High order discretization schemes for stochastic volatility models

Benjamin Jourdain, Mohamed Sbai

High order discretization schemes for stochastic volatility models

Benjamin Jourdain and *Mohamed Sbai*[1]

Abstract

In usual stochastic volatility models, the process driving the volatility of the asset price evolves according to an autonomous one-dimensional stochastic differential equation. We assume that the coefficients of this equation are smooth. Using Itô's formula, we get rid, in the asset price dynamics, of the stochastic integral with respect to the Brownian motion driving this SDE. Taking advantage of this structure, we propose

- a scheme, based on the Milstein discretization of this SDE, with order one of weak trajectorial convergence for the asset price,
- a scheme, based on the Ninomiya-Victoir discretization of this SDE, with order two of weak convergence for the asset price.

We also propose a specific scheme with improved convergence properties when the volatility of the asset price is driven by an Ornstein-Uhlenbeck process. We confirm the theoretical rates of convergence by numerical experiments and show that our schemes are well adapted to the multilevel Monte Carlo method introduced by Giles [2008*b*, 2008*a*].

Introduction

There exists an extensive literature on numerical integration schemes for stochastic differential equations. To start with, we mention, among many others, the work of Talay and Tubaro [1990] who first established an expansion of the weak error of the Euler scheme for polynomially growing functions allowing for the use of Romberg extrapolation. Bally and Taley [1996] extended this result to bounded measurable functions and Guyon [2006] extended it to tempered stable distributions. More recently, many discretization schemes of higher weak convergence order have appeared in the literature. Among others, we cite the work of Kusuoka [2001, 2004], the Ninomiya and Victoir [2008] scheme which we will use hereafter, the Ninomiya and Ninomiya [2009] scheme and the scheme based on cubature on Wiener spaces of Lyons and Victoir [2004].
Concerning strong approximation, the Milstein scheme has order one of strong convergence. Unfortunately, it involves the simulation of iterated Brownian integrals unless a restrictive commutativity condition is satisfied. Under ellipticity, Cruzeiro *et al.* [2004] have recently proposed a discretization scheme which gets rid of these iterated integrals and has nice strong convergence properties. More precisely, for each number of time steps, there exists a Brownian motion different from the one giving the Brownian increments involved in the scheme such that the strong error between the scheme and the stochastic differential equation driven by this new Brownian motion is of order one. We call such a property weak trajectorial convergence of order one. Weak trajectorial error estimation is exactly what is needed to control the discretization bias for the computation of path dependent option prices.

[1]Université Paris-Est, CERMICS, Projet MathFi ENPC-INRIA-UMLV. This research benefited from the support of the "Chaire Risques Financiers", Fondation du Risque and the ANR program BigMC.
Postal address : 6-8 av. Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 Marne-la-Vallée Cedex 2.
E-mails : jourdain@cermics.enpc.fr and sbai@cermics.enpc.fr

Stochastic volatility models, which have now become a standard of the market, are an eloquent example of the use of stochastic differential equations in finance. In our study, we will consider the following specification of a stochastic volatility model for an asset $(S_t)_{t\in[0,T]}$:

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + f(Y_t)S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right); \quad S_0 = s_0 > 0 \\ dY_t &=& b(Y_t)dt + \sigma(Y_t)dW_t; \quad Y_0 = y_0 \end{array}\right. \tag{1}$$

where r the instantaneous interest rate, $(B_t)_{t\in[0,T]}$ and $(W_t)_{t\in[0,T]}$ are independent standard one-dimensional Brownian motions, $\rho \in [-1,1]$ is the correlation between the Brownian motions respectively driving the asset price and the process $(Y_t)_{t\in[0,T]}$ which solves a one-dimensional autonomous stochastic differential equation. The volatility process is $(f(Y_t))_{t\in[0,T]}$ where the transformation function f is usually taken positive and strictly monotonic in order to ensure that the effective correlation between the stock price and the volatility keeps the same sign (the function σ usually takes nonnegative values). This specification nests almost all the known stochastic volatility models:

- Hull&White [1987] model ($\rho = 0$) and Wiggins [1987] ($\rho \neq 0$)

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + \sqrt{Y_t}S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right) \\ dY_t &=& \mu Y_t dt + \zeta Y_t dW_t \end{array}\right.$$

  which can be expressed as (1) with $f(y) = \sqrt{y}$, $b(y) = \mu y$ and $\sigma(y) = \zeta y$. Nlote that it can also be seen as (1) with $f(y) = e^y$, $b(y) = \frac{\mu}{2} - \frac{\zeta^2}{4}$ and $\sigma(y) = \frac{\zeta}{2}$.

- Scott's [1987] model which generalizes Hull&White model

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + \sigma_0 e^{Y_t}S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right) \\ dY_t &=& \kappa(\theta - Y_t)dt + \nu dW_t \end{array}\right. \tag{2}$$
$$\Rightarrow f(y) = \sigma_0 e^y,\ b(y) = \kappa(\theta - y) \text{ and } \sigma(y) = \nu.$$

- Stein&Stein model [1991]

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + Y_t S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right) \\ dY_t &=& \kappa(\theta - Y_t)dt + \nu dW_t \end{array}\right.$$
$$\Rightarrow f(y) = y,\ b(y) = \kappa(\theta - y) \text{ and } \sigma(y) = \nu.$$

- Quadratic Gaussian model

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + Y_t^2 S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right) \\ dY_t &=& \kappa(\theta - Y_t)dt + \nu dW_t \end{array}\right.$$
$$\Rightarrow f(y) = y^2,\ b(y) = \kappa(\theta - y) \text{ and } \sigma(y) = \nu.$$

- Heston's [1993] model

$$\left\{\begin{array}{rcl} dS_t &=& rS_t dt + \sqrt{Y_t}S_t\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right) \\ dY_t &=& \kappa(\theta - Y_t)dt + \nu\sqrt{Y_t}dW_t \end{array}\right.$$
$$\Rightarrow f(y) = \sqrt{y},\ b(y) = \kappa(\theta - y) \text{ and } \sigma(y) = \nu\sqrt{y}.$$

In all but the last example, the volatility of the asset is driven by an Ornstein Uhlenbeck process.

The development of specific discretization schemes for stochastic volatility models has only received little attention. We mention nevertheless the work of Kahl and Jäckel [2006] who discussed various numerical integration methods and proposed a simple scheme with order 1/2 of strong convergence like the standard Euler scheme but with a smaller multiplicative constant. Also the numerical integration of the CIR process and of the Heston model received a particular attention because of the inadequacy of the Euler scheme due to the fact that both f and σ are equal to the square root function (see for example Deelstra and Delbaen [1998], Alfonsi [2005], Kahl and Schurz [2006], Andersen [2007], Berkaoui *et al.* [2008], Ninomiya and Victoir [2008], Lord *et al.* [2008], Alfonsi [2009]). An exact simulation technique for the Heston model was also proposed by Broadie and Kaya [2006].

In the present paper, we assume in return that the functions f, σ and b are smooth and do not deal with the Heston model. Our aim is to take advantage of the structure of (1) to construct and analyse simple and robust ad'hoc discretization schemes which have nice convergence properties.

For a start, we make a logarithmic change of variables for the asset : the two-dimensional process $(X_t := \log(S_t), Y_t)_{t\in[0,T]}$ solves the following SDE

$$\left\{ \begin{array}{rcl} dX_t &=& (r - \frac{1}{2}f^2(Y_t))dt + f(Y_t)\left(\rho dW_t + \sqrt{1-\rho^2}dB_t\right); \quad X_0 = \log(s_0). \\ dY_t &=& b(Y_t)dt + \sigma(Y_t)dW_t; \quad Y_0 = y_0 \end{array} \right. \tag{3}$$

Our main idea is to get rid in the first equality of the stochastic integral involving the common Brownian motion $(W_t)_{t\in[0,T]}$. In all what follows, we assume that

$(\mathcal{A})$ f and σ are $\mathcal{C}^1$ functions and $\sigma > 0$.

One can then define the primitive $F(y) = \int_0^y \frac{f}{\sigma}(z)dz$ and apply Itô's formula to get

$$dF(Y_t) = \frac{f}{\sigma}(Y_t)dY_t + \frac{1}{2}(\sigma f' - f\sigma')(Y_t)dt.$$

Therefore $(X_t, Y_t)_{t\in[0,T]}$ solves

$$\left\{ \begin{array}{rcl} dX_t &=& \rho dF(Y_t) + h(Y_t)dt + \sqrt{1-\rho^2}f(Y_t)dB_t \\ dY_t &=& b(Y_t)dt + \sigma(Y_t)dW_t \end{array} \right. \tag{4}$$

where $h : y \mapsto r - \frac{1}{2}f^2(y) - \rho(\frac{b}{\sigma}f + \frac{1}{2}(\sigma f' - f\sigma'))(y)$. We discretize the autonomous SDE satisfied by Y using a scheme with high order of strong or weak convergence depending on wether one is interested in path-dependent or vanilla options. Then, in the dynamics of X, we only need to discretize the standard integral $\int_0^T h(Y_s)ds$ and the stochastic integral $\int_0^T f(Y_t)dB_t$ where $(Y_t)_{t\in[0,T]}$ and $(B_t)_{t\in[0,T]}$.

We recall that usual weak convergence is the right notion to analyse the discretization bias for plain vanilla options whereas weak trajectorial convergence permits to deal with path-dependent options. The first section of the paper is devoted to path-dependent options. Combining the Milstein discretization of the one-dimensional SDE satisfied by $(Y_t)_{t\in[0,T]}$ with an appropriate discretization of the integral $\int_0^T f(Y_t)dB_t$ based on the independence of $(Y_t)_{t\in[0,T]}$ and $(B_t)_{t\in[0,T]}$, we obtain a scheme with order one of weak trajectorial convergence. In the second section, using the Ninomiya-Victoir discretization of the SDE satisfied by $(Y_t)_{t\in[0,T]}$, we construct a scheme with order two of weak convergence. Since the SDE satisfied by Y is one-dimensional, the Ninomiya-Victoir scheme only involves two one-dimensional ODEs whose solutions are available in closed form. The last section is devoted to numerical experiments which confirm the theoretical rates of convergence. We also show that our schemes are well adapted to the multilevel Monte Carlo method introduced by Giles [2008*b*, 2008*a*].

Notations

We will consider, for a number of time steps $N \geq 1$, the uniform subdivision $\prod_N = \{0 = t_0 < t_1 < \cdots < t_N = T\}$ of $[0,T]$ with the discretization step $\delta_N = \frac{T}{N}$.
We denote by $\underline{\psi}$ the greatest lower bound of the function $\psi : y \mapsto f^2(y)$ and by $\overline{\psi}$ its lowest upper bound. We also introduce the following notation :

$$\widehat{\psi}(y) = \begin{cases} \frac{3}{2}f^2(y) & \text{if } \overline{\psi} = \infty \\ \overline{\psi} & \text{otherwise} \end{cases}$$

1 An efficient scheme for path dependent options pricing

Building a first order strong convergence scheme for a two dimensional SDE is not an obvious task. Even the ad'hoc schemes provided by Kahl and Jäckel [2006] exhibit a strong convergence of order $\frac{1}{2}$.

Actually, the natural candidate for this purpose is the Milstein scheme. Unfortunately, the commutativity condition which permits to implement it amounts to $\sigma f' = 0$ in our setting. This condition is typically true when either f is constant or $\sigma = 0$. Both cases are of no practical interest since they lead to a deterministic volatility.

However, since the inherent Brownian motion is not essential for applications in finance, the usual strong convergence criterion is not adapted for estimating the error of a scheme in pricing a path dependent option. What is more relevant is the approximation in law of the whole trajectory of the process considered for instance by Cruzeiro *et al.* [2004]. Using an ingenious rotation of the Brownian motion, these authors have constructed a discretization scheme allowing for a weak convergence on the whole trajectory of order one which avoids the simulation of the iterated stochastic integrals.

For the SDE (3), the discretization scheme of Cruzeiro, Malliavin and Thalmaier writes as

$$\begin{aligned} X^{CMT}_{t_{k+1}} &= X^{CMT}_{t_k} + \left(r - \frac{f^2(Y^{CMT}_{t_k})}{2}\right)\delta_N + \rho f(Y^{CMT}_{t_k})\Delta W_{k+1} + \tfrac{\rho}{2}\sigma f'(Y^{CMT}_{t_k})\Delta W^2_{k+1} \\ &\quad + \sqrt{1-\rho^2}\sigma f'(Y^{CMT}_{t_k})\Delta W_{k+1}\Delta B_{k+1} + \sqrt{1-\rho^2} f(Y^{CMT}_{t_k})\Delta B_{k+1} - \tfrac{\rho}{2}\sigma f'(Y^{CMT}_{t_k})\Delta B^2_{k+1} \\ Y^{CMT}_{t_{k+1}} &= Y^{CMT}_{t_k} + \left(b(Y^{CMT}_{t_k}) + \tfrac{1}{2}(\tfrac{\sigma^2 f'}{f} - \sigma\sigma')(Y^{CMT}_{t_k})\right)\delta_N + \sigma(Y^{CMT}_{t_k})\Delta W_{k+1} \\ &\quad + \tfrac{1}{2}\sigma\sigma'(Y^{CMT}_{t_k})\Delta W^2_{k+1} - \tfrac{\sigma^2 f'}{2f}\Delta B^2_{k+1} \end{aligned} \tag{5}$$

where $\Delta W_{t_{k+1}} = W_{t_{k+1}} - W_{t_k}$ and $\Delta B_{k+1} = B_{t_{k+1}} - B_{t_k}$ correspond to the Brownian increments.

We set out to construct a much simpler scheme having the same order of weak trajectorial convergence by taking advantage of the particular structure of the SDE defining stochastic volatility models. We first begin with the general case of any process $(Y_t)_{t\in[0,T]}$ driving the volatility and then consider the case of an Ornstein-Uhlenbeck process where we obtain more precise results.

1.1 General case

A discretization scheme will naturally involve the Brownian increments. Thanks to the independence between $(Y_t)_{t\in[0,T]}$ and $(B_t)_{t\in[0,T]}$, we can construct a vector $(\widetilde{X}_{t_0}, \ldots, \widetilde{X}_{t_N})$ using only $(\Delta B_1, \ldots, \Delta B_N)$ and $(Y_t)_{t\in[0,T]}$, which has exactly the same law as $(X_{t_0}, \ldots, X_{t_N})$:

Lemma 1 — *$\forall 0 \leq l < N$, let $v_l = \frac{1}{\delta_N}\int_{t_l}^{t_{l+1}} \psi(Y_s)ds$. The vector $(\widetilde{X}_{t_0}, \ldots, \widetilde{X}_{t_N})$ defined by*

$$\begin{aligned} &\widetilde{X}_{t_0} = X_{t_0} \\ &\forall 1 \leq k \leq N, \widetilde{X}_{t_k} = \widetilde{X}_{t_0} + \rho(F(Y_{t_k}) - F(Y_{t_0})) + \int_0^{t_k} h(Y_s)ds + \sqrt{1-\rho^2}\sum_{l=0}^{k-1}\sqrt{v_l}\,\Delta B_{l+1} \end{aligned}$$

has the same law as $(X_{t_0},\ldots,X_{t_N})$.

Proof : The proof is elementary. Conditionally on Y, the two vectors are Gaussian vectors with the same mean and covariance matrix. □

In order to approximate $(\widetilde{X}_{t_k})_{0\le k\le N}$, one needs to discretize v_k for $k\in\{0,\ldots,N-1\}$. If $(v_k^N)_{0\le k\le N-1}$ is an approximation of $(v_k)_{0\le k\le N-1}$, then by Doob's inequality

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le k\le N-1}\left(\sum_{l=0}^{k}\left(\sqrt{v_l}-\sqrt{v_l^N}\right)\Delta B_{l+1}\right)^2\right] &\le 4\delta_N\sum_{k=0}^{N-1}\mathbb{E}\left[\left(\sqrt{v_l}-\sqrt{v_l^N}\right)^2\right]\\ &\le \frac{1}{\underline{\psi}}\,\delta_N\sum_{k=0}^{N-1}\mathbb{E}\left[\left(v_l-v_l^N\right)^2\right]\end{aligned}$$

as soon as $\underline{\psi}=\inf_x\psi(x)$ is assumed to be positive and, $\forall 0\le k\le N-1$, v_k^N is greater than $\underline{\psi}$. Consequently, to obtain a scheme with order one of strong convergence for $(\widetilde{X}_{t_k})_{0\le k\le N}$, one needs that $\forall 0\le k\le N-1,\mathbb{E}\left[\left(v_k-v_k^N\right)^2\right]=\mathcal{O}\left(\frac{1}{N^2}\right)$. According to the treatment of the term $\overline{I}_2^j$ defined by (9) in the proof of the Theorem 2 below, one has $\forall 0\le k\le N-1$,

$$\mathbb{E}\left[\left(v_k-\left(\psi(Y_{t_k})+\frac{\sigma\psi'(Y_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s-W_{t_k})ds\right)\right)^2\right]=\mathcal{O}\left(\frac{1}{N^2}\right). \tag{6}$$

This equality still holds true when replacing Y by a scheme with order one of strong convergence in the term with sign minus of the left hand side. Better still, $\left(F(Y_{t_k})+\int_0^{t_k}h(Y_s)ds\right)_{0\le k\le N}$ is approximated with strong order one when replacing Y by such a scheme and using a rectangular discretization for the integral in time.

