Option Pricing in Heston Model using Finite Element Methods

Jan. 2010

last update: April 5, 2020

ABSTRACT

Heston stochastic volatility moded has been widely used in financial derivative pricing and risk management. One of the reasons for that is that vanilla options in Heston model have close form solutions. This makes the calibration of the model computationally much efficient and accurate. The essay in the first half introduces Fourier transform method for option pricing and its application in the Heston model. Additionally, a characteristic function based method is also described, which extends the Heston model to have time-dependent model parameters.

In the second half of the essay, we present a detailed implementation of Finite element method (FEM) for option pricing in the Heston model. FEM has been developed for decades for solving partial differential equations arise from various science and engineering problems. It is well known for requiring a low order storage and for its capability to handle complicated irregular computational domains comparing with finite difference method (FDM). This characteristic advantages make FEM an ideal numerical method for pricing exotic options.

TABLE OF CONTENTS

This note is to summarize Heston stochastic volatility model. It was initially prepared as degree essay for my M.S. in Mathematical Finance, which only covered the barrier option pricing by solving PDE using finite element method (i.e. the Chapter 2.5 in this note). I have been continuously expanding it with more mathematical background, such as the derivation of market price of spot/volatility risk, the Fourier transform method for option pricing, the derivation of characteristic function of the joint spot-variance process, the probability distribution of spot return, the piecewise time dependent Heston parameters, etc. The very first topic we want to begin with is the derivation of the Kolmogorov forward and backward equations, which fundamentally govern the transition probability density function of a diffusion process.

1. KOLMOGOROV FORWARD AND BACKWARD EQUATIONS

The time evolution of the transition probability density function is governed by *Kolmogorov* forward and backward equations, which will be introduced as follows, without loss of generality, in multi-dimension.

1.1. Kolmogorov Forward Equation

Let's consider the following m-dimensional stochastic spot process $X_t \in \mathbb{R}^m$ driven by an n-dimensional Brownian motion W_t whose correlation matrix ρ is given by $\rho dt = dW_t dW_t'$

$$\underset{m\times 1}{dX_t} = \underset{m\times 1}{A(t, X_t)}\,\underset{1\times 1}{dt} + \underset{m\times n}{B(t, X_t)}\,\underset{n\times 1}{dW_t} \tag{1}$$

We derive the dynamics of h, where $h: \mathbb{R}^m \longrightarrow \mathbb{R}$ is a scalar-valued Borel-measurable function only on variable X_t

$$\underset{1\times 1}{dh(X_t)} = \underset{1\times m}{J_h}\,\underset{m\times 1}{dX_t} + \frac{1}{2}\underset{1\times m}{dX_t'}\,\underset{m\times m}{H_h}\,\underset{m\times 1}{dX_t} = J_h A dt + J_h B dW_t + \frac{1}{2} dW_t' B' H_h B dW_t \tag{2}$$

where J_h is the $1 \times m$ *Jacobian* (i.e. the same as *gradient* if h is a scalar-valued function) and H_h the $m \times m$ *Hessian* (with subscripts now denoting the indices of vector components)

$$J_h = \left(\frac{\partial h}{\partial X_1} \quad \cdots \quad \frac{\partial h}{\partial X_m}\right), \qquad H_h = \begin{pmatrix} \frac{\partial^2 h}{\partial X_1^2} & \cdots & \frac{\partial^2 h}{\partial X_1 \partial X_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 h}{\partial X_m \partial X_1} & \cdots & \frac{\partial^2 h}{\partial X_m^2} \end{pmatrix} \tag{3}$$

Expanding the expression in (2), we have

$$\begin{aligned} dh(X_t) &= \sum_{i=1}^{m} \frac{\partial h}{\partial X_i} A_i dt + \sum_{i=1}^{m} \frac{\partial h}{\partial X_i} \sum_{k=1}^{n} B_{ik} dW_{k,t} + \frac{1}{2} \sum_{i,j=1}^{m} \frac{\partial^2 h}{\partial X_i \partial X_j} \sum_{k=1}^{n} B_{ik} \rho_{ij} B_{jk} dt \\ &= \left(\sum_{i=1}^{m} A_i \frac{\partial h}{\partial X_i} + \frac{1}{2} \sum_{i,j=1}^{m} \Sigma_{ij} \frac{\partial^2 h}{\partial X_i \partial X_j} \right) dt + \sum_{i=1}^{m} \frac{\partial h}{\partial X_i} \sum_{k=1}^{n} B_{ik} dW_{k,t} \end{aligned} \tag{4}$$

where $\Sigma = B\rho B'$ is the $m \times m$ instantaneous variance-covariance matrix of dX. Integrating on both sides of (4) from initial time s to time t, we have

$$h(X_t) - h(X_s) = \int_s^t \left(\sum_{i=1}^{m} A_i \frac{\partial h}{\partial X_i} + \frac{1}{2} \sum_{i,j=1}^{m} \Sigma_{ij} \frac{\partial^2 h}{\partial X_i \partial X_j} \right) du + \int_s^t \sum_{i=1}^{m} \frac{\partial h}{\partial X_i} \sum_{k=1}^{n} B_{ik} dW_{k,t} \tag{5}$$

Taking expectation on both sides of (5), we get (using notation $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_t]$)

$$\begin{aligned} \text{LHS} &= \mathbb{E}_s[h(X_t)] - h(X_s) = \int_\Omega h_x p_{t,x|s,\alpha} dx - h_\alpha \\ \text{RHS} &= \int_s^t \sum_{i=1}^{m} \mathbb{E}_s \left[A_i \frac{\partial h}{\partial X_i} \right] du + \frac{1}{2} \int_s^t \sum_{i,j=1}^{m} \mathbb{E}_s \left[\Sigma_{ij} \frac{\partial^2 h}{\partial X_i \partial X_j} \right] du \end{aligned} \tag{6}$$

where $p_{t,x|s,\alpha}$ is the transition probability density function having $X_t = x$ at t given $X_s = \alpha$ at s (i.e. if we solve the equation (1) with the initial condition $X_s = \alpha \in \Omega \equiv \mathbb{R}^m$, then the random variable $X_t = x \in \Omega$ has a density $p_{t,x|s,\alpha}$ in the x variable at time t). Differentiating (6) with respect to t on both sides, we get

$$\int_{\Omega} h_x \frac{\partial p_{t,x|s,\alpha}}{\partial t} dx = \sum_{i=1}^{m} \mathbb{E}_s \left[A_i \frac{\partial h}{\partial X_i} \right] + \frac{1}{2} \sum_{i,j=1}^{m} \mathbb{E}_s \left[\Sigma_{ij} \frac{\partial^2 h}{\partial X_i \partial X_j} \right]$$
$$= \sum_{i=1}^{m} \int_{\Omega} p_{t,x|s,\alpha} A_i \frac{\partial h_x}{\partial x_i} dx + \frac{1}{2} \sum_{i,j=1}^{m} \int_{\Omega} p_{t,x|s,\alpha} \Sigma_{ij} \frac{\partial^2 h_x}{\partial x_i \partial x_j} dx \tag{7}$$

If we assume the probability density p and its first derivatives $\partial p/\partial x_i$ vanish at a higher order of rate than h and $\partial h/\partial x_i$ as $x_i \to \pm\infty \; \forall \; i = 1, \cdots, m$, we can integrate by parts for the right hand side of (7), once for the first integral and twice for the second

$$\int_{\Omega} A_i p \frac{\partial h_x}{\partial x_i} dx = \int_{\bar{\Omega}_i} \underbrace{A_i h_x p|_{x_i=-\infty}^{+\infty}}_{=0} d\bar{x}_i - \int_{\Omega} h_x \frac{\partial (A_i p)}{\partial x_i} dx$$

$$\int_{\Omega} \Sigma_{ij} \frac{\partial^2 h_x}{\partial x_i \partial x_j} p dx = \int_{\bar{\Omega}_i} \underbrace{\Sigma_{ij} \frac{\partial h_x}{\partial x_j} p \Big|_{x_i=-\infty}^{+\infty}}_{=0} d\bar{x}_i - \int_{\Omega} \frac{\partial (\Sigma_{ij} p)}{\partial x_i} \frac{\partial h_x}{\partial x_j} dx$$
$$= - \int_{\bar{\Omega}_j} \underbrace{h_x \frac{\partial (\Sigma_{ij} p)}{\partial x_i} \Big|_{x_j=-\infty}^{+\infty}}_{=0} d\bar{x}_j + \int_{\Omega} h_x \frac{\partial^2 (\Sigma_{ij} p)}{\partial x_i \partial x_j} dx \tag{8}$$

where $\int_{\bar{\Omega}_i} (\cdot) d\bar{x}_i = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} (\cdot) dx_1 \cdots dx_{i-1} \, dx_{i+1} \cdots dx_m$

Plugging the results of (8) into (7), we have

$$\int_{\Omega} h_x \frac{\partial p}{\partial t} dx = - \sum_{i=1}^{m} \int_{\Omega} h_x \frac{\partial (A_i p)}{\partial x_i} dx + \frac{1}{2} \sum_{i,j=1}^{m} \int_{\Omega} h_x \frac{\partial^2 (\Sigma_{ij} p)}{\partial x_i \partial x_j} dx$$
$$\Rightarrow \int_{\Omega} h_x \left(\frac{\partial p}{\partial t} + \sum_{i=1}^{m} \frac{\partial (A_i p)}{\partial x_i} - \frac{1}{2} \sum_{i,j=1}^{m} \frac{\partial^2 (\Sigma_{ij} p)}{\partial x_i \partial x_j} \right) dx = 0 \tag{9}$$

By the arbitrariness of function h, we conclude that for any $x \in \Omega$ the density function $p_{t,x|s,\alpha}$ satisfies

$$\frac{\partial p}{\partial t} + \sum_{i=1}^{m} \frac{\partial (A_i p)}{\partial x_i} - \frac{1}{2} \sum_{i,j=1}^{m} \frac{\partial^2 (\Sigma_{ij} p)}{\partial x_i \partial x_j} = 0, \qquad \lim_{t \to s} p_{t,x|s,\alpha} = \delta_{x-\alpha}, \qquad \Sigma = B \rho B' \tag{10}$$

with $\delta(\cdot)$ the Dirac delta function. This is the multi-dimensional *Fokker-Planck Equation (a.k.a. Kolmogorov Forward Equation)* [1]. In this equation, the s and α are held constant, while the t and x are variables (called "forward variables"). In the one-dimensional case, it reduces to

$$\frac{\partial p}{\partial t}+\frac{\partial(Ap)}{\partial x}-\frac{1}{2}\frac{\partial^2(B^2p)}{\partial x^2}=0 \tag{11}$$

where $A=A(t,x)$ and $B=B(t,x)$ are then scalar functions.

1.2. Kolmogorov Backward Equation

Let's express conditional expectation $g(t,X_t)=\mathbb{E}_t[h(X_T)]$. Since for any $t\le v\le T$ we have

$$g(t,X_t)=\mathbb{E}_t[h(X_T)]=\mathbb{E}_t\big[\mathbb{E}_v[h(X_T)]\big]=\mathbb{E}_t[g(v,X_v)] \tag{12}$$

the $g(t,X_t)$ is a martingale by the *tower rule* (i.e. If $\mathcal{H}$ holds less information than $\mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]=\mathbb{E}[X|\mathcal{H}]$). The dynamics of the $g(t,X_t)$ is given by

$$dg=\frac{\partial g}{\partial t}dt+\underset{1\times m}{J_g}\underset{m\times 1}{dX_t}+\frac{1}{2}\underset{1\times m}{dX_t'}\underset{m\times m}{H_g}\underset{m\times 1}{dX_t}=\frac{\partial g}{\partial t}dt+J_gAdt+J_gBdW_t+\frac{1}{2}dW_t'B'H_gBdW_t \tag{13}$$

where J_g is the *Jacobian* and H_g the *Hessian* of g with respect to variable X

$$[J_g]_i=\frac{\partial g}{\partial X_i},\qquad [H_g]_{ij}=\frac{\partial^2 g}{\partial X_i\partial X_j} \tag{14}$$

Expanding (13), we have

$$dg=\left(\frac{\partial g}{\partial t}+\sum_{i=1}^{m}A_i\frac{\partial g}{\partial X_i}+\frac{1}{2}\sum_{i,j=1}^{m}\Sigma_{ij}\frac{\partial^2 g}{\partial X_i\partial X_j}\right)dt+\sum_{i=1}^{m}\frac{\partial g}{\partial X_i}\sum_{k=1}^{n}B_{ik}dW_{k,t} \tag{15}$$

Since $g(t,X_t)$ is a martingale, the dt-term must vanish, which gives

$$\frac{\partial g}{\partial t}+\sum_{i=1}^{m}A_i\frac{\partial g}{\partial X_i}+\frac{1}{2}\sum_{i,j=1}^{m}\Sigma_{ij}\frac{\partial^2 g}{\partial X_i\partial X_j}=0 \tag{16}$$

This is the multi-dimensional *Feynman-Kac* formula[1].

[1] https://en.wikipedia.org/wiki/Feynman-Kac_formula

Using the transition probability density $p_{T,\beta|t,x}$ for $X_t = x$ at t and $X_T = \beta$ at T, we can further write the expectation as

$$g_{t,x} = \mathbb{E}_t[h(X_T)] = \int_\Omega h_\beta p_{T,\beta|t,x} d\beta \tag{17}$$

The formula (16) defines that

$$\begin{aligned} &\left(\frac{\partial}{\partial t} + \sum_{i=1}^{m} A_i \frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{m} \Sigma_{ij} \frac{\partial^2}{\partial x_i \partial x_j}\right) \int_\Omega h_\beta p_{T,\beta|t,x} d\beta = 0 \\ &\Rightarrow \int_\Omega h_\beta \left(\frac{\partial p}{\partial t} + \sum_{i=1}^{m} A_i \frac{\partial p}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{m} \Sigma_{ij} \frac{\partial^2 p}{\partial x_i \partial x_j}\right) d\beta = 0 \end{aligned} \tag{18}$$

By the arbitrariness of h function, the transition probability density $p_{T,\beta|t,x}$must satisfy

$$\frac{\partial p}{\partial t} + \sum_{i=1}^{m} A_i \frac{\partial p}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{m} \Sigma_{ij} \frac{\partial^2 p}{\partial x_i \partial x_j} = 0, \qquad \lim_{t\to T} p_{T,\beta|t,x} = \delta_{x-\beta}, \qquad \Sigma = B\rho B' \tag{19}$$

This is the multi-dimensional *Kolmogorov Backward Equation*. In this equation, the T and β are held constant, while the t and x are variables (called "backward variables"). In the 1-D case, it reduces to

$$\frac{\partial p}{\partial t} + A\frac{\partial p}{\partial x} + \frac{1}{2}B^2 \frac{\partial^2 p}{\partial x^2} = 0 \tag{20}$$

where $A = A(t,x)$ and $B = B(t,x)$ are again scalar functions.

2. HESTON MODEL

In this chapter, we will briefly introduce the Heston stochastic volatility model, which has become quite popular in industry to model volatility smiles. One of the reasons for that is that vanilla options in Heston model have close form solutions. This makes the calibration of the model computationally much efficient and accurate. To understand these closed form formulas, we begin with the Fourier transform method for option pricing and its application to the Heston model. Additionally, a characteristic function based method is also discussed, which extends the Heston model to have piecewise constant time-dependent model parameters.

2.1. Heston Stochastic Volatility Model

The stochastic volatility in Heston's model is a mean-reverting square-root process defined by the following stochastic differential equations (SDE)

$$\begin{aligned} \frac{dX_t}{X_t} &= (\mu - q)dt + \sqrt{v_t}dW_{1,t} \\ dv_t &= \epsilon(\vartheta - v_t)dt + \xi\sqrt{v_t}dW_{2,t} \\ dW_{1,t}dW_{2,t} &= \rho dt \end{aligned} \tag{21}$$

where t denotes the time, X_t the spot process, μ and q the drift rate and dividend rate of the spot (or domestic rate and foreign rate in FX), v_t the variance process, ϵ the mean reversion rate, ϑ the mean variance, ξ the volatility of the variance and $dW_{1,t}$, $dW_{2,t}$ the two Brownian motions correlated by ρ under physical measure $\mathbb{P}$.

2.1.1. Market Price of Risk

In Black-Scholes model, a contingent claim is dependent on tradable asset X_t. The randomness in option value, for example, is solely determined by the randomness of the tradable asset X_t. Such uncertainty in value can be hedged by continuously trading the underlying asset. This makes the market complete (i.e. every contingent claim can be replicated). In a stochastic volatility model, a contingent claim is dependent on both the asset X_t and its associated instantaneous volatility v_t. Since the volatility

itself is not tradable, this renders the market incomplete and results in many implications to the pricing of the options.

Firstly let's consider two arbitrary derivative securities (presume they are traded in markets) whose prices can be written as functions $U(t, X_t, v_t)$ and $V(t, X_t, v_t)$ of variables t, X_t and v_t respectively. We construct a self-financing portfolio with price L_t by taking long on one share of $U(t, X_t, v_t)$, short on Γ_t shares of $V(t, X_t, v_t)$ and short on Δ_t shares of X_t, that is

$$L_t = U_t - \Gamma_t V_t - \Delta_t X_t \tag{22}$$

Following the Ito's lemma, we can derive the price dynamics of the derivative as

$$\begin{aligned} dU &= \frac{\partial U}{\partial t}dt + \frac{\partial U}{\partial X}dX + \frac{1}{2}\frac{\partial^2 U}{\partial X^2}dXdX + \frac{\partial U}{\partial v}dv + \frac{1}{2}\frac{\partial^2 U}{\partial v^2}dvdv + \frac{\partial^2 U}{\partial v \partial X}dvdX \\ &= \frac{\partial U}{\partial t}dt + \frac{\partial U}{\partial X}(\mu - q)Xdt + \frac{\partial U}{\partial X}X\sqrt{v}dW_1 + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2}dt + \frac{\partial U}{\partial v}\epsilon(\vartheta - v)dt + \frac{\partial U}{\partial v}\xi\sqrt{v}dW_2 \\ &\quad + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2}dt + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X}dt \end{aligned} \tag{23}$$

and derive the price dynamics of the self-financing portfolio as

$$\begin{aligned} dL &= dU - \Gamma dV - \Delta dX - \Delta X q dt \\ &= \left(\frac{\partial U}{\partial t} + \frac{\partial U}{\partial X}(\mu - q)X + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2} + \frac{\partial U}{\partial v}\epsilon(\vartheta - v) + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2} + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X}\right)dt \\ &\quad -\Gamma\left(\frac{\partial V}{\partial t} + \frac{\partial V}{\partial X}(\mu - q)X + \frac{vX^2}{2}\frac{\partial^2 V}{\partial X^2} + \frac{\partial V}{\partial v}\epsilon(\vartheta - v) + \frac{v\xi^2}{2}\frac{\partial^2 V}{\partial v^2} + Xv\xi\rho\frac{\partial^2 V}{\partial v \partial X}\right)dt \\ &\quad + \left(\frac{\partial U}{\partial X} - \Delta - \Gamma\frac{\partial V}{\partial X}\right)X\sqrt{v}dW_1 + \left(\frac{\partial U}{\partial v} - \Gamma\frac{\partial V}{\partial v}\right)\xi\sqrt{v}dW_2 - \Delta(\mu - q)Xdt - \Delta X q dt \end{aligned} \tag{24}$$

In order to eliminate both spot and volatility risk, we must have

$$\frac{\partial U}{\partial X} - \Delta - \Gamma\frac{\partial V}{\partial X} = 0, \qquad \frac{\partial U}{\partial v} - \Gamma\frac{\partial V}{\partial v} = 0 \tag{25}$$

and therefore

$$dL = \left(\frac{\partial U}{\partial t} + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2} + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X} - \frac{\partial U}{\partial X}Xq\right)dt$$
$$- \Gamma\left(\frac{\partial V}{\partial t} + \frac{vX^2}{2}\frac{\partial^2 V}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 V}{\partial v^2} + Xv\xi\rho\frac{\partial^2 V}{\partial v \partial X} - \frac{\partial V}{\partial X}Xq\right)dt \quad (26)$$

In this case, the portfolio is riskless and must have a return at risk free rate in order to be arbitrage-free

$$dL = rLdt = r(U - \Gamma V - \Delta X)dt = rUdt - r\Gamma Vdt - r\frac{\partial U}{\partial X}Xdt + r\Gamma\frac{\partial V}{\partial X}Xdt \quad (27)$$

which in turn gives

$$\frac{\frac{\partial U}{\partial t} + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2} + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X} + \frac{\partial U}{\partial X}X(r-q) - rU}{\frac{\partial U}{\partial v}}$$
$$= \frac{\frac{\partial V}{\partial t} + \frac{vX^2}{2}\frac{\partial^2 V}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 V}{\partial v^2} + Xv\xi\rho\frac{\partial^2 V}{\partial v \partial X} + \frac{\partial V}{\partial X}X(r-q) - rV}{\frac{\partial V}{\partial v}} \equiv \eta \quad (28)$$

In the above equation, the left-hand side is a function of U only and the right-hand side is a function of V only. The only way for the equality to hold is for both sides to equal a common function η of the independent variables t, X_t and v_t.

Now let's consider a delta-neutral portfolio Y_t by taking long on one share of $U(t, X_t, v_t)$ and short on Δ_t shares of X_t

$$Y_t = U_t - \Delta_t X_t \quad (29)$$

The price dynamics of the portfolio read

$$dY = dU - \Delta dX - \Delta Xqdt$$
$$= \left(\frac{\partial U}{\partial t} + \frac{\partial U}{\partial X}(\mu - q)X + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2} + \frac{\partial U}{\partial v}\epsilon(\vartheta - v) + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2} + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X}\right)dt \quad (30)$$
$$+ \left(\frac{\partial U}{\partial X} - \Delta\right)X\sqrt{v}dW_1 + \frac{\partial U}{\partial v}\xi\sqrt{v}dW_2 - \Delta(\mu - q)Xdt - \Delta Xqdt$$

Since delta-neutral implies $\partial U/\partial X - \Delta = 0$, we are able to derive the dynamics of the discounted portfolio, a martingale under risk neutral measure, as

$$\frac{d(D_t Y_t)}{D_t} = dY - rYdt = dU - \Delta dX - \Delta X q dt - rUdt + r\Delta X dt$$

$$= \left(\frac{\partial U}{\partial v}\epsilon(\vartheta - v) + \frac{\partial U}{\partial t} + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2} + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X} + \frac{\partial U}{\partial X}X(r-q) - rU\right)dt + \frac{\partial U}{\partial v}\xi\sqrt{v}dW_2 \tag{31}$$

$$= \frac{\partial U}{\partial v}(\epsilon(\vartheta - v) + \eta)dt + \frac{\partial U}{\partial v}\xi\sqrt{v}dW_2 = \frac{\partial U}{\partial v}\xi\sqrt{v}d\widetilde{W}_2$$

by defining

$$d\widetilde{W}_2 = dW_2 + \phi_2 dt, \qquad \phi_2 = \frac{\epsilon(\vartheta - v) + \eta}{\xi\sqrt{v}} \tag{32}$$

In the above, the $\widetilde{W}_2$ is a Brownian motion under risk neutral measure $\mathbb{Q}$ and ϕ_2 is the market price of volatility risk.

