A tree-based method to price American Options in the Heston Model

Michel Vellekoop* and Hans Nieuwenhuis†

March 3, 2020

Abstract

We develop an algorithm to price American options on assets that follow the stochastic volatility model defined by Heston. We use an approach which is based on a modification of an explicitly defined stock price tree where the number of nodes grows quadratically in the number of time steps. We show in a number of numerical tests that we get accurate results in a fast manner. Features which are important for the practical use of option pricing algorithms, such as the incorporation of cash dividends and a term structure of interest rates, can easily be incorporated.

The version implemented in PREMIA is an improved version of the algorithm presented here. Note that the number of timesteps and number of grid points for stock and volatility can be freely specified in the PREMIA implementation but that the theorem in the paper states explicit conditions that need to be satisfied to obtain convergence when increasing these steps.

1 Introduction

The most popular model for equity option pricing under stochastic volatility seems to be the model defined by Heston [9]:

$$\begin{aligned} dS_t &= \mu S_t dt \ + \ \sqrt{V_t} S_t \, dW_t^1 \\ dV_t &= \kappa(\theta - V_t) dt \ + \ \omega \sqrt{V_t} \, dW_t^2 \end{aligned}$$

In this model for the stock price process S and squared volatility process V we denote by W^1 and W^2 standard Brownian Motion processes that may have nonzero correlation, and μ, κ, θ and ω are fixed and known strictly positive parameters.

The popularity of this model can be explained to a large extent by the possibility to derive option price formulas for European options in closed form using Fourier transforms [3]. This closed form actually requires the numerical approximation of a certain integral, and some care has to be taken when dealing with the complex logarithm in this integral (see the excellent recent papers by Lord, Kahl and Jäckel [16, 12] and Albracher et al [1] on this issue). But since this method is still a lot faster than methods in which the corresponding partial differential equation is solved

*Corresponding Author: Financial Engineering Laboratory, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. Tel +31 53 489 2087, Fax + 31 53 489 3800, m.h.vellekoop@math.utwente.nl

†University of Groningen, Faculty of Economics, PO Box 800, 9700 AV Groningen.

numerically, calibration of the Heston model to European options is a lot easier than calibration of other stochastic volatility models which do not admit closed form pricing functions.

For American options, or asset dynamics which involve the payment of cash dividends at fixed dividend dates instead of continuous dividend payments, such closed form pricing functions do not exist. Methods to determine the prices of such options in the Heston model therefore typically focus on solving the Heston partial differential equation under early exercise constraints.

Clarke and Parrott [4] formulate the American put pricing problem as a linear complementarity problem and use an implicit finite difference scheme combined with a multigrid procedure based on earlier work by Brandt and Cryer [2] to find price approximations. Their method was further improved by Oosterlee [18] who used Fourier analysis methods to optimize the smoothing procedure in the multigrid procedure. Forsyth, Vetzal and Zvan have used a penalty method to deal with the early exercise constraint, and showed that in the limit this is in fact equivalent to a linear complementarity formulation [21]. Ikonen and Toivanen [11] used operator splitting methods to price American options in the Heston model. They find that the error induced by the splitting does not reduce the error of accuracy when compared to Crank-Nicholson methods without splitting.

In this paper we will develop an alternative method which is based on building a discrete time process that approximates the dynamics of the Heston model. Not only will we show that American options and options including cash dividends can be priced very fast on such trees, but we will also argue that the structure is very transparent, and therefore particularly easier to implement.

Approximating trees for the Heston model have been considered before, for example in the paper by Leisen [15]. In that paper a multinomial tree is constructed with 4 successor nodes per timestep for the asset price process, and 2 succesor nodes for the volatility process. The goal behind the construction of this tree is weak convergence to the correct joint distribution of asset price and volatility when the number of timesteps grows to infinity. However, Florescu and Viens [6] mention that since a transformation to eliminate the volatility is used, weak convergence cannot be proven along the lines suggested in the paper. They therefore propose a different algorithm [7], which combines a tree with Monte Carlo simulations based on a finite number of particles between time steps, but this is computationaly expensive.

Hilliard and Schwartz [10] use a transformation of the S and V variables to end up with stochastic differential equations for the transformed processes which lead to recombining discretizations in a two-dimensional tree. The transformation itself i based on the one proposed by Nelson and Ramaswamy [17]. However, their method has some numerical problems since negative probabilities may arise in the discretized model if certain bounds on the correlation coefficient are not satisfied. Moreover, since a lot of computational time is used for calculations in parts of the state space which have a low probability, the method is not always very efficient.

Almost all tree methods proposed in the literature use a grid in which both the distance between (transformed) asset price grid points and (transformed) volatility grid points are of the order $\sqrt{\Delta t}$, where Δt denotes the discrete time step. To make sure that the stochastic process defined on the tree converges (weakly) to the correct continuous time process, second order moment matching conditions are used and this generates equations that need to be satisfies by the single time step transition probabilities under the riskneutral measure. In the detailed implementation described by Leisen [15] this leads to a tree with 8 successor nodes per node on a fixed grid.

An exception is the paper by Guan and Xiaoqiang [8], which is based on a scheme introduced in earlier work by Finucane and Thomas [5]. In the Guan and Xiaoqiang paper, a tree is constructed in which the problem that nodes do not recombine is solved by an interpolation/extrapolation scheme. A grid is define, and whenever option prices are needed that or not on the grid, interpolation or extrapolation based on neighbouring points is applied. Negative option prices may occur in the tree, but these are set to zero. The method is shown to work well for short maturity options, where the extrapolation errors are not yet too large.

