VTB Capital

Advanced Volatility Representation Models

First Version : 16th May 2012
This Version : 19th September 2012

Version :

Author(s) : Hayssam Ismail

Abstract :

We study static and dynamic features of the implied volatility surface and the problems arising in its modelling. Non-parametric local volatility can match the smile perfectly, but not its dynamics. Stochastic volatility models like Heston and SABR show better dynamic behaviour. By combining both in the parametric local-stochastic volatility Hyperbolic-Hyperbolic model we can circumvent shortcomings of both approaches. Next, we consider non-parametric adjustments to the latter to fine tune the calibration to market prices.

VTB Capital Quantitative Research
14 Cornhill
London EC3V 3ND
Contact: hayssam.ismail@vtbcapital.com

Contents

Acknowledgements

I would like to thank many people who have helped me through the completion of this dissertation. First and foremost, I am very grateful to my supervisor Charles-Henri Roubinet, who offered me the opportunity of taking part in this research project. This work would not have been possible without his guidance, his invaluable advices being fundamental during the internship. A true embodiment of a mentor, he knew how to teach me important financial and mathematical concepts, encouraging me and supporting me even when accumulating many responsibilities of a team leader.

I am indebted to my colleague Manuel Abellan-Lopez for his untiring support during my internship. I could not count how many times he read drafts of this document, helping me with every inconsistency found, whether mathematical, financial or grammatical. Having had a very similar educational path than I, Manuel understood many of the difficulties I had when trying to fill the gap between theory and practice, and assisted me in deconstructing many problems that I have encountered, what is an inestimable lesson to my future professional life.

I would also like to thank Peter Jäckel, for its sound understanding of both theoretical and practical aspects of the volatility modelling had been instrumental in helping me to understand and overcome many difficulties. He brought to my attention many key concepts in this book.

I could not forget to thanks Ioannis Chryssochoos, Bertrand Étienne and Sérilyne Louk for their precious help during my experience within the VTB quantitative research team. I am eternally thankful to all my professors at University of São Paulo, École polytechnique and UPMC. To my parents, standing by my side in every little moment of my life. And many other friends who helped me along the way, including but not limited to Pedro Natal and Murilo Andrade, that have been motivating me since my first steps in mathematical finance, Augusto dos Anjos for his friendship of many years and Adriana Sousa, for everything.

1

Introduction

Since its publication in 1973, the model developed by Black and Scholes [BS73] has become a standard in the financial industry. Indeed, it provides with a hedging strategy, the **delta-hedging**, a **self-financing** portfolio that guarantees the option pay-off at expiry, with virtually no risk. As we assume that there is no arbitrage, the initial cost of this portfolio must be equal to the price of the option. Although the model has been shown to have some shortcomings due to its restrictive hypotheses, it is still largely used as a reference, notably for its robustness: one can show that even if using the Black-Scholes hedging strategy does not guarantee a perfect replication of the option, we still have a **super-hedging strategy**, as long as we are conservative in the choice of the **volatility** [EKJPS98].

In the Black-Scholes context, provided that the interest rates are deterministic, there is a one-to-one relationship between a Call price and the volatility. We call **implied volatility** the value that, when used as input to price a Call in the Black-Scholes model, yields its corresponding market value. Unfortunately, the use of a unique time-constant volatility as advocated by the model is not consistent with observed market prices. Rigorously, we do not have a Black-Scholes model, but a collection of models with different volatilities for each option strike and expiry, defining an **implied volatility surface**.

Still, traders like to think in terms of implied volatility because it is a unique scale to compare prices of products with different characteristics and to visualize the randomness of the market. However, nothing ensures self-consistence. For instance, a portfolio with 2 calls with different strikes cannot be hedged self-consistently in a Black-Scholes framework. What one is specially careful about is to use a surface that is arbitrage free, unfortunately, arbitrage can come from a variety of sources, including interpolating and extrapolating available market quotes.

The **local volatility** of Dupire [Dup94] and Derman-Kani [DK94] unified implied volatilities, in the sense that with this approach one can match exactly the Call prices for all expiries and strikes. It is arbitrage-free, and it may be a framework good enough to manage vanilla products. However, one shows that it does not reproduce correctly the dynamics of the market, leading to serious mispricing of products depending on the forward smile level. Indeed, predicting the wrong dynamics leads to constant recalibration, hence the need of models that behaves as expected at least approximately.

We could try to model the implied dynamics directly, what was attempted as early as in 1993 by Shimko [Shi93]. However, it is hard to guarantee that the volatility surface remains arbitrage-free. Gatheral [Gat04] shows that respecting even a few of the

necessary conditions that ensure the absence of arbitrage when constructing a volatility surface [CM05], like excluding calendar spreads and negative butterflies, is not an obvious task. Then one attempts to solve this problem by going back to the basic principles of mathematical finance. By the fundamental theorem of asset pricing [HP81], the existence of a **numéraire** probability measure in which traded securities are martingales is equivalent to the non-arbitrage principle. On that account, we look for an underlying price dynamics that is still martingale but is more flexible than the Geometrical Brownian Motion dynamics of Black-Scholes. Then, provided that one manages to compute prices in this model, the resulting volatility surface will be arbitrage-free by construction.

An implied volatility surface resulting from such a model is also expected to have the following properties:

- Be consistent with observed vanilla prices (showing the correct "smile" or "skew" pattern),
- It should also reproduce some known market features, as a relatively correct volatility dynamics, such as sticky-strike or sticky-moneyness.

Allowing for a **Stochastic Volatility**, one can reproduce at least approximately the dynamics of the volatility smile. A standard example of SV model is the Heston Model [Hes93], that has exact formulas for vanilla options. Another important example is the SABR Model [HKL02], that exhibits roughly the correct dynamics of the implied volatility, even though there is no closed formula available and the model is mostly used via an asymptotic expansion for the implied volatility. Nevertheless, one shows that any stochastic volatility can not reproduce short term skew observed in the market.

We attempt to overcome these difficulties by combining the preceding approaches in a **Local Stochastic Volatility** model. Associating both non-parametric local and stochastic volatility models has the advantage of calibrating exactly the current implied volatility surface while having good dynamic features. Yet another class of models known as jump-diffusion generate a very steep skew for shorter maturities (Andersen-Andreasen model [AA00]), and might be combined with stochastic volatility. However, discontinuous models are out of the scope of this document.

2

A non-parametric static representation of the randomness of the market

We consider an arbitrage-free market formed by two assets:

- S is the underlying asset, on which contingent claims are made,
- N is the numéraire process, *i.e.*, the unit of account appropriate to our framework.

Thanks to the numéraire process, this framework is general enough to accommodate many asset classes, as fixed income and commodities, but our nomenclature will come mainly from the equity asset case. The numéraire will be the risk-free interest rate, and the discounted spot, as a traded asset, will be a martingale.

The **Black-Scholes model** [BS73] advocates a price process having a Geometrical Brownian Motion dynamics:

$$\begin{cases} \dfrac{dS_t}{S_t} = r\,dt + \sigma\,dW_t, \\ \dfrac{dN_t}{N_t} = r\,dt \end{cases}$$

where r is the risk-free interest rate and σ is the volatility, both constant. In this model, the volatility is the time-normalized standard deviation of the returns. This is not directly observable and one has to come with an estimate. One possibility is to estimate it historically from the log-returns time-series. However, the market-driven point of view is to imply the volatility from the market quotes of very liquid derivatives, notably, European Calls and Puts. In contrast to the historical approach, this is a forward-looking measure, denoting the expectations of the market participants about the future movements of the underlying (because derivatives contracts are paid at a future date).

Under the assumptions of Black-Scholes model, one easily derives a closed formula for pricing Calls. Assuming that the risk-free interest rate is known, the Call price is a strictly increasing function of volatility, hence it can be inverted, and one can obtain a volatility parameter implied by market prices. However, one observes that every product generally induces a different implied volatility, and rigorously, every option in the market is handled by a different independent Black-Scholes model determined by its implied

volatility. Due to its historical importance, the Black-Scholes model is the default framework in financial modelling: most models are in a way or another extensions of it, and even when using other pricing dynamics, traders often quote prices in implied volatility, that had become a representation of the randomness of the market.

Volatility smile. We call smile a pattern of the implied volatility calculated across a range of options, in which the volatility of the ATM options are below the volatility of the deep ITM or OTM options. Usually when one says smile, one means the case in which both in- and out-of-the-money volatility levels are higher. One calls a downward sloping a *skew*.

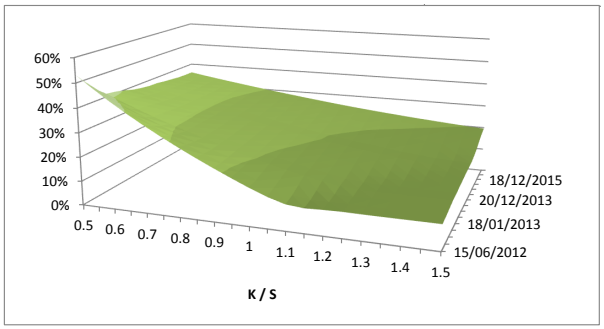

FIGURE 2.1: Implied volatility surface presenting skew phenomenon, FTSE Index, 24 May 2012 market close data. The strikes are relative to the forward price of the underlying.

Skews are commonly observed in equity markets (Figure 2.1). They can be explained by the loss-aversion of the investors: there is a great demand for out-of-the-money puts used as protection against fall of prices. Applying the same reasoning to FX markets, where the market is symmetrical (as a drop in an exchange rate implies a rise in the inverse exchange rate), we should expect to find a (symmetrical) smile, which is verified in practice (Figure 2.2).

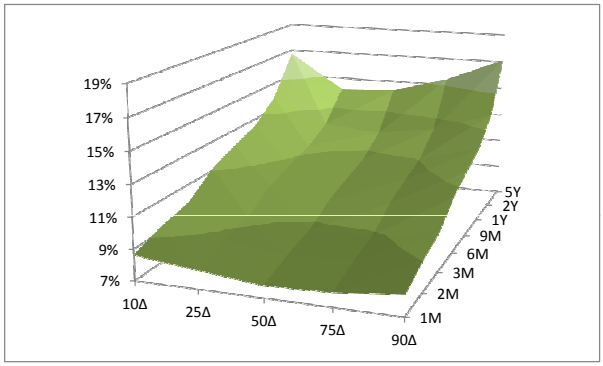

FIGURE 2.2: An example of a smile in the implied volatility in FX markets, exchange rate GBPUSD, 12 May 2012 market close data. The scale is in Black-Scholes deltas. A 50Δ is approximately at-the-money for short maturities and Δ increases as we go in-the-money.

With a small modification, the Black-Scholes model may allow for time-dependent yet deterministic risk-free rate r_t and volatility σ_t. The resulting marginal distributions are still log-normal, but the parameters must be adjusted as follows:

$$r \to \frac{1}{(T-t)} \int_t^T r_s \, ds; \qquad \sigma^2 \to \frac{1}{(T-t)} \int_t^T \sigma_s^2 \, ds.$$

Thus, time-structure of the volatility can be introduced in an *ad hoc* manner. But this is not enough to provide a consistent pricing across strikes. For correctly evaluating a digital option, for example, a more complex model may be needed[1].

Local volatility. Dupire [Dup94] and Derman-Kani [DK94] showed that if we allow the volatility to depend on the underlying price as well,

$$\frac{dS_t}{S_t} = r_t\,dt + \sigma_{Dup}(t, S_t)\,dW_t,$$

under risk neutral probability, there is a unique local volatility function that fits **exactly** all European Call prices. Derman and Kani have done it in a discrete framework, by using local volatility function in binomial trees. Dupire has derived the local volatility function in the continuous framework by using the following equation:

$$\frac{\partial C}{\partial T} = \frac{1}{2}\sigma^2(K, T; S_0)K^2\frac{\partial^2 C}{\partial K^2} - rK\frac{\partial C}{\partial K},$$

that can be found by noting that the second derivative of the call price coincides with the risk-neutral probability density function of the underlying [BL78] and hence should verify the forward Kolmogorov equation. The local volatility function resulting from this equation is called Dupire's volatility and is given by:

$$\sigma^2_{Dup}(T, K) = 2\frac{\partial_T C + r_T K\partial_K C}{K^2\partial^2_K C} \tag{2.2}$$

It is related to the Black-Scholes implied volatility by:

$$\sigma^2_{Dup}(T, K) = \frac{\frac{\partial w}{\partial T}}{1 - \frac{y}{w}\frac{\partial w}{\partial y} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{w} + \frac{y^2}{W^2}\right)\left(\frac{\partial w}{\partial y}\right) + \frac{1}{2}\frac{\partial^2 w}{\partial y^2}}, \tag{2.3}$$

where $y = \log(K/S_T)$ and $w = \Sigma^2 T$.

Despite being able to interpolate vanilla prices in a self-consistent way, Dupire's model is not able to reproduce the market dynamics. Gyöngy [Gyö86] established the theorem popularised in Finance as "Markovian Projection" [Pit07][2] stating that to emulate the continuous stochastic volatility diffusion $X_t = \sigma_t\,dW_t$ with a process $X^{\text{loc}}_t = \sigma_{\text{loc}}(t, X_t)\,dW_t$ one must have:

$$\sigma^2_{\text{loc}}(t, S_t) = \mathbb{E}[\sigma^2_t | S_t].$$

[1] The price of a digital option can be given as a function of the Call price, by differentiation under the expectation sign, as :

$$D(T, K) = \mathbb{E}[\mathbf{1}_{S_T \geq K}] = -\frac{\partial}{\partial K}C(T, K) \tag{2.1}$$

A Call price can be written with the implied volatility as $C(T, K) = C_{BS}(T, K, \Sigma(T, K))$, then:

$$D(T, K) = -\frac{\partial C_{BS}}{\partial K}(T, K, \Sigma) - \frac{\partial C_{BS}}{\partial \sigma}(T, K, \Sigma)\frac{\partial \sigma}{\partial K}(T, K, \Sigma).$$

Hence we notice that this price is directly dependent on the volatility skewness $\frac{\partial \sigma}{\partial K}(T, K, \Sigma)$, which cannot be introduced in the Black-Scholes model

[2] In the context of approximation methods, this result is particularly useful because if we develop a good asymptotic formula for the Call price and/or the implied volatility in a general local volatility framework, we can easily extend our formula to a stochastic volatility framework.

In other words, local volatility is nothing but the projection of the real variance onto the underlying process. It does not imply that the underlying dynamics **is** an one-dimensional diffusion, as noted by Dupire itself in his article. Unfortunately, knowledge about a conditional probability distribution does not provide any information about dynamics of the price (the joint probability distribution of would be necessary instead). Local volatility is a static approach, strictly connected with the current market distribution. In Dupire's model, the volatility smile diffuses in a trivial way, becoming flatter and flatter, which contradicts market observation, where the new volatility surface looks roughly the same as today's. Hence it does not provide with good hedging parameters [DFW95]. Besides, there is no obvious way to obtain implied volatilities from Dupire Local Volatility even if some asymptotic results are available [BBF02]. This approximation process may introduce arbitrage opportunities.

For European options, Dupire's local volatility may suffice, producing a self-consistent system of prices. However, the flattening of the forward smile leads to underpricing smile-dependent products, like Barrier options, forward-starting options[3] and cliquets[4]. The misvaluation of exotic options cliquet-based called Reverse Cliquet and Napoleon had caused big losses to many major institutions, as reported in [Jef04].

[3] A forward-starting option is an option that become active at a future date T_1 and expires further in the date $T_2 > T_1$. Upon activation the underlying level is unknown, so the strike is normally set such that the option will be at-the-money, $K = S_{T_1}$, but we can define other level of moneyness as well. Naturally it highly depends on the smile at the date T_1.

[4] A cliquet option is a chain of forward-starting options, where each new option is activated when the previous one expires. Each option of a cliquet is called a cliquette. The first cliquette is activated at the valuation time and it may be correctly evaluated by a local volatility model. But the remaining cliquettes will accumulate the mispricing of a series of forward-starting option. Let $T_0, \ldots, T_n$ be a set of fixing dates and

$$r_i = \frac{S_{T_i} - S_{T_{i-1}}}{S_{T_{i-1}}}$$

be the performance of the underlying between the dates T_{i-1} and T_i. The Reverse Cliquet guarantees the buyer a very high coupon lowered by the negatives performances of the underlying between:

$$\left(K - \sum_{i=1}^{n} r_i^-\right)^+ \tag{2.4}$$

and a Napoleon option consists in an option that guarantees a high coupon plus the worst performance in the period, that may be negative:

$$\left(K + \min_{i=1,\cdots,n} r_i\right)^+ \tag{2.5}$$

3

Dynamical features of Stochastic Volatility and standard models

The vanilla prices in the market does not have enough information to establish the dynamics of the implied volatility. We may content ourselves with having a static approach like local volatility and imposing some *ad hoc* dynamics based on our past observations. This dynamics can be very complicated in reality, but usually, the practitioners employ some simpler heuristics in order to quickly estimate the behaviour of volatility as the underlying changes. A stylized fact observed in markets is that the forward volatility surface shall look roughly like the current volatility. Two rules-of-thumb reproducing this stylized fact are sticky-strike and sticky-moneyness ([Der99], [DHS07]). One can show that these dynamics are though trivially arbitrageable (see Appendix B).

Instead of modelling directly implied volatility, we relax the assumptions on the underlying dynamics and reconstruct the implied volatility from the resulting prices, which guarantees no arbitrage. Indeed, by the fundamental theorem of pricing the no-arbitrage condition is equivalent to the existence of a numéraire measure in which the discounted asset $\{S_t/N_t\}_{t\in[0,T]}$ is a martingale.