We next consider a vega-neutral portfolio Z_t by taking long on one share of $U(t, X_t, v_t)$ and short on Γ_t shares of $V(t, X_t, v_t)$

$$Z_t = U_t - \Gamma_t V_t \tag{33}$$

The dynamics of the portfolio reads

$$dZ = dU - \Gamma dV$$

$$= \left(\frac{\partial U}{\partial t} + \frac{\partial U}{\partial X}(\mu - q)X + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2} + \frac{\partial U}{\partial v}\epsilon(\vartheta - v) + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2} + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X}\right)dt$$

$$-\Gamma\left(\frac{\partial V}{\partial t} + \frac{\partial V}{\partial X}(\mu - q)X + \frac{vX^2}{2}\frac{\partial^2 V}{\partial X^2} + \frac{\partial V}{\partial v}\epsilon(\vartheta - v) + \frac{v\xi^2}{2}\frac{\partial^2 V}{\partial v^2} + Xv\xi\rho\frac{\partial^2 V}{\partial v \partial X}\right)dt \tag{34}$$

$$+\left(\frac{\partial U}{\partial X} - \Gamma\frac{\partial V}{\partial X}\right)X\sqrt{v}dW_1 + \left(\frac{\partial U}{\partial v} - \Gamma\frac{\partial V}{\partial v}\right)\xi\sqrt{v}dW_2$$

Since vega-neutral implies $\frac{\partial U}{\partial v} - \Gamma\frac{\partial V}{\partial v} = 0$, we can derived the dynamics of the discounted portfolio as

$$\frac{d(D_t Z_t)}{D_t} = dZ - rZdt = dU - \Gamma dV - rUdt + r\Gamma V dt \tag{35}$$

$$= \left(\frac{\partial U}{\partial t} + (\mu - q)X\frac{\partial U}{\partial X} + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2} + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X} - rU\right)dt$$

$$- \Gamma\left(\frac{\partial V}{\partial t} + (\mu - q)X\frac{\partial V}{\partial X} + \frac{vX^2}{2}\frac{\partial^2 V}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 V}{\partial v^2} + Xv\xi\rho\frac{\partial^2 V}{\partial v \partial X} - rV\right)dt$$

$$+ \left(\frac{\partial U}{\partial X} - \Gamma\frac{\partial V}{\partial X}\right)X\sqrt{v}dW_1$$

$$= \left(\frac{\partial U}{\partial t} + \frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2} + Xv\xi\rho\frac{\partial^2 U}{\partial v \partial X} - rU\right)dt$$

$$- \Gamma\left(\frac{\partial V}{\partial t} + \frac{vX^2}{2}\frac{\partial^2 V}{\partial X^2} + \frac{v\xi^2}{2}\frac{\partial^2 V}{\partial v^2} + Xv\xi\rho\frac{\partial^2 V}{\partial v \partial X} - rV\right)dt + \left(\frac{\partial U}{\partial X} - \Gamma\frac{\partial V}{\partial X}\right)dX$$

$$= \left(\eta\frac{\partial U}{\partial v} - X(r-q)\frac{\partial U}{\partial X} - \Gamma\eta\frac{\partial V}{\partial v} + \Gamma X(r-q)\frac{\partial V}{\partial X}\right)dt + \left(\frac{\partial U}{\partial X} - \Gamma\frac{\partial V}{\partial X}\right)dX$$

$$= \left(\frac{\partial U}{\partial X} - \Gamma\frac{\partial V}{\partial X}\right)(dX - X(r-q)dt) = \left(\frac{\partial U}{\partial X} - \Gamma\frac{\partial V}{\partial X}\right)X\left((\mu - r)dt + \sqrt{v}dW_1\right)$$

$$= \left(\frac{\partial U}{\partial X} - \Gamma\frac{\partial V}{\partial X}\right)X\sqrt{v}d\widetilde{W}_1$$

by defining

$$d\widetilde{W}_1 = dW_1 + \phi_1 dt, \qquad \phi_1 = \frac{\mu - r}{\sqrt{v}} \tag{36}$$

where $d\widetilde{W}_1$ is a Brownian motion under risk neutral measure $\mathbb{Q}$ and ϕ_1 is the market price of spot risk.

According to (32) and (36), the Heston SDE (21) under risk neutral measure takes the form

$$\begin{aligned} \frac{dX_t}{X_t} &= (\mu - q)dt + \sqrt{v_t}\left(d\widetilde{W}_{1,t} - \phi_1 dt\right) = (r - q)dt + \sqrt{v_t}d\widetilde{W}_{1,t} \\ dv_t &= \epsilon(\vartheta - v_t)dt + \xi\sqrt{v_t}\left(d\widetilde{W}_{2,t} - \phi_2 dt\right) = \left(\epsilon(\vartheta - v_t) - \phi_2\xi\sqrt{v_t}\right)dt + \xi\sqrt{v_t}d\widetilde{W}_{2,t} \end{aligned} \tag{37}$$

Based on the *Consumption-based Capital Asset Pricing Model*, Heston [24] assumes that the market price of volatility risk is proportional to volatility, that is

$$\phi_2 = c\sqrt{v} \quad \text{for a constant } c \quad \Longrightarrow \quad \phi_2\xi\sqrt{v} = c\xi v = \lambda v \quad \text{where} \quad \lambda = c\xi \tag{38}$$

If we define

$$\kappa = \epsilon + \lambda, \qquad \theta = \frac{\epsilon\vartheta}{\kappa} \tag{39}$$

the Heston SDE under risk neutral measure $\mathbb{Q}$ reads

$$\frac{dX_t}{X_t} = (r-q)dt + \sqrt{v_t}d\widetilde{W}_{1,t}, \qquad dv_t = \kappa(\theta - v_t)dt + \xi\sqrt{v_t}d\widetilde{W}_{2,t}, \qquad d\widetilde{W}_{1,t}d\widetilde{W}_{2,t} = \rho dt \tag{40}$$

which retains the form of the equation under the transformation from the physical measure $\mathbb{P}$ to the risk neutral measure $\mathbb{Q}$.

Since the volatility is not a traded asset, the incompleteness of the market implies the risk neutral measure is not unique and depends on the value of the market price of volatility risk ϕ_2. To estimate the model parameters, one may calibrate the Heston's model using historical spot data, however the historical calibration does not allow for the estimation of ϕ_2. Instead of using the spot data, one may also calibrate the model to the volatility smile (i.e. prices of vanilla options). In this case, the market price of volatility risk has already been implied in the market smile, and consequently embedded into the calibrated model parameters κ and θ through (39).

2.1.2. Radon-Nikodym Derivative

The change of measure from $\mathbb{P}$ to $\mathbb{Q}$ is achieved through *Radon-Nikodym derivative* via multi-dimensional Girsanov's theorem [2]. To derive this derivative, we may write correlated n-dimensional Brownian motions as dW_t and $d\widetilde{W}_t$ under physical measure $\mathbb{P}$ and risk neutral measure $\mathbb{Q}$ respectively. The matrix ρ denotes the instantaneous correlation, e.g. $dW_t dW_t' = \rho dt$. It should be noted that dW_t and $d\widetilde{W}_t$ possess the same correlation structure only if each is under its own probability measure, $\mathbb{P}$ or $\mathbb{Q}$, otherwise this property does not hold. From (32) and (36), we represent the market price of risk by correlation matrix an n-dimensional vector ϕ such that

$$d\widetilde{W}_t = dW_t + \rho\phi_t dt \tag{41}$$

The *Radon-Nikodym derivative* is then given for $t > s$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(-\frac{1}{2}\int_s^t \phi_u'\rho\phi_u du - \int_s^t \phi_u' dW_u\right) \tag{42}$$

To check this, let's assume under $\mathbb{Q}$ we have a martingale process

$$L_t = L_s \exp\left(-\frac{1}{2}\int_s^t \sigma_u' \rho \sigma_u du + \int_s^t \sigma_u' d\widetilde{W}_u\right) \tag{43}$$

where $\sigma_u$is a vector representing an adapted volatility process. According to (41) we have

$$L_t = L_s \exp\left(-\frac{1}{2}\int_s^t \sigma_u' \rho \sigma_u du + \int_s^t \sigma_u' dW_u + \int_s^t \sigma_u' \rho \phi_u du\right) \quad \text{and}$$

$$L_t \frac{d\mathbb{Q}}{d\mathbb{P}} = L_s \exp\left(-\frac{1}{2}\int_s^t \sigma_u' \rho \sigma_u du + \int_s^t \sigma_u' \rho \phi_u du - \frac{1}{2}\int_s^t \phi_u' \rho \phi_u du + \int_s^t (\sigma_u - \phi_u)' dW_u\right) \tag{44}$$

$$= L_s \exp\left(-\frac{1}{2}\int_s^t (\sigma_u - \phi_u)' \rho (\sigma_u - \phi_u) du + \int_s^t (\sigma_u - \phi_u)' dW_u\right)$$

The $L_t \frac{d\mathbb{Q}}{d\mathbb{P}}$ is a martingale under $\mathbb{P}$. Hence we have the following equation

$$\widetilde{\mathbb{E}}_s[L_t] = \mathbb{E}_s\left[L_t \frac{d\mathbb{Q}}{d\mathbb{P}}\right] = L_s \tag{45}$$

as expected.

2.1.3. Feller Condition

The variance process v_t in (40) is known as Cox-Ingersoll-Ross (CIR) process [3]. The distribution of future values of a CIR process can be computed in closed form

$$v_{t+\tau} = \frac{Y}{2c}, \qquad c = \frac{2\kappa}{(1-e^{-\kappa\tau})\xi^2}, \qquad Y \sim \chi^2\left(\frac{4\kappa\theta}{\xi^2}, 2cv_t e^{-\kappa\tau}\right) \tag{46}$$

Where Y is a non-central Chi-Squared distribution with $4\kappa\theta/\xi^2$ degrees of freedom and non-centrality parameter $2cv_t e^{-\kappa\tau}$. Integration of v_t process gives its conditional mean and variance as

$$\mathbb{E}[v_{t+\tau}|v_t] = \theta + (v_t - \theta)e^{-\kappa\tau}, \qquad \mathbb{V}[v_{t+\tau}|v_t] = \frac{v_t\xi^2}{\kappa}(e^{-\kappa\tau} - e^{-2\kappa\tau}) + \frac{\theta\xi^2}{2\kappa}(1-e^{-\kappa\tau})^2 \tag{47}$$

It can be seen that the long term mean for $v_{t+\tau}$ is θ, i.e. the mean reversion level.

Feller observed that the variance process v_t remains strictly positive with probability 1 for all times $t > s$, if $v_s > 0$ and the Feller condition [4] [5] is satisfied

$$2\kappa\theta \geq \xi^2 \tag{48}$$

If the condition is not satisfied, i.e. $0 < 2\kappa\theta < \xi^2$, the variance will visit zero recurrently but will not stay at zero, i.e. the zero boundary is strongly reflecting. In typical applications, the Feller condition is often violated due to the convexities of volatility smiles typically encountered in practice. Indeed the process v_t often has a strong affinity for the area around the origin. However, this is not a complete disaster, as the process v_t can only hit zero for an infinitesimally small amount of time.

2.2. Probability Distribution of Spot Returns

In this section, we present a derivation of the distribution of the spot returns in Heston's model [6]. Let's firstly make a change of variable to have a centered log-spot for $t > s$

$$x_t = \ln\frac{X_t}{F_{s,t}}, \qquad F_{s,t} = X_s\frac{\hat{B}_{s,t}}{B_{s,t}}, \qquad B_{s,t} = \exp\left(-\int_s^t r_u du\right), \qquad \hat{B}_{s,t} = \exp\left(-\int_s^t q_u du\right) \tag{49}$$

The Heston's model (40) under $\mathbb{Q}$ can then take the following form by Ito's lemma

$$dx_t = -\frac{v_t}{2}dt + \sqrt{v_t}d\widetilde{W}_{1,t}, \qquad dv_t = \kappa(\theta - v_t)dt + \xi\sqrt{v_t}d\widetilde{W}_{2,t}, \qquad d\widetilde{W}_{1,t}d\widetilde{W}_{2,t} = \rho dt \tag{50}$$

This defines a 2-D stochastic process characterized by a joint transition probability density function $p(t, x_t, v_t|s, x_s, v_s)$, which is the probability having log-spot x_t and instantaneous variance v_t at time t, conditional on x_s and v_s at outset s.

2.2.1. Derivation of the Transition Probability

We may rewrite (50) in terms of a 2-D Brownian motion $d\widetilde{W}_t$

$$dZ_t = A_t dt + C_t d\widetilde{W}_t \qquad \text{with}$$

$$Z_t = \begin{pmatrix} x_t \\ v_t \end{pmatrix}, \qquad A_t = \begin{pmatrix} -v_t/2 \\ \kappa(\theta - v_t) \end{pmatrix}, \qquad C_t = \begin{pmatrix} \sqrt{v_t} & 0 \\ 0 & \xi\sqrt{v_t} \end{pmatrix}, \qquad d\widetilde{W}_t d\widetilde{W}_t' = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} dt \tag{51}$$

The instantaneous covariance matrix of dZ_t becomes

$$\Sigma_t = C_t \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} C_t = \begin{pmatrix} v_t & \rho\xi v_t \\ \rho\xi v_t & \xi^2 v_t \end{pmatrix} \tag{52}$$

The 2-D Markov process is characterized by the transition probability $p_{t,x,v} = p(t, x_t, v_t|s, x_s, v_s)$. The *Fokker-Planck* equation that governs the time evolution of the transition probability is given by (10)

$$\frac{\partial p}{\partial t} - \frac{1}{2}\frac{\partial (vp)}{\partial x} + \frac{\partial (\kappa(\theta - v)p)}{\partial v} - \frac{1}{2}\frac{\partial^2 (vp)}{\partial x^2} - \rho\xi\frac{\partial^2 (vp)}{\partial x \partial v} - \frac{\xi^2}{2}\frac{\partial^2 (vp)}{\partial v^2} = 0$$

$$\frac{\partial p}{\partial t} - \frac{v}{2}\frac{\partial p}{\partial x} + \kappa(\theta - v)\frac{\partial p}{\partial v} - \frac{v}{2}\frac{\partial^2 p}{\partial x^2} - \rho\xi\frac{\partial^2 (vp)}{\partial x \partial v} - \frac{\xi^2}{2}\frac{\partial^2 (vp)}{\partial v^2} = 0 \tag{53}$$

$$\frac{\partial^2 (vp)}{\partial x \partial v} = \frac{\partial}{\partial x}\frac{\partial (vp)}{\partial v} = \frac{\partial}{\partial x}\left(p + v\frac{\partial p}{\partial v}\right) = \frac{\partial p}{\partial x} + v\frac{\partial^2 p}{\partial x \partial v}$$

with initial condition $p_{s,x,v} = \delta_{x-x_s}\delta_{v-v_s} = \delta_x \delta_{v-v_s}$, where δ is the *Dirac delta function*. The marginal probability density of the variance alone

$$\zeta_{t,v} = \int_{\mathbb{R}} p_{t,x,v} dx \tag{54}$$

satisfies the following *Fokker-Planck equation* obtained from (53) by integration over x

$$\frac{\partial \zeta}{\partial t} = \frac{\partial (\kappa(v - \theta)\zeta)}{\partial v} + \frac{\xi^2}{2}\frac{\partial^2 (v\zeta)}{\partial v^2} \tag{55}$$

Feller has shown that this equation is well defined on the interval $v \in [0, +\infty)$ as long as $\theta > 0$. Equation (55) has a stationary solution, which is a Gamma distribution

$$\zeta_v^* = \frac{\alpha^\alpha}{\Gamma(\alpha)}\frac{v^{\alpha-1}}{\theta^\alpha}\exp\left(-\frac{\alpha v}{\theta}\right) \quad \text{and} \quad \alpha = \frac{2\kappa\theta}{\xi^2} \tag{56}$$

Since x appears in (53) only in the derivative operator, it is convenient to take the *Fourier* transform, such that

$$\hat{p}_{t,\omega,v} = \int_{\mathbb{R}} e^{-i\omega x} p_{t,x,v} dx \quad \text{and} \quad p_{t,x,v} = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\omega x}\hat{p}_{t,\omega,v} d\omega \tag{57}$$

Inserting (57) into (53), we have

$$\frac{\partial \hat{p}}{\partial t} = \frac{\partial (\kappa(v - \theta)\hat{p})}{\partial v} + \frac{i\omega - \omega^2}{2} v\hat{p} + i\rho\xi\omega\frac{\partial (v\hat{p})}{\partial v} + \frac{\xi^2}{2}\frac{\partial^2 (v\hat{p})}{\partial v^2} \tag{58}$$

Since (58) is linear in v and quadratic in $\partial/\partial v$, by taking *Laplace* transform over v, it simplifies to

$$\tilde{p}_{t,\omega,\lambda} = \int_{\mathbb{R}^+} e^{-\lambda v} \hat{p}_{t,\omega,v} dv \tag{59}$$

The PDE satisfied by $\tilde{p}_{t,\omega,\lambda}$ is of the first order

$$\frac{\partial \tilde{p}}{\partial t} = \left(\frac{\omega^2 - \xi^2 \lambda^2 - i\omega}{2} - \gamma\lambda \right) \frac{\partial \tilde{p}}{\partial \lambda} - \kappa\theta\lambda\tilde{p} \tag{60}$$

with initial condition $\tilde{p}_{s,\omega,\lambda} = \exp(-\lambda v_s)$, where $\gamma = \kappa + i\rho\xi\omega$. The solution of this PDE is given by the method of characteristics

$$\tilde{p}_{t,\omega,\lambda} = \exp\left(-\tilde{\lambda}_s v_s - \kappa\theta \int_s^t \tilde{\lambda}_u du \right) \tag{61}$$

where the function $\tilde{\lambda}_t$ is the solution of the characteristic (ordinary) differential equation

$$\frac{d\tilde{\lambda}_u}{du} = \gamma\tilde{\lambda}_u + \frac{\xi^2}{2}\tilde{\lambda}_u^2 - \frac{\omega^2 - i\omega}{2} \tag{62}$$

With a boundary condition $\tilde{\lambda}_t = \lambda$ specified at time t, the (62) is a Riccati equation with constant coefficients and its solution is

$$\tilde{\lambda}_u = \frac{2\Omega}{\xi^2} \frac{1}{\Psi e^{\Omega(t-u)} - 1} - \frac{\gamma - \Omega}{\xi^2} \quad \text{with} \quad \Omega = \sqrt{\gamma^2 + \xi^2(\omega^2 - i\omega)}, \qquad \Psi = 1 + \frac{2\Omega}{\xi^2\lambda + \gamma - \Omega} \tag{63}$$

Plugging (63) into (61), we have

$$\tilde{p}_{t,\omega,\lambda} = \exp\left(-\tilde{\lambda}_s v_s + \frac{\kappa\theta(\gamma - \Omega)t}{\xi^2} - \frac{2\kappa\theta}{\xi^2} \ln \frac{\Psi - e^{-\Omega t}}{\Psi - 1} \right) \tag{64}$$

Normally we are interested only in distribution of log-spot x_t and do not care about variance v_t. Therefore we derive the marginal probability density for x_t with $\lambda = 0$

$$\begin{aligned} p_{t,x} &= \int_{\mathbb{R}^+} p_{t,x,v} dv \\ &= \frac{1}{2\pi} \int_{\mathbb{R}^+} \int_{\mathbb{R}} e^{i\omega x} \hat{p}_{t,\omega,v} d\omega \, dv = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\omega x} \int_{\mathbb{R}^+} \hat{p}_{t,\omega,v} dv \, d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\omega x} \tilde{p}_{t,\omega,\lambda=0} d\omega \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \exp\left(i\omega x - \frac{\omega^2 - i\omega}{\gamma + \Omega \coth \frac{\Omega t}{2}} v_s + \frac{\kappa\theta\gamma t}{\xi^2} - \frac{2\kappa\theta}{\xi^2} \ln\left(\cosh\frac{\Omega t}{2} + \frac{\gamma}{\Omega} \sinh\frac{\Omega t}{2} \right) \right) d\omega \end{aligned} \tag{65}$$

where the last step comes from substitution of $\hat{p}_{t,\omega,\lambda=0}$ in (61) into (57). The derived density function $p_{t,x} = p(t,x|s,x_s,v_s)$ in (65) is still dependent on the unknown initial variance v_s. To remove the dependence, the v_s is assumed to have the stationary distribution density as in (56). Thus the unconditional transition density function $f_{t,x}$, is derived by averaging (65) over v_s with the weight ζ^*

$$f_{t,x} = \int_{\mathbb{R}^+} p(t,x|s,x_s,v)\zeta_v^* dv \tag{66}$$

The integral over v_s is similar to the one of the Gamma function and can be taken explicitly. The final result is the *Fourier* integral

$$f_{t,x} = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\omega x + F_{t,\omega}} d\omega\,, \qquad F_{t,\omega} = \frac{\kappa\theta\gamma t}{\xi^2} - \frac{2\kappa\theta}{\xi^2}\ln\left(\cosh\frac{\Omega t}{2} + \frac{\Omega^2 - \gamma^2 + 2\kappa\gamma}{2\kappa\Omega}\sinh\frac{\Omega t}{2}\right) \tag{67}$$

It is easy to check that $f_{t,x}$ is real, because $\Re\left[F_{t,\omega}\right]$ is an even function of ω and $\Im\left[F_{t,\omega}\right]$ is an odd function. One can also check that $F_{t,\omega=0} = 0$, which implies that $f_{t,x}$ is correctly normalized at all times.

