_i}|^2)} \le C_r h^r, \end{aligned}$$

since $\mathbb{E}(|Y_{t_i}|^2)$ and $\mathbb{E}(|\hat{Y}_{t_i}|^2)$ are finite from [HMS02, Lemma 3.2] and Lemma 6.3.5. □

Define $\delta\check{X}_i := X_{t_i} - \check{X}_{t_i}$, where $\check{X}_{t_i} := \check{Y}_{t_i}^2$, recall Remark 6.2.3. We now consider a general $L^{1+\varepsilon}$-norm for convergence of the discretisation scheme of process X.

Corollary 6.4.2. *Suppose that $\omega > 2$ and fix $\varepsilon \ge 0$.Then*

$$\max_{i=0,\dots,n} \|\delta\check{X}_i\|_{1+\varepsilon} \le C_{r,\varepsilon} h^{r/(1+\varepsilon)},$$

with r defined as in (6.4.4) *and where we set* $\zeta := h^{-\frac{2r}{q'-2}}$, *with* $q' = 3 + 4\epsilon$ *in the definition of* $\check{X} = \check{Y}^2$, *recall Remark 6.2.3.*

Proof. For all $i \geq 0$, we have

$$\begin{aligned} \|\delta\check{X}_i\|_{1+\varepsilon}^{1+\varepsilon} &= \mathbb{E}\big[|X_{t_i} - \check{X}_{t_i}|^{1+\varepsilon}\big] = \mathbb{E}\big[|Y_{t_i} - \check{Y}_{t_i}||Y_{t_i} - \check{Y}_{t_i}|^{\varepsilon}|Y_{t_i} + \check{Y}_{t_i}|^{1+\varepsilon}\big] \\ &\leq \|Y_{t_i} - \check{Y}_{t_i}\|_2 \sqrt{\mathbb{E}\Big[\big(|Y_{t_i}| + |\check{Y}_{t_i}|\big)^{2+4\varepsilon}\Big]}. \end{aligned}$$

From (6.4.5), we have that $\mathbb{E}\big[|Y_{t_i}|^{2+4\varepsilon}\big] < C_\epsilon$. Similarly, since $\mathbb{E}\Big[|Y_{t_i}|^{q'}\Big] < +\infty$, we obtain from Proposition 6.3.3(i), that $\mathbb{E}\big[|\check{Y}_{t_i}|^{2+4\varepsilon}\big] < C_{r,\varepsilon}$. This moment bounds, combined with Corollary 6.3.1 (or Proposition 6.3.2, when $r = 1$) and the above inequality, leads to $\|\delta\check{X}_i\|_{1+\varepsilon}^{1+\varepsilon} \leq C_{r,\varepsilon}h^r$. □

Remark 6.4.1. *We obtain above a rate of convergence for a larger set of parameters compared to the results using an implicit Euler scheme in [NS12], where rates of convergence are proved for* $\omega \geq 3$*; however, we only achieve a convergence rate of* 1 *when* $\omega > 5$.

6.4.2 Locally smooth coefficients

We now consider a stochastic differential equation of the form (6.2.4), with drift function $\mu(x) \equiv \mu_1(x) - \mu_2(x)x$, where $\mu_1, \mu_2 : D \to \mathbb{R}$, and diffusion function $\sigma(x) \equiv \gamma x^\nu$, with $\gamma > 0$ and $\nu \in [1/2, 1]$. This model encompasses the Feller diffusion (see Section 6.4.1) and the CEV model [CR76], both widely used in mathematical finance. For the special case $\nu = 1$, the diffusion function is K-Lipschitz and our scheme applies directly to the process X as long as (6.2.2) and (6.2.3) hold for the drift function μ.

We now focus on the case $\nu \in [1/2, 1)$. The Lamperti transform reads $F(x) \equiv \int^x \mathrm{d}y/\sigma(y) \equiv \frac{1}{\gamma(1-\nu)}x^{1-\nu}$, with inverse $F^{-1}(y) \equiv [\gamma(1-\nu)y]^{\frac{1}{1-\nu}}$. The process $Y = F(X)$ is the solution to $\mathrm{d}Y_t = f(Y_t)\mathrm{d}t + \mathrm{d}W_t$, with $Y_0 = F(x_0)$ and

$$f(y) \equiv \frac{\mu\left(F^{-1}(y)\right)}{\sigma\left(F^{-1}(y)\right)} - \frac{1}{2}\sigma'\left(F^{-1}(y)\right). \tag{6.4.6}$$

In order for the functions μ and σ to satisfy the required conditions, we assume:

($\mathbf{H}s0$): $\nu \in [1/2, 1)$, and μ_1, μ_2 are bounded and belong to $\mathcal{C}_b^2(D)$; furthermore μ_1 is non-negative and non-increasing, and μ_2 is non-decreasing.

We distinguish between two cases for parameter ν:

($\mathbf{H}s1$): $\nu \in (1/2, 1)$ and $\mu_1(0) > 0$.

($\mathbf{H}s2$): $\nu = 1/2$ and there exists $\bar{x} > 0$ such that $2\mu_1(x)/\gamma^2 \geq 1$ for all $0 < x < \bar{x}$.

We now prove a rate of convergence as a corollary of Theorem 6.3.1.

Proposition 6.4.1 (Locally smooth coefficients). *Assume that* ($\mathbf{H}s0$) *holds. Then,*

$$\max_{i=0,\ldots,n} \left(\|\delta Y_i\|_2 + \|\delta X_i\|_1 + \|\delta \breve{X}_i\|_{1+\epsilon}^{1+\epsilon} \right) \leq C_{r,\epsilon} h^r, \ \epsilon \geq 0,$$

with

1. *If* ($\mathbf{H}s1$) *holds,* $r = 1$.
2. *If* ($\mathbf{H}s2$) *and* $2\mu_1(0)/\gamma^2 =: \omega > 3$ *hold,* $r \in (\frac{1}{6}, 1/2 - 1/\omega)$ *if* $3 < \omega \leq 4$, $r = 1/2$ *if* $4 < \omega \leq 6$ *and* $r = 1$ *if* $\omega > 6$.

In both cases, we set $\zeta := h^{-\frac{2r}{q'-2}}$, *with* $q' = 3 + 4\epsilon$ *in the definition of* $\breve{X} = \breve{Y}^2$, *recall Remark 6.2.3.*

Proof. In [DM11, Proposition 3.1], De Marco proves that under ($\mathbf{H}s0$), there exists a unique strong solution to (6.2.4), which stays in $[0, \infty)$ almost surely. In addition, he shows that ($\mathbf{H}s1$) and ($\mathbf{H}s2$) further imply that $\mathbb{P}(\tau_0 = \infty) = 1$, where τ_0 is the first time the process X reaches zero. We recall that once we perform the Lamperti transformation, the diffusion function is a constant.

We divide the proof in several parts: in (i) we show that the drift function f is one-sided Lipschitz continuous; in (ii) we show that f is locally Lipschitz continuous, and hence conclude that (6.2.2) and (6.2.3) hold.

(i) From (6.4.6), it follows that, for all $(x, y) \in D^2$,

$$(x-y)\left(f(x) - f(y)\right) = (x-y)\left(\frac{\mu\big(F^{-1}(x)\big)}{\sigma\big(F^{-1}(x)\big)} - \tfrac{1}{2}\sigma'\left(F^{-1}(x)\right) - \frac{\mu\big(F^{-1}(y)\big)}{\sigma\big(F^{-1}(y)\big)} + \tfrac{1}{2}\sigma'\left(F^{-1}(y)\right) \right).$$

Since $\sigma'(F^{-1}(x)) = \nu/[(1-\nu)x]$, we observe that

$$(x-y)\left(\frac{1}{2}\sigma'\left(F^{-1}(y)\right) - \frac{1}{2}\sigma'\left(F^{-1}(x)\right) \right) = \frac{\nu}{2(1-\nu)}(x-y)\left(\frac{1}{y} - \frac{1}{x}\right) \leq 0,$$

because $x, y > 0$ and $\nu/(2-2\nu) > 0$. Clearly, $\sigma\left(F^{-1}(x)\right) = \gamma\left[\gamma(1-\nu)x\right]^{\frac{\nu}{1-\nu}}$, and

$$\mu\left(F^{-1}(x)\right) = \mu_1\left(\left[\gamma(1-\nu)x\right]^{\frac{1}{1-\nu}}\right) - \mu_2\left(\left[\gamma(1-\nu)x\right]^{\frac{1}{1-\nu}}\right)\left[\gamma(1-\nu)x\right]^{\frac{1}{1-\nu}} .$$

Now, consider the remaining terms, namely

$$(x-y)\left(\frac{\mu\left(F^{-1}(x)\right)}{\sigma\left(F^{-1}(x)\right)}-\frac{\mu\left(F^{-1}(y)\right)}{\sigma\left(F^{-1}(y)\right)}\right).$$

Introduce $\tilde{x} := [\gamma(1-\nu)x]^{\frac{1}{1-\nu}}$ and $\tilde{y} := [\gamma(1-\nu)y]^{\frac{1}{1-\nu}}$. Note that

$$\begin{aligned}
&(x-y)\left(\frac{\mu_1(\tilde{x})}{\sigma(F^{-1}(x))}-\frac{\mu_1(\tilde{y})}{\sigma(F^{-1}(y))}\right)= \\
&(x-y)\mu_1(\tilde{x})\left(\frac{1}{\sigma(F^{-1}(x))}-\frac{1}{\sigma(F^{-1}(y))}\right)+\frac{(x-y)}{\sigma(F^{-1}(y))}\left[\mu_1(\tilde{x})-\mu_1(\tilde{y})\right]\leq 0,
\end{aligned}$$

since μ_1 is non-negative and non-increasing, $\nu/(1-\nu)\geq 1$, and using the fact that the map $\sigma\circ F^{-1}$ is increasing. Additionally,

$$\begin{aligned}
&(x-y)\left(\frac{\mu_2(\tilde{y})\tilde{y}}{\sigma\left(F^{-1}(y)\right)}-\frac{\mu_2(\tilde{x})\tilde{x}}{\sigma\left(F^{-1}(x)\right)}\right)\\
= \quad &(1-\nu)(x-y)\mu_2(\tilde{y})(y-x)+x(x-y)\left[\mu_2(\tilde{y})-\mu_2(\tilde{x})\right]\leq C(x-y)^2,
\end{aligned}$$

since $\sigma\left(F^{-1}(x)\right)\equiv\gamma\left[\gamma(1-\nu)x\right]^{\frac{\nu}{1-\nu}}$, and since μ_2 is bounded and non-decreasing.

Combining these results shows that the function f is one-sided Lipschitz continuous.

(ii) We now show that f is locally Lipschitz continuous. By differentiation, it is clear that $\sigma\left(F^{-1}(x)\right)=\left(F^{-1}\right)'(x)$, and hence

$$f'(x)=\mu'\left(F^{-1}(x)\right)-\frac{\mu\left(F^{-1}(x)\right)\sigma'\left(F^{-1}(x)\right)}{\sigma\left(F^{-1}(x)\right)}-\frac{1}{2}\left(F^{-1}\right)'(x)\sigma''\left(F^{-1}(x)\right). \tag{6.4.7}$$

By ($\mathbf{H}s0$), the first term on the right-hand side can be bounded as follows:

$$|\mu'\left(F^{-1}(x)\right)|\leq|\mu_1'\left(F^{-1}(x)\right)|+|\mu_2\left(F^{-1}(x)\right)|+|\mu_2'\left(F^{-1}(x)\right)F^{-1}(x)|\leq C\left(1+|x|^{1/(1-\nu)}\right).$$

Regarding the second term, since $\sigma'\left(F^{-1}(x)\right)=\gamma\nu\left[\gamma(1-\nu)x\right]^{\frac{\nu-1}{1-\nu}}=\frac{\nu}{(1-\nu)x}$, and

$$\mu\left(F^{-1}(x)\right)=\mu_1\left(\left[\gamma(1-\nu)x\right]^{1/(1-\nu)}\right)-\mu_2\left(\left[\gamma(1-\nu)x\right]^{1/(1-\nu)}\right)\left[\gamma(1-\nu)x\right]^{1/(1-\nu)},$$

we see that

$$\left|\frac{\mu\left(F^{-1}(x)\right)\sigma'\left(F^{-1}(x)\right)}{\sigma\left(F^{-1}(x)\right)}\right|\leq\left|C_1\frac{\mu_1\left(C_2x^{\frac{1}{1-\nu}}\right)}{x^{\frac{1}{1-\nu}}}\right|+\left|C_3\mu_2(C_4x^{\frac{1}{1-\nu}})\right|, \tag{6.4.8}$$

where C_1, C_2, C_3, C_4 are positive constants. By $(\mathbf{H}s0)$ it follows that (6.4.8) is bounded by $C\left(1+x^{-\beta}\right)$, for $\beta = 1/(1-\nu)$.

We finally consider the last term on the right-hand side of (6.4.7). Observe that

$$\sigma''\left(F^{-1}(x)\right) = \gamma\nu(\nu-1)\left[\gamma(1-\nu)x\right]^{\frac{\nu-2}{1-\nu}} = -Cx^{\frac{\nu-2}{1-\nu}}$$

and $|\frac{1}{2}\left(F^{-1}\right)'(x)\sigma''\left(F^{-1}(x)\right)| \leq C/x^2 \leq Cx^{-\beta}$, since $\nu \in [1/2, 1)$. These three bounds yield $|f'(x)| \leq C(1 + x^{1/(1-\nu)} + x^{-1/(1-\nu)})$, and hence the drift function is locally Lipschitz continuous, with $\alpha = \beta = 1/(1-\nu)$. Combining this with (i) allows us to conclude that (6.2.2) and (6.2.3) hold.

We now prove statements 1 and 2 in the corollary.

1) Assume $(\mathbf{H}s1)$. Since the locally Lipschitz exponents are $\alpha = \beta = 1/(1-\nu)$, fix $k = k' = (1-\nu)/2$, so that $(\mathbf{H}p)$ holds. By [DM11], $\mathbb{E}(\sup_{t\in[0,T]}|X_t^p|)$ and $\mathbb{E}(\sup_{t\in[0,T]}|X_t|^{-p})$ are finite for all $p > 0$; therefore $\mathbb{E}(\sup_{t\in[0,T]}|Y_t|^{-q})$ is finite for all $q > 0$ [DM11, Lemma 3.1]. We note that f belongs to the class $\mathcal{C}^2(D)$ and $(\mathbf{H}y2)$ holds, therefore $r = 1$ from Proposition 6.3.2. The proof of the statement for $\|\delta\check{X}_i\|_{1+\epsilon}$ follows from the same arguments as in the proof of Corollary 6.4.2.

2) Assume that $(\mathbf{H}s2)$ holds and let $2\mu_1(0)/\gamma^2 =: \omega > 3$. Here, $\alpha = 0$ an $\beta = 0$. Then, $\max_{t\in[0,T]}\mathbb{E}(|X_t|^{-p})$ is finite for all $p < \omega - 1$ [DM11, Lemma 3.1], and so is $\max_{t\in[0,T]}\mathbb{E}(|Y_t|^{-\ell})$ for all $\ell < 2(\omega-1)$. Fix $q \in (4, 2(\omega-1))$ and set $k = 1/(q+2)$, so that $(\mathbf{H}p)$ and $(\mathbf{H}y1)$ hold. From Theorem 6.3.1, $r = 1/2 - \beta/(q+2) \in (\frac{1}{6}, \frac{1}{2} - \frac{1}{\omega})$ holds.

Further assume that $4 < \omega \leq 6$. Note that the drift function f belongs to the class $\mathcal{C}^2(D)$. Fix $q \in (8, 2\omega)$ and $k = 1/4$, so that $(\mathbf{H}p)$ holds. By the assumptions on the parameters it follows that $\max_{t\in[0,T]}\mathbb{E}(|Y_t|^{-6}) = \max_{t\in[0,T]}\mathbb{E}(|X_t|^{-3})$ is finite, and therefore $(\mathbf{H}y2)$ holds. From Theorem 6.3.1, $r = \min(1/2, (q+2)/8 - 1/2) > 1/2$. Finally, in the case $\omega > 6$, we can apply Proposition 6.3.2, to conclude that $r = 1$.

The proof of the statement for $\|\delta\check{X}_i\|_{1+\epsilon}$ follows from the same arguments as in the proof of Corollary 6.4.2. □

In the CIR model, we obtain $r = 1/2$ for $3 < \omega < 5$, using finite inverse moments of the process Y from [DNS12]. For the general case in Proposition 6.4.1, we assumed that $4 < \omega < 6$ for $r = 1/2$. In the next corollary, we impose additional assumptions in order to recover the same parameter constraints as for the Feller diffusion in the previous section.

