A hybrid tree-finite difference approach for the Heston model

Maya Briani*
Lucia Caramellino†
Antonino Zanette‡

Abstract

We propose a hybrid tree-finite difference method in order to approximate the Heston model. We prove the convergence by embedding the procedure in a bivariate Markov chain and we study the convergence of European and American option prices. We finally provide numerical experiments that give accurate option prices in the Heston model, showing the reliability and the efficiency of the algorithm.

Keywords: tree methods; finite differences; Heston model; pricing European and American options.
2000 MSC: 91G10, 60H30, 65C20.

Premia 22

1 Introduction

Black-Scholes model was the most popular model for derivative pricing and hedging, although it has shown several problems with capturing dramatic moves in financial markets. In fact, the assumption of a constant volatility in the Black-Scholes model over the lifetime of the derivative is not realistic. As an alternative to the Black-Scholes model, stochastic volatility models emerged. The Heston model [17] is perhaps the most popular stochastic volatility model, allowing one to obtain closed-formulae in the European case using Fourier transform. In the American option pricing case, the main algorithms turn out to be tree methods, Fourier-cosine methods and finite difference methods. Approximating trees for the Heston model have been considered in different papers, see e.g. [22], [13] , [14], [18], [15]. The tree approach of Vellekoop and Nieuwenhuis [25] actually provides at our knowledge the best tree procedure in the literature. They use an approach which is based on a modification of an explicitly defined

*Istituto per le Applicazioni del Calcolo, CNR Roma - m.briani@iac.cnr.it
†Dipartimento di Matematica, Università di Roma Tor Vergata - caramell@mat.uniromP2.it
‡Dipartimento di Scienze Economiche e Statistiche, Università di Udine - antonino.zanette@uniud.it

stock price tree where the number of nodes grows quadratically in the number of time steps. Fang and Oosterlee [11] use a Fourier-cosine series expansion approach for pricing Bermudan options under the Heston model. As for finite difference methods for solving the parabolic partial differential equation associated to the option pricing problems, they can be based on implicit, explicit or alternating direction implicit schemes. The implicit scheme requires to solve a sparse system at each time step. Clarke and Parrott [8] and Oosterlee [26] formulate the American put pricing problem as a linear complementarity problem (LCP) and use an implicit finite difference scheme combined with a multigrid procedure, whereas Forsyth, Vetzal and Zvan [33] use a penalty method. The explicit scheme is a quick approach although it requires small time steps to retain the stability. This request brings to a large number of time steps and is not economic in computation. The ADI schemes are good alternative methods. For example, Hout and Foulon [20] investigate four splitting schemes of the ADI type for solving the PDE Heston equation: the Douglas scheme, the Craigh-Sneyd scheme, the Modified Craigh-Sneyd scheme and the Hundsdorfer-Verwer scheme. Ikonen and Tovainen [19] propose a componentwise splitting method for pricing American options in the Heston model. The linear complementarity problem associated to the American option problem is decomposed into a sequence of five one-dimensional LCP's problems at each time step. The advantage is that LCP's need the use of tridiagonal matrices. In Haentjens, Hout and Foulon [20], the splitting method of Ikonen and Tovainen is combined with ADI schemes in order to obtain more efficient numerical results.

In this paper we propose a new approach based both on tree and finite difference methods. Roughly speaking, our method approximates the CIR type volatility process through a tree approach already studied in Appolloni, Caramellino and Zanette [3], which turns out to be very robust and reliable. And at each step, we make use of a suitable transformation of the asset price process allowing one to take care of a new diffusion process with null correlation w.r.t. the volatility process. Then, by taking into account the conditional behavior with respect to the evolution of the volatility process, we consider a finite difference method to deal with the evolution of the (transformed) underlying asset price process.

The paper is organized as follows. In Section 2, we introduce the model, we study in details the partial differential equation associated to the pricing problem (Section 2.1) and then we describe the hybrid-finite difference scheme (Section 2.2). The convergence of the approximating algorithm is studied in Section 3. Section 4 is devoted to numerical results and comparisons.

Acknowledgments. The authors wish to thank Alexander Ern and Giuseppe Ruzzi for useful discussions and comments.

2 Construction of the method

The Heston model [17] concerns with cases where the volatility V is assumed to be stochastic. The dynamics under the risk neutral measure of the share price S and the volatility process V

are governed by the stochastic differential equation system

$$\frac{dS(t)}{S(t)} = (r-\delta)dt + \sqrt{V(t)}\, dZ_S(t),$$
$$dV(t) = \kappa(\theta - V(t))dt + \sigma\sqrt{V(t)}\, dZ_V(t),$$

with $S(0) = S_0 > 0$ and $V(0) = V_0 > 0$, where Z_S and Z_V are Brownian motions with correlation coefficient ρ: $d\langle Z_S, Z_V\rangle(t) = \rho\, dt$. Here r is the risk free rate of interest and δ the continuous dividend rate. We recall that the dynamics of V follows a CIR process with mean reversion rate κ and long run variance θ. The parameter σ is called the volatility of the volatility.

From now on we set

$$\bar\rho = \sqrt{1-\rho^2} \quad \text{and} \quad Z_V = W, \quad Z_S = \rho W + \bar\rho Z,$$

in which (W, Z) denotes a standard 2-dimensional Brownian motion. So, the dynamics can be written as

$$\frac{dS(t)}{S(t)} = (r-\delta)dt + \sqrt{V(t)}\, (\rho dW(t) + \bar\rho dZ(t)), \tag{2.1}$$
$$dV(t) = \kappa(\theta - V(t))dt + \sigma\sqrt{V(t)}\, dW(t). \tag{2.2}$$

We consider the diffusion pair (Y, V), where

$$Y_t = \log S_t - \frac{\rho}{\sigma} V_t. \tag{2.3}$$

One has

$$dY(t) = \Big(r - \delta - \frac{1}{2}V_t - \frac{\rho}{\sigma}\kappa(\theta - V_t)\Big)dt + \bar\rho\sqrt{V(t)}\, dZ(t), \tag{2.4}$$
$$dV(t) = \kappa(\theta - V(t))dt + \sigma\sqrt{V(t)}\, dW(t), \tag{2.5}$$

(recall that W and Z are independent Brownian motions), with

$$Y_0 = \log S_0. \tag{2.6}$$

In the following, we define μ_Y and μ_V to be the drift coefficient of Y_t and V_t respectively, i.e.

$$\mu_Y(v) = r - \delta - \frac{1}{2}v - \frac{\rho}{\sigma}\kappa(\theta - v) \quad \text{and} \quad \mu_V(v) = \kappa(\theta - v). \tag{2.7}$$

This means that any functional of the pair (S_t, V_t) can be written as a suitable functional of the pair (Y_t, V_t) by using the transformation (2.3), so (Y_t, V_t) will be our underlying process of interest.

2.1 The associated pricing PDE in a small time interval

Let $f = f(y, v)$ be a function of the time and the space-variable pair (y, v). For h small, we need to compute (an estimate for) the quantity $u(t, y, v)$ defined through

$$u(t,y,v) = \mathbb{E}\Big(f(Y^{t,y,v}_{t+h}, V^{t,v}_{t+h})\Big),$$

in which $(Y^{t,y,v}, V^{t,v})$ denotes the solution to (2.4) and (2.5) with the starting condition $(Y_t, V_t) = (y, v)$. In our mind, the time instant t plays the role of a dicretization instant in $[0, T]$, that is $t = nh$, so $t + h = (n + 1)h$ stands for the next discretizing time.
We first notice that

$$\mathbb{E}\Big(f(Y^{t,y,v}_{t+h}, V^{t,v}_{t+h})\Big) = \mathbb{E}\Big(\mathbb{E}\Big(f(Y^{t,y,v}_{t+h}, V^{t,v}_{t+h})\,|\,\mathcal{F}^W_{t+h}\Big)\Big)$$

where $\mathcal{F}^W_{t+h} = \sigma(W_u : u \leq t + h)$. But conditionally to $\mathcal{F}^W_{t+h}$, the volatility process V can be considered deterministic and the process Y turns out to have constant coefficients. More precisely, for $g \in L^2([t, t + h])$ with $g \geq 0$ a.e. and $g_t = v$, set

$$U^{(g),t,y}_{t+h} = y + \int_t^{t+h} \mu_Y(g_s)ds + \bar\rho \int_t^{t+h} \sqrt{g_s}\, dZ(s). \tag{2.8}$$

Then

$$\mathbb{E}\Big(f(Y^{t,y,v}_{t+h}, V^{t,v}_{t+h})\,|\,\mathcal{F}^W_{t+h}\Big) = \mathbb{E}\Big(f(U^{(g),t,y}_{t+h}, g_{t+h}\Big)|_{g=V^{t,v}}.$$

We define now

$$\bar u(t,y;g) = \mathbb{E}\Big(f(U^{(g),t,y}_{t+h}, g_{t+h})\Big),$$

so that

$$\mathbb{E}\Big(f(Y^{s,y,v}_{t+h}, V^{s,v}_{t+h})\,|\,\mathcal{F}^W_{t+h}\Big) = \bar u(t,y;V^{t,v})$$

and therefore,

$$u(t,y,v) = \mathbb{E}\Big(\bar u(t,y;V^{t,v})\Big). \tag{2.9}$$

Let us now discuss the quantity $\bar u(t, y; g)$, for g fixed as required above, that is

$$g : [t, t + h] \to \mathbb{R}_+ \text{ is continuous and } g_t = v. \tag{2.10}$$

Set $U^{(g),s,y}_{t+h}$ as the solution $U^{(g)}$ at time $t + h$, with starting condition $U^{(g)}_s = y$ of the following stochastic differential equation *with deterministic (although path dependent) coefficients*:

$$dU^{(g)}_u = \mu_Y(g_u)du + \bar\rho\,\sqrt{g_u}\, dZ_u. \tag{2.11}$$

Recall that the associated infinitesimal generator is given by

$$L^{(g)}_u = \mu_Y(g_u)\partial_y + \frac{1}{2}\,\bar\rho^2 g_u \partial^2_{yy}. \tag{2.12}$$

So, we get

$$\bar u(t,y;g) = \bar u(s,y;g)|_{s=t}, \quad \text{with} \quad \bar u(s,y;g) = \mathbb{E}\Big(f(U^{(g),s,y}_{t+h}; g_{t+h})\Big).$$

Now, from the Feynman-Kac formula, the function $(s, y) \mapsto \bar{u}(s, y; g)$ solves the parabolic PDE Cauchy problem

$$\begin{aligned} &\partial_s \bar{u}(s, y; g) + L_s^{(g)} \bar{u}(s, y; g) = 0 && y \in \mathbb{R}, s \in (t, t+h), \\ &\bar{u}(t+h, y; g) = f(y, g_{t+h}) && y \in \mathbb{R}. \end{aligned} \tag{2.13}$$

Once the problem (2.13) is solved, we can proceed to compute $u(t, y, v)$ by using (2.9). We stress that the fixed path g plays the role of a parameter and the solution to (2.13) depends in general on the whole trajectory of g.

We consider now the case $h \simeq 0$, so that, by (2.10), $g_s \simeq g_t = v$ and $\mu_Y(g_s) \simeq \mu_Y(g_t) = \mu_Y(v)$. This numerically brings to replace (2.13) with a PDE problem *with constant coefficients*. More precisely, we consider the approximation $\hat{u}^h(s, y; v, g_{t+h})$ for $\bar{u}(t, y; g)$ given by the solution to

$$\begin{aligned} &\partial_s \hat{u}^h(s, y; v, g_{t+h}) + L^{(v)} \hat{u}^h(s, y; v, g_{t+h}) = 0 && y \in \mathbb{R}, s \in (t, t+h), \\ &\hat{u}^h(t+h, y; v, g_{t+h}) = f(y, g_{t+h}) && y \in \mathbb{R}, \end{aligned} \tag{2.14}$$

with

$$L^{(v)} = \mu_Y(v)\partial_y + \frac{1}{2}\bar{\rho}^2 v \partial^2_{yy}.$$

Let us remark that the solution to (2.14) actually depends on g only through $v = g_t$ (appearing in the coefficients of the second order operator) and g_{t+h} (appearing in the Cauchy condition), that is why we used the notation $\hat{u}^h(s, y; v, g_{t+h})$. In contrast, the function solving (2.13) depends in principle on the whole trajectory g over the time interval $[t, t+h]$.

Now, problem (2.14) can be easily solved by using a finite difference numerical method. Numerical reasonings suggest the use of an implicit approximation (in time) if v is "far enough" from zero, otherwise an explicit method should be considered - details are given in Section 3.1.1 and 3.1.2. This means that one fixes a space-step Δy_h and a space-grid $\mathcal{Y}^h = \{y_j = Y_0 + j\Delta y_h\}_{j \in \mathbb{Z}}$ splitting the real line and approximates the solution $\hat{u}^h(s, y; v, g_{t+h})$ to (2.14) on the grid $\mathcal{Y}^h$ by means of a linear operator (infinite dimensional matrix) $\Pi^h(v) = (\Pi^h(v)_{i,j})_{i,j \in \mathbb{Z}}$. In other words, one gets

$$\hat{u}^h(s, y_i; v, g_{t+h}) \simeq \sum_{j \in \mathbb{Z}} \Pi^h(v)_{i,j} f(y_j, g_{t+h}), \quad i \in \mathbb{Z}.$$

Now, recalling (2.9) and the fact that $\bar{u}^h \simeq \hat{u}^h$, on the grid $\mathcal{Y}$ the function u is approximated through

$$u(t, y_i, v) \simeq \mathbb{E}\Big(\hat{u}^h(t, y_i; v, V^{t,v}_{t+h})\Big) \simeq \sum_{j \in \mathbb{Z}} \Pi^h(v)_{i,j} \mathbb{E}\Big(f(y_j, V^{t,v}_{t+h})\Big), \quad i \in \mathbb{Z}. \tag{2.15}$$

We stress that the expectation on the r.h.s. above is now written in terms of the process V only, and this is the key point of our story because we can now use the tree method in [3]. But we will examine in depth this point in a moment.

In practice, one cannot solve the PDE problem over the whole real line. So, one takes a positive integer $M_h > 0$ such that $M_h \Delta y_h \to +\infty$ as $h \to 0$ and considers a discretization of the (space) interval $[-M_h\Delta y_h + Y_0, Y_0 + M_h \Delta y_h]$ in $2M_h + 1$ equally spaced points $y_j = Y_0 + j\Delta y_h$, $j \in \mathcal{J}_{M_h} = \{-M_h, \dots, M_h\}$. Then, the grid $\mathcal{Y}^h_{M_h} = \{y_j = Y_0 + j\Delta y_h\}_{j \in \mathcal{J}_{M_h}}$ is finite and the approximation of $\hat{u}^h_n(nh, y; v, g_{t+h})$ is done by adding to (2.14) suitable boundary conditions.