2.2.2. Moment Generating Function (in progress…)

We can integrate (50) over time from s to t to get

$$x_t = x_s - \frac{1}{2}\int_s^t v_u du + \int_s^t \sqrt{v_u} d\widetilde{W}_{1,u}\,, \qquad v_t = v_s + \kappa\theta\tau - \kappa\int_s^t v_u du + \xi\int_s^t \sqrt{v_u} d\widetilde{W}_{2,u} \tag{68}$$

where $\tau = t - s$ and $d\widetilde{W}_{1,t} = \rho d\widetilde{W}_{2,t} + \eta d\tilde{B}_t$ for $\eta = \sqrt{1-\rho^2}$ and the Brownian motion $d\tilde{B}_t$ is independent of $d\widetilde{W}_{2,t}$. Defining a function $h_{t;\alpha,\beta}$ such that

$$\begin{aligned}
h_{t;\alpha,\beta} &:= \widetilde{\mathbb{E}}_s[\exp(\alpha x_t + \beta v_t)] = \widetilde{\mathbb{E}}_s\left[\exp\left(\alpha x_t + \left(\beta + \frac{\alpha\rho}{\xi}\right)v_t - \frac{\alpha\rho}{\xi}v_t\right)\right] \\
&= \widetilde{\mathbb{E}}_s\left[\exp\left(\alpha x_s - \alpha\rho\frac{v_s + \kappa\theta\tau}{\xi} + \left(\beta + \frac{\alpha\rho}{\xi}\right)v_t + \left(\frac{\alpha\rho\kappa}{\xi} - \frac{\alpha}{2}\right)\int_s^t v_u du + \alpha\eta\int_s^t \sqrt{v_u} d\tilde{B}_u\right)\right] \\
&= \exp\left(\alpha x_s - \alpha\rho\frac{v_s + \kappa\theta\tau}{\xi}\right)\widetilde{\mathbb{E}}_s\left[\exp\left(\left(\beta + \frac{\alpha\rho}{\xi}\right)v_t + \left(\frac{\alpha\rho\kappa}{\xi} - \frac{\alpha}{2}\right)\int_s^t v_u du + \alpha\eta\int_s^t \sqrt{v_u} d\tilde{B}_u\right)\right] \\
&= e^{\alpha x_s - \alpha\rho\frac{v_s+\kappa\theta\tau}{\xi}}\widetilde{\mathbb{E}}_s\left[e^{\left(\beta+\frac{\alpha\rho}{\xi}\right)v_t + \left(\frac{\alpha\rho\kappa}{\xi}-\frac{\alpha}{2}\right)\int_s^t v_u du}\widetilde{\mathbb{E}}_s\left[e^{\alpha\eta\int_s^t \sqrt{v_u} d\tilde{B}_u}\right]\right] \quad \left(v_t \text{ is independent of } \tilde{B}_t\right)
\end{aligned} \tag{69}$$

$$= \exp\left(\alpha x_s - \alpha\rho\frac{v_s + \kappa\theta\tau}{\xi}\right)\widetilde{\mathbb{E}}_s\left[\exp\left(\left(\beta + \frac{\alpha\rho}{\xi}\right)v_t + \left(\frac{\alpha\rho\kappa}{\xi} - \frac{\alpha}{2} + \frac{\alpha^2\eta^2}{2}\right)\int_s^t v_u du\right)\right]$$

we can derive the differential of $\exp(\alpha x_t + \beta v_t)$ as

$$\begin{aligned} d\left(e^{\alpha x_t + \beta v_t}\right) &= e^{\alpha x_t + \beta v_t}\left(d(\alpha x_t + \beta v_t) + \frac{1}{2}d(\alpha x_t + \beta v_t)d(\alpha x_t + \beta v_t)\right) \\ &= e^{\alpha x_t + \beta v_t}\left(\left(-\frac{\alpha}{2} + \frac{\alpha^2}{2} + \frac{\beta^2\xi^2}{2} + \alpha\beta\xi\rho - \beta\kappa\right)v_t dt + \beta\kappa\theta dt + \alpha\sqrt{v_t}d\widetilde{W}_{1,t} + \beta\xi\sqrt{v_t}d\widetilde{W}_{2,t}\right) \end{aligned} \tag{70}$$

Integrating both sides gives

$$\begin{aligned} e^{\alpha x_t + \beta v_t} = e^{\alpha x_s + \beta v_s} &+ \left(-\frac{\alpha}{2} + \frac{\alpha^2}{2} + \frac{\beta^2\xi^2}{2} + \alpha\beta\xi\rho - \beta\kappa\right)\int_s^t e^{\alpha x_u + \beta v_u}v_u du \\ &+ \beta\kappa\theta\int_s^t e^{\alpha x_u + \beta v_u}du + \alpha\int_s^t e^{\alpha x_u + \beta v_u}\sqrt{v_u}d\widetilde{W}_{1,u} + \beta\xi\int_s^t e^{\alpha x_u + \beta v_u}\sqrt{v_u}d\widetilde{W}_{2,u} \end{aligned} \tag{71}$$

Taking expectation, we find

$$\begin{aligned} h_{t;\alpha,\beta} &= e^{\alpha x_s + \beta v_s} + \left(-\frac{\alpha}{2} + \frac{\alpha^2}{2} + \frac{\beta^2\xi^2}{2} + \alpha\beta\xi\rho - \beta\kappa\right)\int_s^t \widetilde{\mathbb{E}}_s\left[v_u e^{\alpha x_u + \beta v_u}\right]du \\ &\quad + \beta\kappa\theta\int_s^t h_{u;\alpha,\beta}du \\ \Longrightarrow h_{t;\alpha,\beta} &= h_{s;\alpha,\beta} + \underbrace{\left(-\frac{\alpha}{2} + \frac{\alpha^2}{2} + \frac{\beta^2\xi^2}{2} + \alpha\beta\xi\rho - \beta\kappa\right)}_{p_{\alpha,\beta}}\int_s^t \frac{\partial h_{u;\alpha,\beta}}{\partial\beta}du + \beta\kappa\theta\int_s^t h_{u;\alpha,\beta}du \end{aligned} \tag{72}$$

Differentiating with respect to t, we have the PDE

$$\frac{\partial h_{t;\alpha,\beta}}{\partial t} - p_{\alpha,\beta}\frac{\partial h_{t;\alpha,\beta}}{\partial\beta} = \beta\kappa\theta h_{t;\alpha,\beta} \tag{73}$$

(To be continued, reference [7])

2.3. Analytical Solution of Vanilla Options

Vanilla option price in Heston model can be computed semi-analytically. Shortly we will see that the spot process in Heston model admits a closed-form characteristic function, which allows us to

express the option prices in terms of *Fourier*-inversion integrals that can be evaluated numerically (e.g. using quadrature).

2.3.1. Fourier Transform

In textbooks, *Fourier* transform and its inverse (in n-dimensional case) are often depicted as

$$\begin{aligned} &\text{Forward Transform:} \quad \hat{f}_\xi = \int_{\mathbb{R}^n} e^{-2\pi i \xi' x} f_x dx \quad \forall\ \omega \in \mathbb{R}^n \\ &\text{Inverse Transform:} \quad f_x = \int_{\mathbb{R}^n} e^{2\pi i \xi' x} \hat{f}_\xi d\xi \quad \forall\ x \in \mathbb{R}^n \end{aligned} \tag{74}$$

where $f: \mathbb{R}^n \longmapsto \mathbb{C}$ and $\hat{f}: \mathbb{R}^n \longmapsto \mathbb{C}$ are integrable functions, and ξ is the ordinary frequency measured in hertz. This convention defines a unitary transformation on $L^2(\mathbb{R}^n)$ (i.e. the inner product is preserved before and after the transformation). The transform is equivalent to writing a periodic function f_x in Fourier series expansion within an interval of $T = 1/h$ for $h > 0$. In 1D, this reads

$$f_x = \sum_{k\in\mathbb{Z}} \frac{c_k}{T} e^{2\pi i k \frac{x}{T}} = h \sum_{k\in\mathbb{Z}} c_k e^{2\pi i h k x} \quad \text{and} \quad c_k = \int_{-T/2}^{T/2} f_x e^{-2\pi i k \frac{x}{T}} dx = \int_{-1/2h}^{1/2h} f_x e^{-2\pi i h k x} dx \tag{75}$$

Letting $\xi = hk$ and taking the limit $h \to 0^+$, we have

$$\begin{aligned} &\lim_{h\to 0^+} c_k = \lim_{h\to 0^+} \int_{-1/2h}^{1/2h} f_x e^{-2\pi i h k x} dx = \int_{\mathbb{R}} f_x e^{-2\pi i \xi x} dx = \hat{f}_\xi \qquad \text{and} \\ &f_x = \lim_{h\to 0^+} h \sum_{k\in\mathbb{Z}} c_k e^{2\pi i h k x} = \int_{\mathbb{R}} \hat{f}_\xi e^{2\pi i \xi x} d\xi \end{aligned} \tag{76}$$

The sum over all the integers in the above equation can be regarded as an approximating Riemann sum for the integral [8].

There is no standard definition of a Fourier transform and its inverse. All the definitions are mutually equivalent. The one that we use for characteristic functions follows an angular frequency convention and appears less symmetric, which is defined as follows

$$\text{Forward:} \quad \hat{f}_\omega = \int_{\mathbb{R}^n} e^{i\omega' x} f_x dx \quad \forall\ \omega \in \mathbb{R}^n \tag{77}$$

$$\text{Inverse:}\quad f_x = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{-i\omega' x}\hat{f}_\omega d\omega \quad \forall\ x \in \mathbb{R}^n$$

where the ω is the angular frequency measured in radians per second (e.g. $\omega = 2\pi\xi$) and the $(')$ denotes matrix transpose (and hence the $\omega' x$ denotes a dot product of two column vectors ω and x). The extra term $1/(2\pi)^n$ is introduced due to the change of variable $\omega = 2\pi\xi$ from ordinary frequency to angular frequency. In fact the normalization factors multiplying the forward and inverse transforms (here 1 and $1/(2\pi)^n$, respectively) and the signs of the exponents are merely conventions and differ in treatments. The only requirements for these conventions are: 1) the forward and inverse transforms have opposite-sign exponents, and 2) the product of their normalization factors is $1/(2\pi)^n$. In 1D, the (77) reduces to

$$\begin{aligned}&\text{Forward:}\quad \hat{f}_\omega = \int_{\mathbb{R}} e^{i\omega x} f_x dx \quad \forall\ \omega \in \mathbb{R}\\ &\text{Inverse:}\quad f_x = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\omega x}\hat{f}_\omega d\omega \quad \forall\ x \in \mathbb{R}\end{aligned} \tag{78}$$

2.3.2. Levy's Inversion Formula

Characteristic function ϕ_ω of any random variable X completely defines its probability distribution. On the real line, the characteristic function is given by the formula

$$\phi_\omega \equiv \mathbb{E}\left[e^{i\omega X}\right] = \int_{\mathbb{R}} e^{i\omega x} p_x dx = \int_{\Omega} e^{i\omega x} dP_x \quad \forall\ \omega \in \mathbb{R} \tag{79}$$

where the p_x denotes the probability density function (PDF) and the $P_x = \int_{-\infty}^{x} p_y dy$ is the cumulative density function (CDF). The characteristic function is merely a Fourier transform of the PDF p_x, likewise the p_x can be recovered from ϕ_ω through the inverse Fourier transform [9]

$$p_x = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\omega x}\phi_\omega d\omega \qquad \forall\ \omega \in \mathbb{R} \tag{80}$$

Furthermore, we may compute the CDF P_x from ϕ_ω through *Levy's Inversion Formula* [10] [11] [12] shown below

$$P_x = \frac{\phi_0}{2} + \frac{1}{2\pi}\int_{\mathbb{R}^+} \frac{e^{i\omega x}\phi_{-\omega} - e^{-i\omega x}\phi_\omega}{i\omega} d\omega \tag{81}$$

where $\phi_0 = 1$ if ϕ_ω is a characteristic function of a random variable. Before proving the formula, we need to find Fourier transform of signum function $\mathcal{S}_x$, which is defined as

$$\mathcal{S}_x = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{82}$$

Its transform cannot be obtained via direct integration. However we can consider an odd two-sided exponential function $\mathcal{S}_x^h$ with $h > 0$

$$\mathcal{S}_x^h = \begin{cases} -e^{hx} & \text{if } x < 0 \\ e^{-hx} & \text{if } x > 0 \end{cases} \tag{83}$$

The $\hat{\mathcal{S}}_\omega^h$, i.e. the Fourier transform of $\mathcal{S}_x^h$, can then be derived as

$$\begin{aligned} \hat{\mathcal{S}}_\omega^h &= \int_{\mathbb{R}} e^{i\omega x} \mathcal{S}_x^h dx = -\int_{\mathbb{R}^-} e^{(i\omega+h)x} dx + \int_{\mathbb{R}^+} e^{(i\omega-h)x} dx \\ &= -\frac{e^{(i\omega+h)x}}{i\omega+h}\bigg|_{x=-\infty}^{0} + \frac{e^{(i\omega-h)x}}{i\omega-h}\bigg|_{x=0}^{\infty} = -\frac{1}{i\omega+h} - \frac{1}{i\omega-h} = \frac{2i\omega}{\omega^2+h^2} \end{aligned} \tag{84}$$

The parameter h controls how rapidly the exponential function decays. As $h \to 0$, the exponential function resembles more and more closely the signum function. This suggests that

$$\hat{\mathcal{S}}_\omega = \lim_{h\to 0} \hat{\mathcal{S}}_\omega^h = \lim_{h\to 0} \frac{2i\omega}{\omega^2+h^2} = -\frac{2}{i\omega} \tag{85}$$

Hence the inverse transform of the $\hat{\mathcal{S}}_\omega$ gives

$$\begin{aligned} \mathcal{S}_x &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\omega x} \hat{\mathcal{S}}_\omega d\omega = -\frac{1}{\pi}\int_{\mathbb{R}} \frac{e^{-i\omega x}}{i\omega} d\omega = -\frac{1}{\pi}\int_{\mathbb{R}} \frac{\cos\omega x - i\sin\omega x}{i\omega} d\omega \\ &= -\underbrace{\frac{1}{i\pi}\int_{\mathbb{R}} \frac{\cos\omega x}{\omega} d\omega}_{=0,\ (\text{odd function})} + \frac{1}{\pi}\int_{\mathbb{R}} \frac{\sin\omega x}{\omega} d\omega = \frac{2}{\pi}\int_{\mathbb{R}^+} \frac{\sin\omega x}{\omega} d\omega \end{aligned} \tag{86}$$

With the help of the signum function $\mathcal{S}_x$ in (86), the proof of (81) is given as follows

$$\begin{aligned} &\int_{\mathbb{R}^+} \frac{e^{i\omega x}\phi_{-\omega} - e^{-i\omega x}\phi_\omega}{i\omega} d\omega = \int_{\mathbb{R}^+} \frac{e^{i\omega x}\int_{\mathbb{R}} e^{-i\omega y} p_y dy - e^{-i\omega x}\int_{\mathbb{R}} e^{i\omega y} p_y dy}{i\omega} d\omega \\ &= \int_{\mathbb{R}^+}\int_{\mathbb{R}} \frac{e^{i\omega(x-y)} - e^{-i\omega(x-y)}}{i\omega} p_y dy\, d\omega \end{aligned} \tag{87}$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}^+} \frac{e^{i\omega(x-y)} - e^{-i\omega(x-y)}}{i\omega} d\omega\, p_y dy \qquad \text{(by Fubini's theorem)}$$

$$= 2\int_{\mathbb{R}}\int_{\mathbb{R}^+} \frac{\sin\omega(x-y)}{\omega} d\omega\, p_y dy \qquad \left(\text{by } e^{ix} = \cos x + i\sin x\right)$$

$$= \pi\int_{\mathbb{R}} \mathcal{S}_{x-y} p_y dy = \pi\left(-\int_x^{\infty} p_y dy + \int_{-\infty}^{x} p_y dy\right) = \pi\left(-\int_{\mathbb{R}} p_y dy + 2\int_{-\infty}^{x} p_y dy\right)$$

$$= \pi(-\phi_0 + 2P_x)$$

The *Levy Inversion Formula* can also be written in the following form

$$P_x = \frac{\phi_0}{2} - \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{-i\omega x}\phi_\omega}{i\omega} d\omega \tag{88}$$

which can be proved as

$$\begin{aligned} \int_{\mathbb{R}} \frac{e^{-i\omega x}\phi_\omega}{i\omega} d\omega &= \int_{\mathbb{R}} \frac{e^{-i\omega x}\int_{\mathbb{R}} e^{i\omega y} p_y dy}{i\omega} d\omega = \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{e^{i\omega(y-x)}}{i\omega} p_y dy\, d\omega \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{e^{i\omega(y-x)}}{i\omega} d\omega\, p_y dy = \pi\int_{\mathbb{R}} \mathcal{S}_{x-y} p_y dy = \pi(-\phi_0 + 2P_x) \end{aligned} \tag{89}$$

where $\int_{\mathbb{R}} \frac{e^{i\omega x}}{i\omega} d\omega = \int_{\mathbb{R}^-} \frac{e^{i\omega x}}{i\omega} d\omega + \int_{\mathbb{R}^+} \frac{e^{i\omega x}}{i\omega} d\omega = \int_{\mathbb{R}^+} \frac{e^{-i\omega x}}{-i\omega} d\omega + \int_{\mathbb{R}^+} \frac{e^{i\omega x}}{i\omega} d\omega$

$$= \int_{\mathbb{R}^+} \frac{e^{i\omega x} - e^{-i\omega x}}{i\omega} d\omega = 2\int_{\mathbb{R}^+} \frac{\sin\omega x}{\omega} d\omega = \pi\mathcal{S}_x$$

Note that if p_x is real-valued function (i.e. probability density function), its *Fourier* transform ϕ_ω is then even in its real part and odd in its imaginary part [13], we therefore have $\phi_\omega = \overline{\phi_{-\omega}}$, denoting complex conjugate. The inversion formula (88) becomes identical to (81), which can be further reduced to

$$\begin{aligned} P_x &= \frac{\phi_0}{2} + \frac{1}{2\pi}\int_{\mathbb{R}^+} \frac{e^{i\omega x}\phi_{-\omega} - e^{-i\omega x}\phi_\omega}{i\omega} d\omega = \frac{\phi_0}{2} + \frac{1}{2\pi}\int_{\mathbb{R}^+} \frac{\overline{e^{-i\omega x}\phi_\omega} - e^{-i\omega x}\phi_\omega}{i\omega} d\omega \\ &= \frac{\phi_0}{2} - \frac{1}{\pi}\int_{\mathbb{R}^+} \Re\left[\frac{e^{-i\omega x}\phi_\omega}{i\omega}\right] d\omega = \frac{\phi_0}{2} - \frac{1}{\pi}\int_{\mathbb{R}^+} \Im\left[\frac{e^{-i\omega x}\phi_\omega}{\omega}\right] d\omega \end{aligned} \tag{90}$$

2.3.3. Characteristic Function

In this section, we present a derivation of the closed form characteristic function of the spot in Heston model [14]. The definition of the Heston joint process in (50) will be used. Suppose there exists a payoff function $g(x_T, v_T)$ on x_T and v_T, we may calculate risk neutral expectation of the payoff function as

$$h_t = \widetilde{\mathbb{E}}_t[g(x_T, v_T)] \tag{91}$$

For example, the characteristic function of the joint distribution of x_T and v_T would be given by

$$h_t = \phi_{\alpha,\beta}^{x_T,v_T} = \widetilde{\mathbb{E}}_t\left[g_{\alpha,\beta}(x_T, v_T)\right] \quad \text{where} \quad g_{\alpha,\beta}(x_T, v_T) = \exp(i\alpha x_T + i\beta v_T) \tag{92}$$

The risk neutral expectation of $g(x_T, v_T)$ is a martingale because for $s < t < T$ we have

$$h_s = \widetilde{\mathbb{E}}_s[g(x_T, v_T)] = \widetilde{\mathbb{E}}_s\left[\widetilde{\mathbb{E}}_t[g(x_T, v_T)]\right] = \widetilde{\mathbb{E}}_s[h_t] \tag{93}$$

Applying Ito's lemma to h and forcing the drift to be zero (due to martingale property), we end up with a PDE for h

$$\frac{\partial h}{\partial t} - \frac{v}{2}\frac{\partial h}{\partial x} + \kappa(\theta - v)\frac{\partial h}{\partial v} + \frac{v}{2}\frac{\partial^2 h}{\partial x^2} + \frac{\xi^2 v}{2}\frac{\partial^2 h}{\partial v^2} + \xi\rho v\frac{\partial^2 h}{\partial x \partial v} = 0 \tag{94}$$

To determine the solution of (94), the terminal condition $h_T = \widetilde{\mathbb{E}}_T[g(x_T, v_T)] = g(x_T, v_T)$ at time T must be specified. We will consider that the function at terminal time has the form $g(x_T, v_T) = \exp(\gamma + i\alpha x_T + \delta v_T)$. If $\gamma = 0$ and $\delta = i\beta$, the function becomes $g_{\alpha,\beta}(x_T, v_T)$ as in (92), corresponding to the characteristic function $\phi_{\alpha,\beta}^{x_T,v_T}$ of the joint distribution. If further assuming $\gamma = 0$ and $\delta = 0$, we have $g_\alpha(x_T) = \exp(i\alpha x_T)$, corresponding to the characteristic function $\phi_\alpha^{x_T}$ of the marginal distribution of x_T.