In this paper, we propose an extension to this method in which we use different interpolation schemes in order to exploit the regularity of the option payoff function. We define a flexible grid for the logarithm of the stock price $Z_t = \ln S_t$ and the squared volatility process V_t. We take a meshsize which is of order $o(\Delta t)$ at every timestep. The discrete time stochastic process that we define takes its values on this grid, and has 16 or 64 successor nodes in every point. This obviously means that we require quite some computation time per timestep. But the extra degrees of freedom that we create by our setup allow us to exploit the fact that the American Option price function is once continuously differentiable in both S and V. This smoothness is used to improve the speed of convergence, which means that we require more computations per timestep but far less timesteps than in alternative methods. This means that we retain the flexibility of the Guan and Xiaoqiang method when introducing for example a term structure of interest rates and cash dividends, but at the same time we can get quicker convergence.

2 The Algorithm

The Heston model for squared volatility process V and log stock price process Z is

$$\begin{aligned} dV_t &= \kappa(\theta - V_t)dt \ + \ \omega\sqrt{V_t}dW_t^1 \\ dZ_t &= (r - \tfrac{1}{2}V_t)dt \ + \ \sqrt{V_t}dW_t^2 \end{aligned}$$

for given $(V_0, Z_0) = (\sigma^2, \ln S_0)$, where $\{(W_t^1, W_t^2),\ t \in [0,T]\}$ are correlated standard Brownian Motion processes with correlation coefficient ρ under the riskneutral measure $\mathbb{Q}$, and κ, ω, θ and r are strictly positive given constants. The value of the European option with maturity T and payoff function $\Phi : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ equals (see for example [14])

$$H_t(V_t, Z_t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[\Phi(e^{Z_T}, \sqrt{V_T}) \mid \mathcal{F}_t]$$

with $(\mathcal{F}_t)_{t\in[0,T]}$ the filtration generated by the process $\{(V_t, Z_t),\ t \in [0,T]\}$. The righthand side can be shown to be indeed a function of t, V_t and Z_t only, because of the Markov property of (V_t, Z_t).

We now define the following discrete time stochastic processes ($k = 0..n-1$)

$$\begin{aligned} V_{k+1}^n &= V_k^n + \kappa(\theta - V_k^n)\Delta t^n + Y_{k+1}^{n,1}\omega\sqrt{V_k^n\Delta t^n} \\ Z_{k+1}^n &= Z_k^n + (r - \tfrac{1}{2}V_k^n)\Delta t^n + Y_{k+1}^{n,2}\sqrt{V_k^n\Delta t^n} \end{aligned}$$

with $(V_0^n, Z_0^n) = (V_0, Z_0)$ and where the variables $(Y_k^{n,1}, Y_k^{n,2})$ are i.i.d. distributed in k, with

$$\begin{aligned} \mathbb{Q}^n(Y_k^{n,1} = +1,\ Y_k^{n,2} = +1) &= \tfrac{1}{4}(1+\rho) \\ \mathbb{Q}^n(Y_k^{n,1} = -1,\ Y_k^{n,2} = +1) &= \tfrac{1}{4}(1-\rho) \\ \mathbb{Q}^n(Y_k^{n,1} = +1,\ Y_k^{n,2} = -1) &= \tfrac{1}{4}(1-\rho) \\ \mathbb{Q}^n(Y_k^{n,1} = -1,\ Y_k^{n,2} = -1) &= \tfrac{1}{4}(1+\rho). \end{aligned}$$

under a new pricing measure $\mathbb{Q}^n$. Note that this essentially means that we define a simple Euler simulation scheme for the correlated Brownian Motion processes that drive the riskneutral dynamics. This seems the most natural discrete time stochastic process to approximate the Heston model in continuous time, but it is of course non-recombining. The number of possible values for the process (V^n, Z^n) grows exponentially in the number of timesteps and it is therefore not very well suited to do actual computations.

Let

$$z_k^{n,\max} = \max\{z : \mathbb{Q}^n(Z_k^n = z) > 0\}$$

and define $z_k^{n,\min}$, $v_k^{n,\max}$ and $v_k^{n,\min}$ analogously. We take

$$\begin{aligned} \Delta z_k^n &= (z_k^{n,\max} - z_k^{n,\min})/m_z, \\ \Delta v_k^n &= (v_k^{n,\max} - v_k^{n,\min})/m_v \end{aligned}$$

for certain $m_v, m_z \in \mathbb{N}^+$ which describe how fine the mesh is that we will take, and define the set of gridpoints in $\mathbb{R}^2$:

$$\hat{\mathcal{S}}^n_k = \{(v^{n,\min}_k + i\Delta v^n_k,\ z^{n,\min}_k + j\Delta z^n_k) \ \mid i = 0...m_v,\ j = 0...m_z\}.$$

For any set $\mathcal{S} = \mathcal{S}_x \times \mathcal{S}_y \subset \mathbb{R}^2$ of gridpoints and functions $f : \mathcal{S} \to \mathbb{R}$ we denote by $\mathcal{L}^{\mathcal{S}}[f] : \mathcal{C}(\mathcal{S}) \to \mathbb{R}$ the piecewise bilinear interpolating function corresponding to this function on the grid, where $\mathcal{C}(\mathcal{S})$ denotes the convex hull of $\mathcal{S}$. This means that

$$\begin{aligned}\mathcal{L}^{\mathcal{S}}[f](x,y) \quad &= \quad c^{\mathcal{S}}_{00}(x,y)\, f(x^{\mathcal{S}}_0(x), y^{\mathcal{S}}_0(y)) \ + \ c^{\mathcal{S}}_{10}(x,y)\, f(x^{\mathcal{S}}_1(x), y^{\mathcal{S}}_0(y)) \ + \\ & \quad\ \ c^{\mathcal{S}}_{01}(x,y)\, f(x^{\mathcal{S}}_0(x), y^{\mathcal{S}}_1(y)) \ + \ c^{\mathcal{S}}_{11}(x,y)\, f(x^{\mathcal{S}}_1(x), y^{\mathcal{S}}_1(y))\end{aligned}$$