For all these reasons, we choose the Milstein scheme for Y :

$$\forall 0\le k\le N-1,\ \widetilde{Y}_{t_{k+1}}^N=\widetilde{Y}_{t_k}^N+b(\widetilde{Y}_{t_k}^N)\delta_N+\sigma(\widetilde{Y}_{t_k}^N)\Delta W_{k+1}+\frac{1}{2}\sigma\sigma'(\widetilde{Y}_{t_k}^N)\left(\Delta W_{k+1}^2-\delta_N\right);\quad \widetilde{Y}_{t_0}^N=y_0.$$

and we write our scheme as follows

WeakTraj_1 scheme

$$\begin{aligned}\widetilde{X}_{t_{k+1}}^N=\widetilde{X}_{t_k}^N&+\rho\left(F(\widetilde{Y}_{t_{k+1}}^N)-F(\widetilde{Y}_{t_k}^N)\right)+\delta_N h(\widetilde{Y}_{t_k}^N)\\ &+\sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}_{t_k}^N)+\frac{\sigma\psi'(\widetilde{Y}_{t_k}^N)}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s-W_{t_k})ds\right)\vee\underline{\psi}}\;\Delta B_{k+1}\end{aligned} \tag{7}$$

Note that in order to implement this scheme, one needs to simulate both the Brownian increment ΔW_{k+1} and the random variable $\int_{t_k}^{t_{k+1}}(W_s-W_{t_k})ds$. This is straightforward as one can easily check that

$$\begin{pmatrix}\Delta W_{k+1}\\ \displaystyle\int_{t_k}^{t_{k+1}}(W_s-W_{t_k})ds\end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\delta_N & \delta_N^2/2\\ \delta_N^2/2 & \delta_N^3/3\end{pmatrix}\right)$$

We can now state our first main result :

Theorem 2 — *Under the assumptions of Lemma 3 and if*

(H1) *f and σ are $\mathcal{C}^3$ functions, $\frac{f}{\sigma}$ and ff' are bounded*

(H2) *$\underline{\psi} > 0$*

(H3) *there exists a constant K_1 such that, $\forall (x,y) \in \mathbb{R}^2$,*

$$\begin{aligned}\Big|(bh' + \frac{\sigma^2}{2}h'')(y)\Big| &\leq K_1(1+|y|)\\ \big|\sigma h'(y)\big| &\leq K_1(1+|y|)\\ \Big|h(y) - h(x)\Big| &\leq K_1|y-x|\end{aligned}$$

(H4) *there exists a constant K_2 such that, $\forall (x,y) \in \mathbb{R}^2$,*

$$\begin{aligned}\Big|(b\psi' + \frac{\sigma^2}{2}\psi'')(y)\Big| &\leq K_2(1+|y|)\\ \big|\sigma\psi'(y) - \sigma\psi'(x)\big| &\leq K_2|y-x|\end{aligned}$$

then the WeakTraj_1 scheme has order one of weak trajectorial convergence. More precisely, for each $p \geq 1$, there exists a constant C independent of the number of time steps N such that

$$\mathbb{E}\left[\max_{0\leq k\leq N}\left\|\left(\widetilde{X}_{t_k}, Y_{t_k}\right) - \left(\widetilde{X}^N_{t_k}, \widetilde{Y}^N_{t_k}\right)\right\|^{2p}\right] \leq \frac{C}{N^{2p}}.$$

The proof of the theorem relies on the order one of strong convergence of the Milstein scheme (see Milstein [1995] for the particular case $p = 1$) :

Lemma 3 — *Suppose that*

(H5) *b and σ are $\mathcal{C}^2$ functions with bounded first and second derivatives*

(H6) *there exists a positive constant K such that $\forall (x,y) \in \mathbb{R}^2$*

$$|\sigma\sigma'(x) - \sigma\sigma'(y)| \leq K|x-y|$$

then, $\forall p \geq 1$, there exists a positive constant C_p independent of N such that

$$\mathbb{E}\left(\max_{0\leq k\leq N}\left|Y_{t_k} - \widetilde{Y}^N_{t_k}\right|^{2p}\right) \leq C_p\delta_N^{2p}.$$

The proof for general p is postponed to the appendix.

Remark 4 — *Before giving the proof of the theorem, we make a few comments on its assumptions. (H1) implies that h and ψ are $\mathcal{C}^2$ functions which was implicitly assumed in (H3) and (H4). The latter assumptions are expressed in a reduced form. One can check that the following conditions on the coefficients of the original SDE are sufficient for them to hold :*

- *f and σ are bounded $\mathcal{C}^4$ functions with bounded derivatives.*
- *b is a bounded $\mathcal{C}^3$ function with bounded derivatives.*

- $\exists \sigma_0 > 0$ *such that* $\forall y \in \mathbb{R}$, $\sigma(y) \geq \sigma_0$.

Proof of the theorem : Throughout the proof, we denote by C a constant which can change from one line to another while always being independent of N. Thanks to Lemma 3, we just have to control the error on $\widetilde{X}$:

$$\begin{aligned}\mathbb{E}\Big[\max_{0\leq k\leq N}|\widetilde{X}_{t_k}-\widetilde{X}^N_{t_k}|^{2p}\Big] &= \mathbb{E}\Bigg[\max_{0\leq k\leq N}\Big|\rho(F(Y_{t_k})-F(\widetilde{Y}^N_{t_k}))+\sum_{j=0}^{k-1}\Bigg(\int_{t_j}^{t_{j+1}}h(Y_s)ds-\delta_N h(\widetilde{Y}^N_{t_j})\\ &\quad+\sqrt{\frac{1-\rho^2}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds}\,\Delta B_{j+1}\\ &\quad-\sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}^N_{t_j})+\frac{\sigma\psi'(\widetilde{Y}^N_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\vee\underline{\psi}}\;\Delta B_{j+1}\Big|^{2p}\Bigg)\Bigg]\\ &\leq 3^{2p-1}\left(\rho^{2p}I_0+I_1+(1-\rho^2)^pI_2\right)\end{aligned}$$

where

$$I_0=\mathbb{E}\left[\max_{0\leq k\leq N}\Big|F(Y_{t_k})-F(\widetilde{Y}^N_{t_k})\Big|^{2p}\right]$$

$$I_1=\mathbb{E}\left[\max_{0\leq k\leq N}\Big|\sum_{j=0}^{k-1}\left(\int_{t_j}^{t_{j+1}}h(Y_s)ds-\delta_N h(\widetilde{Y}^N_{t_j})\right)\Big|^{2p}\right]$$

and

$$\begin{aligned}I_2=\mathbb{E}\Bigg[\max_{0\leq k\leq N}\Big|\sum_{j=0}^{k-1}\Bigg(&\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds}-\\ &\sqrt{\left(\psi(\widetilde{Y}^N_{t_j})+\frac{\sigma\psi'(\widetilde{Y}^N_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\vee\underline{\psi}}\;\Delta B_{j+1}\Big|^{2p}\Bigg].\end{aligned}$$

$(\mathcal{H}1)$ yields that F is Lipschitz continuous so using Lemma 3 we show that $I_0\leq\frac{C}{N^{2p}}$. Next, we have that

$$I_1\leq C\left(\mathbb{E}\left[\max_{0\leq k\leq N}\Big|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}h(Y_s)ds-\delta_N h(Y_{t_j})\Big|^{2p}\right]+\delta_N^{2p}\mathbb{E}\left[\max_{0\leq k\leq N}\Big|\sum_{j=0}^{k-1}h(Y_{t_j})-h(\widetilde{Y}^N_{t_j})\Big|^{2p}\right]\right)$$

On one hand, thanks to assumption $(\mathcal{H}1)$ and Lemma 3,

$$\delta_N^{2p}\mathbb{E}\left[\max_{0\leq k\leq N}\Big|\sum_{j=0}^{k-1}h(Y_{t_j})-h(\widetilde{Y}^N_{t_j})\Big|^{2p}\right]\leq C\delta_N\sum_{j=0}^{N-1}\mathbb{E}\left[\Big|h(Y_{t_j})-h(\widetilde{Y}^N_{t_j})\Big|^{2p}\right]\leq\frac{C}{N^{2p}}.$$

On the other hand, using an integration by parts formula,

$$\begin{aligned}\overline{I}_1 &:= \mathbb{E}\left[\max_{0\leq k\leq N}\Big|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}h(Y_s)-h(Y_{t_j})ds\Big|^{2p}\right]\\ &= \mathbb{E}\left[\max_{0\leq k\leq N}\Big|\sum_{j=0}^{k-1}\int_{t_j}^{t_{j+1}}(t_{j+1}-s)\left((bh'+\frac{\sigma^2h''}{2})(Y_s)ds+\sigma h'(Y_s)dW_s\right)\Big|^{2p}\right]\\ &\leq 2^{2p-1}\left(\mathbb{E}\left[\max_{0\leq k\leq N}\Big|\int_0^{t_k}(\tau_s-s)(bh'+\frac{\sigma^2h''}{2})(Y_s)ds\Big|^{2p}\right]+\mathbb{E}\left[\max_{0\leq k\leq N}\Big|\int_0^{t_k}(\tau_s-s)\sigma h'(Y_s)dW_s\Big|^{2p}\right]\right)\end{aligned}$$

where we denoted by τ_s the lowest discretization point greater than s : $\tau_s = \lceil \frac{s}{\delta_N} \rceil \delta_N$. Using Jensen's inequality for the first integral and the Burkholder-Davis-Gundy inequality for the second, we obtain

$$\begin{aligned}
\overline{I}_1 &\leq C\left(\mathbb{E}\left[\max_{0\leq k\leq N} t_k^{2p-1}\int_0^{t_k}(\tau_s-s)^{2p}\left|(bh'+\frac{\sigma^2h''}{2})(Y_s)\right|^{2p}ds\right]\right. \\
&\qquad \left. +\mathbb{E}\left[\left(\int_0^T(\tau_s-s)^2\left|\sigma h'(Y_s)\right|^2ds\right)^p\right]\right) \\
&\leq \frac{C}{N^{2p}}\int_0^T\mathbb{E}\left[\left|(bh'+\frac{\sigma^2h''}{2})(Y_s)\right|^{2p}+\left|\sigma h'(Y_s)\right|^{2p}\right]ds.
\end{aligned}$$

Under the assumptions of Lemma 3, $\sup_{0\leq t\leq T}\mathbb{E}(|Y_s|^{2p}) < \infty$ (see Problem 3.15 p. 306 of Karatzas and Shreve [1991] for example) so, with the help of assumption $(\mathcal{H}3)$, we conclude that $\overline{I}_1 \leq \frac{C}{N^{2p}}$ and hence $I_1 \leq \frac{C}{N^{2p}}$. We now turn to the last term. Using Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned}
I_2 &\leq C\delta_N^p\mathbb{E}\left[\left(\sum_{j=0}^{N-1}\left(\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds}-\sqrt{\left(\psi(\widetilde{Y}_{t_j}^N)+\frac{\sigma\psi'(\widetilde{Y}_{t_j}^N)}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\vee\underline{\psi}}\right)^2\right)^p\right] \\
&\leq \delta_N\sum_{j=0}^{N-1}\mathbb{E}\left[\left|\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds}-\sqrt{\left(\psi(\widetilde{Y}_{t_j}^N)+\frac{\sigma\psi'(\widetilde{Y}_{t_j}^N)}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\vee\underline{\psi}}\right|^{2p}\right]
\end{aligned} \tag{8}$$

Assumption $(\mathcal{H}2)$ yields that the two terms appearing in the square root are bounded from below by $\underline{\psi} > 0$ so we have that

$$\begin{aligned}
I_2 &\leq C\delta_N\sum_{j=0}^{N-1}\mathbb{E}\left[\left|\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds-\left(\psi(\widetilde{Y}_{t_j}^N)+\frac{\sigma\psi'(\widetilde{Y}_{t_j}^N)}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\vee\underline{\psi}\right|^{2p}\right] \\
&\leq CN^{2p-1}\sum_{j=0}^{N-1}\mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}\psi(Y_s)ds-\left(\psi(\widetilde{Y}_{t_j}^N)\delta_N+\sigma\psi'(\widetilde{Y}_{t_j}^N)\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\right|^{2p}\right] \\
&\leq CN^{2p-1}\sum_{j=0}^{N-1}\left(\overline{I}_2^j+\widetilde{I}_2^j\right)
\end{aligned}$$

where

$$\overline{I}_2^j = \mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}\psi(Y_s)ds-\left(\psi(Y_{t_j})\delta_N+\sigma\psi'(Y_{t_j})\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\right|^{2p}\right] \tag{9}$$

and

$$\widetilde{I}_2^j = \mathbb{E}\left[\left|\delta_N\left(\psi(Y_{t_j})-\psi(\widetilde{Y}_{t_j}^N)\right)+\left(\sigma\psi'(Y_{t_j})-\sigma\psi'(\widetilde{Y}_{t_j}^N)\right)\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right|^{2p}\right]$$

Again, integrating by parts yields that

$$\overline{I}_2^j = \mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}(t_{j+1}-s)\left((\sigma\psi'(Y_s)-\sigma\psi'(Y_{t_j}))dW_s+((b\psi'+\frac{\sigma^2}{2}\psi'')(Y_s))ds\right)\right|^{2p}\right]$$

We control the stochastic integral term as follows

$$\begin{aligned}\mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}(t_{j+1}-s)(\sigma\psi'(Y_s)-\sigma\psi'(Y_{t_j}))dW_s\right|^{2p}\right] &\leq C\delta_N^{p-1}\mathbb{E}\left[\int_{t_j}^{t_{j+1}}(t_{j+1}-s)^{2p}|\sigma\psi'(Y_s)-\sigma\psi'(Y_{t_j})|^{2p}ds\right]\\ &\leq C\delta_N^{3p-1}\int_{t_j}^{t_{j+1}}\mathbb{E}\left[\left|\sigma\psi'(Y_s)-\sigma\psi'(Y_{t_j})\right|^{2p}\right]ds\\ &\leq C\delta_N^{3p-1}\int_{t_j}^{t_{j+1}}\mathbb{E}\left[\left|Y_s-Y_{t_j}\right|^{2p}\right]ds\\ &\leq C\delta_N^{3p-1}\int_{t_j}^{t_{j+1}}|s-t_j|^p\,ds\\ &\leq C\delta_N^{4p}\end{aligned}$$

The third inequality is due to assumption $(\mathcal{H}4)$ and the fourth one is a standard result on the control of the moments of the increments of the solution of a SDE with Lipschitz continuous coefficients (see Problem 3.15 p. 306 of Karatzas and Shreve [1991] for example).

We also control the other term thanks to assumption $(\mathcal{H}4)$:

$$\begin{aligned}\mathbb{E}\left[\left|\int_{t_j}^{t_{j+1}}(t_{j+1}-s)(b\psi'+\frac{\sigma^2}{2}\psi'')(Y_s)ds\right|^{2p}\right] &\leq \delta_N^{2p-1}\mathbb{E}\left[\int_{t_j}^{t_{j+1}}(t_{j+1}-s)^{2p}|(b\psi'+\frac{\sigma^2}{2}\psi'')(Y_s)|^{2p}ds\right]\\ &\leq \delta_N^{4p-1}\int_{t_j}^{t_{j+1}}\mathbb{E}\left[\left|(b\psi'+\frac{\sigma^2}{2}\psi'')(Y_s)\right|^{2p}\right]ds\\ &\leq C\delta_N^{4p}\end{aligned}$$

Hence, $\overline{I}_2^j \leq \frac{C}{N^{4p}}$. To conclude the proof of the theorem, it remains to show a similar result for $\widetilde{I}_2^j$:

$$\begin{aligned}\widetilde{I}_2^j &\leq 2^{2p-1}\mathbb{E}\left[\left|\delta_N\left(\psi(Y_{t_j})-\psi(\widetilde{Y}_{t_j}^N)\right)\right|^{2p}+\left|\left(\sigma\psi'(Y_{t_j})-\sigma\psi'(\widetilde{Y}_{t_j}^N)\right)\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right|^{2p}\right]\\ &\leq C\left(\delta_N^{2p}\mathbb{E}\left[\left|Y_{t_j}-\widetilde{Y}_{t_j}^N\right|^{2p}\right]+\frac{\delta_N^{3p}}{3^p}\mathbb{E}\left[\left|Y_{t_j}-\widetilde{Y}_{t_j}^N\right|^{2p}\right]\right)\\ &\leq \frac{C}{N^{4p}}\end{aligned}$$

The second inequality is due to the fact that ψ is Lipschitz continuous (thanks to assumption $(\mathcal{H}1)$) for the first term and to the independence of $\left(\sigma\psi'(Y_{t_j})-\sigma\psi'(\widetilde{Y}_{t_j}^N)\right)$ and $\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds$ for the second term. □

Remark 5 — *Our scheme exhibits the same convergence properties as the Cruzeiro et al. [2004] scheme. Apart from the fact that it involves less terms, it presents the advantage of improving the multilevel Monte Carlo convergence. This method, which is a generalization of the statistical Romberg extrapolation method of Kebaier [2005], was introduced by Giles [2008*b*, 2008*a*].*