Heston [24] guessed a solution for h that has the form

$$h_t = \widetilde{\mathbb{E}}_t[g(x_T, v_T)] = \exp(Q + i\alpha x_t + D v_t), \qquad Q = Q(\tau, \alpha, \gamma, \delta), \qquad D = D(\tau, \alpha, \gamma, \delta) \tag{95}$$

with $\tau = T - t$. Substituting the tentative solution (95) into (94) yields

$$\frac{\partial Q}{\partial t} + v\frac{\partial D}{\partial t} - \frac{i\alpha v}{2} + \kappa(\theta - v)D - \frac{\alpha^2 v}{2} + \frac{\xi^2 v}{2}D^2 + i\alpha D\xi\rho v = 0 \tag{96}$$

$$\Rightarrow \frac{\partial Q}{\partial t} + \kappa\theta D + \left(\frac{\partial D}{\partial t} + \frac{\xi^2}{2}D^2 - mD - \frac{\alpha(i+\alpha)}{2}\right)v = 0 \quad \text{for} \quad m = \kappa - \alpha\xi\rho i$$

As the v is an independent variable, (96) is zero only if

$$\frac{\partial Q}{\partial t} + \kappa\theta D = 0 \quad \text{and} \quad \frac{\partial D}{\partial t} + \frac{\xi^2}{2}D^2 - mD - \frac{\alpha(i+\alpha)}{2} = 0 \tag{97}$$

Changing the variable t to $\tau = T - t$, we have

$$\frac{\partial Q}{\partial \tau} = \kappa\theta D \quad \text{and} \quad \frac{\partial D}{\partial \tau} = \frac{\xi^2}{2}D^2 - mD - \frac{\alpha(i+\alpha)}{2} \tag{98}$$

The terminal condition for these equations is given by $Q_{\tau=0} = \gamma$ and $D_{\tau=0} = \delta$. The ODE for D is a Riccati equation that only depends on D. This Riccati equation can be turned into an ODE through the change of variable, $Z = \left(D - \widehat{D}\right)^{-1}$, where $\widehat{D}$ is a particular solution to the second equation in (98)

$$\begin{gathered}\frac{\partial Z}{\partial \tau} = -\frac{1}{\left(D-\widehat{D}\right)^2}\frac{\partial\left(D-\widehat{D}\right)}{\partial \tau} = -Z^2\left(\frac{\xi^2}{2}D^2 - mD - \frac{\xi^2}{2}\widehat{D}^2 + m\widehat{D}\right) = -\left(\xi^2\widehat{D} - m\right)Z - \frac{\xi^2}{2} \\ \Rightarrow \frac{\partial Z}{\partial \tau} + BZ + A = 0 \quad \text{where} \quad A = \frac{\xi^2}{2}, \quad B = \xi^2\widehat{D} - m\end{gathered} \tag{99}$$

The solution to (99) is given by

$$\begin{gathered}Z = -\frac{A}{B} + \left(Z_{\tau=0} + \frac{A}{B}\right)e^{-B\tau} \quad \text{with} \quad Z_{\tau=0} = \frac{1}{D_{\tau=0} - \widehat{D}} = \frac{1}{\delta - \widehat{D}} \\ \Rightarrow D = \frac{1}{-\frac{A}{B} + \left(Z_{\tau=0} + \frac{A}{B}\right)e^{-B\tau}} + \widehat{D}\end{gathered} \tag{100}$$

The particular solution $\widehat{D}$ can be as simple as a constant, which implies by the second equation in (98) it could be the solution of the quadratic equation

$$\frac{\xi^2}{2}\widehat{D}^2 - m\widehat{D} - \frac{\alpha(i+\alpha)}{2} = 0 \Rightarrow \widehat{D} = \frac{m+d}{\xi^2} \quad \text{where} \quad d = \pm\sqrt{m^2 + \xi^2\alpha(i+\alpha)} \tag{101}$$

This particular solution for $\widehat{D}$ makes $B = d$. The solution for D in (100) can then be derived as

$$D = \frac{1}{-\frac{\xi^2}{2d} + \left(\frac{-\xi^2}{m+d-\delta\xi^2} + \frac{\xi^2}{2d}\right)e^{-d\tau}} + \frac{m+d}{\xi^2} \tag{102}$$

$$= \frac{1}{\xi^2}\left(m + d - \frac{2d(m + d - \delta\xi^2)}{(m + d - \delta\xi^2) - (m - d - \delta\xi^2)e^{-d\tau}}\right)$$

$$= \frac{1}{\xi^2}\left(m + d - \frac{2d}{1 - \hat{g}e^{-d\tau}}\right) \qquad \left(\text{by defining } \hat{g} = \frac{m - d - \delta\xi^2}{m + d - \delta\xi^2}\right)$$

$$= \frac{1}{\xi^2}\frac{m - d - (m + d)\hat{g}e^{-d\tau}}{1 - \hat{g}e^{-d\tau}} = \frac{m + d}{\xi^2}\frac{g - \hat{g}e^{-d\tau}}{1 - \hat{g}e^{-d\tau}} \qquad \left(\text{by defining } g = \frac{m - d}{m + d}\right)$$

The D can then be plugged into the ODE for Q in (98), such that

$$\frac{\partial Q}{\partial \tau} = \kappa\theta\frac{m + d}{\xi^2}\frac{g - \hat{g}e^{-d\tau}}{1 - \hat{g}e^{-d\tau}} \quad \Rightarrow \quad Q = \kappa\theta\frac{m + d}{\xi^2}\int\frac{g - \hat{g}e^{-d\tau}}{1 - \hat{g}e^{-d\tau}}d\tau + \bar{Q} \tag{103}$$

where $\bar{Q}$ is a constant to be fixed by terminal condition. The indefinite integral in (103) can be solved through change of variable $u = \exp(-d\tau)$ where $\partial u/\partial\tau = -ud$, such that

$$\int\frac{g - \hat{g}e^{-d\tau}}{1 - \hat{g}e^{-d\tau}}\partial\tau = -\int\frac{g - \hat{g}u}{1 - \hat{g}u}\frac{1}{ud}\partial u = -\frac{1}{d}\int\frac{\frac{g}{u} - \hat{g}}{1 - \hat{g}u}\partial u$$

$$= -\frac{1}{d}\int\left(\frac{\frac{g}{u} - \hat{g} - \frac{g}{u}(1 - \hat{g}u)}{1 - \hat{g}u} + \frac{g}{u}\right)\partial u = -\frac{1}{d}\int\left(\frac{(g - 1)\hat{g}}{1 - \hat{g}u} + \frac{g}{u}\right)\partial u \tag{104}$$

$$= \frac{g - 1}{d}\ln(1 - \hat{g}u) - \frac{g}{d}\ln u = \frac{g - 1}{d}\ln(1 - \hat{g}e^{-d\tau}) + g\tau$$

which gives

$$Q = \kappa\theta\frac{m + d}{\xi^2}\left(\frac{g - 1}{d}\ln(1 - \hat{g}e^{-d\tau}) + g\tau\right) + \bar{Q} = \frac{\kappa\theta}{\xi^2}(-2\ln(1 - \hat{g}e^{-d\tau}) + (m - d)\tau) + \bar{Q} \tag{105}$$

We then fix $\bar{Q}$ by $Q_{\tau=0} = \gamma$

$$Q_{\tau=0} = -2\frac{\kappa\theta}{\xi^2}\ln(1 - \hat{g}) + \bar{Q} = \gamma \quad \Rightarrow \quad \bar{Q} = 2\frac{\kappa\theta}{\xi^2}\ln(1 - \hat{g}) + \gamma \tag{106}$$

Therefore we have the solution for Q as below

$$Q = \frac{\kappa\theta}{\xi^2}\left(2\ln\frac{1 - \hat{g}}{1 - \hat{g}e^{-d\tau}} + (m - d)\tau\right) + \gamma \tag{107}$$

Combining the solutions in (102) and (107), the solution h_t to the PDE (94) is of the following form

$$h_t = \widetilde{\mathbb{E}}_t\left[g_{x_T,v_T}\right] = \exp(Q + i\alpha x_t + Dv_t) \qquad \text{where}$$

$$Q = \frac{\kappa\theta}{\xi^2}\left(2\ln\frac{1-\hat{g}}{1-\hat{g}e^{-d\tau}} + (m-d)\tau\right) + \gamma, \qquad D = \frac{m+d}{\xi^2}\frac{g-\hat{g}e^{-d\tau}}{1-\hat{g}e^{-d\tau}} \tag{108}$$

$$m = \kappa - \alpha\xi\rho i, \qquad d = \pm\sqrt{m^2 + \xi^2\alpha(i+\alpha)}, \qquad g = \frac{m-d}{m+d}, \qquad \hat{g} = \frac{m-d-\delta\xi^2}{m+d-\delta\xi^2}$$

The joint characteristic function of x_T and v_T is obtained by having $\gamma = 0$ and $\delta = i\beta$

$$\phi_{\alpha,\beta}^{x_T,v_T} = \exp\left(\frac{\kappa\theta}{\xi^2}\left(2\ln\frac{1-\hat{g}}{1-\hat{g}e^{-d\tau}} + (m-d)\tau\right) + i\alpha x_t + \frac{m+d}{\xi^2}\frac{g-\hat{g}e^{-d\tau}}{1-\hat{g}e^{-d\tau}}v_t\right) \qquad \text{with}$$

$$m = \kappa - \alpha\xi\rho i, \qquad d = \sqrt{m^2 + \xi^2\alpha(i+\alpha)}, \qquad g = \frac{m-d}{m+d}, \qquad \hat{g} = \frac{m-d-i\beta\xi^2}{m+d-i\beta\xi^2} \tag{109}$$

whilst the marginal characteristic function of x_T is given by $\gamma = 0$ and $\delta = 0$ (which implied $\hat{g} = g$)

$$\phi_{\alpha}^{x_T} = \exp\left(\frac{\kappa\theta}{\xi^2}\left(2\ln\frac{1-g}{1-ge^{-d\tau}} + (m-d)\tau\right) + i\alpha x_t + \frac{m-d}{\xi^2}\frac{1-e^{-d\tau}}{1-ge^{-d\tau}}v_t\right) \qquad \text{with}$$

$$m = \kappa - \alpha\xi\rho i, \qquad d = \sqrt{m^2 + \xi^2\alpha(i+\alpha)}, \qquad g = \frac{m-d}{m+d} \tag{110}$$

Further notice that we have $x_t = 0$ almost surely, the term $i\alpha x_t$ in (109) and (110) can be omitted when we evaluate the characteristic function.

In Heston's original derivation [24], the d in (108) takes the negative square root, which makes the calculation of the complex logarithm prone to numerical instabilities. This is because taking the principal value of the logarithm causes Q to jump discontinuously each time the imaginary part of the argument of the logarithm crosses the negative real axis (i.e. discontinuity due to branch cut of complex numbers), especially for long maturities. Albrecher et al. [15] presents an extensive study proving that both positive and negative roots are completely equivalent from a theoretical point of view, it is also mentioned that rather than using the negative root, the positive square root for d guarantees numerical stability of the resulting formula under a full dimensional and unrestricted parameter space.

2.3.4. Vanilla Option Prices

Once we know the analytical form of the characteristic function $\phi_\alpha^{x_T}$ of the centered log-spot x_T, we are able to compute the vanilla option prices using inversion methods. Here, we are going to discuss two methods, which treat the option price function analogous to the cumulative density function or the probability density function, respectively. In addition, we also summarize the original Heston's method [24].

2.3.4.1. *Analogy to Cumulative Density Function*

Assuming deterministic r, the call option is priced by the formula below with x_t defined in (49)

$$C_K = \widetilde{\mathbb{E}}_t\left[\frac{M_t}{M_T}(X_T - K)^+\right] = B_{t,T}F_{t,T}\widetilde{\mathbb{E}}_t[(e^{x_T} - e^k)^+], \qquad k = \ln\frac{K}{F_{t,T}} \tag{111}$$

where k is the log moneyness and the cash account $M_t = \exp\left(\int_s^t r_u du\right)$. We may further define a forward option value in percentage of the underlying forward as

$$\mathcal{C}_k = \frac{C_K}{B_{t,T}F_{t,T}} = \widetilde{\mathbb{E}}_t[(e^{x_T} - e^k)^+] = \widetilde{\mathbb{E}}_t[(e^{x_T} - e^k)\mathbb{1}\{x_T > k\}], \qquad \mathbb{1}\{x > k\} = \begin{cases}1 & \text{if } x > k\\ 0 & \text{else}\end{cases} \tag{112}$$

As the characteristic function of terminal distribution of x_T is already known, we may derive *Fourier* transform of the call option, and then perform inversion to gain its value [16] [17]. Since $\mathcal{C}_k \in [0,1]$ with $\mathcal{C}_{-\infty} = 1$ and $\mathcal{C}_\infty = 0$, we can devise a function $\Theta_h = \mathcal{C}_{k=-h} = \widetilde{\mathbb{E}}_t[(e^{x_T} - e^{-h})\mathbb{1}\{-x_T < h\}]$ such that $\Theta_{-\infty} = 0$ and $\Theta_\infty = 1$, which behaves just like a cumulative density function in h. We derive the *Fourier* transform of Θ_h by

$$\begin{aligned}
\chi_\omega^c &= \int_{k\in\mathbb{R}} e^{i\omega h} d\Theta_h = e^{i\omega h}\Theta_h\big|_{k=-\infty}^{\infty} - \int_{\mathbb{R}} i\omega e^{i\omega h}\Theta_h dh \\
&= e^{i\omega\infty} - \int_{\mathbb{R}} i\omega e^{i\omega h}\int_\Omega (e^x - e^{-h})\mathbb{1}\{-x < h\} dP_x^{x_T}\, dh \\
&= e^{i\omega\infty} - \int_\Omega\int_{\mathbb{R}} i\omega\left(e^{i\omega h + x} - e^{(i\omega-1)h}\right)\mathbb{1}\{-x < h\} dh\, dP_x^{x_T}, \qquad \text{(by Fubini's theorem)} \\
&= e^{i\omega\infty} - \int_\Omega\int_{-x}^{\infty} i\omega\left(e^{i\omega h + x} - e^{(i\omega-1)h}\right) dh\, dP_x^{x_T}
\end{aligned} \tag{113}$$

$$= e^{i\omega\infty} - \int_\Omega \left(e^{i\omega h + x} - \frac{i\omega e^{(i\omega-1)h}}{i\omega - 1} \right) \Bigg|_{h=-x}^{\infty} dP_x^{x_T}$$

$$= e^{i\omega\infty} - \int_\Omega \left(e^{i\omega\infty + x} - e^{(-i\omega+1)x} - \underbrace{\frac{i\omega e^{(i\omega-1)\infty}}{i\omega - 1}}_{=0} + \frac{i\omega e^{-(i\omega-1)x}}{i\omega - 1} \right) dP_x^{x_T}$$

$$= e^{i\omega\infty} - e^{i\omega\infty} \int_\Omega e^x dP_x^{x_T} + \int_\Omega \left(e^{(-i\omega+1)x} - \frac{i\omega e^{-(i\omega-1)x}}{i\omega - 1} \right) dP_x^{x_T}$$

$$= \frac{1}{1 - i\omega} \int_\Omega e^{i(-\omega-i)x} dP_x^{x_T} \qquad \left(\text{by } \int_\Omega e^x dP_x^{x_T} = \frac{1}{F_{t,T}} \widetilde{\mathbb{E}}_t[X_T] = 1 \right)$$

$$= \frac{\phi_{-\omega-i}^{x_T}}{1 - i\omega}$$

The χ_ω^c is the characteristic function of the call variant Θ_h. The option value $\mathcal{C}_k$ is then given by the *Levy's Inversion Formula* (81) [18] through below steps

$$\begin{aligned}
\Theta_h &= \frac{\chi_0^c}{2} + \frac{1}{2\pi} \int_{\mathbb{R}+} \frac{e^{i\omega h} \chi_{-\omega}^c - e^{-i\omega h} \chi_\omega^c}{i\omega} d\omega \\
&= \frac{1}{2} + \frac{1}{2\pi} \int_{\mathbb{R}+} \frac{e^{i\omega h} \frac{\phi_{\omega-i}^{x_T}}{i\omega + 1} + e^{-i\omega h} \frac{\phi_{-\omega-i}^{x_T}}{i\omega - 1}}{i\omega} d\omega \qquad \left(\text{by } \chi_0^c = \phi_{-i}^{x_T} = 1 \right) \\
&= \frac{1}{2} - \frac{1}{2\pi} \int_{\mathbb{R}+} \left(e^{i\omega h} \frac{\phi_{\omega-i}^{x_T}}{\omega^2 - i\omega} + e^{-i\omega h} \frac{\phi_{-\omega-i}^{x_T}}{\omega^2 + i\omega} \right) d\omega \\
\Longrightarrow \mathcal{C}_k &= \Theta_{h=-k} = \frac{1}{2} - \frac{1}{2\pi} \int_{\mathbb{R}+} \left(e^{i\omega k} \frac{\phi_{-\omega-i}^{x_T}}{\omega^2 + i\omega} + e^{-i\omega k} \frac{\phi_{\omega-i}^{x_T}}{\omega^2 - i\omega} \right) d\omega
\end{aligned} \tag{114}$$

where the characteristic function $\phi_\alpha^{x_T}$ of the terminal distribution of x_T is given in (110). The inversion formula in (114) involves evaluation of the $\phi_\alpha^{x_T}$ function twice, once with $\alpha = \omega - i$ and the other with $\alpha = -\omega - i$, which is less efficient. Since $\mathcal{C}_k$ is real-valued, so is Θ_h, we may just use (90) to perform the inversion, that is

$$\Theta_h = \frac{\chi_0^c}{2} - \frac{1}{\pi} \int_{\mathbb{R}+} \Im \left[\frac{e^{-i\omega h} \chi_\omega^c}{\omega} \right] d\omega = \frac{1}{2} + \frac{1}{\pi} \int_{\mathbb{R}+} \Im \left[e^{-i\omega h} \frac{\phi_{-\omega-i}^{x_T}}{i\omega^2 - \omega} \right] d\omega \tag{115}$$

$$\Longrightarrow \mathcal{C}_k = \Theta_{h=-k} = \frac{1}{2} + \frac{1}{\pi}\int_{\mathbb{R}^+} \Im\left[e^{i\omega k}\frac{\phi_{-\omega-i}^{x_T}}{i\omega^2-\omega}\right]d\omega$$

The integral can be estimated numerically, e.g. by *Gauss-Laguerre* quadrature. Once we have $\mathcal{C}_k$ in log-moneyness k, the option value can be easily derived from relation in (112), e.g. $C_K = B_{t,T}F_{t,T}\mathcal{C}_k$ for a call, or $P_K = C_K - B_{t,T}\left(F_{t,T}-K\right)$ for a put by call-put parity.

2.3.4.2. *Analogy to Probability Density Function*

In this section, we will treat $\mathcal{C}_k$ analogous to a probability density [19] [20] and define in terms of generalized *Fourier* transform with $z = z_r - iz_i,\ \ z_r \in \mathbb{R},\ \ z_i \in \mathcal{D} \subseteq \mathbb{R}^+$

$$\begin{aligned}
&\chi_z^p = \int_{\mathbb{R}} e^{izk}\mathcal{C}_k dk = \int_{\mathbb{R}} e^{izk}\widetilde{\mathbb{E}}_t[(e^{x_T}-e^k)\mathbb{1}\{x_T > k\}]dk \\
&= \widetilde{\mathbb{E}}_t\left[\int_{\mathbb{R}} e^{izk}(e^{x_T}-e^k)\mathbb{1}\{x_T>k\}dk\right] = \widetilde{\mathbb{E}}_t\left[\int_{-\infty}^{x_T} e^{izk}(e^{x_T}-e^k)dk\right] \\
&= \widetilde{\mathbb{E}}_t\left[\left(\frac{e^{izk+x_T}}{iz} - \frac{e^{(iz+1)k}}{iz+1}\right)\Bigg|_{k=-\infty}^{x_T}\right] = \widetilde{\mathbb{E}}_t\left[\frac{e^{(iz+1)x_T}}{iz-z^2}\right] \quad \left(\text{by } \lim_{k\to-\infty} e^{izk} = 0\right) \\
&= \widetilde{\mathbb{E}}_t\left[\frac{e^{i(z-i)x_T}}{iz-z^2}\right] = \frac{\phi_{z-i}^{x_T}}{iz-z^2}
\end{aligned} \tag{116}$$

For some distributions, the transform $\phi_{z-i}^{x_T}$ is well-defined only when z_i is in a subset of the real line. We use $\mathcal{D} \subseteq \mathbb{R}^+$ to denote the subset that both guarantees the convergence of e^{izk} and $e^{i(z-i)k}$ as $k \longrightarrow -\infty$, and assures the finiteness of the transform $\phi_{z-i}^{x_T}$.

The forward option value $\mathcal{C}_k$ is then given by the inverse *Fourier* transform

$$\begin{aligned}
\mathcal{C}_k &= \frac{1}{2\pi}\int_{-\infty-iz_i}^{\infty-iz_i} e^{-izk}\chi_z^p dz = \frac{1}{2\pi}\int_{-\infty-iz_i}^{\infty-iz_i} e^{-izk}\chi_z^p d(z_r - iz_i) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-izk}\chi_z^p dz_r \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-izk}\frac{\phi_{z-i}^{x_T}}{iz-z^2}dz_r = \frac{1}{\pi}\int_{\mathbb{R}^+} \Re\left[e^{-izk}\frac{\phi_{z-i}^{x_T}}{iz-z^2}\right]dz_r
\end{aligned} \tag{117}$$

where $\phi_\alpha^{x_T}$ is given in (110). The last equality holds because C_k is real, which implies that the function χ_z^p is odd in its imaginary part and even in its real part. The functional form of $\phi_\alpha^{x_T}$ can be derived from (110) in a similar manner.