where the functions $x^{\mathcal{S}}_0 : \mathcal{C}(\mathcal{S}) \to \mathbb{R}$ and $y^{\mathcal{S}}_0 : \mathcal{C}(\mathcal{S}) \to \mathbb{R}$ are defined in such a way that

$$(x^{\mathcal{S}}_0(x), y^{\mathcal{S}}_0(y)) \in \mathcal{S}, \qquad (x^{\mathcal{S}}_1(x), y^{\mathcal{S}}_1(y)) \in \mathcal{S}, \qquad x^{\mathcal{S}}_0(x) \le x \le x^{\mathcal{S}}_1(x), \quad y^{\mathcal{S}}_0(y) \le y \le y^{\mathcal{S}}_1(y),$$

i.e., x and y are always in between the adjacent grid points $x^{\mathcal{S}}_0(x)$, $x^{\mathcal{S}}_1(x)$ and $y^{\mathcal{S}}_0(y)$, $y^{\mathcal{S}}_1(y)$ respectively. The functions $c^{\mathcal{S}}_{ij}$ are defined by

$$\begin{aligned}c^{\mathcal{S}}_{00}(x,y) \quad &= \quad (1-\tilde{x})(1-\tilde{y}), \qquad \tilde{x} = \tfrac{x - x^{\mathcal{S}}_0(x)}{x^{\mathcal{S}}_1(x) - x^{\mathcal{S}}_0(x)}, \quad \tilde{y} = \tfrac{y - y^{\mathcal{S}}_0(y)}{y^{\mathcal{S}}_1(y) - y^{\mathcal{S}}_0(y)} \\ c^{\mathcal{S}}_{10}(x,y) \quad &= \quad \tilde{x}(1-\tilde{y}) \\ c^{\mathcal{S}}_{01}(x,y) \quad &= \quad (1-\tilde{x})\tilde{y} \\ c^{\mathcal{S}}_{11}(x,y) \quad &= \quad \tilde{x}\tilde{y}.\end{aligned}$$

We define for $u \in \{0, 1\}$

$$\begin{aligned}h^{k,n}_u(x) \quad &= \quad x - x^{\hat{\mathcal{S}}^n_k}_u(x), \\ g^{k,n}_u(y) \quad &= \quad y - y^{\hat{\mathcal{S}}^n_k}_u(y).\end{aligned}$$

Obviously, $|h^{k,n}_u(x)| < \Delta v^n_k$ and $|g^{k,n}_u(x)| < \Delta z^n_k$, and we will need these expressions in the proof of Theorem 2 in the next section.

We take a new process $(\tilde{V}^n, \tilde{Z}^n)$ on the space $\hat{\mathcal{S}}^n_k$, as follows. First we define

$$\begin{aligned}v^n(y,v,z) \quad &= \quad v + \kappa(\theta - v)\Delta t^n + y\omega\sqrt{v\Delta t^n} \\ z^n(y,v,z) \quad &= \quad z + (r - \tfrac{1}{2}v)\Delta t^n + y\sqrt{v\Delta t^n}\end{aligned}$$

where $\Delta t^n = T/n$. Let

$$(\tilde{V}^n_{k+1},\ \tilde{Z}^n_{k+1}) \quad = \quad \left(x^{\hat{\mathcal{S}}^n_{k+1}}_{Y^{n,3}_{k+1}}(v^n(Y^{n,1}_{k+1}, \tilde{V}^n_k, \tilde{Z}^n_k)),\ y^{\hat{\mathcal{S}}^n_{k+1}}_{Y^{n,4}_{k+1}}(z^n(Y^{n,2}_{k+1}, \tilde{V}^n_k, \tilde{Z}^n_k)) \right)$$

where under a new measure $\tilde{\mathbb{Q}}^n$, the $(Y^{n,1}, Y^{n,2})$ have the same distribution as under $\mathbb{Q}^n$

$$\tilde{\mathbb{Q}}^n(Y^{n,1}_{k+1} = i,\ Y^{n,2}_{k+1} = j) \quad = \quad \tfrac{1}{4}(1 + ij\rho) \qquad i,j \in \{-1, 1\} \tag{2.1}$$

and are independent for different values of k, while the $(Y^{n,3}, Y^{n,4})$ have the following conditional distribution:

$$\begin{aligned}&\tilde{\mathbb{Q}}^n(Y^{n,3}_{k+1} = i,\ Y^{n,4}_{k+1} = j \mid (Y^{n,1}_{k+1}, Y^{n,2}_{k+1}, \tilde{V}^n_k, \tilde{Z}^n_k)) \\ &= \quad c^{\hat{\mathcal{S}}^n_{k+1}}_{ij}(v^n(Y^{n,1}_{k+1}, \tilde{V}^n_k, \tilde{Z}^n_k), z^n(Y^{n,2}_{k+1}, \tilde{V}^n_k, \tilde{Z}^n_k)) \qquad i,j \in \{0,1\}\end{aligned}$$

and are conditionally independent for different values of k. Note that this means that in every time step there are four times four is 16 new possible values for $(\tilde{V}^n_{k+1},\ \tilde{Z}^n_{k+1})$ based on the current value of $(\tilde{V}^n_k,\ \tilde{Z}^n_k)$, since every one of the four next time step values is split over four new points.

Having thus defined a discrete-time homogeneous Markov process, we will now investigate the weak convergence of this stochastic process in the next section.

3 Weak Convergence

From now on, we take $T = 1$ without loss of generality. We will use the following theorem by Stroock and Varadhan [20] (see their section 11.2).