Proposition 6.4.2. *Assume* $(\mathbf{H}s0)$ *and* $(\mathbf{H}s2)$*, and let* $a^*, b^* > 0$ *be such that* $\mu_1(x) \geq a^*$ *and* $\mu_2(x) \leq b^*$ *for all* $x \in D = (0, \infty)$*. Then,*

$$\max_{i=0,\ldots,n} \left(\|\delta Y_i\|_2 + \|\delta X_i\|_1 + \|\delta \check{X}\|_{1+\epsilon}^{1+\epsilon} \right) \leq C_{r,\epsilon} h^r, \ \epsilon \geq 0 \,,$$

with $r = 1/2$ *if* $3 < \omega := 2\mu_1(0)/\gamma^2 \leq 5$*, and* $r = 1$ *if* $\omega > 5$*.*
We set $\zeta := h^{-\frac{2r}{q'-2}}$*, with* $q' = 3 + 4\epsilon$ *in the definition of* $\check{X} = \check{Y}^2$*, recall Remark 6.2.3.*

Proof. From the assumptions on μ_1 and μ_2, there exists $a^*, b^* > 0$ such that the inequality $\mu_1(x) - \mu_2(x)x \geq a^* - b^*x$ holds in the domain D. We define Z as the process with drift $a^* - b^*x$ (instead of $\mu_1(x) - \mu_2(x)x$), and diffusion $\sigma(x) \equiv \gamma x^{1/2}$. Therefore, by the Comparison Theorem (see [KS91, Section 5.2]) the inequality $X_t \geq Z_t$ holds for all $t \in [0, T]$ almost surely, and hence $\mathbb{E}(|X_t|^{-p}) \leq \mathbb{E}(|Z_t|^{-p})$ is true for all $p > 0$. Now, Z is clearly a Feller diffusion and, from the assumption on ω, it follows that $\max_{t\in[0,T]} \mathbb{E}(|Z_t|^{-3})$ is finite. The result then follows directly from the second part of Corollary 6.4.1.
The proof of the statement for $\|\delta \check{X}_i\|_{1+\epsilon}$ follows from the same arguments as in the proof of Corollary 6.4.2. □

6.4.3 3/2 model

The 3/2 process $X = (X_t)_{t\geq 0}$ [Hes97] is the solution to

$$\mathrm{d}X_t = c_1 X_t (c_2 - X_t)\mathrm{d}t + c_3 X_t^{3/2} \mathrm{d}W_t, \quad X_0 = x_0 > 0, \tag{6.4.9}$$

with $c_1, c_2, c_3 > 0$. Introduce the quantity $\omega := 2 + 2c_1/c_3^2$. The Feller diffusion and the 3/2 process are related as follows: the map $F(y) \equiv y^{-1/2}$ yields the Lamperti transformed CIR process $Y := F(X)$, as in (6.4.2) and (6.4.3), with parameters, $a := (4c_1 + 3c_3^2)/8$, $b := -c_1c_2/2$ and $c := -c_3/2$. Existence and uniqueness can be retrieved from the properties of the Feller diffusion, and $\max_{t\in[0,T]} \mathbb{E}(|X_t|^p)$ is finite for all $p < \omega$.

Corollary 6.4.3 (3/2 model)**.** *Let* $Y := X^{-1/2}$*. Then,* $\max_{i=0,\ldots,n} \|\delta Y_i\|_2 \leq Ch^r$*, with* $r \in (\frac{1}{6}, \frac{1}{2} - \frac{1}{w+1})$ *if* $\omega \in (2, 3]$*,* $r = 1/2$ *if* $3 < \omega \leq 5$ *and* $r = 1$ *if* $\omega > 5$*.*

Proof. In terms of the CIR coefficients, we have $\omega = 2 + 2c_1/c_3^2 = 2\kappa\theta/\xi^2$. We directly apply Corollary 6.4.1 to get the desired results. □

We now establish a convergence result for the 3/2 process X, using the modification $\tilde{X}$ (recall Remark 6.2.3).

Proposition 6.4.3. *Let $\omega > 3$ and fix $\varepsilon \geq 0$. If $3 + 2\varepsilon < \omega$, then*

$$\max_{i=0,\ldots,n} \|X_{t_i} - \tilde{X}_{t_i}\|_{1+\varepsilon} \leq C_{r,\varepsilon} h^{\frac{r}{2(1+\varepsilon)}},$$

with $r = 1/2$ for $\omega \leq 5$ and $r = 1$ for $\omega > 5$, where $\eta = h^{r/(2\omega)}$.

Proof. It follows that

$$\begin{aligned}
\|X_{t_i} - \tilde{X}_{t_i}\|_{1+\varepsilon}^{1+\varepsilon} &= \mathbb{E}\big[|X_{t_i} - \tilde{X}_{t_i}|^{1+\varepsilon}\big] \\
&= \mathbb{E}\Big[\Big|\frac{1}{Y_{t_i}^2} - \frac{1}{\tilde{Y}_{t_i}^2}\Big|^{1+\varepsilon}\Big] \\
&= \mathbb{E}\left[\left|\frac{(Y_{t_i} - \tilde{Y}_{t_i})(Y_{t_i} + \tilde{Y}_{t_i})}{Y_{t_i}^2 \tilde{Y}_{t_i}^2}\right|^{1+\varepsilon}\right] \\
&\leq C_\varepsilon \|Y_{t_i} - \tilde{Y}_{t_i}\|_2 \sqrt{\mathbb{E}\left[\frac{(Y_{t_i} + \tilde{Y}_{t_i})^{2+4\varepsilon}}{|Y_{t_i}|^{4+4\varepsilon}\tilde{Y}_{t_i}^{4+4\varepsilon}}\right]} \\
&\leq C_\varepsilon \|Y_{t_i} - \tilde{Y}_{t_i}\|_2 \sqrt{\mathbb{E}\big[|Y_{t_i}|^{-(6+4\varepsilon)} + \mathbb{E}\big[|\tilde{Y}_{t_i}|^{-(6+4\varepsilon)}\big]\big]};
\end{aligned}$$

since $3 + 2\varepsilon < \omega$ it follows that $\mathbb{E}\Big[|Y_{t_i}|^{-(6+4\varepsilon)}\Big]$ is bounded by a constant. Furthermore, for $\eta = h^{r/(2\omega)}$ (q is such that $q < 2\omega$), it follows that $\mathbb{E}\Big[|\tilde{Y}_{t_i}|^{-(6+4\varepsilon)}\Big] \leq \eta^{-(6+4\varepsilon)}$, therefore $\sqrt{\mathbb{E}\big[|\tilde{Y}_{t_i}|^{-(6+4\varepsilon)}\big]} \leq C_{\varepsilon,\omega} h^{-r/2}$, which together with $3 + 2\varepsilon < \omega$ and Corollary 6.3.1 (or Proposition 6.3.2, if $r = 1$), conclude the result. □

Remark 6.4.2. *The last corollary proves L^p-bounds ($p > 1$) for the 3/2 model. In [NS12, Proposition 3.2] the authors prove strong convergence for the 3/2 process using a drift-implicit scheme when $\omega > 6$ holds; our results above improve this by yielding strong rates of convergence for $\omega > 3$.*

Alternatively, we could indeed use Proposition 6.3.3 for a higher rate of convergence, however the parameter ω required is larger:

Corollary 6.4.4. *Let $\omega > \frac{9+4\varepsilon}{2} \vee 5$ for some fixed $\varepsilon \geq 0$. Then*

$$\max_{i=0,\ldots,n} \|X_{t_i} - \tilde{X}_{t_i}\|_{1+\varepsilon} \leq C_{\varepsilon,\omega} h^{1/(1+\varepsilon)}.$$

Proof. From the computation in the proof of Proposition 6.4.3, we have

$$\|X_{t_i} - \tilde{X}_{t_i}\|_{1+\varepsilon}^{1+\varepsilon} \le C_\varepsilon \|Y_{t_i} - \tilde{Y}_{t_i}\|_2 \sqrt{\mathbb{E}\big[|Y_{t_i}|^{-(6+4\varepsilon)} + \mathbb{E}\big[|\tilde{Y}_{t_i}|^{-(6+4\varepsilon)}\big]\big]};$$

Using Proposition 6.3.3(ii), the term $\mathbb{E}\Big[|\tilde{Y}_{t_i}|^{-(6+4\varepsilon)}\Big]$ is bounded by a constant depending on ω and ε, since $6 + 4\varepsilon < q - 3 < 2\omega - 3$. Moreover, since $\omega > 5$, we get that $\|Y_{t_i} - \tilde{Y}_{t_i}\|_2 \le Ch$, from (6.4.4) and the same arguments as in the proof of Proposition 6.3.2. □

6.4.4 Ait-Sahalia model

In the Ait-Sahalia interest rate model [AS96], X is the solution to

$$\mathrm{d}X_t = \left(\frac{a_{-1}}{X_t} - a_0 + a_1 X_t - a_2 X_t^{\varrho}\right)\mathrm{d}t + \gamma X_t^{\rho}\mathrm{d}W_t, \quad X_0 = x_0 > 0, \tag{6.4.10}$$

where all constant parameters are non-negative, and $\rho, \varrho > 1$. From [SMHP11], there exists a strong solution on $(0,\infty)$, and the Lamperti transformation $Y := X^{1-\rho}$ satisfies

$$\mathrm{d}Y_t = f(Y_t)\mathrm{d}t + (1-\rho)\gamma\mathrm{d}W_t, \quad Y_0 = x_0^{1-\rho} > 0, \tag{6.4.11}$$

with

$$f(x) \equiv (1-\rho)\left(a_{-1}x^{\frac{-1-\rho}{1-\rho}} - a_0 x^{\frac{-\rho}{1-\rho}} + a_1 x - a_2 x^{\frac{-\rho+\varrho}{1-\rho}} - \frac{\rho\gamma^2}{2}x^{-1}\right).$$

Corollary 6.4.5. *If $\varrho + 1 > 2\rho$, then $\max_{i=0,\dots,n} \|\delta Y_i\|_2 \le Ch$.*

Proof. Straightforward differentiation yields

$$f'(x) = -a_{-1}(1+\rho)x^{\frac{2}{\rho-1}} + a_0\rho x^{\frac{1}{\rho-1}} + a_1(1-\rho) - a_2(-\rho+\varrho)x^{-\frac{r-1}{\rho-1}} - \frac{\rho\gamma^2}{2}(\rho-1)x^{-2}.$$

We have $\lim_{x\downarrow 0} f'(x) = \lim_{x\uparrow\infty} f'(x) = -\infty$, hence $\sup_{0<x<\infty} f'(x)$ is finite by continuity and therefore f is one-sided Lipschitz continuous. In addition, $|f'(x)| \le C(1 + x^{\frac{2}{\rho-1}} + x^{-\frac{\varrho-1}{\rho-1}})$ for $x > 0$, so f is locally Lipschitz continuous with $\alpha = 2/(\rho-1)$ and $\beta = (\varrho-1)/(\rho-1)$. The diffusion is constant, hence Lipschitz continuous. Using the locally Lipschitz continuous properties of the drift, fix $k = 1/(2\beta)$ and $k' = 1/(2\alpha)$. We recall that if $\varrho + 1 > 2\rho$, then $\max_{t\in[0,T]} \mathbb{E}(|X_t|^p)$ and $\max_{t\in[0,T]} \mathbb{E}(|X_t|^{-p})$ are finite for all $p \neq 0$ [SMHP11, Lemma 2.1] so that $(\mathbf{H}y1)$ holds. Differentiation yields

$$f''(x) = \frac{-2a_{-1}(\rho+1)}{\rho-1}x^{\frac{3-\rho}{\rho-1}} + \frac{a_0\rho}{\rho-1}x^{\frac{2-\rho}{\rho-1}} + a_2\frac{(-\rho+\varrho)(\varrho-1)}{\rho-1}x^{-\frac{\varrho+\rho-2}{\rho-1}} + \rho\gamma^2(\rho-1)x^{-3}.$$

Since f belongs to $\mathcal{C}^2(D)$ and (6.2.6) is finite by [SMHP11, Lemma 2.3], then $(\mathbf{H}y2)$ holds. Fix $q > 6\beta - 2$ and $q' > 6\alpha + 2$. Then, by Proposition 6.3.2, the statement is proved. □

We now compute a strong rate of convergence for the Ait-Sahalia process X. We need to control the behaviour of the approximation near 0 and at ∞. In order to do that, we introduce modification $\check{X}_{t_i} := \check{Y}^{\frac{1}{1-\rho}}$ where $\check{Y}_{t_i} = p_{\bar{D}_\eta} \circ p_{\check{D}_\zeta}(\hat{Y}_{t_i}) = p_{\check{D}_{\eta,\zeta}}(\hat{Y}_{t_i})$, for η and ζ to be determined later on.

Corollary 6.4.6. *If $\varrho + 1 > 2\rho$, then for $\epsilon \geq 0$,*

$$\max_{i=0,\dots,n} \|X_{t_i} - \check{X}_{t_i}\|_{1+\epsilon} \leq Ch^{\frac{1}{1+\epsilon}}$$

with $\eta := h^{2/q}$, $\zeta = h^{-\frac{2}{q'-2}}$ and $q = 3 + 4\rho(1+\epsilon)/(1-\rho)$, $q' = 4\epsilon + 1$.

Proof. A similar approach to Proposition 6.3.3 yields

$$\mathbb{E}[|\delta\check{X}_{t_i}|^{1+\epsilon}] \leq C\left(\mathbb{E}\left[|Y_{t_i}|^{4\rho(1+\epsilon)/(1-\rho)} + |Y_{t_i}|^{4\epsilon} + |\check{Y}_{t_i}|^{4\rho(1+\epsilon)/(1-\rho)} + |\check{Y}_{t_i}|^{4\epsilon}\right]\right)^{\frac{1}{2}} (\mathbb{E}|\delta\check{Y}_{t_i}|^2)^{\frac{1}{2}},$$

where $\delta\check{X}_{t_i} = X_{t_i} - \check{X}_{t_i}$ and $\delta\check{Y}_{t_i} = Y_{t_i} - \check{Y}_{t_i}$. Since $\rho > 1$ and $\varrho + 1 > 2\rho$, $\mathbb{E}[|Y_{t_i}|^{4\rho(1+\epsilon)/(1-\rho)} + |Y_{t_i}|^{4\epsilon}]$ is finite. Observing that $\check{Y} \leq \check{Y} + \eta$, $\frac{1}{\check{Y}} \leq \frac{1}{\check{Y}} + \frac{1}{\zeta}$ and using Proposition 6.3.3, we get $\mathbb{E}\Big[|\check{Y}_{t_i}|^{4\rho(1+\epsilon)/(1-\rho)} + |\check{Y}_{t_i}|^{4\epsilon}\Big] \leq C$. Also, we compute

$$\begin{aligned}|Y_{t_i} - \check{Y}_{t_i}| &\leq |Y_{t_i} - p_{\bar{D}_\eta}(Y_{t_i})| + |p_{\bar{D}_\eta}(Y_{t_i}) - p_{\bar{D}_\eta} \circ p_{\check{D}_\zeta}(\hat{Y}_{t_i})| \\ &\leq |Y_{t_i} - p_{\bar{D}_\eta}(Y_{t_i})| + |Y_{t_i} - p_{\check{D}_\zeta}(Y_{t_i})| + |Y_{t_i} - \hat{Y}_{t_i}|\end{aligned}$$

recalling that $p_{\bar{D}_\eta}$ and $p_{\check{D}_\zeta}$ are 1-Lipschitz. Using similar arguments as in the proof of Corollary 6.3.1, we then obtain $(\mathbb{E}|\delta\check{Y}_{t_i}|^2)^{\frac{1}{2}} \leq Ch$ and the result follows. □

6.5 Numerical results

In this section, we numerically confirm the strong convergence rate of the modified Euler scheme for the CIR model, the one-dimensional stochastic Ginzburg-Landau equation with multiplicative noise, and the Ait-Sahalia model. For a process X, denote by $\hat{X}_T^{(j)}$ the modified Euler-Maruyama approximation at time T and $X_T^{(j)}$ the closed-form solution (or reference solution), using the same Brownian motion path (the j^{th} path). The empirical average absolute error $\mathcal{E}$ is defined by

$$\mathcal{E} := \frac{1}{M}\sum_{j=1}^{M} |X_T^{(j)} - \hat{X}_T^{(j)}|,$$

over M sample paths, which we will set to $M = 10000$. An equidistant time grid is used, with step sizes $h := T/2^N$, for different values of N. The strong error rates are computed by plotting $\mathcal{E}$ against the number of discretisation steps on a log-log scale, and the strong rate of convergence r is then retrieved using linear regression.