Indeed, consider the discretization scheme with time step $\delta_{2N}=\frac{T}{2N}$:

$$\begin{aligned}\forall 0\leq k\leq 2N-1,\ \widetilde{X}^{2N}_{\frac{(k+1)T}{2N}} &= \widetilde{X}^{2N}_{\frac{kT}{2N}}+\rho\left(F(\widetilde{Y}^{2N}_{\frac{(k+1)T}{2N}})-F(\widetilde{Y}^{2N}_{\frac{kT}{2N}})\right)+\delta_{2N}h(\widetilde{Y}^{2N}_{\frac{kT}{2N}})+\sqrt{1-\rho^2}\\ &\times\sqrt{\left(\psi(\widetilde{Y}^{2N}_{\frac{kT}{2N}})+\frac{\sigma\psi'(\widetilde{Y}^{2N}_{\frac{kT}{2N}})}{\delta_{2N}}\int_{\frac{kT}{2N}}^{\frac{(k+1)T}{2N}}(W_s-W_{\frac{kT}{2N}})ds\right)\vee\underline{\psi}}\left(B_{\frac{(k+1)T}{2N}}-B_{\frac{kT}{2N}}\right)\end{aligned}$$

Denote by $v_k^{2N} = \sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}^{2N}_{\frac{kT}{2N}}) + \frac{\sigma\psi'(\widetilde{Y}^{2N}_{\frac{kT}{2N}})}{\delta_{2N}}\int_{\frac{kT}{2N}}^{\frac{(k+1)T}{2N}}(W_s - W_{\frac{kT}{2N}})ds\right)\vee\underline{\psi}}$ *the random variable which multiplies the increment of the Brownian motion* $\left(B_{\frac{(k+1)T}{2N}} - B_{\frac{kT}{2N}}\right)$. *Because of the independence properties,* $\left(\widetilde{X}^N_{t_k}\right)_{0\le k\le N}$ *has the same distribution law as the vector* $\left(\widetilde{\widetilde{X}}^N_{t_k}\right)_{0\le k\le N}$ *defined inductively by* $\widetilde{\widetilde{X}}^N_{t_0} = \log(s_0)$ *and*

$$\forall 0\le k\le N-1,\ \widetilde{\widetilde{X}}^N_{t_{k+1}} = \widetilde{\widetilde{X}}^N_{t_k} + \rho\left(F(\widetilde{Y}^N_{t_{k+1}}) - F(\widetilde{Y}^N_{t_k})\right) + \delta_N h(\widetilde{Y}^N_{t_k}) + \sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}^N_{t_k}) + \frac{\sigma\psi'(\widetilde{Y}^N_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds\right)\vee\underline{\psi}}\,\Delta\widetilde{B}^N_{k+1}$$

where

$$\Delta\widetilde{B}^N_{k+1} = \sqrt{2}\left[\frac{v_{2k}^{2N}\left(B_{\frac{(2k+1)T}{2N}} - B_{\frac{2kT}{2N}}\right) + v_{2k+1}^{2N}\left(B_{\frac{(2k+2)T}{2N}} - B_{\frac{(2k+1)T}{2N}}\right)}{\sqrt{\left(v_{2k}^{2N}\right)^2 + \left(v_{2k+1}^{2N}\right)^2}}\right] \tag{10}$$

Going over the proof of the theorem, one can show in the same way that

$$\mathbb{E}\left[\max_{0\le k\le N}\left|\widetilde{\widetilde{X}}^N_{t_k} - \widetilde{X}^{2N}_{t_k}\right|^2\right] = \mathcal{O}(N^{-2}) \tag{11}$$

Hence, one can apply the multilevel Monte Carlo method to compute the expectation of a Lipschitz continuous functional of X *and reduce the computational cost to achieve a desired root-mean-square error of* $\epsilon > 0$ *to a* $\mathcal{O}(\epsilon^{-2})$.

As a matter of fact, the particular structure of our scheme enabled us to reconstruct the coupling which allows to efficiently control the error between the scheme with time step $\frac{T}{N}$ *and the one with time step* $\frac{T}{2N}$. *This does not seem possible with the Cruzeiro et al. [2004] scheme.*

From a practical point of view, it is more interesting to obtain a convergence result for the stock price. It is also more challenging because the exponential function is not globally Lipschitz continuous. We can nevertheless state the following corollary with some general assumptions and we will see in the next section that we can make them more precise in case $(Y_t)_{t\in[0,T]}$ is an Ornstein-Uhlenbeck process.

Corollary 6 — *Let* $p \ge 1$. *Under the assumptions of Theorem 2 and if*

(H7)

$$\exists\epsilon > 0 \text{ such that } \mathbb{E}\left[\max_{0\le k\le N} S_{t_k}^{2p+\epsilon}\right] + \mathbb{E}\left[\max_{0\le k\le N} e^{(2p+\epsilon)\widetilde{X}^N_{t_k}}\right] < \infty$$

then there exists a positive constant C *independent of* N *such that*

$$\mathbb{E}\left[\max_{0\le k\le N}\left|e^{\widetilde{X}_{t_k}} - e^{\widetilde{X}^N_{t_k}}\right|^{2p}\right] \le \frac{C}{N^{2p}}$$

Proof : Using Hölder inequality we have that

$$\begin{aligned}\mathbb{E}\left[\max_{0\le k\le N}\left|e^{\widetilde{X}_{t_k}}-e^{\widetilde{X}^N_{t_k}}\right|^{2p}\right] &\le \mathbb{E}\left[\max_{0\le k\le N}\left(e^{2p\widetilde{X}_{t_k}}\vee e^{2p\widetilde{X}^N_{t_k}}\right)\left|\widetilde{X}_{t_k}-\widetilde{X}^N_{t_k}\right|^{2p}\right]\\ &\le \left(\mathbb{E}\left[\max_{0\le k\le N}S^{2p+\epsilon}_{t_k}\right]+\mathbb{E}\left[\max_{0\le k\le N}e^{(2p+\epsilon)\widetilde{X}^N_{t_k}}\right]\right)^{\frac{2p}{2p+\epsilon}}\\ &\quad\times\left(\mathbb{E}\left[\max_{0\le k\le N}\left|\widetilde{X}_{t_k}-\widetilde{X}^N_{t_k}\right|^{\frac{2p\epsilon+4p^2}{\epsilon}}\right]\right)^{\frac{\epsilon}{2p+\epsilon}}\end{aligned}$$

We conclude by assumption ($\mathcal{H}7$) and Theorem 2. □

Remark 7 — *Had we introduced a new cut-off to our scheme as follows*

$$\begin{aligned}\widetilde{X}^N_{t_{k+1}} &= \widetilde{X}^N_{t_k}+\rho\left(F(\widetilde{Y}^N_{t_{k+1}})-F(\widetilde{Y}^N_{t_k})\right)+\delta_N h(\widetilde{Y}^N_{t_k})\\ &\quad+\sqrt{1-\rho^2}\sqrt{\left(\psi(\widetilde{Y}^N_{t_k})+\frac{\sigma\psi'(\widetilde{Y}^N_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s-W_{t_k})ds\right)\wedge\overline{\psi}\vee\underline{\psi}}\;\Delta B_{k+1}\end{aligned}$$

assumption ($\mathcal{H}7$) would have been induced by assuming that the functions F, f and h are bounded.

1.2 Special case of an Ornstein-Uhlenbeck process driving the volatility

For many stochastic volatility models, the process $(Y_t)_{t\in[0,T]}$ which drives the volatility is an Ornstein-Uhlenbeck process. For example, this is the case for all the models cited in the introduction but the Heston model. Therefore, it is useful to focus on this particular case. We will hereafter suppose that $(Y_t)_{t\in[0,T]}$ is solution of the following SDE

$$dY_t=\nu dW_t+\kappa(\theta-Y_t)dt \tag{12}$$

with ν,κ and θ three positive constants. Since exact simulation is possible, we can replace the Milstein discretization by the true solution in our previous scheme :

WeakTraj_1 scheme when Y is an O-U process

$$\begin{aligned}\widetilde{X}^N_{t_{k+1}} &= \widetilde{X}^N_{t_k}+\rho\left(F(Y_{t_{k+1}})-F(Y_{t_k})\right)+\delta_N h(Y_{t_k})\\ &\quad+\sqrt{1-\rho^2}\sqrt{\left(\psi(Y_{t_k})+\frac{\nu\psi'(Y_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s-W_{t_k})ds\right)\vee\underline{\psi}}\;\Delta B_{k+1}\end{aligned} \tag{13}$$

Note that we require the exact simulation of both $(Y_{t_k},Y_{t_{k+1}})$ and $\int_{t_k}^{t_{k+1}}(W_s-W_{t_k})ds$. The unique solution of (12) is $Y_t=y_0e^{-\kappa t}+\theta(1-e^{-\kappa t})+\nu\int_0^t e^{-\kappa(t-s)}dW_s$ and one can easily deduce that, $\forall k\in\{0,\dots,N-1\}$,

$$\begin{pmatrix} Y_{t_{k+1}}-e^{-\kappa\delta_N}Y_{t_k}\\ \int_{t_k}^{t_{k+1}}(W_s-W_{t_k})ds\end{pmatrix}\sim\mathcal{N}(M,\Gamma)$$

where $M = \begin{pmatrix} \theta(1-e^{-\kappa\delta_N}) \\ 0 \end{pmatrix}$ and $\Gamma = \begin{pmatrix} \frac{\nu^2}{2\kappa}(1-e^{-2\kappa\delta_N}) & \frac{\nu}{\kappa^2}(1-e^{-\kappa\delta_N}(1+\kappa\delta_N)) \\ \frac{\nu}{\kappa^2}(1-e^{-\kappa\delta_N}(1+\kappa\delta_N)) & \frac{\delta_N^3}{3} \end{pmatrix}$.

We first state the following technical lemma whose proof is postponed to the appendix :

Lemma 8 — $\forall\, c_1 > 0, c_2 \in [0,1)$,

$$\mathbb{E}\left(e^{c_1 \sup_{0\le t\le T} |Y_t|^{1+c_2}}\right) < \infty.$$

As might be expected, it is possible to weaken the assumptions of Theorem 2. In particular, we relax the assumption on the lower bound of the volatility ($\mathcal{H}2$) and replace it with a weaker one (see assumption ($\mathcal{H}10$) below).

Theorem 9 — *Let $p \ge 1$. Suppose that Y is solution of (12) and that the scheme is defined by (13). Under assumption ($\mathcal{H}2$) of Theorem 2 and if*

($\mathcal{H}8$) *f is a $\mathcal{C}^3$ function*

($\mathcal{H}9$) *there exist three constants $c_0 > 0, c_1 > 0$ and $c_2 \in [0,1)$ such that, $\forall y \in \mathbb{R}$,*

$$\begin{aligned}
&\left|\kappa(\theta-y)h'(y) + \frac{\nu^2}{2}h''(y)\right| \le c_0 e^{c_1|y|^{1+c_2}} \\
&\left|h'(y)\right| \le c_0 e^{c_1|y|^{1+c_2}} \\
&\left|\kappa(\theta-y)\psi'(y) + \frac{\nu^2}{2}\psi''(y)\right| \le c_0 e^{c_1|y|^{1+c_2}} \\
&\left|\psi''(y)\right| \le c_0 e^{c_1|y|^{1+c_2}}
\end{aligned}$$

then there exists a constant C independent of the number of time steps N such that

$$\mathbb{E}\left[\max_{0\le k\le N}\left|\widetilde{X}_{t_k} - \widetilde{X}^N_{t_k}\right|^{2p}\right] \le \frac{C}{N^{2p}}$$

The same result holds true when we replace assumption ($\mathcal{H}2$) by

($\mathcal{H}10$) *There exist two positive constants C and ϵ such that $\forall y \in \mathbb{R}$,*

$$\begin{aligned}
\psi(y) &> 0 \\
|\psi'(y)| &\le C\psi(y) \\
\sup_{t\le T}\mathbb{E}\left(\psi^{p(1+\epsilon)}(Y_t)\right) &< \infty \\
\sup_{t\le T}\mathbb{E}\left(\frac{1}{\psi^{p(1+\epsilon)}(Y_t)}\right) &< \infty.
\end{aligned}$$

Proof : The proof of the first part of the theorem repeats the proof of Theorem 2 with fewer terms to control because of the exact simulation of $(Y_t)_{t\in[0,T]}$. At the places where we used assumptions ($\mathcal{H}3$) and ($\mathcal{H}4$), we use assumption ($\mathcal{H}9$) together with Lemma 8.

We now focus on the second part of the theorem. According to equation (8), all we have to show is the existence of a positive constant C independent of N such that $\forall j \in \{0, \dots, N-1\}$

$$\mathbb{E}\left[\left|\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds} - \sqrt{\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right)\vee\underline{\psi}}\right|^{2p}\right] \leq \frac{C}{N^{2p}}$$

We will adopt the following notations

- $A_j = \frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds$
- $D_j = \left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right)\vee\underline{\psi}$

Thanks to assumption ($\mathcal{H}10$), we have that $\forall j \in \{0, \dots, N-1\}$, $A_j > 0$ and $D_j \geq 0$. The idea of the proof is to isolate the case where D_j is small which is problematic since the square root is not Lipschitz continuous in the neighborhood of 0 :

$$\begin{aligned}
\left|\sqrt{A_j} - \sqrt{D_j}\right|^{2p} &= \left|\sqrt{A_j} - \sqrt{D_j}\right|^{2p}\mathbb{1}_{\{D_j \leq \psi(Y_{t_j})/2\}} + \left|\sqrt{A_j} - \sqrt{D_j}\right|^{2p}\mathbb{1}_{\{D_j > \psi(Y_{t_j})/2\}} \\
&\leq 2^{2p-1}\left(A_j^p + \frac{\psi^p(Y_{t_j})}{2^p}\right)\mathbb{1}_{\{D_j \leq \psi(Y_{t_j})/2\}} \\
&\quad + 2^{2p-2}\left(\frac{1}{A_j^p} + \frac{2^p}{\psi^p(Y_{t_j})}\right)|A_j - D_j|^{2p}\mathbb{1}_{\{D_j > \psi(Y_{t_j})/2\}}
\end{aligned}$$

We take the expectation and apply Hölder inequality to obtain

$$\mathbb{E}\left[\left|\sqrt{A_j} - \sqrt{D_j}\right|^{2p}\right] \leq C(\epsilon_1 + \epsilon_2)$$

with

$$\epsilon_1 = \left(\mathbb{E}\left[\left(A_j^p + \frac{\psi^p(Y_{t_j})}{2^p}\right)^{1+\epsilon}\right]\right)^{\frac{1}{1+\epsilon}}\left(\mathbb{P}\left(D_j \leq \frac{\psi(Y_{t_j})}{2}\right)\right)^{\frac{\epsilon}{1+\epsilon}}$$

and

$$\epsilon_2 = \left(\mathbb{E}\left[\left(\frac{1}{A_j^p} + \frac{2^p}{\psi^p(Y_{t_j})}\right)^{1+\epsilon}\right]\right)^{\frac{1}{1+\epsilon}}\left(\mathbb{E}\left[|A_j - D_j|^{2p\frac{1+\epsilon}{\epsilon}}\right]\right)^{\frac{\epsilon}{1+\epsilon}}.$$

Let us begin with the second term. Following the estimation of $\overline{I}_2^j$ in the proof of Theorem 2, we show that

$$\begin{aligned}
\mathbb{E}\left[|A_j - D_j|^{2p\frac{1+\epsilon}{\epsilon}}\right] &= \mathbb{E}\left[\left|\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds - \left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right)\vee\underline{\psi}\right|^{2p\frac{1+\epsilon}{\epsilon}}\right] \\
&\leq C\delta_N^{2p\frac{1+\epsilon}{\epsilon}}
\end{aligned}$$

Thanks to assumption ($\mathcal{H}10$) and Jensen's inequality, we also have that

$$\begin{aligned}
\left(\mathbb{E}\left[\left(\frac{1}{A_j^p} + \frac{2^p}{\psi^p(Y_{t_j})}\right)^{1+\epsilon}\right]\right)^{\frac{1}{1+\epsilon}} &\leq 2^{\frac{\epsilon}{1+\epsilon}}\left(\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\mathbb{E}\left(\frac{1}{\psi^{p(1+\epsilon)}(Y_s)}\right)ds + 2^{p(1+\epsilon)}\mathbb{E}\left(\frac{1}{\psi^{p(1+\epsilon)}(Y_{t_j})}\right)\right)^{\frac{1}{1+\epsilon}} \\
&\leq C
\end{aligned}$$

Hence $\epsilon_2 \leq \frac{C}{N^{2p}}$. Now let us turn to ϵ_1. Note first that assumption $(\mathcal{H}10)$ enables us to show that there exists a positive constant C independent of N such that $\left(\mathbb{E}\left[\left(A_j^p + \frac{\psi^p(Y_{t_j})}{2^p}\right)^{1+\epsilon}\right]\right)^{\frac{1}{1+\epsilon}} \leq C$. Finally, what is left to prove is that $\mathbb{P}\left(D_j \leq \frac{\psi(Y_{t_j})}{2}\right) \leq \frac{C}{N^{2p\frac{1+\epsilon}{\epsilon}}}$. In fact, we can show that $\forall \alpha > 0, \exists C_\alpha > 0$ such that $\mathbb{P}\left(D_j \leq \frac{\psi(Y_{t_j})}{2}\right) \leq \frac{C_\alpha}{N^\alpha}$:

$$\begin{aligned}\mathbb{P}\left(D_j \leq \frac{\psi(Y_{t_j})}{2}\right) &\leq \mathbb{P}\left(\frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds \leq -\frac{\psi(Y_{t_j})}{2}\right) \\ &= \mathbb{P}\left(|G| \geq \frac{\sqrt{3}\psi(Y_{t_j})}{2\sqrt{\delta_N}\nu|\psi'(Y_{t_j})|}\right)\end{aligned}$$

where G is a centered reduced Gaussian random variable independent of Y_{t_j}.