By calling again $\Pi^h(v)$ the matrix (now, finite dimensional) giving the solution from the finite difference approach, we still obtain

$$\mathbb{E}\Big(f(Y^{t,y,v}_{t+h}, V^{t,v}_{t+h})\Big)\Big|_{y=y_i} \simeq \sum_{j\in\mathcal{J}_{M_h}} \Pi^h(v)_{i,j}\mathbb{E}\Big(f(y_j, V^{t,v}_{t+h})\Big), \quad i\in\mathcal{J}_{M_h}. \tag{2.16}$$

2.2 The hybrid tree-finite difference approach

We describe the main ideas of our approximating algorithm by means of an example, which was our starting point.
Consider an American option with maturity T and payoff function $(\Phi(S_t))_{t\in[0,T]}$. First of all, by using (2.3) we replace the pair (S,V) with the pair (Y,V), so the obstacle will be given by

$$\Psi(Y_t, V_t) = \Phi(e^{Y_t+\frac{\rho}{\sigma}V_t}), \quad t\in[0,T].$$

The price at time 0 of such an option is then approximated by a backward dynamic programming algorithm, working as follows. First, consider a discretization of the time interval $[0,T]$ into N subintervals of length $h=T/N$: $[0,T]=\cup_{n=0}^{N-1}[nh,(n+1)h]$. Then the price $P(0,Y_0,V_0)$ of such an American option is numerically approximated through the quantity $P_h(0,Y_0,V_0)$ which is iteratively defined as follows: for $(y,v)\in\mathbb{R}\times\mathbb{R}_+$,

$$\begin{cases} P_h(T,y,v) = \Psi(y,v) \quad \text{and as } n=N-1,\dots,0\\ P_h(nh,y,v) = \max\Big\{\Psi(y,v), e^{-rh}\mathbb{E}\Big(P_h\Big((n+1)h, Y^{nh,y,v}_{(n+1)h}, V^{nh,v}_{(n+1)h}\Big)\Big)\Big\}.\end{cases}$$

From the financial point of view, this means to allow the exercise at the fixed times nh, $n=0,\dots,N$. Now, what we are going to set up is a contamination of a tree method for the process V with a finite difference method to handle the noise in Y (which is independent of the noise driving V). In fact, the expectations appearing in the backward induction can be written as expectations of functions of the process V only, such functions being solution to parabolic PDE's. So, we proceed as described in the previous section: we fix a grid on the y-axis $\mathcal{Y}^h_{M_h}=\{y_j=Y_0+j\Delta y_h\}_{j\in\mathcal{J}_{M_h}}$, with $\mathcal{J}_{M_h}=\{-M_h,\dots,M_h\}$, and we approximate the above conditional expectations for $y=y_i\in\mathcal{Y}^h_{M_h}$ by using the matrix $\Pi^h(v)$ from the finite difference method. So, as already seen in (2.16), we write

$$\mathbb{E}\Big(P_h\Big((n+1)h, Y^{nh,y,v}_{(n+1)h}, V^{nh,v}_{(n+1)h}\Big)\Big)\Big|_{y=y_i} \simeq \sum_{j\in\mathcal{J}_{M_h}} \Pi^h(v)_{i,j}\mathbb{E}\Big(P_h\Big((n+1)h, y_j, V^{nh,v}_{(n+1)h}\Big), \quad i\in\mathcal{J}_{M_h}. \tag{2.17}$$

By resuming, the price $P(0,y,v)$ in $y\in\mathcal{Y}^h_{M_h}$ can be numerically computed from the function $\hat P_h(0,y,v)$ defined on the grid $\mathcal{Y}^h_{M_h}$ as follows:

$$\begin{cases} \hat P_h(T,y_i,v) = \Psi(y_i,v), \quad i\in\mathcal{J}_{M_h} \text{ and as } n=N-1,\dots,0,\\ \hat P_h(nh,y_i,v) = \max\Big\{\Psi(y_i,v), e^{-rh}\sum_{j\in\mathcal{J}_{M_h}}\Pi^h(v)_{i,j}\mathbb{E}\Big(\hat P_h\Big((n+1)h, y_j, V^{nh,v}_{(n+1)h}\Big)\Big)\Big\},\ i\in\mathcal{J}_{M_h}.\end{cases} \tag{2.18}$$

We stress that the backward induction (2.18) is now written in terms of the process V only, and here the binomial tree method in [3] comes on, we see how in a moment. First, let us briefly recall the binomial tree procedure in [3].

For $n = 0, 1, \dots, N$, consider the lattice

$$\mathcal{V}_n^h = \{v_{n,k}\}_{k=0,1,\dots,n} \quad \text{with} \quad v_{n,k} = \Big(\sqrt{V_0} + \frac{\sigma}{2}(2k-n)\sqrt{h}\Big)^2 \mathbb{1}_{\sqrt{V_0}+\frac{\sigma}{2}(2k-n)\sqrt{h}>0} \tag{2.19}$$

(notice that $v_{0,0} = V_0$) and for each fixed $v_{n,k} \in \mathcal{V}_n^h$, we define

$$k_d^h(n,k) = \max\{k^* \,:\, 0 \le k^* \le k \text{ and } v_{n,k} + \mu_V(v_{n,k})h \ge v_{n+1,k^*}\}, \tag{2.20}$$

$$k_u^h(n,k) = \min\{k^* \,:\, k+1 \le k^* \le n+1 \text{ and } v_{n,k} + \mu_V(v_{n,k})h \le v_{n+1,k^*}\} \tag{2.21}$$

with the understanding $k_d^h(n,k) = 0$ if $\{k^* \,:\, 0 \le k^* \le k \text{ and } v_{n,k} + \mu_V(v_{n,k})h \ge v_{n+1,k^*}\} = \emptyset$ and $k_u^h(n,k) = n+1$ if $\{k^* \,:\, k+1 \le k^* \le n+1 \text{ and } v_{n,k} + \mu_V(v_{n,k})h \le v_{n+1,k^*}\} = \emptyset$. The transition probabilities are defined as follows: starting from the node (n,k) the probability that the process jumps to $k_u^h(n,k)$ at time-step $n+1$ is set as

$$p^h_{k_u^h(n,k)} = 0 \vee \frac{\mu_V(v_{n,k})h + v_{n,k} - v_{n+1,k_d^h(n,k)}}{v_{n+1,k_u^h(n,k)} - v_{n+1,k_d^h(n,k)}} \wedge 1. \tag{2.22}$$

And of course, the jump to $(n+1, k_d^h(n,k))$ happens with probability $p^h_{k_d^h(n,k)} = 1 - p^h_{k_u^h(n,k)}$. This gives rise to a Markov chain $(\bar V_n^h)_{n=0,\dots,N}$ that weakly converges, as $h \to 0$, to the diffusion process $(V_t)_{t\in[0,T]}$ and turns out to be a robust tree approximation for the CIR process V. This means that we can approximate the expectation of suitable functionals of the diffusion V with the same expectation evaluated on the Markov chain $\bar V^h$. In particular, for a function g we write

$$\mathbb{E}\Big(g(V^{nh,v_{n,k}}_{(n+1)h})\Big) \simeq \mathbb{E}\Big(g(\bar V^h_{n+1}) \,|\, \bar V^h_n = v_{n,k}\Big) = g(v_{n+1,k_u^h(n,k)})p^h_{k_u^h(n,k)} + g(v_{n+1,k_d^h}(n,k))p^h_{k_d^h(n,k)}.$$

So, at step n, we can numerically compute the expectation in the backward induction (2.18) on the lattice $\mathcal{V}_n^h$ as

$$\mathbb{E}\Big(\hat P_h\Big((n+1)h, y_j, V^{nh,v}_{(n+1)h}\Big)\Big)\Big|_{v=v_{n,k}} \simeq \sum_{k^*\in\{k_u^h(n,k),k_d^h(n,k)\}} \hat P_h\Big((n+1)h, y_j, v_{n+1,k^*}\Big)p^h_{k^*}.$$

We can finally write the backward induction giving our approximating algorithm: for $n = 0, 1, \dots, N$, we define $\tilde P_h(nh, y, v)$ for $(y,v) \in \mathcal{Y}^h_{M_h} \times \mathcal{V}^h_n$ by

$$\begin{cases} \tilde P_h(T, y_i, v_{N,k}) = \Psi(y_i, v_{N,k}) \quad i \in \mathcal{J}_{M_h} \text{ and } v_{N,k} \in \mathcal{V}_n^h, \text{ and as } n = N-1, \dots, 0 \\ \tilde P_h(nh, y_i, v_{n,k}) = \max\Big\{\Psi(y_i, v), e^{-rh}\displaystyle\sum_{k^*,j} \Pi^h(v_{n,k})_{i,j}\tilde P_h\Big((n+1)h, y_j, v_{n+1,k^*}\Big)p^h_{k^*}\Big\}, \\ \qquad\qquad i \in \mathcal{J}_{M_h} \quad \text{and} \quad v_{n,k} \in \mathcal{V}_n^h, \end{cases} \tag{2.23}$$

where the sum above is done for $k^* \in \{k_u^h(n,k), k_d^h(n,k)\}$ and $j \in \mathcal{J}_{M_h}$. Notice that, at time step n, for every fixed $i \in \mathcal{J}_{M_h}$ and $k = 0, \dots, n$ the sum in (2.23) can be seen as an integral w.r.t. the measure

$$\mu^h(y_i, v_{n,k}; A \times B) = \sum_{k^* \in \{k_u^h(n,k), k_d^h(n,k)\}} \sum_{j \in \mathcal{J}_{M_h}} \Pi^h(v_{n,k})_{i,j} p_{k^*}^h \delta_{\{y_j\}}(A) \delta_{\{v_{n+1,k^*}\}}(B) \qquad (2.24)$$

for every Borel sets A and B, $\delta_{\{a\}}$ denoting the Dirac mass in a, so that $\mu^h(y_i, v_{n,k}; \cdot)$ is a discrete measure on $\mathcal{Y}_{M_h}^h \times \mathcal{V}_{n+1}^h$.

Now, in next Section 3 we shall be able to prove that, for small values of h, $\Pi^h(v)$ is a stochastic matrix. This gives that $\mu^h(y_i, v_{n,k}; \cdot)$ is actually a probability measure, that can be interpreted as a transition probability measure. Thus, we are doing expectations on a Markov chain $(\bar{Y}_n^h, \bar{V}_n^h)_{n=0,1,\dots,N}$, whose state-space, at time step n, is given by $\mathcal{Y}_{M_h}^h \times \mathcal{V}_n^h$ and whose transition probability measure at time step n is given by $\mu^h(y_i, v_{n,k}; \cdot)$ in (2.24). Moreover, we shall prove that, under appropriate conditions on Δy_h and M_h such that, as $h \to 0$, $\Delta y_h \to 0$ and $M_h \Delta y_h \to \infty$ (see next (3.21) and (3.22)), the family of Markov chains $(\bar{Y}^h, \bar{V}^h)_h$ weakly converges to the diffusion process (Y, V) . And this gives the convergence of our hybrid tree-finite difference algorithm approximating the Heston model. We shall finally discuss the convergence of the backward induction (2.23) to the price of the associated American option written on the Heston model.

3 The convergence of the algorithm

We first set up the finite difference method we take into account. Then, in Section 3.2, we formally define the approximating Markov chain and prove the weak convergence to the Heston model in the path space.

For ease of notation, for a while we drop the dependence on the time-step h for the space-step Δy_h and the number M_h related to the points of the space-grid, so we simply write Δy and M, as well as $\mathcal{J}_M = \{-M, \dots, M\}$ and $\mathcal{Y}_M = \{y_i = Y_0 + i\Delta y\}_{i \in \mathcal{J}_M}$.

3.1 The finite difference scheme for the PDE problem (2.14)

As described in Section 2.2, at each time step n we need to numerically solve (2.14) for $t = t_n = nh$. So, we briefly describe the finite difference method we apply to problem (2.14), outlining some key properties of the associated operator allowing us to prove the convergence. For further information on finite difference methods for partial differential equations we refer for instance to [28].

Let $t = nh$, v and $v^* = g_{nh+h}$ be fixed and let us set $u_j^n = \tilde{u}_n^h(nh, y_j)$ the discrete solution of (2.14) at time nh on the point y_j of the grid $\mathcal{Y}_M$ - for simplicity of notations, we do not stress on u_j^n the dependence on v (from the coefficients of the PDE), v^* (from the Cauchy problem) and h.

It is well known that the behavior of the solution to (2.14) changes with respect the magnitude of the rate between the diffusion coefficient $(\rho^2 v/2)$ and the advection term $(\mu_Y(v))$. To deal with these cases, we fix a small real threshold $\epsilon > 0$ and in the following we shall describe how

to solve both the case $v < \epsilon$ and $v > \epsilon$ by applying an explicit in time and an implicit in time approximation respectively.
It is well known that for a big enough diffusion coefficient, to avoid over-restrictive conditions on the grid steps, it is suggested to apply implicit finite differences to problem (2.14). In this case, the discrete solution $\{u_j^n\}_{j\in\mathbb{Z}}$ at time nh will then be computed in terms of the solution $\{u_j^{n+1}\}_{j\in\mathbb{Z}}$ at time $(n+1)h$ by solving the following discrete problem:

$$\frac{u_j^{n+1}-u_j^n}{h} + \mu_Y(v)\frac{u_{j+1}^n - u_{j-1}^n}{2\Delta y} + \frac{1}{2}\bar{\rho}^2\ v\ \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\Delta y^2} = 0, \quad j \in \mathbb{Z} \tag{3.1}$$

where $\Delta y = y_j - y_{j-1}$, $\forall j \in \mathbb{Z}$.
On the other hand, when the diffusion coefficient is small compared with the reaction one, it is suggested to apply an explicit in time approximation coupled with a forward or backward finite differences for the first order term u_x depending on the sign of the reaction coefficient.
Specifically, for v close to 0, that is $v < \epsilon$, we solve the problem by the following approximation schemes: when $\mu_Y(v) > 0$,

$$\frac{u_j^{n+1}-u_j^n}{h} + \mu_Y(v)\frac{u_{j+1}^{n+1} - u_{j}^{n+1}}{\Delta y} + \frac{1}{2}\bar{\rho}^2\ v\ \frac{u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}}{\Delta y^2} = 0, \quad j \in \mathbb{Z}, \tag{3.2}$$

while, when $\mu_Y(v) < 0$,

$$\frac{u_j^{n+1}-u_j^n}{h} + \mu_Y(v)\frac{u_{j}^{n+1} - u_{j-1}^{n+1}}{\Delta y} + \frac{1}{2}\bar{\rho}^2\ v\ \frac{u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}}{\Delta y^2} = 0, \quad j \in \mathbb{Z}. \tag{3.3}$$

As previously mentioned at the end of the Section 2.1, for the numerical tests one has to deal with a finite grid $\mathcal{Y}_M^h = \{y_j\}_{j\in\mathcal{J}_M}$ (for simplicity, here we simply write M instead of M_h) and problems (3.1) and (3.2) have to be coupled with suitable numerical boundary conditions. Here, we assume that the two-off the domain values are defined by the following relations (Neumann-type boundary conditions): in the implicit case we set

$$u_{-M-1}^n = u_{-M+1}^n, \quad u_{M+1}^n = u_{M-1}^n, \tag{3.4}$$

whereas in the explicit case we set

$$u_{-M-1}^{n+1} = u_{-M+1}^{n+1}, \quad u_{M+1}^{n+1} = u_{M-1}^{n+1} \tag{3.5}$$

Other conditions can surely be selected, for example the two *boundary* values u_{-M}^{n+1} and u_M^{n+1} may be a priori fixed by a known constant (this typically appears in financial problems). All arguments that follow apply as well.
So, hereafter we set

$$\alpha = \frac{h}{2\Delta y}\,\mu_Y(v) \quad \text{and} \quad \beta = \frac{h}{2\Delta y^2}\,\bar{\rho}^2 v. \tag{3.6}$$

3.1.1 The case $v > \epsilon$

By applying implicit finite differences (3.1) coupled with boundary conditions (3.4), we get the solution $u^n = (u^n_{-M}, \dots, u^n_M)^T$ by solving the following linear system

$$A\,u^n = u^{n+1}, \tag{3.7}$$

where A is the $(2M+1) \times (2M+1)$ tridiagonal real matrix given by

$$A = \begin{pmatrix} 1+2\beta & -2\beta & & & \\ \alpha-\beta & 1+2\beta & -\alpha-\beta & & \\ & \ddots & \ddots & \ddots & \\ & & \alpha-\beta & 1+2\beta & -\alpha-\beta \\ & & & -2\beta & 1+2\beta \end{pmatrix}. \tag{3.8}$$

We immediately prove that the solution u^n to (3.7) actually exists at least when $\beta > |\alpha|$ (we will see later that this is not a restrictive condition).