Remark In the suitable numéraire, the underlying process is a drift-less process. As we are only interested in modelling the implied volatility, we will assume that the drift term is always null without loss of generality, by modelling the dynamic of the discounted underlying S_t/N_t that we will write also S_t for economy of notation.

3.1 Barrier options

Barrier options are an important yet simple example of why a dynamical approach must be needed in volatility modelling. Let consider the of a one-touch option, that is, an option that pays 1 if the underlying touches a given barrier during the life of the option.

Let us consider the log-price process in the Black-Scholes case. We assume in a first instant that the process have no cost of carry. Our dynamics is a Brownian motion:

$$dX_t = \sigma\, dW_t.$$

The Brownian motion has an interesting property of symmetry, as we see in the Fig. 3.1. It is a Markovian process and given that the underlying have touched the barrier, both red and blue trajectories have the same probability of happening. By this principle, we

have as many trajectories touching the barrier and finishing above it than trajectories touching the barrier and finishing below it. A digital Call is an option that pays 1 when the underlying expiries above a barrier. We conclude that, in this dynamics, the price of the One-Touch option is exactly twice the price of the Digital Call. The principle of symmetry can be easily generalized to the Geometrical Brownian motion and we will still have a relation holding between digital options and one-touch option in Black-Scholes model.

We have already seen that the Digital Call depends on the implied volatility smile (footnote in p. 6), but only statically: once we have a model that incorporates the volatility smile, its price is independent of dynamical features. This is easily understood if we visualize the Digital Call as a difference of two call options with infinitely close strikes. So it can be perfectly hedged by this Call spread, which price is completely determined by the static smile. However, one observes that the relation between digital prices and one-touch prices often does not hold in markets, which implies that the one-touch option is very dependent on the smile dynamics. It is a minimal example of why dynamical implied volatility models are needed, and one-touch options can be used as a verification of the quality of a model.

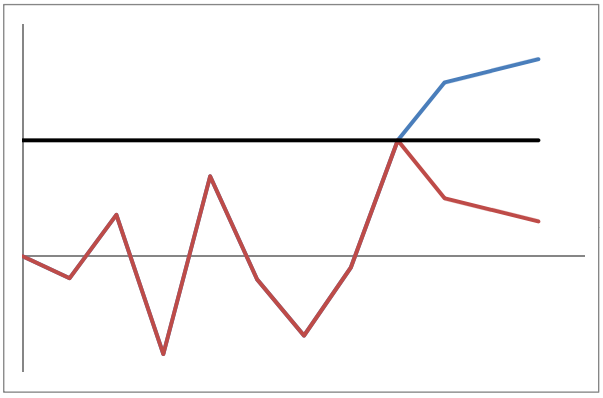

FIGURE 3.1: Principle of symmetry of the Brownian motion. By symmetry and Markovian property, the process have the same probability of taking the blue or the red path.

3.2 Heston Model

The **Heston model** [Hes93] has extended the Black-Scholes model by introducing a variance driven by a Square-Root process, obtained by assuming that the volatility is mean-reverting normal (*i.e.*, follows a Ornstein-Uhlenbeck process):

$$\begin{aligned} dX_t &= -\frac{\nu}{2}\,dt + \sqrt{\nu_t}\,dW_t \\ d\nu_t &= \kappa(\theta - \nu_t)\,dt + \alpha\sqrt{\nu_t}\,dB_t, \\ \langle dW, dB\rangle &= \rho dt \end{aligned}$$

Here, $X_t = \log S_t e^{-rt}$ is the log-discounted-price, α is the volatility of variance, θ is the long-term variance and κ is the mean-reversion coefficient. In order to have a strictly positive variance, parameters must respect the Feller condition:

$$\kappa\theta > \frac{1}{2}\alpha^2.$$

Pros

The principal asset of the Heston model is having a **semi-analytical form** for its calibration. As the characteristic function of the underlying is known, the Call price can be found through a Fourier inversion technique:

$$C_{\text{Heston}} = S_t - \frac{Ke^{-r(T-t)}}{2\pi} \int_{\frac{i}{2}-\infty}^{\frac{i}{2}+\infty} e^{-izX} \varphi_T(-z) \frac{dz}{z^2 - iz} \tag{3.1}$$

where X is the log-forward-price and $\varphi_T(z) = \mathbb{E}[e^{iz(X_T - X_t)} \mid \mathcal{F}_t]$. If we are able to calculate this characteristic function, we can compute prices by inverting this Fourier transform, which can be done by numerical integration or Fast Fourier Transform. Heston found that:

$$\varphi_T(z) = e^{C(z,\tau)+D(z,\tau)\nu_0} \tag{3.2}$$

with D satisfying a Riccati equation and C being easily calculated once D is obtained.

Cons

Despite the simple look of the closed formula, its implementation has caused issues during many years. The original version suggested by Heston in 1993 can be very unstable for long maturities or strong mean reversion, due to some issues raised by the calculation of complex logarithms. Only in 2005 Kahl and Jäckel have suggested an astute implementation that solved this problem [KJ05].

Bakshi et al. [BCC97] suggested an alternative implementation with a slightly different characteristic function, that seemed empirically to be stable without the modification of Kahl-Jäckel. Still, this conjecture remained unproven until 2007, when a definitive proof that both implementations were equivalent under any set of parameters and lead to stable results under some conditions was furnished by Kahl and Lord [KL07].

Andersen and Piterbarg have shown that some moments of the distribution of the Heston model explode in finite time [AP05b], and may produce divergent prices for many financial products, like CMS swaps. Kahl and Lord [KL07] show that this problem is strictly linked to the instability of the implementation of Heston model. We remark that the semi-analytical formula works only for European options, thus are only useful for calibration. Moreover, the calibrated parameters often do not respect Feller condition and the variance can assume negative values, which is simply inconsistent with the theory [Jäc04].

As any pure stochastic volatility model, Heston's is unable to calibrate well the short term skewness. Capturing the market skew requires a very small κ, a large vovol and a too strong correlation between volatility and option price (leverage effect), which is supported by no econometric analysis, as observed by Jäckel [Jäc04]. Extensions of Heston with local volatility have been suggested, notably with Constant Elasticity of Variance (CEV)[1] (hereafter CEV) and quadratic volatility [Lip02].

[1]The local volatility form: $dS_t = \sigma_t S^\beta \, dW_t$, is known as Constant Elasticity of Variance and has been introduced by Cox [Cox75] in 1975. It exhibits leverage effect and has a positive probability of attaining 0.

3.3 SABR

The **SABR model**, which stands for Stochastic Alpha-Beta-Rho, was introduced by Hagan et al. in their paper [HKL02] and is another SV model that is largely used, specially in interest rates markets. It considers a log-normal volatility, which has been studied before (in a more general case with mean-reversion and time-dependent coefficients) by Hull-White [HW87], Scott [Sco87] and [Wig87]. Its local volatility has a CEV form:

$$dS_t = \sigma_t S_t^{\beta}\, dW_t, \tag{3.3}$$
$$d\sigma_t = \alpha\sigma_t\, dB_t, \tag{3.4}$$
$$\langle dW_t, dB_t\rangle = \rho\, dt \tag{3.5}$$

Pros

Its implied dynamics is realistic, with the implied volatility moving in the same direction than the underlying (sticky-moneyness). The volatility process was inspired by an analysis of the historical volatility-of-volatility against the volatility level, obtaining a reasonable fit. But apart from some very particular cases ($\beta = 0$ or 1), the model does not have a closed form solution. In fact, the model itself is not often used, one uses instead an asymptotic formula for the implied volatility in the short maturity limit derived by Hagan et al. in the original paper. When one refers to SABR, much probably one is considering this approximation formula.

Cons

We have two kind of problems, those who refer to the model itself and those who refer to the approximation formula. Concerning the exact model, the local volatility form of CEV leads to a price process attaining zero with positive probability. It is an undesirable feature in most financial applications. While the authors have expressed a concern about the finiteness of the moments of the SABR model in the 2004 version of [AP05b][2], they could prove that the model has finite moments. Still, higher order moments may be very large and lead to numerical misbehaviour. The asymptotic derived by Hagan et al. with singular perturbation technique when the parameter α is small does not allow for either term-structure of parameters or mean reversion of volatility, which is often necessary for having a decent calibration. A volatility that is not mean reverting can grow arbitrarily high which is not economically reasonable. A remedy to the latter was the mean-reverting $\lambda-$SABR model, that have an asymptotic formula as well, see [HL09].

3.4 Hyp-Hyp model

In order to avoid some of the problems of the most used models Heston and SABR, Kahl and Jäckel [KJ07] analysed a model where we combine a (parametric-)local volatility and

[2] Although this version is not available on SSRN any more, this concern is cited in [KJ07].

a stochastic volatility component, both having a hyperbolic functional form. Its SDE is:

$$
\begin{aligned}
dS_t &= \sigma_0 S_0 \cdot f\left(\frac{S_t}{S_0}\right) \cdot g(Y_t)\, dW_t \\
dY_t &= -\kappa Y_t\, dt + \alpha\sqrt{2\kappa}\, dB_t \\
\langle dW_t, dB_t \rangle &= \rho\, dt
\end{aligned}
$$

with

$$
\begin{aligned}
f(x) &= \left[(1-\beta+\beta^2)\cdot x + (\beta-1)\left(\sqrt{x^2+\beta^2(1-x)^2}-\beta\right)\right]/\beta \\
g(y) &= y + \sqrt{y^2+1}
\end{aligned}
$$

Jäckel first studied the benefits of a hyperbolic local volatility in [Jäc06]. This form was chosen due to its resemblance to a CEV local volatility, which is also used in the SABR model, but it avoids the problematic asymptotic behaviour of the latter. Hyperbolic local volatility has finite slope in 0 and positive slope as S approaches ∞, which prevents the underlying from attaining 0, a property very suitable numerically.

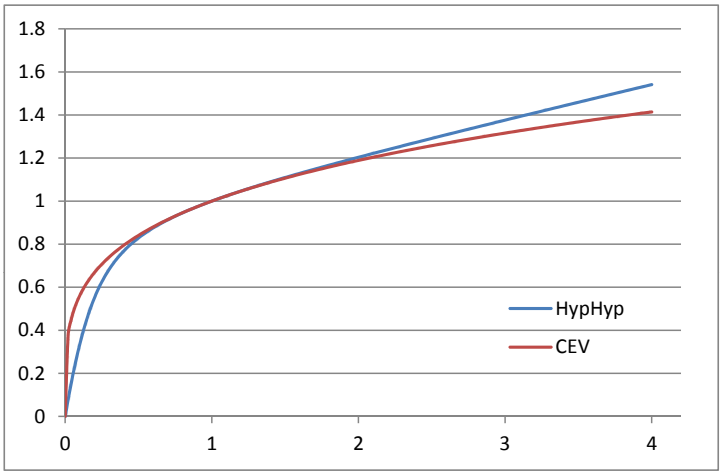

FIGURE 3.2: Hyperbolic local volatility function compared against CEV local volatility function.

The volatility driving function is a conventional Ornstein-Uhlenbeck process, so, a functional form $g(y) = e^y$ would lead to log-normal volatilities, as in SABR, for example. The hyperbolic functional form is close to the log-normal form in the center of the distribution, but have a tail behaviour much more convenient. As it approaches $+\infty$, the tails are thinner than the log-normal model, implying better behaved moments. Indeed, the right tails are asymptotically normal. As it approaches 0, the distribution is asymptotically normal inverse Gaussian, so it converges to 0 and have all of its derivatives converging to zero either. So the function is almost flat, therefore null volatility level cannot be attained.

In the original paper, the authors have derived an asymptotic formula that allows for the computation of approximated implied volatilities generated by the model. With some adjustments developed by the authors in the same article, the formula has a very good fitting for many different regimes of maturities and strikes.

The authors have developed the model based on its fundamental characteristics and not on the easiness of the analytical calculations. We expect that the model will be able

to fit better the market prices than the preceding models, without the computational shortcomings. Still, it is a fully parametric model, so its degrees of freedom may not suffice to calibrate. We may consider adding a local noise, an additional non-parametric local volatility term that would permit a full calibration of the vanilla prices. The Hyp-Hyp model would be used in a first calibration, while the local noise would be a fine tuning, when necessary.

4

Pricing and calibration with Finite Differencing Method

Once we have chosen a model, our problem is to price financial products in this set-up. In addition, we also need to calibrate the model, i.e., find the parameters of the model to match prices of relevant instruments. We recall that by the fundamental theorem of asset pricing [HP81], the present value of the contingent claim that pays $\pi(S_T)$ at T is

$$V_0 = N_0 \cdot \mathbb{E}\left[\frac{\pi(S_T)}{N_T}\right]. \tag{4.1}$$

Most of the time we need to evaluate (4.1) numerically.

There is two generic approaches, Monte Carlo methods or a Partial Differential Equation approach. Monte Carlo methods are specially appropriate to high-dimensional problems. However, PDE approach is more efficient in low-dimensional context and handles efficiently with a wider class of problems, including options with conditional exercise features. As the stochastic volatility models that we have considered are two-dimensional only, we have chosen PDE methods

Different methods to numerically solve a PDE exist, as finite difference (FD) and finite element (FE) methods. FEM handles complicated domains and geometry, specially important in subjects like mechanics and fluid dynamics, however, it requires more complex theoretical and numerical mathematics. In finance, domain geometry is not crucial and FDM suffices. Moreover, FDM allows for methods of dimension splitting which will increase substantially the performance.

4.1 PDE formulation of the pricing problem and its inverse

4.1.1 Backward Kolmogorov

Let us consider a generic diffusion underlying dynamics:

$$dS_t = \mu(t, S_t)\, dt + \sigma(t, S_t)\, dW_t. \tag{4.2}$$

The Feynman-Kać formula links the problems of finding the conditional expectation

$$u(t, S_t) = \mathbb{E}\left[e^{-\int_t^T r(s,S_s)\, ds}\Pi(S_T) \mid S_t\right], \tag{4.3}$$

with the Parabolic Partial Differential Equation:

$$\begin{cases} \partial_t u(t,s) + L \cdot u(t,s) = 0, \\ u(T,s) = \Pi(s) \end{cases} \tag{4.4}$$

where L is an elliptic differential operator:

$$L = \frac{1}{2}\sigma^2(t,s)\frac{\partial^2}{\partial s^2} + \mu(t,s)\frac{\partial}{\partial s} - r(t,s).$$

In this backward approach, we will **fix a terminal condition** and possibly some boundary depending on the specific features of the product. The price of the product is $u(0, S_0)$.

As we have obtained all the prices at the initial time as a function of the initial value of the underlying, we can obtain the price sensibilities with respect to the initial parameters, the *Greeks*[1]. Hence, this approach is more convenient for pricing and hedging. Moreover, this approach naturally handles early exercise features.

It is specially interesting to consider the PDE derived from Black-Scholes SDE (2) in log-coordinates, as we will have:

$$L = \frac{1}{2}\sigma^2\left(\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x}\right).$$

This is a convection-diffusion equation with constant coefficients.

In a more general context of stochastic volatility, we will have additional dimensions in our problem. In this work we will only consider two-factor models, where the additional factor corresponds to a diffusion process leading the volatility dynamics. We assume:

$$L = a_{11}\partial_{11} + a_{12}\partial_{12} + a_{22}\partial_{22} + b_1\partial_1 + b_2\partial_2 + c \tag{4.5}$$

4.1.2 Forward Kolmogorov

We could use the dual PDE computing the forward probability density and calculating prices starting from a **fixed initial condition**. Indeed, denoting backward variables (s, y) and forward variables (s', y'), the probability transition is written:

$$p(t,s,v;T,S,V)\,ds\,dy = P[S_T, \nu_T \in \mathcal{V}_{ds,dv}(S,V) \mid S_t = s, \nu_t = v]$$

i.e., the probability of the diffusion starting in (s, y) to finish in the infinitesimal neighbourhood $\mathcal{V}_{ds,dv}(S,V)$ of (S,V). As a function of the forward variables (T,S,V) with the backward variables fixed, it is known as Green's function and we will note it $G(T,S,V)$. G is a solution of the forward Kolmogorov equation:

$$\begin{cases} \partial_T G(T,S,V) - L^\dagger \cdot G(T,S,V) = 0, \\ G(t,S,V) = \delta(S_t - S)\delta(\nu_t - V), \end{cases} \tag{4.6}$$

where $L^\dagger$ is the adjoint operator of the diffusion generator L, which is formally given by:

$$L^\dagger = \frac{1}{2}\frac{\partial^2}{\partial s^2}(\sigma^2(t,s)) - \frac{\partial}{\partial s}(\mu(t,s)) - r(t,s)$$

[1]However, with FDM, the numerical solution accuracy is strictly related to the grid choice and its accuracy is maximum in the center of the grid, where the boundary effects are smaller. We may calculate Greeks by considering displaced grids to avoid grid effects.

To completely define the adjoint operator we need the boundary conditions. Let us assume that the Green's Function is a rapidly decreasing function (such as the function and all its derivatives go to zero as the space coordinates approaches infinity). If we take the backward equation, multiply by the Green's Function and integrate, we will have:

$$\begin{aligned}
0 &= \int_t^T \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left[\partial_t u(t,s,v) + L \cdot u(t,s,v)\right] G(t,s,v)\, ds\, dv\, d\tau \\
&= \int_t^T \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} u(t,s,v) \left[-\partial_t G(t,s,v) + L^\dagger \cdot G(t,s,v)\right]\, ds\, dv\, d\tau \\
&+ \int_{-\infty}^{+\infty} \left[u(T,s,v)G(T,s,v) - u(t,s,v)G(t,s,v)\right] ds\, dv,
\end{aligned}$$

that implies:

$$u(t; S_t, \nu_t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} u(T,S,V)G(T,S,V)\, dS\, dV. \tag{4.7}$$

To get the price, we integrate (4.7) numerically. Notice that the operator involves derivatives of the coefficients. If they cannot be analytically calculated, they must be interpolated and approximated with finite differences.