2.3.4.3. *Heston's Original Solution*

At time t, the European call value with spot X_t and strike K is given by the no-arbitrage formula

$$\begin{aligned} C_K &= \widetilde{\mathbb{E}}_t\left[\frac{M_t}{M_T}(X_T-K)^+\right] = \widetilde{\mathbb{E}}_t\left[\frac{M_t}{M_T}X_T\mathbb{1}\{X_T>K\}\right] - K\widetilde{\mathbb{E}}_t\left[\frac{M_t}{M_T}\mathbb{1}\{X_T>K\}\right] \\ &= \mathbb{E}_t^{\hat{B}X}\left[\frac{\hat{B}_{t,T}X_t}{\hat{B}_{T,T}X_T}X_T\mathbb{1}\{X_T>K\}\right] - K\mathbb{E}_t^T\left[\frac{B_{t,T}}{B_{T,T}}\mathbb{1}\{X_T>K\}\right], \qquad \begin{matrix}\text{change of numeraire}\\ M_t \text{ to } \hat{B}_{t,T}X_t \text{ and } B_{t,T}\end{matrix} \\ &= \hat{B}_{t,T}X_t\mathbb{E}_t^{\hat{B}X}[\mathbb{1}\{X_T>K\}] - B_{t,T}K\mathbb{E}_t^T[\mathbb{1}\{X_T>K\}] \\ &= B_{t,T}\left(F_{t,T}\mathbb{P}_t^{\hat{B}X}[X_T>K] - K\mathbb{P}_t^T[X_T>K]\right) \end{aligned} \tag{118}$$

where $\mathbb{P}_t^{\hat{B}X}[X_T>K]$ and $\mathbb{P}_t^T[X_T>K]$ are both conditional probabilities of spot finishing in-the-money at maturity. The $\mathbb{P}_t^{\hat{B}X}[X_T>K]$ is computed under the measure associated with numeraire $\hat{B}_{t,T}X_t$, whereas the $\mathbb{P}_t^T[X_T>K]$ is computed under T-forward measure associated with zero coupon bond $B_{t,T}$ as the numeraire [21]. In FX markets, the $\hat{B}_{t,T}$ denotes the foreign zero coupon bond. In stock markets, the spot X_t is assumed to be *non-dividend-bearing* in order to qualify for a numeraire, which makes $\hat{B}_{t,T}=1$. In Black-Scholes model

$$C_K^{BS} = \hat{B}_{t,T}X_t\Phi(d_+) - KB_{t,T}\Phi(d_-), \qquad d_\pm = \frac{1}{\sigma\sqrt{\tau}}\ln\frac{F_{t,T}}{K} \pm \frac{\sigma\sqrt{\tau}}{2} \tag{119}$$

the $\mathbb{P}_t^{\hat{B}X}[X_T>K]$ and $\mathbb{P}_t^T[X_T>K]$ are computed as $\Phi(d_+)$ and $\Phi(d_-)$ respectively, where Φ is the standard normal cumulative density. Since the drift adjustment due to change of numeraire is

$$\underset{\text{Under }\mathbb{N}}{dW_t^{\mathbb{N}}} = \underset{\text{Under }\mathbb{Q}}{d\widetilde{W}_t} - \sigma_N dt \tag{120}$$

where $\mathbb{N}$ denotes the measure associated with numeraire N and $\mathbb{Q}$ the risk neutral measure. Assuming zero dividend for stock (where $\hat{B}_{t,T} = 1$) or deterministic foreign rate q for FX (where $\hat{B}_{t,T}$ is deterministic), the process X_t under the measure associated with itself as the numeraire would be

$$\frac{dX_t}{X_t} = (r-q)dt + \sigma d\widetilde{W}_t = (r-q+\sigma^2)dt + \sigma dW_t^X \tag{121}$$

The total drift adjustment $\sigma^2\tau$ for period $\tau = T - t$ is then normalized by the total volatility $\sigma\sqrt{\tau}$ of the stock to give a shift term $\sigma\sqrt{\tau}$ as the difference between d_+ and d_- in the classic Black-Scholes formula.

Suppose we use the definition of $\mathcal{C}_k$ in (112) for a call, the two conditional probabilities can be expressed as

$$\begin{aligned} &\mathcal{C}_k = \widetilde{\mathbb{E}}_t[(e^{x_T} - e^k)^+] = \widetilde{\mathbb{E}}_t[e^{x_T}\mathbb{1}\{x_T > k\}] - e^k\widetilde{\mathbb{E}}_t[\mathbb{1}\{x_T > k\}] = P_k^+ - e^k P_k^- \quad \text{with} \\ &P_k^+ = \widetilde{\mathbb{E}}_t[e^{x_T}\mathbb{1}\{x_T > k\}], \qquad P_k^- = \widetilde{\mathbb{E}}_t[\mathbb{1}\{x_T > k\}] \end{aligned} \tag{122}$$

Because in Heston model the P_k^+ and P_k^- are not available in closed form, Heston [24] sought to firstly derive the characteristic functions (i.e. the *Fourier* transforms) of P_k^+ and P_k^- by solving the PDE (94) for each of them, and then obtain the inverse of the two characteristic functions for the option price. Since we already have the characteristic function of x_T as in (110), we can easily derive those for P_k^+ and P_k^- respectively. For example, the $\Theta_h^+ = P_{k=-h}^+ = \widetilde{\mathbb{E}}_t[e^{x_T}\mathbb{1}\{-x_T < h\}] \in [0,1]$ with $\Theta_{-\infty}^+ = 0$ and $\Theta_\infty^+ = 1$ can be treated as a CDF and then the characteristic function of P_h^+ by following similar derivations in (113) becomes

$$\begin{aligned} \chi_\omega^+ &= \int_{k\in\mathbb{R}} e^{i\omega h} d\Theta_h^+ = e^{i\omega h}\Theta_h^+\big|_{h=-\infty}^{\infty} - \int_{\mathbb{R}} i\omega e^{i\omega h}\Theta_h^+ dh \\ &= e^{i\omega\infty} - \int_{\mathbb{R}} i\omega e^{i\omega h}\int_\Omega e^x \mathbb{1}\{-x < h\} dP_x^{x_T}\, dh = e^{i\omega\infty} - \int_\Omega\int_{\mathbb{R}} i\omega e^{i\omega h + x}\mathbb{1}\{-x < h\} dh\, dP_x^{x_T} \\ &= e^{i\omega\infty} - \int_\Omega\int_{-x}^{\infty} i\omega e^{i\omega h + x} dh\, dP_x^{x_T} = e^{i\omega\infty} - \int_\Omega e^{i\omega h + x}\big|_{h=-x}^{\infty} dP_x^{x_T} \\ &= e^{i\omega\infty} - \int_\Omega e^{i\omega\infty + x} dP_x^{x_T} + \int_\Omega e^{-i\omega x + x} dP_x^{x_T} = \int_\Omega e^{i(-\omega - i)x} dP_x^{x_T} \end{aligned} \tag{123}$$

$$= \phi_{-\omega-i}^{x_T}$$

Similarly, $\Theta_h^- = P_{k=-h}^- = \widetilde{\mathbb{E}}_t[\mathbb{1}\{-x_T < h\}] \in [0,1]$ with $\Theta_{-\infty}^- = 0$ and $\Theta_{\infty}^- = 1$, the characteristic function of Θ_h^- is derived as

$$\begin{aligned}
\chi_\omega^- &= \int_{k\in\mathbb{R}} e^{i\omega h} d\Theta_h^- = e^{i\omega h}\Theta_h^-\Big|_{h=-\infty}^{\infty} - \int_{\mathbb{R}} i\omega e^{i\omega h}\Theta_h^- dh \\
&= e^{i\omega\infty} - \int_{\mathbb{R}} i\omega e^{i\omega h} \int_\Omega \mathbb{1}\{-x < h\} dP_x^{x_T}\, dh = e^{i\omega\infty} - \int_\Omega \int_{\mathbb{R}} i\omega e^{i\omega h}\mathbb{1}\{-x < h\} dh\, dP_x^{x_T} \\
&= e^{i\omega\infty} - \int_\Omega \int_{-x}^{\infty} i\omega e^{i\omega h} dh\, dP_x^{x_T} = e^{i\omega\infty} - \int_\Omega e^{i\omega h}\Big|_{h=-x}^{\infty} dP_x^{x_T} = \int_\Omega e^{-i\omega x} dP_x \\
&= \phi_{-\omega}^{x_T}
\end{aligned} \tag{124}$$

The $P_k^+ = \Theta_{h=-k}^+$ and $P_k^- = \Theta_{h=-k}^-$ are then obtained through the inverse of χ_ω^+ and χ_ω^- respectively. Given that $\chi_0^+ = \chi_0^- = 1$, the Heston's vanilla option price formula can be derived from $\phi_\alpha^{x_T}$ in (110) along with the inverse formula in (90). We summarize the final solution below

$$\mathcal{C}_k = P_k^+ - e^k P_k^- \quad \text{with} \quad P_k^\pm = \frac{1}{2} - \frac{1}{\pi}\int_{\mathbb{R}^+} \Im\left[e^{i\omega k}\frac{\chi_\omega^\pm}{\omega}\right] d\omega\,, \qquad \chi_\omega^+ = \phi_{-\omega-i}^{x_T}, \qquad \chi_\omega^- = \phi_{-\omega}^{x_T} \tag{125}$$

2.4. Piecewise Time Dependent Heston Model

The time-dependent Heston model we present here was proposed by Elices [22] in 2008. The model relies on the joint characteristic function of the two-dimensional Markov process $Y_t = \begin{pmatrix} x_t \\ v_t \end{pmatrix}$, which we have derived in (110). It bootstraps a series of piecewise constant Heston parameters starting from the earliest maturity, each set of parameters for one period of time, which allows the model to fit to a term structure of the implied volatility surfaces.

Without loss of generality, let us consider an n-dimensional Markov stochastic process Y_t. We define $p_{y_2}^{1,2}$ the transition probability density having $Y = y_2$ at time t_2 conditional on $Y = y_1$ at time t_1 for $t_0 \le t_1 \le t_2$. Its joint characteristic function $\phi_\omega^{1,2}$ is the multi-dimensional *Fourier* transform of the density $p_{y_2}^{1,2}$, such that

$$\phi_\omega^{0,2} = \int_{\mathbb{R}^n} e^{i\omega' y_2} p_{y_2}^{0,2} dy_2 = \int_{\mathbb{R}^n} e^{i\omega' y_2} \int_{\mathbb{R}^n} p_{y_2}^{1,2} p_{y_1}^{0,1} dy_1\, dy_2 = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} e^{i\omega' y_2} p_{y_2}^{1,2} p_{y_1}^{0,1} dy_1\, dy_2$$
$$= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} e^{i\omega' y_2} p_{y_2}^{1,2} dy_2\, p_{y_1}^{0,1} dy_1 = \int_{\mathbb{R}^n} \phi_\omega^{1,2} p_{y_1}^{0,1} dy_1 \tag{126}$$

Consider a family of exponential characteristic functions with exponent linear in the state y_1 at time t_1

$$\phi_\omega^{1,2} = \exp(A^{1,2}(\omega) + B^{1,2}(\omega)' y_1) \tag{127}$$

we have

$$\phi_\omega^{0,2} = \int_{\mathbb{R}^n} \exp(A^{1,2}(\omega) + B^{1,2}(\omega)' y_1)\, p_{y_1}^{0,1} dy_1 = \exp\big(A^{1,2}(\omega)\big) \int_{\mathbb{R}^n} \exp(B^{1,2}(\omega)' y_1)\, p_{y_1}^{0,1} dy_1$$
$$= \exp\big(A^{1,2}(\omega)\big) \int_{\mathbb{R}^n} \exp\Big(i\big(-iB^{1,2}(\omega)\big)' y_1\Big) p_{y_1}^{0,1} dy_u = \exp\big(A^{1,2}(\omega)\big)\, \phi^{0,1}\big(-iB^{1,2}(\omega)\big) \tag{128}$$
$$= \exp\Big(A^{1,2}(\omega) + A^{0,1}\big(-iB^{1,2}(\omega)\big) + B^{0,1}\big(-iB^{1,2}(\omega)\big)' y_0\Big)$$

where A is a complex scalar function and B is an n-dimensional complex vector function. Identifying terms between the (127) when $t_1 = t_0$ and the (128), we find that

$$A^{0,2}(\omega) = A^{1,2}(\omega) + A^{0,1}\big(-iB^{1,2}(\omega)\big), \qquad B^{0,2}(\omega) = B^{0,1}\big(-iB^{1,2}(\omega)\big) \tag{129}$$

Formula (129) gives a recursive definition of $A^{0,2}(\omega)$ and $B^{0,2}(\omega)$, which can be used to bootstrap the Heston model parameters for each time period starting from the earliest maturity. For illustrative purpose, the first 3 periods are presented below

Tenure	A	B
$t_0 \to t_1$	$A^{0,1}(\omega)$	$B^{0,1}(\omega)$
$t_0 \to t_2$	$A^{1,2}(\omega) + A^{0,1}\big(-iB^{1,2}(\omega)\big)$	$B^{0,1}\big(-iB^{1,2}(\omega)\big)$
$t_0 \to t_3$	$A^{2,3}(\omega) + A^{1,2}\big(-iB^{2,3}(\omega)\big) + A^{0,1}\Big(-iB^{1,2}\big(-iB^{2,3}(\omega)\big)\Big)$	$B^{0,1}\Big(-iB^{1,2}\big(-iB^{2,3}(\omega)\big)\Big)$

In Heston model, we see from (109) that these parameters correspond to

$$\omega = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \qquad A(\omega) = \frac{\kappa\theta}{\xi^2}\left(2\ln\frac{1-\hat{g}}{1-\hat{g}e^{-d\tau}} + (m-d)\tau\right), \qquad B(\omega) = \begin{pmatrix} i\alpha \\ \dfrac{m+d}{\xi^2}\dfrac{g-\hat{g}e^{-d\tau}}{1-\hat{g}e^{-d\tau}}\end{pmatrix} \tag{130}$$

It can be seen that, for the first entry, we always have $(-iB)_1 = \omega_1 = \alpha$.

2.5. Forward Equation

Heston model

$$\frac{dS_t}{S_t} = (r - \hat{r})dt + \sqrt{v_t}dW_{1,t}, \qquad dv_t = \kappa(\theta - v_t)dt + \xi\sqrt{v_t}dW_{2,t}, \qquad dW_{1,t}dW_{2,t} = \rho dt$$

$$dX_t = \left(r - \hat{r} - \frac{v_t}{2}\right)dt + \sqrt{v_t}dW_{1,t}, \qquad X_t = \ln S_t$$

$$Y_t = \ln v_t \tag{131}$$

$$dY_t = \frac{1}{v_t}dv_t - \frac{1}{2v_t^2}dv_t dv_t = \frac{\kappa(\theta - v_t)}{v_t}dt + \frac{\xi}{\sqrt{v_t}}dW_{2,t} - \frac{\xi^2}{2v_t}dt$$

$$= \left(\frac{2\kappa\theta - \xi^2}{2v_t} - \kappa\right)dt + \frac{\xi}{\sqrt{v_t}}dW_{2,t}$$

Stochastic Local Volatility model

$$dX_t = \left(r - \hat{r} - \frac{\sigma_{t,X}^2\eta_t^2 v_t}{2}\right)dt + \sigma_{t,X}\eta_t\sqrt{v_t}dW_{1,t},$$

$$dY_t = \left(\frac{2\kappa\theta - \xi^2}{2v_t} - \kappa\right)dt + \frac{\xi}{\sqrt{v_t}}dW_{2,t} \tag{132}$$

Let $\delta(t, X_t, Y_t) = \sigma_{t,X}\eta_t\sqrt{v_t}$, we can rewrite the model

$$dH_t = A(t, H_t)dt + B(t, H_t)dW_t, \qquad H_t = \begin{pmatrix} X_t \\ Y_t \end{pmatrix}, \qquad dW_t dW_t' = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}dt$$

$$A(t, H_t) = \begin{pmatrix} r - \hat{r} - \frac{1}{2}\delta^2 \\ \frac{2\kappa\theta - \xi^2}{2v} - \kappa \end{pmatrix}, \qquad B(t, H_t) = \begin{pmatrix} \delta & 0 \\ 0 & \frac{\xi}{\sqrt{v}} \end{pmatrix} \tag{133}$$

The *forward* PDE for the transition probability density $p = p(t, H_t|s, H_s)$ corresponding to the 2D process H_t can be derived from (10)

$$\frac{\partial p}{\partial t} + \sum_{i=1}^{2}\frac{\partial(pA_i)}{\partial H_i} - \frac{1}{2}\sum_{i,j=1}^{2}\frac{\partial^2\left(p\sum_{k=1}^{2}\rho_{ij}B_{ik}B_{jk}\right)}{\partial H_i \partial H_j} = 0 \tag{134}$$

where $\rho_{ij} = \rho$ if $i \neq j$ or 1 otherwise. This translates into

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2(\delta^2 p)}{\partial X^2} + \frac{1}{2}\frac{\partial^2\left(\frac{\xi^2}{v}p\right)}{\partial Y^2} - \frac{\partial\left(\left(r - \hat{r} - \frac{1}{2}\delta^2\right)p\right)}{\partial X} - \frac{\partial\left(\left(\frac{2\kappa\theta - \xi^2}{2v} - \kappa\right)p\right)}{\partial Y} + \frac{\partial^2\left(\rho\frac{\xi\delta}{\sqrt{v}}p\right)}{\partial X \partial Y} \tag{135}$$

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2(\delta^2 p)}{\partial X^2} + \frac{1}{2}\frac{\partial^2\left(\frac{\xi^2}{v}p\right)}{\partial Y^2} - \frac{\partial\left(\left(r - \hat{r} - \frac{1}{2}\delta^2\right)p\right)}{\partial X} - \frac{\partial\left(\left(\frac{2\kappa\theta - \xi^2}{2v} - \kappa\right)p\right)}{\partial Y} + \frac{\partial^2(\rho\xi\sigma\eta p)}{\partial X \partial Y}$$

From (47),

$$\begin{aligned} \mathbb{E}[v_t|v_s] &= \theta + (v_s - \theta)e^{-\kappa(t-s)} \\ \mathbb{V}[v_t|v_s] &= \frac{v_s\xi^2}{\kappa}\left(e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}\right) + \frac{\theta\xi^2}{2\kappa}\left(1 - e^{-\kappa(t-s)}\right)^2 \end{aligned} \tag{136}$$

3. HESTON MODEL: PDE BY FINITE ELEMENT METHOD

In 1973, Black and Scholes introduced a simple formula [23] to price European-style options under a few strong assumptions. It was the first successful attempt to provide an arbitrage free valuation of financial derivatives. However due to limitations, the model fails to capture some critical features observed in financial markets, such as heavy tails of return, skewness and smile in implied volatility, clustering and autocorrelation in volatility, etc. Many approaches have been proposed to address these issues. One of such approaches is to assume the volatility process is stochastic and correlated with spot process. Heston [24] proposed a stochastic volatility model in 1993. It extends the Black-Scholes model and includes it as a special case. One major advantage of the Heston model is that it can be solved in closed-form for vanilla options. For exotic products, the prices are usually obtained numerically. Traditionally, the PDE's that arise from option pricing are generally solved by finite difference method

(FDM). FDM is straightforward to implement, but it also imposes many strong constraints. For example, it may demand sufficiently smooth terminal and boundary conditions, rectangular domains, etc. To relax these constraints, finite element methods (FEM) appear to be promising candidates for solving such PDE's. Many studies have been focused on this topic. Topper [25] provides an excellent introduction to the FEM in the context of financial engineering applications. Previous work by Winkler et al. [26] illustrates an application of FEM to valuation of vanilla option in Heston stochastic volatility model. Achdou and Tchou [27] Propose a finite element analysis for Black-Scholes equation with stochastic volatility process. Miglio and Sgarra [28] discuss an application of finite element method for option pricing in a stochastic volatility model with jumps, known as the Bates model.

In this chapter, we will present an implementation of finite element method in the Heston model for pricing exotic options. Our journey starts with the PDE arising from the Heston model.

3.1. The Partial Differential Equation

Let $U(t,v_t,X_t)=\frac{1}{D_t}\widetilde{\mathbb{E}}_t[D_T U(T,v_T,X_T)]$ be the price of a contingent payment $U(T,v_T,X_T)$ that occur at maturity T. Assuming deterministic interest rate, we have the dynamics of $D_tU(t,v_t,X_t)$ in Heston model (40) under risk neutral measure $\mathbb{Q}$, that is

$$\begin{aligned}
&\frac{1}{D_t}d\big(D_tU(t,v_t,X_t)\big)=dU-rUdt\\
&=\frac{\partial U}{\partial t}dt+\frac{\partial U}{\partial X}dX+\frac{1}{2}\frac{\partial^2 U}{\partial X^2}dXdX+\frac{\partial U}{\partial v}dv+\frac{1}{2}\frac{\partial^2 U}{\partial v^2}dvdv+\frac{\partial^2 U}{\partial v\partial X}dvdX-rUdt\\
&=\frac{\partial U}{\partial t}dt+\frac{\partial U}{\partial X}(r-q)Xdt+\frac{\partial U}{\partial X}X\sqrt{v}d\widetilde{W}_1+\frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2}dt+\frac{\partial U}{\partial v}\kappa(\theta-v)dt+\frac{\partial U}{\partial v}\xi\sqrt{v}d\widetilde{W}_2\\
&\qquad+\frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2}dt+Xv\xi\rho\frac{\partial^2 U}{\partial v\partial X}dt-rUdt
\end{aligned}\tag{137}$$

Since the $D_tU(t,v_t,X_t)$ is a martingale under $\mathbb{Q}$, the dt-term in (137) must vanish, which defines a parabolic diffusion-convection-reaction PDE that a derivative price U must follow

$$\frac{\partial U}{\partial t}+\frac{v\xi^2}{2}\frac{\partial^2 U}{\partial v^2}+Xv\xi\rho\frac{\partial^2 U}{\partial v\partial X}+\frac{vX^2}{2}\frac{\partial^2 U}{\partial X^2}+\kappa(\theta-v)\frac{\partial U}{\partial v}+(r-q)X\frac{\partial U}{\partial X}-rU=0\tag{138}$$

By change of variable $y_t = \ln X_t$, PDE (138) can be further transformed into

$$\frac{\partial w}{\partial t} + \frac{\xi^2 v}{2}\frac{\partial^2 w}{\partial v^2} + \rho\xi v\frac{\partial^2 w}{\partial v \partial y} + \frac{v}{2}\frac{\partial^2 w}{\partial y^2} + \kappa(\theta - v)\frac{\partial w}{\partial v} + \left(r - q - \frac{v}{2}\right)\frac{\partial w}{\partial y} - rw = 0 \tag{139}$$

where $w(t, v_t, y_t) = U(t, v_t, X_t)$. The PDE (139) is actually identical to our previously derived (94) knowing that

$$w(t, v_t, y_t) = e^{rt}h(t, v_t, x_t), \qquad x_t = \ln\frac{X_t}{X_0} - (r - q)t \tag{140}$$

This can be shown by using the chain rule to derive partial derivatives after change of variables

$$\begin{aligned}
&\frac{\partial w}{\partial t} = \left(\frac{\partial}{\partial t} + \frac{\partial x}{\partial t}\frac{\partial}{\partial x}\right)(e^{rt}h) = e^{rt}\left(rh + \frac{\partial h}{\partial t} + \frac{\partial x}{\partial t}\frac{\partial h}{\partial x}\right) = e^{rt}\left(rh + \frac{\partial h}{\partial t} - (r - q)\frac{\partial h}{\partial x}\right)\\
&\frac{\partial w}{\partial v} = e^{rt}\frac{\partial h}{\partial v}, \qquad \frac{\partial^2 w}{\partial v^2} = e^{rt}\frac{\partial^2 h}{\partial v^2}, \qquad \frac{\partial^2 w}{\partial v \partial y} = \frac{\partial}{\partial v}\left(\frac{\partial (e^{rt}h)}{\partial x}\frac{\partial x}{\partial y}\right) = e^{rt}\frac{\partial^2 h}{\partial v \partial x}\\
&\frac{\partial w}{\partial y} = \frac{\partial (e^{rt}h)}{\partial x}\frac{\partial x}{\partial y} = e^{rt}\frac{\partial h}{\partial x}, \qquad \frac{\partial^2 w}{\partial y^2} = e^{rt}\frac{\partial}{\partial y}\left(\frac{\partial h}{\partial t}\frac{\partial t}{\partial y} + \frac{\partial h}{\partial x}\frac{\partial x}{\partial y}\right) = e^{rt}\frac{\partial^2 h}{\partial x^2}
\end{aligned} \tag{141}$$