Proposition 3.1. *Assume that $\beta > |\alpha|$. Then A is invertible and A^{-1} is a stochastic matrix, that is all entries are non negative and, for $\mathbf{1} = (1, ..., 1)^T$, $A\mathbf{1} = \mathbf{1}$.*

Proof. The matrix $A = (a_{ij})_{i,j\in\mathcal{J}_M}$ satisfies

(P1) $A\mathbf{1} = \mathbf{1}$, i.e. $\sum_{j=-M}^{M} a_{ij} = 1$ for $i \in \mathcal{J}_M$

and for $\beta > |\alpha|$, one has also

(P2) $a_{ii} > 0$ for $i \in \mathcal{J}_M$ and for $j \in \mathcal{J}_M$, $j \neq i$, $a_{ij} \leq 0$,

(P3) A is strict or irreducibly diagonally dominant, i.e. $\sum_{j\in\mathcal{J}_M, j\neq i} |a_{ij}| < a_{ii}$ for $i \in \mathcal{J}_M$.

(P2)-(P3) give that A is an invertible M-matrix (see for instance [4]), so that A^{-1} is non-negative (i.e. $a^{-1}_{ij} \geq 0$, $i, j = -M, ..., M$). Moreover, by (P1), $\mathbf{1} = A^{-1}\mathbf{1}$. □

For each $l \in \mathbb{N}$ and $y_i \in \mathcal{Y}_M$, we consider the polynomial $(y - y_i)^l$ and we call $\psi^i_l(y) \in \mathbb{R}^{2M+1}$ the associated (vector) function of $y \in \mathcal{Y}_M$:

$$\left(\psi^i_l(y)\right)_k = (y_k - y_i)^l = \Delta y^l (k-i)^l, \quad k \in \mathcal{J}_M. \tag{3.9}$$

In next Section 3.2 we need to deal with $A^{-1}\psi^i_l(y)$ for $l \leq 4$ and $i \in \mathcal{J}_M$. So, we study such objects. By Proposition 3.1, for $\beta > |\alpha|$ one has that A is invertible and we may then compute $A^{-1}\psi^i_l(y)$. We also notice that $\psi^i_0(y) = \mathbf{1}$, so that $A^{-1}\psi^i_0(y) = A^{-1}\mathbf{1} = \mathbf{1}$.

In the following, the symbol $[\cdot]$ will stand for the floor function and we use the understanding $\sum_{k=1}^{0} (\cdot)_k := 0$. Moreover, we let $\mathbf{e}_i$ denote the standard orthonormal basis: for $i, k \in \mathcal{J}_M$, $(\mathbf{e}_i)_k = 0$ for $k \neq i$ and $(\mathbf{e}_i)_k = 1$ if $k = i$.

Lemma 3.2. *Let $\psi^i_l(y)$ be defined in (3.9). Then for every $l \in \mathbb{N}$ and $i \in \mathcal{J}_M$ one has*

$$A\psi^i_l(y) = \psi^i_l(y) - \sum_{j=0}^{l-1} \binom{l}{j} a_{l-j} \Delta y^{l-j} \psi^i_j(y) + b^{-M}_{l,i} \mathbf{e}_{-M} + b^{M}_{l,i} \mathbf{e}_M, \tag{3.10}$$

where

$$a_n = (\beta - \alpha)(-1)^n + (\beta + \alpha), \quad n \in \mathbb{N}, \tag{3.11}$$

that is $a_n = 2\beta$ *if* n *is even and* $a_n = 2\alpha$ *if* n *is odd, and the coefficients* $b_{l,i}^{\pm M}$ *are given by*

$$b_{l,i}^{\pm M} = \pm 2 \sum_{j=0, l-j \ odd}^{l-1} \binom{l}{j} (\beta \pm \alpha) \Delta y^{l-j} (y_{\pm M} - y_i)^j. \tag{3.12}$$

Moreover, $b_{l,i}^{\pm M}$ *can be bounded as follows:*

$$|b_{l,i}^{\pm M}| \leq 2(\beta \pm \alpha)(\Delta y + |y_{\pm M} - y_i|)^l \tag{3.13}$$

Proof. For $k \in \mathcal{J}_M$ with $k \neq \pm M$, we have

$$\begin{aligned}
\Big(A\psi_l^i(y)\Big)_k &= -(\beta - \alpha)\Big(\psi_l^i(y)\Big)_{k-1} + (1 + 2\beta)\Big(\psi_l^i(y)\Big)_k - (\beta + \alpha)\Big(\psi_l^i(y)\Big)_{k+1} \\
&= -(\beta - \alpha)(y_k - y_i - \Delta y)^l + (1 + 2\beta)(y_k - y_i)^l - (\beta + \alpha)(y_k - y_i + \Delta y)^l \\
&= (1 + 2\beta)(y_k - y_i)^l - \sum_{j=0}^{l} \binom{l}{j} (y_k - y_i)^j \Big((\beta - \alpha)(-1)^{l-j} + (\beta + \alpha)\Big)\Delta y^{l-j} \\
&= (y_k - y_i)^l - \sum_{j=0}^{l-1} \binom{l}{j} (y_k - y_i)^j a_{l-j} \Delta y^{l-j}.
\end{aligned}$$

On the other hand, following the same reasonings for $k = \pm M$ we easily obtain

$$\Big(A\psi_l^i(y)\Big)_{\pm M} = (y_{\pm M} - y_i)^l - 2\beta \sum_{j=0}^{l-1} \binom{l}{j} (\mp \Delta y)^{l-j} (y_{\pm M} - y_i)^j$$

and by using (3.12) we get the full form (3.10). Finally, the estimate in (3.13) immediately follows from the Newton's binomial formula. $\square$

We are now ready to characterize $A^{-1}\psi_l^i(y)$, for every polynomial $\psi_l^i(y)$ as in (3.9).

Proposition 3.3. *Suppose that* $\beta > |\alpha|$ *and for* $l \geq 1$ *let* $\gamma_{l,k}$, $k = 0, 1, \ldots, l-1$, *be iteratively (backwardly) defined as follows:*

$$\gamma_{l,k} = \binom{l}{k} a_{l-k} \Delta y^{l-k} + \sum_{j=k+1}^{l-1} \gamma_{l,j} \binom{j}{k} a_{j-k} \Delta y^{j-k}, \quad k = l-1, \ldots, 0, \tag{3.14}$$

where the coefficients a_n *are given in* (3.11). *Then*

$$A^{-1}\psi_l^i(y) = \psi_l^i(y) + \sum_{j=0}^{l-1} \gamma_{l,j} \psi_j^i(y) + \phi_{l,M}^i(y),$$

in which

$$\phi_{l,M}^i(y) = T_{l,i}^{-M} A^{-1} \mathbf{e}_{-M} + T_{l,i}^{M} A^{-1} \mathbf{e}_M \quad \text{with} \quad T_{l,i}^{\pm M} = -b_{l,i}^{\pm M} - \sum_{j=0}^{l-1} \gamma_{l,j} b_{j,i}^{\pm M}, \tag{3.15}$$

the $b_{j,i}^{\pm M}$*'s being given in* (3.12). *Moreover, if* $\Delta y \leq 1$, $M\Delta y \geq 1$ *and* $l 2^{l+2}(\beta \Delta y^2 + |\alpha| \Delta y) \leq 1$, *the following estimate holds for* $T_{l,i}^{\pm M}$*: for every* $i \in \mathcal{J}_M$,

$$|T_{l,i}^{\pm M}| \leq 4(\beta \pm \alpha) \Delta y^l (1 + 2M)^l. \tag{3.16}$$

Proof. It is clear that $A^{-1}\psi_l^i(y) = \psi_l^i(y) + \sum_{j=1}^{l-1} \gamma_{l,j}\psi_j^i(y) + \phi_l^i(y)$ if and only if

$$\psi_l^i(y) = A\psi_l^i(y) + \sum_{j=0}^{l-1} \gamma_{l,j} A\psi_j^i(y) + A\phi_{l,M}^i(y).$$

We call $(*)$ the r.h.s. above. By using Lemma 3.2 and setting

$$\theta_l^i = b_{l,i}^{-M}\mathbf{e}_{-M} + b_{l,i}^{M}\mathbf{e}_M,$$

one has

$$\begin{aligned}
(*) =&\psi_l^i(y) - \sum_{k=0}^{l-1} \binom{l}{k} a_{l-k}\Delta y^{l-k}\psi_k^i(y) + \theta_l^i + \\
&+ \sum_{j=0}^{l-1} \gamma_{l,j}\Big(\psi_j^i(y) - \sum_{k=0}^{j-1}\binom{j}{k} a_{j-k}\Delta y^{j-k}\psi_k^i(y) + \theta_j^i\Big) + A\phi_{l,M}^i(y) \\
=&\psi_l^i(y) - \sum_{k=0}^{l-1} \binom{l}{k} a_{l-k}\Delta y^{l-k}\psi_k^i(y) + \sum_{j=0}^{l-1}\gamma_{l,j}\psi_j^i(y) \\
&- \sum_{j=0}^{l-1}\gamma_{l,j}\sum_{k=0}^{j-1}\binom{j}{k} a_{j-k}\Delta y^{j-k}\psi_k^i(y) + \theta_l^i + \sum_{j=0}^{l-1}\gamma_{l,j}\theta_j^i + A\phi_{l,M}^i(y) \\
=&\psi_l^i(y) + \sum_{k=0}^{l-1}\Big(- \binom{l}{k} a_{l-k}\Delta y^{l-k} + \gamma_{l,k} - \sum_{j=k+1}^{l-1}\gamma_{l,j}\binom{j}{k} a_{j-k}\Delta y^{j-k}\Big)\psi_k^i(y) + \\
&+ \theta_l^i + \sum_{j=0}^{l-1}\gamma_{l,j}\theta_j^i + A\phi_{l,M}^i(y).
\end{aligned}$$

By the definition of the $\gamma_{j,k}$'s and $\phi_{l,M}^i(y)$'s, each coefficients in the above (first) sum is null and the sum of the last three terms is zero, so that $(*) = \psi_l^i(y)$. Let us discuss (3.16). By using (3.13) and the fact that $|y_{\pm M} - y_i| \le 2M\Delta y$, since $\Delta y(1+2M) \ge 1$ we can write

$$|T_{l,i}^{\pm M}(y)| \le 2(\beta \pm \alpha)\Delta y^l (1+2M)^l \Big(1 + \sum_{j=0}^{l-1}|\gamma_{l,j}|\Big).$$

It remains to prove that, under our constraints, $\sum_{j=0}^{l-1}|\gamma_{l,j}| \le 1$. For every $k = 0, 1, \dots, l-1$ and $j = k+1, \dots, l-1$ we consider the estimates

$$\binom{l}{k} \le 2^l \quad \text{and} \quad \binom{j}{k} \le 2^l.$$

By inserting in (3.14) we can write

$$|\gamma_{l,k}| \le 2^l |a_{l-k}|\Delta y^{l-k} + 2^l \sum_{j=k+1}^{l-1} |\gamma_{l,j}|\,|a_{j-k}|\Delta y^{j-k}, \quad k = l-1, \dots, 0.$$

We also notice that, for $m \ge 1$ and $\Delta y \le 1$,

$$|a_m|\Delta y^m \le 2(\beta\Delta y^2 + |\alpha|\Delta y)$$

so that we get

$$|\gamma_{l,k}| \leq 2^{l+1}(\beta\Delta y^2 + |\alpha|\Delta y) + 2^{l+1}(\beta\Delta y^2 + |\alpha|\Delta y)\sum_{j=k+1}^{l-1}|\gamma_{l,j}|, \quad k = l-1,\dots,0.$$

Now, if $l2^{l+2}(\beta\Delta y^2 + |\alpha|\Delta y) \leq 1$ we get

$$|\gamma_{l,k}| \leq \frac{1}{2l} + \frac{1}{2l}\sum_{j=k+1}^{l-1}|\gamma_{l,j}|, \quad k = l-1,\dots,0$$

and by summing over k,

$$\sum_{k=0}^{l-1}|\gamma_{l,k}| \leq \frac{1}{2} + \frac{1}{2l}\sum_{k=0}^{l-1}\sum_{j=k+1}^{l-1}|\gamma_{l,j}| = \frac{1}{2} + \frac{1}{2l}\sum_{j=1}^{l-1}\sum_{k=0}^{j-1}|\gamma_{l,j}| \leq \frac{1}{2} + \frac{1}{2}\sum_{j=0}^{l-1}|\gamma_{l,j}|$$

from which it follows that $\sum_{k=0}^{l-1}|\gamma_{l,k}| \leq 1$ and the statement holds. □

Remark 3.4. *For further use, we write down explicitly the vector $A^{-1}\psi_l^i(y)$ for $l = 1,2,4$. Straightforward computations give the following:*

$$\begin{aligned}
A^{-1}\psi_1^i(y) =&\psi_1^i(y) + 2\alpha\Delta y\mathbf{1} + \phi_{1,M}^i(y),\\
A^{-1}\psi_2^i(y) =&\psi_2^i(y) + 4\alpha\Delta y\psi_1^i(y) + 2(\beta+2\alpha)\Delta y^2\mathbf{1} + \phi_{2,M}^i(y),\\
A^{-1}\psi_4^i(y) =&\psi_4^i(y) + 8\alpha\Delta y\psi_3^i(y) + 12(\beta+4\alpha^2)\Delta y^2\psi_2^i(y) + 8(\alpha+12\alpha^2+18\alpha\beta)\Delta y^3\psi_1^i(y)+\\
&+2(\beta+16\alpha^2+96\alpha^3+12\beta^2+192\alpha^3\beta)\Delta y^4\mathbf{1} + \phi_{4,M}^i(y).
\end{aligned}$$

In particular, since $(\psi_l^i(y))_i = 0$ for every $l \geq 1$, the ith entry of the above sequences are given by

$$\begin{aligned}
(A^{-1}\psi_1^i(y))_i &= 2\alpha\Delta y + (\phi_{1,M}^i(y))_i,\\
(A^{-1}\psi_2^i(y))_i &= 2(\beta+2\alpha)\Delta y^2 + (\phi_{2,M}^i(y))_i,\\
(A^{-1}\psi_4^i(y))_i &= 2(\beta+16\alpha^2+96\alpha^3+12\beta^2+192\alpha^3\beta)\Delta y^4 + (\phi_{4,M}^i(y))_i.
\end{aligned}$$

By replacing the α and β expressions (3.6), we get the formulas

$$\begin{aligned}
(A^{-1}\psi_1^i(y))_i =& \; h\mu_Y(v) + (\phi_{1,M}^i)_i,\\
(A^{-1}\psi_2^i(y))_i =& \; h\bar\rho^2 v + 2h\Delta y\mu_Y(v) + (\phi_{2,M}^i)_i,\\
(A^{-1}\psi_4^i(y))_i =& \; h\Delta y^2\bar\rho^2 v + 8h^2\Delta y^2\mu_Y(v)^2 + 24h^3\mu_Y(v)^3 + 6h^2\bar\rho^4 v^2+\\
&+24\tfrac{h^4}{\Delta y}\,\bar\rho^2 v\mu_Y(v)^3 + (\phi_{4,M}^i)_i.
\end{aligned} \tag{3.17}$$

Furthermore, to deal with the numerical boundary conditions, as those given in (3.4), we need to study the behavior of the ith component of the boundary term $\phi_{l,M}^i(y)$ in (3.15) as i is "far from the boundary" and $l = 1,2,4$. Here, we use a quite general result (allowing one to set up different boundary conditions) whose precise statement and proof are postponed in Appendix A.1.