Thanks to assumption $(\mathcal{H}10)$, $\exists C > 0$ s.t. $\mathbb{P}\left(|G| \geq \frac{\sqrt{3}\psi(Y_{t_j})}{2\sqrt{\delta_N}\nu|\psi'(Y_{t_j})|}\right) \leq 2\mathbb{P}\left(G \geq \frac{C}{\sqrt{\delta_N}}\right)$ and using the following standard upper bound of the Gaussian tail probability : $\forall t > 0, \mathbb{P}(G \geq t) \leq \frac{e^{-\frac{t^2}{2}}}{t\sqrt{2\pi}}$, we conclude. □

Remark 10 —

- *The fact that we can simulate exactly the volatility process without affecting the order of convergence of the scheme is yet another advantage of our approach over the Cruzeiro et al. [2004] scheme. On the other hand, the Kahl and Jäckel [2006] scheme allows the exact simulation of $(Y_t)_{t\in[0,T]}$. Applied to the SDE (3), it writes as*

$$\begin{aligned}X_{t_{k+1}}^{IJK} = X_{t_k}^{IJK} &+ \left(r - \frac{f^2(Y_{t_{k+1}}) + f^2(Y_{t_k})}{4}\right)\delta_N + \rho f(Y_{t_k})\Delta W_{k+1} \\ &+\sqrt{1-\rho^2}\frac{f(Y_{t_{k+1}}) + f(Y_{t_k})}{2}\Delta B_{k+1} + \frac{\rho\nu}{2}f'(Y_{t_k})\left((\Delta W_{k+1})^2 - \delta_N\right)\end{aligned} \tag{14}$$

  Note that it is close to our scheme insofar as it takes advantage of the structure of the SDE (for example, unlike the Cruzeiro et al. [2004] scheme, it allows the use of the coupling introduced in Remark 5). The main difference, which explains why our scheme has better weak trajectorial convergence order, is that we discretize more accurately the integral of $f(Y_t)$ with respect to the Brownian motion $(B_t)_{t\in[0,T]}$. If, instead of a trapezoidal method, one uses the same discretization as for the WeakTraj_1 scheme, then it can be shown that this modified IJK scheme will exhibit a first order weak trajectorial convergence.

- *One can easily check that this theorem applies for the Scott's [1987] model (and therefore for the Hull and White [1987] model) where we have $h(y) = r - \frac{\sigma_0^2 e^{2y}}{2} - \rho\sigma_0 e^y(\frac{\kappa}{\nu}(\theta - y) + \frac{\nu}{2})$ and $\psi(y) = \sigma_0^2 e^{2y}$. The Stein and Stein [1991] and the quadratic Gaussian models do not satisfy the assumption $|\psi'(y)| \leq C\psi(y)$.*

- *It is possible to improve the convergence at fixed times up to the order $\frac{3}{2}$. Following Lapeyre and Temam [2001] who approximate an integral of the form $\int_{t_k}^{t_{k+1}} g(Y_s)ds$ for a twice differentiable function g by $\delta_N g(Y_{t_k}) + \nu g'(Y_{t_k})\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds + (\kappa(\theta - Y_{t_k})g'(Y_{t_k}) + \frac{\nu^2}{2}g''(Y_{t_k}))\frac{\delta_N^2}{2}$, we obtain the following scheme*

  OU_Improved scheme

$$\widetilde{X}_{t_{k+1}}^N = \widetilde{X}_{t_k}^N + \rho\left(F(Y_{t_{k+1}}) - F(Y_{t_k})\right) + \widetilde{h}_k + \sqrt{1-\rho^2}\sqrt{\widetilde{\psi}_k}\ \Delta B_{k+1} \tag{15}$$

where $\widetilde{h}_k = \delta_N h(Y_{t_k}) + \nu h'(Y_{t_k}) \int_{t_k}^{t_{k+1}} (W_s - W_{t_k})ds + (\kappa(\theta - Y_{t_k})h'(Y_{t_k}) + \frac{\nu^2}{2}h''(Y_{t_k}))\frac{\delta_N^2}{2}$ *and* $\widetilde{\psi}_k = \left(\psi(Y_{t_k}) + \frac{\nu\psi'(Y_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds + (\kappa(\theta - Y_{t_k})\psi'(Y_{t_k}) + \frac{\nu^2}{2}\psi''(Y_{t_k}))\frac{\delta_N}{2}\right) \vee \underline{\psi}$.

Mimicking the proof of Theorem 2, one can show that

$$\max_{0\leq k\leq N} \mathbb{E}\left[\left|\widehat{X}_{t_k} - \widehat{X}^N_{t_{k+1}}\right|^2\right] = \mathcal{O}\left(N^{-3}\right)$$

where $\widehat{X}_{t_k}$ *and* $\widehat{X}^N_{t_{k+1}}$ *have respectively the same distribution as* X_{t_k} *and* $\widetilde{X}^N_{t_k}$ *:*

$$\widehat{X}_{t_k} = X_0 + \rho(F(Y_{t_k}) - F(y_0)) + \int_0^{t_k} h(Y_s)ds + \sqrt{1-\rho^2}\sqrt{\frac{1}{t_k}\int_0^{t_k}\psi(Y_s)ds}\; B_{t_k}$$

and

$$\widehat{X}^N_{t_k} = X_0 + \rho\left(F(Y_{t_k}) - F(y_0)\right) + \sum_{j=0}^{k-1}\widetilde{h}_j + \sqrt{1-\rho^2}\sqrt{\frac{\delta_N}{t_k}\sum_{j=0}^{k-1}\widetilde{\psi}_j}\; B_{t_k}.$$

As for the stock, we can prove the same convergence result under some additional assumptions which are more explicit than assumption ($\mathcal{H}7$) of Corollary 6. To do so, let us make the following changes in our scheme so that we can control its exponential moments :

$$\begin{aligned}\widetilde{X}^N_{t_{k+1}} &= \widetilde{X}^N_{t_k} + \rho\left(F(Y_{t_{k+1}}) - F(Y_{t_k})\right) + \delta_N h(Y_{t_k}) \\ &+\sqrt{1-\rho^2}\sqrt{\left(\psi(Y_{t_k}) + \frac{\nu\psi'(Y_{t_k})}{\delta_N}\int_{t_k}^{t_{k+1}}(W_s - W_{t_k})ds\right) \wedge \widehat{\psi}(Y_{t_k}) \vee \underline{\psi}}\;\Delta B_{k+1}\end{aligned} \tag{16}$$

Proposition 11 — *Suppose that* Y *is solution of (12) and that the scheme is defined by (16). Under the assumptions* ($\mathcal{H}8$)*,* ($\mathcal{H}9$) *and* ($\mathcal{H}10$) *of Theorem 9 and if*

(*$\mathcal{H}$*11) *there exists* $\beta \in (0,1)$ *and* $K > 0$ *such that* $\forall y \in \mathbb{R}$

$$\begin{aligned}|h(y)| + |F(y)| + |f'(y)| &\leq K(1 + |y|^{1+\beta}) \\ |f(y)| &\leq K(1 + |y|^{\beta})\end{aligned}$$

then, $\forall p \geq 1$*, there exists a positive constant* C *independent of* N *such that*

$$\mathbb{E}\left[\max_{0\leq k\leq N}\left|e^{\widetilde{X}_{t_k}} - e^{\widetilde{X}^N_{t_k}}\right|^{2p}\right] \leq \frac{C}{N^{2p}}.$$

The same result holds true if one replaces assumption ($\mathcal{H}10$) *by assumption* ($\mathcal{H}2$) *together with the assumption that* $\exists C > 0$ *for which* $\forall y \in \mathbb{R}, |\psi'(y)| \leq C\psi(y)$.

Proof : We go over the proof of Corollary 6. The fact that $\mathbb{E}\left[\max_{0\leq k\leq N}\left|\widetilde{X}_{t_k} - \widetilde{X}^N_{t_k}\right|^{4p}\right] = \mathcal{O}(\frac{1}{N^{4p}})$ is not a straightforward consequence of Theorem 9 anymore because we have introduced some changes in

our scheme. However, looking through the proof of the theorem, one can see that it is enough to prove the following inequality : $\forall j \in \{0, \dots, N-1\}$

$$\mathbb{E}\left[\left|\sqrt{\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds} - \sqrt{\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right) \wedge \widehat{\psi}(Y_{t_j}) \vee \underline{\psi}}\right|^{2p}\right] \leq \frac{C}{N^{2p}} \tag{17}$$

When $\overline{\psi}$ is finite, since $\frac{1}{\delta_N}\int_{t_j}^{t_{j+1}}\psi(Y_s)ds$ is smaller than $\widehat{\psi}(Y_{t_k}) = \overline{\psi}$, we can remove the new cut-off from the left hand side of (17) and then proceed like in Theorem 9. When $\overline{\psi} = +\infty$, on the event $\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right) \leq \widehat{\psi}(Y_{t_j})$, we recover our original scheme and we prove (17) like in Theorem 9. Then, using the Gaussian arguments developed in the end of the proof of Theorem 9, we control the probability of the complementary event to conclude.

Now, what is left to prove is that assumption ($\mathcal{H}7$) is satisfied. On the one hand, we have that

$$\begin{aligned}
\mathbb{E}\left[\max_{0\leq k\leq N} S_{t_k}^{4p}\right] &= \mathbb{E}\left[\max_{0\leq k\leq N}\left(S_0 + \int_0^{t_k} rS_s ds + \int_0^{t_k} f(Y_s)S_s\left(\rho dW_s + \sqrt{1-\rho^2}dB_s\right)\right)^{4p}\right] \\
&\leq C\left(1 + \int_0^T \mathbb{E}\left(S_t^{4p}(1 + f^{4p}(Y_t))\right)dt\right) \\
&\leq C\left(1 + \int_0^T \sqrt{\mathbb{E}(S_t^{8p})}\sqrt{\mathbb{E}\left((1+f^{4p}(Y_t))^2\right)}dt\right)
\end{aligned}$$

Thanks to assumption ($\mathcal{H}11$) and Lemma 8, there exists $C > 0$ such that $\sqrt{\mathbb{E}\left((1+f^{4p}(Y_t))^2\right)} \leq C$. Observe that conditionally on $(Y_t)_{t\in[0,T]}$,

$$X_t \sim \mathcal{N}\left(\log(s_0) + \rho(F(Y_t) - F(y_0)) + \int_0^t h(Y_s)ds\,,\,(1-\rho^2)\int_0^t f^2(Y_s)ds\right) \tag{18}$$

so, by Jensen's inequality and assumption ($\mathcal{H}11$)

$$\begin{aligned}
\mathbb{E}\left(S_t^{8p}\right) &= \mathbb{E}\left(e^{8p\left(\log(s_0)+\rho(F(Y_t)-F(y_0))+\int_0^t h(Y_s)ds\right)}e^{32p^2(1-\rho^2)\int_0^t f^2(Y_s)ds}\right) \\
&\leq \mathbb{E}\left(e^{8p(\log(s_0)+\rho(F(Y_t)-F(y_0)))}\frac{1}{t}\int_0^t e^{t\left(8ph(Y_s)+32p^2(1-\rho^2)f^2(Y_s)\right)}ds\right) \\
&\leq C\mathbb{E}\left(e^{C\sup_{0\leq t\leq T}|Y_t|^{1+\beta}}\right)
\end{aligned}$$

Using Lemma 8, we deduce that $\mathbb{E}\left[\max_{0\leq k\leq N} S_{t_k}^{4p}\right] < \infty$.

On the other hand, using Cauchy-Schwartz inequality, we have that

$$\begin{aligned}
\mathbb{E}\left[\max_{0\leq k\leq N} e^{4p\widetilde{X}_{t_k}^N}\right] &= \mathbb{E}\left[\max_{0\leq k\leq N}\exp\left(4p\left(X_0 + \rho(F(Y_{t_k}) - F(y_0)) + \sum_{j=0}^{k-1}\delta_N h(Y_{t_j}) + \sum_{j=0}^{k-1}\sqrt{1-\rho^2}\right.\right.\right. \\
&\qquad \left.\left.\left.\times\sqrt{\left(\psi(Y_{t_j}) + \frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right) \wedge \widehat{\psi}(Y_{t_j}) \vee \underline{\psi}}\,\Delta B_{j+1}\right)\right)\right] \\
&\leq \sqrt{\widetilde{E}_1^N}\sqrt{\widetilde{E}_2^N}
\end{aligned}$$

where

$$\widetilde{E}_1^N = \mathbb{E}\left[\max_{0\leq k\leq N} e^{8p\left(X_0+\rho(F(Y_{t_k})-F(y_0))+\sum_{j=0}^{k-1}\delta_N h(Y_{t_j})\right)}\right]$$

and

$$\widetilde{E}_2^N = \mathbb{E}\left[\max_{0\le k\le N} e^{8p\sqrt{1-\rho^2}\sum_{j=0}^{k-1}\sqrt{\left(\psi(Y_{t_j})+\frac{\nu\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s-W_{t_j})ds\right)\wedge\widehat{\psi}(Y_{t_j})\vee\underline{\psi}}\Delta B_{j+1}}\right].$$

Using the same argument as before, we show that $\widetilde{E}_1^N \le C\mathbb{E}\left(e^{C\sup_{0\le t\le T}|Y_t|^{1+\beta}}\right) < \infty$.

Denote by $D_j = \left(\psi(Y_{t_j}) + \frac{\sigma\psi'(Y_{t_j})}{\delta_N}\int_{t_j}^{t_{j+1}}(W_s - W_{t_j})ds\right) \wedge \widehat{\psi}(Y_{t_j}) \vee \underline{\psi}$. Using Doob's maximal inequality for the positive submartingale $\left(e^{4p\sqrt{1-\rho^2}\sum_{j=0}^{k-1}\sqrt{D_j}\Delta B_{j+1}}\right)_{0\le k\le N}$ (see Theorem 3.8 p. 13 of Karatzas and Shreve [1991] for example), we also have that

$$\begin{aligned}
\widetilde{E}_2^N &\le 4\mathbb{E}\left(e^{8p\sqrt{1-\rho^2}\sum_{j=0}^{N-1}\sqrt{D_j}\,\Delta B_{j+1}}\right)\\
&= 4\mathbb{E}\left(\prod_{j=0}^{N-1} e^{32p^2\delta_N(1-\rho^2)D_j}\right)\\
&\le 4\mathbb{E}\left(\max_{0\le k\le N-1} e^{32p^2(1-\rho^2)\widehat{\psi}(Y_{t_j})}\right)
\end{aligned}$$

By virtue of assumption $(\mathcal{H}11)$, $\widetilde{E}_2^N < \infty$ which concludes the proof. □

2 A second order weak scheme

Integrating the first stochastic differential equation in (4) gives

$$X_t = \log(s_0) + \rho(F(Y_t) - F(y_0)) + \int_0^t h(Y_s)ds + \sqrt{1-\rho^2}\int_0^t f(Y_s)dB_s \tag{19}$$

We are only left with an integral with respect to time which can be handled by the use of a trapezoidal scheme and a stochastic integral where the integrand is independent of the Brownian motion. Hence, conditionally on $(Y_t)_{t\in[0,T]}$,

$$X_T \sim \mathcal{N}\left(\log(s_0) + \rho(F(Y_T) - F(y_0)) + m_T\,,\,(1-\rho^2)v_T\right) \tag{20}$$

where $m_T = \int_0^T h(Y_s)ds$ and $v_T = \int_0^T f^2(Y_s)ds$. This suggests that, in order to properly approximate the law of X_T, one should accurately approximate the law of Y_T and carefully handle integrals with respect to time of functions of the process $(Y_t)_{t\in[0,T]}$. We thus define our weak scheme as follows

Weak_2 scheme

$$\overline{X}_T^N = \log(s_0) + \rho(F(\overline{Y}_T^N) - F(y_0)) + \overline{m}_T^N + \sqrt{(1-\rho^2)\overline{v}_T^N}G \tag{21}$$

where $\overline{m}_T^N = \delta_N\sum_{k=0}^{N-1}\frac{h(\overline{Y}_{t_k}^N)+h(\overline{Y}_{t_{k+1}}^N)}{2}$, $\overline{v}_T^N = \delta_N\sum_{k=0}^{N-1}\frac{f^2(\overline{Y}_{t_k}^N)+f^2(\overline{Y}_{t_{k+1}}^N)}{2}$, $(\overline{Y}_{t_k}^N)_{0\le k\le N}$ is the Ninomiya-Victoir scheme of $(Y_t)_{t\in[0,T]}$ and G is an independent centered reduced Gaussian random variable. Note that, conditionally on $(\overline{Y}_{t_k}^N)_{0\le k\le N}$, $\overline{X}_t^N$ is also a Gaussian random variable with mean $\log(s_0) + \rho(F(\overline{Y}_T^N) - F(y_0)) + \overline{m}_T^N$ and variance $(1-\rho^2)\overline{v}_T^N$.