4.1.3 Kolmogorov equations as transport problem

We remark that both Kolmogorov equations have the form of a convection-diffusion equation with non-constant diffusion and convection parameters. In its standard form, the convection-diffusion equation is written as:

$$\partial_t u = \operatorname{div}(\mathbf{D}\nabla u) - \operatorname{div}(\mathbf{v}u) + f \tag{4.8}$$

where $\mathbf{D}$ is the diffusion matrix, $\mathbf{v}$ is the velocity field and f is the source term. The convection-diffusion equation is largely studied in the fluid dynamics literature, long before the advent of the financial mathematics.

It can be written as well as:

$$\partial_t u + \operatorname{div}(\mathbf{j}) = f \tag{4.9}$$

where $\mathbf{j} := -\mathbf{D}\nabla u + \mathbf{v}u$ is the flow vector. It is the *continuity equation* and it expresses that the time variation of quantity inside an infinitesimal region is equal to the flow across the boundary of this region (plus any external source f).

This interpretation of the equation as convection and diffusion can give us a better qualitative understanding of a model. The analogy with the fluid mechanics will permit us to remedy some problems that we will encounter.

4.2 Numerical resolution of Kolmogorov equations

4.2.1 One-dimensional Finite Differencing

Let us consider a domain $[0,T] \times \Omega, \Omega = [A,B]$. We pose the following general problem – finding the initial time-slice $u(0,x)$ of a function that follows the dynamics determined by the partial differential equation:

$$\partial_t u(t,x) + \left[\mu(t,x)\partial_x + \frac{1}{2}\sigma^2(t,x)\partial_{xx} - r(t,x)\right] u(t,x) = 0, \quad (t,x) \in \Omega$$

with boundary conditions:

$$\begin{cases} u(T,x) = g(x), & \forall x \in [A,B] \\ \dfrac{\partial^2 u}{\partial x^2}(t,A) = 0, & \forall t \in [0,T] \\ \dfrac{\partial^2 u}{\partial x^2}(t,B) = 0, & \forall t \in [0,T] \end{cases}$$

A grid for the domain Ω is a simply a finite set of points $\Omega_h = \{x_i : i = 0, \ldots, m\}$ with $x_0 = A$ and $x_m = B$. We replace all the considered functions in our PDE by grid functions, their values on the grid points,

$$u_i = u(x_i), \qquad \mu_i = \mu(x_i), \qquad \sigma_i = \sigma(x_i), \qquad r_i = r(x_i).$$

Let us assume that the grid is uniform. By expanding the function u in Taylor series around the point x_i, we obtain *central finite difference schemes*:

$$\delta \cdot u_i := \frac{u_{i+1} - u_{i-1}}{2\Delta x} = \frac{\partial u}{\partial x}(x_i,t) + O(\Delta x^2)$$

$$\delta^2 \cdot u_i := \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} = \frac{\partial^2 u}{\partial x^2}(x_i,t) + O(\Delta x^2).$$

Therefore, we can replace the derivatives in the equation with finite difference schemes still being accurate to the second-order in Δx^2.

We have considered symmetrical schemes, not applicable on the boundaries, where we consider rather one-sided schemes. The relations:

$$\delta^+ := \frac{u_{i+1} - u_i}{\Delta x} = \frac{\partial u}{\partial x}(x_i,t) + O(\Delta x)$$

$$\delta^- := \frac{u_i - u_{i-1}}{\Delta x} = \frac{\partial u}{\partial x}(x_i,t) + O(\Delta x)$$

provides us with left and right sided schemes for the first derivative. However, we see that they are less accurate then the central scheme. By a procedure detailed further, we can obtain other schemes, more accurate and/or applicable to non-uniform meshes.

The *semi-discretized problem* is obtained by replacing the operator L by its discretization:

$$\partial_t \bar{u} + (\bar{L} \cdot \bar{u} + \bar{f}) = 0$$

where f is necessary to possibly include boundary conditions. Here the boundary conditions were linear functions of u and could be inserted directly into $\bar{L}$ (by eliminating the second derivative and approximating the first derivative by one-sided schemes). Then, $f \equiv 0$ and

$$\begin{cases} \bar{L} \cdot u_i = \mu_i(\delta_i \cdot u_i) + \dfrac{1}{2}\sigma_i^2(\delta^2 \cdot u_i) - r_i u_i \\ \bar{L} \cdot u_0 = \mu_0(\delta^+ \cdot u_0) - r_0 u_0 \\ \bar{L} \cdot u_m = \mu_m(\delta^- \cdot u_m) - r_m u_m \end{cases}$$

$\bar{L}$ becomes a linear application in a finite-dimensional vector space. It can be also expressed as a matrix:

$$\bar{L} = \begin{pmatrix} c_0 & u_0 & 0 & 0 & \dots & 0 \\ l_1 & c_1 & u_1 & 0 & \dots & 0 \\ 0 & l_2 & c_2 & u_2 & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & l_{M-1} & c_{M-1} & u_{M-1} \\ 0 & 0 & 0 & 0 & l_M & c_M \end{pmatrix}.$$

4.2.2 Time discretization schemes

With the space discretization, L becomes a finite-dimensional linear operator. Our problem becomes solving a system of ordinary differential equations:

$$\frac{du}{dt} + \bar{L} \cdot u = 0. \tag{4.10}$$

The solution of (4.10) is, formally[2]:

$$\bar{u}(t - \Delta t) = e^{\int_{t-\delta t}^{t} \bar{L}_s \, ds} \cdot \bar{u}(t),$$

which would require the full diagonalization of the operator $\bar{L}$. One will rather try to solve the system approximately.

For integrating these equations, we have different schemes. A scheme consists in an approximation of the propagator $A(t, \delta t, \bar{L}) := \exp(\int_t^T \bar{L}_s \, ds)$. Some notable schemes are the forward Euler, the backward Euler and the trapezoid rule. All the definitions about stability and convergence of a scheme can be found in Appendix D.

Forward Euler (or explicit) scheme is first order accurate in time. It is written:

$$\bar{u}(t - \Delta t) = [1 + \Delta t \cdot \bar{L}(t)] \cdot \bar{u}(t).$$

It is not unconditionally stable though, so the time step need to follow a too restrictive CFL condition in order to ensure convergence. For that reason, **we will not discuss fully-explicit schemes further**.

Using the backward Euler (or implicit) scheme, we will have:

$$[1 - \Delta t \cdot \bar{L}(t)] \cdot \bar{u}(t - \Delta t) = \bar{u}(t)$$

It is a unconditionally stable scheme, quiet for all modes, first order accurate. As the propagator appears on the left-hand side, we need to solve a linear system to obtain $\bar{u}(t - \Delta t)$.

The trapezoid rule (or Crank-Nicolson scheme) is given by:

$$\left[1 - \frac{1}{2}\Delta t \cdot \bar{L}(t)\right] \cdot \bar{u}(t - \Delta t) = \left[1 + \frac{1}{2}\Delta t \cdot \bar{L}(t)\right] \cdot \bar{u}(t),$$

[2] The semi-discretization of the forward operator is $\partial_\tau G(\tau) - \bar{L}^\dagger \cdot G(\tau) = 0$, with formal solution

$$\bar{G}(\tau + \Delta\tau) = e^{\int_\tau^{\tau+\Delta\tau} \bar{L}_s^\dagger \, ds} \cdot \bar{G}(\tau).$$

and the same integration schemes can be used to obtain the forward time-slice $\bar{G}(\tau + \Delta\tau)$ from $\bar{G}(\tau)$.

It is a unconditionally stable scheme, second order accurate. However, it is ringing, which may lead to oscillations. Its eigenvalues tend to -1 to the very high-frequency modes, so, these modes takes virtually an infinite amount of time to be damped. These oscillations will be perceived as we propagate discontinuous functions, which high-frequency modes are populated.

Other integration schemes

The preceding schemes can be seen under the perspective of the functional analysis as Padé approximants of the exponential functional. We call a Padé approximant of order (m, n) the "best" approximation of a function f by a rational function of the form

$$R_{m,n}(x)[f] = \frac{a_0 + a_1 x + \ldots + a_m \theta^m}{1 + b_1 x + \ldots + b_n \theta^n},$$

in the sense that f and its Padé approximant have the same derivatives at the origin, up to the (m+n)-th order. The coefficients of $R_{m,n}$ are explicitly know for the exponential function. Thus, Padé approximants (0, 1), (1, 0) and (1, 1) are respectively the Implicit, the Explicit and Crank-Nicolson schemes. The (0, 2) approximant, or second order fully implicit scheme is second-order accurate, unconditionally stable and quiet:

$$\left[1 - \Delta t \cdot \bar{L}(t) + \frac{1}{2}\Delta t^2 \cdot \bar{L}(t)^2\right] \cdot \bar{u}(t - \Delta t) = \bar{u}(t).$$

The fully-implicit, fully-explicit and Crank-Nicolson schemes are, in a way, fundamental blocks, and we have great flexibility to build other schemes, mainly due to the exponential functional form of the propagator. It satisfies the property valid to **commuting**[3] operators:

$$\exp(t(A + B)) = \exp(tA) \cdot \exp(tB), \text{ if } AB = BA.$$

We can discretize different terms of the operator separately. The so-called IMEX schemes for example, use differents schemes for the convection and for diffusion. In the multi-dimensional context, the same reasoning makes possible the use of *splitting schemes*, that consider the terms in different directions separately, which greatly simplifies the computations.

4.2.3 Solving sparse linear systems

We have already seen that except in the case of fully explicit schemes, we are required to solve a linear system. A direct elimination technique like Gauss Elimination is computationally expensive, as its memory complexity is $O(N^2)$ and its arithmetical complexity is $O(N^3)$.

Fortunately, as L is a differential operator, its approximations in a given point will involve only the point and its neighbours. Hence the matrix form of the operator will have mostly null entries. We say that the operator matrix (and consequently, the propagator matrix) is sparse. We need to use solvers that take advantage of this sparsity in order to obtain faster computations.

[3]Even though the operator do not commute, the equality is still true up to the order t^2 by the Baker-Campbell-Hausdorff formula.

Direct solvers solve the problem in a finite number of steps, that will depend on the sparsity structure of the matrix and in its size. It is only useful in cases where the matrix satisfies some very strict conditions, however in these cases, it offers the best performance. Iterative solvers may be useful for more general operators, not requiring a structured matrix form[4]. It will successively approximate the solution until it satisfies a tolerance criterion.

Direct solvers

The sparsity structure that arises when discretizing operators is very special. Non-null elements populate bands at and around the diagonal, the number of bands, or *bandwidth*, being determined by the stencil of the finite differences we used. Slightly modifying Gaussian elimination to take advantage of this structure, called banded matrix, can radically improve its performance.

Consider the discretization of the operator with central finite differences in the interior and BC1 condition in the borders, which constitutes a three-stencil discretization for all the points. It implies that the resulting operator matrix will be tri-diagonal Hence, the first-order propagator $A_{(0,1)}(t, \theta, \bar{L})$ will be represented as a tri-diagonal matrix, and the second-order implicit propagator $A_{(0,2)}(t, \theta, \bar{L})$ will be a penta-diagonal matrix.

A Gaussian Elimination adapted to a banded linear system will provide a solution with $O(nm^2)$ operations where m is the matrix bandwidth. In the usual cases, tri- or penta-diagonal, the square bandwidth is still $O(1)$ and we have a final arithmetic complexity of $O(n)$ instead of $O(n^3)$, which is a huge improvement. For the tri-diagonal matrices, the adaptation of the Gaussian Elimination is widely known as Thomas algorithm.

Iterative solvers

These methods will improve an initial guess by considering a series of successive iterations:

$$\mathbf{x}^{(k+1)} = T \cdot \mathbf{x}^{(k)},$$

with T a contracting operator. We refers the reader to the excellent book of Saad [Saa03]. A simple example is the Jacobi method. We write the matrix $A = D + R$, where $D = \mathrm{diag}(A)$. Then, a series of approximations of x is given by:

$$\mathbf{x}^{(k+1)} = D^{-1}\left[\mathbf{b} - R\mathbf{x}^{(k)}\right]$$

and the inverse of D is trivial as it is diagonal. A necessary condition for the convergence is that the spectral norm $\rho(D^{-1}R) < 1$, because the problem becomes finding the fixed point of a continuous contracting map. The convergence is guaranteed for diagonally dominant matrices.

More complex methods include the Conjugate Gradient, that requires a self-adjoint matrix, or its modification for general matrices Bi-conjugate Gradient. Both of these methods are not numerically stable, then, they can magnify small deviations instead of damp them. Hence, we have used the BiCGStab(ℓ) [dV92, SF93] method, with two

[4]Strictly, one do not even need to express the operator as a matrix to use iterative methods. It can be useful to implement directly the action of the operator on a vector, as efficiently implementing sparse matrices is a complex issue.

improvements suggested by G.Sleijpen and H.van der Vorst. In this method, every l steps of the Bi-conjugate Gradient method is followed by a GMRES(ℓ) step (see restarted GMRES in [Saa03]), in order to smooth the convergence of the algorithm. We used an adaptation of the algorithm [M.A98] for the version ($\ell = 1, 2$) by Botchev and Fokkema.

4.2.4 Multi-dimensional Finite Differencing

In the multi-dimensional framework, the approach is same. For discretizing the spatial coordinates, we will still use the same procedures. We will consider the coordinates $(t, x_1, x_2, \ldots, x_N)$ and the spatial discretization grid $(0, \ldots, m_1) \times \ldots \times (0, \ldots, m_N)$. What will change is that, a fixed time-slice is not a vector of $\mathbf{R}^{m_1}$ any more, but an element of $\mathbf{R}^{m_1} \otimes \ldots \otimes \mathbf{R}^{m_N}$. If we want to recover a matrix representation of the propagator, we must write the tensor $u_{i_1,\ldots,i_N}$ as a vector by relabelling the grid points ("flatten the tensor"). Let us consider the case $N = 2$.

In this case, $u_{i,j} = u(x_i, y_j)$ and $(u_{i,j})_{\substack{i=0,\ldots,M \\ j=0,\ldots,N}}$ which would be naturally represented as a matrix. Provided that we re-index this matrix, what is normally done by rows

$$(i, j) \mapsto i \times M + j,$$

or by columns

$$(i, j) \mapsto j \times N + i,$$

the operator will admit a representation as a matrix of dimensions $MN \times MN$. Naturally, each different remapping will entail a distinct representation of the operator.

For example, let us consider the matrix form of the central finite difference ∂_{xx} operator. It is a local operator, uni-directional. When we write the (i, j) component of $\delta_{xx}v$, the only components of v that intervenes are $(i-1, j)$, (i, j) and $(i+1, j)$. If we flatten it along the x-direction (*i.e.*, by columns), the x-neighbour points of $u(i, j)$ will still be neighbouring points in the flattened vector. However, when the flattening is done by the y-direction (by rows), the x-neighbour elements (i + 1, j) and (i - 1, j) will not be neighbours of (i, j). Instead, the neighbours elements will be remapped to $(i-1)N + j$, $iN + j$ and $(i+1)N + j$. The resulting matrix will be have only three non-empty bands, but with N empty bands in-between.

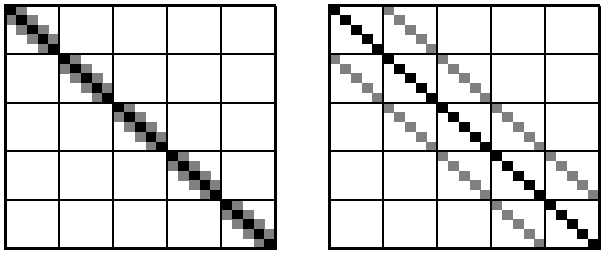

FIGURE 4.1: Sparsity structure of the tensor $(\partial_{xx})_{ii'jj'}$ flattened in x-direction (left) or y-direction (right). Black dots are entries $-2/\Delta_x^2$, grey dots are entries $1/\Delta_x^2$.

The first matrix is 3-banded, whilst the second is 3N-banded. The resulting system for the former can be solved with arithmetical complexity $O(9MN)$ by Gaussian elimination. But the same technique applied to the latter 3N-banded matrix have a complexity of $O(9MN^3)$, even if only three of the bands are populated, for the algorithm will fill the empty bands in its intermediate steps. We conclude that relabelling the indices in a

proper way can dramatically reduce the arithmetical complexity of our problem. When solving a local operator that affects only the one direction, perform flattening in this direction can drastically reduce the complexity of solving the linear system.

4.2.5 Splitting schemes

Suppose that our equation has no cross-derivative term. We split the operator in two uni-directional operators:

$$L = L_1 + L_2$$

with

$$L_i = a_{ii}\partial_{ii} + b_i\partial_i + c_i.$$

It is not really important how we split the c term between L_1 and L_2. We have chosen to always split half the term in each operator. Writing

$$\bar{u}(t - \Delta t) = e^{\int_t^T \bar{L}_{1,s}\, ds} e^{\int_t^T \bar{L}_{2,s}\, ds} \cdot \bar{u}(t),$$

we see that we can discretize the operator differently in each dimension. We can do, for example, a two-step scheme composed by two implicit steps alternating the implicit direction:

Peaceman-Rachford
Accuracy: $O(\Delta t^2)$

$$\begin{cases} \left(1 - \frac{\Delta t}{2} \cdot \bar{L}_1\right) \cdot v = \left(1 + \frac{\Delta t}{2} \cdot \bar{L}_2\right) \cdot \bar{u}(t), \\ \left(1 - \frac{\Delta t}{2} \cdot \bar{L}_2\right) \cdot \bar{u}(t - \Delta t) = \left(1 + \frac{\Delta t}{2} \cdot \bar{L}_1\right) \cdot v. \end{cases} \tag{4.11}$$

A discretization method like this is known as Alternating Direction Implicit method. Its advantage is that, if we have originally used a two-dimensional compact stencil (9-stencil) for discretization, L_1 will be a tri-diagonal operator when column flattening is done by columns, and L_2 will be a tri-diagonal operator when u is flattened by rows. Writing it explicitly:

1. We compute the right-hand-side of (4.11). It is an explicit computation that does not present any special difficulties.