6.5.1 CIR model

The Lamperti-transformed drift-implicit square-root Euler method (see [DNS12, NS12]) has a unique strictly positive solution defined for $i = 0, \dots, n-1$ by

$$Y_{t_{i+1}} = \frac{Y_{t_i} + c\Delta W_{i+1}}{2(1-bh_{i+1})} + \sqrt{\frac{(Y_{t_i} + c\Delta W_{i+1})^2}{4(1-bh_{i+1})^2} + \frac{ah_{i+1}}{1-bh_{i+1}}}, \qquad Y_0 = \sqrt{x_0} > 0,$$

with a, b, c defined in (6.4.3). The CIR/Feller diffusion is recovered by setting $X_{t_i} = Y_{t_i}^2$ for $i \le n$, and we compare the modified explicit Euler scheme with this implicit scheme used as a reference solution (with a large number of time steps).

We compute the strong rates of convergence for the CIR process, where the implicit scheme is used as a reference solution. Set $(\kappa, \theta, \xi, T, x_0) = (0.125\omega, 1, 0.5, 1, 1)$, such that $2\kappa\theta/\xi^2 = \omega$. The cases $\omega = (1, 1.5, 2, 2.5, 3, 3.5, 4)$ are considered. The reference solution is computed using $N = 12$. Figure 6.1 shows the rates of convergence r achieved for the CIR process, where $k = 1/4$ in the modified scheme, according to Corollary 6.4.1. In the corollary, we prove a strong rate of convergence of $1/2$ when $3 < \omega \le 5$, and $r = 1$ for $\omega > 5$. The coefficient of determination R^2, for the goodness of the fit of the straight line, is above 0.998 for all ω. We observe that numerically order 1 is achieved by our scheme for $\omega > 1$, which is better than the bound we proved.

Remark 6.5.1. *The projection introduced in Definition 6.2.1 can be modified to* $\tilde{p}_n(x) := Ln^{-k} \vee x \wedge Un^{k'}$*, with* $L, U > 0$ *suitably chosen constant. This is beneficial if the process has extreme initial conditions or average state, and does not impact the convergence results.*

For small x_0, it is intuitive to use the projection in Remark 6.5.1 to achieve faster convergence (albeit without affecting the asymptotic behaviour). Set $(\kappa, \theta, \xi, T) = (0.375, 1, 0.5, 1)$, such

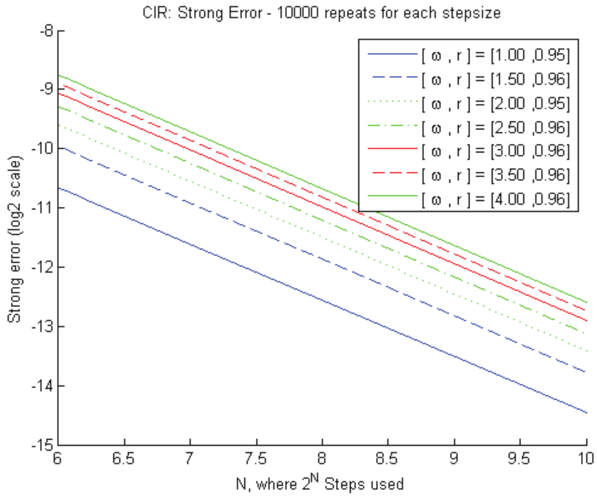

Figure 6.1: CIR model: $\mathcal{E}$ against number of steps ($\log_2$ scale).

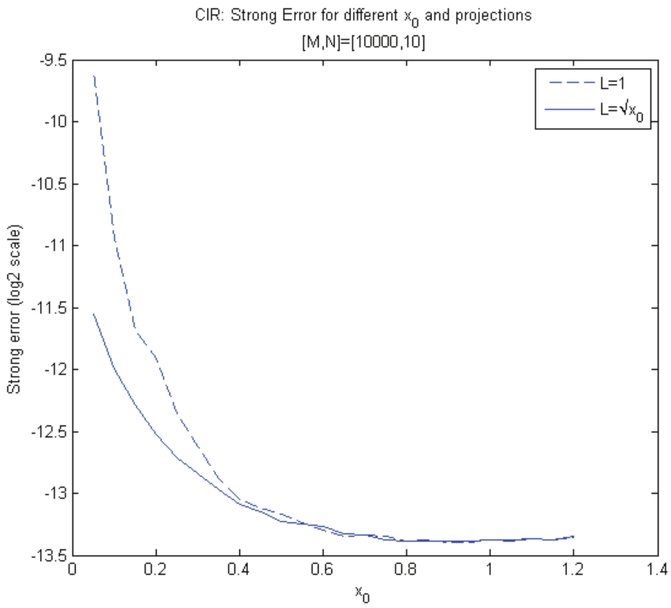

Figure 6.2: Absolute error ($\log_2$ scale) for $N = 10$.

that $2\kappa\theta/\xi^2 = 3$. In Figure 6.2, we let x_0 vary between 0.05 and 1.2 in increments of 0.05. We compare the errors achieved for $k = 1/4$, using the projections $p_n(x) = n^{-k} \vee x$ and $\tilde{p}_n(x) = \sqrt{x_0} n^{-k} \vee x$. By using the projection $\tilde{p}_n$, smaller errors can be achieved for small x_0.

6.5.2 Ginzburg-Landau

Consider the one-dimensional stochastic Ginzburg-Landau SDE [KP92, Chapter 4], where the process X is the unique strong solution to

$$\mathrm{d}X_t = \left[-X_t^3 + \left(\lambda + \frac{1}{2}\sigma^2\right) X_t\right] \mathrm{d}t + \sigma X_t \mathrm{d}W_t, \quad X_0 = x_0 > 0,$$

for $\lambda, \sigma \geq 0$, which admits the closed-form solution

$$X_t = \frac{x_0 \exp(\lambda t + \sigma W_t)}{\sqrt{1 + 2x_0^2 \int_0^t \exp(2\lambda s + 2\sigma W_s)\mathrm{d}s}}. \tag{6.5.1}$$

This SDE is a special case of the Ait-Sahalia process with $(a_{-1}, a_0, a_1, a_2, \varrho, \rho) = (0, 0, \lambda + \sigma^2/2, 1, 3, 1)$. For this choice of parameters, $\varrho + 1 > 2\rho$, hence the moments and inverse moments of X_t are finite for all $t \in [0, T]$, and the solution stays in $(0, \infty)$ almost surely. The drift function satisfies (6.2.2), with $(\alpha, \beta) = (2, 0)$, e.g. set $k' = 1/4$ in the modified scheme. In addition, the drift is one-sided Lipschitz continuous and the diffusion is K-Lipschitz. As a result, theoretical convergence for this example can be obtained with rate $r = 1$, recall also Remark 6.3.1.

Ginzburg-Landau strong convergence: For this SDE, the closed-form solution is used in the definition of $\mathcal{E}$ to compute the strong rate of convergence r. Figure 6.3 shows the average absolute error $\mathcal{E}$ using the modified scheme, for parameters $(\sigma, \lambda, T, x_0) = (1, 1/2, 1, 1)$. The empirical rate achieved of 0.53 (same as the standard Euler scheme) which is lower than the predicted rate of 1. This can be explained since we are approximating the integral in (6.5.1) as a summation.

Ginzburg-Landau Euler-Maruyama divergence: We consider an example of the Ginzburg-Landau SDE for which the standard Euler-Maruyama scheme diverges, and compare the results with the modified explicit scheme. Fix parameters $(\sigma, \lambda, T, x_0) = (7, 0, 3, 1)$ as

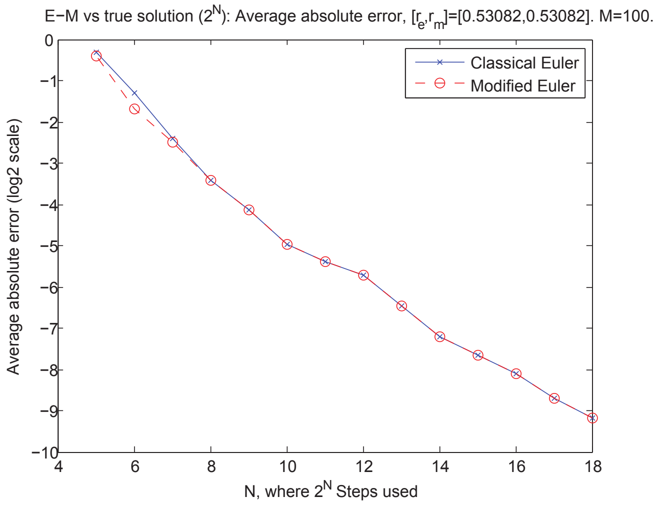

Figure 6.3: Ginzburg-Landau model: average absolute error $\mathcal{E}$ vs N ($\log_2$ scale).

in [HJK11], for which the authors prove moment explosion for the classical Euler-Maruyama scheme, see [HJK11, Table 1]. Figure 6.4 shows the error $\mathcal{E}$ for the classical and the modified schemes, for different N. For the modified scheme, set $k' = 1/4$. The modified Euler scheme converges with a rate $r_m = 0.43$. For a range of step sizes, the classical Euler scheme explodes, as proven in [HJK11] (N.B. very large and NaN values are set to 2^{20} in the figure, to illustrate the explosions for the classical scheme). The modified scheme appears to be more robust.

6.5.3 Ait-Sahalia model

The strong rate of convergence for the Ait-Sahalia model is computed using a reference solution with a large number of steps. Consider the parameters $(a_{-1}, a_0, a_1, a_2, \gamma, x_0) = (1, 1, 1, 1, 1, 1)$, and $(\varrho, \rho, T) = (2, 3/2, 1)$. From these parameters, note that $\alpha = 4$ and $\beta = 2$. Fix k and k', such that $2\beta k = 1$ and $2\alpha k' = 1$, so that $(\mathbf{H}y1)$ holds. Figure 6.5 shows $\mathcal{E}$ against the number of steps (log-log plot), where 2^{12} steps are used for the reference solution. The Ait-Sahalia empirical rate of convergence $r = 1.25$ could be justified by the fact that we used a reference solution instead of the true solution.

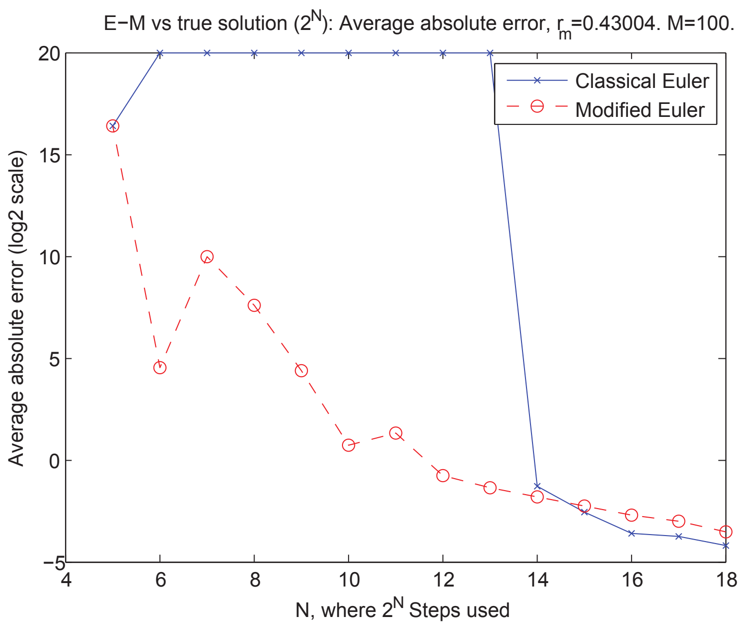

Figure 6.4: Average absolute error $\mathcal{E}$ vs number of steps ($\log_2$ scale).

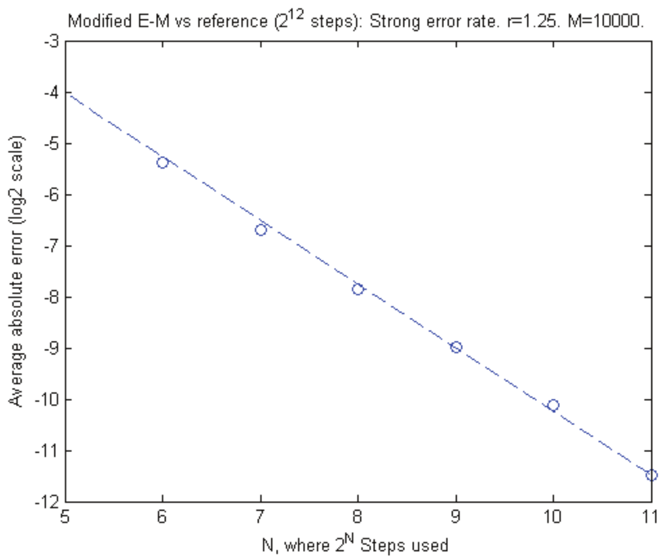

Figure 6.5: Ait-Sahalia model: average absolute error vs N ($\log_2$ scale).

7. Examples and extensions

Let $\|x\|$ be the Euclidean distance of a vector $x \in \mathbb{R}^d$,

$$\|x\| := \sqrt{\sum_{i=1}^{d} x_i^2}\ .$$

We consider a d-dimensional drift vector function $f : \mathbb{R}^d \to \mathbb{R}^d$ and a $d \times m$-matrix diffusion function $\gamma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$. We denote $\gamma^{i,\cdot}$ the vector function that returns the ith column of γ; in other words $\gamma^{i,\cdot} : \mathbb{R}^d \to \mathbb{R}^d$. Let the process $Y = (Y_t)_{t\geq 0}$ be the solution to the stochastic differential equation

$$\mathrm{d}Y_t = f(Y_t)\mathrm{d}t + \gamma(Y_t)\mathrm{d}W_t, \qquad Y_0 = y_0 \in \mathbb{R}^d, \tag{7.0.1}$$

where W is an m-dimensional Brownian motion. We shall consider the following assumptions:
$(\mathbf{H}D)$: the solution of (7.0.1) takes values in $D \subseteq \mathbb{R}^d$, almost surely.
$(\mathbf{H}f)$: f is locally Lipschitz continuous and globally one-sided Lipschitz continuous on D, namely there exist $\alpha, \beta \geq 0$ and $K > 0$, such that for all $(x, y) \in D^2$:

$$\|f(x) - f(y)\| \leq K(1 + \|x\|^\alpha + \|y\|^\alpha + \frac{1}{\|x\|^\beta} + \frac{1}{\|y\|^\beta})\|x - y\|,$$
$$\langle x - y, f(x) - f(y)\rangle \leq K\|x - y\|^2\ .$$

$(\mathbf{H}\gamma)$: $\gamma^{i,\cdot}$ is Lipschitz continuous on D for all dimensions $i = 1, \ldots, d$: there exists $K > 0$ such that for all $(x, y) \in D^2$,

$$\|\gamma^{i,\cdot}(x) - \gamma^{i,\cdot}(y)\| \leq K\|x - y\|\ .$$

7.1 Singularities on the closure of D

In the one-dimensional case, we introduced a projection map of the state-space to an interval D_n. The projection map in a multi-dimensional example is $p_{n,d} : \mathbb{R}^d \to D_{n,d}$, with $D_{n,d} := \{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_i > 0, i = 1, \ldots, d$ and $n^{-k} \leq \|x\| \leq n^{k'}\}$. Consider a sequence of convex sub-domains, $D_{n,2}$, such that the set-theoretic limit of D_n exists and equals to D, and a one-Lipschitz projection map $p_{n,d}$. In Figure 7.1, we consider $D_{n,2}$ for $D = (0, \infty)^2$, with

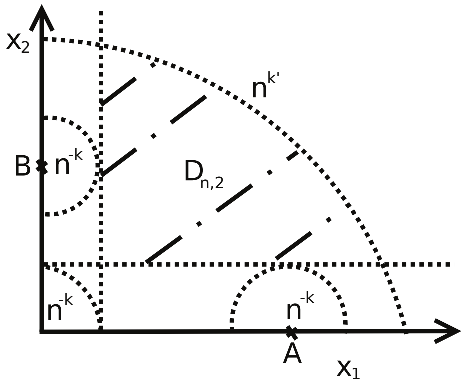

Figure 7.1: Sub-domain $D_{n,2}$, with singularities at the origin, and at points A and B.

singularities on the closure of D.