Theorem 1. *Let $\mathbb{Q}^n$ be a probability measure and Π_n a transition function on $\mathbb{R}^d$ such that $\mathbb{Q}^n$-a.s. $X_0^n = x_0$ and for all $k > 0$*

$$X_t^n = (k+1-tn)X_{k/n} + (tn-k)X_{(k+1)/n}, \qquad \tfrac{k}{n} \le t \le \tfrac{k+1}{n}.$$

Define for all $k > 0$ and all Borel sets A in $\mathbb{R}^d$

$$\Pi_n(X_{k/n}, A) \quad = \quad \mathbb{Q}^n(X_{(k+1)/n} \in A \mid \mathcal{F}_{k/n}),$$

where $\mathcal{F}_{k/n} = \sigma(\{X_t,\ 0 \le t \le k/n\})$ and assume that for all $R > 0$, $\epsilon > 0$ and certain continuous functions $a : \mathbb{R}^d \to S_d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$:

$$\lim_{n\to\infty} \sup_{\|x\|\le R} \|a_n(x) - a(x)\| = 0, \qquad a_n^{ij}(x) = n \int_{\|y-x\|\le 1} (y_i - x_i)(y_j - x_j)\Pi_n(x, dy)$$

$$\lim_{n\to\infty} \sup_{\|x\|\le R} \|b_n(x) - b(x)\| = 0, \qquad b_n^{i}(x) = n \int_{\|y-x\|\le 1} (y_i - x_i)\Pi_n(x, dy)$$

$$\lim_{n\to\infty} \sup_{\|x\|\le R} \Delta_n^{\epsilon}(x) = 0, \qquad \Delta_n^{\epsilon}(x) = n\,\Pi^n(x,\ \{y \in \mathbb{R}^d :\ \|y-x\| > \epsilon\}).$$

If the martingale problem for a and b has exactly one solution $\mathbb{Q}_x$ for every initial value $x_0 = x$ then $\mathbb{Q}^n$ converges weakly to $\mathbb{Q}_{x_0}$, uniformly on compact subsets of $\mathbb{R}^d$.

We take $\hat{V}^n_{k/n} = \tilde{V}^n_k$ and $\hat{Z}^n_{k/n} = \tilde{Z}^n_k$ and define

$$\begin{aligned}
\hat{V}_t^n &= (k+1-tn)\hat{V}^n_{k/n} + (tn-k)\hat{V}^n_{(k+1)/n}, \qquad \tfrac{k}{n} \le t \le \tfrac{k+1}{n} \\
\hat{Z}_t^n &= (k+1-tn)\hat{Z}^n_{k/n} + (tn-k)\hat{Z}^n_{(k+1)/n}.
\end{aligned}$$

Our main result is the following.

Theorem 2. *Assume that*

$$\lim_{n\to\infty} \Delta t^n = 0, \qquad \lim_{n\to\infty} \max_{k=1..n-1} \frac{\Delta v_k^n}{\Delta t^n} = 0, \qquad \lim_{n\to\infty} \max_{k=1..n-1} \frac{\Delta z_k^n}{\Delta t^n} = 0.$$

Then the processes $(\hat{V}^n, \hat{Z}^n)$ converge weakly to the process (V, Z), i.e.

$$((\hat{V}^n, \hat{Z}^n) \mid \tilde{\mathbb{Q}}^n) \qquad \overset{\mathcal{W}}{\to} \qquad ((V, Z) \mid \mathbb{Q})$$

where $\mathcal{W}$ stands for weak convergence in the space of continuous processes on $[0, 1]$.

Proof.
Define $X^n_{k/n} = (\tilde{V}^n_k, \tilde{Z}^n_k)$ and

$$\Delta X_k^n \quad = \quad X^n_{(k+1)/n} - X^n_{k/n}.$$

Take $x = (v, z)$. Since the drift and diffusion functions of the Heston-model

$$b(x) \quad = \quad \begin{bmatrix} \kappa(\theta - v) \\ r - \frac{1}{2}v^2 \end{bmatrix}, \qquad a(x) \quad = \quad v \begin{bmatrix} \omega^2 & \rho\omega \\ \rho\omega & 1 \end{bmatrix}.$$

define a unique solution for the associated martingale problem (see for example [13]), we only need to check the conditions stated in the previous theorem. This boils down to proving that for all $R > 0$ and all $\epsilon > 0$ we have

$$
\begin{aligned}
0 &= \lim_{n\to\infty} \sup_{\|x\|\le R} \left\| b(x) - n\,\mathbb{E}^{\tilde{\mathbb{Q}}^n}\left[(\Delta X_k^n)\,\mathbf{1}_{\{\|\Delta X_k^n\|\le 1\}} \middle|\, X_{k/n}^n = x \right] \right\| \\
0 &= \lim_{n\to\infty} \sup_{\|x\|\le R} \left\| a(x) - n\,\mathbb{E}^{\tilde{\mathbb{Q}}^n}\left[(\Delta X_k^n)\,(\Delta X_k^n)^T\,\mathbf{1}_{\{\|\Delta X_k^n\|\le 1\}} \middle|\, X_{k/n}^n = x \right] \right\| \\
0 &= \lim_{n\to\infty} \sup_{\|x\|\le R} n\,\tilde{\mathbb{Q}}^n(\,\|\Delta X_k^n\| > \epsilon \mid X_{k/n}^n = x)
\end{aligned}
$$

Let $X_{k/n}^n = x = (v, z)$. Then

$$
\Delta X_k^n = \begin{bmatrix} \tilde{V}_{k+1}^n - \tilde{V}_k^n \\ \tilde{Z}_{k+1}^n - \tilde{Z}_k^n \end{bmatrix} = \begin{bmatrix} x_{Y_{k+1}^{n,3}}^{\hat{S}_{k+1}^n}(v^n(Y_{k+1}^{n,1}, v, z)) - v \\ y_{Y_{k+1}^{n,4}}^{\hat{S}_{k+1}^n}(z^n(Y_{k+1}^{n,2}, v, z)) - z \end{bmatrix} = \begin{bmatrix} v^n(v, z, Y_{k+1}^{n,1}) - v + h_{Y_{k+1}^{n,3}}^{k+1,n}(v) \\ z^n(v, z, Y_{k+1}^{n,2}) - z + g_{Y_{k+1}^{n,4}}^{k+1,n}(z) \end{bmatrix}
$$

so

$$
\Delta X_k^n = \begin{bmatrix} \kappa(\theta - v)\Delta t^n + Y_{k+1}^{n,1}\,\omega\sqrt{v\Delta t^n} \\ (r - \frac{1}{2}v)\Delta t^n + Y_{k+1}^{n,2}\,\sqrt{v\Delta t^n} \end{bmatrix} + \begin{bmatrix} h_{Y_{k+1}^{n,3}}^{k+1,n}(v) \\ g_{Y_{k+1}^{n,4}}^{k+1,n}(z) \end{bmatrix} \tag{3.2}
$$

with

$$
\tilde{\mathbb{Q}}^n(Y_{k+1}^{n,1} = i,\ Y_{k+1}^{n,2} = j) = \tfrac{1}{4}(1 + ij\rho) \qquad i, j \in \{-1, 1\}
$$