2. We flatten the result by columns. In this new indexing, L_1 will be a tri-diagonal matrix. The first equation in (4.11) is a tri-diagonal linear system, and we solve it by Thomas algorithm.

3. We rebuild a matrix v de-flattening the vector v by columns.

4. We compute the right-hand-side of the second equation in (4.11). It is an explicit computation that does not present any special difficulties.

5. We flatten the resulting matrix by rows. In this indexing, L_2 will be a tri-diagonal matrix, and the second equation in (4.11) can be easily solved.

6. We rebuild a matrix $u(t - \Delta t)$ by de-flattening the vector $u(t - \Delta t)$.

There are many extensions to the basic ADI scheme in order to include the cross-derivative term and have better accuracy. A first approach would be introduce the mixed derivative term explicitly. However, introducing explicit discretizations will reduce the temporal order of convergence of the scheme. Besides, its stability will be conditioned by a CFL-like condition, requiring the time steps to be very small. A better way to deal with the cross-term would be using a predictor-corrector scheme, like **Craig-Sneyd**:

Craig-Sneyd
Accuracy: $O(\Delta t^2)$ **iff** $\theta = \mathbf{1/2}$

$$\begin{cases} \textit{Predictor} \\ Y_0 = (1 + \Delta t \cdot \bar{L}) \cdot \bar{u}(t) \\ \left(1 - \theta\Delta t \cdot \bar{L}_j\right) \cdot Y_j = Y_{j-1} - \theta\Delta t \cdot \bar{L}_j \cdot \bar{u}(t), \quad (j = 1, 2) \\ \textit{Corrector} \\ \tilde{Y}_0 = Y_0 + \dfrac{1}{2}\Delta t \cdot \bar{L}_{12} \cdot (Y_2 - \bar{u}(t)) \\ \left(1 - \theta\Delta t \cdot \bar{L}_j\right) \cdot \tilde{Y}_j = \tilde{Y}_{j-1} - \theta\Delta t \cdot \bar{L}_j \cdot \bar{u}(t), \quad (j = 1, 2) \\ \textit{End} \\ \bar{u}(t - \Delta t) = \tilde{Y}_2 \end{cases} \tag{4.12}$$

Much like the Crank-Nicolson scheme, Craig-Sneyd may present some instabilities. Moreover, the convergence in second order is only attained for $\theta = 0.5$. Some attempts have been done in order to find better schemes. We have considered Modified Craig-Sneyd and Hundsdorfer-Verwer, advocated by Foulon and in 't Hout and leading to good and stable results to Heston PDE [itHF10]. They are both consistent of order two for any value of θ and are written as:

Modified Craig-Sneyd – MCS
Accuracy: $O(\Delta t^2)$

$$\begin{cases} \textit{Predictor} \\ Y_0 = (1 + \Delta t \cdot \bar{L}) \cdot \bar{u}(t) \\ \left(1 - \theta\Delta t \cdot \bar{L}_j\right) \cdot Y_j = Y_{j-1} - \theta\Delta t \cdot \bar{L}_j \cdot \bar{u}(t), \quad (j = 1, 2) \\ \textit{Corrector} \\ \hat{Y}_0 = Y_0 + \theta\Delta t \cdot \bar{L}_{12} \cdot (Y_2 - \bar{u}(t)) \\ \tilde{Y}_0 = \hat{Y}_0 + \left(\dfrac{1}{2} - \theta\right) \Delta t \cdot \bar{L} \cdot (Y_2 - \bar{u}(t)) \\ \left(1 - \theta\Delta t \cdot \bar{L}_j\right) \cdot \tilde{Y}_j = \tilde{Y}_{j-1} - \theta\Delta t \cdot \bar{L}_j \cdot \bar{u}(t), \quad (j = 1, 2) \\ \textit{End} \\ \bar{u}(t - \Delta t) = \tilde{Y}_2 \end{cases} \tag{4.13}$$

and

Hundsdorfer-Verwer – HV
Accuracy: $O(\Delta t^2)$

$$\begin{cases} \textit{Predictor} \\ Y_0 = (1 + \Delta t \cdot \bar{L}) \cdot \bar{u}(t) \\ \left(1 - \theta \Delta t \cdot \bar{L}_j\right) \cdot Y_j = Y_{j-1} - \theta \Delta t \cdot \bar{L}_j \cdot \bar{u}(t), \quad (j = 1, 2) \\ \textit{Corrector} \\ \tilde{Y}_0 = Y_0 + \frac{1}{2}\Delta t \cdot \bar{L} \cdot (Y_2 - \bar{u}(t)) \\ \left(1 - \theta \Delta t \cdot \bar{L}_j\right) \cdot \tilde{Y}_j = \tilde{Y}_{j-1} - \theta \Delta t \cdot \bar{L}_j \cdot Y_2, \quad (j = 1, 2) \\ \textit{End} \\ \bar{u}(t - \Delta t) = \tilde{Y}_2. \end{cases} \tag{4.14}$$

4.2.6 Details and improvements

Non uniform grids

Choosing an appropriate grid can have a positive impact in our approximation. We may want to have grid more dense around critical points: a strike, a barrier, etc. to improve the accuracy of the solution. Moreover, to avoid boundary effects and to have reasonable boundary conditions (which are often based on asymptotics) we want to have a large domain. But having a fine grid all along a great domain is too costly. Non-uniform grid ally reasonable boundary conditions and a good central resolution.

We have some different approaches, namely, grid refinement, where we add points to an uniform mesh, or grid displacement, where we remap the points of the grid concentrating them in suitable locations. We remark that when we write non-uniform finite differences, we have a term $O(\Delta x^+ - \Delta x^-)$, hence, if the difference between the time-steps is large, this term will be relevant and we will lose accuracy. We will consider then smooth grids. We define a grid generating function $g : [0, 1] \mapsto [0, 1]$, bijective, differentiable, strictly increasing. The uniform mesh $x_i = i/n$ will be then remapped to $y_i = y_{\min} + g(x_i)(y_{\max} - y_{\min})$. Considering that:

$$y_{i+1} - y_i \approx cg'(x_i)\delta x = cg'(g^{-1}(y_i))\Delta x,$$

we define a distance ratio function $r(y) = g'(g^{-1}(y))$.

Using $r(y) := \sqrt{c^2 + p^2(y - y^*)^2}$ as considered by Kluge and Tavella Randall leads to a hyperbolic sine grid generating function that concentrate the points around y^* with a concentration c (Fig. 4.2):

$$g(x) = y^* + \frac{c}{p} \sinh\left(px + \operatorname{arcsinh}\left(-\frac{p}{c}y^*\right)\right). \tag{4.15}$$

The concentration means that, around the anchor, the step size will be approximately $c(y_{\max} - y_{\min})/n$. This same function generates another type of mesh if $c = 0$ and $y^* = 0$,

$$g(x) = pe^x,$$

hence transforming the grid in this manner is equivalent to transform the equation to log-coordinates.

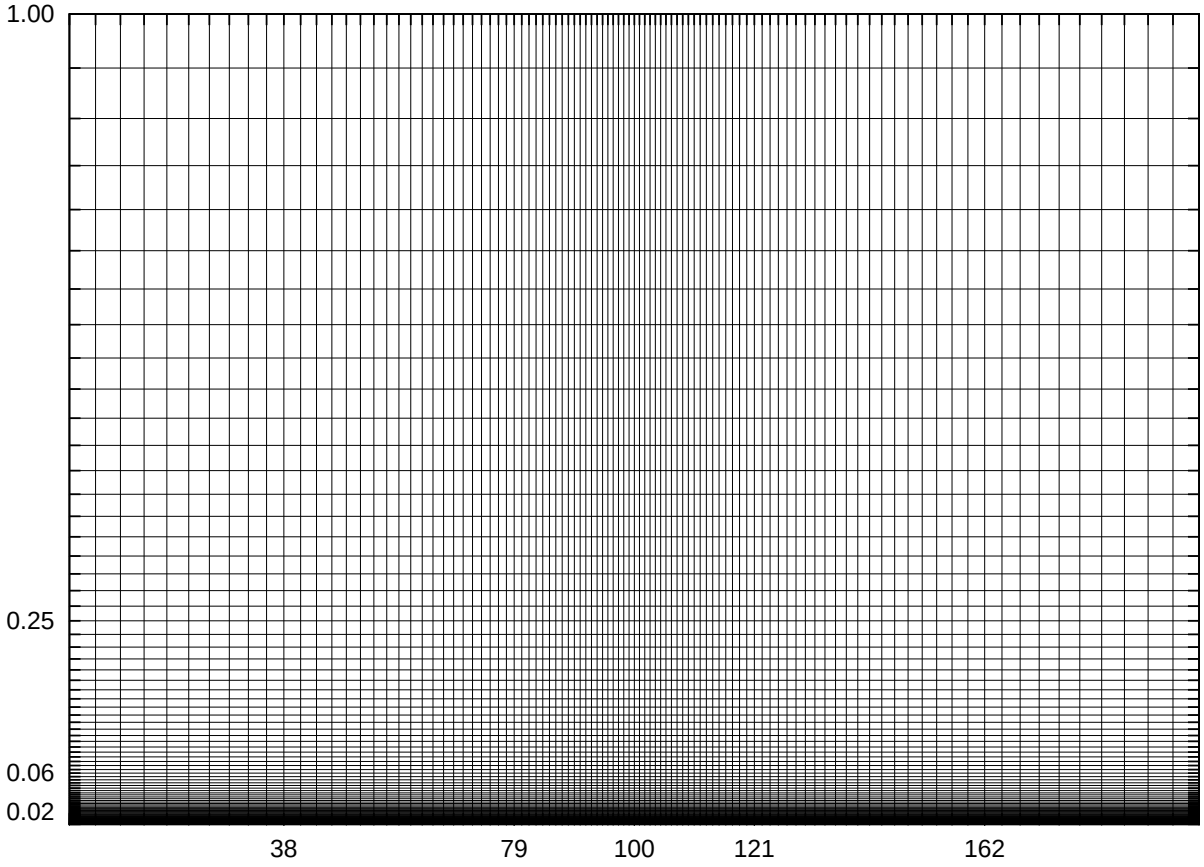

FIGURE 4.2: 2-D grid obtained used the distance ratio function $r(y) := \sqrt{c^2 + p^2(y - y^*)^2}$.

Boundary conditions discretization

Depending on the problem, we can be lead to treat the boundaries differently. If the problem does not have natural boundaries, we shall found the solution in the artificial boundaries as part of the solution of the PDE. But we have no access to the points beyond the boundaries, necessary to central FD, so it must be done with one-sided difference schemes. Some boundary conditions are homogeneous in u, *i.e.*, a linear functional of u. Then, when discretized, they can be represented as a matrix and be inserted into the operator $\bar{L}$. Some other conditions cannot be inserted into the operator matrix, that is why we have as well the f term that can include this class of conditions.

Homogeneous boundary conditions

BC1 condition. A first approach is to assume that, in the boundaries, the function is linear. We call this condition **BC1**, and it is the boundary condition used in the first example of the chapter. That is reasonable for many options, the simplest being a Call. If we are deep in-the-money, the Call price approximates its intrinsic value $S - K$ which is linear in S. In the boundary, we need to solve:

$$\partial_t + \mu \partial_x u - ru = 0 \tag{4.16}$$

And that can be inserted into the operator $\bar{L}$ by using one-sided finite differences for the discretization of the derivative. The interest of this condition is that the tri-diagonal structure of $\bar{L}$ is preserved.

When we consider log-coordinates for the spot, the boundary condition is transformed likely-wise:

$$\frac{\partial^2 u}{\partial s^2}(t,s) = \frac{1}{e^{2x}}\left(\frac{\partial^2 u}{\partial x^2}(t,x) - \frac{\partial u}{\partial x}(t,x)\right) \tag{4.17}$$

BC2 condition. Discretizing both the first and second derivatives with one-sided FD schemes will be called **BC2 condition**. It may be a good approach when we are considering an artificial boundary. The operator matrix will have an extra element in its boundary rows, due to the larger stencil used for the second derivative. It cannot be solved with Thomas algorithm any more, and using a penta-diagonal Gauss elimination would be too costly. It would be better to recur to iterative methods. A simple one that performs nicely is a modification of Jacobi method where we write the matrix A as $T + R$, where T is the tri-diagonal elements of A and then the successive iterations become:

$$T\mathbf{x}^{(k+1)} = \mathbf{b} - R\mathbf{x}^{(k)}.$$

General operator. Naturally, any boundary condition that is expressed as differential operator *without source term* applied to u can be discretized and inserted to the operator $\bar{L}$.

In this class, we have particularly the no-flux condition. For the convection-diffusion equation (4.8), it is written as

$$\mathbf{j} := -\mathbf{D}\nabla u + \mathbf{v}u = 0,$$

and represents that we have no flux of mass across this boundary. It can be useful when solving the Forward Kolmogorov equation to ensure that there is no flow of probability mass out of the domain.

Non-homogeneous boundary conditions Let Γ_1 be a part of the boundary, $\mathbf{n}(x)$ is normal vector pointing outside the region in the point $x \in \Gamma_1$. A conditions over the function itself is called Dirichlet condition:

$$u(t,x) = f_1(x), \qquad x \in \Gamma_1 \tag{4.18}$$

If the condition is over the derivative, it is a Neumann condition:

$$\mathbf{n}(x) \cdot \nabla u(t,x) = f_1(x), \qquad x \in \Gamma_1. \tag{4.19}$$

We will limit our presentation for these two boundary types. Dirichlet boundaries appears naturally in some options, like barriers. A knock-out barrier option at H, for example, is written:

$$\begin{cases} u(t,0) = 0, \\ u(t,H) = 0. \end{cases}$$

Most problem in Finance are considered in semi-infinite region and we need to truncate the domain to apply the FDM. We introduce artificial boundaries, one possibility being to truncate the domain far from the region of interest and impose *a priori* conditions based in our knowledge about the asymptotic behaviour of the problem. For example, for a Black-Scholes Call, as S approaches 0 or $+\infty$, the Call price approaches

its intrinsic value, that will be 0 or $S-K$ respectively. Then, if we consider a $S_{\min}$ small enough and a $S_{\max}$ big enough, the following conditions will only induce a small error:

$$\begin{cases} u(t, S_{\min}) = 0, \\ u(t, S_{\max}) = S - K. \end{cases}$$

Neumann conditions will appear less frequently. An important type of Boundary condition that is represented by a Neumann condition is the reflecting boundary:

$$\vec{n}(x) \cdot \nabla u(t, x) = 0, x \in \Gamma_1. \tag{4.20}$$

Non-homogeneous conditions can be discretized as an additive f term to the discrete operator. We have chosen another approach for the implementation. We **imposed** the conditions to the propagator, instead. After applying the propagator to u_{t_i}, we will have $u_{t_{i-1}}$. So we impose the Dirichlet condition by replacing the boundary elements in $u_{t_{i-1}}$. The same can be done with Neumann conditions, but we use the derivative at the boundary and the inner neighbour points to extrapolate the value of the boundary element.

Convection-dominated transports and up-winding

Let us consider a simple convection equation:

$$\partial_t u - v \cdot \partial_x u = 0. \tag{4.21}$$

which admits the solution

$$u = f(x - vt). \tag{4.22}$$

As $v > 0$, we note that the velocity field of this operator is directed to the right. It means that the velocity field transport a quantity from point i to the point $i + 1$.

We consider the explicit scheme. Using a left-sided or right-sided scheme for the spatial derivative may lead to very different results. The right-sided scheme $v \cdot \partial_x u = v\frac{u_{i+1}-u_i}{\Delta x}$. leads to a unstable scheme. The same way, the central derivative, that can be seen as $0.5(\delta^+ + \delta^-)$, will lead to the same instability problem. The idea of the up-winding comes from the transport interpretation of the equation. As the flux goes toward the right, the scheme should look for the information that is coming from the left. The left-sided scheme translates this idea

$$v \cdot \partial_x u = v\frac{u_i - u_{i-1}}{\Delta x},$$

and the resulting scheme is stable in L^2 and L^∞ under a CFL condition.