7.2 Domain $D = (-\infty, \infty)$

We define the sub-domain $D_n := D_n^- \bigcup D_n^+$ where $D_n^- := [-n^{k'}, -n^{-k}]$ and $D_n^+ := [n^{-k}, n^{k'}]$. We also introduce $E_n := E_n^- \bigcup E_n^0 \bigcup E_n^+$, where $E_n^- := (-\infty, -n^{k'})$, $E_n^0 := (-n^{-k}, n^{-k})$ and $E_n^+ := (n^{k'}, \infty)$. However, the projection map

$$\tilde{p}_n(x) := \begin{cases} x & \text{if } x \in D_n, \\ -n^{k'} & \text{if } x \in E_n^-, \\ n^{k'} & \text{if } x \in E_n^+, \\ -n^{-k} & \text{if } 0 > x > -n^{-k}, \\ n^{-k} & \text{if } 0 \le x < n^{-k}, \end{cases}$$

is not one-Lipschitz and D_n is not an interval.

Example 7.2.1 (SDEs on $D = (-\infty, \infty)$). *We consider the stochastic differential equation*

$$\mathrm{d}Y_t = \mu f(Y_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad Y_0 = y \in \mathbb{R}, \tag{7.2.1}$$

where μ and σ are positive constants and

$$f(x) = \begin{cases} -\sqrt{x} & \textit{if } x \ge 0, \\ \sqrt{-x} & \textit{if } x < 0. \end{cases}$$

The first derivative of the drift function is

$$f'(x) = \begin{cases} -1/(2\sqrt{x}) & \text{if } x > 0, \\ -1/(2\sqrt{-x}) & \text{if } x < 0, \\ \text{undefined} & \text{if } x = 0. \end{cases}$$

We require the projection map to allow the process to "cross" the singularity at the origin. For the Euler scheme $y = p_n(x) + f(p_n(x))h + \gamma(p_n(x))\Delta W$, we require $|p_n(x) - p_n(y)| \leq |x - y|$, i.e. the map to be one-Lipschitz. The two regions D_n^- and D_n^+, are separated by E_n^0.

Definition 7.2.1. *In the case where $D = \mathbb{R}$, we define the projection map $p_{n,x} : \mathbb{R} \to D_n$ using the next step y given by $y = x + f(x)h + \gamma(x)\Delta W$— using strictly positive k, k' and $x \in D_n$ — by*

$$p_{n,x}(y) := \begin{cases} y & \text{if } y \in D_n, \\ n^{-k} & \text{if } y \in E_n^0, x \in D_n^+, \\ -n^{-k} & \text{if } y \in E_n^0, x \in D_n^-, \\ n^{k'} & \text{if } y \in E_n^+, x \in D_n, \\ -n^{k'} & \text{if } y \in E_n^-, x \in D_n. \end{cases}$$

Remark 7.2.1.

(i) For $x \in D_n$, set $y = x + f(x)h + \gamma(x)\Delta W$; if $y \in D_n$, the projection $p_{n,x}(y)$ is the identity map. s

(ii) For $y \in E_n$; if $y \in E_n^+$ (or E_n^-), then $p_{n,x}(y) = n^{k'}$ (resp. $-n^{k'}$), if $y \in E_n^0$, and $x \in D_n^+$ (or $x \in D_n^-$), then $p_{n,x}(y) = n^{-k}$ (resp. $-n^{-k}$).

(iii) As we increase the number of steps, n, the region E_n^0 becomes smaller. The aim is to allow a transition of the process from D_n^+ to D_n^- and vice versa, and the projection allows the process to cross E_n^0 in one time step.

For $x \in D_n$, note that $|p_{n,x}(x) - p_{n,x}(y)| \leq |x - y|$, and we now define the discretisation scheme:

Definition 7.2.2. *Set $\hat{Y}_0 = Y_0 \in D_n$ and for $i = 0, \ldots, n-1$,*

$$\hat{Y}_{t_{i+1}} := p_{n,\hat{Y}_{t_{i-1}}}(\hat{Y}_{t_i}) + f_{n,\hat{Y}_{t_{i-1}}}(\hat{Y}_{t_i})h_{i+1} + \gamma_{n,\hat{Y}_{t_{i-1}}}(\hat{Y}_{t_i})\Delta W_{i+1},$$

with $h_{i+1} := t_{i+1} - t_i$, $\Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}$, $f_{n,x} \equiv f \circ p_{n,x}$ and $\gamma_{n,x} \equiv \gamma \circ p_{n,x}$.

Example 7.2.2. *Consider the SDE in* (7.2.1), *with parameters* $(\mu, \sigma, Y_0, T) = (2, 2, 1, 1)$. *The diffusion function is Lipschitz continuous and we apply the discretisation scheme in Definition 7.2.2. Fix* $\alpha = 0$ *and* $\beta = 1/2$ *as the scheme parameters, and choose* $k = 1$ *(the upper bound, since* $2\kappa\beta \leq 1$*), and* $k' = 10$ *(arbitrary choice).*

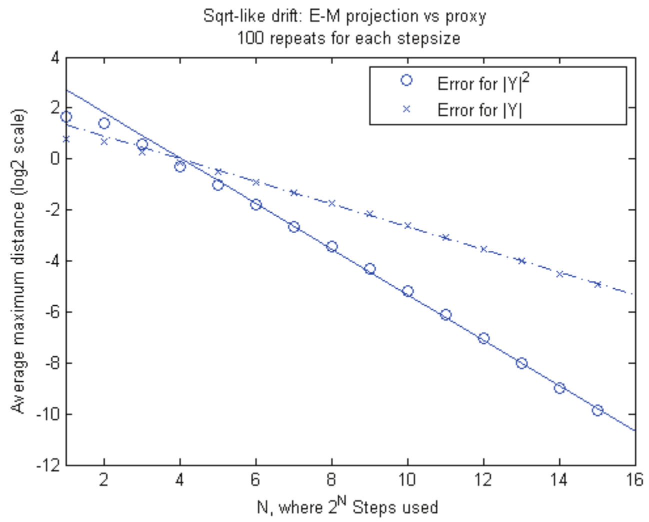

Figure 7.2: Example 7.2.2: Mean absolute error and MSE for process Y, using different step sizes.

In Figure 7.2, we show the rates of convergence for the mean absolute error (MAE) and the MSE, using a reference solution with 2^{16} steps. The rates obtained are 0.45 and 0.90.

7.3 Discontinuous drift function

We consider the stochastic differential equation

$$dY_t = \mu f(Y_t)dt + \sigma dW_t, \qquad Y_0 \in \mathbb{R}, \tag{7.3.1}$$

where μ and σ are strictly positive parameters and $f(x) = 2(1/2 - \mathbf{1}_{x \geq 0})$. Clearly, we have a discontinuity at $x = 0$ for the drift function; more so, the function is one-sided Lipschitz.

We consider an approximation using the drift function, f_ε, using a cubic spline in an ε-

neighbourhood of the discontinuity at the origin. We define this approximation by

$$f_\varepsilon(x) = \begin{cases} 1 & \text{if } x \leq -\varepsilon, \\ \frac{x^3}{2\varepsilon^3} - \frac{3x}{2\varepsilon} & \text{if } -\varepsilon < x < \varepsilon, \\ -1 & \text{if } x \geq \varepsilon. \end{cases}$$

Note that $f'_\varepsilon(x) \leq C_\varepsilon(1 + |x|^2)$. For this example the locally Lipschitz parameters are $\alpha = 2$ and $\beta = 0$. Fix the explicit Euler scheme parameters such that $0 < k' \leq 1/4$, and k arbitrarily. For $\varepsilon = n^{-k}$, the map $p_{n,x}$ will project points in E_n^0 to D_n^+ or D_n^-. We consider the SDE in (7.3.1) with parameters $(\mu, \sigma, Y_0, T) = (2, 2, 1, 1)$. The MAE and MSE rates for this example are 0.47 and 0.96.

8. Monte Carlo Acceleration

We present applications using the strong rate of convergence for the modified Euler scheme introduced in Chapter 6. We consider the multilevel Monte Carlo (MLMC) technique which requires this strong rate of convergence, and an accelerating scheme for a stochastic volatility model.

8.1 MLMC

We combine the modified Euler scheme and the multilevel Monte Carlo approach introduced by Giles [Gil08b, GS12]. The original paper focused on approximating the expected value of Lipschitz continuous payoffs. The MLMC method has also been justified for digitals, lookback and barrier options [GHM09]. Multischeme MLMC techniques use different discretisation schemes in order to further improve the computational efficiency [Abe11]. The use of MLMC techniques has also been applied to compute Greeks [BG12].

We target a root mean squared error (RMSE) of $\mathcal{O}(\varepsilon)$ for the option price. Using an Euler-Maruyama scheme, the MSE of an option price is $C_1/N + C_2h^2$, where N is the number of Monte Carlo paths, and h is the step size of the discretisation. By choosing $N := \mathcal{O}(\varepsilon^{-2})$, and $h := \mathcal{O}(\varepsilon)$, the total cost is $\mathcal{O}(\varepsilon^{-3})$.

The idea behind MLMC is to use different time steps, at different levels of the simulation. We increase the number of time steps at each level by a factor M, where level l uses M^l steps of size $h_l := T/M^l$. We define P_l to be the numerical approximation of the payoff at level l, for $l = 0, \ldots, L$, where L is the maximum number of levels. By linearity of the expectation operator we note that

$$\mathbb{E}\left[P_L\right] = \mathbb{E}\left[P_0\right] + \sum_{l=1}^{L} \mathbb{E}\left[P_l - P_{l-1}\right] \,, \tag{8.1.1}$$

where the difference in the payoff approximation on levels l and $l-1$ is estimated using the same Brownian path, for both levels. The variance of the payoff difference, $V_l := \mathbb{V}(P_l - P_{l-1})$, decreases quickly with increasing levels, and it has been shown that for European options with Lipschitz continuous payoffs, V_l converges to zero twice as fast as the strong convergence rate of the scheme. At each level l, we simulate N_l paths and estimate $\mathbb{E}\left[P_l - P_{l-1}\right]$. The multilevel estimator has variance $1/N_l \sum_{l=0}^{L} V_l$, and $N_l := C\sqrt{V_l h_l}$ minimises the computational cost [Gil08b], to achieve a RMSE of $\mathcal{O}(\varepsilon)$. The strong convergence rate is required for the MLMC techniques, and the complexity theorem provides a general result for the computational cost of the MLMC method [Gil08b]. MLMC methods have been shown to improve the computational efficiency using an Euler-Maruyama discretisation to $\mathcal{O}\left(\varepsilon^{-2}(\log\varepsilon)^2\right)$, and $\mathcal{O}(\varepsilon^{-2})$ for a Milstein scheme [Gil08b, Gil08a].

8.1.1 CIR model ZCB with MLMC

We consider the Cox-Ingersoll-Ross model (6.4.1) for the process $(v_t)_{t\geq 0}$ [CIR85]; the price of a zero-coupon bond (ZCB) with maturity T, at time t, reads

$$B(t,T) = \mathbb{E}\left[\exp\left(-\int_t^T v_s \mathrm{d}s\right)\bigg|\mathcal{F}_t\right],$$

which admits a closed-form solution [CIR85, BM07]. This solution at time zero is $B(0,T) = A\exp(-Cv_0)$, where $\Lambda := \sqrt{\kappa^2 + 2\xi^2}$ and

$$A := \left(\frac{2\Lambda\exp\left[(\kappa+\Lambda)T/2\right]}{2\Lambda + (\kappa+\Lambda)(\exp T\Lambda - 1)}\right)^{2\kappa\theta/\xi^2}, \qquad C := \frac{2(\exp(T\Lambda)-1)}{2\Lambda + (\kappa+\Lambda)(\exp(T\Lambda)-1)}.$$

We consider a CIR model with parameters $(\kappa,\theta,\xi,v_0,T) = (2,1,0.5,1,1)$, $(N,M,L) = (2000000,4,5)$, and RMSE thresholds $(0.001, 0.0005, 0.0002, 0.0001, 0.00005)$.

In Figure 8.1, we compute the standard Monte Carlo, and MLMC approximations for the ZCB. The first plot demonstrates the average variance for the approximations P_l and the differences $P_l - P_{l-1}$. Observe that the variance of the differences decreased roughly twice as fast as the rate of weak convergence of an Euler scheme. Also, the variance of P_l is asymptotically a constant. The second plot shows the mean of P_l and the mean of $P_l - P_{l-1}$. The third plot shows how decreasing the target ε, require more steps N_l and the number of levels increasing from 3 to 5. The fourth plot shows the ratio of savings between the standard Monte Carlo approach for approximating the bond price (Std MC), and the MLMC counterpart. The ratio of savings is a factor of 27 for $\varepsilon = 0.00005$ between the standard Monte Carlo and the MLMC approach. We adapt code freely available from [Gil08b].

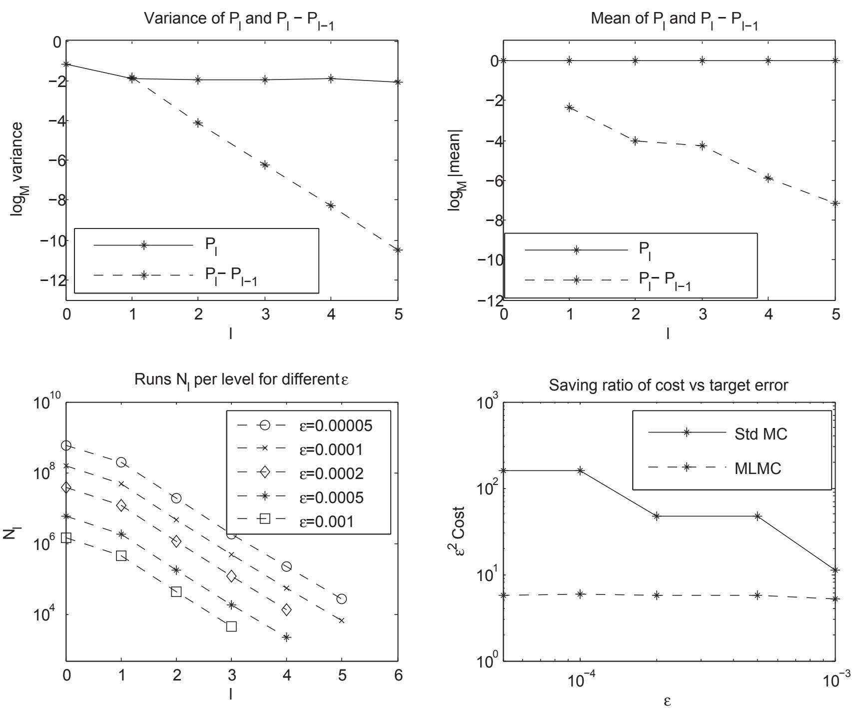

Figure 8.1: CIR model, and ZCB pricing using MLMC.

8.2 Accelerating the modified Euler-Maruyama scheme

Accelerated Euler-Maruyama schemes are studied in [TY12]. Suppose that the process X^ε depends on some small parameter ε, and consider a discretisation $\hat{X}^\varepsilon$. Let X^0 be another process with parameter $\varepsilon = 0$, and let $\hat{X}^0$ be its discretised process. Suppose that the bias of the process $X^\varepsilon - \hat{X}^\varepsilon$ is similar to the bias of $X^0 - \hat{X}^0$; then we can consider $\hat{X}^\varepsilon - \hat{X}^0 + X^0$ as an approximation of X^ε, which is a control variate method.

Example 8.2.1. *Consider the solution to the stochastic differential equations*

$$\begin{aligned} dS_t &= \sqrt{\alpha_t} S_t^\beta dB_t^1 \,, & S_0 = s_0 > 0 \,, \\ d\alpha_t &= \varepsilon \alpha_t (\rho dB_t^1 + \sqrt{1-\rho^2} dB_t^2) \,, & \alpha_0 > 0 \,. \end{aligned} \tag{8.2.1}$$

We define $\acute{S}_t^\varepsilon := \hat{S}_t^\varepsilon - \hat{S}_t^0 + S_t^0$, where $\hat{S}^\varepsilon$ is the discretised version using the modified Euler scheme, for some ε, and $(S_t^0)_{t\geq 0}$ is simulated using the Milstein scheme. We compare $(\hat{S}_t^\varepsilon)_{t\geq 0}$ and the accelerated $(\acute{S}_t^\varepsilon)_{t\geq 0}$ against S_t^ε (using a large number of time steps and the Milstein scheme). In Figure 8.2, we consider the model with with parameters $(S_0, \beta, \alpha_0, \varepsilon, \rho, T) =$

$(100, 0.9, 0.4, 0.1, -0.7, 1)$. We compute the strong error using $M = 10000$ and 2^{11} steps for the Milstein scheme which is used as a reference solution. The constant for the accelerated scheme is much smaller, demonstrating the merit of this approach.