It is well known that the Ninomiya and Victoir [2008] scheme is of weak order two. For the sake of completeness, we give its definition in our setting :

$$\left\{ \begin{array}{l} \overline{Y}_0^N = y_0 \\ \forall 0 \leq k \leq N-1, \overline{Y}_{t_{k+1}}^N = \exp\left(\frac{T}{2N} V_0\right) \exp\left((W_{t_{k+1}} - W_{t_k}) V\right) \exp\left(\frac{T}{2N} V_0\right) (\overline{Y}_{t_k}^N) \end{array} \right.$$

where $V_0 : x \mapsto b(x) - \frac{1}{2}\sigma\sigma'(x)$ and $V : x \mapsto \sigma(x)$. The notation $\exp(tV)(x)$ stands for the solution, at time t and starting from x, of the ODE $\eta'(t) = V(\eta(t))$. What is nice with our setting is that we are in dimension one and thus such ODEs can be solved explicitly. Indeed, if ζ is a primitive of $\frac{1}{V}$: $\zeta(t) = \int_0^t \frac{1}{V(s)} ds$, then the solution writes as $\eta(t) = \zeta^{-1}\left(t + \zeta(x)\right)$.

Note that our scheme can be seen as a splitting scheme for the SDE satisfied by $(Z_t = X_t - \rho F(Y_t), Y_t)$:

$$\left\{ \begin{array}{rcl} dZ_t &=& h(Y_t)dt + \sqrt{1-\rho^2} f(Y_t) dB_t \\ dY_t &=& b(Y_t)dt + \sigma(Y_t) dW_t \end{array} \right. \tag{22}$$

The differential operator associated to (22) writes as

$$\mathcal{L}v(z,y) = h(y)\frac{\partial v}{\partial z} + b(y)\frac{\partial v}{\partial y} + \frac{\sigma^2(y)}{2}\frac{\partial^2 v}{\partial y^2} + \frac{(1-\rho^2)}{2} f^2(y)\frac{\partial^2 v}{\partial z^2} = \mathcal{L}_Y v(z,y) + \mathcal{L}_Z v(z,y)$$

where $\mathcal{L}_Y v(z,y) = b(y)\frac{\partial v}{\partial y} + \frac{\sigma^2(y)}{2}\frac{\partial^2 v}{\partial y^2}$ and $\mathcal{L}_Z v(z,y) = h(y)\frac{\partial v}{\partial z} + \frac{(1-\rho^2)}{2} f^2(y)\frac{\partial^2 v}{\partial z^2}$. One can check that our scheme amounts to first integrate exactly $\mathcal{L}_Z$ over a half time step then apply the Ninomiya-Victoir scheme to $\mathcal{L}_Y$ over a time step and finally integrate exactly $\mathcal{L}_Z$ over a half time step. According to results on splitting (see Alfonsi [2009] or Tanaka and Kohatsu-Higa [2009] for example) one expects this scheme to exhibit second order weak convergence. We will not use this point of view to prove our convergence result stated in the next theorem, since we need to apply test functions with exponential growth to X_T to be able to analyse weak convergence of the stock price.

Theorem 12 — *Suppose that $\rho \in (-1,1)$. If the following assumptions hold*

(H12) *b and σ are respectively $\mathcal{C}^4$ and $\mathcal{C}^5$, with bounded derivatives of any order greater or equal to 1.*

(H13) *h and f are $\mathcal{C}^4$ and F is $\mathcal{C}^6$. The three functions are bounded together with all their derivatives.*

(H14) *$\underline{\psi} > 0$*

then, for any measurable function g verifying $\exists c \geq 0, \mu \in [0,2)$ such that $\forall x \in \mathbb{R}, |g(x)| \leq c e^{|x|^\mu}$, there exists $C > 0$ such that

$$\left| \mathbb{E}\Big(g(X_T)\Big) - \mathbb{E}\left(g(\overline{X}_T^N)\right) \right| \leq \frac{C}{N^2}$$

In terms of the asset price, we easily deduce the following corollary :

Corollary 13 — *Under the assumptions of Theorem 12, for any measurable function α verifying $\exists c \geq 0, \mu \in [0,2)$ such that $\forall y > 0, |\alpha(y)| \leq c e^{|\log(y)|^\mu}$, there exists $C > 0$ such that*

$$\left| \mathbb{E}\left(\alpha(S_T)\right) - \mathbb{E}\left(\alpha(e^{\overline{X}_T^N})\right) \right| \leq \frac{C}{N^2}$$

Proof of the theorem : The idea of the proof consists in conditioning by the Brownian motion which drives the volatility process and then applying the weak error analysis of Talay and Tubaro [1990].

As stated above, conditionally on $(W_t)_{t\in[0,T]}$, both X_T and $\overline{X}_T^N$ are Gaussian random variables and one can easily show that

$$\begin{aligned}\epsilon &:= \left|\mathbb{E}\left[g(X_T) - g(\overline{X}_T^N)\right]\right| \\ &= \left|\int_{\mathbb{R}} g(x)\mathbb{E}\left[\frac{\exp\left(-\frac{(x-\log(s_0)+\rho F(y_0)-\rho F(Y_T)-m_T)^2}{2(1-\rho^2)v_T}\right)}{\sqrt{2\pi(1-\rho^2)v_T}} - \frac{\exp\left(-\frac{(x-\log(s_0)+\rho F(y_0)-\rho F(\overline{Y}_T^N)-\overline{m}_T^N)^2}{2(1-\rho^2)\overline{v}_T^N}\right)}{\sqrt{2\pi(1-\rho^2)\overline{v}_T^N}}\right] dx\right|\end{aligned}$$

For $x \in \mathbb{R}$, denote by γ_x the function

$$\begin{aligned}\gamma_x : \mathbb{R}\times\mathbb{R}\times\mathbb{R}_+^* &\to \mathbb{R} \\ (y,m,v) &\mapsto \frac{\exp\left(-\frac{(x-\log(s_0)+\rho F(y_0)-\rho F(y)-m)^2}{2(1-\rho^2)v}\right)}{\sqrt{2\pi(1-\rho^2)v}}\end{aligned}$$

so that $\epsilon \le \int_{\mathbb{R}} g(x)\left|\mathbb{E}\left[\gamma_x(Y_T,m_T,v_T) - \gamma_x(\overline{Y}_T^N,\overline{m}_T^N,\overline{v}_T^N)\right]\right| dx$. Consequently, it is enough to show the following intermediate result : $\exists C, K > 0$ and $p \in \mathbb{N}$ such that

$$\forall x \in \mathbb{R}, \left|\mathbb{E}\left[\gamma_x(Y_T,m_T,v_T) - \gamma_x(\overline{Y}_T^N,\overline{m}_T^N,\overline{v}_T^N)\right]\right| \le \frac{C}{N^2}e^{-Kx^2}(1+|x|^p). \tag{23}$$

We naturally consider the following 3-dimensional degenerate SDE:

$$\left\{\begin{array}{rcl} dY_t &=& \sigma(Y_t)dW_t + b(Y_t)dt; \quad Y_0 = y_0 \\ dm_t &=& h(Y_t)dt; \quad m_0 = 0 \\ dv_t &=& f^2(Y_t)dt; \quad v_0 = 0 \end{array}\right. \tag{24}$$

Note that $(\overline{Y}_T^N,\overline{m}_T^N,\overline{v}_T^N)$ is close to the terminal value of the Ninomiya-Victoir scheme applied to this 3-dimensional SDE. In order to prove (23), we need to analyse the dependence of the error on x and not only on N. That is why we resume the error analysis of Ninomiya and Victoir [2008] in a more detailed fashion.

For $x \in \mathbb{R}$, let us define the function $u_x : [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}_+^* \to \mathbb{R}$ by

$$u_x(t,y,m,v) = \mathbb{E}\left[\gamma_x\left((Y_{T-t},m_{T-t},v_{T-t})^{(y,m,v)}\right)\right]$$

where we denote by $(Y_{T-t},m_{T-t},v_{T-t})^{(y,m,v)}$ the solution at time $T-t$ of (24) starting from (y,m,v).

The remainder of the proof leans on the following lemmas. We will use the standard notation for partial derivatives: for a multi-index $\alpha = (\alpha_1,\ldots,\alpha_d) \in \mathbb{N}^d$, d being a positive integer, we denote by $|\alpha| = \alpha_1 + \cdots + \alpha_d$ its length and by ∂_α the differential operator $\partial^{|\alpha|}/\partial_1^{\alpha_1}\ldots\partial_d^{\alpha_d}$.

Lemma 14 — *Under assumptions (H12), (H13) and (H14), we have that*

i) u_x is $\mathcal{C}^3$ with respect to the time variable and $\mathcal{C}^6$ with respect to the space variable. Moreover, it solves the following PDE

$$\left\{\begin{array}{rcl} \partial_t u_x + \mathcal{L}u_x &=& 0 \\ u_x(T,y,m,v) &=& \gamma_x(y,m,v) \end{array}\right. \tag{25}$$

where $\mathcal{L}$ is the differential operator associated to (24):

$$\mathcal{L}u(y,m,v) = \frac{\sigma^2(y)}{2}\frac{\partial^2 u}{\partial y^2} + b(y)\frac{\partial u}{\partial y} + h(y)\frac{\partial u}{\partial m} + f^2(y)\frac{\partial u}{\partial v}.$$

ii) For any multi-index $\alpha \in \mathbb{N}^3$ and integer l such that $2l + |\alpha| \leq 6$, there exists $C_{l,\alpha}, K_{l,\alpha} > 0$ and $(p_{l,\alpha}, q_{l,\alpha}) \in \mathbb{N}^2$ such that

$$\forall (t,y,m,v) \in [0,T] \times \mathcal{D}_t, \quad \left|\partial_t^l \partial_\alpha u_x(t,y,m,v)\right| \leq C_{l,\alpha} e^{-K_{l,\alpha} x^2} \left(1 + |x|^{p_{l,\alpha}}\right)\left(1 + |y|^{q_{l,\alpha}}\right)$$

where $\mathcal{D}_t$ is the set $\mathbb{R} \times [-t \sup_{z\in\mathbb{R}} |h(z)|, t \sup_{z\in\mathbb{R}} |h(z)|] \times [t\underline{\psi}, t\overline{\psi}]$. Note that $\underline{\psi}$ and $\overline{\psi}$ are finite by virtue of assumptions ($\mathcal{H}13$) and ($\mathcal{H}14$).

Lemma 15 — *Under assumption ($\mathcal{H}12$),*

$$\forall q \in \mathbb{N}, \sup_{0\leq k\leq N} \mathbb{E}\left(\left|\overline{Y}^N_{t_k}\right|^q\right) < \infty$$

Now, following the error analysis of Talay and Tubaro [1990], we write that

$$\left|\mathbb{E}\left[\gamma_x(Y_T, m_T, v_T) - \gamma_x(\overline{Y}^N_T, \overline{m}^N_T, \overline{v}^N_T)\right]\right| \leq \sum_{k=0}^{N-1} \eta_k(x)$$

where $\eta_k(x) = \left|\mathbb{E}\left[u_x(t_{k+1}, \overline{Y}^N_{t_{k+1}}, \overline{m}^N_{t_{k+1}}, \overline{v}^N_{t_{k+1}}) - u_x(t_k, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k})\right]\right|$ and $\forall 0 \leq k \leq N$, $\overline{m}^N_{t_k} = \delta_N \sum_{j=0}^{k-1} \frac{h(\overline{Y}^N_{t_j}) + h(\overline{Y}^N_{t_{j+1}})}{2}$ and $\overline{v}^N_{t_k} = \delta_N \sum_{j=0}^{k-1} \frac{f^2(\overline{Y}^N_{t_j}) + f^2(\overline{Y}^N_{t_{j+1}})}{2}$. Using the Markov property for the first term in the expectation and Taylor's formula together with PDE (25) for the second, we get

$$\begin{aligned}
\eta_k(x) &= \Big|\mathbb{E}\Big[\phi_x(t_{k+1}, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k}) - u_x(t_{k+1}, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k}) - \delta_N \mathcal{L} u_x(t_{k+1}, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k}) \\
&\quad - \frac{\delta_N^2}{2} \mathcal{L}^2 u_x(t_{k+1}, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k}) + \frac{1}{2}\int_{t_k}^{t_{k+1}} \frac{\partial^3 u_x}{\partial t^3}(t, \overline{Y}^N_{t_k}, \overline{m}^N_{t_k}, \overline{v}^N_{t_k})(t - t_k)^2 dt\Big]\Big|
\end{aligned}$$

where

$$\phi_x(t_{k+1}, y, m, v) = \mathbb{E}\left[u_x(t_{k+1}, \overline{Y}^{N,y}_{t_1}, m + \delta_N \frac{h(\overline{Y}^{N,y}_{t_1}) + h(y)}{2}, v + \delta_N \frac{f^2(\overline{Y}^{N,y}_{t_1}) + f^2(y)}{2})\right]$$

Denote by Γ_y the function $z \mapsto u_x(t_{k+1}, z, m + \delta_N \frac{h(z)+h(y)}{2}, v + \delta_N \frac{f^2(z)+f^2(y)}{2})$. Using Taylor's formula we can show that $\forall z \in \mathbb{R}$,

$$\Gamma_y(z) = \Gamma_{y,1}(z) + \delta_N \Gamma_{y,2}(z) + \frac{\delta_N^2}{2}\Gamma_{y,3}(z) + R_0(z)$$

where

$$\begin{aligned}
\Gamma_{y,1}(z) &= u_x(t_{k+1}, z, m, v) \\
\Gamma_{y,2}(z) &= \frac{h(z) + h(y)}{2}\frac{\partial u_x}{\partial m}(t_{k+1}, z, m, v) + \frac{f^2(z) + f^2(y)}{2}\frac{\partial u_x}{\partial v}(t_{k+1}, z, m, v) \\
\Gamma_{y,3}(z) &= \left(\frac{h(z) + h(y)}{2}\right)^2 \frac{\partial^2 u_x}{\partial m^2}(t_{k+1}, z, m, v) + \left(\frac{f^2(z) + f^2(y)}{2}\right)^2 \frac{\partial^2 u_x}{\partial v^2}(t_{k+1}, z, m, v) \\
&\quad + 2\frac{h(z) + h(y)}{2}\frac{f^2(z) + f^2(y)}{2}\frac{\partial^2 u_x}{\partial m \partial v}(t_{k+1}, z, m, v)
\end{aligned}$$

and

$$
\begin{aligned}
R_0(z) = \int_0^{\delta_N} \tfrac{(\delta_N - t)^2}{2} dt \Bigg(& \left(\tfrac{h(z)+h(y)}{2}\right)^3 \tfrac{\partial^3 u_x}{\partial m^3}\left(t_{k+1}, z, m + t\tfrac{h(z)+h(y)}{2}, v + t\tfrac{f^2(z)+f^2(y)}{2}\right) \\
& + \left(\tfrac{f^2(z)+f^2(y)}{2}\right)^3 \tfrac{\partial^3 u_x}{\partial v^3}\left(t_{k+1}, z, m + t\tfrac{h(z)+h(y)}{2}, v + t\tfrac{f^2(z)+f^2(y)}{2}\right) \\
& + 3\left(\tfrac{f^2(z)+f^2(y)}{2}\right)^2 \left(\tfrac{h(z)+h(y)}{2}\right) \tfrac{\partial^3 u_x}{\partial m \partial v^2}\left(t_{k+1}, z, m + t\tfrac{h(z)+h(y)}{2}, v + t\tfrac{f^2(z)+f^2(y)}{2}\right) \\
& + 3\left(\tfrac{h(z)+h(y)}{2}\right)^2 \left(\tfrac{f^2(z)+f^2(y)}{2}\right) \tfrac{\partial^3 u_x}{\partial m^2 \partial v}\left(t_{k+1}, z, m + t\tfrac{h(z)+h(y)}{2}, v + t\tfrac{f^2(z)+f^2(y)}{2}\right)\Bigg)
\end{aligned} \tag{26}
$$

So,

$$
\phi_x(t_{k+1}, y, m, v) = \underbrace{\mathbb{E}\left[\Gamma_{y,1}(\overline{Y}^{N,y}_{t_1})\right]}_{\phi_{x,1}(t_{k+1},y,m,v)} + \underbrace{\delta_N \mathbb{E}\left[\Gamma_{y,2}(\overline{Y}^{N,y}_{t_1})\right]}_{\phi_{x,2}(t_{k+1},y,m,v)} + \underbrace{\frac{\delta_N^2}{2}\mathbb{E}\left[\Gamma_{y,3}(\overline{Y}^{N,y}_{t_1})\right]}_{\phi_{x,3}(t_{k+1},y,m,v)} + \mathbb{E}\left[R_0(\overline{Y}^{N,y}_{t_1})\right] \tag{27}
$$