The most common approach in financial literature to FDM is to use Crank-Nicolson scheme with central finite differences, because of this order 2 in both space and time. But for any scheme that contains an explicit part, if the convection term is large with respect to the diffusion term, *i.e.*, the equation is **convection-dominated**, spurious oscillations may appear due to the central spatial differencing. A better choice is to discretize the convection term with an up-winding scheme:

$$\begin{cases} v \cdot \partial_x u = v \cdot \delta^- u, \text{ if } v > 0 \\ v \cdot \partial_x u = v \cdot \delta^+ u, \text{ if } v < 0. \end{cases} \tag{4.23}$$

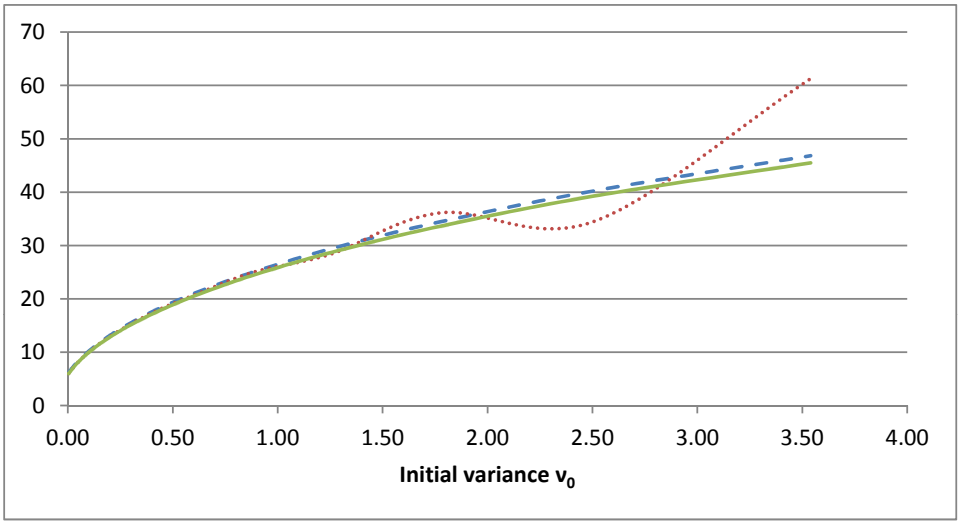

FIGURE 4.3: Ripples in Heston PDE in the high-initial variance region. Analytical value (dashed), FD without up-winding (dotted), FD with up-winding (solid). $\kappa = 2$, $\theta = 0.04$, $\eta = 0.01$.

An up-winding scheme may be required for example in the Black-Scholes discretization if the interest rates are too large or the volatility too small. In a mean-reverting volatility model, like Heston or Hyp-Hyp, if the vo-vol is small, the mean-reverting coefficient can grow larger than the diffusion term, and up-winding may be necessary as well.

Initial conditions smoothing for partial explicit schemes

The temporal accuracy of 2 of the Crank-Nicolson scheme is attained in L^2. For non-smooth initial conditions, which are recurrent in finance (e.g. call options), this rate of convergence may not be achieved. We may observe some instability in the discontinuity region: as we commented before, the damping factor for Crank-Nicolson scheme tends to -1 for the high frequencies:

$$A(\omega) = \frac{1 + \frac{1}{2}\omega}{1 - \frac{1}{2}\omega} \xrightarrow{|\omega| \to +\infty} -1. \tag{4.24}$$

The Shannon sampling theory states that when sampling a signal with sampling frequency $1/\Delta x$ we can only reconstruct it up to the frequency $1/2\Delta x$. If the time step is too coarse relatively to the spatial step, we are guaranteed to have numerical errors in the higher frequencies, that will take very long time to be dissipated, and we will notice small ripples around the regions of non-smoothness. The scheme is numerical stable, but this dissipation time may be so long that it will not be numerically accurate. Moreover, even if we have a solution stable enough, these ripples will have a dramatic effect on the sensibilities Δ and Γ. Even if our analysis is restricted to the Crank-Nicolson case, we shall expect having this kind of spurious oscillations with every scheme that are not full implicit, as the numerator of the damping factor will not be 1. A remedy to this problem is to smooth the initial conditions, analytically or numerically.

Averaging the initial conditions. A possible solution would be simply apply a method known as continuity correction, where we substitute the value of the pay-off in

the grid point x_i by the average of the function over the interval $x_{i-1/2}$ and $x_{i+1/2}$, the grid mid-points. So, if the pay-off is given by a function g(x), the initial condition of the PDE becomes:

$$u(T, x_i) = \frac{1}{\frac{1}{2}(\Delta x_i^- + \Delta x_i^+)} \int_{x_i - \frac{1}{2}\Delta x_i^-}^{x_i + \frac{1}{2}\Delta x_i^+} g(x)\, dx. \tag{4.25}$$

We note that this averaging process does not affect a linear function. A piece-wise linear function will only be changed in this discontinuities points.

Critical grid points positioning. Tavella and Randall suggests repositioning the grid points so that the mid-points will match non-smooth points of the function [TR00]. For a piece-wise linear function, this will have exactly the same effect than the preceding method.

Projection in a set of basis functions. Rannacher [Ran84] suggests L^2-projecting the function in the set of the linear Lagrange basis functions (the triangle hat functions). This also smooths the pay-off. If the function is **continuous** piece-wise linear, as the Call pay-off, and its points of discontinuous slope coincides with the grid points, it already in the linear span of the basis, so it is not affected by the projection. A discontinuous pay-off like a digital option will always be affected.

Rannacher stepping. Rannacher [Ran84] proposes a numerical smoothing of the initial conditions by substituting two fully-implicit half-time-steps for the first Crank-Nicolson time step. The implicit scheme does not present the same issue of the Crank-Nicolson dealing with non-smooth functions. This is a result coming from the finite element method, that allows for non-smooth initial conditions. FDM can be seen as a particular case of the FEM, so the result is very useful here. But in FEM, the initial condition must be projected in the space of the basis functions. The full modification of Rannacher is:

1. Project L^2 the terminal pay-off onto the space of Lagrange basis functions
2. Perform two fully implicit half time-steps, followed by normal Crank-Nicolson steps.

Pooley et al. [PVF03] shows that proceeding with Rannacher time-stepping without smoothing the initials conditions will typically not suffice to restore the order of convergence $O(\Delta t^2)$ of Crank-Nicolson.

Giles and Carter [GC06] have studied different modifications of Rannacher time-stepping and have concluded that the optimum time-stepping is preceding the Crank-Nicolson stepping with 4 fully implicit half time-steps, which will improve the damping both in high and low-frequencies. Their analysis considers Dirac Delta functions, that are discretized as hat functions and hence do not need to be L^2-projected. However, they do not seem to have carried out any smoothing of initial conditions when considering digital Call options, which needs to be projected as we see in Pooley et al., but they allegedly attain the desired convergence rate.

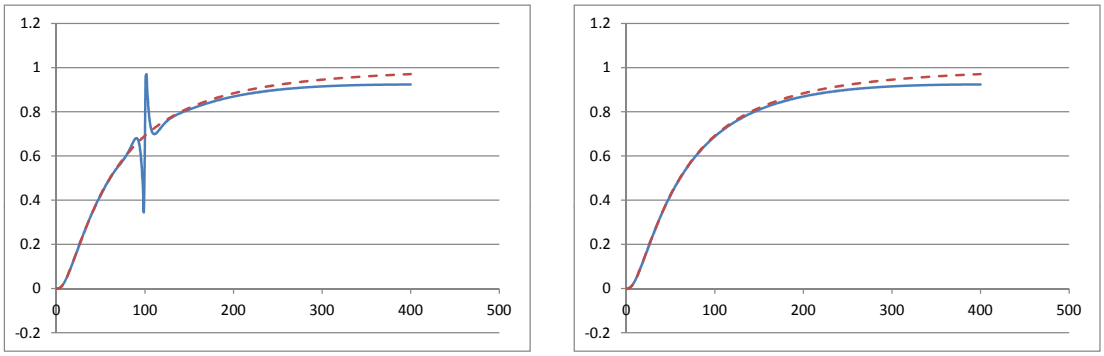

FIGURE 4.4: Black-Scholes Call Delta obtained by Crank-Nicolson scheme (solid) and the analytical value (dashed), without Rannacher stepping (left) and with Rannacher stepping (right).

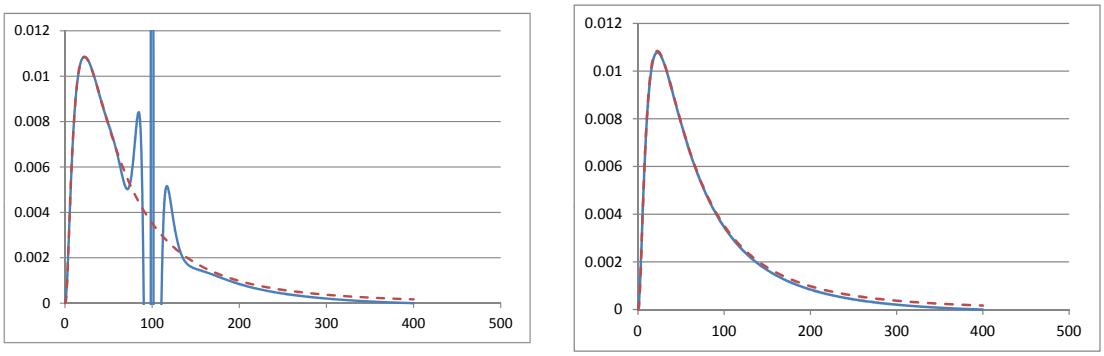

FIGURE 4.5: Black-Scholes Call Gamma obtained by Crank-Nicolson scheme (solid) and the analytical value (dashed), without Rannacher stepping (left) and with Rannacher stepping (right).

Negative probabilities in forward equation

Another characteristic of a integration scheme is positivity-preservation, *i.e.*, if a time-slice is strictly positive, the next time-slice will still be. When solving the forward equation, this is a highly desirable feature, as a probability must be positive to have a physical sense. Unfortunately, among all considered schemes, only the full implicit have this characteristic and this scheme is considerable slow compared to splitting schemes. To mitigate the issue, in every situation that a negative probability arisen when solving a step of the forward equation, we set it to 0. Normally, it can lead to loss of mass, as the probability will not sum to 1 any more. We could have renormalised the function but it introduces non-linearity, as we solve a local issue with a global correction. We shall study this issue further.

4.3 Calibration

In the ideal case, the problem of calibration is to choose the parameters of the model such that the prices constructed with the model match the market prices. We have seen until now our model as a dynamic equation. Let us consider, in an more general and abstract framework, a model as an operator T that has as input:

- the models parameters $\mathbf{p}$
- the details of a financial product $\mathbf{x}$

and returns the product price $\mathbf{y}$. Ideally, we would like to find $\mathbf{p}$ such that $T(\mathbf{p};\mathbf{x})$ such that model outputs are the "correct" prices, that are only defined for some products in a calibration set $\mathcal{C}$, constituted usually of European Calls and Puts. Hence, for a product in a calibration set, given the product parameters $\mathbf{x}$ and its price $\mathbf{y}$, we want to find the parameters $\mathbf{p}$ that reproduce this price. This problem is not evidently well posed. The prices are not necessarily in the image of the operator, meaning that there is possibly no parameter p leading to y. Solutions need to be defined in a generalised sense, and still, they might not be unique and might not depend continuously of $\mathbf{p}$.

We define a generalised solution to the inverse problem, in the sense that we consider that p is a solution of the problem if it minimises some distance

$$d\left[(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n), (T(\mathbf{p}, \mathbf{y}_1), T(\mathbf{p}, \mathbf{y}_2), \ldots, T(\mathbf{p}, \mathbf{y}_n)\right],$$

for $1, \ldots, n$ representing the inputs and prices for every product in the calibration set. We shall find the parameters that best fit the market prices, but if the solutions of inverse problem are not continuous on their parameters, the problem will not be stable, and trying to increase the accuracy of our calibration will normally increase its instability. This distance often consider regularisation techniques to add stability to the solutions, among them, Tikhonov regularisation [LO97].

One of the main reasons that we have chosen to consider only continuous stochastic volatility and a calibration set of only European Calls and Puts is that in this quite simple context, the problem is often well-behaved by considering d as Euclidean distance,

$$d(x, y) = \sum (x_i - y_i)^2,$$

such that the problem of calibration becomes a Non-Linear Least Squares optimisation. Parameters still might be subject to a constraint in order to have keep model consistence. A penalty regularisation term could be included to improve stability but we keep this simpler version in a first moment. Considering, for example, jump-diffusion models [CT04] or a calibration set with American options [AP05a] leads to strongly ill-posed problems.

4.3.1 Calibration of parametric models

Knowing the current underlying level and volatility (S, ν), we solve the forward Kolmogorov equation with initial condition $G(t, s, v) = \delta(s - S)\delta(v - \nu)$ and we will find that prices will be, for Calls and Puts of $(T, K) \in \mathcal{C}$:

$$u^{Call}(t, S, \nu) = \int G(T, s, v)(s - K)^+ \, ds \, dv.$$

$$u^{Put}(t, S, \nu) = \int G(T, s, v)(K - s)^+ \, ds \, dv.$$

If we want to calibrate our model with time-constant parameters:

- Set the parameters of the model $(\alpha^{(0)}, ..., \omega^{(0)})$
- Due to the step-by-step solving of the PDEs, we can adjust the time-steps in order to obtain all $G^{(0)}(T_i, S, V)$ at once.

- With the Green's function, calculate the prices in the model.
- Repeat till the convergence of the parameters.

otherwise we can do a boot-strapping procedure to obtain step-wise constant parameters, that change in each one of the available expiries:

- Guess an initial set $(\alpha_1^{(0)}, ..., \omega_1^{(0)})$ of model parameters
- Solve the PDE for $G^{(0)}(T_1, S, V)$ at once.
- Integrate G against the pay-off to obtain the prices at the first maturity.
- Repeat until the convergence of the parameters of the first time slice.
- Store $(\alpha_1^{(*)}, ..., \omega_1^{(*)})$ and $G^{(*)}(T_1, S, V)$ when convergence is attained.
- Knowing the semi-group property of probability transition functions

$$G_{t,s,y}(T_2, s', y') = G_{t,s,y}(T_1, s', y') \cdot G_{T_1,s,y}(T_2, s', y'),$$

  we solve for the next step starting from T_1.

A simple improvement to this routine is to consider a first calibration using asymptotic closed formulas instead of the exact solution by PDE. Then, our initial guess of parameters is improved and we may attain a good accuracy in less iterations of the exact routine, that is considerably costly.

4.3.2 Calibration of non-parametric models

Let us consider the following non-parametric local stochastic volatility model, where Y_t is a stochastic process driving the stochastic volatility:

$$dS_t = \sigma_L(t, S_t)g(Y_t)dW_t.$$

We have by Markovian Projection that the Dupire's local volatility that replicates the market is given by:

$$\sigma^2_{\text{Dup}}(T, K) = \mathbb{E}[\sigma^2_L(T, S_T)g^2(T, Y_T) \mid S_T = K] = \sigma^2_L(T, K)\mathbb{E}[g^2(T, Y_T) \mid S_T = K],$$

and we can perfectly calibrate our model to the vanilla market prices by setting:

$$\sigma^2_L(T, K) = \frac{\sigma^2_{\text{Dup}}(T, K)}{V(T, K)}. \tag{4.26}$$

where

$$V(T, K) = \frac{\int G(T, K, v) \cdot g^2(T, v)\, dv}{\int G(T, K, v)\, dv.}$$

An idea for the calibration algorithm is:

1. Set $\sigma^2_L \equiv 1$. The model can be parametrically calibrated. In the last step of the calibration, we store the Green's function, $G^{(0)}$.

2. Calculate the conditional variance $V^{(0)}$ using $G^{(i)}$, and by (4.26) obtain the first estimate of $\sigma_L^{(i+1)}$.

3. Solve the Forward-Kolmogorov equation with $\sigma_L^{(i+1)}$, obtaining the new Green's Function $G^{(i+1)}$

4. Iterate (2-3) until it converges.

It is a fixed-point search algorithm, one should still demonstrate that this is a well-posed problem, with a unique fixed point, and that the algorithm converges.

5

Numerical analysis of SV models

5.1 Heston model - Pricing Calls

In order to verify the accuracy of our PDE solver, we test it in a context where analytical formulas for pricing vanillas exist. We choose the Heston model. For its popularity, many expositions of its implementations with Finite Differencing Method exists. It presents some difficulties that will be outlined here.

For the Heston model, in the coordinates (s, ν) the backward operator becomes:

$$L = \frac{1}{2}\nu s^2 \partial_{ss} + \rho\nu\alpha s \partial_{s\nu} + \frac{1}{2}\alpha^2 \partial_{\nu\nu} + \kappa(\theta - \nu)\partial_\nu. \tag{5.1}$$

We started by semi-discretizing the equation with central finite differences. To simplify its implementation, we ignored the boundary terms in the operator and imposed artificial Dirichlet conditions set to the intrinsic value, expecting that the approximation would be good enough to not interfere drastically with the accuracy of the solution. The domain was truncated to $[0, S = 2K] \times [0, V = 1]$ and uniform grids were used. This naïve implementation though leaded to a misbehaved solution. By improving our understanding of the qualitative properties of the Heston model, we were able to implement it better numerically as well. In order to do so, we put the Heston equation in log-coordinates $(x = \log S, \nu)$ and write it on the form (4.8). We obtain:

$$\begin{cases} D(x,\nu) = \frac{1}{2}\nu \begin{pmatrix} 1 & \rho\alpha \\ \rho\alpha & \alpha^2 \end{pmatrix} \\ v(x,\nu) = \nu \begin{pmatrix} \frac{1}{2} \\ \kappa \end{pmatrix} + \begin{pmatrix} \frac{1}{2}\rho\alpha \\ \frac{1}{2}\alpha^2 - \kappa\theta \end{pmatrix} \\ f = \kappa. \end{cases}$$

Coordinates and grid choice. Tavella-Randall consider Heston in spot-coordinates and a uniform grid, as we did initially. We believe that using a non-uniform grid may improve the accuracy of the solution, as typical values of variance in markets are quite low, however the domain must be extended to large values to have a good upper boundary behaviour.

In 't Hout and Foulon adopt a grid generating function like (4.15), and Heston equation in (S, ν) coordinates, truncating the domain in the empirical threshold of $8S$ and $5K$. It results in a bigger concentration of grid points around $(S = K, \nu = 0)$.