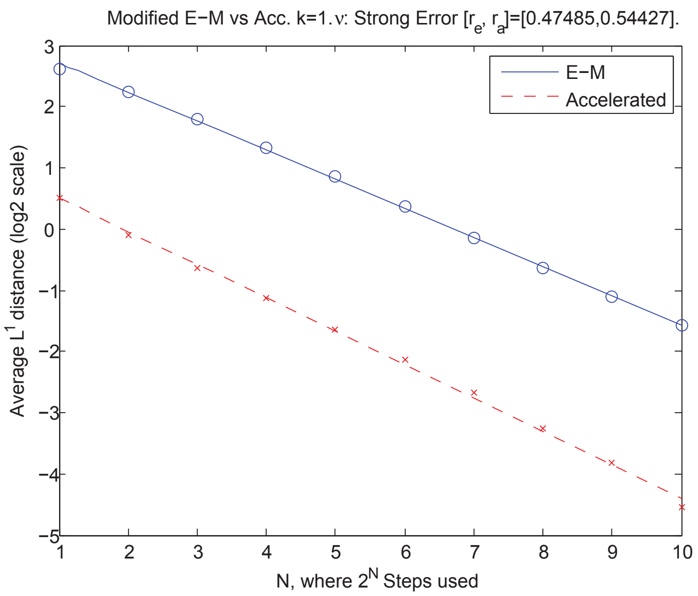

Figure 8.2: Strong convergence for the modified Euler and the Accelerated scheme.

Appendices

A. A class of approximate Greek weights

A.1 Proofs from Chapter 2

Proof of Proposition 2.2.1: We show the $d = m = 2$ case, which naturally extends to the d-dimensional case. Note that $u_\vartheta = \mathbb{E}\left[g(X_T)|\mathcal{F}_\vartheta\right]$ from the conditional expectation. Multiplying the Itô expansion of u_ϑ by weights $I_\vartheta^{(1)}$ and $I_\vartheta^{(2)}$, taking expectations and using Lemma 2.2.1 yields

$$\begin{aligned}
\mathbb{E}\left[u_\vartheta I_\vartheta^{(1)}\right] &= \mathbb{E}\left[I_\vartheta^{(1)}\left[u_0^{(1)}\right]I_\vartheta^{(1)}\right] \\
&\quad +\mathbb{E}\left[I_\vartheta^{(0,0)}\left[u_\cdot^{(0,0)}\right]I_\vartheta^{(1)}\right]+\mathbb{E}\left[I_\vartheta^{(0,1)}\left[u_\cdot^{(0,1)}\right]I_\vartheta^{(1)}\right]+\mathbb{E}\left[I_\vartheta^{(1,0)}\left[u_\cdot^{(1,0)}\right]I_\vartheta^{(1)}\right] \\
&= \vartheta u_0^{(1)} \\
&\quad +\mathbb{E}\left[I_\vartheta^{(0,0)}\left[u_0^{(0,0)}+I_s^{(0)}\left[u_\cdot^{(0,0,0)}\right]+I_s^{(1)}\left[u_\cdot^{(1,0,0)}\right]+I_s^{(2)}\left[u_\cdot^{(2,0,0)}\right]\right]I_\vartheta^{(1)}\right] \\
&\quad +\mathbb{E}\left[I_\vartheta^{(0,1)}\left[u_0^{(0,1)}+I_s^{(0)}\left[u_\cdot^{(0,0,1)}\right]+I_s^{(1)}\left[u_\cdot^{(1,0,1)}\right]+I_s^{(2)}\left[u_\cdot^{(2,0,1)}\right]\right]I_\vartheta^{(1)}\right] \\
&\quad +\mathbb{E}\left[I_\vartheta^{(1,0)}\left[u_0^{(1,0)}+I_s^{(0)}\left[u_\cdot^{(1,0,1)}\right]+I_s^{(1)}\left[u_\cdot^{(1,1,0)}\right]+I_s^{(2)}\left[u_\cdot^{(2,1,0)}\right]\right]I_\vartheta^{(1)}\right] \\
&= hu_0^{(1)}+\frac{\vartheta^2}{2}\left[u_0^{(0,1)}+u_0^{(1,0)}\right] \\
&\quad +\mathbb{E}\left[I_\vartheta^{(0,0,1)}\left[u_\cdot^{(0,0,1)}\right]I_\vartheta^{(1)}\right]+\mathbb{E}\left[I_\vartheta^{(0,1,0)}\left[u_\cdot^{(0,1,0)}\right]I_\vartheta^{(1)}\right]+\mathbb{E}\left[I_\vartheta^{(1,0,0)}\left[u_\cdot^{(1,0,0)}\right]I_\vartheta^{(1)}\right] \\
&= \vartheta u_0^{(1)}+\mathcal{O}(\vartheta^2).
\end{aligned}$$

Dividing through by ϑ yields the first result $\mathbb{E}[u_\vartheta I_\vartheta^{(1)}/\vartheta] = u_0^{(1)} + \mathcal{O}(\vartheta)$. Similar analysis yields $\mathbb{E}[u_\vartheta I_\vartheta^{(2)}/\vartheta] = u_0^{(2)} + \mathcal{O}(\vartheta)$ and this completes the proof of Proposition 2.2.1. □

Proof of Proposition 2.3.1: Using Lemma 2.2.1 and an expansion similar to (2.2.7), with $\beta = (l)$ and $k_j(\beta) = 0$ (for j=0,1), it follows that

$$\begin{aligned}
\mathbb{E}\left[g(X_T)\frac{W_\vartheta^{(l)}}{\vartheta}\right] &= u_0^{(l)}+\frac{1}{\vartheta}\left(\sum_{i=2}^k\sum_{\alpha\in\mathcal{M}_{i,1,l}}u_0^\alpha\frac{\vartheta^{w(\alpha,\beta)}}{w(\alpha,\beta)!}\prod_{j=0}^{l\left(\alpha^+\right)}C_{k_j(\alpha)+k_j(\beta)}^{k_j(\alpha)}\right)+\mathcal{O}\left(\vartheta^k\right) \\
&= u_0^{(l)}+\frac{1}{\vartheta}\left(\sum_{i=2}^k\sum_{\alpha\in\mathcal{M}_{i,1,l}}u_0^\alpha\frac{\vartheta^i}{i!}\prod_{j=0}^{l\left(\alpha^+\right)}C_{k_j(\alpha)}^{k_j(\alpha)}\right)+\mathcal{O}\left(\vartheta^k\right) \\
&= u_0^{(l)}+\frac{1}{\vartheta}\left(\sum_{i=2}^k\sum_{\alpha\in\mathcal{M}_{i,1,l}}u_0^\alpha\frac{\vartheta^i}{i!}\right)+\mathcal{O}\left(\vartheta^k\right),
\end{aligned}$$

where the convention $\sum_{i=2}^1 i = 0$ is used. The proof is completed by noting that $u_0^{\alpha*(0)}$ is equal to zero for all $\alpha \in \mathcal{M}$, from the partial differential equation (2.1.1). $\square$

Proof of Proposition 2.3.2: By continuing from (2.2.7) and recalling the hierarchical set $\mathcal{D}_2$, u_ϑ has the following Itô-Taylor expansion:

$$u_\vartheta = \sum_{\alpha \in \mathcal{D}_2} I_\vartheta^\alpha \left[u_0^\alpha\right] + \sum_{\alpha \in \mathcal{B}(\mathcal{D}_2)} I_\vartheta^\alpha \left[u_\cdot^\alpha\right]. \tag{A.1.1}$$

Using Lemma 2.2.1 to simplify the expectation of $\mathbb{E}[u_\vartheta I_\vartheta^\alpha]$, for $\alpha = (1,1)$, it follows that

$$\begin{aligned}
\mathbb{E}\left[u_\vartheta I_\vartheta^{(1,1)}\right] = \;& \mathbb{E}\left[I_\vartheta^{(1,1)}\left[u_0^{(1,1)}\right] I_\vartheta^{(1,1)}\right] + \mathbb{E}\left[I_\vartheta^{(0,0,0)}\left[u_\cdot^{(0,0,0)}\right] I_\vartheta^{(1,1)}\right] + \mathbb{E}\left[I_\vartheta^{(1,0,0)}\left[u_\cdot^{(1,0,0)}\right] I_\vartheta^{(1,1)}\right] \\
& + \mathbb{E}\left[I_\vartheta^{(0,0,1)}\left[u_\cdot^{(0,0,1)}\right] I_\vartheta^{(1,1)}\right] + \mathbb{E}\left[I_\vartheta^{(1,0,1)}\left[u_\cdot^{(1,0,1)}\right] I_\vartheta^{(1,1)}\right] \\
& + \mathbb{E}\left[I_\vartheta^{(0,1,0)}\left[u_\cdot^{(0,1,0)}\right] I_\vartheta^{(1,1)}\right] + \mathbb{E}\left[I_\vartheta^{(1,1,0)}\left[u_\cdot^{(1,1,0)}\right] I_\vartheta^{(1,1)}\right] \\
& + \mathbb{E}\left[I_\vartheta^{(0,1,1)}\left[u_\cdot^{(0,1,1)}\right] I_\vartheta^{(1,1)}\right] \\
= \;& \mathbb{E}\left[I_\vartheta^{(1,1)}\left[u_0^{(1,1)}\right] I_\vartheta^{(1,1)}\right] + \sum_{\alpha \in \mathcal{M}_{3,2,1}} \mathbb{E}\left[I_\vartheta^\alpha\left[u_\cdot^\alpha\right] I_\vartheta^{(1,1)}\right].
\end{aligned}$$

In addition, the following equalities can be shown

$$\mathbb{E}\left[I_\vartheta^{(1,1)} I_\vartheta^{(1,1)}\right] = \frac{\vartheta^2}{2!}, \qquad \mathbb{E}\left[I_\vartheta^{(1,1,0)} I_\vartheta^{(1,1)}\right] = \mathbb{E}\left[I_\vartheta^{(1,0,1)} I_\vartheta^{(1,1)}\right] = \mathbb{E}\left[I_\vartheta^{(0,1,1)} I_\vartheta^{(1,1)}\right] = \frac{\vartheta^3}{3!},$$
$$\mathbb{E}\left[I_\vartheta^{(1,1,0,0)} I_\vartheta^{(1,1)}\right] = \mathbb{E}\left[I_\vartheta^{(1,0,1,0)} I_\vartheta^{(1,1)}\right] = \ldots = \mathbb{E}\left[I_\vartheta^{(0,0,1,1)} I_\vartheta^{(1,1)}\right] = \frac{\vartheta^4}{4!},$$

and the expansion of $\mathbb{E}[u_\vartheta I_\vartheta^{(1,1)}]$ simplifies to $\frac{\vartheta^2}{2} u_0^{(1,1)} + \mathcal{O}(\vartheta^3)$. It follows that

$$\mathbb{E}\left[u_\vartheta \frac{2 I_\vartheta^{(1,1)}}{\vartheta^2}\right] = u_0^{(1,1)} + \mathcal{O}(\vartheta). \tag{A.1.2}$$

We can continue to do further expansions, in addition to noting that $u_0^{\alpha*(0)}$ is equal to zero for all $\alpha \in \mathcal{M}$. This concludes the proof of Proposition 2.3.2. $\square$

A.2 Proofs from Chapter 4

Proof of Theorem 4.3.1: i) By a telescoping sum it follows that

$$\mathbb{E}\Big[\Gamma_h^\phi g(\hat{X}_T)\Big] = \mathbb{E}\Big[\Gamma_h^\phi u(t_n, \hat{X}_{t_n})\Big] = \mathbb{E}\Bigg[\Gamma_h^\phi \sum_{i=1}^{n-1} \Big\{u(t_{i+1}, \hat{X}_{t_{i+1}}) - u(t_i, \hat{X}_{t_i})\Big\}\Bigg] + \mathbb{E}\Big[\Gamma_h^\phi u(t_1, \hat{X}_{t_1})\Big], \tag{A.2.1}$$

from the second part of Lemma 4.1.4 note that $\mathbb{E}\Big[\Gamma_h^\phi u(h, \hat{X}_h)\Big] = \hat{L}_x^{(1,1)} u(0,x) + C_1 h + \mathcal{O}(h^2)$, where $h = t_1$.

ii) We recall (3.3.2). For $(s,x) \in [0,T] \times \mathbb{R}^d$, define

$$\phi_e^1(s,x) := \frac{1}{2}\hat{L}_x^{(0,0)} u(s,x)\ ,\ \phi_e^2(s,x) := \frac{1}{6}\hat{L}_x^{(0,0,0)} u(s,x).$$

With this notation, we obtain,

$$\mathbb{E}\Big[\Gamma_h^\phi \{g(\hat{X}_T) - u(h, \hat{X}_h)\}\Big] = \mathbb{E}\Bigg[\Gamma_h^\phi \Bigg(\sum_{i=1}^{n-1} \mathbb{E}_{t_1}\Big[h^2 \phi_e^1(t_i, \hat{X}_{t_i}) + h^3 \phi_e^2(t_i, \hat{X}_{t_i})\Big] + \mathcal{O}(h^4)\Bigg)\Bigg]. \tag{A.2.2}$$

From [TT90, Theorem 1], we know that

$$\mathbb{E}_{t_1}\Big[\phi_e^1(t_i, \hat{X}_{t_i})\Big] = \mathbb{E}_{t_1}\Big[\phi_e^1(t_i, X_{t_i}^{t_1, \hat{X}_{t_1}})\Big] + h\tilde{\phi}_{e,i}^1(t_1, \hat{X}_{t_1}) + \mathcal{O}(h^2)\ ,$$

for some bounded function $\tilde{\phi}_{e,i}^1$, and

$$\mathbb{E}_{t_1}\Big[\phi_e^2(t_i, \hat{X}_{t_i})\Big] = \mathbb{E}_{t_1}\Big[\phi_e^2(t_i, X_{t_i}^{t_1, \hat{X}_{t_1}})\Big] + \mathcal{O}(h).$$

Combining these equalities with (A.2.2), we obtain

$$\begin{aligned}\mathbb{E}\Big[\Gamma_h^\phi \{g(\hat{X}_T) - u(h, \hat{X}_h)\}\Big] &= \mathcal{O}(h^2) + h^3 \mathbb{E}\Big[\Gamma_h^\phi \textstyle\sum_{i=1}^{n-1} \tilde{\phi}_{e,i}^1(t_1, \hat{X}_{t_1})\Big] \\ &+ \mathbb{E}\Bigg[\Gamma_h^\phi \Bigg(\textstyle\sum_{i=1}^{n-1} \mathbb{E}_{t_1}\Big[h^2 \phi_e^1(t_i, X_{t_i}^{t_1, \hat{X}_{t_1}}) + h^3 \phi_e^2(t_i, X_{t_i}^{t_1, \hat{X}_{t_1}})\Big]\Bigg)\Bigg],\end{aligned} \tag{A.2.3}$$

using the Cauchy-Schwarz inequality and the variance of weight Γ_h^ϕ. Using a natural extension to Lemma 3.2.1, observe that

$$\mathbb{E}\Bigg[\Gamma_h^\phi \sum_{i=1}^{n-1} \tilde{\phi}_{e,i}^1(t_1, \hat{X}_{t_1})\Bigg] = \sum_{i=1}^{n-1} \Big[L^{(j,j)} \tilde{\phi}_{e,i}^1(0,x) + \mathcal{O}(h)\Big] = \mathcal{O}\left(\frac{1}{h}\right).$$