Proposition 3.5. *Suppose that $\beta > |\alpha|$. Let $l \in \mathbb{N}$, $i \in \mathcal{J}_M$ and let $\phi^i_{l,M}(y)$ denote the boundary term in* (3.15). *Assume that $\Delta y \leq 1$, $M\Delta y \geq 1$ and $l2^{l+2}(\beta\Delta y^2 + |\alpha|\Delta y) \leq 1$. Then one has*

$$|(\phi^i_{l,M}(y))_i| \leq 8(\beta + |\alpha|)\Delta y^l (1 + 2M)^l \Big(1 - \frac{1}{1 + \beta + |\alpha|}\Big)^{M-i}.$$

Proof. Since $\beta > |\alpha|$, A satisfies the requirements in Proposition A.1. So, we use such result, that has been specialized to our type of matrix in Remark A.2: taking $a = 1 + 2\beta$, $b = -\beta + \alpha$ and $c = -\beta - \alpha$, we obtain

$$|(A^{-1}\mathbf{e}_{\pm M})_i| \leq \Big(\frac{\beta \pm \alpha}{\gamma^*_{\pm M}}\Big)^{M-i}$$

where

$$\gamma^*_{\pm M} = \min\Big(1 + 2\beta - \frac{2\beta(\beta \mp \alpha)}{1 + 2\beta}, \frac{1 + 2\beta + \sqrt{(1 + 2\beta)^2 - 4(\beta^2 - \alpha^2)}}{2}\Big).$$

Straightforward computations give $\gamma^*_{\pm M} \geq 1 + \beta \pm \alpha$, so that

$$|(A^{-1}\mathbf{e}_{\pm M})_i| \leq \Big(\frac{\beta \pm \alpha}{1 + \beta \pm \alpha}\Big)^{M-i}.$$

Now, since $\beta > |\alpha|$ we can write $\frac{\beta \pm \alpha}{1+\beta\pm\alpha} = 1 - \frac{1}{1+\beta\pm\alpha} \leq 1 - \frac{1}{1+\beta+|\alpha|}$, so that

$$|(A^{-1}\mathbf{e}_{\pm M})_i| \leq \Big(1 - \frac{1}{1 + \beta + |\alpha|}\Big)^{M-i}.$$

We insert now the above estimate and estimate (3.16) in (3.15), and we get the result. □

3.1.2 The case $v < \epsilon$

Here we go through the already described procedure for the explicit in time approximation. We recall that, for $v < \epsilon$, we consider the forward finite differences (3.2) or the backward finite differences (3.3) for the first order term depending on the sign of the reaction coefficient: $\mu_Y(v) > 0$ or $\mu_Y(v) < 0$ respectively, and from (3.6) this means that $\alpha > 0$ or $\alpha < 0$ respectively. So, by considering also the case $\alpha = 0$ and by coupling with the boundary conditions (3.5), straightforward computations give that the solution u^n of the explicit in time scheme satisfies the condition

$$u^n = Cu^{n+1}$$

where

$$C = \begin{pmatrix} 1-2\beta-2|\alpha| & 2\beta+2|\alpha| & & & \\ \beta+2|\alpha|1\!\!1_{\{\alpha<0\}} & 1-2\beta-2|\alpha| & \beta+2|\alpha|1\!\!1_{\{\alpha>0\}} & & \\ & \ddots & \ddots & \ddots & \\ & & \beta+2|\alpha|1\!\!1_{\{\alpha<0\}} & 1-2\beta-2|\alpha| & \beta+2|\alpha|1\!\!1_{\{\alpha>0\}} \\ & & & 2\beta+2|\alpha| & 1-2\beta-2|\alpha| \end{pmatrix}, \tag{3.18}$$

α and β being given in (3.6) and $1\!\!1$ denoting the indicator function ($1\!\!1_\Gamma = 1$ on Γ and $1\!\!1_\Gamma = 0$ otherwise). We remark that C is a stochastic matrix if and only if

$$2\beta + 2|\alpha| \leq 1. \tag{3.19}$$

We also notice that if ϵ is small enough such a condition is not restrictive, but we will discuss deeper this point later.
In next Section 3.2 we need to deal with $C\psi_l^i(y)$ for $l \leq 4$ and $i \in \mathcal{J}_M$, where the function $\psi_l^i(y) \in \mathbb{R}^{2M+1}$ are defined in (3.9). So, we get

Lemma 3.6. *Let $\psi_l^i(y)$ be defined in (3.9). Then for every $l \in \mathbb{N}$ and $i \in \mathcal{J}_M$ one has*

$$C\psi_l^i(y) = \psi_l^i(y) + \sum_{j=0}^{l-1} \binom{l}{j} d_{l-j} \Delta y^{l-j} \psi_j^i(y) + c_{l,i}^{-M} \mathbf{e}_{-M} + c_{l,i}^{M} \mathbf{e}_M,$$

where

$$d_n = (-1)^n(\beta + 2|\alpha| 1\!\!1_{\{\alpha<0\}}) + \beta + 2|\alpha| 1\!\!1_{\{\alpha>0\}}, \quad n \in \mathbb{N},$$

that is $d_n = 2(\beta + |\alpha|)$ if n is even and $d_n = 2\alpha$ if n is odd, and $c_{l,i}^{\pm M}$ are given by

$$c_{l,i}^{\pm M} = 2(\beta + 2|\alpha| 1\!\!1_{\pm\alpha>0}) \sum_{j=0,\, l-j \in \mathcal{A}_\pm}^{l-1} \binom{l}{j} \Delta y^{l-j} (y_{\pm M} - y_i)^j$$

where $\mathcal{A}_+$ and $\mathcal{A}_-$ denotes the set of the even and the odd numbers respectively.

The proof is straightforward (it suffices to follow the same arguments developed for Lemma 3.2), so we omit it.

Remark 3.7. *In the special case $l = 1, 2, 4$, Lemma 3.6 gives*

$$\begin{aligned}
C\psi_1^i(y) = &\ \psi_1^i(y) + 2\alpha\Delta y \mathbf{1} + c_{1,i}^{-M} \mathbf{e}_{-M} + c_{1,i}^{M} \mathbf{e}_M, \\
C\psi_2^i(y) = &\ \psi_2^i(y) + 4\alpha\Delta y \psi_1^i(y) + 2(\beta + |\alpha|)\Delta y^2 \mathbf{1} + c_{2,i}^{-M} \mathbf{e}_{-M} + c_{2,i}^{M} \mathbf{e}_M \\
C\psi_4^i(y) = &\ \psi_4^i(y) + 8\alpha\Delta y \psi_3^i(y) + 12(\beta + |\alpha|)\Delta y^2 \psi_2^i(y) + 8\alpha\Delta y^3 \psi_1^i(y) + 2(\beta + |\alpha|)\Delta y^4 \mathbf{1} + \\
&+ c_{4,i}^{-M} \mathbf{e}_{-M} + c_{4,i}^{M} \mathbf{e}_M.
\end{aligned}$$

In particular, since $(\psi_l^i(y))_i = 0$ for every $l \geq 1$ and assuming $|i| < M$, the ith entries of the above sequences are given by

$$\begin{aligned}
(C\psi_1^i(y))_i &= h\mu_Y(v), \\
(C\psi_2^i(y))_i &= h\bar{\rho}^2 v + h\Delta y |\mu_Y(v)|, \\
(C\psi_4^i(y))_i &= h\Delta y^2 \bar{\rho}^2 v + h\Delta y^3 |\mu_Y(v)|,
\end{aligned} \tag{3.20}$$

in which we have inserted the formulas for α and β in (3.6),

3.2 The associated Markov chain and the convergence of the hybrid algorithm

We denote, as in Section 2.2, $(\bar{V}_n^h)_{n=0,1,\dots,N}$, with $Nh = T$, the Markov chain approximating the volatility process V introduced in [3]. We recall that the state-space at step n is given by $\mathcal{V}_n^h$ defined in (2.19). We define now the Y-component of our Markov chain.
First, given the time-step h, we set up the dependence on h for the space-step Δy, the number M giving the points of the grid $\mathcal{Y}_M = \{y_i = Y_0 + i\Delta y\,;\, i = -M,\dots,M\}$ and the threshold ϵ that allows us to use the explicit or the implicit finite difference method. So, we assume that

$$\Delta y \equiv \Delta y_h = c_y h^p, \qquad M \equiv M_h = c_M h^{-q}, \qquad \epsilon \equiv \epsilon_h = c_\epsilon h^p \tag{3.21}$$

where $c_M > 0$ and the constants $c_y, c_\epsilon, p, q > 0$ are chosen as follows

$$\begin{aligned} &p < 1, \qquad q > p, \qquad \frac{2c_y}{\bar\rho^2}\left|r - \delta - \frac{\rho}{\sigma}\kappa\theta\right| < c_\epsilon, \quad \text{or}\\ &p = 1, \qquad q > p, \qquad \frac{2c_y}{\bar\rho^2}\left|r - \delta - \frac{\rho}{\sigma}\kappa\theta\right| < c_\epsilon < \left(\frac{1}{2} - \frac{1}{c_y}\left|r - \delta - \frac{\rho}{\sigma}\kappa\theta\right|\right)\frac{c_y^2}{\bar\rho^2}, \end{aligned} \tag{3.22}$$

in which the parameters $\kappa, \theta, r, \delta, \sigma$ and ρ come from our model, see Section 2. Let us stress that the last constraint in (3.22) can be really satisfied, for example by choosing $c_y > 4|r - \delta - \frac{\rho}{\sigma}\kappa\theta|$. We also notice that (3.21) and (3.22) give $M_h \Delta y_h = O(h^{-q+p}) \to \infty$ as $h \to 0$, so that $\mathcal{Y}_{M_h}^h = \{y_i = Y_0 + i\Delta y_h\,;\, i = -M_h,\dots,M_h\} \uparrow \mathbb{R}$ as $h \to 0$. Moreover, one has

Proposition 3.8. *Let $\beta = \beta_h$ and $\alpha = \alpha_h$ be given in (3.6) with the constraints (3.21) and (3.22). Then there exists $h_0 > 0$ such that for every $h < h_0$ one has*

i) if $v > \epsilon_h$ then $\beta_h > |\alpha_h|$;

ii) if $v \le \epsilon_h$ then $2\beta_h + 2|\alpha_h| < 1$.

Proof. By formula (2.7), we write $|\mu_Y(v)| \le a_Y + b_Y v$, with $a_Y = |r - \delta - \frac{\rho}{\sigma}\kappa\theta|$ and $b_Y = |\frac{\rho}{\sigma}\kappa - \frac{1}{2}|$.
i) One has $\beta_h = \frac{h^{1-2p}}{2c_y^2}\bar\rho^2 v$ and $|\alpha_h| \le \frac{h^{1-p}}{2c_y}(a_Y + b_Y v)$, so $\beta_h > |\alpha_h|$ if $\frac{h^{1-2p}}{2c_y^2}\bar\rho^2 v > \frac{h^{1-p}}{2c_y}(a_Y + b_Y v)$. But this holds if $v(\frac{\bar\rho^2}{c_y} - h^p b_Y) > a_Y h^p$, which in turn is implied by $v > \epsilon_h$ when h is small enough.
ii) One has $2\beta_h + 2\alpha_h \le \frac{h^{1-2p}}{c_y^2}\bar\rho^2 c_\epsilon h^p + \frac{h^{1-p}}{c_y}(a_Y + b_Y c_\epsilon h^p) = h^{1-p}(\frac{\bar\rho^2}{c_y^2}c_\epsilon + \frac{a_Y}{c_y} + \frac{b_Y}{c_\epsilon}h^p)$. If $p < 1$, $2\beta_h + 2\alpha_h < 1$ definitely. If instead $p = 1$, (3.22) gives $\frac{\bar\rho^2}{c_y^2}c_\epsilon + \frac{a_Y}{c_y} < \frac{1}{2}$, so that $2\beta_h + 2\alpha_h \le \frac{1}{2} + \frac{b_Y}{c_\epsilon}h < 1$ for every h small enough. $\square$

Now, Proposition 3.8 ensures that there exists $h_0 > 0$ such that for every $h < h_0$ and for every $v \in \cup_{n=0}^N \mathcal{V}_n^h$, the matrix A^{-1} discussed in Section 3.1.1 and the matrix C discussed in Section 3.1.2 are both well defined and stochastic matrices. So, for h small, that is $h < h_0$ with h_0 as in Proposition 3.8, we define $\Pi^h(v)$ as follows:

- if $v > \epsilon_h$, $\Pi^h(v)$ is the inverse of the matrix A in (3.8),

- if $v \le \epsilon_h$, $\Pi^h(v)$ is the matrix C in (3.18).

As a consequence, we can assert that for every $v \in \cup_{n=0}^N \mathcal{V}_n^h$, $\Pi^h(v) = (\Pi^h(v)_{ij})_{i,j\in\mathcal{J}_{M_h}}$ is a stochastic matrix. We now define the following transition probability law: at time-step $n \in \{0,1,\dots,N\}$, for $(y_i, v_{n,k}) \in \mathcal{Y}_{M_h}^h \times \mathcal{V}_n^h$ we set $\mu^h(y_i, v_{n,k};\cdot)$ the law in $\mathbb{R}^2$ as in (2.24), that is

$$\mu^h(y_i, v_{n,k}; A\times B) = \sum_{k^*\in\{k_u^h(n,k),k_d^h(n,k)\}} \sum_{j\in\mathcal{J}_{M_h}} \Pi^h(v_{n,k})_{i,j} p_{k^*}^h \delta_{\{y_j\}}(A)\delta_{\{v_{n+1,k^*}\}}(B). \tag{3.23}$$

So, we call $\bar X^h = (\bar X_n^h)_{n=0,1,\dots,N}$ the 2-dimensional Markov chain having (3.23) as its transition probability law at time-step $n \in \{0,1,\dots,N\}$, that is

$$\mathbb{P}\Big(\bar X_{n+1}^h = (y_j, v_{n+1,k^*}) \mid \bar X_n^h = (y_i, v_{n,k})\Big) = \begin{cases} \Pi(v_{n,k})_{ij}\, p_{k_u^h(n,k)}^h & \text{if } k^* = k_u^h(n,k) \\ \Pi(v_{n,k})_{ij}\, p_{k_d^h(n,k)}^h & \text{if } k^* = k_d^h(n,k) \\ 0 & \text{otherwise,} \end{cases}$$

for every $(y_i, v_{n,k}) \in \mathcal{Y}^h \times \mathcal{V}_n^h$ and $(y_j, v_{n+1,k^*}) \in \mathcal{Y}^h \times \mathcal{V}_{n+1}^h$. Since $\sum_j \Pi(v)_{ij} = 1$, one gets that the second component of $\bar X^h$ is a Markov chain itself and it equals, in law, to $\bar V^h$. So, we write $\bar X^h = (\bar Y^h, \bar V^h)$ and for every function $f : \mathbb{R}^2 \to \mathbb{R}$ we have

$$\mathbb{E}(f(\bar Y_{n+1}^h, \bar V_{n+1}^h) \mid (\bar Y_n^h, \bar V_n^h) = (y_i, v_{n,k})\Big) = \sum_{k^*\in\{k_u^h(n,k),k_d^h(n,k)\}} \sum_{j\in\mathcal{J}_{M_h}} f(y_j, v_{n+1,k^*})\Pi^h(v_{n,k})_{i,j} p_{k^*}^h. \tag{3.24}$$

So, coming back to the discussion in Section 2.2, by (3.24) we can assert that our algorithm is actually given by approximating the diffusion pair $X = (Y, V)$ with the Markov chain $\bar X^h = (\bar Y^h, \bar V^h)$. So, we set $X^h = (Y^h, V^h)$ as the piecewise constant and càdlàg interpolation in time of $\bar X^h$, that is

$$X_t^h = \bar X_n^h, \quad t \in [nh, (n+1)h), \quad n = 0,1,\dots,N-1. \tag{3.25}$$

We set $\mathrm{D}([0,T];\mathbb{R}^2)$ the space of the $\mathbb{R}^2$-valued and càdlàg functions on the interval $[0,T]$, that we assume to be endowed with the Skorohod topology (see e.g. Billingsley [5]). Our main result is the following:

Theorem 3.9. *Suppose that (3.21) and (3.22) hold. Then as $h \to 0$, the sequence $\{X^h\}_h = \{(Y^h, V^h)\}_h$ weakly converges in the space $\mathrm{D}([0,T];\mathbb{R}^2)$ to the diffusion process $X = (Y, V)$ solution to (2.4)-(2.5).*

Proof. The idea of the proof is standard, see e.g. Nelson and Ramaswamy [24] or also classical books such as Billingsley [5], Ethier and Kurtz [10] or Stroock and Varadhan [29].