Since $\Delta t^n = 1/n$ we see from (3.2) that for $n > n_1$ with n_1 large enough we have $\|\Delta X_k^n\| \le 1$ since $|Y_{k+1}^{n,u}| = 1$ and at the same time

$$
|h_u^{k+1,n}(v)| < \Delta v_{k+1}^n, \qquad |g_u^{k+1,n}(z)| < \Delta z_{k+1}^n.
$$

which implies that $h_u^{k+1,n} = o(\Delta t^n)$ and $g_u^{k+1,n} = o(\Delta t^n)$ for $u \in \{0, 1\}$. But this means that for $n > n_1$we find

$$
b(x) - n\,\mathbb{E}^{\tilde{\mathbb{Q}}^n}\left[(\Delta X_k^n)\,\mathbf{1}_{\{\|\Delta X_k^n\|\le 1\}} \middle|\, X_{k/n}^n = x \right] = \mathbb{E}^{\tilde{\mathbb{Q}}^n} \begin{bmatrix} h_{Y_{k+1}^{n,3}}^{k+1,n}(v) \\ g_{Y_{k+1}^{n,4}}^{k+1,n}(z) \end{bmatrix}
$$

and the norm of this clearly converges to zero. A similar argument shows that the second condition is satisfied as well, because

$$
a(x) - n\,\mathbb{E}^{\tilde{\mathbb{Q}}^n}\left[(\Delta X_k^n)\,(\Delta X_k^n)^T\,\mathbf{1}_{\{\|\Delta X_k^n\|\le 1\}} \middle|\, X_{k/n}^n = x \right] = \begin{bmatrix} \tilde{a}_{11}^n & \tilde{a}_{12}^n \\ \tilde{a}_{12}^n & \tilde{a}_{22}^n \end{bmatrix}
$$

for $n > n_1$, with

$$
\begin{aligned}
\tilde{a}_{11}^n &= v\omega^2 - \mathbb{E}^{\tilde{\mathbb{Q}}^n}\tfrac{1}{\Delta t^n}\left(\kappa(\theta - v)\Delta t^n + Y_{k+1}^{n,1}\,\omega\sqrt{v\Delta t^n} + o(\Delta t^n)\right)^2 \\
\tilde{a}_{12}^n &= v\rho\omega - \mathbb{E}^{\tilde{\mathbb{Q}}^n}\tfrac{1}{\Delta t^n}\left((r - \tfrac{1}{2}v)\Delta t^n + Y_{k+1}^{n,2}\,\sqrt{v\Delta t^n} + o(\Delta t^n)\right)\left(\kappa(\theta - v)\Delta t^n + Y_{k+1}^{n,1}\,\omega\sqrt{v\Delta t^n} + o(\Delta t^n)\right) \\
\tilde{a}_{22}^n &= v - \mathbb{E}^{\tilde{\mathbb{Q}}^n}\tfrac{1}{\Delta t^n}\left((r - \tfrac{1}{2}v)\Delta t^n + Y_{k+1}^{n,2}\,\sqrt{v\Delta t^n} + o(\Delta t^n)\right)^2
\end{aligned}
$$

which shows convergence in norm. The third condition is trivial since when $\|x\| < R$

$$
\begin{aligned}
&\tilde{\mathbb{Q}}^n(\|\Delta X_k^n\| > \epsilon \mid X_{k/n}^n = x) \\
&\quad = \tilde{\mathbb{Q}}^n\left(\left\| \begin{bmatrix} \kappa(\theta - v)\Delta t^n + Y_{k+1}^{n,1}\,\omega\sqrt{v\Delta t^n} \\ (r - \frac{1}{2}v)\Delta t^n + Y_{k+1}^{n,2}\,\sqrt{v\Delta t^n} \end{bmatrix} + \begin{bmatrix} h_{Y_{k+1}^{n,3}}^{k+1,n}(v) \\ g_{Y_{k+1}^{n,4}}^{k+1,n}(z) \end{bmatrix} \right\| > \epsilon\right)
\end{aligned}
$$

and this is actually equal to zero if we take n large enough, since the vector of which the norm is taken can be made arbitrarily small since $|Y_{k+1}^{n,u}| = 1$ for $u \in \{1,2\}$ while $h_u^{k+1,n} = o(\Delta t^n)$ and $g_u^{k+1,n} = o(\Delta t^n)$ for $u \in \{0,1\}$. ■