We also compute

$$\sum_{i=1}^{n-1} h\mathbb{E}_{t_1}\Big[\phi_e^2(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\Big] = \mathbb{E}_{t_1}\Big[\int_{t_1}^T \phi_e^2(s, X_s^{t_1,\hat{X}_{t_1}})\mathrm{d}s\Big] + \mathcal{O}(h) = \varphi_e^2(t_1, \hat{X}_{t_1}) + \mathcal{O}(h),$$

leading to

$$h^2\mathbb{E}\Bigg[\Gamma_h^\phi h \sum_{i=1}^{n-1} \phi_e^2(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\Bigg] = h^2\left(L^{(j,j)}\varphi_e^2(0,x) + \mathcal{O}(1)\right) = \mathcal{O}(h^2),$$

where $\varphi_e^2(t,x) := \mathbb{E}\Big[\int_t^T \phi_e^2(s, X_s)\mathrm{d}s\Big]$. Now we have, using Lemma 3.3.1,

$$\begin{aligned}\sum_{i=1}^{n-1} h\mathbb{E}_{t_1}\Big[\phi_e^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\Big] &= \mathbb{E}_{t_1}\Big[\int_{t_1}^T \phi_e^1(s, X_s^{t_1,\hat{X}_{t_1}})\mathrm{d}s\Big] + \frac{h}{2}\mathbb{E}_{t_1}\Big[\int_{t_1}^T L^{(0)}\phi_e^1(s, X_s^{t_1,\hat{X}_{t_1}})\mathrm{d}s\Big] + \mathcal{O}(h^2)\\ &= \varphi_e^1(t_1, \hat{X}_{t_1}) + \tilde{\varphi}_e^1(t_1, \hat{X}_{t_1})h + \mathcal{O}(h^2),\end{aligned}$$

such that $\varphi_e^1 \in \mathcal{G}_b^2$ and $\tilde{\varphi}_e^1 \in \mathcal{G}_b^1$. We compute

$$h\mathbb{E}\Bigg[\Gamma_h^\phi h \sum_{i=1}^{n-1} \phi_e^1(t_i, X_{t_i}^{t_1,\hat{X}_{t_1}})\Bigg] = hL^{(j,j)}\varphi_e^1(0,x) + \mathcal{O}(h^2);$$

combining the above results yields

$$\mathbb{E}\Big[\Gamma_h^\phi\{g(\hat{X}_T) - u(h, \hat{X}_h)\}\Big] = hL^{(1,1)}\varphi_e^1(0,x) + \mathcal{O}(h^2). \tag{A.2.4}$$

The proof is completed since (A.2.1) and (A.2.4) justify the extrapolation. □

A.3 Proofs from Chapter 5

Proof of Corollary 5.1.1: Multiplying the value function by $I_\vartheta^{(1)}$ and $I_\vartheta^{(2)}$, taking expectations, applying Lemma 2.2.1, and expansion in ε yield

$$\begin{aligned}\mathbb{E}\left[u_\vartheta^\varepsilon I_\vartheta^{(1)}\right] &= \mathbb{E}\left[I_\vartheta^{(1)}\left[L^{(1)}u_0^\varepsilon\right]I_\vartheta^{(1)}\right] + \mathbb{E}\left[I_\vartheta^{(1,0)}\left[L^{(1,0)}u_\cdot^\varepsilon\right]I_\vartheta^{(1)}\right] + \mathbb{E}\left[I_\vartheta^{(0,1)}\left[L^{(0,1)}u_\cdot^\varepsilon\right]I_\vartheta^{(1)}\right]\\ &= \vartheta L^{(1)}u_0^0 + \mathcal{O}(\vartheta\varepsilon) + \mathcal{O}(\vartheta^2) + \mathcal{O}(\vartheta^2\varepsilon),\\ \mathbb{E}\left[u_\vartheta^\varepsilon I_\vartheta^{(2)}\right] &= \mathbb{E}\left[I_\vartheta^{(2)}\left[L^{(2)}u_0^\varepsilon\right]I_\vartheta^{(2)}\right] + \mathbb{E}\left[I_\vartheta^{(2,0)}\left[L^{(2,0)}u_\cdot^\varepsilon\right]I_\vartheta^{(2)}\right] + \mathbb{E}\left[I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]I_\vartheta^{(2)}\right]\\ &= \vartheta L^{(2)}u_0^\varepsilon + \mathbb{E}\left[I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]I_\vartheta^{(2)}\right],\end{aligned} \tag{A.3.1}$$

since u_0^ε satisfies (2.1.1): $L^{(0)}u_0^\varepsilon = 0$. Thus, the Delta of the perturbed system reads

$$\Delta := \partial_x u_0^0 = \mathbb{E}\left[g(X_T^\varepsilon)\frac{I_\vartheta^{(1)}}{\vartheta\theta}\right] + \mathcal{O}(\varepsilon) + \mathcal{O}(\vartheta) + \mathcal{O}(\vartheta\varepsilon).$$

To compute the Vega, one needs an expansion of $\mathbb{E}\left[g(X_T^\varepsilon)\frac{I_\vartheta^{(2)}}{\vartheta\varepsilon}\right] = \mathbb{E}\left[u_\vartheta^\varepsilon\frac{I_\vartheta^{(2)}}{\vartheta\varepsilon}\right]$ obtained by expanding (A.3.1) up to at least orders $\mathcal{O}(\vartheta\varepsilon^2)$, $\mathcal{O}(\vartheta^2\varepsilon)$ and $\mathcal{O}(\vartheta^2\varepsilon^2)$ terms. First,

$$L^{(2)}u_0^\varepsilon = \varepsilon\partial_\theta u_0^\varepsilon = \varepsilon\left[\partial_\theta u_0^0 + \varepsilon\partial_\theta\left(\partial_\varepsilon u_0^\varepsilon\right)\Big|_{\varepsilon\downarrow 0} + \mathcal{O}(\varepsilon^2)\right], \tag{A.3.2}$$

then

$$\begin{aligned} L^{(0,2)}u_0^\varepsilon &= L^{(0)}\left[\varepsilon\partial_\theta u_0^\varepsilon\right] \\ &= \partial_t\left(\varepsilon\partial_\theta u_0^\varepsilon\right) + \tfrac{1}{2}\theta^2\partial_{xx}\left(\varepsilon\partial_\theta u_0^\varepsilon\right) + \tfrac{1}{2}\varepsilon^2\partial_{\theta\theta}\left(\varepsilon\partial_\theta u_0^\varepsilon\right) + \theta\varepsilon\partial_{x\theta}\left(\varepsilon\partial_\theta u_0^\varepsilon\right) \\ &= \varepsilon\partial_{t\theta}u_0^\varepsilon + \tfrac{1}{2}\varepsilon\theta^2\partial_{xx\theta}u_0^\varepsilon + \mathcal{O}(\varepsilon^2) \\ &= \varepsilon\partial_{t\theta}u_0^0 + \tfrac{1}{2}\varepsilon\theta^2\partial_{xx\theta}u_0^0 + \mathcal{O}(\varepsilon^2), \end{aligned}$$

and similarly $L^{(0)}L^{(0)}L^{(2)}u_0^\varepsilon = \mathcal{O}(\varepsilon)$, thus

$$\begin{aligned} \mathbb{E}\left[I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]I_\vartheta^{(2)}\right] &= \tfrac{\vartheta^2}{2}\left[L^{(0,2)}u_0^\varepsilon\right] + \mathbb{E}\left[I_\vartheta^{(0,0,2)}\left[L^{(0,0,2)}u_\cdot^\varepsilon\right]I_\vartheta^{(2)}\right] \\ &= \tfrac{\vartheta^2\varepsilon}{2}\left(\partial_{t\theta}u_0^0 + \tfrac{1}{2}\theta^2\partial_{xx\theta}u_0^0 + \mathcal{O}(\varepsilon)\right) + \tfrac{\vartheta^3}{3!}L^{(0,0,0)}u_0^\varepsilon + \mathcal{O}(\vartheta^4\varepsilon). \end{aligned} \tag{A.3.3}$$

Therefore, combining (A.3.2) and (A.3.3), it follows that

$$\begin{aligned} \mathbb{E}\left[u_\vartheta^\varepsilon\frac{I_\vartheta^{(2)}}{\vartheta\varepsilon}\right] &= \tfrac{1}{\vartheta\varepsilon}\left(\vartheta L^{(2)}u_0^\varepsilon + \mathbb{E}\left[I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]I_\vartheta^{(2)}\right]\right) \\ &= \partial_\theta u_0^0 + \mathcal{O}(\varepsilon) + \tfrac{1}{\vartheta\varepsilon}\left(\mathcal{O}(\vartheta^2\varepsilon) + \mathcal{O}(\vartheta^2\varepsilon^2)\right) \\ &= \partial_\theta u_0^0 + \mathcal{O}(\varepsilon) + \mathcal{O}(\vartheta) + \mathcal{O}(\vartheta\varepsilon). \end{aligned}$$

Thus, the Vega of the perturbed system is $\mathcal{V} := \mathbb{E}\left[g(X_T^\varepsilon)\frac{I_\vartheta^{(2)}}{\vartheta\varepsilon}\right] + \mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon)$.

For higher-order Greeks, multiply the Itô-Taylor expansion of u_ϑ^ε by $I_\vartheta^{(1,1)}$ to obtain

$$\begin{aligned} \mathbb{E}\left[u_\vartheta^\varepsilon I_\vartheta^{(1,1)}\right] = &\ \mathbb{E}\left[\left(I_\vartheta^{(1,1)}\left[L^{(1,1)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,1)}\right] + \mathbb{E}\left[\left(I_\vartheta^{(0,0)}\left[L^{(0,0)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,1)}\right] \\ &+\mathbb{E}\left[\left(I_\vartheta^{(0,1)}\left[L^{(0,1)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,1)}\right] + \mathbb{E}\left[\left(I_\vartheta^{(1,0)}\left[L^{(1,0)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,1)}\right] \\ = &\ \mathbb{E}\left[\left(I_\vartheta^{(1,1)}\left[L^{(1,1)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,1)}\right] + \mathbb{E}\left[\left(I_\vartheta^{(0,1)}\left[L^{(0,1)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,1)}\right]. \end{aligned} \tag{A.3.4}$$

The following expansions,

$$L^{(1,1)}u_0^\varepsilon = \theta^2\partial_{xx}u_0^\varepsilon = \theta^2\partial_{xx}u_0^0 + \theta^2\varepsilon\partial_\varepsilon\left(\partial_{xx}u_0^\varepsilon\right)\Big|_{\varepsilon\downarrow 0} + \mathcal{O}(\varepsilon^2)$$

and $L^{(0,1,1)}u_0^\varepsilon = \theta^2\partial_{txx}u_0^0 + \frac{1}{2}\theta^2\partial_{xxxx}u_0^0 + \mathcal{O}(\varepsilon)$, yield

$$\begin{aligned}\mathbb{E}\left[I_\vartheta^{(1,1)}\left[L^{(1,1)}u_\cdot^\varepsilon\right]I_\vartheta^{(1,1)}\right] &= \mathbb{E}\left[I_\vartheta^{(1,1)}\left[L^{(1,1)}u_0^\varepsilon\right]I_\vartheta^{(1,1)}\right] + \mathbb{E}\left[I_\vartheta^{(0,1,1)}\left[L^{(0,1,1)}u_\cdot^\varepsilon\right]I_\vartheta^{(1,1)}\right] \\ &= \tfrac{\vartheta^2}{2}L^{(1,1)}u_0^\varepsilon + \mathcal{O}(\vartheta^3).\end{aligned} \tag{A.3.5}$$

Rearranging (A.3.5) yields an expression for the Gamma:

$$\Gamma := \partial_{xx}u_0^0 = \mathbb{E}\left[g(X_T^\varepsilon)\frac{2I_\vartheta^{(1,1)}}{\vartheta^2\theta^2}\right] + \mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon) + \mathcal{O}(\vartheta\varepsilon).$$

Now, multiply u_ϑ^ε by $I_\vartheta^{(2,2)}$, and consider the non-zero terms in the following expansion

$$\begin{aligned}\mathbb{E}\left[u_\vartheta^\varepsilon I_\vartheta^{(2,2)}\right] &= \mathbb{E}\left[\left(I_\vartheta^{(2,2)}\left[L^{(2,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right] + \mathbb{E}\left[\left(I_\vartheta^{(0,0)}\left[L^{(0,0)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right] \\ &\quad + \mathbb{E}\left[\left(I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right] + \mathbb{E}\left[\left(I_\vartheta^{(2,0)}\left[L^{(2,0)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right] \\ &= \mathbb{E}\left[\left(I_\vartheta^{(2,2)}\left[L^{(2,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right] + \mathbb{E}\left[\left(I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right].\end{aligned} \tag{A.3.6}$$

Consider the expansion

$$I_\vartheta^{(2,2)}\left[L^{(2,2)}u_\cdot^\varepsilon\right] = I_\vartheta^{(2,2)}\left[L^{(2,2)}u_0^\varepsilon\right] + I_\vartheta^{(0,2,2)}\left[L^{(0,2,2)}u_\cdot^\varepsilon\right] + I_\vartheta^{(1,2,2)}\left[L^{(1,2,2)}u_\cdot^\varepsilon\right] + I_\vartheta^{(2,2,2)}\left[L^{(2,2,2)}u_\cdot^\varepsilon\right],$$

and multiplying it by $I_\vartheta^{(2,2)}$ yields

$$\mathbb{E}\left[\left(I_\vartheta^{(2,2)}\left[L^{(2,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right] = \frac{\vartheta^2}{2}L^{(2,2)}u_0^\varepsilon + \mathbb{E}\left[I_\vartheta^{(0,2,2)}\left[L^{(0,2,2)}u_\cdot^\varepsilon\right]I_\vartheta^{(2,2)}\right].$$

Use the following expansions,

$$L^{(2,2)}u_0^\varepsilon = \varepsilon^2\partial_{\theta\theta}u_0^\varepsilon = \varepsilon^2\left(\partial_{\theta\theta}u_0^0 + \varepsilon\partial_\varepsilon(\partial_{\theta\theta}u_0^\varepsilon)\Big|_{\varepsilon\searrow 0} + \mathcal{O}(\varepsilon^2)\right),$$

and

$$\begin{aligned}L^{(0,2,2)}u_0^\varepsilon &= \partial_t\left(L^{(2,2)}u_0^\varepsilon\right) + \tfrac{1}{2}\theta^2\partial_{xx}\left(L^{(2,2)}u_0^\varepsilon\right) + \theta\varepsilon\partial_{x\theta}\left(L^{(2,2)}u_0^\varepsilon\right) + \tfrac{1}{2}\varepsilon^2\partial_{\theta\theta}\left(L^{(2,2)}u_0^\varepsilon\right) \\ &= \varepsilon^2\left(\partial_{t\theta\theta}u_0^0 + \tfrac{1}{2}\theta^2\partial_{xx\theta\theta}u_0^0\right) + \mathcal{O}(\varepsilon^3),\end{aligned}$$

it can be shown that the expansion of $\mathbb{E}\left[\left(I_\vartheta^{(2,2)}\left[L^{(2,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right]$ is

$$
\begin{aligned}
&\tfrac{\vartheta^2\varepsilon^2}{2}\left(\partial_{\theta\theta}u_0^0+\mathcal{O}(\varepsilon)\right)+\mathbb{E}\left[I_\vartheta^{(0,2,2)}\left[L^{(0,2,2)}u_\cdot^\varepsilon\right]I_\vartheta^{(2,2)}\right]\\
=\;&\tfrac{\vartheta^2\varepsilon^2}{2}\left(\partial_{\theta\theta}u_0^0+\mathcal{O}(\varepsilon)\right)+\tfrac{\vartheta^3}{3!}L^{(0,2,2)}u_0^\varepsilon+\mathbb{E}\left[I_\vartheta^{(0,0,2,2)}\left[L^{(0,0,2,2)}u_\cdot^\varepsilon\right]I_\vartheta^{(2,2)}\right]\\
=\;&\tfrac{\vartheta^2\varepsilon^2}{2}\partial_{\theta\theta}u_0^0+\mathcal{O}(\vartheta^2\varepsilon^3)+\tfrac{\vartheta^3\varepsilon^2}{3!}\left(\partial_{t\theta\theta}u_0^0+\tfrac{1}{2}\theta^2\partial_{xx\theta\theta}u_0^0\right)+\mathcal{O}(\vartheta^3\varepsilon^3)+\mathcal{O}(\vartheta^4\varepsilon^2).
\end{aligned}
$$

We now consider the second term in (A.3.6), after expanding and taking expectation with $I_\vartheta^{(2,2)}$, to obtain

$$
\mathbb{E}\left[\left(I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right] = \tfrac{\vartheta^3}{3!}L^{(2,0,2)}u_0^\varepsilon+\mathbb{E}\left[\left(I_\vartheta^{(2,0,2)}\left[L^{(2,0,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(2,2)}\right].
$$