To simplify the notations, let us set

$$\begin{aligned}
\mathcal{M}_{n,i,k}^Y(h;l) &= \mathbb{E}\Big((\bar Y_{n+1}^h - y_i)^l \,|\, (\bar Y_n^h, \bar V_n^h) = (y_i, v_{n,k})\Big), \quad l = 1,2,4,\\
\mathcal{M}_{n,i,k}^V(h;l) &= \mathbb{E}\Big((\bar V_{n+1}^h - v_{n,k})^l \,|\, (\bar Y_n^h, \bar V_n^h) = (y_i, v_{n,k})\Big), \quad l = 1,2,4,\\
\mathcal{M}_{n,i,k}^{Y,V}(h) &= \mathbb{E}\Big((\bar Y_{n+1}^h - y_i)(\bar V_{n+1}^h - v_{n,k}) \,|\, (\bar Y_n^h, \bar V_n^h) = (y_i, v_{n,k})\Big).
\end{aligned}$$

It is clear that $\mathcal{M}^Y_{n,i,k}(h;l)$ is the local moment of order l at time nh related to Y, as well as $\mathcal{M}^V_{n,k}(h;l)$ is related to the component V, and $\mathcal{M}^{Y,V}_{n,i,k}(h)$ is the local cross-moment of the two components at the generic time step n. So, the proof reduces in checking that, for fixed $r_*, v_* > 0$ and setting $\Lambda_* = \{(n,i,k) : v_{n,k} \le v_*, |y_i| \le r_*\}$, then the following properties $i)$, $ii)$ and $iii)$ hold:

$i)$ (convergence of the local drift)

$$\lim_{h\to 0} \sup_{(n,i,k)\in\Lambda_*} \frac{1}{h} \left|\mathcal{M}^Y_{n,i,k}(h;1) - (\mu_Y)_{n,k} h\right| = 0,$$
$$\lim_{h\to 0} \sup_{(n,i,k)\in\Lambda_*} \frac{1}{h} \left|\mathcal{M}^V_{n,i,k}(h;1) - (\mu_V)_{n,k} h\right| = 0;$$

where we have set $(\mu_Y)_{n,k} = \mu_Y(v_{n,k})$ and $(\mu_V)_{n,k} = \mu_V(v_{n,k})$;

$ii)$ (convergence of the local diffusion coefficient)

$$\lim_{h\to 0} \sup_{(n,i,k)\in\Lambda_*} \frac{1}{h} \left|\mathcal{M}^Y_{n,i,k}(h;2) - \bar\rho^2 v_{n,k} h\right| = 0,$$
$$\lim_{h\to 0} \sup_{(n,i,k)\in\Lambda_*} \frac{1}{h} \left|\mathcal{M}^V_{n,k}(h;2) - \sigma^2 v_{n,k} h\right| = 0$$
$$\lim_{h\to 0} \sup_{(n,i,k)\in\Lambda_*} \frac{1}{h} \left|\mathcal{M}^{Y,V}_{n,i,k}(h)\right| = 0;$$

$iii)$ (fast convergence to 0 of the fourth order local moments)

$$\lim_{h\to 0} \sup_{(n,i,k)\in\Lambda_*} \frac{1}{h} \mathcal{M}^Y_{n,i,k}(h;4) = 0,$$
$$\lim_{h\to 0} \sup_{(n,i,k)\in\Lambda_*} \frac{1}{h} \mathcal{M}^V_{n,i,k}(h;4) = 0.$$

We recall that the V-component of the 2-dimensional Markov chain is a Markov chain itself and we have

$$\mathcal{M}^V_{n,i,k}(h;l) \equiv \mathcal{M}^V_{n,k}(h;l) = \mathbb{E}\Big((\bar V^h_{n+1} - v_{n,k})^l \,|\, \bar V^h_n = v_{n,k}\Big), \quad l = 1,2,4.$$

The limits in $i)$, $ii)$ and $iii)$ containing $\mathcal{M}^V_{n,k}(h;l)$ for $l = 1,2,4$, have been already proved in [3] (see Theorem 7 therein), so we prove the validity of the remaining limits.
We set $\psi^i_l(y)$ the vector in $\mathbb{R}^{2M_h+1}$ given by $(\psi^i_l(y))_j = (y_j - y_i)^l$, $j \in \mathcal{J}_{M_h}$. From (3.24) we get

$$\mathcal{M}^Y_{n,i,k}(h;l) = \sum_{y_j\in\mathcal{Y}^h} \Pi^h(v_{n,k})_{i,j}\psi^i_l(y_j) = \Big(\Pi^h(v_{n,k})\psi^i_l(y)\Big)_i$$

and we notice that the above quantity has been already discussed in the previous sections. We set $\Lambda_* = \Lambda_{*,1,h} \cup \Lambda_{*,2,h}$, with

$$\Lambda_{*,1,h} = \{(n,i,k) : \epsilon_h < v_{n,k} \le v_*, |y_i| \le r_*\},$$
$$\Lambda_{*,2,h} = \{(n,i,k) : v_{n,k} \le \epsilon_h, |y_i| \le r_*\}.$$

For $(n,i,k)\in\Lambda_{*,1,h}$, $\Pi^h(v_{n,k})$ is the inverse of the matrix A in (3.8). So, by using (3.17), we have

$$
\begin{aligned}
\mathcal{M}^Y_{n,i,k}(h;1)=(A^{-1}\psi^i_1(y))_i= &\ \ h(\mu_Y)_{n,k}+(\phi^i_{1,M_h})_i,\\
\mathcal{M}^Y_{n,i,k}(h;2)=(A^{-1}\psi^i_2(y))_i= &\ \ h\bar\rho^2 v_{n,k}+2h\Delta y_h(\mu_Y)_{n,k}+(\phi^i_{2,M_h})_i,\\
\mathcal{M}^Y_{n,i,k}(h;4)=(A^{-1}\psi^i_4(y))_i= &\ \ h\Delta y_h^2\bar\rho^2 v_{n,k}+8h^2\Delta y_h^2(\mu_Y)^2_{n,k}+24h^3(\mu_Y)^3_{n,k}+6h^2\bar\rho^4 v^2_{n,k}+\\
&+24\tfrac{h^4}{\Delta y_h}\,\bar\rho^2 v_{n,k}(\mu_Y)^3_{n,k}+(\phi^i_{4,M_h})_i,
\end{aligned}
$$

$\phi^i_{l,M_h}(y)$ being given in (3.15). In Lemma 3.10 below, we prove that for $l\le 4$,

$$
\sup_{(n,i,k)\in\Lambda_{*,1,h}}\frac{1}{h}\big|(\phi^i_{l,M_h}(y))_i\big|\to 0\quad\text{as}\quad h\to 0.
$$

And since $(\mu_Y)_{n,k}$ is bounded on Λ_*, the limits in $i)$, $ii)$ and $iii)$ associated to $\mathcal{M}^Y_{n,i,k}(h;l)$, $l=1,2,4$, hold uniformly in $\Lambda_{*,1,h}$. We prove the same on the set $\Lambda_{*,2,h}$. For $(n,i,k)\in\Lambda_{*,2,h}$, the matrix $\Pi^h(v_{n,k})$ to be taken into account is given by the matrix C in (3.18). Moreover, for h small enough, we have that if $(n,i,k)\in\Lambda_{*,2,h}$ then $|i|<M_h$. Therefore, by (3.20) we obtain

$$
\begin{aligned}
\mathcal{M}^Y_{n,i,k}(h;1)=(C\psi^i_1(y))_i &\ =h(\mu_Y)_{n,k},\\
\mathcal{M}^Y_{n,i,k}(h;2)=(C\psi^i_2(y))_i &\ =h\bar\rho^2 v_{n,k}+h\Delta y_h|(\mu_Y)_{n,k}|,\\
\mathcal{M}^Y_{n,i,k}(h;4)=(C\psi^i_4(y))_i &\ =h\Delta y_h^2\bar\rho^2 v_{n,k}+h\Delta y_h^3|(\mu_Y)_{n,k}|
\end{aligned}
$$

and again the limits in $i)$, $ii)$ and $iii)$ concerning $\mathcal{M}^Y_{n,i,k}(h;l)$, $l=1,2,4$, hold uniformly in $\Lambda_{*,2,h}$. It remains to study the cross-moment. By using (3.24), it is given by

$$
\mathcal{M}^{Y,V}_{n,i,k}(h)=\mathcal{M}^Y_{n,i,k}(h;1)\mathcal{M}^V_{n,k}(h;1)
$$

and the convergence as in $ii)$ immediately follows from the already proved limits in $i)$. $\square$

In order to conclude, we only need to prove next

Lemma 3.10. *Assume that (3.21) and (3.22) both hold. Let $v_*,r_*>0$ and set*

$$
\Lambda_{*,1,h}=\Lambda_*=\{(n,i,k)\,:\,\epsilon_h<v_{n,k}\le v_*,|y_i|\le r_*\}
$$

Then one has

$$
\lim_{h\to0}\frac{1}{h}\sup_{(n,i,k)\in\Lambda_{*,1,h}}\frac{1}{h}|(\phi^i_{l,M_h}(y))_i|=0, \tag{3.26}
$$

for every $l\le 4$, where $\phi^i_{l,M_h}(y)$ is defined in (3.15) with $M=M_h$.

Proof. We use Proposition 3.5. Here, C denotes a positive constant that may vary from line to line.

Under (3.21) and (3.22), for $(n,i,k)\in\Lambda_{*,1,h}$ we have already observed that $\beta_h>|\alpha_h|$, α_h,β_h being given in (3.6), and the constraints $\Delta y_h\le 1$ and $M_h\Delta y_h\ge 1$ are trivially satisfied for h small. Moreover, on the set $\Lambda_{*,1,h}$, there exists $C>0$ such that for every $l\le 4$,

$$
l2^{l+2}(\beta_h\Delta y_h^2+|\alpha_h|\Delta y_h)\le Ch\le 1.
$$

for h small enough. And since $\beta_h + |\alpha_h| \leq Ch^{1-2p}$, we also have

$$1 - \frac{1}{1+\beta+|\alpha|} \leq 1 - \frac{1}{1+Ch^{1-2p}}.$$

Then, by applying Proposition 3.5, we can write

$$|(\phi^i_{l,M}(y))_i| \leq Ch^{-(q-p)l+1-2p}\Big(1 - \frac{1}{1+Ch^{1-2p}}\Big)^{M_h - i}.$$

Now, on the set $\Lambda_{*,1,h}$ we have $|y_i| \leq r_*$, so that $|i| \leq \frac{r_*+Y_0}{\Delta y_h} \leq Ch^{-p}$. And by recalling that $q > p$, we get $M_h - i \geq C(h^{-q} - h^{-p}) \geq ch^{-q}$ for some $c > 0$. So, we have proved that there exist $C, c > 0$ such that for every h close to 0

$$\sup_{(n,i,k)\in\Lambda_{*,1,h}} \frac{1}{h}|(\phi^i_{l,M}(y))_i| \leq Ch^{-(q-p)l-2p}\Big(1 - \frac{1}{1+Ch^{1-2p}}\Big)^{ch^{-q}} \quad \text{for} \quad l \leq 4,$$

Now, straightforward computations give that, as $h \to 0$, the r.h.s. above tends to 0 if $q > p$ and $p \leq 1$, and this completes the proof. $\square$

Remark 3.11. *Theorem 3.9 proves the convergence of the algorithm in the case we introduce suitable boundary conditions in the finite difference component of our procedure. We stress that our Neumann-type conditions (3.4) and (3.5) may be replaced by other types of boundary conditions, that can be handled by using Proposition A.1 (see Appendix A.1).*
We also recall that the use of a boundary is a numerical requirement which is necessary to solve the problem in practice. However, one could prove the convergence result even in the whole grid $\mathcal{Y}^h = \{y_i = Y_0 + i\Delta y_h\}_{i\in\mathbb{Z}}$, that is by considering the solutions to (3.1) and (3.2) without linking to these equations any boundary condition. In fact, under the requirements of Theorem 3.9, the inverse of the implicit difference linear operator A associated to (3.1) remains a Markovian transition function, as well as the explicit difference linear operator C related to (3.2). Moreover, formulas similar to (3.17) and (3.20) hold, with a simplification from the fact that boundary contributions do not exist. For the sake of completeness, we give these proofs in Appendix A.2. This means that the arguments used to prove Theorem 3.9 can be applied also on the infinite grid $\mathcal{Y}^h = \{y_i = Y_0 + i\Delta y_h\}_{i\in\mathbb{Z}}$.