With a slight abuse of notations, we define the first order differential operators V_0 and V acting on $\mathcal{C}^1$ functions by $V_0\xi(x) = V_0(x)\xi'(x)$ and $V\xi(x) = V(x)\xi'(x)$ for $\xi \in \mathcal{C}^1(\mathbb{R})$. We make the same expansions as in Ninomiya and Victoir [2008] but with making the remainder terms explicit in order to check if they have the good behavior with respect to x. We can show after tedious but simple computations that

$$
\begin{aligned}
\phi_{x,1}(t_{k+1}, y, m, v) &= \Gamma_{y,1}(y) + \frac{\delta_N}{2}\left(V^2\Gamma_{y,1}(y) + 2V_0\Gamma_{y,1}(y)\right) \\
&\quad + \frac{\delta_N^2}{8}\left({V_0}^2\Gamma_{y,1}(y) + 2V_0V^2\Gamma_{y,1}(y) + 2V^2V_0\Gamma_{y,1}(y) + V^4\Gamma_{y,1}(y)\right) + \mathbb{E}\left(R_1(y)\right) \\
\phi_{x,2}(t_{k+1}, y, m, v) &= \delta_N\Gamma_{y,2}(y) + \frac{\delta_N^2}{2}\left(V^2\Gamma_{y,2}(y) + 2V_0\Gamma_{y,2}(y)\right) + \mathbb{E}\left(R_2(y)\right) \\
\phi_{x,3}(t_{k+1}, y, m, v) &= \frac{\delta_N^2}{2}\Gamma_{y,3}(y) + \mathbb{E}\left(R_3(y)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
R_1(y) &= \int_0^{\frac{\delta_N}{2}}\int_0^{s_1}\int_0^{s_2} {V_0}^3\Gamma_{y,1}(e^{s_3V_0}e^{W_{\delta_N}V}e^{\frac{\delta_N}{2}V_0}(y))ds_3ds_2ds_1 \\
&+ \int_0^{W_{\delta_N}}\int_0^{s_1}\int_0^{s_2}\int_0^{s_3}\int_0^{s_4}\int_0^{s_5} V^6\Gamma_{y,1}(e^{s_6V}e^{\frac{\delta_N}{2}V_0}(y))ds_6ds_5ds_4ds_3ds_2ds_1 \\
&+ \tfrac{\delta_N}{2}\int_0^{W_{\delta_N}}\int_0^{s_1}\int_0^{s_2}\int_0^{s_3} V^4V_0\Gamma_{y,1}(e^{s_4V}e^{\frac{\delta_N}{2}V_0}(y))ds_4ds_3ds_2ds_1 \\
&+ \tfrac{\delta_N^2}{8}\int_0^{W_{\delta_N}}\int_0^{s_1} V^2{V_0}^2\Gamma_{y,1}(e^{s_2V}e^{\frac{\delta_N}{2}V_0}(y))ds_2ds_1 \\
&+ \int_0^{\frac{\delta_N}{2}}\int_0^{s_1}\int_0^{s_2} {V_0}^3\Gamma_{y,1}(e^{s_3V_0}(y))ds_3ds_2ds_1 + \tfrac{\delta_N}{2}\int_0^{\frac{\delta_N}{2}}\int_0^{s_1} {V_0}^2V^2\Gamma_{y,1}(e^{s_2V_0}(y))ds_2ds_1 \\
&+ \tfrac{\delta_N^2}{8}\int_0^{\frac{\delta_N}{2}} V_0V^4\Gamma_{y,1}(e^{s_1V_0}(y))ds_1 + \tfrac{\delta_N}{2}\int_0^{\frac{\delta_N}{2}}\int_0^{s_1} {V_0}^3\Gamma_{y,1}(e^{s_2V_0}(y))ds_2ds_1 \\
&+ \tfrac{\delta_N^2}{4}\int_0^{\frac{\delta_N}{2}} V_0V^2V_0\Gamma_{y,1}(e^{s_1V_0}(y))ds_1 + \tfrac{\delta_N^2}{8}\int_0^{\frac{\delta_N}{2}} {V_0}^3\Gamma_{y,1}(e^{s_1V_0}(y))ds_1
\end{aligned} \tag{28}
$$

$$
\begin{aligned}
R_2(y) &= \delta_N\Bigg(\int_0^{\frac{\delta_N}{2}}\int_0^{s_1} {V_0}^2\Gamma_{y,2}(e^{s_2V_0}e^{W_{\delta_N}V}e^{\frac{\delta_N}{2}V_0}(y))ds_2ds_1 \\
&+ \int_0^{W_{\delta_N}}\int_0^{s_1}\int_0^{s_2}\int_0^{s_3} V^4\Gamma_{y,2}(e^{s_4V}e^{\frac{\delta_N}{2}V_0}(y))ds_4ds_3ds_2ds_1 \\
&+ \tfrac{\delta_N}{2}\int_0^{W_{\delta_N}}\int_0^{s_1} V^2V_0\Gamma_{y,2}(e^{s_2V}e^{\frac{\delta_N}{2}V_0}(y))ds_2ds_1 + \int_0^{\frac{\delta_N}{2}}\int_0^{s_1} {V_0}^2\Gamma_{y,2}(e^{s_2V_0}(y))ds_2ds_1 \\
&+ \tfrac{\delta_N}{2}\int_0^{\frac{\delta_N}{2}} V_0V^2\Gamma_{y,2}(e^{s_1V_0}(y))ds_1 + \tfrac{\delta_N}{2}\int_0^{\frac{\delta_N}{2}} {V_0}^2\Gamma_{y,2}(e^{s_1V_0}(y))ds_1\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
R_3(y) &= \tfrac{\delta_N^2}{2}\Bigg(\int_0^{\frac{\delta_N}{2}} V_0\Gamma_{y,3}(e^{s_1V_0}e^{W_{\delta_N}V}e^{\frac{\delta_N}{2}V_0}(y))ds_1 + \int_0^{W_{\delta_N}}\int_0^{s_1} V^2\Gamma_{y,3}(e^{s_2V}e^{\frac{\delta_N}{2}V_0}(y))ds_2ds_1 \\
&+ \int_0^{\frac{\delta_N}{2}} V_0\Gamma_{y,3}(e^{s_1V_0}(y))ds_1\Bigg)
\end{aligned}
$$

Putting all the terms together, one can check that

$$\phi_x(t_{k+1},y,m,v) = u_x(t_{k+1},y,m,v) + \delta_N\mathcal{L}u_x(t_{k+1},y,m,v) + \frac{\delta_N^2}{2}\mathcal{L}^2u_x(t_{k+1},y,m,v) + R(y)$$

where $R(y) = \mathbb{E}\left[R_0(\overline{Y}_{t_1}^{N,y}) + R_1(y) + R_2(y) + R_3(y)\right]$. Finally,

$$\left|\mathbb{E}\left[\gamma_x(Y_T,m_T,v_T) - \gamma_x(\overline{Y}_T^N,\overline{m}_T^N,\overline{v}_T^N)\right]\right| \le \sum_{k=0}^{N-1}\mathbb{E}\left[\left|\frac{1}{2}\int_{t_k}^{t_{k+1}}\frac{\partial^3u_x}{\partial t^3}(t,\overline{Y}_{t_k}^N,\overline{m}_{t_k}^N,\overline{v}_{t_k}^N)(t-t_k)^2dt\right| + \left|R(\overline{Y}_{t_k}^N)\right|\right]$$

From Lemmas 14 and 15, we deduce that there exists $C_1, K_1 > 0$ and $p_1 \in \mathbb{N}$ such that

$$\sum_{k=0}^{N-1}\mathbb{E}\left[\left|\frac{1}{2}\int_{t_k}^{t_{k+1}}\frac{\partial^3u_x}{\partial t^3}(t,\overline{Y}_{t_k}^N,\overline{m}_{t_k}^N,\overline{v}_{t_k}^N)(t-t_k)^2dt\right|\right] \le \frac{1}{N^2}C_1e^{-K_1x^2}(1+|x|^{p_1}) \tag{29}$$

On the other hand, a close look to (26) and (28) convinces us that the term $\mathbb{E}\left[\left|R(\overline{Y}_{t_k}^N)\right|\right]$ is of order $\frac{1}{N^3}$ and that it involves only derivatives of u_x and of the coefficients of the SDE (24). So, thanks Lemmas 14 and 15, there exists $C_2, K_2 > 0$ and $p_2 \in \mathbb{N}$ such that

$$\sum_{k=0}^{N-1}\mathbb{E}\left[\left|R(\overline{Y}_{t_k}^N)\right|\right] \le \frac{1}{N^2}C_2e^{-K_2x^2}(1+|x|^{p_2}) \tag{30}$$

From (29) and (30) we deduce the desired result (23) to conclude.

□

Remark 16 —

- *The theorem does not cover the case of perfectly correlated or uncorrelated stock and volatility which is not very interesting from a practical point of view.*
- *As for plain vanilla options pricing, observe that, by the Romano and Touzi [1997] formula,*

$$\mathbb{E}\left(e^{-rT}\alpha(S_T)|(Y_t)_{t\in[0,T]}\right) = BS_{\alpha,T}\left(s_0 e^{\rho(F(Y_T)-F(y_0))+m_T+(\frac{(1-\rho^2)v_T}{2T}-r)T}, \frac{(1-\rho^2)v_T}{T}\right)$$

where $BS_{\alpha,T}(s,v)$ stands for the price of a European option with pay-off α and maturity T in the Black & Scholes model with initial stock price s, volatility $\sqrt{v}$ and constant interest rate r. When, like for a call or a put option, $BS_{\alpha,T}$ is available in a closed form, one should approximate $\mathbb{E}\left(e^{-rT}\alpha(S_T)\right)$ by

$$\frac{1}{M}\sum_{i=1}^{M} BS_{\alpha,T}\left(s_0 e^{\rho(F(\overline{Y}_T^{N,i})-F(y_0))+\overline{m}_T^{N,i}+(\frac{(1-\rho^2)\overline{v}_T^{N,i}}{2T}-r)T}, \frac{(1-\rho^2)\overline{v}_T^{N,i}}{T}\right)$$

where M is the total number of Monte Carlo samples and the index i refers to independent draws.

Indeed, the conditioning provides a variance reduction. We also note that what is most important is to have a scheme with a high order weak convergence on the triplet $(Y_t, m_t, v_t)_{t\in[0,T]}$ solution of the SDE (24), which is the case for our scheme.

- *In the special case of an Ornstein-Uhlenbeck process driving the volatility (i.e $(Y_t)_{t\in[0,T]}$ is solution of the SDE (12)), one should replace the Ninomiya-Victoir scheme by the true solution. We can then prove more easily the same weak convergence result: at step (27) of the preceding proof, we apply Itô's formula instead of carrying out the Ninomiya-Victoir expansion. Moreover, we can prove, following the same error analysis, that the OU_Improved scheme (15) also exhibits a second order weak convergence property. Better still, it achieves a weak trajectorial convergence of order $\frac{3}{2}$ on the triplet $(Y_t, m_t, v_t)_{t\in[0,T]}$ which allows for a significant improvement of the multilevel Monte Carlo method, as we shall check numerically.*

3 Numerical results

For numerical computations, we are going to consider Scott's model (2). We use the same set of parameters as in Kahl and Jäckel [2006] : $S_0 = 100, r = 0.05, T = 1, \sigma_0 = 0.25, y_0 = 0, \kappa = 1, \theta = 0, \nu = \frac{7\sqrt{2}}{20}, \rho = -0.2$ and $f : y \mapsto \sigma_0 e^y$.

We are going to compare our schemes (WeakTraj_1, Weak_2 and OU_Improved) to the Euler scheme with exact simulation of the volatility (hereafter denoted Euler), the Kahl and Jäckel [2006] scheme (IJK) and the Cruzeiro *et al.* [2004] scheme (CMT).

3.1 Numerical illustration of strong convergence properties

In order to illustrate the strong convergence rate of a discretization scheme $\widehat{X}^N$, we consider the squared L^2-norm of the supremum of the difference between the scheme with time step $\frac{T}{N}$ and the one with time step $\frac{T}{2N}$:

$$\mathbb{E}\left[\max_{0\le k\le N}\left|\widehat{X}_{t_k}^N - \widehat{X}_{t_k}^{2N}\right|^2\right] \tag{31}$$

This quantity will exhibit the same asymptotic behavior with respect to N as the squared L^2-norm of the difference between the scheme with time step $\frac{T}{N}$ and the limiting process towards which it converges (see Alfonsi [2005]).

In Figure 1, we draw the logarithm of the Monte Carlo estimation of (31) as a function of the logarithm of the number of time steps. The number of Monte Carlo samples used is equal to $M = 10000$ and the number of discretization steps is a power of 2 varying from 2 to 256. We also consider the strong convergence of the schemes on the asset itself (see Figure 2) by computing $\mathbb{E}\left[\max_{0\leq k\leq N}\left|e^{\widehat{X}^N_{t_k}} - e^{\widehat{X}^{2N}_{t_k}}\right|^2\right]$.

The slopes of the regression lines are reported in Table 1. We see that, both for the logarithm of the asset and for the asset itself, all the schemes exhibit a strong convergence of order $\frac{1}{2}$. Our schemes only have a better constant.

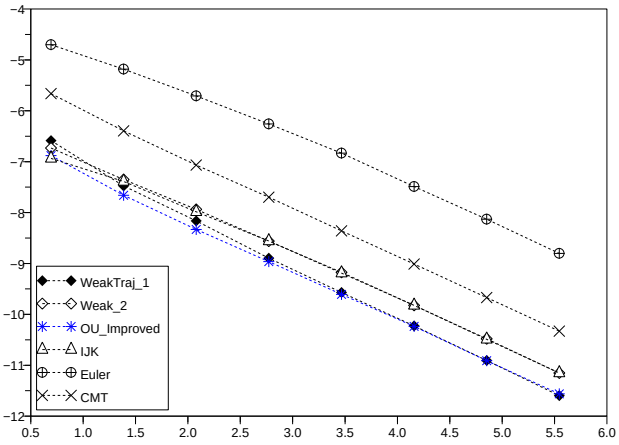

Figure 1: Strong convergence on the log-asset

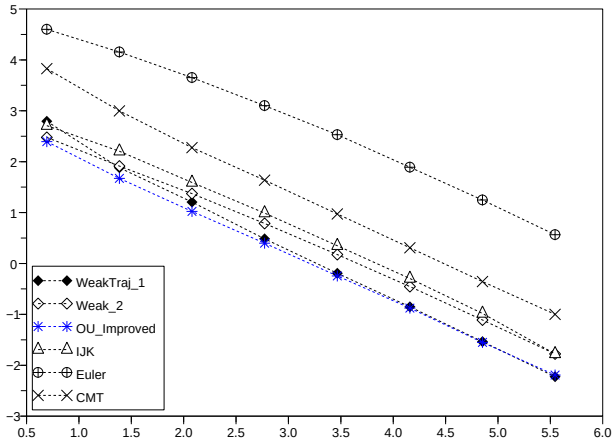

Figure 2: Strong convergence on the asset

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-1.01	-0.88	-0.94	-0.92	-0.98	-0.84
Asset	-1.01	-0.91	-0.95	-0.88	-0.95	-0.85

Table 1: Slopes of the regression lines (Strong convergence)

3.1.1 Weak trajectorial convergence

Nevertheless, as explained in Remark 5, for the scheme with time step $\frac{1}{N}$, one can replace the increments of the Brownian motion $(B_t)_{t\in[0,T]}$ by a sequence of Gaussian random variables smartly constructed from the scheme with time step $\frac{1}{2N}$. This particular coupling is possible whenever the independence structure between $(B_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ is preserved by the discretization of the latter process, which is the case for all the schemes but the CMT scheme. So we carry out this coupling and we repeat the preceding numerical experiment. The results are put together in Figures 3 and 4 and in Table 2.

As expected, we see that the WeakTraj_1 and the OU_Improved schemes exhibit a first order convergence rate whereas the other schemes exhibit a $\frac{1}{2}$ order convergence rate. Note that the CMT scheme has a weak trajectorial convergence of order one but it is much more difficult to implement the coupling for which the convergence order is indeed equal to one.

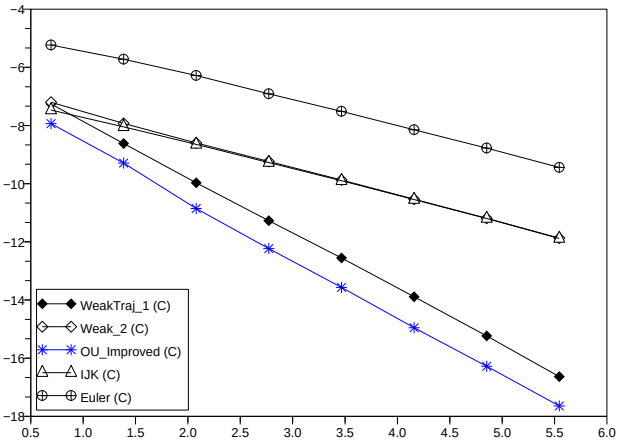

Figure 3: Weak trajectorial convergence on the log-asset (with coupling)

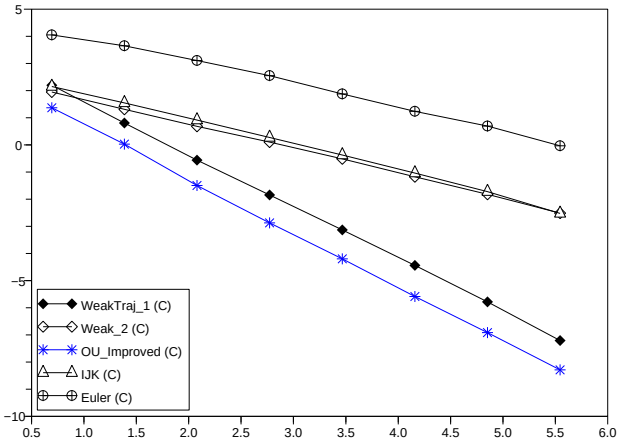

Figure 4: Weak trajectorial convergence on the asset (with coupling)

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-1.92	-0.91	-1.99	-0.95	–	-0.85
Asset	-1.92	-0.95	-2	-0.91	–	-0.87

Table 2: Slopes of the regression lines (Weak trajectorial convergence)

3.1.2 Convergence at terminal time

We consider now convergence at terminal time, precisely the squared L^2-norm of the difference between the terminal values of the schemes with time steps $\frac{T}{N}$ and $\frac{T}{2N}$:

$$\mathbb{E}\left[\left|\widehat{X}_T^N - \widehat{X}_T^{2N}\right|^2\right]. \tag{32}$$

Note that we introduce a coupling : we write the schemes straight at the terminal time as we did for the Weak_2 scheme (see (21)) and we generate the terminal values of the schemes with time steps $\frac{T}{N}$ and $\frac{T}{2N}$ using the same single normal random variable to simulate the stochastic integral w.r.t. $(B_t)_{t\in[0,T]}$. Once again, it is possible to proceed alike for all the schemes but the CMT scheme. For the latter, we simulate the scheme at all the intermediate discretization times to obtain the value at terminal time.