With the aid of the expansions

$$
L^{(2,0,2)}u_0^\varepsilon=\varepsilon^2\partial_{t\theta\theta}u_0^0+\mathcal{O}(\varepsilon^3),\qquad L^{(0,2,0,2)}u_0^\varepsilon=\mathcal{O}(\varepsilon^2),
$$

one can obtain the following expansion for (A.3.6):

$$
\mathbb{E}\left[g(X_T^\varepsilon)I_\vartheta^{(2,2)}\right]=\tfrac{\vartheta^2\varepsilon^2}{2}\partial_{\theta\theta}u_0^0+\mathcal{O}(\vartheta^2\varepsilon^3)+\mathcal{O}(\vartheta^3\varepsilon^2)+\mathcal{O}(\vartheta^3\varepsilon^3).
$$

Therefore, the weight for the Vomma approximation is $\frac{2I_\vartheta^{(2,2)}}{\vartheta^2\varepsilon^2}$, so that

$$
\text{Vomma} := \partial_{\theta\theta}u_0^0=\mathbb{E}\left[g(X_T^\varepsilon)\frac{2I_\vartheta^{(2,2)}}{\vartheta^2\varepsilon^2}\right]+\mathcal{O}(\varepsilon)+\mathcal{O}(\vartheta)+\mathcal{O}(\vartheta\varepsilon). \tag{A.3.7}
$$

For cross-terms, multiply the value function by $I_\vartheta^{(1,2)}$ and consider the expectation

$$
\begin{aligned}
\mathbb{E}\left[g(X_T^\varepsilon)I_\vartheta^{(1,2)}\right] =\;& \mathbb{E}\left[\left(I_\vartheta^{(1,2)}\left[L^{(1,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,2)}\right]+\mathbb{E}\left[\left(I_\vartheta^{(0,0)}\left[L^{(0,0)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,2)}\right]\\
&+\mathbb{E}\left[\left(I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,2)}\right]+\mathbb{E}\left[\left(I_\vartheta^{(2,0)}\left[L^{(2,0)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,2)}\right]\\
=\;& \mathbb{E}\left[\left(I_\vartheta^{(1,2)}\left[L^{(1,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,2)}\right]+\mathbb{E}\left[\left(I_\vartheta^{(0,2)}\left[L^{(0,2)}u_\cdot^\varepsilon\right]\right)I_\vartheta^{(1,2)}\right].
\end{aligned}\tag{A.3.8}
$$

By considering the expansions,

$$
\begin{aligned}
L^{(1,2)}u_0^\varepsilon &= \theta\varepsilon\partial_{x\theta}u_0^\varepsilon=\theta\varepsilon\left(\partial_{x\theta}u_0^0+\varepsilon\partial_\varepsilon\left(\partial_{x\theta}u_0^\varepsilon\right)\Big|_{\varepsilon\searrow 0}+\mathcal{O}(\varepsilon^2)\right),\\
L^{(0,2)}u_0^\varepsilon &= \varepsilon\left(\partial_{t\theta}u_0^\varepsilon+\tfrac{\theta^2}{2}\partial_{xx\theta}u_0^\varepsilon\right)+\mathcal{O}(\varepsilon^2),\\
L^{(1,0,2)}u_0^\varepsilon &= \theta\partial_x\left[\varepsilon(\partial_{t\theta}u_0^\varepsilon+\tfrac{\theta^2}{2}\partial_{xx\theta}u_0^\varepsilon)+\mathcal{O}(\varepsilon^2)\right]\\
&= \varepsilon\left(\theta\partial_{tx\theta}u_0^0+\tfrac{\theta^3}{2}\partial_{xxx\theta}u_0^0\right)+\mathcal{O}(\varepsilon^2),
\end{aligned}
$$

the expansions of the two terms in (A.3.8) are computable. Using

$$\begin{aligned}
\mathbb{E}\left[I_\vartheta^{(1,2)}\left[L^{(1,2)}u^\varepsilon_\cdot\right]I_\vartheta^{(1,2)}\right] &= \mathbb{E}\left[I_\vartheta^{(1,2)}\left[L^{(1,2)}u^\varepsilon_\cdot\right]I_\vartheta^{(1,2)}\right]+\mathbb{E}\left[I_\vartheta^{(0,1,2)}\left[L^{(0,1,2)}u^\varepsilon_\cdot\right]I_\vartheta^{(1,2)}\right]\\
&= \tfrac{\vartheta^2}{2}L^{(1,2)}u_0^\varepsilon+\tfrac{\vartheta^3}{3!}L^{(0,1,2)}u_0^\varepsilon+\mathcal{O}(\vartheta^4\varepsilon)\\
&= \tfrac{\vartheta^2\theta\varepsilon}{2}\partial_{x\theta}u_0^\varepsilon+\mathcal{O}(\vartheta^3\varepsilon)\\
&= \tfrac{\vartheta^2\theta\varepsilon}{2}\left(\partial_{x\theta}u_0^0+\mathcal{O}(\varepsilon)\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[I_\vartheta^{(0,2)}\left[L^{(0,2)}u^\varepsilon_\cdot\right]I_\vartheta^{(1,2)}\right] &= \mathbb{E}\left[I_\vartheta^{(1,0,2)}\left[L^{(1,0,2)}u^\varepsilon_\cdot\right]I_\vartheta^{(1,2)}\right]+\mathbb{E}\left[I_\vartheta^{(0,0,2)}\left[L^{(0,0,2)}u^\varepsilon_\cdot\right]I_\vartheta^{(1,2)}\right]\\
&= \mathbb{E}\left[I_\vartheta^{(1,0,2)}\left[L^{(1,0,2)}u^\varepsilon_0\right]I_\vartheta^{(1,2)}\right]+\mathcal{O}(\vartheta^4\varepsilon)\\
&= \tfrac{\vartheta^3\varepsilon}{3!}\left(\theta\partial_{txx}u_0^0+\tfrac{\theta^3}{2}\partial_{xxx\theta}u_0^0\right)+\mathcal{O}(\vartheta^3\varepsilon^2),
\end{aligned}$$

we can express (A.3.8) as

$$\mathbb{E}\left[g(X_T^\varepsilon)I_\vartheta^{(1,2)}\right] = \tfrac{\vartheta^2\varepsilon\theta}{2}\partial_{x\theta}u_0^0+\mathcal{O}(\vartheta^3\varepsilon)+\mathcal{O}(\vartheta^2\varepsilon^2)+\mathcal{O}(\vartheta^3\varepsilon^2),$$

from which it follows that

$$\mathbb{E}\left[g(X_T^\varepsilon)\frac{2I_\vartheta^{(1,2)}}{\vartheta^2\varepsilon\theta}\right] = \partial_{x\theta}u_0^0+\mathcal{O}(\vartheta)+\mathcal{O}(\varepsilon)+\mathcal{O}(\vartheta\varepsilon).$$

For the second cross term, consider

$$\begin{aligned}
\mathbb{E}\left[g(X_T^\varepsilon)I_\vartheta^{(2,1)}\right] &= \mathbb{E}\left[\left(I_\vartheta^{(2,1)}\left[L^{(2,1)}u^\varepsilon_\cdot\right]\right)I_\vartheta^{(2,1)}\right]+\mathbb{E}\left[\left(I_\vartheta^{(0,0)}\left[L^{(0,0)}u^\varepsilon_\cdot\right]\right)I_\vartheta^{(2,1)}\right]\\
&\quad+\mathbb{E}\left[\left(I_\vartheta^{(0,1)}\left[L^{(0,1)}u^\varepsilon_\cdot\right]\right)I_\vartheta^{(2,1)}\right]+\mathbb{E}\left[\left(I_\vartheta^{(1,0)}\left[L^{(1,0)}u^\varepsilon_\cdot\right]\right)I_\vartheta^{(2,1)}\right]\\
&= \mathbb{E}\left[\left(I_\vartheta^{(2,1)}\left[L^{(2,1)}u^\varepsilon_\cdot\right]\right)I_\vartheta^{(2,1)}\right]+\mathbb{E}\left[\left(I_\vartheta^{(0,1)}\left[L^{(0,1)}u^\varepsilon_\cdot\right]\right)I_\vartheta^{(2,1)}\right].
\end{aligned}\tag{A.3.9}$$

The following expansions hold:

$$\begin{aligned}
L^{(2,1)}u_0^\varepsilon &= \varepsilon(\theta\partial_{x\theta}u_0^\varepsilon+\partial_x u_0^\varepsilon)\\
&= v\varepsilon\left(\partial_{x\theta}u_0^0+\varepsilon\partial_\varepsilon\left(\partial_{x\theta}u_0^\varepsilon\right)\Big|_{\varepsilon\searrow 0}+\mathcal{O}(\varepsilon^2)\right)+\varepsilon\left(\partial_x u_0^0+\varepsilon\partial_\varepsilon(\partial_x u_0^\varepsilon)\Big|_{\varepsilon\searrow 0}+\mathcal{O}(\varepsilon^2)\right),
\end{aligned}$$

and

$$\begin{aligned}
L^{(0,1)}u_0^\varepsilon &= \left(\theta\partial_{tx}u_0^\varepsilon+\tfrac{\theta^3}{2}\partial_{xxx}u_0^\varepsilon\right)+\varepsilon\theta^2\partial_{xxx}u_0^\varepsilon+\mathcal{O}(\varepsilon^2),\\
L^{(2,0,1)}u_0^\varepsilon &= \mathcal{O}(\varepsilon).
\end{aligned}$$

Now consider the two terms in (A.3.9) and compute their expansion. Using

$$\begin{aligned}\mathbb{E}\left[I_\vartheta^{(2,1)}\left[L^{(2,1)}u^\varepsilon_\cdot\right]I_\vartheta^{(2,1)}\right] &= \mathbb{E}\left[I_\vartheta^{(2,1)}\left[L^{(2,1)}u^\varepsilon_\cdot\right]I_\vartheta^{(2,1)}\right] + \mathbb{E}\left[I_\vartheta^{(0,2,1)}\left[L^{(0,2,1)}u^\varepsilon_\cdot\right]I_\vartheta^{(2,1)}\right]\\ &= \tfrac{\vartheta^2}{2}L^{(2,1)}u_0^\varepsilon + \tfrac{\vartheta^3}{3!}L^{(0,2,1)}u_0^\varepsilon + \mathcal{O}(\vartheta^4\varepsilon)\\ &= \tfrac{\vartheta^2}{2}(\theta\varepsilon\partial_{x\theta}u_0^\varepsilon + \varepsilon\partial_x u_0^\varepsilon) + \mathcal{O}(\vartheta^3\varepsilon)\\ &= \tfrac{\vartheta^2}{2}\left(\varepsilon\theta\partial_{x\theta}u_0^0 + \varepsilon\partial_x u_0^0 + \mathcal{O}(\varepsilon^2)\right)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[I_\vartheta^{(0,1)}\left[L^{(0,1)}u^\varepsilon_\cdot\right]I_\vartheta^{(2,1)}\right] &= \mathbb{E}\left[I_\vartheta^{(2,0,1)}\left[L^{(2,0,1)}u^\varepsilon_\cdot\right]I_\vartheta^{(2,1)}\right] + \mathbb{E}\left[I_\vartheta^{(0,0,1)}\left[L^{(0,0,1)}u^\varepsilon_\cdot\right]I_\vartheta^{(2,1)}\right]\\ &= \mathbb{E}\left[I_\vartheta^{(2,0,1)}\left[L^{(2,0,1)}u^\varepsilon_0\right]I_\vartheta^{(2,1)}\right] + \mathcal{O}(\vartheta^4\varepsilon)\\ &= \mathcal{O}(\vartheta^3\varepsilon),\end{aligned}$$

we can expand (A.3.9) as

$$\mathbb{E}\left[g(X_T^\varepsilon)I_\vartheta^{(2,1)}\right] = \frac{\vartheta^2}{2}(\varepsilon\theta\partial_{x\theta}u_0^0 + \varepsilon\partial_x u_0^0) + \mathcal{O}(\vartheta^3\varepsilon) + \mathcal{O}(\vartheta^2\varepsilon^2) + \mathcal{O}(\vartheta^3\varepsilon^2).$$

By recalling Remark 2.3.1, the cross-term sensitivity can be approximated by

$$\text{Vanna} := \partial_{x\theta}u_0^0 = \mathbb{E}\left[g(X_T^\varepsilon)\left(\frac{I_\vartheta^{(1)}I_\vartheta^{(2)}}{\vartheta^2\theta\varepsilon} - \frac{I_\vartheta^{(1)}}{2\vartheta\theta^2}\right)\right] + \mathcal{O}(\vartheta) + \mathcal{O}(\varepsilon).$$

Bibliography

[AA13] S. Alanko and M. Avellaneda. Reducing variance in the numerical solution of BSDEs. *Comptes Rendus Mathematique*, 351(3):135–138, 2013.

[Abe11] K. Abe. Pricing exotic options using MSL-MC. *Quantitative Finance*, 11(9):1379–1392, 2011.

[ABT08] F. Aiube, T. Baidya, and E. Tito. Stochastic processes of commodity prices: a particle filter approach. *Energy Economics*, 30:597–605, 2008.

[Alf05] A. Alfonsi. On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods and Applications*, 11(4):355–384, 2005.

[Alf08] A. Alfonsi. A second-order discretization scheme for the CIR process: application to the Heston model. *Preprint CERMICS hal-00143723*, 14, 2008.

[Alf13a] A. Alfonsi. Strong order one convergence of a drift implicit Euler scheme: Application to the CIR process. *Statistics & Probability Letters*, 83(2):602–607, 2013.

[Alf13b] A. Alfonsi. Strong order one convergence of a drift implicit euler scheme: Application to the cir process. *Statistics & Probability Letters*, 83(2):602–607, 2013.

[AP07] L. Andersen and V. Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007.

[AS96] Y. Ait-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial studies*, 9(2):385–426, 1996.

[BBD08] A. Berkaoui, M. Bossy, and A. Diop. Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. *ESAIM: Probability and Statistics*, 12, 2008.

[BC09] A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering*. Springer, 2009.

[BD04] M. Bossy and A. Diop. An efficient discretization scheme for one dimensional SDEs with a diffusion coefficient of the form $|x|^\alpha$, $\alpha \in [1/2, 1]$. Technical report, INRIA working paper, 2004.

[BDGT00] M. Broadie, J. Detemple, E. Ghysels, and O. Torrés. American options with stochastic dividends and volatility: A nonparametric investigation. *Journal of Econometrics*, 94(1-2):53–92, 2000.

[BDM10] M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state–space models. *Annals of the Institute of Statistical Mathematics*, 62:61–89, 2010.

[Bec80] S. Beckers. The Constant Elasticity of Variance Model and Its Implications For Option Pricing. *The Journal of Finance*, 35(3):661–673, 1980.

[Ben01] E. Benhamou. *Application of Malliavin calculus and Wiener chaos to option theory*. PhD thesis, London School of Economics and Political Science (University of London), 2001.

[BG96] M. Broadie and P. Glasserman. Estimating security price derivatives using simulation. *Management Science*, 42(2):pp. 269–285, 1996.

[BG12] S. Burgos and M. Giles. Computing Greeks using multilevel path simulation. In *Monte Carlo and Quasi-Monte Carlo Methods 2010*, pages 281–296. Springer, 2012.

[BLB08] P. Bickel, B. Li, and T. Bengtsson. Sharp failure rates for the bootstrap particle filter in high dimensions. *IMS Collections*, 3:318–329, May 2008.

[BM07] D. Brigo and F. Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Finance, 2007.

[BMV06] N. Batalova, V. Maroussov, and F. Viens. Selection of an optimal portfolio with stochastic volatility and discrete observations. *Transactions of the Wessex Institute on Modelling and Simulation*, 43:371–380, 2006.

[BS73] F. Black and M. Scholes. Pricing options on corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[BT04] B. Bouchard and N. Touzi. Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their applications*, 111(2):175–206, 2004.

[Cap08] L. Capriotti. Reducing the variance of likelihood Ratio Greeks in Monte Carlo. In *Simulation Conference, 2008. WSC 2008. Winter*, pages 587–593. IEEE, 2008.

[Car96] J. Carriere. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19(1):19 – 30, 1996.

[CC14] J.-F. Chassagneux and D. Crisan. Runge–Kutta schemes for backward stochastic differential equations. *The Annals of Applied Probability*, 24(2):679–720, 2014.

[CD00] D. Crisan and A. Doucet. Convergence of sequential Monte Carlo methods. *Signal Processing Group, Department of Engineering, University of Cambridge, Technical Report*, 1, 2000.

[CD02] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *Signal Processing, IEEE Transactions on*, 50(3):736 –746, March 2002.