Thanks to Theorem 3.9 we can deal with the convergence of the price evaluated on our Markov chain to the one computed on the Heston model.
The problem of pricing European options is simple when the payoff is not too complicated. In fact, let T denote the maturity date and f be the payoff function, that is $f : \mathrm{D}([0,T];\mathbb{R}) \to \mathbb{R}_+$. First, we write the payoff in terms of the transformation (2.3), so we get the transformed payoff-function $g(y,v) = f(e^{y+\frac{\varrho}{\sigma}v})$, $(y,v) \in \mathrm{D}([0,T];\mathbb{R}^2)$, and the associated option prices on the continuous and the discrete model as seen at time 0 are given by

$$P_{Eu} = \mathbb{E}\Big(\tilde{g}(Y,V)\Big) \quad \text{and} \quad P^h_{Eu} = \mathbb{E}\Big(\tilde{g}(Y^h,V^h)\Big),$$

respectively, $\tilde{g}$ denoting the discounted payoff, i.e. $\tilde{g} = e^{-rT}g$ (it is clear that the writing $\mathbb{E}$ for both prices is an abuse of notations, since in principle one should use the notations $\mathbb{E}_{\mathbb{P}}$ and

$\mathbb{E}_{\mathbb{P}^h}$ related to the measures $\mathbb{P}$ and $\mathbb{P}^h$ of the probability space where the processes (Y, V) and (Y^h, V^h) are defined). Now, the weak convergence in Theorem 3.9 ensures the convergence $P^h_{Eu} \to P_{Eu}$ of the European price when the discounted payoff-function fulfills the following requests: $(y, v) \mapsto \tilde{g}(y, v)$ is continuous and there exists $a > 0$ and $h_* > 0$ such that

$$\sup_{h<h_*} \mathbb{E}\Big(|\tilde{g}(Y^h, V^h)|^{1+a}\Big) < \infty.$$

This a consequence of standard results on the convergence of the expectations for sequences of random variables which are weakly convergent and satisfy uniform integrability properties.
As for American style options, even for simple payoffs things are more difficult because of the presence of optimal stopping times. However, due to the results in Amin and Khanna [2], we can deduce the convergence of the prices for suitable payoffs. In fact, let $f : [0, T] \times \mathrm{D}([0, T]; \mathbb{R}) \to \mathbb{R}_+$ denote a payoff function. By passing to the pair (Y, V) and by considering the resulting discounted payoff function $\tilde{g}(t, y, v) = e^{-rt} f(t, e^{y+\frac{\rho}{\sigma}v})$, the associated option prices on the continuous and the discrete model as seen at time 0 are given by

$$P_{Am} = \sup\nolimits_{\tau\in\mathcal{G}_{0,T}} \mathbb{E}\Big(\tilde{g}(\tau, Y, V)\Big) \quad \text{and} \quad P^h_{Am} = \sup\nolimits_{\sigma\in\mathcal{G}^h_{0,T}} \mathbb{E}\Big(\tilde{g}(\sigma, Y^h, V^h)\Big),$$

where $\mathcal{G}_{0,T}$ and $\mathcal{G}^h_{0,T}$ denote the stopping times in $[0, T]$ w.r.t. the filtration $\mathcal{F}_t = \sigma((Y_s, V_s) : s \le t)$ and $\mathcal{F}^h_t = \sigma((Y^h_s, V^h_s) : s \le t)$ respectively.
Consider the following assumptions:

(H1) $\tilde{g} : [0, T] \times \mathrm{D}([0, T]; \mathbb{R}^2) \to \mathbb{R}_+$ is a continuous function (in the product topology) and for every $\xi, \eta \in \mathrm{D}([0, T]; \mathbb{R}^2)$ such that $\xi_s = \eta_s$ for every $s \in [0, t]$ then $\tilde{g}(t, \xi) = \tilde{g}(t, \eta)$;

(H2) there exist $a > 0$ and $h_* > 0$ such that $\sup_{h<h_*} \mathbb{E}\Big(\sup_{t\le T} |\tilde{g}(t, Y^h, V^h)|^{1+a}\Big) < \infty$.

Let ρ^h denote the optimal stopping time related to the discrete problem, that is

$$P^h_{Am} = \mathbb{E}\Big(\tilde{g}(\rho^h, Y^h, V^h)\Big).$$

Then by using the arguments in Amin and Khanna [2], properties *i*)-*iii*) in the proof of Theorem 3.9 ensure the existence of $\rho \in \mathcal{G}_{0,T}$ such that the triple (ρ^h, Y^h, V^h) converges in law to (ρ, Y, V) - this is a very roughly speaking: in the words by Amin and Khanna, ρ is "in some appropriate sense a legitimate stopping time w.r.t. the filtration generated by (Y, V)", and to be precise one should refer to a further probability space (for technical details, see the discussion at pages 299-300 and in particular Theorem 4.1 in [2]). We stress that in [2] this result is based on two starting assumptions: the (global) Lipschitz continuity and the sublinearity property for both the drift and the diffusion coefficient of the pair (Y, V). Here, the Lipschitz continuity property does not hold because of the presence of the square-root function, so the diffusion coefficient is only Hölder continuous. Nevertheless, they use such a property just to ensure the existence of the strong solution, a condition that here holds, and use specifically the sublinearity property of the drift coefficients, that here holds. So, their arguments apply as well. Once the weak convergence of the triple is achieved, the plan to prove the convergence of the prices is the following. Under (H1) one gets that $\{\tilde{g}(\rho^h, Y^h, V^h)\}_h$ converges in law, as $h \to 0$, to

$\tilde{g}(\rho, Y, V)$. Moreover, (H2) implies that the $\{\tilde{g}(\rho^h, Y^h, V^h)\}_h$ is a uniformly integrable family of random variables, and this suffices to get the convergence of the expectations. Finally, one gets $P_{Am} = \mathbb{E}(\tilde{g}(\rho, Y, V))$, and the convergence $P^h_{Am} \to P_{Am}$ of the American prices is achieved. As an immediate consequence, both European and American put options can be numerically evaluated by means of the approximating algorithm described in Section 2.2 (and, due to the call-put parity formula, also European call option can be numerically priced with our method). However, in next Section 4 we apply our procedure to numerically price European/American barrier options written on the Heston model, that is options which are much more sophisticated and, for example, do not fulfill (H1). Nevertheless, the numerical experiments give a real evidence of the goodness (that is, convergence) of our algorithm also in this case.

4 Numerical results

In this section we provide numerical results in order to asses the efficiency and the robustness of our hybrid tree-finite difference method in the case of plain vanilla options and barrier options.

4.1 European and American vanilla options

We compare here the performance of the hybrid tree-finite difference algorithm (called **HTFD**) introduced in Section 2.2 with the tree method of Vellekoop Nieuwenhuis (called **VN**) in [25] for the computation of European and American options in the Heston model. All the numerical results (except for our method) are obtained by using the Premia software [27]. In the European and American option contracts we are dealing with, we consider the following set of parameters: initial price $S_0 = 100$, strike price $K = 100$, maturity $T = 1$, interest rate $r = \log(1.1)$, dividend rate $\delta = 0$, initial volatility $V_0 = 0.1$, long-mean $\theta = 0.1$, speed of mean-reversion $\kappa = 2$, correlation $\rho = -0.5$. In order to study the numerical robustness of the algorithms, we choose three different values for σ: we set $\sigma = 0.04, 0.5, 1$. We first consider the case $\sigma = 0.04$, that is σ close to zero (which implies that the Heston PDE is convection-dominated in the V-direction). Moreover, for $\sigma = 1$, we stress that the Feller condition $2\kappa\theta \geq \sigma^2$ is not satisfied.
In the (pure) tree method **VN**, we fix the number of points in the V coordinate as $N_V = 50$, with varying number of time and space steps: $N_t = N_S = 50, 100, 200, 400$.
The numerical study of the hybrid tree-finite difference method **HTFD** is split in two cases: **HTFD1** refers to the (fixed) number of time steps $N_t = 100$ and varying number of space steps $N_S = 50, 100, 200, 400$; we add the situation **HTFD2** where the number of time steps is equal to the number of space steps $N_t = N_S = 50, 100, 200, 400$.
Table 1 reports European put option prices. Comparisons are given with a benchmark value obtained using the Heston closed formula **CF** in [17].
In Table 2 we provide results for American put option prices. In this case we use a benchmark from the Monte Carlo Longstaff-Schwartz algorithm, called **MC-LS**, as in [23], with a huge number of Monte Carlo simulation (1 million iterations) which are done by means of the accurate Alfonsi [1] discretization scheme for the CIR process with $M = 100$ discretization time steps and bermudan exercise dates. We recall that the Alfonsi method provides a Monte Carlo weak second-order scheme for the CIR process, without any restriction on its parameters.

Table 3 refers to the computational time cost (in seconds) of the different algorithms for $\sigma = 0.5$ in the European case.

	N_S	VN	HTFD1	HTFD2	CF
σ =0.04	50	8.040982	7.934492	7.911034	
	100	8.021780	7.970437	7.970437	
	200	8.003938	7.978890	7.983188	7.994716
	400	7.984248	7.980984	7.990825	
σ =0.50	50	8.148234	7.758954	7.746533	
	100	7.727191	7.804520	7.804520	
	200	7.813599	7.816749	7.821404	7.8318540
	400	7.910909	7.818596	7.827805	
σ =1.00	50	6.586889	7.214303	7.247748	
	100	7.114225	7.225292	7.225292	
	200	7.964052	7.228235	7.229139	7.2313083
	400	6.639931	7.224356	7.233742	

Table 1: *Prices of European put options.* $\sigma = 0.04, 0.5, 1$. $S_0 = 100$, $K = 100$, $T = 1$, $r = \log(1.1)$, $\delta = 0$, $V_0 = 0.1$, $\theta = 0.1$, $\kappa = 2$, $\rho = -0.5$.

	N_S	VN	HTFD1	HTFD2	MC-LS
σ =0.04	50	9.100312	8.966651	8.932445	
	100	9.086233	9.016732	9.016732	
	200	9.073722	9.028866	9.042581	9.074102
	400	9.063396	9.031881	9.054538	
σ =0.50	50	9.150887	8.763369	8.731867	
	100	8.892206	8.841776	8.841776	
	200	8.981855	8.862606	8.878530	8.904514
	400	9.058313	8.866911	8.892583	
σ =1.00	50	8.588392	8.185052	8.206052	
	100	9.020989	8.263395	8.263395	
	200	9.251595	8.281755	8.290371	8.277985
	400	9.102788	8.283214	8.304415	

Table 2: *Prices of American put options.* $\sigma = 0.04, 0.5, 1$. $S_0 = 100$, $K = 100$, $T = 1$, $r = \log(1.1)$, $\delta = 0$, $V_0 = 0.1$, $\theta = 0.1$, $\kappa = 2$, $\rho = -0.5$.

The numerical results show that the hybrid tree-finite difference method is very accurate, reliable and efficient.

4.2 European and American barrier options

We study here the continuously monitered barrier options case and we compare our hybrid tree-finite difference algorithm with the numerical results of the method of lines provided in Chiarella et al. [7]. So, we consider European and American up-and-out call options with the following set of parameters: $K = 100$, $T = 0.5$, $r = 0.03$, $\delta = 0.05$, $V_0 = 0.1$, $\theta = 0.1$, $\kappa = 2$, $\rho = -0.5$. The up barrier is $H = 130$. We choose different values for S_0: $S_0 = 80, 100, 120$.
We also compare with a benchmark value obtained by using the method of lines, called **MOL**, with mesh parameters $100, 200, 6400$ (see Chiarella et al. [7]).

N_S	VN	HTFD1	HTDF2
50	0.11	0.02	0.007
100	0.42	0.04	0.040
200	1.73	0.08	0.380
400	7.06	0.16	3.040

Table 3: *Computational times (in seconds) for European put options for* $\sigma = 0.5$.

Table 4 and Table 6 report European and American Up-and-Out option prices respectively, while Table 5 refers to the computational time cost (in seconds) of the various algorithms for the European barrier case.

	N_S	HTFD1	HTFD2	MOL
S_0 =80	50	0.913861	0.875374	
	100	0.893484	0.893484	
	200	0.895127	0.900893	0.9029
	400	0.897820	0.902770	
S_0 =100	50	2.635396	2.583568	
	100	2.606249	2.606249	
	200	2.597363	2.591857	2.5908
	400	2.603679	2.594134	
S_0 =120	50	1.417225	1.438429	
	100	1.485704	1.485704	
	200	1.500692	1.482193	1.4782
	400	1.504755	1.486212	

Table 4: *Prices of European call up-and-out options. Up barrier is* $H = 130$. $K = 100$, $T = 0.5$, $r = 0.03$, $\delta = 0.05$, $V_0 = 0.1$, $\theta = 0.1$, $\kappa = 2$, $\rho = -0.5$.

N_S	HTFD1	HTDF2
50	0.007	0.017
100	0.132	0.132
200	0.284	1.079
400	0.535	8.901

Table 5: *Computational times (in seconds) for European Barrier options.*

	N_S	HTFD1	HTFD2	MOL
S_0 =80	50	1.199802	1.285959	
	100	1.369914	1.369914	
	200	1.400823	1.396628	1.4012
	400	1.400710	1.401111	
S_0 =100	50	8.274116	8.269779	
	100	8.286667	8.286667	
	200	8.284054	8.294226	8.3003
	400	8.283815	8.296745	
S_0 =120	50	21.943742	21.884228	
	100	21.820015	21.820015	
	200	21.785274	21.815989	21.8216
	400	21.779648	21.804518	

Table 6: *Prices of American call up-and-out options. Up barrier is $H = 130$. $K = 100$, $T = 0.5$, $r = 0.03$, $\delta = 0.05$, $V_0 = 0.1$, $\theta = 0.1$, $\kappa = 2$, $\rho = -0.5$.*

A Appendix

A.1 Boundary sensitivity for the implicit finite difference operator

We study here the behavior of the solution $x = (x_1, \dots, x_N)^T$ of the two following linear systems

$$Ax = \mathbf{v}_1 \tag{A.1}$$

$$Ax = \mathbf{v}_N, \tag{A.2}$$

where $\mathbf{v}_i$, $i = 1, \dots, N$, denotes the standard orthonormal basis in $\mathbb{R}^N$, i.e. $(\mathbf{v}_i)_k = 0$ for $k \neq i$ and $(\mathbf{v}_i)_i = 1$, $i = 1, \dots, N$, and where A has the following general tridiagonal form

$$A = \begin{pmatrix} a_1 & c_1 & & & \\ b & a & c & & \\ & \ddots & \ddots & \ddots & \\ & & b & a & c \\ & & & b_N & a_N \end{pmatrix}. \tag{A.3}$$

The result we are going to present is due for matrices A as in (A.3) and that satisfy the hypotheses (P2)-(P3) in the proof of Proposition 3.1, ensuring that they are invertible M-matrices (see for instance [4]).