The proof given above relies on the fact that we use a bilinear interpolation for option price values which fall outside grid points, as can clearly be seen from the functions $c_{ij}^{\mathcal{S}}$ that we defined in Section 2. Obviously, we could use other approximations as well. For example, interpolation based on bicubic splines instead of bilinear ones would use

$$\mathcal{L}^{\mathcal{S}}[f](x,y) = \sum_{i=-1}^{2} \sum_{j=-1}^{2} c_{ij}^{\mathcal{S}}(x,y)\, f(x_{\mathrm{i}}^{\mathcal{S}}(x), y_{\mathrm{j}}^{\mathcal{S}}(y))$$

where the functions $x_0^{\mathcal{S}}$ and $y_0^{\mathcal{S}}$ are defined as before, $x_{-1}^{\mathcal{S}}(x) = x_0^{\mathcal{S}}(x) - \Delta x$, $x_2^{\mathcal{S}}(x) = x_1^{\mathcal{S}}(x) + \Delta x$, and similar definitions hold for $y_{-1}^{\mathcal{S}}$ and $y_2^{\mathcal{S}}$, while

$$\begin{aligned} c_{00}^{\mathcal{S}}(x,y) &= \tfrac{1}{4}(\tilde{x}-1)(\tilde{y}-1)(3\tilde{x}^2 - 2\tilde{x} - 2)(3\tilde{y}^2 - 2\tilde{y} - 2) \\ c_{-1,-1}^{\mathcal{S}}(x,y) &= \tfrac{1}{4}\tilde{x}(\tilde{x}-1)^2\tilde{y}(\tilde{y}-1)^2 \\ c_{0,-1}^{\mathcal{S}}(x,y) &= -\tfrac{1}{4}\tilde{x}(\tilde{x}-1)^2\tilde{y}(3\tilde{y}^2 - 2\tilde{y} - 2) \end{aligned}$$

while the rest follows from symmetries: $c_{i,j} \to c_{j,i}$ means $(\tilde{x},\tilde{y}) \to (\tilde{y},\tilde{x})$, $c_{i,j} \to c_{1-i,j}$ means $(\tilde{x},\tilde{y}) \to (1-\tilde{x},\tilde{y})$ and finally $c_{i,j} \to c_{i,1-j}$ means $(\tilde{x},\tilde{y}) \to (\tilde{x}, 1-\tilde{y})$.

For this bicubic interpolation we cannot give a similar convergence proof as for the bilinear case. In the bilinear case we use the fact that the functions c_{ij} all map into $[0,1]$ while the sum over all i and j of $c_{ij}(x,y)$ equals 1 for all x and y. This allows us to use the bilinear interpolation to redistribute probability mass over neighbouring points, to create a stochastic process and use weak convergence results for such processes. However, when we use bicubic interpolation some of the weights $c_{ij}(x,y)$ can be negative so we cannot use the probabilistic result we used earlier.

Numerical error analysis helps to illustrate what may go wrong. If some weights $c_{ij}(x,y)$ are negative, numerical errors that we introduce in our scheme may increase over time. Indeed, we would like to have that $\mathbb{E}^{\tilde{\mathbb{Q}}^n}\mathcal{L}^{\mathcal{S}}[\Delta f] \le \sup_{\mathcal{S}} \Delta f$ for an error function Δf on the grid. This is satisfied for bilinear interpolation, but not for bicubic interpolation. It would therefore be better to use a local scheme which gives positive values for all positive functions and is C^1 at the same time. Such grids exist [19] but are very computationally intensive. Therefore we prefer to use the bicubic interpolation. Notice that if the errors Δf do not change signs in neighbouring points all the time, the errors can be expected to decrease on average and we will see in numerical results that this indeed what is observed in practice.

4 Numerical Results

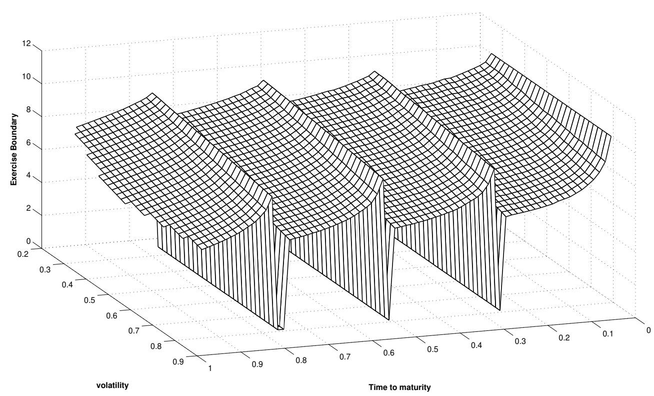
In Tables 1 to 8 we report the results of some experiments that we performed with our new numerical method. The test case we took is the same as in earlier papers that deal with American options in the Heston model [2, 4, 11, 18, 21]. The relevant parameter values are

$$\kappa = 5, \qquad \theta = 0.16, \qquad \omega = 0.9, \qquad \rho = 0.1, \qquad r = 0.10.$$

The strike was chosen to be $K = 10.0$ and different values of the starting volatility V_0 and initial stock price S_0 were considered.

Tables 1 and 2 give the option prices, and relative errors for the method involving 16 successors for every state, while Tables 3 and 4 provide the same for the model with 64 successors per state. To get error estimates we used exact Heston option prices for the European case, and the results by Zven, Forsyth and Vetzal in the American case; the results reported by other authors are close to these. Results are provided for two initial volatility values, $\sqrt{V_0} = 25\%$ and $\sqrt{V_0} = 50\%$, and for initial stock prices $S_0 \in \{8, 9, 10, 11, 12\}$, i.e. the values that were used earlier in the papers mentioned above. Per option that we consider, we present the results for our algorithm with four different number of timesteps and grid sizes to show the speed of convergence. The results by other authors are indicated by ZFV (Zven, Forsyth, Vetzal), IT-PSOR (Ikonen and Toivanen), OO (Oosterlee) and CP (Clarke and Parott). Errors are based on the Heston closed-form formula for European options, and on the results by Zven, Forsyth, and Vetzal for American options; these values are close to the other values reported.

We see from these tables that the bicubic method indeed peforms considerably better than the bilinear method. Indeed, for the bicubic method we find option prices with errors smaller than one cent (or maximally 1%) even for the smallest number of timesteps. The errors converge to zero in all cases but it is clear that the bicubic method converges a lot faster, while the computational time is of the same order of magnitude as for the bilinear calculations. The CPU times for the smallest grid are approximately 0.19 seconds for the method with 16 successors and 0.35 seconds for the method with 64 successors, while on the largest grids they are 17 and 27 seconds respectively[a].