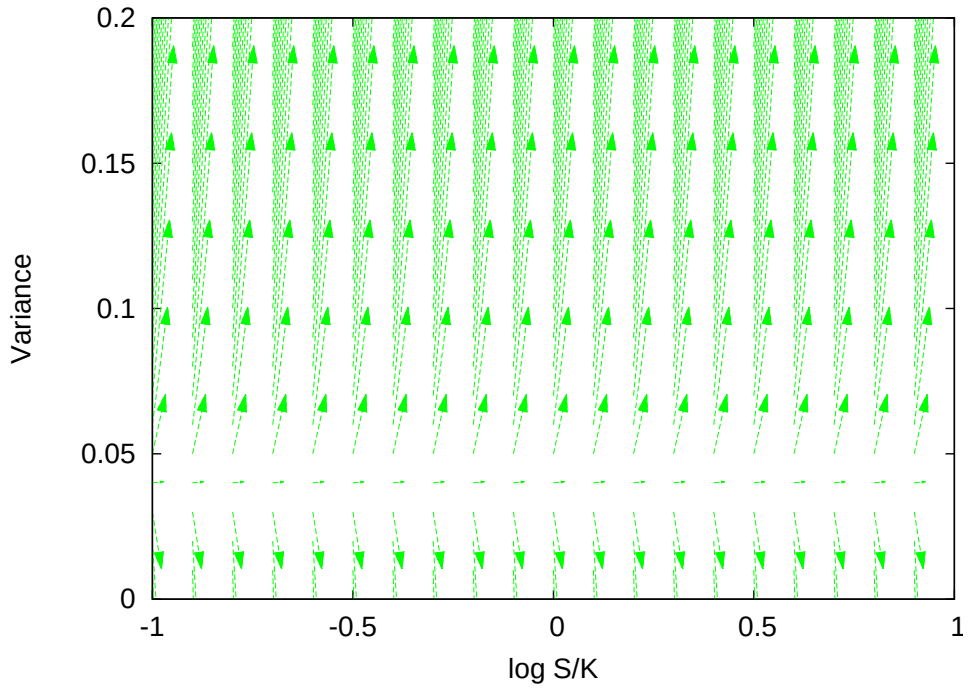

FIGURE 5.1: Typical Heston convection field. Note that the boundary $\nu = 0$ is an outflow boundary. $\theta = 0.04$, $\kappa = 2$, $\eta = 0.3$

Kluge considers Heston in log-coordinates (x, ν) and a grid generating function (4.15). Kluge derives a "good" truncation region by approximating the operator L in each direction and estimating optimality conditions in these simpler cases.

Boundary conditions. Heston [Hes93] considers for the Call option the following boundary conditions:

$$u(T, s, v) = (S - K)^+ \tag{5.2a}$$
$$u(t, 0, v) = 0, \tag{5.2b}$$
$$\frac{\partial u}{\partial s}(t, \infty, v) = 1, \tag{5.2c}$$
$$\left(\partial_t + \frac{1}{2}\nu\partial_{xx} + \frac{1}{2}\alpha^2\partial_{\nu\nu} - \frac{1}{2}\nu\partial_x + \kappa\theta\partial_\nu\right) u(t, s, 0) = 0 \tag{5.2d}$$
$$u(t, s, \infty) = 0 \tag{5.2e}$$

The equation (5.2a) is simply the Call pay-off, which gives as well the equation (5.2b). For $S \to +\infty$, the pay-off approaches its intrinsic value $S - K$ and we expect (5.2c) to hold.

The pay-off do not explicitly depend on volatility and there are not natural boundaries conditions in this dimension. If Feller condition is respected, the lower boundary is an outflow border. Close to $\nu = 0$, diffusion is almost absent and the equation is convection dominated. Intuitively, the values in these regions are determined by the transport of the neighbouring inner values to the outer region, and it not seems very reasonable imposing any condition. Instead we obtain it by solving the degenerate equation (5.2d) as part of the problem with an one-sided upward scheme. Indeed, it is the choice in [TR00], [Klu02], [itHF10], [S.08].

	κ	θ	α	ρ	K	ν_0
Set 1	1.5	0.04	0.3	-0.9	100	0.04
Set 2	2	0.1	0.2	0	100	0.01

TABLE 5.1: Set of parameters to Heston simulation

Kluge and Tavella-Randall consider BC1 condition for all boundaries, that is, assume linearity (in spot coordinates). It is a 3×3-stencil discretization and we may then use ADI and tri-diagonal system solvers to have very fast performance for the PDE solver.

For the a Up-Out Call barrier, that we will consider here, we set the upper boundary in the S dimension to the barrier level H. Then, instead of (5.2c), we consider a Dirichlet condition equals to the rebate, that we will w.l.o.g consider 0.

Summary. We solve the equation (5.1) with the parameters displayed on Table 5.1. We consider two kinds of boundary conditions:

- BC1,
- or (5.2);

in two kind of $[0, S] \times [0, V]$ grids:

- uniform, centred in (K, V/2)
- non-uniform sinh (4.15) anchored in $(K, 0)$ and with concentration $(K/5, V/500)$.

The spatial discretization is done with central finite differences in the interior domain and one-sided first-order schemes in the boundaries when necessary. We will also consider up-winding schemes for the convection in the ν direction in the sub-domain $\nu > 1$. Additionally, we consider many different integration schemes for the system of ODE.

5.1.1 Numerical study of the convergence

Spatial errors. We firstly study the error on the spatial discretization. In order to do so, we compare the semi-analytical solution of Heston model with the numerical value obtained by integration using HV($1/2 - \sqrt{3}/6$) and fairly small time steps ($\Delta t = 1/200$), so we avoid to observe a stalling on the convergence rate due to a coarse time stepping[1].

Our region of interest is given by the points $\{\frac{1}{2}K < s_i < \frac{3}{2}K, 0 < \nu_j < 1\}$. The S coordinate is not problematic because the desired points will normally coincide with the center of the discretized domain. However, for the variance coordinate, if we consider uniform grids, we will always be considering points close to the lower boundary.

We will compare error at the point $(100, 0.04)$ in the following domains, considering the discretization with boundaries Call asymptotics :

- Regular mesh $[0, 400] \times [0, 0.5]$
- Sinh mesh $[0, 800] \times [0, 5]$, anchored in $(100, 0)$ with concentration $(20, 0.01)$.

[1] The convergence order of the solution is $\mathcal{O}(\Delta x^p + \Delta t^q) = \mathcal{O}(\Delta t^q)$ if the time step is too large.

Both give good results at the center, as we can see in the Fig. 5.2. However, we can go much further in the domain with the irregular mesh, and we can consider a quite large region without incurring in a big loss of accuracy. To illustrate that, we have inserted in the same figure the global error for the irregular mesh given by:

$$e(m1, m2) = \max\{|u(0, s_i, \nu_j) - \bar{u}(0, s_i, \nu_j)| : \tfrac{1}{2}K < s_i < \tfrac{3}{2}K, 0 \leq \nu_j \leq 1\}) \qquad (5.3)$$

This effective domain covers most of actual market cases and we see that wee have still a very good accuracy, hence we retain the sinh mesh for the rest of the analysis.

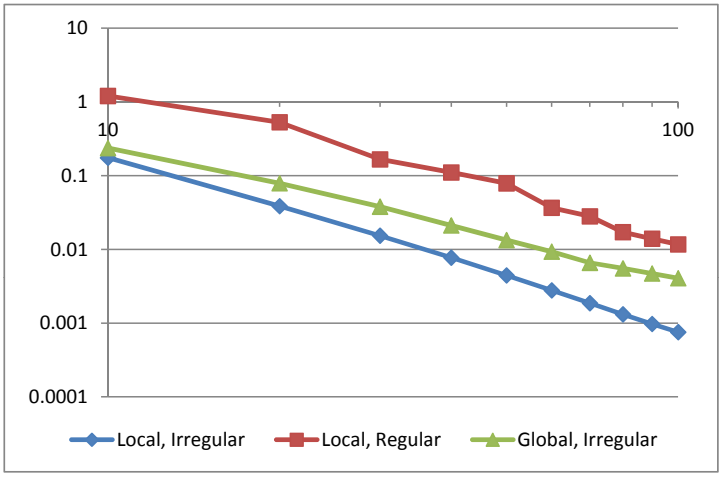

FIGURE 5.2: Global and local error vs m for Heston Call prices with Set 1 parameters.

Then, having selected the sinh mesh, we have compared the effect of the different boundary conditions. The figure exhibit the global errors to three different cases. We observed that effectively the Heston boundaries (5.2) and the BC1 lead to the same results, both cases giving almost identical results. Imposing the intrinsic value as we did initially leads to a stronger boundary effect that results in loss of accuracy.

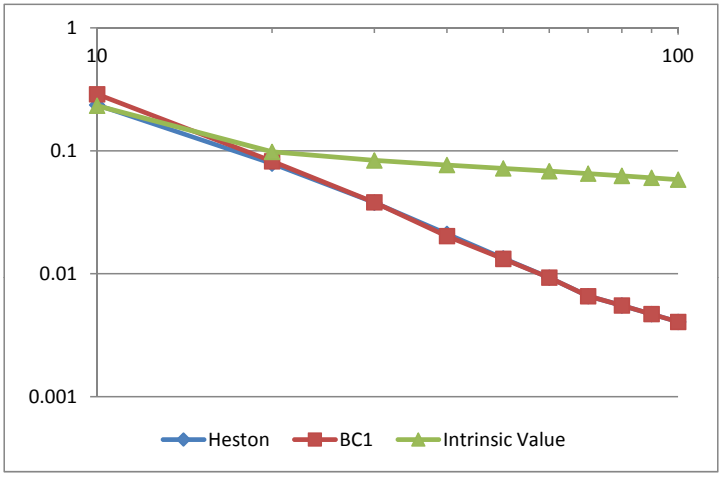

FIGURE 5.3: Global errors $e(2m, m)$ vs m for Heston Call prices considering a sinh mesh and different boundary conditions with Set 1 parameters.

Lastly, we compare the effects of considering the underlying in log-coordinates, as Kluge have done. We consider:

- Log-coordinates, sinh mesh $[\log 12.5, \log 800] \times [0, 0.5]$, anchored in $(100, 0)$ with concentration $(0.001, 0.01)$.

- Log-coordinates, sinh mesh $[\log 12.5, \log 800] \times [0, 0.5]$, anchored in $(100, 0)$ with concentration $(0.001, 0.001)$.
- Spot-coordinates sinh mesh $[0, 800] \times [0, 5]$, anchored in $(100, 0)$ with concentration $(20, 0.01)$.

We observe that, for effects of order of convergence, they are equivalent. The log-coordinates has the effect of concentrating the points for the small values of the grid, which is quite redundant, as the grid is already concentrated due to the sinh transformation. Indeed, we observe that we have a larger error in the log-coordinates, because the far-region too lightly sampled. We will obtain a better accuracy around the money, but we could have obtained the same effect by increasing the grid concentration c. Anyway, the log-transformation is still useful to understand qualitative properties easier.

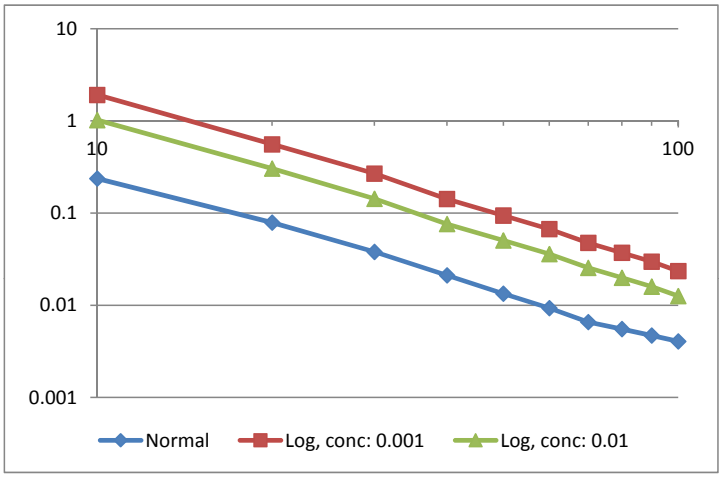

FIGURE 5.4: Global errors $e(2m, m)$ vs m for Heston Call prices using normal and log-coordinates for the underlying and Set 1 parameters.

On account of the preceding discussion, we will consider the spatial discretization in spot coordinates, with a mom-uniform sinh mesh.

Temporal errors. Now, we consider a fixed spatial mesh 40×20. We use the same error measure and we consider the convergence of the schemes:

- First order implicit
- Crank-Nicolson
- Second order implicit
- Craig-Sneyd ($1/2$)
- Modified Craig-Sneyd ($1/3$)
- HV ($1 - \sqrt{2}/2 \approx 0.29$)
- HV ($1/2 + \sqrt{3}/6 \approx 0.78$)

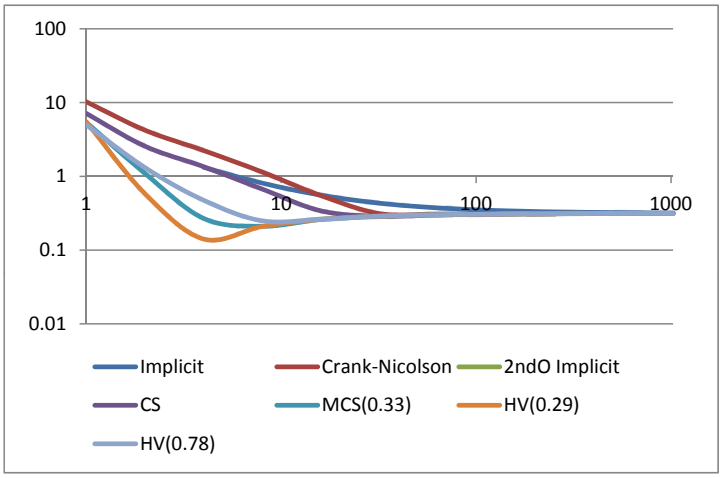

FIGURE 5.5: Global errors $e(40, 20)$ vs number of time steps for Heston Call prices with Set 1 parameters. The error is considered with respect the analytical value .

We consider the **global** errors with respect to the analytical solution firstly, to see the general behaviour of the solution. We observe a stalling of the convergence by the spatial step size and the solution converges to a biased value. The comparison of the different schemes are displayed on Fig. 5.5.

Next, we consider the **global** error with respect to the asymptotic solution (which we approximate by the solution with $N_{\max} = 1024$) to analyse the order of convergence 5.6. We observe a erratic behaviour in the schemes Crank-Nicolson and Craig-Sneyd. HV schemes are rather stable, specially HV(0.78). But the schemes do not attain their theoretical order of convergence, probably because the mesh is too coarse and the boundary effects are still large. If we consider a finer spatial mesh 100×50, we have the results 5.7. Direct schemes performs less than ADI schemes. Crank-Nicolson still presents some instability. We remark that stability issues can be mitigated by Rannacher stepping. All ADI schemes attain its theoretical order of convergence, with HV(0.78) being the most stable. We retain scheme for the rest of the analysis.

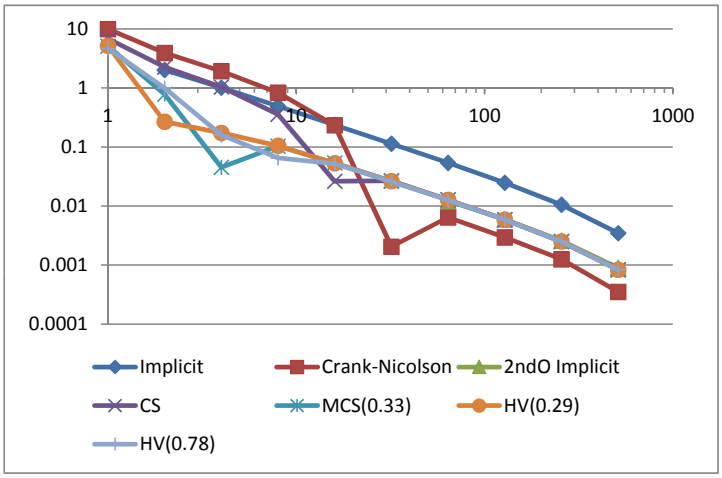

FIGURE 5.6: Global errors $e(40, 20)$ vs number of time steps for Heston Call prices with Set 1 parameters. The error is considered with respect the numerical value with 1024 time steps.

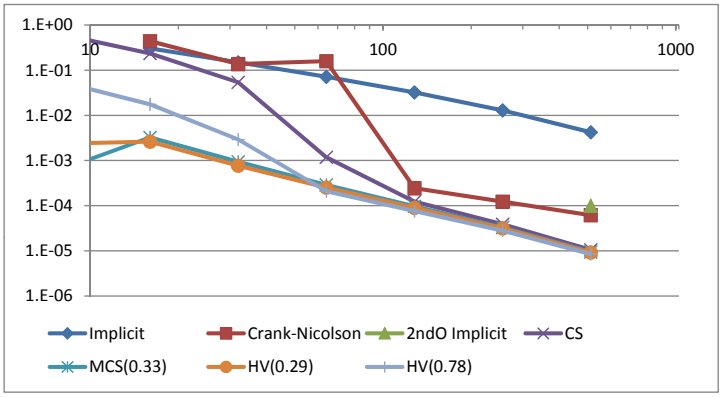

FIGURE 5.7: Global errors $e(100, 50)$ vs number of time steps for Heston Call prices with Set 1 parameters. The error is considered with respect the numerical value with 1024 time steps.