Proposition A.1. *Suppose that the matrix A in (A.3) satisfies*

$$a, a_1, a_N > 0, \quad b, c \leq 0, \quad c_1, b_N \leq 0, \quad a > |b| + |c|, \quad a_1 > |c_1|, \quad a_N > |b_N|. \tag{A.4}$$

Assume moreover the following stability conditions on the "boundary" values a_1, a_N, c_1 and b_N:

$$\frac{|bc_1|}{a_1} < z_+ \quad \text{and} \quad \frac{|b_N c|}{a_N} < z_+, \tag{A.5}$$

where $z_+ = (a + \sqrt{a^2 - 4|bc|})/2$. Then the solution x of (A.1) is defined by a sequence $\{x_k\}_{k=1,\dots,N}$ of positive terms and there exists a positive value $\tilde{\gamma}^* > |b|$ such that, for $k = 2, \dots, N-1$

$$x_{N+1-k} \leq x_1 \left(\frac{|b|}{\tilde{\gamma}^*}\right)^{N-k} \quad \text{and} \quad x_N \leq x_1 \frac{|b_N|}{a_N} \left(\frac{|b|}{\tilde{\gamma}^*}\right)^{N-2}. \tag{A.6}$$

Similarly, for the solution x of (A.2) it holds $x_k > 0$, for all $k = 1, \dots, N$ and there exists a positive value $\gamma^* > |c|$ such that for $k = 2, \dots, N-1$,

$$x_k \leq x_N \left(\frac{|c|}{\gamma^*}\right)^{N-k} \quad \text{and} \quad x_1 \leq x_N \frac{|c_1|}{a_1} \left(\frac{|c|}{\gamma^*}\right)^{N-2}. \tag{A.7}$$

Proof. Let us start by estimating first the solution x of system (A.2). By applying the Thomas algorithm, also known as tridiagonal matrix algorithm [30], the solution is given by back substitutions:

$$x_N = \frac{1}{\gamma_N}, \quad x_k = \frac{|c|}{\gamma_k} x_{k+1} \text{ for } k = N-1, \dots, 2, \quad x_1 = \frac{|c_1|}{\gamma_1} x_2,$$

where the coefficients γ_k are recursively defined by

$$\gamma_1 = a_1, \quad \gamma_2 = a - \frac{|bc_1|}{\gamma_1}, \quad \gamma_k = a - \frac{|bc|}{\gamma_{k-1}} \text{ for } k = 3, \dots, N-1, \quad \gamma_N = a_N - \frac{|b_N c|}{\gamma_{N-1}}. \tag{A.8}$$

It is easy to verify that under assumptions (A.4), for $k = 3, \dots, N-1$, the sequence $\{\gamma_k\}$ has two strictly positive fixed points $z_\pm = (a \pm \sqrt{a^2 - 4|bc|})/2$. Moreover, z_- is an unstable fixed point while z_+ is stable. Since $\gamma_2 = a - |bc_1|/a_1$ and by (A.5), we have that $\gamma_2 > z_-$. So, starting from γ_2 the sequence converges to z_+ and we have that for $\gamma^* = \min\{\gamma_2, z_+\}$

$$\gamma_k \geq \gamma_*, \quad k = 2, \dots, N-1. \tag{A.9}$$

Furthermore, by (A.4) we have that $\gamma_N > 0$.
Going back to the sequence $\{x_k\}_{i=1,\dots,N}$, we first notice that since $\gamma_N > 0$ then $x_N > 0$ and accordingly $x_k > 0$ for all $k = 1, ..., N$. Moreover, from condition (A.9) we obtain (A.7). In fact, for $k = 2, \dots, N-1$ we have

$$x_k = \frac{|c|}{\gamma_k} x_{k+1} \leq \frac{|c|}{\gamma^*} x_{k+1} \leq \dots \leq \left(\frac{|c|}{\gamma^*}\right)^{N-k} x_N$$

and thus

$$x_1 = \frac{|c_1|}{\gamma_1} x_2 \leq \frac{|c_1|}{a_1} \left(\frac{|c|}{\gamma^*}\right)^{N-2} x_N.$$

It remains to verify that $\gamma^* > |c|$. We have already seen that $\gamma_2 > |c|$; so, it remains to verify the case $\gamma^* = z_+$ which is obtained as a result of the strictly diagonal dominant condition on A.
To obtain the estimate (A.6) we introduce the $N \times N$ matrix U satisfying $U\mathbf{v}_i = \mathbf{v}_{N+1-i}$, $i = 1, \dots, N$, so

$$U = \begin{pmatrix} 0 & & 0 & 1 \\ & & 1 & \\ & \iddots & & \\ 1 & 0 & & 0 \end{pmatrix}.$$

Since $U\mathbf{v}_N = \mathbf{v}_1$ and $UU = I$ (i.e. $U^{-1} = U$), to compute (A.6) we use that

$$Ax = \mathbf{v}_1 \Leftrightarrow \tilde{A}\tilde{x} = \mathbf{v}_N,$$

where

$$\tilde{A} = UAU = \begin{pmatrix} a_N & b_N & & & \\ c & a & b & & \\ & \ddots & \ddots & \ddots & \\ & & c & a & b \\ & & & c_1 & a_1 \end{pmatrix}$$

and

$$\tilde{x} = Ux = (x_N, x_{N-1}, \dots, x_1)^T.$$

So, following the same reasoning as above, we get $\tilde{\gamma}_2 = a - \frac{|b_N c|}{a_N}$ and $\tilde{\gamma}^* = \min(\tilde{\gamma}_2, z_+) > |b|$ such that

$$\tilde{x}_k \leq \tilde{x}_N \left(\frac{|b|}{\tilde{\gamma}^*}\right)^{N-k}, \quad k = 2, \dots, N-1$$

i.e.

$$x_{N+1-k} \leq x_1 \left(\frac{|b|}{\tilde{\gamma}^*}\right)^{N-k}, \quad k = 2, \dots, N-1.$$

□

Remark A.2. *Assume that A has the form (A.3) with $a > 0$, $b, c < 0$, $a + b + c = 1$ and $a_1 = a_N = a$, $c_1 = b_N = 1 - a$ - this is actually the type of matrix to which we apply Proposition A.1. One can easily check that both (A.4) and the the boundary requirements in (A.5) hold, so Proposition A.1 can be applied. Moreover, estimates (A.6) and (A.7) can be rewritten as follows: for $k = 2, \dots, N-1$,*

$$|(A^{-1}\mathbf{v}_1)_{N+1-k}| \leq \left(\frac{|b|}{\tilde{\gamma}^*}\right)^{N-k} \qquad \text{with} \quad \tilde{\gamma}^* = \min\left(a - \frac{|c(1-a)|}{a}, \frac{a + \sqrt{a^2 - 4|bc|}}{2}\right),$$
$$|(A^{-1}\mathbf{v}_N)_k| \leq \left(\frac{|c|}{\gamma^*}\right)^{N-k} \qquad \text{with} \quad \gamma^* = \min\left(a - \frac{|b(1-a)|}{a}, \frac{a + \sqrt{a^2 - 4|bc|}}{2}\right).$$

In fact, as for the second inequality, (A.7) gives $|(A^{-1}\mathbf{v}_N)_k| \leq x_N (\frac{|c|}{\gamma^*})^{N-k}$, where $x_N = \frac{1}{\gamma_N}$ and γ_N is defined in (A.8), together with $\gamma_1, \dots, \gamma_{N-1}$. Since $\gamma_{N-1} > |c|$, (A.8) gives

$$\gamma_N = a - \frac{|(1-a)c|}{\gamma_{N-1}} \geq a - |1-a|.$$

But $1 - a = b + c \leq 0$, so $\gamma_N \geq 1$ and then $|(A^{-1}\mathbf{v}_N)_k| \leq (\frac{|c|}{\gamma^*})^{N-k}$. Similarly, one has $\tilde{\gamma}_N \geq 1$, and the first inequality holds as well.

A.2 The finite difference operators on the infinite grid

For fixed $\Delta y > 0$ and $y_0 \in \mathbb{R}$, we consider the grid $\mathcal{Y} = \{y_k\}_{k\in\mathbb{Z}}$ on $\mathbb{R}$ by setting

$$y_k = y_0 + k\Delta y, \quad k \in \mathbb{Z}.$$

Any real function φ on $\mathcal{Y}$ can be seen as a sequence on $\mathbb{R}$. So, as usual we set $\mathbb{R}^{\mathbb{Z}} = \{x = (x_k)_{k\in\mathbb{Z}} : x_k \in \mathbb{R}\ \forall k\}$,and we think to $x \in \mathbb{R}^{\mathbb{Z}}$ as "column-type" vectors. We also set $\varphi(y)$, $y \in \mathcal{Y}$, as the point in $\mathbb{R}^{\mathbb{Z}}$ defined as

$$(\varphi(y))_k = \varphi(y_k), \quad k \in \mathbb{Z}.$$

We split our discussion in two parts: the case $v > \epsilon$, giving the implicit finite difference operator, and $v \leq \epsilon$, related to the explicit finite difference operator

A.2.1 The case $v > \epsilon$

For the implicit in time approximation, equation (3.1) can be written as $u^n = Au^{n+1}$, in which $u^n = (u_j^n)_{j\in\mathbb{Z}}$ is the unknown, $u^{n+1} = (u_j^{n+1})_{j\in\mathbb{Z}}$ is given and $A = (A_{kj})_{k,j\in\mathbb{Z}}$ is the infinite dimensional matrix given by

$$A_{kj} = \begin{cases} -\beta+\alpha & \text{if } j = k-1 \\ 1+2\beta & \text{if } j = k \\ -\beta-\alpha & \text{if } j = k+1 \end{cases} \qquad \text{and } A_{kj} = 0 \text{ for } |j-k| > 1.$$

A is actually a linear operator on $\mathbb{R}^{\mathbb{Z}}$: by using the standard row/column product, we get $A : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ and $y = Ax$ if and only if $y_k = (Ax)_k = \sum_{j\in\mathbb{Z}} A_{kj}x_j = (-\beta+\alpha)x_{k-1} + (1+2\beta)x_k + (-\beta-\alpha)x_{k+1}$, $x \in \mathbb{R}^{\mathbb{Z}}$. So, we write

$$A = (1+2\beta)I - T$$

where I denotes the identity map over $\mathbb{R}^{\mathbb{Z}}$ and $T : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ is the linear operator defined through the infinite dimensional matrix

$$T_{kj} = \begin{cases} \beta-\alpha & \text{if } j = k-1 \\ \beta+\alpha & \text{if } j = k+1 \end{cases} \qquad \text{and } T_{kj} = 0 \text{ for } |j-k| \neq 1.$$

Throughout this section, we assume

$$\beta > |\alpha| \geq 0.$$

This gives that $T \geq 0$, in the sense that $T_{ij} \geq 0$ for every $i, j \in \mathbb{Z}$.
First, we want to show that A is an invertible operator, at least over a nice subset of $\mathbb{R}^{\mathbb{Z}}$. Secondly, we want to show that the inverse is a stochastic infinite dimensional matrix, that is it defines a Markov transition function over $\mathbb{Z}$.
So, for $a > 0$, we set

$$\ell_a^2 = \Big\{ x \in \mathbb{R}^{\mathbb{Z}} \,:\, \sum_{k\in\mathbb{Z}} a^{2|k|}|x_k|^2 < \infty \Big\}.$$

Remark A.3. *If $a < 1$ then ℓ_a^2 contains all the sequences with polynomial growth: for any fixed $l \in \mathbb{N}$, the sequence $x = (x_k)_k$ such that $|x_k| \leq C(1+|k|^l)$, $k \in \mathbb{Z}$, belongs to ℓ_a^2. Moreover, $\ell_{a_1}^2 \subset \ell_{a_2}^2$ if $a_1 \leq a_2 \leq 1$. Finally, we notice that, for $a < 1$, ℓ_a^2 contains also special sequences with exponential growth, namely points $x \in \mathbb{R}^{\mathbb{Z}}$ such that $|x_k| \leq C(1+e^{\gamma|k|})$ for every $\gamma < -2\log a$.*

On ℓ_a^2, we define the inner product and the norm by

$$\langle x, y\rangle_a = \sum_{k\in\mathbb{Z}} a^{2|k|}x_k y_k \quad \text{and} \quad |x|_a^2 = \sum_{k\in\mathbb{Z}} a^{2|k|}x_k^2$$

respectively. When $a = 1$ one has

$$\ell_1^2 = \ell^2 = \Big\{ x \in \mathbb{R}^{\mathbb{Z}} \,:\, \sum_{k\in\mathbb{Z}} |x_k|^2 < \infty \Big\},$$

which is the standard ell-2 space of sequences.
For $a > 0$ the map

$$D_a : \ell^2_a \to \ell^2, \quad (D_a x)_k = a^{|k|} x_k, \quad k \in \mathbb{Z}, \tag{A.10}$$

is well defined and invertible, with $D_a^{-1} = D_{a^{-1}}$. Moreover, since

$$\langle x, y\rangle_a = \langle D_a x, D_a y\rangle, \quad x, y \in \ell^2_a, \tag{A.11}$$

$\langle\cdot,\cdot\rangle$ denoting the standard scalar product in ℓ^2, D_a is an isometry between ℓ^2_a and ℓ^2. Therefore, the Hilbert structure of ℓ^2 can be transferred to ℓ^2_a through D_a, so that the space ℓ^2_a, endowed with the scalar product $\langle\cdot,\cdot\rangle_a$, is a Hilbert space.
Since the objects of our interest will belong to ℓ^2_a with $a < 1$, we study the invertibility of A when it is restricted on ℓ^2_a.

Proposition A.4. *Assume that*

$$a > \frac{2\beta}{1+2\beta}.$$

Then $A : \ell^2_a \to \ell^2_a$. Moreover, A is invertible and the inverse A^{-1} can be identified with an infinite dimensional matrix $(\Pi_{i,j})_{i,j\in\mathbb{Z}}$ that does not depend on the choice of a: for every $x \in \ell^2_a$,

$$(A^{-1}x)_k = \sum_{j\in\mathbb{Z}} \Pi_{kj} x_j \quad \textit{for every } k \in \mathbb{Z}.$$

Finally, Π defines a Markov transition function, that is $\Pi_{kj} \geq 0$ and $\sum_{j\in\mathbb{Z}} \Pi_{kj} = 1$ for every $k, j \in \mathbb{Z}$.

Proof. For $x \in \ell^2_a$, we set $(I_\pm x)_k = x_{k\pm1}$, $k \in \mathbb{Z}$. So, we have

$$Tx = (\beta - \alpha) I_- x + (\beta + \alpha) I_+ x.$$

Straightforward computations give $|I_- x|_a^2 = \frac{1}{a^2}\sum_{k\leq 0} a^{2|k|} x_k^2 + a^2 \sum_{k\geq 1} a^{2|k|} x_k^2$, so that $|I_- x|_a \leq \frac{1}{a}|x|_a$. And similarly, one gets $|I_+ x|_a \leq \frac{1}{a}|x|_a$. Therefore,

$$|Tx|_a \leq (\beta - \alpha)|I_- x|_a + (\beta + \alpha)|I_+ x|_a \leq \frac{2\beta}{a}\,|x|_a$$

and this gives $T : \ell^2_a \to \ell^2_a$, so that $A : \ell^2_a \to \ell^2_a$. Moreover, one has $\|T\|_a \leq \frac{2\beta}{a}$, $\|\cdot\|_a$ denoting the standard operator norm on ℓ^2_a. As a consequence one gets

$$\|T^n\|_a \leq \left(\frac{2\beta}{a}\right)^n \quad \text{for every } n \geq 0.$$

We prove now that A is invertible on ℓ^2_a. For $N \in \mathbb{N}$, we set $S_N : \ell^2_a \to \ell^2_a$ through

$$S_N = \frac{1}{1+2\beta}\sum_{n=0}^{N} \frac{1}{(1+2\beta)^n}\, T^n.$$

We notice that

$$\|S_N\|_a \leq \frac{1}{1+2\beta}\sum_{n=0}^{N} \frac{1}{(1+2\beta)^n}\,\|T^n\|_a \leq \frac{1}{1+2\beta}\sum_{n=0}^{N}\left(\frac{2\beta}{a(1+2\beta)}\right)^n.$$

Since $\frac{2\beta}{a(1+2\beta)} < 1$, the series converges and the limit (linear) operator $S : \ell^2_a \to \ell^2_a$ exists. Moreover, we can write

$$S = \frac{1}{1+2\beta} \sum_{n=0}^{\infty} \frac{1}{(1+2\beta)^n} T^n.$$

Let us prove that $S = A^{-1}$. For every N, one has

$$S_N A = S_N\Big((1+2\beta)I - T\Big) = I - \frac{1}{(1+2\beta)^{n+1}} T^{n+1}$$

and by taking the limit as $N \to \infty$ one gets $SA = I$. Similarly, one proves that $AS = I$, so that $S = A^{-1}$ on ℓ^2_a.

Now, let $\{\mathbf{e}_i\}_{i\in\mathbb{Z}} \subset \ell^2$ denote the standard orthonormal basis: $(\mathbf{e}_i)_k = 0$ for $k \neq i$ and $(\mathbf{e}_i)_i = 1$. Since $\ell^2 = \ell^2_a$ with $a = 1 > \frac{2\beta}{1+2\beta}$, we can set

$$\Pi_{kj} = \langle \mathbf{e}_k, S\mathbf{e}_j \rangle, \quad k, j \in \mathbb{Z}.$$

We prove now that Π is actually the matrix we are looking for: if $x \in \ell^2_a$ then $(Sx)_k = \sum_{j\in\mathbb{Z}} \Pi_{kj} x_j$, $k \in \mathbb{Z}$.