For simplicity, the PDE (139) can be further expressed in terms of the *gradient* and *divergence* operator

$$(\partial_t + \boldsymbol{\nabla}\cdot\mathbf{A}\boldsymbol{\nabla} - \mathbf{b}\cdot\boldsymbol{\nabla} - r)w = 0, \qquad \mathbf{A} = \frac{v}{2}\begin{pmatrix}\xi^2 & \rho\xi\\ \rho\xi & 1\end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix}\frac{\xi^2}{2} - \kappa(\theta - v)\\ \frac{v + \rho\xi}{2} - (r - q)\end{pmatrix} \tag{142}$$

where $\partial_t = \partial/\partial t$, the gradient operator $\boldsymbol{\nabla} = \begin{pmatrix}\partial_v\\ \partial_y\end{pmatrix}$, and the divergence operator $\boldsymbol{\nabla}\cdot = \begin{pmatrix}\partial_v\\ \partial_y\end{pmatrix}\cdot$. Note that we have the following identities

$$\begin{aligned}
\boldsymbol{\nabla}\cdot\mathbf{A}\boldsymbol{\nabla} &= \begin{pmatrix}\partial_v\\ \partial_y\end{pmatrix}^T\frac{v}{2}\begin{pmatrix}\xi^2 & \rho\xi\\ \rho\xi & 1\end{pmatrix}\begin{pmatrix}\partial_v\\ \partial_y\end{pmatrix} = \begin{pmatrix}\partial_v\\ \partial_y\end{pmatrix}^T\frac{v}{2}\begin{pmatrix}\xi^2\partial_v + \rho\xi\partial_y\\ \rho\xi\partial_v + \partial_y\end{pmatrix}\\
&= \frac{\xi^2\partial_v + \rho\xi\partial_y}{2} + \frac{v}{2}\left(\xi^2\partial_{vv} + \rho\xi\partial_{vy}\right) + \frac{v}{2}\left(\rho\xi\partial_{vy} + \partial_{yy}\right)\\
&= \frac{\xi^2}{2}\partial_v + \frac{\rho\xi}{2}\partial_y + \rho\xi v\partial_{vy} + \frac{v\xi^2}{2}\partial_{vv} + \frac{v}{2}\partial_{yy}
\end{aligned} \tag{143}$$

$$\mathbf{b}\cdot\boldsymbol{\nabla}=\begin{pmatrix}\frac{\xi^2}{2}-\kappa(\theta-v)\\ \frac{v+\rho\xi}{2}-(r-q)\end{pmatrix}^T\begin{pmatrix}\partial_v\\ \partial_y\end{pmatrix}=\frac{\xi^2}{2}\partial_v-\kappa(\theta-v)\partial_v+\frac{v+\rho\xi}{2}\partial_y-(r-q)\partial_y$$

$$\boldsymbol{\nabla}\cdot\mathbf{A}\boldsymbol{\nabla}-\mathbf{b}\cdot\boldsymbol{\nabla}=\rho\xi v\partial_{vy}+\frac{v\xi^2}{2}\partial_{vv}+\frac{v}{2}\partial_{yy}+\kappa(\theta-v)\partial_v+\left(r-q-\frac{v}{2}\right)\partial_y$$

Henceforth, we will use $(\cdot)$ to denote "dot product" of vectors. Here, we may think of the divergence operator as a transpose of the gradient operator (a bit abuse of notation). Derivative products can be priced by solving the PDE (142) under different terminal and boundary conditions. We first define a rectangle bounded domain

$$\Omega\coloneqq\{(v,y)\in\mathbb{R}^2:v\in(v_{\min},v_{\max}),\ \ y\in(y_{\min},y_{\max})\} \tag{144}$$

with the boundary of Ω denoted by $\partial\Omega$ and the closure $\bar{\Omega}=\Omega\cup\partial\Omega$. The rectangle domain have four pieces of boundaries defined as follows

$$\begin{aligned}\Gamma_1&\coloneqq\{(v,y)\in\mathbb{R}^2:v=v_{\min},\ \ y\in(y_{\min},y_{\max})\}\\ \Gamma_2&\coloneqq\{(v,y)\in\mathbb{R}^2:v=v_{\max},\ \ y\in(y_{\min},y_{\max})\}\\ \Gamma_3&\coloneqq\{(v,y)\in\mathbb{R}^2:v\in[v_{\min},v_{\max}],\ \ y=y_{\min}\}\\ \Gamma_4&\coloneqq\{(v,y)\in\mathbb{R}^2:v\in[v_{\min},v_{\max}],\ \ y=y_{\max}\}\end{aligned} \tag{145}$$

Here we assume that any points where y meets v belong to boundaries associated with y.

The PDE (142) shows a 2-D dynamic problem. To solve this problem, a semi-discretization in time is applied, which yields a series of 2-D boundary value problems. These boundary value problems are then solved numerically using 2-D finite element method as time advances.

3.2. Temporal Discretization

It should be noted that the time advances *backwards*, such that the initial value is given at the terminal $t_n=T$ and the solution is sought at $t_0=0$. The partial derivative with respect to time is approximated by finite difference method. According to (142), the z-weighted finite difference scheme for time is given as

$$\frac{\widetilde{w}-w}{t_{k+1}-t_k}+(1-z)(\boldsymbol{\nabla}\cdot\mathbf{A}\boldsymbol{\nabla}-\mathbf{b}\cdot\boldsymbol{\nabla}-r)\widetilde{w}+z(\boldsymbol{\nabla}\cdot\mathbf{A}\boldsymbol{\nabla}-\mathbf{b}\cdot\boldsymbol{\nabla}-r)w=0 \tag{146}$$

where we denote $w=w(t_k)$ and $\widetilde{w}=w(t_{k+1})$. The quantity associated with time-step t_{k+1} will be denoted with an extra "~" accent sign. In the above equation, the scheme becomes purely explicit when $z=0$, purely implicit when $z=1$, and becomes Crank-Nicolson scheme when $z=0.5$. In practice, Crank-Nicolson scheme is often in favor due to its superior second order convergence. However, it is also well known that the Crank-Nicolson scheme may exhibit localized oscillations for discontinuous terminal conditions if the time step is too coarse relative to the spatial step. A remedy proposed by Rannacher is to take two fully implicit time steps ($z=1$) before we switch to Crank-Nicolson ($z=0.5$) time-stepping. This solution is also known as Rannacher time-stepping [29], which will be used in our implementation.

The computation starts from $t_n=T$ and advances backwards. At maturity, the $w(T)$ is known and given by terminal condition. At time-step t_k, the $\widetilde{w}$ is already known and hence the w can derived from solving (146). The whole process is repeated until $t=0$ to get the final solution. Rearrangement of (146) gives a new equation, in which we collect all the $\widetilde{w}$ terms on the right hand side and all the w terms on the left

$$\begin{aligned}&(z\boldsymbol{\nabla}\cdot\mathbf{A}\boldsymbol{\nabla}-z\mathbf{b}\cdot\boldsymbol{\nabla}-c)w=\ (\tilde{z}\boldsymbol{\nabla}\cdot\mathbf{A}\boldsymbol{\nabla}-\tilde{z}\mathbf{b}\cdot\boldsymbol{\nabla}-\tilde{c})\widetilde{w}\\&\text{where}\qquad c=zr+\frac{1}{t_{k+1}-t_k},\qquad \tilde{z}=z-1,\qquad \tilde{c}=c-r\end{aligned} \tag{147}$$

3.3. Finite Element Method in 2D

Finite element method is based on weak formulation. It is ideal for solving PDE when the solution lacks smoothness and when the domain is irregular, dynamically changing, and/or unevenly spaced. Valuation of complex exotic options can often exhibit these properties, which makes FEM an ideal tool for this type of applications.

3.3.1. Weak Formulation

let's first define a few functional spaces for the solution w and the test function ψ

$$L^2(\Omega) = \left\{ \text{ƒ}: \Omega \to \mathbb{R}: \int_\Omega \text{ƒ}^2 < \infty \right\}$$
$$H^1(\Omega) = \{\text{ƒ} \in L^2(\Omega) : \mathcal{D}^1\text{ƒ} \in L^2(\Omega)\}, \qquad H_0^1(\Omega) = \{\text{ƒ} \in H^1(\Omega) : \text{ƒ} = 0 \text{ on } \partial\Omega\} \tag{148}$$

where Ω is the open domain given in (144), $\mathcal{D}^1\text{ƒ}$ denotes the first order weak partial derivatives of function ƒ. The $L^2(\Omega)$ is the Lebesgue space with Euclidean norm (which coincides with Hilbert space here). The $H^1(\Omega)$ and $H_0^1(\Omega)$ are Sobolev space. The (147) can be transformed to a weak formulation by multiplying both sides with a scalar-valued test function $\psi \in H^1(\Omega)$(which in turn can be constructed as a linear combination of basis functions on the 2-D domain). That is, find $w \in H^1(\Omega)$ such that

$$\int_\Omega \psi(z\boldsymbol{\nabla} \cdot \mathbf{A}\boldsymbol{\nabla} - z\mathbf{b} \cdot \boldsymbol{\nabla} - c)w = \int_\Omega \psi(\tilde{z}\boldsymbol{\nabla} \cdot \mathbf{A}\boldsymbol{\nabla} - \tilde{z}\mathbf{b} \cdot \boldsymbol{\nabla} - \tilde{c})\tilde{w} \tag{149}$$

For all $\psi \in H^1(\Omega)$. The weak form can be derived using (first) Green's identity, which is the multidimensional analogue of integration by parts. Assuming that u is a scalar function and $\mathbf{v}$ a vector-valued function, both are continuously differentiable, the integration by parts in multi-dimension follows

$$\int_\Omega u\boldsymbol{\nabla} \cdot \mathbf{v} = \int_{\partial\Omega} u\mathbf{v} \cdot \mathbf{n} - \int_\Omega \boldsymbol{\nabla} u \cdot \mathbf{v} \tag{150}$$

where $\mathbf{n}$ denotes the outward unit surface normal to $\partial\Omega$. Applying the transformation (150) to both sides of (149) we have

$$\begin{aligned} &\int_\Omega \psi(z\boldsymbol{\nabla} \cdot \mathbf{A}\boldsymbol{\nabla} - z\mathbf{b} \cdot \boldsymbol{\nabla} - c)w = z\int_{\partial\Omega} \psi\mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} - \int_\Omega \mathcal{R}w \qquad \text{and} \\ &\int_\Omega \psi(\tilde{z}\boldsymbol{\nabla} \cdot \mathbf{A}\boldsymbol{\nabla} - \tilde{z}\mathbf{b} \cdot \boldsymbol{\nabla} - \tilde{c})\tilde{w} = \tilde{z}\int_{\partial\Omega} \psi\mathbf{A}\boldsymbol{\nabla}\tilde{w} \cdot \mathbf{n} - \int_\Omega \tilde{\mathcal{R}}\tilde{w} \end{aligned} \tag{151}$$

where we define two more operators

$$\mathcal{R} = z(\boldsymbol{\nabla}\psi \cdot \mathbf{A}\boldsymbol{\nabla} + \psi\mathbf{b} \cdot \boldsymbol{\nabla}) + c\psi, \qquad \tilde{\mathcal{R}} = \tilde{z}(\boldsymbol{\nabla}\psi \cdot \mathbf{A}\boldsymbol{\nabla} + \psi\mathbf{b} \cdot \boldsymbol{\nabla}) + \tilde{c}\psi \tag{152}$$

The (149) eventually becomes

$$\int_\Omega \mathcal{R}w - z\int_{\partial\Omega} \psi\mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} = \int_\Omega \tilde{\mathcal{R}}\tilde{w} - \tilde{z}\int_{\partial\Omega} \psi\mathbf{A}\boldsymbol{\nabla}\tilde{w} \cdot \mathbf{n} \tag{153}$$

Note that both $\mathcal{R}$ and $\tilde{\mathcal{R}}$ operators can be time-dependent. The (153) is also called the weak form (or variational form) of the PDE (142). This is because the requirements of the solution w in (153) have been considerably weakened over the strong form in (142). In the weak formulation of the problem, the solution w and the test function ψ must belong to $H^1(\Omega)$, however it is not necessary that all functions and derivatives be continuous.

3.3.2. Terminal Conditions

The terminal condition is defined at option maturity by payoff function of a product. For instance, a vanilla call $(\eta = 1)$ or put $(\eta = -1)$ that matures in $\tau = T - t$ would have a terminal condition defined by the payoff function as

$$w(T, v_T, y_T) = \left(\eta(\exp(y_T) - K)\right)^+ \tag{154}$$

3.3.3. Boundary Conditions

The boundary conditions, on the other hand, define how a PDE should behave at domain boundaries. The boundary integral in (153)

$$\int_{\partial\Omega} \psi \mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} \tag{155}$$

must be correctly handled to be consistent with the boundary conditions. In the following, we will introduce four types of typical boundary conditions.

3.3.3.1. *Homogeneous Dirichlet Boundary Condition*

Homogeneous Dirichlet boundary condition requires the function value at domain boundary be zero, i.e. $w = 0$ on $\partial\Omega$. Under this condition, we have $w \in H_0^1(\Omega)$ and $\psi \in H_0^1(\Omega)$. The boundary integral (155) in (153) is zero as test function ψ is zero on $\partial\Omega$. The weak form of the PDE (153) is to find $w \in H_0^1(\Omega)$, such that for all $\psi \in H_0^1(\Omega)$ we have [30]

$$\int_\Omega \mathcal{R} w = \int_\Omega \tilde{\mathcal{R}} \tilde{w}, \qquad w, \psi \in H_0^1(\Omega) \tag{156}$$

where the operators $\mathcal{R}$ and $\tilde{\mathcal{R}}$ are defined in (152).

3.3.3.2. *Homogeneous Neumann Boundary Condition*

Homogeneous Neumann boundary condition requires the function derivative at domain boundary be zero, i.e. $\mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} = 0$ on $\partial\Omega$. Under this condition, we have $w \in H^1(\Omega)$ and $\psi \in H^1(\Omega)$. The boundary integral (155) vanishes because $\mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} = 0$ on $\partial\Omega$. The weak form of the PDE (153) is to find $w \in H^1(\Omega)$, such that for all $\psi \in H^1(\Omega)$ we have [31]

$$\int_\Omega \mathcal{R} w = \int_\Omega \tilde{\mathcal{R}} \tilde{w}\,, \qquad w, \psi \in H^1(\Omega) \tag{157}$$

Though (157) looks almost identical to (156), the function spaces for the solution and the test functions are different. In homogeneous Dirichlet condition, it is the test function $\psi = 0$ on $\partial\Omega$ that causes the boundary integral to vanish, while in homogeneous Neumann condition, it is the function derivative $\mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} = 0$ on $\partial\Omega$ that erases the boundary integral.

3.3.3.3. *Inhomogeneous Dirichlet Boundary Condition*

Inhomogeneous Dirichlet boundary condition requires the solution $w = g$ on $\partial\Omega$ where g is a function defined on $\partial\Omega$. The function g must satisfy some regularity conditions, and we will assume that there exists a function $l \in \{u \in H^1(\Omega) : u = g \text{ on } \partial\Omega\}$. It turns out that the correct space of test functions is still $H_0^1(\Omega)$, i.e. $\psi \in H_0^1(\Omega)$, the same as in homogeneous Dirichlet condition. However, since the desired solution w does not satisfy homogeneous Dirichlet condition, this means that $w \notin H_0^1(\Omega)$. Instead, the function $u = w - l$ for $u \in H_0^1(\Omega)$ is zero on $\partial\Omega$, and hence the solution has the form $w = u + l$ where l is assumed to be known and u is unknown. The weak form of the PDE (153) is then to find $w = u + l$ for $u \in H_0^1(\Omega)$, such that for all $\psi \in H_0^1(\Omega)$ we have [32]

$$\int_\Omega \mathcal{R}(u + l) = \int_\Omega \tilde{\mathcal{R}} \tilde{w}\,, \qquad u, \psi \in H_0^1(\Omega) \tag{158}$$

In Section 3.3.6, we will show how to find a suitable function $l \in \{\not{w} \in H^1(\Omega) : \not{w} = g \text{ on } \partial\Omega\}$ for the purpose of finite element method that satisfies Dirichlet condition $l = g$ on $\partial\Omega$ (note that in a boundary value problem, only g is given, not l).

3.3.3.4. *Inhomogeneous Neumann Boundary Condition*

Inhomogeneous Neumann boundary condition requires the function derivative at domain boundary to be a known function, i.e. $\mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} = h$ on $\partial\Omega$. Under this condition, we have $w \in H^1(\Omega)$ and $\psi \in H^1(\Omega)$. The weak form of the PDE (153) is to find $w \in H^1(\Omega)$, such that for all $\psi \in H^1(\Omega)$ we have [33]

$$\int_\Omega \mathcal{R} w - z \int_{\partial\Omega} \psi h = \int_\Omega \tilde{\mathcal{R}} \tilde{w} - \tilde{z} \int_{\partial\Omega} \psi \tilde{h} \tag{159}$$

The weak form is the same as for the homogeneous Neumann condition, except for the extra boundary integral terms $z \int_{\partial\Omega} \psi h$ and $\tilde{z} \int_{\partial\Omega} \psi \tilde{h}$.

Boundaries involve a combination of these boundary conditions can be treated in a similar manner. Since Dirichlet conditions must be explicitly imposed in the weak form while Neumann conditions are rather implied, Dirichlet conditions are often called essential boundary conditions while Neumann conditions are instead called natural boundary conditions.

3.3.4. Boundary Conditions for Vanilla Options

For a vanilla call option, we employ a rectangle domain consisting of 4 pieces of boundaries, $\partial\Omega = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$, as defined in (145). On Γ_1 where $v = v_{\min}$, we choose $v_{\min} = 0$ and the boundary integral (155) becomes zero as the $\mathbf{A}$ term varnishes. On Γ_2 where $v = v_{\max}$, we may assume inhomogeneous Neumann boundary condition with derivative $\partial w(t, v_{\max}, y)/\partial v$ approximated in Black-Scholes model (equivalent to the vega sensitivity). We estimate the bound $v_{\max}$ using the conditional mean and variance of v given in (47), such that

$$v_{\max} = \mathbb{E}[v_{t+\tau}|v_t] + n\sqrt{\mathbb{V}[v_{t+\tau}|v_t]} \tag{160}$$

where we set $n = 8$ to make the $v_{\max}$ sufficiently large. On Γ_3 where $y = y_{\min}$, an inhomogeneous Dirichlet boundary condition is devised with boundary values approximated again in Black-Scholes model. On Γ_4 where $y = y_{\max}$, we manage to implement an inhomogeneous Neumann boundary

condition with derivative $\partial w(t,v,y_{\max})/\partial y$ approximated in Black-Scholes model (equivalent to the delta sensitivity). The $y_{\min}$ and $y_{\max}$ are also determined by the conditional mean of v, such that

$$y_{\min} = \ln F_{t,t+\tau} - m\sqrt{\mathbb{E}[v_{t+\tau}|v_t]}, \qquad y_{\max} = \ln F_{t,t+\tau} + m\sqrt{\mathbb{E}[v_{t+\tau}|v_t]} \tag{161}$$

where we set $m = 4$ to make both $y_{\min}$ and $y_{\max}$ sufficiently far away from the spot. These boundary conditions are summarized below

$$\begin{aligned}
&\text{For } (v,y) \in \Gamma_1, \qquad v_{\min} = 0 \Longrightarrow \mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} = 0 \\
&\text{For } (v,y) \in \Gamma_2, \qquad \boldsymbol{\nabla} w(t, v_{\max}, y) = \begin{pmatrix} \partial w/\partial v \\ 0 \end{pmatrix} \Longrightarrow \mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} = h_v(t,v,y) \\
&\text{For } (v,y) \in \Gamma_3, \qquad w(t,v,y_{\min}) = g(t,v,y) \\
&\text{For } (v,y) \in \Gamma_4, \qquad \boldsymbol{\nabla} w(t,v,y_{\max}) = \begin{pmatrix} 0 \\ \partial w/\partial y \end{pmatrix} \Longrightarrow \mathbf{A}\boldsymbol{\nabla} w \cdot \mathbf{n} = h_y(t,v,y)
\end{aligned} \tag{162}$$

where the Dirichlet condition comes from the Black-Scholes call value

$$g(t,v,y) = \exp(y - q\tau)\,\Phi(d_+) - K\exp(-r\tau)\,\Phi(d_-), \qquad d_\pm = \frac{y - \ln K + (r+q)\tau}{\sqrt{v\tau}} \pm \frac{\sqrt{v\tau}}{2} \tag{163}$$

The Neumann conditions are further derived as follows

$$\begin{aligned}
h_v(t,y,v) &= \left(\frac{v}{2} \begin{pmatrix} \xi^2 & \rho\xi \\ \rho\xi & 1 \end{pmatrix} \begin{pmatrix} \partial w/\partial v \\ 0 \end{pmatrix} \right) \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{v\xi^2}{2}\frac{\partial w}{\partial v} \\
h_y(t,v,y) &= \left(\frac{v}{2} \begin{pmatrix} \xi^2 & \rho\xi \\ \rho\xi & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \partial w/\partial y \end{pmatrix} \right) \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{v}{2}\frac{\partial w}{\partial y}
\end{aligned} \tag{164}$$

where the derivatives are also stemmed from the Black-Scholes call value

$$\begin{aligned}
\frac{\partial w}{\partial v} &= \frac{-\exp(y - q\tau)\,\phi(d_+)d_- + K\exp(-r\tau)\,\phi(d_-)d_+}{2v} = \exp(y - q\tau)\,\phi(d_+)\frac{d_+ - d_-}{2v} \\
&= \frac{\exp(y - q\tau)\,\phi(d_+)\sqrt{\tau}}{2\sqrt{v}} \\
\frac{\partial w}{\partial y} &= \exp(y - q\tau)\,\Phi(d_+) + \frac{\exp(y - q\tau)\,\phi(d_+)}{\sqrt{v\tau}} - \frac{K\exp(-r\tau)\,\phi(d_-)}{\sqrt{v\tau}} \\
&= \exp(y - q\tau)\,\Phi(d_+)
\end{aligned} \tag{165}$$

using the fact that

$$\exp(y-q\tau)\,\phi(d_+) = K\exp(-r\tau)\,\phi(d_-), \qquad \frac{\partial d_\pm}{\partial v} = -\frac{d_\mp}{2v}, \qquad \frac{\partial d_\pm}{\partial y} = \frac{1}{\sqrt{v\tau}} \tag{166}$$