5 Conclusions

In this paper we have shown that tree-based methods can be used to find the price of American Options in the Heston model. We believe that the speed and flexibility of this method may make it a suitable alternative for the PDE-based methods that have been mentioned in this paper. Moreover, the basic idea behind our method, to approximate a diffusion process on a grid which has a finer mesh for ΔS and ΔV for the stock and volatility than the usual square root of the time spacing Δt, may be applied to other methods as well. As indicated in the paper, the precise properties of the interpolation schemes, such as whether strictly positive functions are mapped to strictly positive interpolating functions, will be essential for the further analysis of such schemes. Interpolation chemes exist which preserve positivity, monotonicity, differentiability and local convexity in the space variables and once these become computationally feasible. possibly even better methods to approximate American options in the Heston model may become available.

References

[1] H. Albrecher, P. Mayer, W. Schoutens, and J. Tistaert. The little Heston trap. Technical Report K.U. Leuven, Section of Statistics, 2006. 1

[2] Brandt C. and Cryer C.W. Multigrid algorithms for the solution of linear complimentarity problems arising from free boundary problems. *SIAM Journal of Scientific and Statistical Computing*, 4:655–684, 1983. 2, 7

[a] Algorithms were implemented in Matlab on a Pentium 4, 2GHz machine

[3] P. Carr and D. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2:61–73, 1998. 1

[4] N. Clarke and K. Parrott. Multigrid for American option pricing with stochastic volatility. *Applied Mathematical Finance*, 6(3):177–195, 1999. 2, 7

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

[6] I. Florescu and F. Viens. A binomial tree approach to stochastic volatility driven model of the stock price. *An. Univ. Craiova Ser. Mat. Inform.*, 32:126–142, 2005. 2

[7] I. Florescu and F. Viens. Stochastic volatility: option pricing using a multinomial recombining tree. Preprint Department of Statistics, Purdue University, 2006. 2

[8] L.K. Guan and G. Xiaoqiang. Pricing American options with stochastic volatility: Evidence from S&P futures options. *Journal of Futures Markets*, 20:625–659, 2000. 2

[9] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993. 1

[10] J.E. Hilliard and A. Schwartz. Binomial option pricing under stochastic volatility and correlated state variables. *Journal of Derivatives*, pages 23–39, 1996. 2

[11] S. Ikonen and J. Toivanen. Operator splitting methods for pricing American options with stochastic volatility. In *European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)*, Jyväskylä, 2004. 2, 7

[12] C. Kahl and P. Jackel. Not-so-complex logarithms in the Heston model. *Wilmott Magazine*, pages 94–103, 2005. 1

[13] J. Kallsen. A didactic note on affine stochastic volatility models. Preprint, TU München, 2004. 6

[14] I. Karatzas and S. Shreve. *Methods of Mathematical Finance*. Springer-Verlag, 1998. 3

[15] D.P.J. Leisen. Stock evolution under stochastic volatility: A discrete approach. *Journal of Derivatives*, 8:8–27, 2000. 2

[16] R. Lord and C. Kahl. Why the rotation count algorithm works. Working Paper, University of Wuppertal (Submitted for Publication), 2006. 1

[17] D. Nelson and K. Ramaswamy. Simple binomial processes as diffusion approximations in financial models. *Review of Financial Studies*, 3:393–430, 1990. 2

[18] C.W. Oosterlee. On multigrid for linear complimentarity problems with application to American-style options. *Electronic Transactions on Numerical Analysis*, 15:165–185, 2003. 2, 7

[19] J.W. Schmidt. Positive and S-convex C-1-interpolation of gridded 3-dimensional data. *Numerical functional analysis and optimization*, 16:233–246, 1995. 7

[20] D.W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*. Springer, Berlin, 1979. 5

[21] R. Zvan, P. Forsyth, and K. Vetzal. A penalty method for American options with stochastic volatility. *J. Comp. Appl. Math.*, 92:199–218, 1998. 2, 7

			S_0				
steps s	steps v	steps t	8	9	10	11	12
125	6	25	1.8427	1.0572	0.5153	0.2267	0.0975
250	12	35	1.8417	1.0543	0.5103	0.2206	0.0920
500	24	50	1.8408	1.0527	0.5077	0.2169	0.0884
1000	48	71	1.8402	1.0515	0.5060	0.2142	0.0857
		(Exact)	1.8389	1.0483	0.5015	0.2082	0.0804
		(Error)	0.2%	0.8%	2.8%	8.9%	21.3%
			0.2%	0.6%	1.8%	5.9%	14.3%
			0.1%	0.4%	1.2%	4.2%	9.9%
			0.1%	0.3%	0.9%	2.9%	6.6%

Table 1: European Option, 16 Successors, $\sqrt{V_0} = 0.25$

			S_0					
steps s	steps v	steps t	8	9	10	11	12	
125	6	25	1.9933	1.1189	0.5366	0.2333	0.0995	
250	12	35	1.9948	1.1149	0.5307	0.2268	0.0938	
500	24	50	1.9960	1.1128	0.5276	0.2230	0.0902	
1000	48	71	1.9970	1.1112	0.5254	0.2201	0.0875	
			2.0000	1.1076	0.5202	0.2138	0.0821	ZFV
			2.0000	1.1074	0.5190	0.2130	0.0818	IT-PSOR
			2.0000	1.1070	0.5170	0.2120	0.0815	Oo
			2.0000	1.1080	0.5316	0.2261	0.0907	CP
		(Error)	-0.3%	1.0%	3.2%	9.1%	21.2%	(ZFV)
			-0.3%	0.7%	2.0%	6.1%	14.3%	
			-0.2%	0.5%	1.4%	4.3%	9.9%	
			-0.1%	0.3%	1.0%	3.0%	6.6%	