Let D_a denote the linear operator in (A.10). Since $D_a^{-1} = D_{a^{-1}}$, (A.11) immediately gives that $\{D_{a^{-1}}\mathbf{e}_i\}_i$ is an orthonormal basis in ℓ^2_a. Moreover, for $x \in \ell^2_a$ one has

$$x = \sum_{j\in\mathbb{Z}} x_j \mathbf{e}_j = \sum_{j\in\mathbb{Z}} a^{|j|} x_j a^{-|j|} \mathbf{e}_j = \sum_{j\in\mathbb{Z}} (D_a x)_j D_{a^{-1}} \mathbf{e}_j$$

and since $Sx \in \ell^2_a$, we can also write

$$\langle D_{a^{-1}}\mathbf{e}_k, Sx \rangle_a = a^{|k|} (Sx)_k.$$

So,

$$\begin{aligned}
(Sx)_k &= a^{-|k|} \langle D_{a^{-1}}\mathbf{e}_k, Sx \rangle_a = a^{-|k|} \sum_{j\in\mathbb{Z}} (D_a x)_j \langle D_{a^{-1}}\mathbf{e}_k, S D_{a^{-1}}\mathbf{e}_j \rangle_a \\
&= a^{-|k|} \sum_{j\in\mathbb{Z}} a^{|j|} x_j \langle D_{a^{-1}}\mathbf{e}_k, S D_{a^{-1}}\mathbf{e}_j \rangle_a = \sum_{j\in\mathbb{Z}} a^{|j|-|k|} \langle D_{a^{-1}}\mathbf{e}_k, S D_{a^{-1}}\mathbf{e}_j \rangle_a x_j.
\end{aligned}$$

Therefore, we get $(Sx)_k = \sum_j \Pi_{kj} x_j$ with

$$\Pi_{kj} = a^{|j|-|k|} \langle D_{a^{-1}}\mathbf{e}_k, S D_{a^{-1}}\mathbf{e}_j \rangle_a.$$

We show now that actually Π is independent of a and moreover $\Pi_{kj} = \langle \mathbf{e}_k, A\mathbf{e}_j \rangle$. Since $D_{a^{-1}}\mathbf{e}_j = a^{-|j|}\mathbf{e}_j$ and $\mathbf{e}_j \in \ell^2_a$, using (A.11) and the fact that $D_a x \in \ell^2$ if $x \in \ell^2_a$, we can write

$$\begin{aligned}
\Pi_{kj} &= a^{-|k|} \langle D_{a^{-1}}\mathbf{e}_k, S\mathbf{e}_j \rangle_a = a^{-|k|} \langle D_{a^{-1}}\mathbf{e}_k, D_{a^{-1}} D_a S\mathbf{e}_j \rangle_a = a^{-|k|} \langle \mathbf{e}_k, D_a S\mathbf{e}_j \rangle \\
&= a^{-|k|} (D_a S\mathbf{e}_j)_k = a^{-|k|} a^{|k|} (S\mathbf{e}_j)_k = \langle e_k, S\mathbf{e}_j \rangle.
\end{aligned}$$

In order to show that Π is a Markov transition function on $\mathbb{Z}$, we proceed as follows. One has

$$\Pi_{kj} = \langle \mathbf{e}_k, S\mathbf{e}_j \rangle = \lim_{N\to\infty} \langle \mathbf{e}_k, S_N\mathbf{e}_j \rangle = \lim_{N\to\infty} \frac{1}{1+2\beta} \sum_{n=0}^{N} \frac{1}{(1+2\beta)^n} (T^n)_{i,j} \geq 0$$

because the entries of T are all ≥ 0, and so it is for T^n, for every n. Finally, let $\mathbf{1} \in \mathbb{R}^{\mathbb{Z}}$ be the unit vector, that is $(\mathbf{1})_k = 1$ for every $k \in \mathbb{Z}$. Notice that $\mathbf{1} \in \ell^2_a$ for every $a < 1$ and $T\mathbf{1} = 2\beta\mathbf{1}$. Then $S\mathbf{1}$ is well defined and

$$S\mathbf{1} = \frac{1}{1+2\beta} \sum_{n=0}^{\infty} \left(\frac{2\beta}{1+2\beta} \right)^n \mathbf{1} = \mathbf{1}$$

and this automatically gives $\sum_{j\in\mathbb{Z}} \Pi_{kj} = 1$ for every $k \in \mathbb{Z}$. $\square$

For each $l \in \mathbb{N}$, we consider the polynomial $(y - y_i)^l$ and we call $\psi^i_l(y) \in \mathbb{R}^{\mathbb{Z}}$ the associated (vector) function of $y \in \mathcal{Y}$:

$$\left(\psi^i_l(y)\right)_k = (y_k - y_i)^l = \Delta y^l (k-i)^l, \quad k \in \mathbb{Z}. \tag{A.12}$$

By Remark A.3, one has $\psi^i_l(y) \in \ell^2_a$ for every fixed $a < 1$, $i \in \mathbb{Z}$ and $l \in \mathbb{N}$, so the quantity $A^{-1}\psi^i_l(y)$ makes sense. We also notice that $\psi^i_0(y) = \mathbf{1}$, so that $A^{-1}\psi^i_0(y) = A^{-1}\mathbf{1} = \mathbf{1}$.
As already developed in Section 3.1.1, we need to deal with $A^{-1}\psi^i_l(y)$ for $l \leq 4$ and $i \in \mathbb{Z}$.
First of all we have

Lemma A.5. *Let $\psi^i_l(y)$ be defined in (A.12). Then for every $l \in \mathbb{N}$ and $i \in \mathbb{Z}$ one has*

$$A\psi^i_l(y) = \psi^i_l(y) - \sum_{j=0}^{l-1} \binom{l}{j} a_{l-j} \Delta y^{l-j} \psi^i_j(y).$$

where

$$a_n = (\beta - \alpha)(-1)^n + (\beta + \alpha), \quad n \in \mathbb{N}, \tag{A.13}$$

that is $a_n = 2\beta$ if n is even and $a_n = 2\alpha$ if n is odd.

The proof is identical to the one of Proposition 3.2, so we omit it. We only notice that, due to the use of the grid on the whole real line, we do not have any contribution from the boundary - see the terms $b^{\pm M}_{l,i} \mathbf{e}_{\pm M}$ in (3.10). This gives that the inverse of A on any polynomial $\psi^i_l(y)$ is much simpler, and in fact we have

Proposition A.6. *For $l \geq 1$ let $\gamma_{l,k}$, $k = 0, 1, \ldots, l$, be iteratively (backwardly) defined as follows:*

$$\gamma_{l,k} = \binom{l}{k} a_{l-k} \Delta y^{l-k} + \sum_{j=k+1}^{l-1} \gamma_{l,j} \binom{j}{k} a_{j-k} \Delta y^{j-k}, \quad k = l-1, \ldots, 0,$$

where the coefficients a_n are given in (A.13). Then

$$A^{-1}\psi^i_l(y) = \psi^i_l(y) + \sum_{j=0}^{l-1} \gamma_{l,j} \psi^i_j(y).$$

For the proof, we again refer to the finite case, that is Proposition 3.3.

As a consequence, for $l = 1, 2, 4$ and $i \in \mathbb{Z}$, the formulas for $(A^{-1}\psi_l^i(y))_i$ are identical to the ones for the finite grid (see (3.17)) except for the boundary terms, which are null here. In fact, by inserting the formula (3.6) for β and α, we get

$$\begin{aligned} (A^{-1}\psi_1^i(y))_i = &\quad h\mu_Y(v), \\ (A^{-1}\psi_2^i(y))_i = &\quad h\bar{\rho}^2 v + 2h\Delta y \mu_Y(v), \\ (A^{-1}\psi_4^i(y))_i = &\quad h\Delta y^2 \bar{\rho}^2 v + 8h^2\Delta y^2 \mu_Y(v)^2 + 24h^3\mu_Y(v)^3 + 6h^2\bar{\rho}^4 v^2 + 24\tfrac{h^4}{\Delta y}\,\bar{\rho}^2 v \mu_Y(v)^3, \end{aligned} \tag{A.14}$$

for every $i \in \mathbb{Z}$.

A.2.2 The case $v \le \epsilon$

The operator associated to the the explicit in time approximation is given by

$$C_{kj} = \begin{cases} \beta + 2|\alpha| 1\!\!1_{\{\alpha<0\}} & \text{if } j = k-1 \\ 1 - 2\beta - 2|\alpha| & \text{if } j = k \\ \beta + 2|\alpha| 1\!\!1_{\{\alpha>0\}} & \text{if } j = k+1 \end{cases} \qquad \text{and } C_{kj} = 0 \text{ for } |j-k| > 1.$$

α and β being given in (3.6), and the solution $u^n \in \mathbb{R}^{\mathbb{Z}}$ is given by $u^n = Cu^{n+1}$. The infinite dimensional matrix C defines a Markov transition function if and only if

$$2\beta + 2|\alpha| \le 1.$$

Here, nothing chances with respect to what developed in Section 3.1.2. So, for every $i \in \mathbb{Z}$ we get the same formulas in (3.20), that is

$$\begin{aligned} (C\psi_1^i(y))_i &= h\mu_Y(v), \\ (C\psi_2^i(y))_i &= h\rho^2 v + h\Delta y|\mu_Y(v)|, \\ (C\psi_4^i(y))_i &= h\Delta y^2\rho^2 v + h\Delta y^3|\mu_Y(v)|, \end{aligned} \tag{A.15}$$

in which we have inserted the formulas for α and β in (3.6).

Remark A.7. *By using* (A.14) *and* (A.15)*, we can prove the convergence result in Theorem 3.9 also in the case of the infinite grid. The proof is identical, even simpler because we do not have here contributions from the boundary.*

References

[1] A. Alfonsi (2010): High order discretization schemes for the CIR process: application to affine term structure and Heston models, *Mathematics of Computation*, **79**, 209-237. 22

[2] K. Amin, A. Khanna (1994): Convergence of American option values from discrete-to continuous-time financial models , *Mathematical Finance*, **4**, 289-304. 21

[3] E. Appolloni, L. Caramellino, A. Zanette (2013): A robust tree method for pricing American options with CIR stochastic interest rate. Preprint `arXiv:1305.0479v1`. 2, 5, 7, 16, 18

[4] A. Berman, R. J. Plemmons (1994): Nonnegative matrices in the mathematical sciences, *Society for Industrial and Applied Mathematics (SIAM)*, Philadelphia, PA. 10, 26

[5] P. Billingsley (1968): *Convergence of probability measures.* Wiley, New York. 17

[6] M.J. Brennan, E.S. Schwartz (1976): The pricing of equity-linked life insurance policies with an asset value guarantee, *Journal of Financial Economics*, **3**, 195-213.

[7] C. Chiarella, B.Kang, G.H. Meyer (2012): The evaluation of barrier option prices under stochastic volatility, *Computers and Mathematics with Applications*, **64**, 2034-2048. 23

[8] N. Clarke, K. Parrott. (1999): Multigrid for American option pricing with stochastic volatility. *Applied Mathematical Finance*, **6**, 177–195. 2

[9] J. C. Cox, J. Ingersoll, S. Ross (1985): A theory of the term structure of interest rates, *Econometrica*, **53**, 385-407.

[10] S.N. Ethier, T. Kurtz (1986): *Markov processes: characterization and convergence.* John Wiley & Sons, New York. 17

[11] F. Fang, C.W. Oosterlee (2011): A Fourier-based valuation method for Bermudan and barrier options under Heston's model, *SIAM J. Fin. Math*, to appear.

[12] T.J. Finuance, M.J. Tomas (1996): American stochastic volatility call option pricing: a lattice based approach. *Review of Derivatives Research*, **1**, 183–201. 2

[13] I. Florescu, F. Viens (2005): A binomial tree approach to stochastic volatility driven model of the stock price. *An. Univ. Craiova Ser. Mat. Inform.*, **32**, 126–142.

[14] I. Florescu, F. Viens (2006): Stochastic volatility: option pricing using a multinomial recombining tree. Preprint Department of Statistics, Purdue University. 1

1

[15] L.K. Guan, G. Xiaoqiang (2000): Pricing American options with stochastic volatility: Evidence from S&P futures options. *Journal of Futures Markets*, **20**, 625–659. 1

[16] T. Haentjens, K.J. in't Hout, K. Volders (2010): ADI schemes with Ikonen-Toivanen splitting for pricing American put options in the Heston model. In: *Numerical Analysis and Applied Mathematics, eds. T. E. Simos et. al., AIP Conf. Proc.* 1281, 231-234.

[17] S.L. Heston (1993): A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, **6**, 327-343. 1, 2, 22

[18] J.E. Hilliard, A. Schwartz (1996): Binomial option pricing under stochastic volatility and correlated state variables. *Journal of Derivatives*, 23–39. 1

[19] S. Ikonen, J. Toivanen (2009): Operator splitting methods for pricing American options under stochastic volatility, *Numer. Math.*, **113**, 299-324. 2

[20] K.J. in't Hout, S. Foulon (2010): ADI finite difference schemes for option pricing in the Heston model with correlation.*Int. J. Numer. Anal. Mod.* , **7**, 303-320. 2

[21] H. Kushner, P.G. Dupuis (1992): *Numerical methods for stochastic control problems in continous time.* Springer Verlag.

[22] D.P.J. Leisen (2000): Stock evolution under stochastic volatility: A discrete approach. *Journal of Derivatives*, **8**, 8–27. 1

[23] F.A. Longstaff, E.S. Schwartz (2001): Valuing American options by simulations: a simple least squares approach. *The Review of Financial Studies,* **14**, 113-148. 22

[24] D.B. Nelson, K. Ramaswamy (1990): Simple binomial processes as diffusion approximations in financial models. *The Review of Financial Studies*, **3**, 393-430. 17

[25] H. Nieuwenhuis, M. Vellekoop (2009): A tree-based method to price American Options in the Heston Model. *The Journal of Computational Finance*, **13**, 1–21. 1, 22

[26] C.W. Oosterlee (2003): On multigrid for linear complimentarity problems with application to American-style options. *Electronic Transactions on Numerical Analysis*, **15**, 165–185. 2

[27] PREMIA: An Option Pricer. `http://www.premia.fr` 22

[28] J. C. Strikwerda (2004): Finite difference schemes and partial differential equations, *Society for Industrial and Applied Mathematics (SIAM)*, Philadelphia, PA. 8

[29] D.W. Stroock, S.R.S. Varadhan (1979): *Multidimensional Diffusion Processes.* Springer, Berlin. 17

[30] L. H. Thomas (1949): Elliptic Problems in Linear Differential Equations over a Network, *Watson Sci. Comput. Lab Report*, Columbia University, New York. 26

[31] Y. Tian (1992): A simplified binomial approach to the pricing of interest-rate contingent claims. *Journal of Financial Engineering*, **1**, 4-37.

[32] Y. Tian (1994): A reexamination of lattice procedures for interest rate-contingent claims. *Advances in Futures and Options Research*, **7**, 87-110.

[33] R. Zvan, P. Forsyth, K. Vetzal (1998): A penalty method for American options with stochastic volatility. *J. Comp. Appl. Math.*, **92**, 199–218. 2