[CDM09] P. Coquelin, R. Deguest, and R. Munos. Sensitivity analysis in HMMs with application to likelihood maximization. *Advances in Neural Information Processing Systems*, 22:387–395, 2009.

[CDML99] D. Crisan, P. Del Moral, and T. Lyons. Discrete filtering using branching and interacting particle systems. *Markov Processes and Related Fields*, 5:293–318, 1999.

[CG07] N. Chen and P. Glasserman. Malliavin Greeks without Malliavin calculus. *Stochastic Processes and their Applications*, 117(11):1689 – 1723, 2007.

[CGM07] O. Cappé, S. Godsill, and E. Moulines. An overview of existing methods and recent advances in Sequential Monte Carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007.

[Cha14] J.-F. Chassagneux. Linear Multistep Schemes for BSDEs. *SIAM Journal on Numerical Analysis*, 52(6):2815–2836, 2014.

[Cho05] N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Annals of Statistics*, 32(6):2385–2411, 2005.

[CIR85] J. Cox, J. Ingersoll, and S. Ross. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407, 1985.

[CLS99] R. Craine, L. Lochstoer, and K. Syrtveit. Estimation of Stochastic-Volatility Jump Diffusion Model. *Risk in Emerging Financial Markets: Prediction, Pricing, and Policy*, Sept, 1999.

[CMR05] O. Cappé, É. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer-Verlag, 2005.

[Cox75] J. Cox. Notes on Option Pricing I: Constant Elasticity of Variance Diffusions. *Working Paper Stanford University*, 1975.

[Cox96] J. Cox. The constant elasticity of variance option pricing model. *The Journal of Portfolio Management Special Issue*, 23(15-17), 1996.

[CR76] J. Cox and S. Ross. The Valuation of Options for Alternative Stochastic Processes. *Journal of Financial Economics*, 3:145–166, 1976.

[Cre12] D. Creal. A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews*, 31:245–296, 2012.

[CSTV07] P. Cheridito, H. Soner, N. Touzi, and N. Victoir. Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs. *Communications on Pure and Applied Mathematics*, 60(7):1081–1110, 2007.

[DdFG01] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[DGA00] A. Doucet, S. Godsill, and C. Andrieu. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[DGMO11] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo Smoothing For General State Space Hidden Markov Models. 21(6):2109–2145, 2011.

[DJ06] M. Davis and M. Johansson. Malliavin Monte Carlo Greeks for jump diffusions. *Stochastic Processes and their Applications*, 116(1):101 – 129, 2006.

[DJ08] A. Doucet and A. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *The Oxford Handbook of Nonlinear Filtering*, pages 656–704, 2008.

[DM04] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Spinger: New York, 2004.

[DM11] S. De Marco. Smoothness and asymptotic estimates of densities for SDEs with locally smooth coefficients and applications to square root-type diffusions. *The Annals of Applied Probability*, 21:1282–1321, 2011.

[DMD03] P. Del Moral and A. Doucet. On a class of genealogical and interacting Metropolis models. *Lecture Notes in Mathematics*, 1832:415–446, 2003.

[DMDJ12] P. Del Moral, A. Doucet, and A. Jasra. On Adaptive Resampling Procedures for Sequential Monte Carlo Methods. *Bernoulli*, 18(1):47, 2012.

[DMDS09] P. Del Moral, A. Doucet, and S. Singh. Forward Smoothing using Sequential Monte Carlo. Technical report, 2009.

[DMS14] R. Douc, E. Moulines, and D. Stoffer. *Nonlinear Time Series: Theory, Methods and Applications with R Examples*. CRC Press, 2014.

[DNS12] S. Dereich, A. Neuenkirch, and L. Szpruch. An Euler-type method for the strong approximation of the Cox–Ingersoll–Ross process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 468(2140):1105–1115, 2012.

[Dou05] R. Douc. Comparison of Resampling Schemes for Particle Filtering. In *4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 64–69, 2005.

[Duf06] D. Duffy. *Finite Difference methods in financial engineering: a Partial Differential Equation approach*. John Wiley & Sons, 2006.

[EKHM08] N. El Karoui, S. Hamadène, and A. Matoussi. Backward stochastic differential equations and applications. *Indifference pricing: theory and applications*, pages 267–320, 2008.

[EKP04] Y. El-Khatib and N. Privault. Computations of Greeks in a market with jumps via the Malliavin calculus. *Finance and Stochastics*, 8(2):161–179, 2004.

[EKPQ97] N. El Karoui, S. Peng, and M. Quenez. Backward stochastic differential equations in finance. *Mathematical finance*, 7(1):1–71, 1997.

[Fel54] W. Feller. Diffusion Processes in One Dimension. *Transactions of the American Mathematical Society*, 77(1):pp. 1–31, 1954.

[FLL+99] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3(4):391–412, 1999.

[FSW12] M. Freidlin, J. Szücs, and A. Wentzell. *Random perturbations of dynamical systems*, volume 260. Springer, 2012.

[FTW11] A. Fahim, N. Touzi, and X. Warin. A probabilistic numerical method for fully nonlinear parabolic PDEs. *The Annals of Applied Probability*, 21(4):1322–1364, 2011.

[FWT10] P. Fearnhead, D. Wyncoll, and J. Tawn. A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2):447–464, 2010.

[Gat11] J. Gatheral. *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons, 2011.

[GC00] S. Godsill and T. Clapp. Improvement Strategies for Monte Carlo Particle Filters. In *Sequential Monte Carlo Methods in Practice*, pages 139–158. Springer-Verlag, 2000.

[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.

[GHS00] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating value-at-risk. *Management Science*, 46:1349–1364, 2000.

[Gil08a] M. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, 2008.

[Gil08b] M. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[Gla03] P. Glasserman. *Monte Carlo methods in Financial Engineering*, volume 53. Springer, 2003.

[Gly89] P. Glynn. Optimization of stochastic systems via simulation. *Proceedings of the 21st conference on Winter simulation*, WSC '89:90–105, 1989.

[GM02] E. Gobet and R. Munos. Sensitivity analysis using Itô-Malliavin calculus and martingales. Application to stochastic optimal control. In *SIAM journal on Control and Optimization*. Citeseer, 2002.

[GR11] I. Gyöngy and M. Rásonyi. A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients. *Stochastic processes and their applications*, 121(10):2189–2200, 2011.

[GS12] M. Giles and L. Szpruch. Multilevel Monte Carlo methods for applications in finance. *arXiv:1212.1377*, 2012.

[GSS93] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2):107–113, April 1993.

[GY92] P. Glasserman and D. Yao. Some guidelines and guarantees for common random numbers. *Management Science*, 38(6):884–908, 1992.

[Gyö98] I. Gyöngy. A note on Euler's approximations. *Potential Analysis*, 8(3):205–216, 1998.

[Hes93] S. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[Hes97] S. Heston. A simple new formula for options with stochastic volatility. Course notes of Washington University in St. Louis, Missouri, 1997.

[HJ12] M. Hutzenthaler and A. Jentzen. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *arXiv:1203.5809*, 2012.

[HJK11] M. Hutzenthaler, A. Jentzen, and P. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 467(2130):1563–1576, 2011.

[HJK12] M. Hutzenthaler, A. Jentzen, and P. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *The Annals of Applied Probability*, 22(4):1611–1641, 2012.

[HKLW02] P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward. Managing Smile Risk. *Wilmott magazine*, 2002.

[HMS02] D. Higham, X. Mao, and A. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM Journal on Numerical Analysis*, 40(3):1041–1063, 2002.

[HSL08] X. Hu, T. Schon, and L. Ljung. A Basic Convergence Result for Particle Filtering. *Signal Processing, IEEE Transactions on*, 56(4):1337–1348, April 2008.

[Hul14] J. Hull. *Options, futures and other derivatives, 9th Edition*. Prentice Hall, 2014.

[HW87] J. Hull and A. White. The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance*, 42:281–300, 1987.

[IO05] P. Iseger and E. Oldenkamp. Cliquet Options: Pricing and Greeks in Deterministic and Stochastic Volatility Models. *Technical Report, SSRN eLibrary*, 2005.

[Jac05] P. Jackel. Stochastic Volatility Models: Past, Present and Future. *The Best of Wilmott I: Incorporating the Quantitative Finance Review*, pages 355–378, 2005.

[JDM10] A. Jasra and P. Del Moral. Sequential Monte Carlo Methods for Option Pricing. *Stoch. Analysis and Appl.*, May 2010.

[JKN09] A. Jentzen, P. Kloeden, and A. Neuenkirch. Pathwise approximation of stochastic differential equations on domains: higher order convergence rates without global Lipschitz coefficients. *Numerische Mathematik*, 112(1):41–64, 2009.

[JM11] A. Jacquier and C. Martini. Heston 2010. *Available at SSRN 1769744*, 2011.

[JMMS12] A. Jasra, J. Martin, E. McCoy, and S. Singh. Filtering via approximate Bayesian computation. *Statistics and Computing*, pages 1–15, 2012.

[JN11] J. Jiang and W. Ng. Tracking Order Book Price Densities with Kalman Filters. *International Journal of Information Processing and Management*, 2(2):15–22, 2011.

[JSDT11] A. Jasra, D. Stephens, A. Doucet, and T. Tsagaris. Inference for Lévy-Driven Stochastic Volatility Models via Adaptive Sequential Monte Carlo. *Scandinavian Journal of Statistics*, 38(1), 2011.

[JU97] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls*, 1997.

[Kal60] R. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME- Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[KdFD05] M. Klaas, N. de Freitas, and A. Doucet. Toward Practical N2; Monte Carlo: the Marginal Particle Filter. In *Uncertainty in Artificial Intelligence*. AUAI Press, 2005.

[KDSM09] N. Kantas, A. Doucet, S. Singh, and J. Maciejowski. An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. In *15th IFAC Symposium on System Identification (SYSID), Saint-Malo, France.*, volume 102, page 117, 2009.

[Kit96] G. Kitagawa. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

[KLW94] A. Kong, J. Liu, and W. Wong. Sequential imputation and Bayesian missing data problems. *J. Am. Statist. Assoc.*, 89:278–288, 1994.

[KN12] P. Kloeden and A. Neuenkirch. Convergence of numerical methods for stochastic differential equations in mathematical finance. *arXiv:1204.6620*, 2012.

[Koe11] J. Koebrugge. Estimating default probabilities for CDO's : a regime switching model. March 2011.

[KP92] P. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer Verlag, 1992.

[Kry90] N. Krylov. A simple proof of the existence of a solution of Itô's equation with monotone coefficients. *Theory of Probability & Its Applications*, 35:583–587, 1990.

[KS91] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Springer, 1991.

[KT01] N. Kunitomo and A. Takahashi. The asymptotic expansion approach to the valuation of interest rate contingent claims. *Mathematical Finance*, 11(1):117–151, 2001.

[KT03] N. Kunitomo and A. Takahashi. On validity of the asymptotic expansion approach in contingent claim analysis. *The Annals of Applied Probability*, 13(3):914–952, 2003.

[LC98] J. Liu and R. Chen. Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association*, 93(443):1032–1044, 1998.

[Lew00] A. Lewis. *Option valuation under stochastic volatility*. Finance press, 2000.

[LP14] V. Lemaire and G. Pages. Multilevel Richardson-Romberg extrapolation. *arXiv:1401.1177*, 2014.

[LS01] F. Longstaff and E. Schwartz. Valuing american options by simulation: A simple least-squares approach. *Review of Financial studies*, 14(1):113–147, 2001.

[LS07] M. Lombardi and S. Sgherri. (Un)Naturally Low? Sequential Monte Carlo Tracking of the US Natural Interest Rate. *ECB Working Paper Series*, (No. 794), 2007.

[MN03] S. Mikhailov and U. Nögel. Heston's Stochastic Volatility Model: Implementation, Calibration and Some Extensions. *Wilmott Magazine*, July, 2003.

[MTU04] R. Matsuoka, A. Takahashi, and Y. Uchida. A New Computational Scheme for Computing Greeks by the Asymptotic Expansion Approach. *Asia-Pacific Financial Markets*, 11(4):393–430, 2004.

[MY99] J. Ma and J. Yong. *Forward-Backward Stochastic Differential Equations and their Applications*, volume Lecture Notes in Mathematics 1702. Springer, 1999.

[NS12] A. Neuenkirch and L. Szpruch. First order strong approximations of scalar SDEs with values in a domain. *arXiv:1209.0390*, 2012.

[NT13] H. Ngo and D. Taguchi. Strong rate of convergence for the Euler-Maruyama approximation of stochastic differential equations with irregular coefficients. *arXiv:1311.2725*, 2013.

[OCDM08] J. Olsson, O. Cappé, R. Douc, and E. Moulines. Sequential Monte Carlo smoothing with application to parameter estimation in non-linear state space models. *Bernoulli*, 14(1):155–179, 2008.

[OTdF+04] K. Okuma, A. Taleghani, N. de Freitas, O. de Freitas, J. Little, and D. Lowe. A

Boosted Particle Filter: Multitarget Detection and Tracking. In *ECCV*, volume 1, pages 28–39, 2004.

[Pag07] G. Pages. Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity. *Monte Carlo Methods and Applications*, 13(1):37, 2007.

[Poy06] G. Poyiadjis. *Particle methods for parameter estimation in general state space models*. PhD thesis, University of Cambridge, 2006.

[Poy11] Poyiadjis, G. and Doucet, A. and Singh, S. Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, 98(1):65–80, 2011.

[PP92] E. Pardoux and S. Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. In *Stochastic partial differential equations and their applications*, pages 200–217. Springer, 1992.

[PS99] M. Pitt and N. Shephard. Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.

[RB10] B. Rambharat and A. Brockwell. Sequential Monte Carlo pricing of American-style options under stochastic volatility models. *Ann. Appl. Stat.*, 4(1):222–265, 2010.

[RC05] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[Rub89] R. Rubinstein. Sensitivity analysis and performance extrapolation for computer simulation models. *Operations Research*, 37(1):72–81, 1989.

[RV08] L. Rogers and L. Veraart. A Stochastic Volatility Alternative to SABR. *Journal of Applied Probability*, 45(4):1071–1085, 2008.

[RW00] L. Rogers and D. Williams. *Diffusions, Markov processes, and martingales*, volume 2. Cambridge University Press, Cambridge, 2000.

[Sab13] S. Sabanis. A note on tamed Euler approximations. *Electronic Communications in Probability*, 18:1–10, 2013.

[Sco87] L. Scott. Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application. *The Journal of Financial and Quantitative Analysis*, 22(4):419–438, 1987.

[SD08] S. Sivanandam and S. Deepa. *Introduction to genetic algorithms*. Springer, 2008.

[SMHP11] L. Szpruch, X. Mao, D. Higham, and J. Pan. Numerical simulation of a strongly nonlinear Ait-Sahalia-type interest rate model. *BIT Numerical Mathematics*, 51(2):405–425, 2011.

[Til93] J. Tilley. Valuing american options in a path simulation model. *Transactions of the Society of Actuaries*, 45(83):104, 1993.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic analysis and applications*, 8(4):483–509, 1990.

[TY12] H. Tanaka and T. Yamada. Strong Convergence for Euler-Maruyama and Milstein Schemes with Asymptotic Method. *ArXiv e-prints*, October 2012.

[VDP03] J. Vermaak, A. Doucet, and P. Pérez. Maintaining Multi-Modality through Mixture Tracking. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, 2003.

[Wes93] M. West. Approximating Posterior Distributions by Mixture. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(2):409–422, 1993.

[WH97] M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer Series in Statistics, Springer-Verlag, New York, 1997.

[Whi] D. Whitley. A genetic algorithm tutorial. *Statistics and Computing*, 4:65–85.

[Wil01] P. Wilmott. Cliquet Options and Volatility Models. *Wilmott Magazine*, pages 78–83, 2001.

[Wil06] P. Wilmott. *The Best of Wilmott*, volume 2. John Wiley & Sons, 2006.

[WSG10] N. Whiteley, S. Singh, and S. Godsill. Auxiliary Particle Implementation of the Probability Hypothesis Density Filter. *IEEE Transactions on Aerospace and Electronic Systems*, 46:1437–1454, 2010.

[Yan02] L. Yan. The Euler scheme with irregular coefficients. *The Annals of Probability*, 30(3):1172–1194, 2002.

[YK91] J. Yang and H. Kushner. A Monte Carlo method for sensitivity analysis and parametric optimization of nonlinear stochastic systems. *SIAM journal on control and optimization*, 29(5):1216–1249, 1991.

[Zha04] J. Zhang. A numerical scheme for BSDEs. *The Annals of Applied Probability*, 14(1):459–488, 2004.