Table 2: American Option, 16 Successors, , $\sqrt{V_0} = 0.25$

			S_0				
steps s	steps v	steps t	8	9	10	11	12
125	6	25	1.8378	1.0474	0.5001	0.2080	0.0802
250	12	35	1.8378	1.0479	0.5012	0.2079	0.0797
500	24	50	1.8380	1.0481	0.5015	0.2079	0.0798
1000	48	71	1.8382	1.0481	0.5016	0.2080	0.0799
		(Exact)	1.8389	1.0483	0.5015	0.2082	0.0804
		(Error)	-0.1%	-0.1%	-0.3%	-0.1%	-0.3%
			-0.1%	0.0%	0.0%	-0.1%	-0.9%
			0.0%	0.0%	0.0%	-0.1%	-0.8%
			0.0%	0.0%	0.0%	-0.1%	-0.6%

Table 3: European Option, 64 Successors, $\sqrt{V_0} = 0.25$

			S_0					
steps s	steps v	steps t	8	9	10	11	12	
125	6	25	1.9922	1.1073	0.5192	0.2133	0.0815	
250	12	35	1.9942	1.1075	0.5199	0.2133	0.0812	
500	24	50	1.9957	1.1076	0.5201	0.2133	0.0813	
1000	48	71	1.9968	1.1076	0.5202	0.2134	0.0815	
			2.0000	1.1076	0.5202	0.2138	0.0821	ZFV
			2.0000	1.1074	0.5190	0.2130	0.0818	IT-PSOR
			2.0000	1.1070	0.5170	0.2120	0.0815	Oo
			2.0000	1.1080	0.5316	0.2261	0.0907	CP
		(Error)	-0.4%	0.0%	-0.2%	-0.2%	-0.7%	(ZFV)
			-0.3%	0.0%	-0.1%	-0.2%	-1.1%	
			-0.2%	0.0%	0.0%	-0.2%	-1.0%	
			-0.2%	0.0%	0.0%	-0.2%	-0.7%	

Table 4: American Option, 64 Successors, $\sqrt{V_0} = 0.25$

steps s	steps v	steps t	S_0 = 8	9	10	11	12
125	6	25	1.9837	1.2933	0.7887	0.4582	0.2599
250	12	35	1.9816	1.2886	0.7821	0.4506	0.2520
500	24	50	1.9803	1.2864	0.7786	0.4462	0.2472
1000	48	71	1.9794	1.2846	0.7760	0.4430	0.2438
		(Exact)	1.9773	1.2800	0.7697	0.4361	0.2373
		(Error)	0.3%	1.0%	2.5%	5.1%	9.5%
			0.2%	0.7%	1.6%	3.3%	6.2%
			0.2%	0.5%	1.2%	2.3%	4.2%
			0.1%	0.4%	0.8%	1.6%	2.8%

Table 5: European Option, 16 Successors, $\sqrt{V_0} = 0.50$

steps s	steps v	steps t	S_0 = 8	9	10	11	12	
125	6	25	2.0873	1.3514	0.8190	0.4729	0.2669	
250	12	35	2.0839	1.3450	0.8109	0.4644	0.2584	
500	24	50	2.0820	1.3416	0.8064	0.4594	0.2533	
1000	48	71	2.0809	1.3392	0.8033	0.4559	0.2498	
		(Exact)	2.0784	1.3337	0.7961	0.4483	0.2428	ZFV
			2.0783	1.3335	0.7958	0.4481	0.2427	IT-PSOR
			2.0790	1.3340	0.7960	0.4490	0.2430	OO
			2.0733	1.3290	0.7992	0.4536	0.2502	CP
		(Error)	0.4%	1.3%	2.9%	5.5%	9.9%	(ZFV)
			0.3%	0.8%	1.9%	3.6%	6.4%	
			0.2%	0.6%	1.3%	2.5%	4.3%	
			0.1%	0.4%	0.9%	1.7%	2.9%	

Table 6: American Option, 16 Successors, $\sqrt{V_0} = 0.50$

steps s	steps v	steps t	S_0 = 8	9	10	11	12
125	6	25	1.9781	1.2832	0.7740	0.4396	0.2395
250	12	35	1.9777	1.2823	0.7730	0.4387	0.2386
500	24	50	1.9776	1.2816	0.7720	0.4379	0.2382
1000	48	71	1.9775	1.2811	0.7713	0.4374	0.2379
		(Exact)	1.9773	1.2800	0.7697	0.4361	0.2373
		(Error)	0.0%	0.2%	0.6%	0.8%	1.0%
			0.0%	0.2%	0.4%	0.6%	0.6%
			0.0%	0.1%	0.3%	0.4%	0.4%
			0.0%	0.1%	0.2%	0.3%	0.3%

Table 7: European Option, 64 Successors, $\sqrt{V_0} = 0.50$

steps s	steps v	steps t	S_0 = 8	9	10	11	12	
125	6	25	2.0797	1.3377	0.8011	0.4522	0.2452	
250	12	35	2.0793	1.3365	0.7996	0.4510	0.2441	
500	24	50	2.0790	1.3357	0.7986	0.4502	0.2437	
1000	48	71	2.0789	1.3351	0.7978	0.4497	0.2435	
			2.0784	1.3337	0.7961	0.4483	0.2428	ZFV
			2.0783	1.3335	0.7958	0.4481	0.2427	IT-PSOR
			2.0790	1.3340	0.7960	0.4490	0.2430	Oo
			2.0733	1.3290	0.7992	0.4536	0.2502	CP
		(Error)	0.1%	0.3%	0.6%	0.9%	1.0%	(ZFV)
			0.0%	0.2%	0.4%	0.6%	0.5%	
			0.0%	0.1%	0.3%	0.4%	0.4%	
			0.0%	0.1%	0.2%	0.3%	0.3%	

Table 8: American Option, 64 Successors, $\sqrt{V_0} = 0.50$