More adaptive boundary conditions may also be used. However, in the case of vanilla options, the choice of boundary conditions appears immaterial when the bounded domain becomes sufficiently large. Knowing these boundary conditions, we can define a function space

$$V = \{\not{p} \in H^1(\Omega) : \not{p} = 0 \text{ on } \Gamma_3\} \tag{167}$$

The weak form of the PDE (153) is to find $w = u + l$ for $u \in V$ and $l \in \{\not{p} \in H^1(\Omega) : \not{p} = g \text{ on } \Gamma_3\}$, such that for all $\psi \in V$ we have

$$\int_\Omega \mathcal{R}(u+l) - z\int_{\Gamma_2}\psi h_v - z\int_{\Gamma_4}\psi h_y = \int_\Omega \tilde{\mathcal{R}}\tilde{w} - \tilde{z}\int_{\Gamma_2}\psi\tilde{h}_v - \tilde{z}\int_{\Gamma_4}\psi\tilde{h}_y \tag{168}$$

In the case of a vanilla put, similar boundary conditions can be employed

$$\begin{aligned}
&\text{For } (v,y)\in\Gamma_1, && v_{\min} = 0 \Longrightarrow \mathbf{A}\boldsymbol{\nabla} w\cdot\mathbf{n} = 0\\
&\text{For } (v,y)\in\Gamma_2, && \boldsymbol{\nabla} w(t,v_{\max},y) = \begin{pmatrix}\partial w/\partial v\\ 0\end{pmatrix} \Longrightarrow \mathbf{A}\boldsymbol{\nabla} w\cdot\mathbf{n} = h_v(t,v,y)\\
&\text{For } (v,y)\in\Gamma_3, && \boldsymbol{\nabla} w(t,v,y_{\min}) = \begin{pmatrix}0\\ \partial w/\partial y\end{pmatrix} \Longrightarrow \mathbf{A}\boldsymbol{\nabla} w\cdot\mathbf{n} = h_y(t,v,y)\\
&\text{For } (v,y)\in\Gamma_4, && w(t,v,y_{\max}) = g(t,v,y)
\end{aligned} \tag{169}$$

where the functions are given below

$$g(t,y,v) = K\exp(-r\tau)\,\Phi(-d_-) - \exp(y-q\tau)\,\Phi(-d_+)$$

$$h_v(t,y,v) = \left(\frac{v}{2}\begin{pmatrix}\xi^2 & \rho\xi\\ \rho\xi & 1\end{pmatrix}\begin{pmatrix}\partial w/\partial v\\ 0\end{pmatrix}\right)\cdot\begin{pmatrix}1\\0\end{pmatrix} = \frac{v\xi^2}{2}\frac{\partial w}{\partial v}, \qquad \frac{\partial w}{\partial v} = \frac{\exp(y-q\tau)\,\phi(d_+)\sqrt{\tau}}{2\sqrt{v}} \tag{170}$$

$$h_y(t,v,y) = \left(\frac{v}{2}\begin{pmatrix}\xi^2 & \rho\xi\\ \rho\xi & 1\end{pmatrix}\begin{pmatrix}0\\ \partial w/\partial v\end{pmatrix}\right)\cdot\begin{pmatrix}0\\-1\end{pmatrix} = -\frac{v}{2}\frac{\partial w}{\partial y}, \qquad \frac{\partial w}{\partial y} = -\exp(y-q\tau)\,\Phi(-d_+)$$

The resulted weak form of the PDE (153) is similar to the one we have derived for a vanilla call, which is to find $w = u + l$ for $u \in V = \{\not{p} \in H^1(\Omega) : \not{p} = 0 \text{ on } \Gamma_4\}$ and $l \in \{\not{p} \in H^1(\Omega) : \not{p} = g \text{ on } \Gamma_4\}$, such that for all $\psi \in V$ we have

$$\int_{\Omega} \mathcal{R}(u+l) - z\int_{\Gamma_2} \psi h_v - z\int_{\Gamma_3} \psi h_y = \int_{\Omega} \tilde{\mathcal{R}}\tilde{w} - \tilde{z}\int_{\Gamma_2} \psi \tilde{h}_v - \tilde{z}\int_{\Gamma_3} \psi \tilde{h}_y \tag{171}$$

We would like to demonstrate the algorithm using vanilla call as an example. In order to find numerical solution, the continuous operator problem (168) must be converted to a discrete one. We will use Galerkin method [34] [35] to perform the conversion, which computes the best approximation to the true solution from a given finite-dimensional subspace.

3.3.5. Mesh and Basis Functions

The finite element method is a general and systematic technique for constructing basis functions of the finite-dimensional subspace for Galerkin approximations on a mesh that discretizes the continuous domain into a union of discrete geometric cells. For our problem, we construct a triangular mesh (Figure 1), which can be either structured or unstructured. The intersection of any two triangles in the mesh must be a common vertex or a common edge. For example, we denote the closed domain $\bar{\Omega} = \Omega \cup \partial\Omega = [v_{\min}, v_{\max}] \times [y_{\min}, y_{\max}]$. The $\bar{\Omega}$ is discretized into a triangulation mesh $\mathcal{T}_h$ consisting of triangles $T_i,\ i = 1, \cdots, N_t$ as the elements and their nodes $D_i,\ i = 1, \cdots, N_d$. The triangulation $\mathcal{T}_h$ is labelled with mesh size h, which is the maximum diameter of any triangle in the triangulation.

The mesh generator keeps two arrays: a list of nodes and a list of triangles (i.e. elements). In addition, it also groups boundary nodes into four node sets, one for each boundary (we assume that any points where v and y meet are nodes in the triangulation, and any such nodes belong to boundaries associated with y). Each triangle element is uniquely defined by its three vertex nodes, while each node records relevant information including: index of the node in the node list, coordinates of the node, left and right neighbors if the node lies on a boundary (such that the connections of the node to its neighbors form part of the boundary. This information is used for boundary integral calculation). Figure 1 illustrates a simple evenly spaced structured mesh used in the computation. It can be seen that the mesh consists of total $I \times J$ squares, $N_t = 2IJ$ triangles and $N_d = (I+1)(J+1)$ nodes, where I and J are the spatial resolution for variable v and y, respectively.

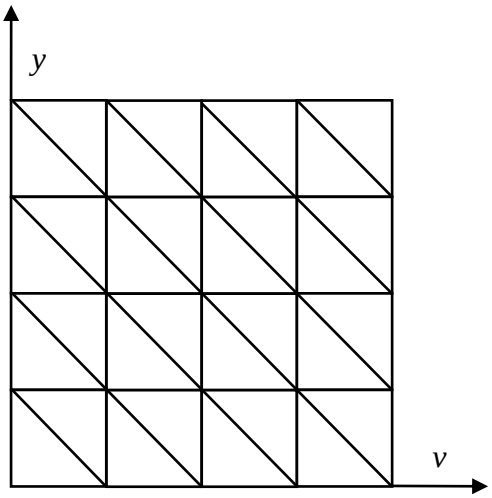

Figure 1. A structured triangular mesh

The basis functions for the finite dimensional subspace on the 2-D mesh are defined as piecewise linear functions, $\phi_i,\ i = 1, \cdots, N_d$, such that they take value 1 at node D_i and 0 at all other nodes. This basis is usually called Lagrange basis or nodal basis. To allow generic computations over any arbitrary triangles, each triangle is mapped by affine transformation to a reference triangle $\hat{T} = \{(\hat{v}, \hat{y}) \in \mathbb{R}^2 : \hat{v} \in [0,1],\ \hat{y} \in [0, 1-\hat{v}]\}$, where a "hat" accent is used to denote quantities defined on the reference triangle [36]. The reference triangle is shown in Figure 2. The three vertices $\widehat{D}_i, i = 1,2,3$ are $(0,0)$, $(1,0)$ and $(0,1)$, respectively. The open circle denotes $\hat{\phi}_i = 0$ at node $\widehat{D}_j$ if $i \neq j$, while the full circle denotes $\hat{\phi}_i = 1$ at node $\widehat{D}_j$ if $i = j$. In summary, the three basis functions on the reference triangle read

$$\hat{\phi}_0(\hat{v}, \hat{y}) = 1 - \hat{v} - \hat{y}, \qquad \hat{\phi}_1(\hat{v}, \hat{y}) = \hat{v}, \qquad \hat{\phi}_2(\hat{v}, \hat{y}) = \hat{y} \tag{172}$$

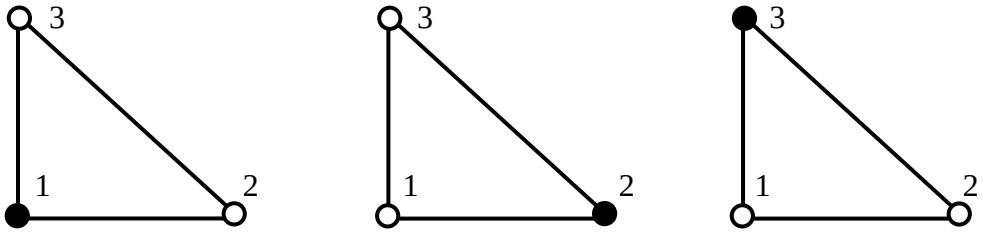

Figure 2. The reference triangle and the basis functions

The mapping from an arbitrary triangle to a reference triangle is achieved via affine transformation. Considering an arbitrary triangle T with vertices $D_0 = (v_0, y_0)$, $D_1 = (v_1, y_1)$ and $D_2 = (v_2, y_2)$, the transformation maps the reference triangle $\hat{T}$ onto the triangle T through

$$(v,y) = F(\hat{v},\hat{y}) = \mathbf{J}\begin{pmatrix}\hat{v}\\ \hat{y}\end{pmatrix} + \begin{pmatrix}v_0\\ y_0\end{pmatrix}, \qquad \mathbf{J} = \frac{\partial F(\hat{v},\hat{y})}{\partial(\hat{v},\hat{y})} = \begin{pmatrix}v_1 - v_0 & v_2 - v_0\\ y_1 - y_0 & y_2 - y_0\end{pmatrix} \tag{173}$$

where **J** is the Jacobian matrix. Denoting the function defined on the reference triangle with a "hat", we have $f(v,y) = \hat{f}(\hat{v},\hat{y})$ whenever $(v,y) = F(\hat{v},\hat{y})$. In particular, the following three equations hold

$$\hat{f}(0,0) = f(v_0,y_0), \qquad \hat{f}(1,0) = f(v_1,y_1), \qquad \hat{f}(0,1) = f(v_2,y_2) \tag{174}$$

Additionally, the integration of a function $f(v,y)$ over an arbitrary triangle T can be transformed into an equivalent integration over the reference triangle $\hat{T}$ via the integral of a function

$$\int_T f(v,y)dvdy = \int_{\hat{T}} f\big(F(\hat{v},\hat{y})\big)\left|\frac{\partial F(\hat{v},\hat{y})}{\partial(\hat{v},\hat{y})}\right| d\hat{v}d\hat{y} = \mathcal{J}\int_{\hat{T}} \hat{f}(\hat{v},\hat{y})d\hat{v}d\hat{y}, \qquad \mathcal{J} = |\det \mathbf{J}| \tag{175}$$

where $\mathcal{J}$ is the absolute of Jacobian determinant of **J**, also known as *Jacobian factor*. The affine transformation greatly simplifies the calculation and renders the use of unstructured mesh possible. In our application, we use a mesh grid that is non-uniformly spaced for improved numerical accuracy with denser grid points around key values, e.g. spot, strike, initial variance, etc.

3.3.6. Stiffness Matrix and Load Factor

Given the boundary conditions (162), we now can divide the mesh nodes (denoted by set N_d; a bit abuse of notation here) into two groups, the nodes lying on $\Gamma_1 \cup \Gamma_2 \cup \Gamma_4$ or in Ω that are called free nodes (denoted by set N_f) and the nodes on Γ_3 that are called constrained nodes (denoted by set N_c). With the piecewise linear basis functions in Section 3.3.5, the function space P_h of all continuous piecewise linear functions defined on the mesh $\mathcal{T}_h$ is a finite dimensional vector space with dimension N_d, having basis functions $\{\phi_1, \cdots, \phi_{N_d}\}$. Each function $\mathfrak{f} \in P_h$ can be identified as a linear combination of the basis functions through formula

$$\mathfrak{f} = \sum_{i \in N_d} \varphi_i \phi_i \tag{176}$$

with a vector $\varphi \in \mathbb{R}^{N_d}$ consisting of the nodal values of $\mathfrak{f}$.

Using Galerkin method, the weak form (168) is to find $w = u + l$ for $u \in V_h = \{\phi \in P_h : \phi = 0 \text{ on } \Gamma_3\}$ and $l \in \{\phi \in P_h : \phi = g \text{ on } \Gamma_3\}$, such that for all $\psi \in V_h$ we have

$$a(u, \psi) = \ell(\psi) \tag{177}$$

where the bilinear functional $a(\cdot,\cdot)$ and linear functional $\ell(\cdot)$ are taken from (168) as

$$a(u, \psi) = \int_\Omega \mathcal{R}u\,, \qquad \ell(\psi) = \int_\Omega \tilde{\mathcal{R}}\tilde{w} - \tilde{z}\int_{\Gamma_2} \psi\tilde{h}_v - \tilde{z}\int_{\Gamma_4} \psi\tilde{h}_y + z\int_{\Gamma_2} \psi h_v + z\int_{\Gamma_4} \psi h_y - \int_\Omega \mathcal{R}l \tag{178}$$

We may write the solution and the test function as linear combinations of basis functions

$$u = \sum_{i \in N_f} \mu_i \phi_i\,, \qquad \psi = \sum_{i \in N_f} c_i \phi_i \tag{179}$$

For the inhomogeneous Dirichlet boundary condition on Γ_3, the function l is any function in P_h that satisfies the boundary condition $l = g$ on Γ_3. It may not be easy to find such l function exactly, but it is easy to define a function l that approximately satisfies the boundary condition. Indeed, the continuous piecewise linear function l defined by

$$l = \sum_{i \in N_c} \lambda_i \phi_i \tag{180}$$

where λ_i is the value of function g evaluated at node i, agrees with g at the endpoints of every constrained edge and therefore interpolates g on Γ_3. It is a sufficiently good approximation to g for the purposes of the finite element method [37].

Using linearity properties of bilinear functional and linear functional we can derive

$$\begin{aligned} & a(u, \psi) = \ell(\psi) \Longrightarrow a\left(\sum_{j \in N_f} \mu_j \phi_j\,, \sum_{j \in N_f} c_i \phi_i\right) = \ell\left(\sum_{i \in N_f} c_i \phi_i\right) \\ & \Longrightarrow \sum_{i \in N_f} c_i \sum_{j \in N_f} \mu_j a(\phi_j, \phi_i) = \sum_{i \in N_f} c_i \ell(\phi_i) \\ & \Longrightarrow \sum_{i \in N_f} c_i \left(\sum_{j \in N_f} R_{ij}\mu_j - f_i\right) = 0, \qquad R_{ij} = a(\phi_j, \phi_i), \qquad f_i = \ell(\phi_i) \end{aligned} \tag{181}$$

Since c_i are arbitrary, it is certainly sufficient to satisfy

$$\sum_{j \in N_f} R_{ij}\mu_j - f_i = 0 \tag{182}$$

or write it in vector-matrix form

$$R\mu = f \tag{183}$$

where R is called the stiffness matrix and f the load factor. Both are known and can be calculated directly. The $\mu \in \mathbb{R}^{N_f}$ is a vector that can be solved from inverting the stiffness matrix. It consists of the nodal values of the approximate solution in the finite-dimensional subspace.

In order to simplify the implementation, we transform (183) into an equivalent linear system, such that the solution vector consists of nodal values of the whole domain (rather than just free nodes) including the nodes on Dirichlet boundaries. We first define matrix Q, vectors λ and α as

$$\begin{aligned} &\underset{N_f \times N_f}{R}\ \underset{N_f \times 1}{\mu} = \int_\Omega \mathcal{R}u\,, \qquad \underset{N_f \times N_c}{Q}\ \underset{N_c \times 1}{\lambda} = \int_\Omega \mathcal{R}l \\ &\underset{N_f \times 1}{\alpha} = \int_\Omega \tilde{\mathcal{R}}\tilde{w} - \tilde{z}\int_{\Gamma_2} \psi\tilde{h}_v - \tilde{z}\int_{\Gamma_4} \psi\tilde{h}_y + z\int_{\Gamma_2} \psi h_v + z\int_{\Gamma_4} \psi h_y \end{aligned} \tag{184}$$

where the $R\mu$ is the same as in (183) and $\alpha - Q\lambda = f$. With these definitions, the new linear system, which is equivalent to (183), would be

$$S\omega = \beta, \qquad \underset{N_d \times N_d}{S} = \begin{pmatrix} R & Q \\ 0 & I \end{pmatrix}, \qquad \underset{N_d \times 1}{\omega} = \begin{pmatrix} \mu \\ \lambda \end{pmatrix}, \qquad \underset{N_d \times 1}{\beta} = \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \tag{185}$$

where the I is $N_c \times N_c$ identity matrix and the ω vector consists of nodal values of the whole domain including all the boundary nodes. We can see that the sub-vector λ is known and given by the Dirichlet boundary conditions. The sub-matrices R and Q and the sub-vector α can be evaluated through the integrals in (184), which are then used to construct the matrix S and vector β. We solve the linear system $S\omega = \beta$ for the vector ω, which consists of the coefficients of the basis functions, provides an approximate solution to the continuous function w at time-step t_k. We repeat the process backwards

from maturity until $t = 0$, where we can then extract the call option value in the model from the final solution vector.

3.3.6.1. *Computation of Integrals*

From previous discussion, we know that in order to obtain the matrix H and the vector z, we must calculate the integrals in (184). For example, the R_{ij}, as shown in (181), can be calculated as

$$R_{ij} = a(\phi_j, \phi_i) = z\int_{\Omega} \nabla\phi_i \cdot \mathbf{A}\nabla\phi_j + z\int_{\Omega} \phi_i \mathbf{b} \cdot \nabla\phi_j + c\int_{\Omega} \phi_i\phi_j\,, \qquad \forall\ i,j \in N_f \tag{186}$$

We must evaluate three integrals over the whole triangular mesh. However, because of the local support of the basis functions, the R_{ij} is evaluated to zero unless the node i and j are vertices of the same triangle of the mesh. Instead of computing the integral over the whole domain Ω directly, we calculate the integrals over each triangular element and then assemble over all the triangle elements of the domain. Given an arbitrary triangular element T, we can see that there are only three basis functions (corresponding to the three nodes/vertices) are nonzero. There will be 9 combinations of the bilinear form to be evaluated on each element. To make the implementation simpler and yet more adaptive, the integrals on an arbitrary triangle T can be calculated through the mapping of the reference triangle $\hat{T}$. That is, we calculate the integrals on a reference triangle $\hat{T}$, which can be standardized, and the results are then linearly transformed back to the arbitrary triangle T using the aforesaid affine transformation.