We also consider the convergence at terminal time of the asset itself. We report the numerical results in Figures 5 and 6 and give the slopes of the regression lines in Table 3.

We observe that, as stated in Remark 10, the OU_Improved scheme exhibits a convergence rate of order $\frac{3}{2}$, outperforming all the other schemes. As previously, the WeakTrak_1 scheme exhibits a first order convergence

	WeakTraj_1	Weak_2	OU_Improved	IJK	CMT	Euler
Log-asset	-2.03	-2	-2.97	-1.97	-1.05	-1.34
Asset	-2.02	-1.98	-2.97	-1.95	-1.08	-1.34

Table 3: Slopes of the regression lines (Convergence at terminal time)

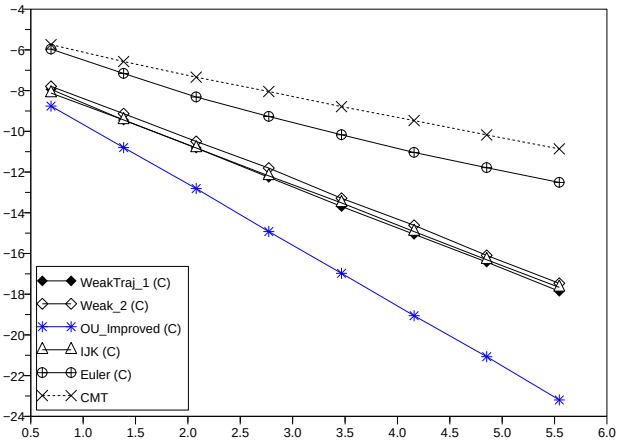

Figure 5: Convergence at terminal time for the log-asset

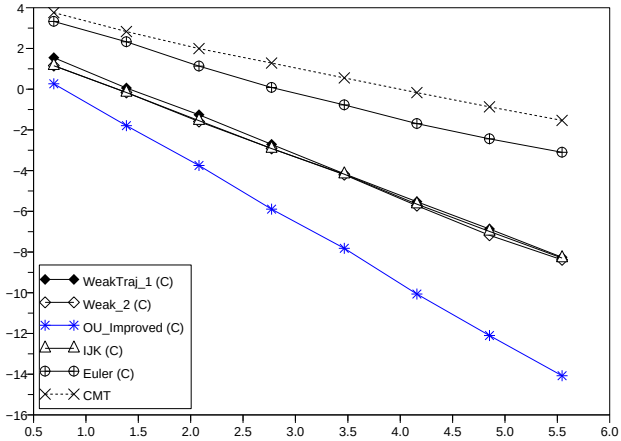

Figure 6: Convergence at terminal time for the asset

rate. Note also that this new coupling at terminal time improved the convergence rate of the Weak_2 and the IJK schemes up to order one and, surprisingly, it improved the convergence rate of the Euler scheme up to an order strictly greater than the expected $\frac{1}{2}$, approximately 0.67.

3.2 Standard call pricing

3.2.1 Numerical illustration of weak convergence

We compute the price of a call option with strike $K = 100$ and maturity $T = 1$. For all the schemes but the CMT scheme, we use the conditioning variance reduction technique presented in Remark 16.

In Figure 7, we draw the price as a function of the number of time steps for each scheme and in Figure 8 we draw the logarithm of the pricing error : $\log\left(\left|P_{\text{exact}} - P^N_{\text{scheme}}\right|\right)$ where $P_{\text{exact}} \approx 12.82603$ is obtained by a multilevel Monte Carlo with an accuracy of $5bp$, as a function of the logarithm of the number of times steps.

We see that, as expected, the Weak_2 scheme and the OU_Improved scheme exhibit a weak convergence of order two and converge much faster than the others. The weak scheme already gives an accurate price with only four time steps. The WeakTraj_1 scheme has a weak convergence of order one like the Euler and the IJK scheme, but it has a greater leading error term. Fortunately, its better strong convergence properties enable it to catch up with the multilevel Monte Carlo method as we will see hereafter.

Finally, note that the weak scheme does not require the simulation of additional terms when compared to the Euler or the IJK schemes. Combined with its second order weak convergence order, this makes the Weak_2 scheme very competitive for the pricing of plain vanilla European option.

3.2.2 Multilevel monte carlo

Let us now apply the multilevel Monte Carlo method of Giles [2008*b*] to compute the Call price. As previously, we consider the schemes straight at the terminal time and use a conditioning variance reduction technique. We give the CPU time as a function of the root mean square error in Figure 9 (see Giles [2008*b*] for details on the heuristic numerical algorithm which is used).

We observe that both the Weak_2 and the OU_Improved scheme are great time-savers. For the OU_Improved scheme, the effect coming from its good strong convergence properties is somewhat offset by the additional

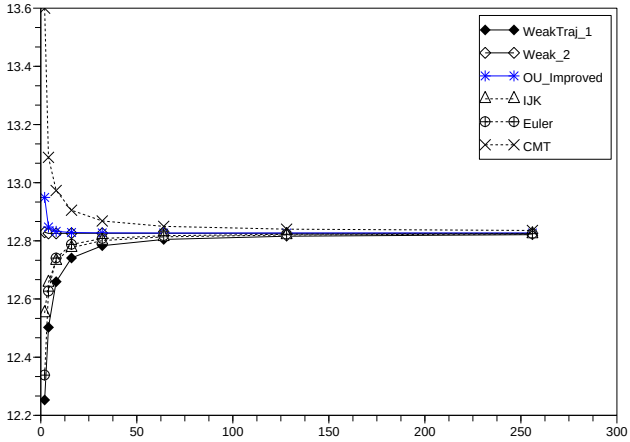

Figure 7: Convergence of the call price with respect to N

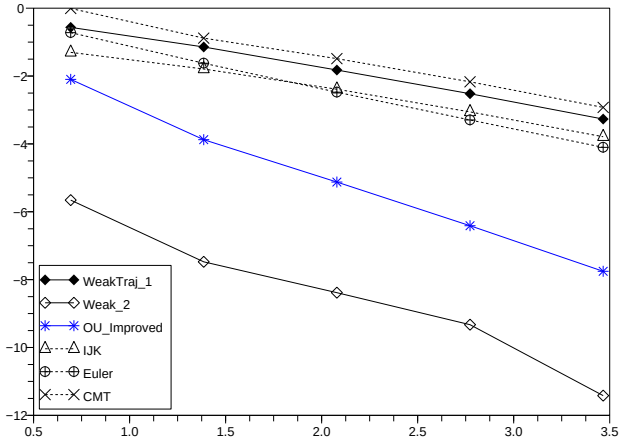

Figure 8: Illustration of the convergence rate for the call option

terms it requires to simulate. We can see nevertheless that it is going to overcome the Weak_2 scheme for bigger accuracy levels.

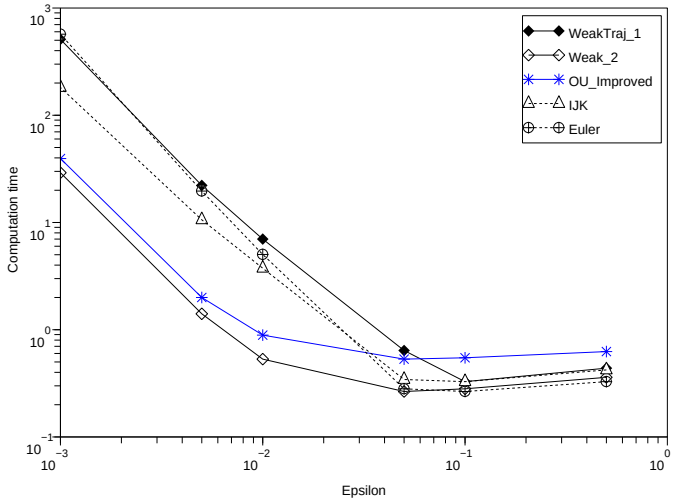

Figure 9: Multilevel Monte Carlo method for a Call option using different schemes

3.3 Lookback option pricing and multilevel Monte Carlo

Finally, we consider an example of path-dependent option pricing : the Lookback option. More precisely, we compute the price of the option whose pay-off is equal to $S_T - \min_{t\in[0,T]} S_t$.

In order to take full advantage of the good convergence properties of our schemes, we approximate the minimum of the scheme by the minimum of a drifted Brownian motion. This is similar to what is done in Giles [2008*a*].

More precisely, for the WeakTraj_1 scheme, consider the interval $[k\frac{T}{N},(k+1)\frac{T}{N}]$.

Scheme with time step δ_{2N} :
We approximate $\min_{t\in[k\frac{T}{N},(k+1)\frac{T}{N}]} S_t$ by $\widetilde{m}_{2k}^{2N}\wedge\widetilde{m}_{2k+1}^{2N}$ where, $\forall 0\leq j\leq 2N-1$,

$$\widetilde{m}_j^{2N}=\frac{1}{2}\left(e^{\widetilde{X}^{2N}_{j\frac{T}{2N}}}+S^{e,2N}_{(j+1)\frac{T}{2N}}-\sqrt{\left(e^{\widetilde{X}^{2N}_{j\frac{T}{2N}}}-S^{e,2N}_{(j+1)\frac{T}{2N}}\right)^2-2e^{2\widetilde{X}^{2N}_{j\frac{T}{2N}}}f^2(Y_{j\frac{T}{2N}})\frac{T}{2N}\ln(U_j)}\right)$$

where $S^{e,2N}_{(j+1)\frac{T}{2N}}=e^{\widetilde{X}^{2N}_{j\frac{T}{2N}}}\left(1+r\frac{T}{2N}+f(Y_{j\frac{T}{2N}})\left(\rho(W_{(j+1)\frac{T}{2N}}-W_{j\frac{T}{2N}})+\sqrt{1-\rho^2}(B_{(j+1)\frac{T}{2N}}-B_{j\frac{T}{2N}})\right)\right)$ and $(U_j)_{0\leq j\leq 2N-1}$ is an independent sequence of independent random variable uniformly distributed.

Scheme with time step δ_N :
According to Remark 5, $\widetilde{\widetilde{X}}^N_{t_{k+1}}$ is computed using the Brownian increment $\Delta\widetilde{B}^N_{k+1}$ given by a linear combination of $\left(B_{(2k+1)\frac{T}{2N}}-B_{k\frac{T}{N}},B_{(k+1)\frac{T}{N}}-B_{(2k+1)\frac{T}{2N}}\right)$ (see (10)). Now, to prevent bias, we are going to approximate $\min_{t\in[k\frac{T}{N},(k+1)\frac{T}{N}]} S_t$ by the minimum $\min_{[t_k,t_{k+1}]}\widetilde{S}^{e,2N}_t$ of some Euler scheme $\widetilde{S}^{e,2N}_t$ like in the scheme with time step δ_{2N}. To remain consistent, we have to choose

$$\widetilde{S}^{e,2N}_{t_{k+1}}=e^{\widetilde{\widetilde{X}}^N_{t_k}}\left(1+r\frac{T}{N}+f(Y_{k\frac{T}{N}})\left(\rho(W_{t_{k+1}}-W_{t_k})+\sqrt{1-\rho^2}\Delta\widetilde{B}^N_{k+1}\right)\right)$$

In order to ensure a good strong coupling with the scheme with time step δ_{2N}, we need to compute the intermediate value $\widetilde{S}^{e,2N}_{(2k+1)\frac{T}{2N}}=e^{\widetilde{\widetilde{X}}^N_{t_k}}\left(1+r\frac{T}{N}+f(Y_{k\frac{T}{N}})\left(\rho(W_{t_{k+1}}-W_{t_k})+\sqrt{1-\rho^2}\Delta\widetilde{\widetilde{B}}^{2N}_{2k+1}\right)\right)$ using some Brownian increment $\Delta\widetilde{\widetilde{B}}^{2N}_{2k+1}$ as close as possible to $B_{(2k+1)\frac{T}{2N}}-B_{k\frac{T}{N}}$ but such that $\Delta\widetilde{B}^N_{k+1}-\Delta\widetilde{\widetilde{B}}^{2N}_{2k+1}$ is independent of $\Delta\widetilde{\widetilde{B}}^{2N}_{2k+1}$ and distributed according to $\mathcal{N}(0,\frac{T}{2N})$. Choosing $\Delta\widetilde{\widetilde{B}}^{2N}_{2k+1}$ of the form $\frac{a\left(B_{(2k+1)\frac{T}{2N}}-B_{k\frac{T}{N}}\right)+b\left(B_{(k+1)\frac{T}{N}}-B_{(2k+1)\frac{T}{2N}}\right)}{\sqrt{a^2+b^2}}$ and maximizing $Cov\left(\Delta\widetilde{\widetilde{B}}^{2N}_{2k+1},B_{(2k+1)\frac{T}{2N}}-B_{k\frac{T}{N}}\right)=\frac{a}{\sqrt{a^2+b^2}}$ leads to $a=v^{2N}_{2k}+v^{2N}_{2k+1}$ and $b=v^{2N}_{2k+1}-v^{2N}_{2k}$ (see Remark 5 for the definition of $v^{2N}_{\cdot}$).

Finally, we approximate $\min_{t\in[k\frac{T}{N},(k+1)\frac{T}{N}]} S_t$ by $\widetilde{m}^N_k\wedge\widetilde{m}^N_{k+1}$ where

$$\widetilde{m}^N_k=\frac{1}{2}\left(e^{\widetilde{\widetilde{X}}^N_{k\frac{T}{N}}}+\widetilde{S}^{e,2N}_{(2k+1)\frac{T}{2N}}-\sqrt{\left(e^{\widetilde{\widetilde{X}}^N_{k\frac{T}{N}}}-\widetilde{S}^{e,2N}_{(2k+1)\frac{T}{2N}}\right)^2-2e^{2\widetilde{\widetilde{X}}^N_{k\frac{T}{N}}}f^2(Y_{k\frac{T}{N}})\frac{T}{2N}\ln(U_{2k})}\right)$$

and

$$\widetilde{m}^N_{k+1}=\frac{1}{2}\left(\widetilde{S}^{e,2N}_{(2k+1)\frac{T}{2N}}+\widetilde{S}^{e,2N}_{(k+1)\frac{T}{N}}-\sqrt{\left(\widetilde{S}^{e,2N}_{(2k+1)\frac{T}{2N}}-\widetilde{S}^{e,2N}_{(k+1)\frac{T}{N}}\right)^2-2e^{2\widetilde{\widetilde{X}}^N_{k\frac{T}{N}}}f^2(Y_{k\frac{T}{N}})\frac{T}{2N}\ln(U_{2k+1})}\right).$$

The numerical results we obtain are very satisfactory. In figure 10, we draw the CPU time times the mean square error against the root mean square error. We see that our schemes perform much better than the others.

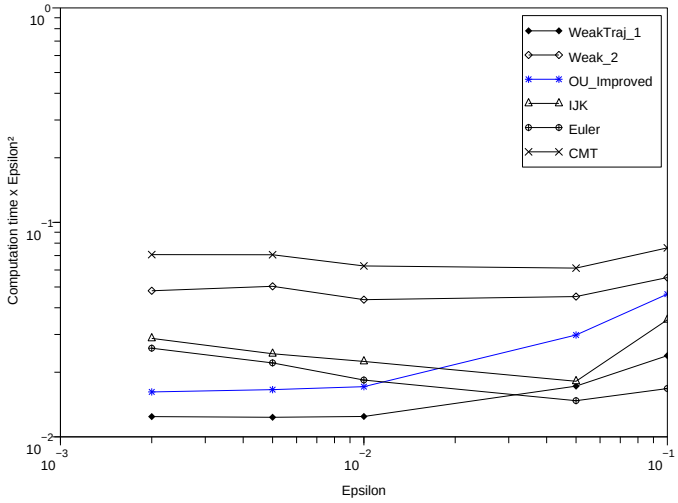

Figure 10: Multilevel Monte Carlo method for a Lookback option using different schemes.

4 Conclusion

In this article, we have capitalized on the particular structure of stochastic volatility models to propose and discuss two simple and yet competitive discretization schemes. The first one exhibits first order weak trajectorial convergence and has the advantage of improving multilevel Monte Carlo methods for the pricing of path dependent options. The second one is rather useful for pricing European options since it has a second order weak convergence rate.