5.2 Heston model - Calibration

To improve our knowledge about the problems that can arise calibrating a model, we have studied Heston model once more. The Forward equation is:

$$\partial_t G = \partial_{xx}\left[\frac{1}{2}\nu G\right] + \partial_{x\nu}[\rho\alpha\nu G] + \partial_{\nu\nu}\left[\frac{1}{2}\alpha^2\nu G\right] + \partial_\nu\left[\frac{1}{2}\nu G\right] - \partial_\nu\left[(\kappa(\theta-\nu)G\right]. \quad (5.4)$$

With some straight-forward calculations lead to an equation in the default form (4.5). Doing these calculations instead of implementing directly (5.4) contributes for reducing the discretization errors.

For the spatial discretization, we considered a sinh non-uniform grid centred in the initial condition values, as it will concentrate the grid around the "bulk" of the probability transition function and may improve the accuracy of the integration necessary to obtain the prices in the last time-slice. In Fig. 5.9, we observe the errors with respect to the same values found by solving the forward equation.

The initial condition is a Dirac Function (cell-averaged) that will diffuse and cover the whole domain. The boundary conditions are somewhat easier to handle because the probability density may not even attain the boundaries in the lifetime of the solution. For the sake of clarity, we show an example of the evolution of the Green's function in Fig. 5.8. We will just take particular care of the boundary $\nu = 0$, because attaining or crossing that border would be fundamentally inconsistent with the theory. We considered the zero-flow conditions, as recommended by [Luc08]. It is written:

$$\left[\frac{\alpha^2}{2}\frac{\partial}{\partial\nu}(\nu G) - \kappa(\theta-\nu)G + \alpha\rho\nu\frac{\partial G}{\partial x}\right]_{\nu=0} = 0 \quad (5.5)$$

which simplifies to $G(S, \nu = 0) = 0$ because G is null at $\nu = 0$ as we respect the Feller condition and the process has no probability of attaining zero variance.

Then, let us consider a fictitious calibration set, created with the semi-analytical prices of the Heston model. Finally, we use the parametric calibration procedure to calibrate a set of market values. To solve the minimisation problem we have many methods. We have chosen the algorithm NL2Sol, an adaptative solver which implementation was adapted from Numerical Recipes [PTVF92]. We will need to compute the prices in each iteration of the algorithm. We have used HV(0.7) as integrator with 50 time-steps and a spatial grid of 50×50 spatial grid, and we performed the 10 calibration steps. The

results are reproduced in the figure 5.10. The analytical target was generated with the set 2 parameters. The last forward resolution in the calibration process is shown in the graph. The fit seems excellent, but when we plot an analytical curve with the obtained parameters, we observe a bias of approximately 0.5 all across the curve. Moreover, if we consider the same graph in implied volatilities, we observe a pattern in the forward solution. The bias and the pattern in forward resolution are probably due to the flooring of the probabilities and the domain truncation, and must be studied further.

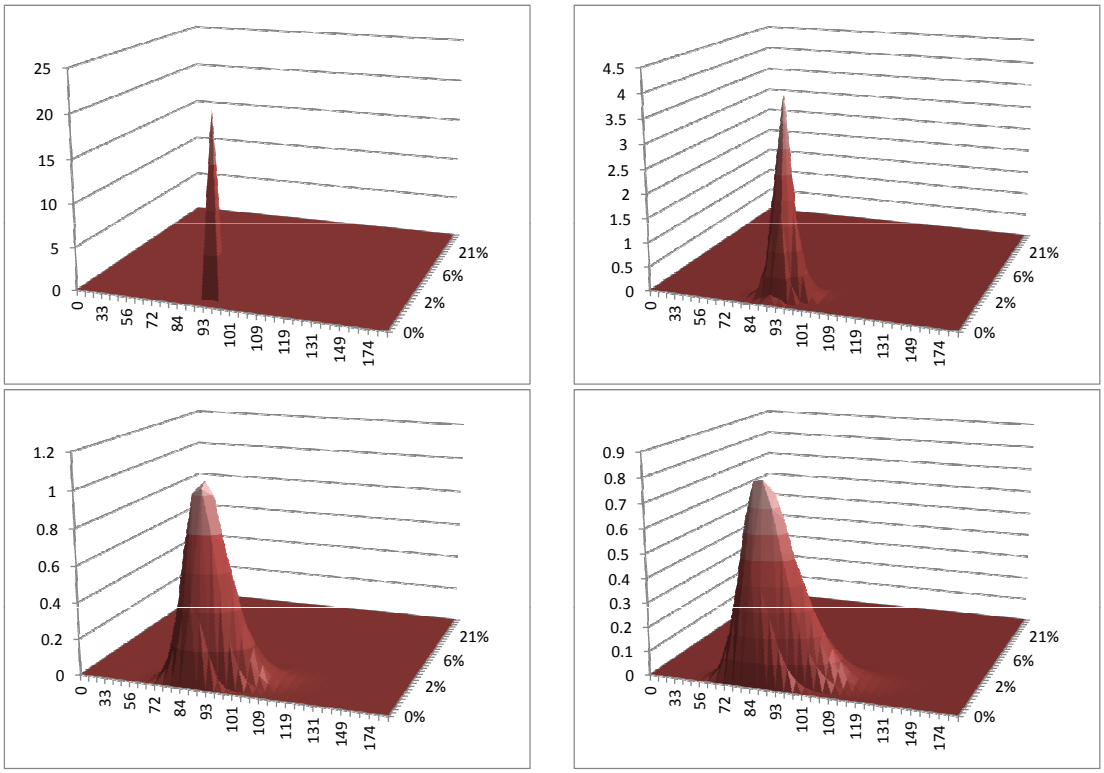

FIGURE 5.8: Evolution of Heston Green's function with Set 1 parameters after 0, 10, 50 and 100 time-steps.

5.3 Hyp-Hyp model analysis

The backward Kolmogorov operator for the Hyp-Hyp process is:

$$L = \frac{1}{2}\mathcal{L}^2\mathcal{V}^2\partial_{ss} + \rho\mathcal{L}\mathcal{V}\partial_{sy} + \alpha^2\kappa\partial_{yy} + \kappa y\partial_y. \tag{5.6}$$

with

$$\mathcal{V} = \sigma^2 g(y)$$

and

$$\mathcal{L} = S_0 \cdot f\left(\frac{s}{S_0}\right)$$

being the local and stochastic volatility terms of Hyp-Hyp.

The coordinate and boundary conditions choice will be slightly different as the y process is not the variance process but a volatility driver. The grid for the x coordinate will be the same, but for the second we will consider an uniform grid centred in 0 with

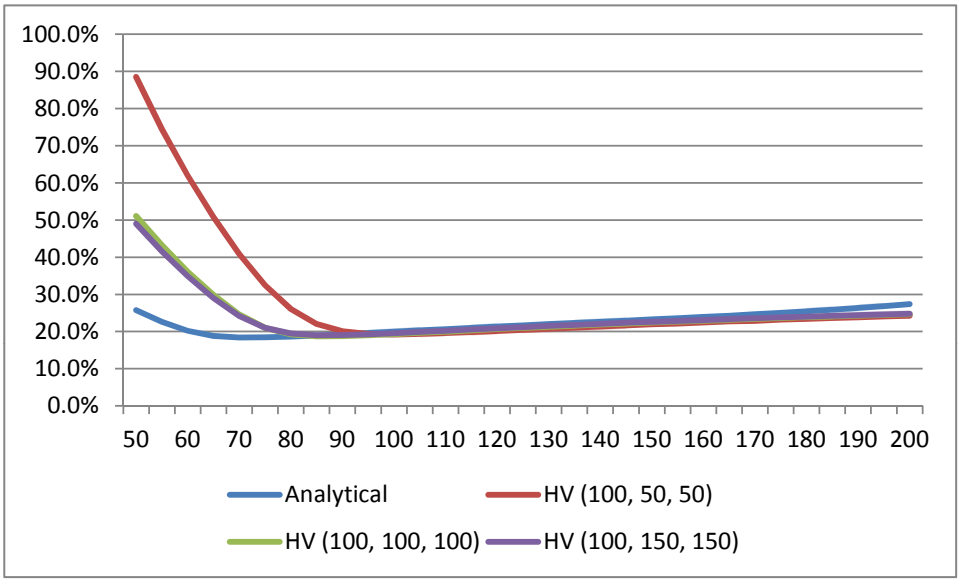

FIGURE 5.9: Errors in implied volatility with forward pricing, Set 1 parameters.

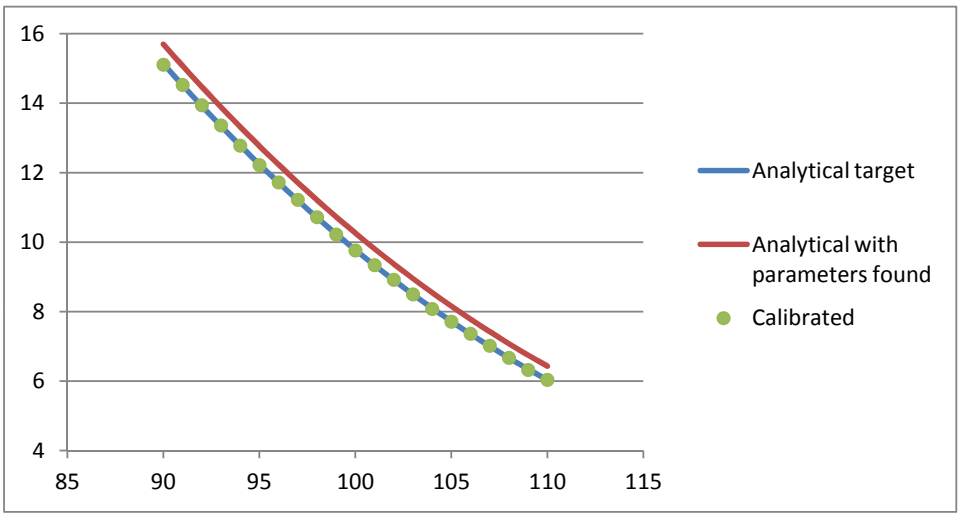

FIGURE 5.10: Calibrated prices of Heston model to the a fictitious set of market prices created with semi-analytical Heston formula using the parameters of Set 2. The semi-analytical prices obtained with calibrated parameters are shown, and we observe a bias of approximately 5% all along the curve.

boundaries [-5 σ_o, +5 σ_o]. The truncation region have been chosen taking on account that y follows a Ornstein-Uhlenbeck process and then we know its distribution, normal with standard deviation σ_o.

On account of the results to the Heston model, we have solved the equation (5.6) with the parameters displayed on Table 5.2. The integration was performed using the HV(0.7) scheme. The results seems correct when visually compared to those of Kahl-Jäckel.

5.4 Barrier prices

We have implemented barrier options in our engine to test the quality of the dynamical features of the studied models. However, as our calibration framework is still biased and we did not implement non-parametric calibration necessary to local volatility calibration,

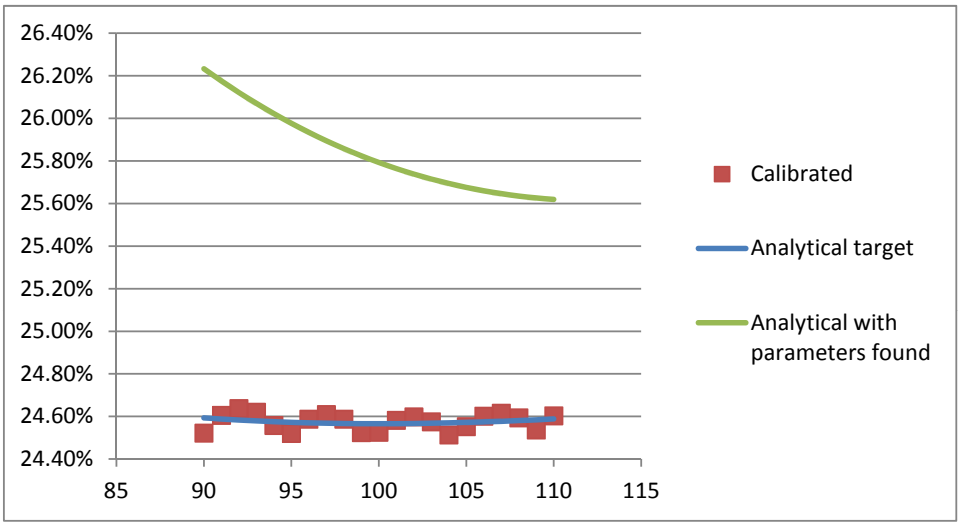

FIGURE 5.11: Calibrated implied volatilities of Heston model to the a fictitious set of market prices created with semi-analytical Heston formula using the parameters of Set 2. The semi-analytical prices obtained with calibrated parameters are shown, and we observe a bias of approximately 1.5% all along the curve.

	α	β	κ	σ_0	ρ
Set 1	3/5	1	1/2	1/5	0
Set 2	2/5	1	1	1/10	-0.5

TABLE 5.2: Set of parameters to Hyp-Hyp simulation.

we could not exhibit that even if vanilla prices may be very closely matched by different underlying dynamics, we may still have very different barrier prices. For now, we have only analysed numerical prices in Black-Scholes model, where analytical prices for Barrier options are known, obtaining a good accuracy. As our studies are still inconclusive, we will not detail our results here.

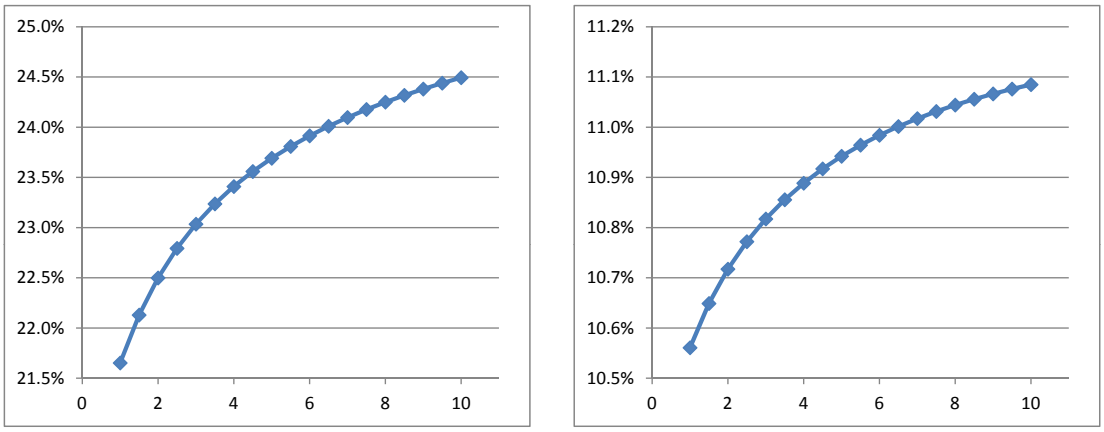

FIGURE 5.12: Term structure of the ATM implied volatility in the Hyp-Hyp model with parameters of the Set 1 (left) and set 2(right).

6

Conclusion and future developments

In this work, we have considered the modelling of the volatility from a static and a dynamic point of view. To ensure no-arbitrage, we have chosen to model the underlying price process and derive the implied volatility from it. After introducing non-parametric static approach, Dupire's volatility, that reprices any non-arbitrage surface of Calls and Puts, we considered two widely used stochastic volatility models, Heston and SABR. In a first attempt to combine goodness of calibration and dynamical features, we considered also Hyp-Hyp model, that has been developed in an attempt to remedy issues of both Heston and SABR.

The numerical pricing and calibration was done using Finite Difference Methods. We have dealt with different kinds of spatial discretization and time integration schemes. For the spatial discretization, we studied regular grids and non-uniform grids which grid generation function has a hyperbolic sine form. Combined to a boundary condition based on the call pay-off asymptotics, this choice has lead to accurate results for every initial price in the region $\{0.5K < s < 1.5K, 0 < \nu < 1\}$ and to a second-order spatial convergence. We have considered several integration schemes, notably splitting schemes. Among them, we observed that HV(0.78) performed very well, regarding both accuracy and speed, attaining its theoretical order of convergence and not presenting erratic behaviour even for coarse grids. The preceding choices had been defined as our default finite difference framework.

Future developments We shall study further how the truncation of the domain influences the discretization. For now, the cut-off has been chosen in an *ad hoc* manner. We shall attempt to develop a systematic method, in order to have a more robust discretization.

The second order implicit method have not been fully exploited. It is a promising method, but is not very numerical performing in our implementation with iterative methods (BiCGStab). That is due to the discontinuity of the Call pay-off: the second order implicit scheme have a term proportional to the fourth derivative of the pay-off, so any discontinuity becomes very pronounced, and an iterative method is prone to experience numerical instability. The spurious oscillations led to a very slow convergence. We shall study this issue further. A possible idea is to consider a direct method. It is numerically costly, but considering that most considered operators are time-constant, we can calculate the propagator matrix and perform a LU decomposition only once. And solving the linear systems with decomposed L and U matrices considerably attenuates the problem. We may consider as well other iterative methods and/or preconditioning.

In general, improving the non-split solving routines is desirable, as we may use full implicit schemes to Rannacher stepping.

We may also consider using the second order integrators combined with splitting methods, what we did not find in the literature. Every splitting scheme that we have confronted considered only combinations of first order implicit and explicit approximations of the operator.

We may consider implementing a second-order up-winding scheme. Our current first-order implementation has indeed a stabilizing effect, but led to a loss of accuracy.

Finally and most importantly, we shall improve our forward solver and calibration routine. While the backward solving is already quite robust, the set of choices to the forward solving is not yet optimized, specially considering the treatment of negative probabilities. For now, we only have considered parametric models, and hence, we have only implemented the parametric calibration algorithm. We may as well implement the non-parametric calibration algorithm in order to add a local noise to Hyp-Hyp model. As our finite difference framework handles general two-factor stochastic volatility models quite robustly, the implementation of Hyp-Hyp model with local noise shall pose no further problems in the PDE solving, but it introduces some complications regarding interpolation/extrapolation of local volatility.