We have also focused on the special case of an Ornstein-Uhlenbeck process driving the volatility, which encompasses many stochastic volatility models such as the Scott's [1987] model or the quadratic Gaussian model. Then, the convergence properties of the previous schemes are preserved when simulating $(Y_t)_{0\leq t\leq T}$ exactly. We have also proposed an improved scheme exhibiting both weak trajectorial convergence of order one and weak convergence of order two.

The numerical experiments show that our schemes are very competitive for the pricing of plain vanilla and path-dependent options. Their use with multilevel Monte Carlo gives satisfactory results too. We should also mention that the main purpose of our study was the convergence order with respect to the time step. It would be of great interest to carry out an extensive numerical study of the computational complexity of the schemes presented in this paper. This will be the subject of future research.

References

Alfonsi, A. [2005], 'On the discretization schemes for the CIR (and Bessel squared) processes', *Monte Carlo Methods and Applications* **11**(4), 355–384.

Alfonsi, A. [2009], 'High order discretization schemes for the CIR process: Application to affine term structure and Heston models', *Mathematics of Computations* **79**, 209–237.

Andersen, L. [2007], 'Efficient simulation of the heston stochastic volatility model', *SSRN eLibrary* .

Bally, V. & Talay, D. [1996], 'The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function', *Probability Theory and Related Fields* **104**(1), 43–60.

Berkaoui, A., Bossy, M. & Diop, A. [2008], 'Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence', *ESAIM. Probability and Statistics* **12**, 1–11.

Broadie, M. & Kaya, Ö. [2006], 'Exact simulation of stochastic volatility and other affine jump diffusion processes', *Operations Research* **54**(2), 217–231.

Cruzeiro, A., Malliavin, P. & Thalmaier, A. [2004], 'Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation', *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique* **338**(6), 481–486.

Deelstra, G. & Delbaen, F. [1998], 'Convergence of discretized stochastic (interest rate) processes with stochastic drift term', *Applied Stochastic Models and Data Analysis* **14**(1), 77–84.

Doss, H. [1977], 'Liens entre équations différentielles stochastiques et ordinaires', *Ann. Inst. H. Poincaré Sect. B (N.S.)* **13**(2), 99–125.

Giles, M. [2008*a*], Improved multilevel Monte Carlo convergence using the Milstein scheme, *in* 'Monte Carlo and quasi-Monte Carlo methods 2006', Springer, Berlin, pp. 343–358.

Giles, M. [2008*b*], 'Multilevel Monte Carlo path simulation', *Operations Research* **56**(3), 607–617.

Guyon, J. [2006], 'Euler scheme and tempered distributions', *Stochastic Processes and their Applications* **116**(6), 877–904.

Heston, S. [1993], 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *Review of Financial Studies* **6**(2), 327–43.

Hull, J. & White, A. [1987], 'The pricing of options on assets with stochastic volatilities', *The Journal of Finance* **42**(2), 281–300.

Kahl, C. & Jäckel, P. [2006], 'Fast strong approximation Monte Carlo schemes for stochastic volatility models', *Quantitative Finance* **6**(6).

Kahl, C. & Schurz, H. [2006], 'Balanced Milstein methods for ordinary SDEs', *Monte Carlo Methods and Applications* **12**(2), 143–170.

Karatzas, I. & Shreve, S. [1991], *Brownian motion and stochastic calculus*, 2nd edn, Springer-Verlag New-York.

Kebaier, A. [2005], 'Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing', *The Annals of Applied Probability* **15**(4), 2681–2705.

Kunita, H. [1984], 'Stochastic differential equations and stochastic flows of diffeomorphisms', **1097**, 143–303.

Kusuoka, S. [2001], 'Approximation of expectation of diffusion process and mathematical finance', *Taniguchi Conference on Mathematics Nara '98* **31**, 147–165.

Kusuoka, S. [2004], 'Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus', *Advances in mathematical economics* **6**, 69–83.

Lapeyre, B. & Temam, E. [2001], 'Competitive Monte Carlo methods for pricing Asian options', *Journal of Computational Finance* **5**(1).

Lord, R., Koekkoek, R. & Van Dijk, D. [2008], 'A comparison of biased simulation schemes for stochastic volatility models', *SSRN eLibrary* .

Lyons, T. & Victoir, N. [2004], 'Cubature on Wiener space', *Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences* **460**(2041), 169–198.

Milstein, G. [1995], *Numerical Integration of Stochastic Differential Equations*, Vol. 313, Kluwer Academic Publishers.

Ninomiya, S. & Ninomiya, M. [2009], 'A new higher-order weak approximation scheme for stochastic differential equations and the Runge-Kutta method', *Finance and Stochastics* **13**, 415–443.

Ninomiya, S. & Victoir, N. [2008], 'Weak approximation of stochastic differential equations and application to derivative pricing', *Applied Mathematical Finance* **15**(1-2), 107–121.

Romano, M. & Touzi, N. [1997], 'Contingent claims and market completeness in a stochastic volatility model', *Mathematical Finance* **7**(4), 399–412.

Scott, L. [1987], 'Option pricing when the variance changes randomly: theory, estimation, and an application', *The Journal of Financial and Quantitative Analysis* **22**(4), 419–438.

Stein, E. & Stein, J. [1991], 'Stock price distributions with stochastic volatility: an analytic approach', **4**(4), 727–752.

Talay, D. & Tubaro, L. [1990], 'Expansion of the global error for numerical schemes solving stochastic differential equations', *Stochastic Analysis and Applications* **8**(4), 483–509.

Tanaka, H. & Kohatsu-Higa, A. [2009], 'An operator approach for Markov chain weak approximations with an application to infinite activity Lévy driven SDEs', *The Annals of Applied Probability* **19**(3), 1026–1062.

Wiggins, J. [1987], 'Option values under stochastic volatility: Theory and empirical estimates', *Journal of Financial Economics* **19**(2), 351–372.

5 Appendix

5.1 Proof of Lemma 3

We first suppose that $p = 1$. According to Theorem 5.2 page 72 of Milstein [1995], it suffices to check that there exists a positive constant C independent of N such that

$$\begin{aligned} \left|\mathbb{E}\left(Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right)\right| &\leq C\delta_N^2 \\ \left|\mathbb{E}\left(\left(Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right)^2\right)\right|^{\frac{1}{2}} &\leq C\delta_N^{\frac{3}{2}} \\ \left|\mathbb{E}\left(\left(Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right)^4\right)\right|^{\frac{1}{4}} &\leq C\delta_N^{\frac{5}{4}} \end{aligned} \tag{33}$$

First note that

$$Y_{\delta_N} - \overline{Y}^N_{\delta_N} = \int_0^{\delta_N} b(Y_s) - b(y_0)ds + \int_0^{\delta_N} \left(\int_0^s (b\sigma' + \frac{1}{2}\sigma^2\sigma'')(Y_r)dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0))dW_r\right) dW_s$$

Thanks to Itô's formula and to assumption $(\mathcal{H}5)$, we have that

$$\begin{aligned} \left|\mathbb{E}\left(Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right)\right| &= \left|\int_0^{\delta_N}\int_0^s \mathbb{E}\left((bb' + \frac{1}{2}b''\sigma^2)(Y_r)\right) drds\right| \\ &\leq C\left|\int_0^{\delta_N}\int_0^s C(1 + \mathbb{E}(|Y_r|^2))drds\right| \\ &\leq C\delta_N^2 \end{aligned}$$

Using assumptions ($\mathcal{H}5$) and ($\mathcal{H}6$), we also have $\forall p \geq 1$

$$\begin{aligned}
\mathbb{E}\left(\left|Y_{\delta_N} - \overline{Y}^N_{\delta_N}\right|^{2p}\right) &\leq 2^{2p-1}\mathbb{E}\left[\left|\int_0^{\delta_N} b(Y_s) - b(y_0)ds\right|^{2p}\right. \\
&\qquad \left. + \left|\int_0^{\delta_N}\left(\int_0^s (b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r)dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0))dW_r\right)dW_s\right|^{2p}\right] \\
&\leq 2^{2p-1}\left[\delta_N^{2p-1}\int_0^{\delta_N}\mathbb{E}\left(|b(Y_s) - b(y_0)|^{2p}\right)ds\right. \\
&\qquad \left. + C\delta_N^{p-1}\int_0^{\delta_N}\mathbb{E}\left(\left|\int_0^s (b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r)dr + (\sigma\sigma'(Y_r) - \sigma\sigma'(y_0))dW_r\right|^{2p}\right)ds\right] \\
&\leq C\left[\delta_N^{2p-1}\int_0^{\delta_N} s^p ds + \delta_N^{p-1}\int_0^{\delta_N} s^{2p-1}\int_0^s \mathbb{E}\left(\left|(b\sigma' + \frac{1}{2}\sigma\sigma'')(Y_r)\right|^{2p}\right)dr\,ds\right. \\
&\qquad \left. + \delta_N^{p-1}\int_0^{\delta_N} s^{p-1}\int_0^s \mathbb{E}\left(\left|\sigma\sigma'(Y_r) - \sigma\sigma'(y_0)\right|^{2p}\right)dr\,ds\right] \\
&\leq C\delta_N^{3p}
\end{aligned}$$

This implies both the second and the third inequality of (33). This estimation is also sufficient to extend the result of Milstein [1995] to the L^{2p} norm and conclude the proof.

5.2 Proof of Lemma 8

One can easily check that $(Y_t)_{0\leq t\leq T}$ is a Gaussian process which has the same distribution law as the process $(y_0e^{-\kappa t} + \theta(1-e^{-\kappa t}) + \frac{\nu e^{-\kappa t}}{\sqrt{2\kappa}}W_{e^{2\kappa t}-1})_{0\leq t\leq T}$. So,

$$\begin{aligned}
\mathbb{E}\left(e^{c_1\sup_{0\leq t\leq T}|Y_t|^{1+c_2}}\right) &= \mathbb{E}\left(e^{c_1\sup_{0\leq t\leq T}|y_0e^{-\kappa t}+\theta(1-e^{-\kappa t})+\frac{\nu e^{-\kappa t}}{\sqrt{2\kappa}}W_{e^{2\kappa t}-1}|^{1+c_2}}\right) \\
&\leq C\mathbb{E}\left(e^{C\sup_{0\leq t\leq T}|W_{e^{2\kappa t}-1}|^{1+c_2}}\right)
\end{aligned}$$

Since $\sup_{0\leq t\leq e^{2\kappa T}-1}|W_t| = \left(\sup_{0\leq t\leq e^{2\kappa T}-1} W_t\right) \vee \left(-\inf_{0\leq t\leq e^{2\kappa T}-1} W_t\right)$, we deduce from the symmetry property of the Brownian motion that

$$\begin{aligned}
\mathbb{E}\left(e^{c_1\sup_{0\leq t\leq T}|Y_t|^{1+c_2}}\right) &\leq C\mathbb{E}\left(e^{C|\sup_{0\leq t\leq e^{2\kappa T}-1} W_t|^{1+c_2}} + e^{C|\inf_{0\leq t\leq e^{2\kappa T}-1} W_t|^{1+c_2}}\right) \\
&\leq 2C\mathbb{E}\left(e^{C|\sup_{0\leq t\leq e^{2\kappa T}-1} W_t|^{1+c_2}}\right)
\end{aligned}$$

The probability density function of $\sup_{0\leq t\leq T} W_t$ is equal to $y \mapsto \sqrt{\frac{2}{\pi T}}e^{-\frac{y^2}{2T}}\mathbb{1}_{\{y>0\}}$ (see for example problem 8.2 p. 96 of Karatzas and Shreve [1991]) which permits to conclude.

5.3 Proof of Lemma 14

The first point is an obvious consequence of the Feynman-Kac theorem. In order to prove the second one, let us first check the following result :

$$\begin{aligned}
&\text{For any multi-index } \beta \in \mathbb{N}^3 \text{ such that } \beta_1 \leq 6, \exists C_\beta, K_\beta \geq 0 \text{ and } p_\beta \in \mathbb{N} \text{ such that} \\
&\forall (y,m,v) \in \mathcal{D}_T, \quad |\partial_\beta \gamma_x(y,m,v)| \leq C_\beta e^{-K_\beta x^2}\left(1 + |x|^{p_\beta}\right)
\end{aligned} \tag{34}$$

Indeed, using Leibniz's formula, one can show that $\partial_\beta \gamma_x(y,m,v)$ can be written as a weighted sum of terms of the form

$$\zeta_k = \frac{(x-\log(s_0)+\rho F(y_0)-\rho F(y)-m)^{k_2}}{v^{k_1+\frac{1}{2}}} \exp\left(-\frac{(x-\log(s_0)+\rho F(y_0)-\rho F(y)-m)^2}{2(1-\rho^2)v}\right) \prod_{i=0}^{k_3} a_i F^{(i)}(y)$$

where $k=(k_1,k_2,k_3)$ belongs to a finit set $I_\beta \subset \mathbb{N}^3$ and $(a_i)_{0\leq i\leq k_3}$ are constants taking value in $\{0,1\}$. Using assumption $(\mathcal{H}13)$ and $(\mathcal{H}14)$ and Young's inequality, we show that $\exists C_k, K_k > 0$ and $p_k \in \mathbb{N}$ such that $|\zeta_k| \leq C_k e^{-K_k x^2}(1+|x|^{p_k})$ which yields the desired result.

Now, let us fix $\alpha \in \mathbb{N}^3, l \in \mathbb{N}$ such that $2l+|\alpha| \leq 6$ and $(t,y,m,v) \in [0,T]\times \mathcal{D}_t$. Thanks to PDE (24), $\partial_t^l \partial_\alpha u_x(t,y,m,v) = (-1)^l \partial_\alpha \mathcal{L}^l u_x(t,y,m,v)$. One can check that the right hand side is equal to a weighted sum of terms of the form $\partial_{\beta_1} u_x(t,y,m,v) \times \pi_{\beta_2}(b,\sigma,f,h)$ where $\beta_1 \in \mathbb{N}^3$ is multi-index belonging to a finite set $I^1_{\alpha,l}$, β_2 is a suffix belonging to a finite set $I^2_{\alpha,l}$ and $\pi_{\beta_2}(b,\sigma,f,h)$ is a product of terms involving the functions b,σ,f,h and their derivatives up to order 4.

On the first hand, assumptions $(\mathcal{H}12)$ and $(\mathcal{H}13)$ yield that $\exists c^2_{l,\alpha} \geq 0$ and $q_{l,\alpha} \in \mathbb{N}$ such that

$$\forall \beta_2 \in I^2_{\alpha,l}, |\pi_{\beta_2}(b,\sigma,f,h)| \leq c^2_{l,\alpha}(1+|y|^{q_{l,\alpha}}). \tag{35}$$

On the other hand, by inverting expectation and differentiations, we see that $\partial_{\beta_1} u_x(t,y,m,v)$ is equal to the expectation of a product between derivatives of the flow $(y,m,v) \to (Y_{T-t}, m_{T-t}, v_{T-t})^{(y,m,v)}$ and derivatives of the function γ_x evaluated at $(Y_{T-t}, m_{T-t}, v_{T-t})^{(y,m,v)} \in \mathcal{D}_T$. Using result (34) and the fact that, under assumptions $(\mathcal{H}12)$ and $(\mathcal{H}13)$, the derivatives of the flow satisfy a system of SDEs with Lipschitz continuous coefficients (see for example Kunita [1984]) we show that $\exists c^1_{l,\alpha}, K_{l,\alpha} > 0$ and $p_{l,\alpha} \in \mathbb{N}$ such that

$$\forall \beta_1 \in I^1_{\alpha,l}, |\partial_{\beta_1} u_x(t,y,m,v)| \leq c^1_{l,\alpha} e^{-K_{l,\alpha}x^2}(1+|x|^{p_{l,\alpha}}). \tag{36}$$

Gathering (35) and (36) enables us to conclude.

5.4 Proof of Lemma 15

Making the link between ODEs and SDEs (see Doss [1977]), one can check that $(\overline{Y}^N_{t_1}, \dots, \overline{Y}^N_{t_N})$ has the same distribution law as $(\overline{\overline{Y}}_{2t_1}, \dots, \overline{\overline{Y}}_{2t_N})$ where $(\overline{\overline{Y}}_t)_{t\in[0,2T]}$ is solution of the following inhomogeneous SDE $\overline{\overline{Y}}_t = y_0 + \int_0^t \overline{b}(s,\overline{\overline{Y}}_s)ds + \int_0^t \overline{\sigma}(s,\overline{\overline{Y}}_s)dW_s$ with,
$\forall (s,y) \in [0,2T]\times\mathbb{R}$,

$$\overline{b}(s,y) = \begin{cases} b(y) - \frac{1}{2}\sigma\sigma'(y) & \text{if } s \in \bigcup_{k=0}^{N-1}\left[\frac{(4k+1)T}{2N}, \frac{(4k+3)T}{2N}\right] \\ -\frac{1}{2}\sigma\sigma'(y) & \text{otherwise} \end{cases}$$

and

$$\overline{\sigma}(s,y) = \begin{cases} 0 & \text{if } s \in \bigcup_{k=0}^{N-1}\left[\frac{(4k+1)T}{2N}, \frac{(4k+3)T}{2N}\right] \\ \sigma(y) & \text{otherwise} \end{cases}$$

Since these coefficient have a uniform in time linear growth in the spatial variable, one easily concludes.