Appendix

A Notations

- ∂_x, ∂_i: partial derivative with respect to the x-coordinate, $i-th$ coordinate.
- B_t: standard Brownian motion
- $\delta^-, \delta, \delta^+$: left-sided, central and right-sided finite differences.
- Δx: uniform step of spatial discretization in x-direction.
- $\Delta^- x, \Delta^+ x$: non-uniform left and right steps of spatial discretization in x-direction.
- L, $L^\dagger$, $\bar{L}$, $\bar{L}^\dagger$: spatial operator of backward Kolmogorov PDE, forward Kolmogorov PDE and its discretized versions with bar.
- S_t: underlying asset price process.
- X_t: log of underlying asset price process (possibly discounted).
- W_t: standard Brownian motion
- ADI: alternating directions implicit (4.2.5).
- BC1: linear boundary conditions (sec. 4.2.6).
- BC2: second order boundary conditions (sec. 4.2.6).
- CFL: Courant-Friedrichs-Levy condition (appendix D).
- CEV: constant elasticity of variance model (footnote in p. 10).
- CS: Craig-Sneyd integrator (sec 4.2.5).
- Damping factor: see appendix D.
- FDM: finite difference methods.
- FEM: finite element methods.
- MCS: Hundsdorfer-Verwer integrator (sec 4.2.5).
- MCS: modified Craig-Sneyd integrator (sec 4.2.5).
- PDE: partial differential equation.
- SDE: stochastic differential equation.
- Sinh mesh: irregular grid created by using the grid-generating function (4.15).
- Stable, quiet, ringing: see appendix D.
- SV: Stochastic volatility (model).
- Time-slice: the vector representing the function in a fixed instant T, *i.e.*, the discretization of $u(T, \cdot)$.

B Sticky dynamics and arbitrage

A **sticky-strike** dynamics means that the implied volatility level will be constant across time, when seen as a function of the strike K.

$$\Sigma_{S_t}(K,T) = \Sigma_{S_{t+\delta t}}(K,T).$$

As the strike of an option does not change, its implied volatility is also constant. If we have a sticky-strike heuristics, each option behave independently of the others. It is an attempt to preserve the Black-Scholes Model. A **sticky-moneyness** behaviour implies that the volatility remains the same if the ratio price over strike (the *moneyness*) is preserved:

$$\Sigma_{S_t}(K,T) = \Sigma_{S_{t+\delta t}}\left(K \cdot \frac{S_{t+\delta t}}{S_t}, T\right)$$

In other words, the implied smile $\Sigma_S(K,T)$ is a function $f(S/K)$. The base idea is having a constant level of volatility at-the-money, where the liquidity is greater. Hence, the implied volatility is shifted upon a change in the underlying price (Figure .1). Both dynamics are unfortunately trivially arbitrageable.

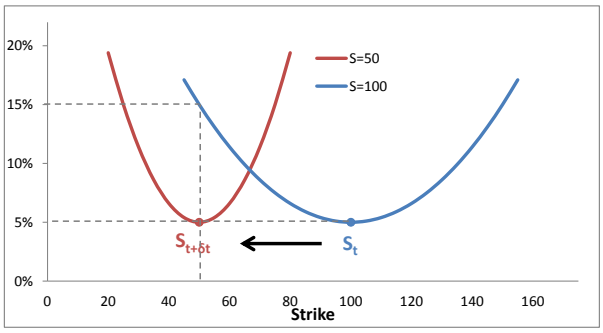

FIGURE .1: Sticky-moneyness heuristics. The surface moves in the same direction than the underlying, such that its value at-the-money is preserved.

Proposition 1. *A global sticky-strike implied volatility dynamics is arbitrageable.*

Proof. If we follow a sticky-strike heuristics, it means that the implied volatility of a Call option is constant. The P&L of a delta-hedged portfolio hedged with a constant volatility Σ instead of the market volatility σ_t is given by[1]:

$$d(\text{P\&L})_t = \frac{1}{2}\Gamma^{\Sigma}(\Sigma^2 - \sigma_t^2)S_t^2\,dt.$$

The cost of this portfolio is 0 by definition: the price of the option is defined as being the price of its hedge, so buying the option and selling the hedging portfolio must have no cost. Then, choosing two options whose strikes are K_1 and K_2 and the (constant) implied volatilities are Σ_1 and Σ_2, the P&L of the respective hedged portfolios P1 and P2 will be

$$d(\text{P1})_t = \frac{1}{2}\Gamma^{\Sigma_1}(\Sigma_1^2 - \sigma_t^2)S_t^2\,dt,$$

and

$$d(\text{P2})_t = \frac{1}{2}\Gamma^{\Sigma_2}(\Sigma_2^2 - \sigma_t^2)S_t^2\,dt.$$

[1] See eq. (6.5) of [EKJPS98] for a proof.

Then the position constituted by long Γ_2 of the first option and short Γ_1 of the second one will have a total P&L of:

$$\boxed{d(\Gamma_2\,\mathrm{P1} - \Gamma_1\,\mathrm{P2})_t = \frac{1}{2}\Gamma^{\Sigma_1}\Gamma^{\Sigma_2}(\Sigma_1^2 - \Sigma_2^2)S_t^2\,dt.}$$

The initial cost of this strategy is 0, as we have delta-hedged both portfolios. But as both Γ are strictly positive, choosing $\Sigma_1 > \Sigma_2$ ensures that the derivative of the P&L is strictly positive and so that we have a strictly positive profit. It characterizes an arbitrage. □

Proposition 2. *A global sticky-moneyness implied volatility dynamics is arbitrageable.*

Proof. (Intuition) We will show an arbitrage strategy to the particular case of downward skew. Let $K_1 < K_2$. We consider then a **delta-hedged** portfolio composed by a long call with strike K_2 and a short put with strike K_1. The cost of this strategy is null as the portfolio is delta-hedged, and the next factor contributing to the P&L is the sensibility of the price with respect the volatility, the vega Greek $\mathcal{V}$. Then, if the implied volatility follow a sticky-moneyness heuristic, we have at each instant, one of the following cases (Figure .2):

- The underlying rises. So the volatility skew shifts to the right, and the implied volatility for each strike rises as well. Then we have $\mathcal{V}_{K_1} > \mathcal{V}_{K_2}$ and then the portfolio has a positive instantaneous P&L.
- The underlying falls. So the volatility skew shifts to the left, and the implied volatility for each strike falls as well. Then we have $\mathcal{V}_{K_1} < \mathcal{V}_{K_2}$ and then the portfolio has a positive instantaneous P&L.

It constitutes an arbitrage because the P&L is positive in all the states of the world. □

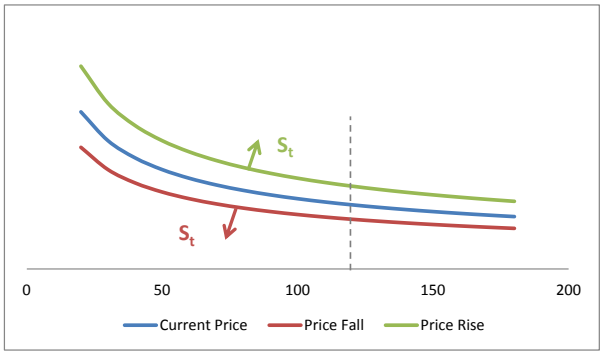

FIGURE .2: Volatility skew shift upon a rise (resp. fall) of the underlying price. The curve is shifted to the right (resp. left). As we can observe, for a fixed Call/Put option strike level (dashed line), its implied volatility will increase (resp. decrease).

C General procedure for constructing finite difference schemes

First, we consider a subset of the points used for the approximation, $x_{j_1}, x_{j_2}, ...x_{j_N}$, that is called the **stencil** of the scheme. Then, we simply write

$$\frac{\partial^l u}{\partial x^l}(x_i, t) = \sum_1^N \alpha_i u(x_{j_i})$$

and expand the right-hand-side in Taylor Series around the point x_i. Matching the terms of both sides up to the $O(\Delta x^p)$, whenever it is possible, gives a p-th order scheme.

A stencil is **compact** when it covers only the point and its first neighbours. For example, in one dimension, it corresponds to $(i-1, i, i+1)$, also called 3-stencil. The 5-stencil would corresponds to $(i-2, i-1, i, i+1, i+2)$, and is not compact. Central finite differences for first and second derivatives requires a 3-stencil. However, to approximate a second derivative in the boundary, we must use a second neighbour of the point (either 2 or $M-2$), which requires a non-compact stencil. The stencil of a scheme controls how many bands are filled in the matrix representation of the operator.

D Stability of time-discretization schemes

A scheme will be said **consistent** when it coincides with the propagator in the limit that the temporal step tends to 0.

The concept of **stability** is concerned with how a scheme deals with small perturbations on the initial conditions. It is directly related with the eigenvalues and modes of the operator. A scheme can be:

- *Stable*: an *unconditionally stable* scheme is such that the eigenvalues of the approximated propagator will be contained inside the unit circle or exactly equal to 1 with multiplicity 1. Such operator will guarantee that small errors will not be magnified. Sometimes, the schema is stable under a condition relating the spatial and temporal time-steps. The scheme is then said *conditionally stable* under a *CFL condition*.
- *Unstable*: the modes of the function will grow with each iteration of the propagator and small numerical error will be magnified and lead to diverging oscillations.

and yet a stable scheme can be:

- *Quiet*: All the eigenvalues have positive real part. Hence, in each iteration, the sign of the modes are preserved.
- *Ringing*: In each iteration, the sign of the modes are inverted, leading to oscillation.

We can consider a scheme as a function acting on a differential operator L. We define the **damping factor** of a scheme this same function when seen as a function over the complex field. It explains how the propagator affects the eigenvalues of the differential operator. For example, for the Crank-Nicolson scheme, the damping factor is:

$$A(\omega) = \frac{1 + \frac{1}{2}\omega}{1 - \frac{1}{2}\omega}.$$

The closer the damping factor is to 0, the better the scheme damps small numerical errors in the mode of frequency ω.

A **major detail** that is often neglected in the financial literature is that the stability is a concept depending on the **norm** considered and in most of stability considerations are done considering that the scheme is applied to a **particular type of operator L**. As the operator norm of L and $L^{\dagger}$ are the same, the convergence and the stability of the discretization scheme of the former assure the convergence and the stability of the discretization of the latter. Mostly results of stability consider a set of hypothesis too restrictive to the financial context (for example, that L is pure-diffusive, have constant coefficients, etc.). For lack of an alternative, we may content ourselves of a numerical study of stability.

Stability and consistency are important as the Lax-Richtmyer Equivalence theorem states that a stable scheme that is consistent converges.

Bibliography

[AA00] L. Andersen and J. Andreasen. Jump-Diffusion Processes: Volatility Smile Fitting and Numerical Methods for Option Pricing. *Review of Derivatives Research*, 4:231–262, 2000.

[AP05a] Yves Achdou and Olivier Pironneau. Numerical procedure for calibration of volatility with american options. *Applied Mathematical Finance*, 12(3):201–241, 2005.

[AP05b] L. Andersen and V. Piterbarg. Moment Explosions in Stochastic Volatilty Models. Technical report, Bank of America, 2005. ssrn.com/abstract=559481.

[BBF02] H. Berestycki, J. Busca, and I. Florent. Asymptotics and calibration of local volatility models. *Quantitative Finance*, 2:61–69, 2002. www.iop.org/EJ/article/1469-7688/2/1/305/qf2105.pdf.

[BCC97] G. Bakshi, C. Cao, and Z. Chen. Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5):2003–2049, 1997.

[BL78] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1978.

[BS73] F. Black and M. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81:637–654, 1973.

[CM05] P. Carr and D. Madan. A Note on Sufficient Conditions for No Arbitrage. *Finance Research Letters*, 2005.

[Cox75] J. C. Cox. Notes on option pricing I: Constant elasticity of variance diffusions. Working paper, Stanford University, 1975.

[CT04] R. Cont and P. Tankov. *Financial modelling with jump processes*. Chapman & Hall / CRC, 2004.

[Der99] E. Derman. Regimes of Volatility: Some observations on the variation of S&P 500 implied volatilities. *Risk*, 12(4):55–59, 1999.

[DFW95] B. Dumas, J. Fleming, and R. E. Whaley. Implied volatility functions: Empirical tests. Preprint, Fuqua School of Business, Duke University, 1995.

[DHS07] T. Daglish, J. Hull, and W. Suo. Volatility surfaces: theory, rules of thumb, and empirical evidence. *Quantitative Finance*, 7(5):507–524, 2007.

[DK94] E. Derman and I. Kani. Riding on a Smile. *Risk*, 7(2):32–39, 1994.

[Dup94] B. Dupire. Pricing with a Smile. *Risk*, 7(1):18–20, 1994.

[dV92] H. Van der Vorst. A fast and smoothly converging variant of BiCG for the solution of nonsysmmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 13:631–644, 1992.

[EKJPS98] N. El-Karoui, M. Jeanblanc-Pique, and S. Shreve. Robustness of the Black and Scholes formula . *Mathematical Finance*, 8(2):93–126, April 1998.

[Gat04] J. Gatheral. A parsimonious arbitrage-free implied volatility parameterisation with application to the valuation of volatility derivatives. In *TDTF Derivatives Day Amsterdam*, 2004.

[GC06] M.B. Giles and R. Carter. Convergence analysis of Crank-Nicolson and Rannacher time-marching. *J. Comp. Fin.*, 9(4):89–112, 2006.

[Gyö86] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itō differential. *Probability Theory and Related Fields*, 71:501–516, 1986.

[Hes93] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6:327–343, 1993.

[HKL02] P. Hagan, D. Kumar, and A. S. Lesniewski. Managing Smile Risk. *Wilmott Magazine*, pages 84–108, September 2002.

[HL09] P. Henry-Labordère. *Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing*. Chapman & Hall, 2009.

[HP81] J. M. Harrison and S. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11:215–260, 1981.

[HW87] J. Hull and A. White. The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance*, 42(2):281–300, June 1987. faculty.baruch.cuny.edu/lwu/890/HullWhite87.pdf.

[itHF10] K.J. in 't Hout and S. Foulon. ADI finite difference schemes for option pricing in the Heston model with correlation. *Int. J. Numer. Anal. Mod.*, 7:303–320, 2010.

[Jäc04] P. Jäckel. Stochastic volatility models - past, present, and future. pages 355–377. John Wiley and Sons, 2004. The original presentation given at the "Quantitative Finance Review" conference in November 2003 in London can be found at www.jaeckel.org/StochasticVolatilityModels-PastPresentAndFuture.pdf.

[Jäc06] P. Jäckel. Hyperbolic local volatility. www.jaeckel.org/HyperbolicLocalVolatility.pdf, November 2006.

[Jef04] C. Jeffery. Reverse cliquets: end of the road. *Risk*, February:20–22, 2004.

[KJ05] C. Kahl and P. Jäckel. Not-so-complex logarithms in the Heston model. *Wilmott Magazine*, September:94–103, September 2005.

[KJ07] C. Kahl and P. Jäckel. Hyp Hyp Hooray, June 2007. www.jaeckel.org/HypHypHooray.pdf.

[KL07] C. Kahl and R. Lord. Why the Rotation Count Algorithm Works. Technical report, 2007. ssrn.com/abstract=921335.

[Klu02] T. Kluge. Pricing derivatives in stochastic volatility models using the finite difference method. Master's thesis, Technische Universität Chemnitz, September 2002. kluge.in-chemnitz.de/documents/diploma/diploma.pdf.

[Lip02] A. Lipton. The vol smile problem. *Risk*, February 2002.

[LO97] R. Lagnado and S. Osher. A technique for calibrating derivative security pricing models: Numerical solution of an inverse problem. Computing in Economics and Finance 101, Society for Computational Economics, 1997.

[Luc08] V. Lucic. Boundary Conditions for Computing Densities in Hybrid Models via PDE Methods. Technical report, 2008. ssrn.com/abstract=1191962.

[M.A98] M.A.Botchev. Simple bicgstab(l) iterative method, 1998. www.staff.science.uu.nl/~vorst102/bcg2.f.

[Pit07] V. Piterbarg. Markovian projection for volatility calibration. *Risk*, 4:84–89, 2007.

[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992. www.nrbook.com/a/bookcpdf.php.

[PVF03] D.M. Pooley, K. R. Vetzal, and P.A Forsyth. Convergence Remedies For Non-Smooth Payoffs in Option Pricing. *J. Comp. Fin.*, 6(4):25–40, 2003.

[Ran84] R. Rannacher. Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, 43:309–327, 1984.

[S.08] Lin S. Finite Difference Schemes for Heston Model. Master's thesis, University of Oxford, 2008. eprints.maths.ox.ac.uk/718/1/Sensen_Lin_thesis.pdf.

[Saa03] Y. Saad. *Iterative Methods for Sparse Linear Systems, 2nd edition*. SIAM, Philadelpha, PA, 2003.

[Sco87] L. Scott. Option Pricing When the Variance Changes Randomly – Theory, Estimation and An Application - Scott. *Journal of Financial and Quantitative Analysis*, 22:419–438, December 1987.

[SF93] Gerard L.G. Sleijpen and Diederik R. Fokkema. BiCGStab(L) for Linear Equations involving unsymmetric matrices with complex spectrum. *Electronic Transactions on Numerical Analysis*, pages 11–32, September 1993.

[Shi93] D. Shimko. Bounds of probability. *Risk*, 6(4):33–37, 1993.

[TR00] D. Tavella and C. Randall. *Pricing Financial Instruments: The Finite Difference Method*. John Wiley and Sons, April 2000. ISBN 0471197602.

[Wig87] J. Wiggins. Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics*, 19:351–372, 1987.