Accounting for the 9 combinations of the bilinear form among the 3 basis functions, The first integral denoted by I_1 appearing in the right hand side of (186) is a 3×3 matrix. Over an arbitrary triangle T, the integral is evaluated as

$$I_1 = \left[\int_T \nabla\phi_i \cdot \mathbf{A}\nabla\phi_j\right]_{i,j=0,1,2} = \left[\int_{\hat{T}} \left(\mathbf{J}^{-T}\nabla\hat{\phi}_i\right)^T \left(\mathbf{A}\mathbf{J}^{-T}\nabla\hat{\phi}_j\right)\left|\frac{\partial F(\hat{v},\hat{y})}{\partial(\hat{v},\hat{y})}\right|\right]_{i,j=0,1,2}$$

$$= \mathcal{J}\left[\int_{\hat{T}} \left(\mathbf{J}^{-T}\nabla\hat{\phi}_i\right)^T \left(\mathbf{A}\mathbf{J}^{-T}\nabla\hat{\phi}_j\right)\right]_{i,j=0,1,2} = \mathcal{J}\left(\mathbf{J}^{-T}\widehat{\mathbf{G}}\right)^T \left(\int_{\hat{T}} \mathbf{A}\right)\mathbf{J}^{-T}\widehat{\mathbf{G}} \approx \frac{1}{2}\mathcal{J}\widehat{\mathbf{G}}^T\mathbf{J}^{-1}\bar{\mathbf{A}}\mathbf{J}^{-T}\widehat{\mathbf{G}} \tag{187}$$

where $\widehat{\mathbf{G}} = \left[\nabla\hat{\phi}_j\right]_{j=0,1,2} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$ and

$$\int_{\hat{T}} \mathbf{A} \approx \bar{\mathbf{A}} \int_{\hat{T}} d\hat{T} = \begin{pmatrix} \text{average of } \mathbf{A} \text{ over} \\ \text{triangular element} \end{pmatrix} \times \begin{pmatrix} \text{area of} \\ \text{reference triangle} \end{pmatrix} = \frac{\sum_{n=0}^{2} \mathbf{A}_n}{3} \frac{1}{2}$$

In (187), the subscript i and j denote the matrix row and column index respectively. The $\mathcal{J}$ is the Jacobian factor defined in (175). The $\widehat{\mathbf{G}}$ is the matrix of the gradients of the three basis functions (172) on reference triangle $\hat{T}$. Since $\widehat{\mathbf{G}}$ is constant, it can move out of the integral. The $\mathbf{A}_n,\ n = 0, 1, 2$ are the matrix $\mathbf{A}$ evaluated at the three vertices of the triangle T. To compute the integral $\int_{\hat{T}} \mathbf{A}$, matrix $\mathbf{A}$ is assumed to be constant within T and takes an average over the three vertices. The second integral I_2 is computed as

$$I_2 = \left[\int_T \phi_i \mathbf{b} \cdot \boldsymbol{\nabla} \phi_j\right]_{i,j=0,1,2} \approx \mathcal{J} \left[\int_{\hat{T}} \hat{\phi}_i\right]_{i=0,1,2} \bar{\mathbf{b}}^T \mathbf{J}^{-T} \widehat{\mathbf{G}} = \frac{\mathcal{J}}{6} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \bar{\mathbf{b}}^T \mathbf{J}^{-T} \widehat{\mathbf{G}} \quad \text{and} \quad \bar{\mathbf{b}} = \frac{1}{3} \sum_{n=0}^{2} \mathbf{b}_n \tag{188}$$

In (188), the $\left[\int_{\hat{T}} \hat{\phi}_i\right]_{i=0,1,2}$ can be regarded as the volumes of the triangular pyramids formed by the 3 basis functions integrated over the reference triangle $\hat{T}$. Similar to matrix $\mathbf{A}$, The vector $\mathbf{b}$ is assumed to be constant within a element triangle and takes an average over the three vertices. Hence it can also move out of the integral along with the matrix $\widehat{\mathbf{G}}$. The computation of the third integral I_3 is even simpler

$$I_3 = \left[\int_T \phi_i \phi_j\right]_{i,j=0,1,2} = \mathcal{J} \left[\int_{\hat{T}} \hat{\phi}_i \hat{\phi}_j\right]_{i,j=0,1,2} = \frac{\mathcal{J}}{24} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \tag{189}$$

It can be seen that I_1, I_2 and I_3 are 3×3 matrix, while I_2 appears to be asymmetric. In the process we determine the nine combinations of the i and j vertex for each triangle T. We then map the local vertex indices i and j of the triangle T to the node indices of the mesh. The global stiffness matrix R is then assembled over the mesh $\mathcal{T}_h$ from the integrals computed on each triangle of the mesh. That is, by assuming i and j are node indices of the mesh (e.g. $i, j \in N_f$), we can find the global stiffness matrix

$$R_{ij} = \sum_{T \in \mathcal{T}_h} \left[z \int_T \boldsymbol{\nabla} \phi_i \cdot \mathbf{A} \boldsymbol{\nabla} \phi_j + z \int_T \phi_i \mathbf{b} \cdot \boldsymbol{\nabla} \phi_j + c \int_T \phi_i \phi_j \right]_{i,j}, \qquad \forall\ i, j \in N_f \tag{190}$$

The entries of matrix Q in (184) can be derived in the same manner, except that for Q we have $i \in N_f$ and $j \in N_c$. In our implementation, we actually construct matrix S directly, in a simpler way. We first initiate S as an $N_d \times N_d$ identity matrix and then set its entries as follows

$$S_{ij} = \sum_{T \in \mathcal{T}_h} \left[z \int_T \boldsymbol{\nabla}\phi_i \cdot \mathbf{A}\boldsymbol{\nabla}\phi_j + z \int_T \phi_i \mathbf{b} \cdot \boldsymbol{\nabla}\phi_j + c \int_T \phi_i \phi_j \right]_{i,j}, \qquad \forall\ i \in N_f,\ j \in N_d \tag{191}$$

The computation of vector β in (185) is a bit more complicated. In our implementation, we compute the β in a few steps. We first obtain the vector β as

$$\beta = \underset{N_d \times N_d}{\tilde{S}} \ \underset{N_d \times 1}{\widetilde{\omega}} = \int_\Omega \tilde{\mathcal{R}}\tilde{w} \tag{192}$$

where the vector $\tilde{w}$ is known and the matrix $\tilde{S}$ is constructed in a way very similar to the matrix S. We initiate $\tilde{S}$ as an $N_d \times N_d$ identity matrix and then update its entries by

$$\tilde{S}_{ij} = \sum_{T \in \mathcal{T}_h} \left[\tilde{z} \int_T \boldsymbol{\nabla}\phi_i \cdot \mathbf{A}\boldsymbol{\nabla}\phi_j + \tilde{z} \int_T \phi_i \mathbf{b} \cdot \boldsymbol{\nabla}\phi_j + \tilde{c} \int_T \phi_i \phi_j \right]_{i,j}, \qquad \forall\ i \in N_f,\ j \in N_d \tag{193}$$

The resulted vector β must be updated according to the Dirichlet boundary conditions, that is we assign

$$\beta_i = \lambda_i, \qquad \forall\ i \in N_c \tag{194}$$

Lastly, the boundary integrals resulted from inhomogeneous Neumann boundary conditions must also be taken into account. This is relatively easy on a rectangular domain as the basis functions are just triangle functions at domain boundary. Assuming Neumann condition on boundary Γ with function h, we can estimate the boundary integral as

$$\int_\Gamma \psi h = \gamma, \qquad \gamma_i = \int_\Gamma h_i \phi_i = \frac{1}{2} h_i \delta_i, \qquad \forall\ i \in \Gamma \tag{195}$$

where δ_i is the base of function ϕ_i along Γ. For example, in the case of Γ_2 (where $v = v_{\max}$), we have $\delta_i = y_{i+1} - y_{i-1}$, and the y_{i+1} and y_{i-1} are the y coordinates of the right and left neighbor of the node i along Γ_2. This is equivalent to calculating the area of a triangle determined by vertices h_i, y_{i+1} and y_{i-1}. Once γ is estimated, we can finalize the vector β by taking

$$\beta_i = \beta_i + \gamma_i, \qquad \forall\ i \in \Gamma \tag{196}$$

for each of the inhomogeneous Neumann boundaries.

In our implementation, we assume time-invariant (constant) model parameters for simplicity and hence the matrix R, Q and $\tilde{R}$ remain unchanged through time. In practical application where time-dependent model parameters are assumed, these matrices must be constructed dynamically for each time-step.

3.3.7. Iterative Linear Solvers

Because of local support of basis functions, a basis function can interact with at most, including itself, seven basis functions. The matrix S in (185) can therefore be sparse. To avoid the unnecessary memory storage for the zeros, we use sparse matrix implementation provided by GNU Scientific Library (GSL) [38]. The linear systems resulted from this problem can be quite large. Traditional direct method and simple iterative methods will not work well for this type of applications. Additionally, since the matrix S in (185) is asymmetric (due to asymmetric I_2 in (188)), the conjugate gradient method, an effective method for symmetric positive definite systems, is no longer applicable. Instead, we have implemented conjugate gradient squared (CGS) method [39] and (restarted) generalized minimal residual (GMRES) method [40] to solve the linear systems. Interested readers may refer to Kelley's book [41] for details of the two algorithms. In our application, we will primarily use the GMRES solver with a trivial preconditioner $M = \left(\text{diag}(S)\right)^{-1}$ [42] for its high robustness and efficiency. GSL also provides an implementation of the GMRES solver, but currently there is no preconditioners available [43]. Numerical experiments also show that our implementation of GMRES (with the trivial preconditioner) converges much faster than the GSL counterpart.

3.3.8. Interpolation of Numerical Solution

Numerical solution to the PDE is obtained as an approximate solution vector ω consisting of the nodal values of function w. To extract a solution for an arbitrary point in the domain, interpolation

among the nodal values must be performed. To do this, we convert the point coordinate to a barycentric coordinate in a triangle of the mesh. Providing the coordinate (v, y) of the point, its barycentric coordinates λ_0, λ_1, and λ_2 within a triangle can be determined from the three triangular vertices $\big((v_0, y_0), (v_1, y_1), (v_2, y_2)\big)$, such that

$$\begin{pmatrix}\lambda_0\\ \lambda_1\end{pmatrix} = \begin{pmatrix}v_0 - v_2 & v_1 - v_2\\ y_0 - y_2 & y_1 - y_2\end{pmatrix}^{-1}\begin{pmatrix}v - v_2\\ y - y_2\end{pmatrix}, \qquad \lambda_2 = 1 - \lambda_0 - \lambda_1 \tag{197}$$

If $\lambda_i \in [0,1]\ \forall\ i = 0,1,2$, then the point is inside the triangle and the solution can be interpolated by

$$w(v, y) = \lambda_0 w(v_0, y_0) + \lambda_1 w(v_1, y_1) + \lambda_2 w(v_2, y_2) \tag{198}$$

3.3.9. Numerical Solution of Vanilla Options

We attempt to compute the PV as well as the implied volatility (inverted from PV) of vanilla call and put using the FEM method along with the boundary conditions described above. The model and the FEM PDE parameters used in the experiments are shown in Table 1. The solution (PV) vectors at $t = 0$ are obtained for both the call and put option. They are shown as a surface, a function of initial variance and initial spot, in Figure 3. The PV of the two products can be interpolated from the two surfaces using $X_0 = 100$ and $v_0 = 0.12$.

Table 1. Model and PDE parameters used for vanilla call and put valuation

Heston	Initial Spot Price	X_0	100
	Initial Variance	v_0	0.12
	Mean Variance/Reversion Level	θ	0.10
	Mean Reversion Rate	κ	2.0
	Volatility of Variance	ξ	0.4
	Correlation	ρ	-0.50
	Risk Free Rate/Domestic Rate	r	5%
	Dividend Rate/Foreign Rate	q	3%
	Call/Put Strike	K	100
	Option Expiry	T	1Y
PDE	Variance Resolution	N_v	50
	Log Spot Resolution	N_y	90
	Temporal Resolution	N_t	60

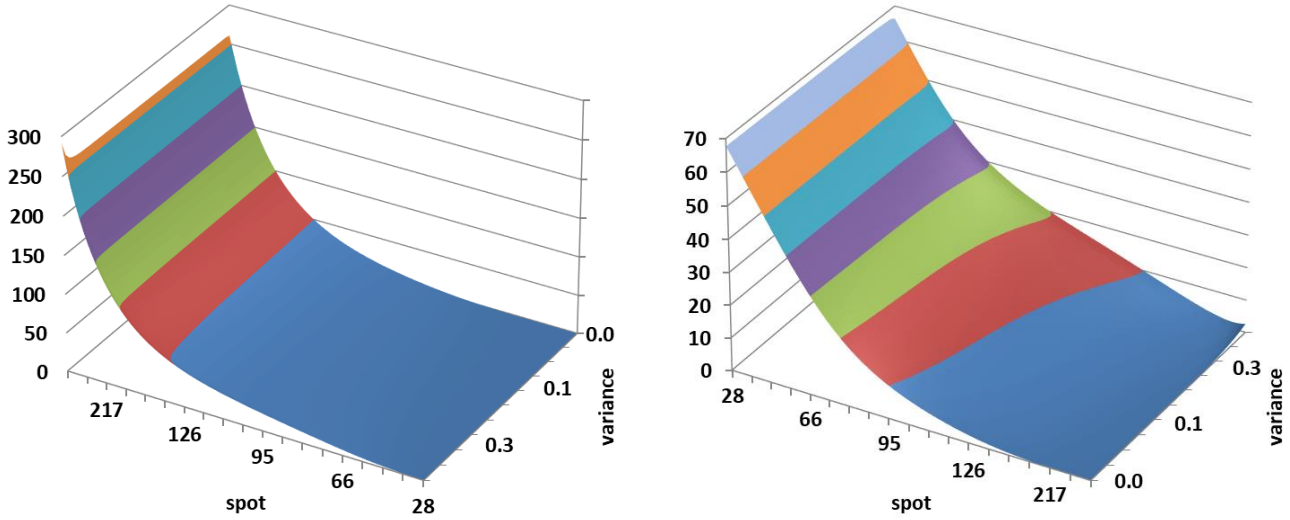

Figure 3. Solution surface for the call (left) and the put (right) using parameters in Table 1

The implied volatility as a function of strike is also derived and presented in Table 2. As a comparison, we have also calculated the analytic results using 3 different types of Fourier inversion method (described in Section 0). It can be seen that numerical results from FEM PDE follow closely with the analytic results. The maximal absolute difference is well within 1 basis point.

Table 2. Implied volatilities: Fourier inversion methods vs FEM PDE

	Strike	**Fourier**	**Inversion**	**Method**	**FEM**	**PDE**
	K	Heston	CDF	PDF	PV	Imp. Vol.
Put	50	38.47%	38.47%	38.47%	0.3205	38.46%
Put	60	36.79%	36.79%	36.79%	0.9013	36.78%
Put	70	35.34%	35.34%	35.34%	2.0649	35.35%
Put	80	34.09%	34.09%	34.09%	4.0575	34.08%
Put	90	32.99%	32.99%	32.99%	7.0957	33.00%
Call	100	32.05%	32.05%	32.05%	13.2184	32.05%
Call	120	30.56%	30.56%	30.56%	6.0256	30.57%
Call	140	29.54%	29.54%	29.54%	2.4237	29.55%
Call	160	28.88%	28.88%	28.88%	0.8948	28.89%
Call	180	28.51%	28.51%	28.51%	0.3171	28.52%
Call	200	28.32%	28.32%	28.32%	0.1113	28.32%

3.4. Double Barrier and Double-No-Touch Products

Double barrier knock-out options and double-no-touch products share great similarity. They differ only in payoff functions (i.e. terminal conditions). For both products, we use again a rectangular domain with $v_{\min} = 0$ and $v_{\max}$ given in (160). Unlike vanilla trades, we set the $y_{\min}$ and $y_{\max}$ of the

domain to the upper and lower barrier (log) levels. The treatment of boundary conditions is essential the same for both. Specifically, for an up-and-out-down-and-out double barrier call, its boundary conditions can be treated as follows

$$\begin{aligned}
&\text{For } (v,y) \in \Gamma_1, \quad && v_{\min} = 0 \Longrightarrow \mathbf{A}\nabla w \cdot \mathbf{n} = 0 \\
&\text{For } (v,y) \in \Gamma_2, \quad && v_{\max} = 0 \Longrightarrow \mathbf{A}\nabla w \cdot \mathbf{n} = 0 \\
&\text{For } (v,y) \in \Gamma_3, \quad && w(t,v,y_{\min}) = g \\
&\text{For } (v,y) \in \Gamma_4, \quad && w(t,v,y_{\max}) = g
\end{aligned} \tag{199}$$

where g is a fixed amount of rebate paid upon knock-out event. For boundary Γ_2, the exact value of $\partial w/\partial v$ is difficult to estimate, therefore we simply assume a homogeneous Neumann condition $\partial w/\partial v = 0$ for this boundary, which removes the boundary integral. This assumption appears justifiable as vega sensitivity diminishes as volatility approaches to a very large number. With these boundary conditions, we can define a function space

$$V = \{\phi \in H^1(\Omega) : \phi = 0 \text{ on } \Gamma_3 \cup \Gamma_4\} \tag{200}$$

The weak form of the PDE is then to find $w = u + l$ for $u \in V$ and $l \in \{\phi \in H^1(\Omega) : \phi = g \text{ on } \Gamma_3 \cup \Gamma_4\}$, such that for all $\psi \in V$ we have

$$\int_\Omega \mathcal{R}(u+l) = \int_\Omega \tilde{\mathcal{R}}\tilde{w} \tag{201}$$

The problem can then be solved accordingly using the FEM method.

In the following, we will demonstrate some numerical results using the FEM PDE solver. We first present the PV benchmarking against analytic solution to confirm the correctness of the implementation. Then we show a typical solution surface for a double barrier option using the established FEM solver.

3.4.1. PV Comparison against Analytic Solution

We attempt to compute the PV of some sample double barrier options. The model and PDE parameters used in the experiments are shown in Table 3. Lipton [44] proposed a (semi-)analytical

solution for both double barrier and double-no-touch products in Heston model, however the analytical formula exists only if we assume zero correlation $\rho = 0.0$ and equal rates $r = q$. Faulhaber's Excel spreadsheet [45], which implements Lipton's algorithm, is used to compute the benchmark PV for comparison. Figure 4 and Figure 5 show the solution (PV) surface at $t = 0$ for the sample barrier call and put, respectively. The PV surface is illustrated as a function of initial variance v_0 and initial spot X_0, from which we can interpolate the product PV using $v_0 = 0.12$ and $X_0 = 100$.

Table 3. Model and PDE parameters used for double barrier knock-out call and put

Heston	Initial Spot Price	X_0	100
	Initial Variance	v_0	0.12
	Mean Variance/Reversion Level	θ	0.10
	Mean Reversion Rate	κ	1.50
	Volatility of Variance	ξ	0.50
	Correlation	ρ	0.00
	Risk Free Rate/Domestic Rate	r	3%
	Dividend Rate/Foreign Rate	q	3%
Trade	Call Strike	K	100
	Option Expiry	T	1Y
	Lower Barrier		70
	Upper Barrier		130
	Rebate		0.00
PDE	Variance Resolution	N_v	50
	Log Spot Resolution	N_y	60
	Temporal Resolution	N_t	50

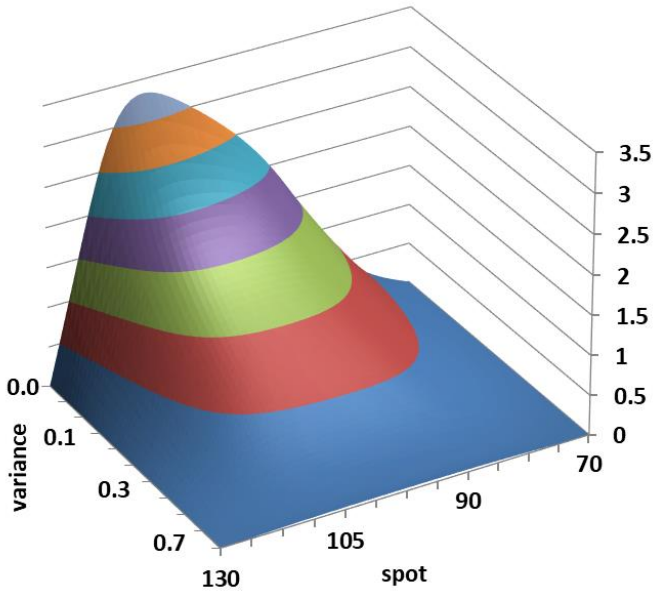

Figure 4. Solution surface of the sample barrier call ($K = 100$)
PV comparison: 1.3499 (FEM) vs 1.3501 (Analytic)

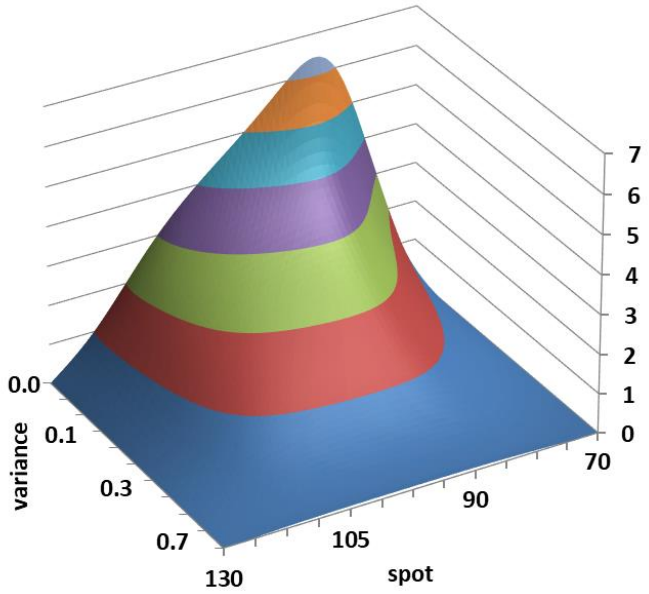

Figure 5. Solution surface of the sample barrier put ($K = 100$)
PV comparison: 2.7917 (FEM) vs 2.7919 (Analytic)

Using the same model parameters in Table 3, we have also calculated the PV of barrier options at various strike and barrier levels. The PV comparison is shown in Table 4. As can be seen, the PV differences between the FEM PDE method and Lipton's analytic formula are typically quite small.

Table 4. PV of Double Barrier Options: FEM PDE vs Lipton's Formula

Option	Strike	Lower Barrier	Upper Barrier	FEM PDE	Analytic
Call	80	60	140	9.5476	9.5499
Call	85	65	135	5.9997	6.0000
Call	90	70	130	3.2028	3.2036
Call	95	75	125	1.2963	1.2969
Call	100	80	120	0.3090	0.3090
Call	105	85	115	0.0206	0.0206
Put	120	60	140	15.6883	15.6854
Put	115	65	135	10.1213	10.1208
Put	110	70	130	5.4893	5.4900
Put	105	75	125	2.2328	2.2334
Put	100	80	120	0.5297	0.5297
Put	95	85	115	0.0350	0.0349

A similar PV comparison is also obtained for double-no-touch products at various barrier levels. The result is shown in Table 5. The PV differences between the FEM PDE method and Lipton's analytic

formula are negligibly small. Figure 6 displays the solution surface of a double-no-touch with lower barrier 70 and upper barrier 130.

Table 5. PV of Double No Touch: FEM PDE vs Lipton's Formula

Lower Barrier	Upper Barrier	FEM PDE	Analytic
60	140	0.5981	0.5982
65	135	0.4908	0.4909
70	130	0.3659	0.3660
75	125	0.2334	0.2335
80	120	0.1128	0.1128
85	115	0.0312	0.0312

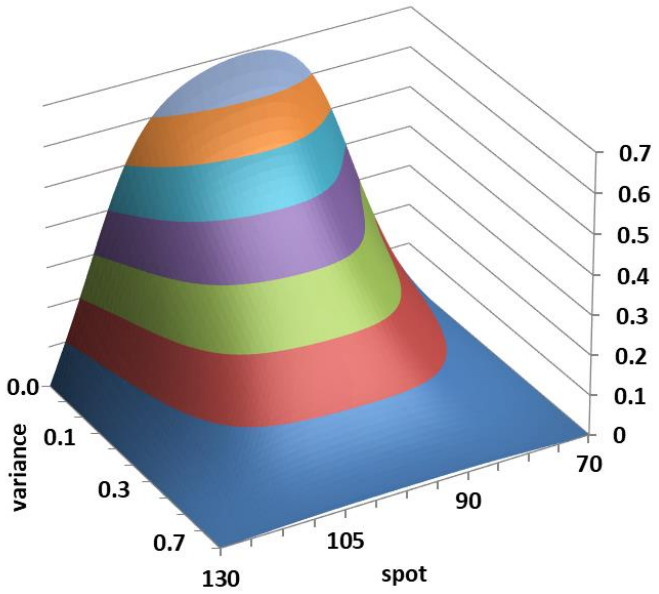

Figure 6. Solution surface of a double-no-touch (lower barrier 70 and upper barrier 130)
PV comparison: 0.3659 (FEM) vs 0.3660 (Analytic)

3.4.2. Crank-Nicolson Oscillation and Remedy

Terminal conditions in barrier options can be discontinuous, pure Crank-Nicolson scheme time-stepping may exhibit localized oscillations in solution. To demonstrate this issue, we have temporarily disabled the Rannacher time-stepping in our implementation and obtained the solution surface for the same sample barrier call. The result is shown in Figure 7, which clearly exhibits kinks at the vicinity of the strike. This issue is mitigated by Rannacher time-stepping, with which we end up with a smooth surface shown in Figure 4.

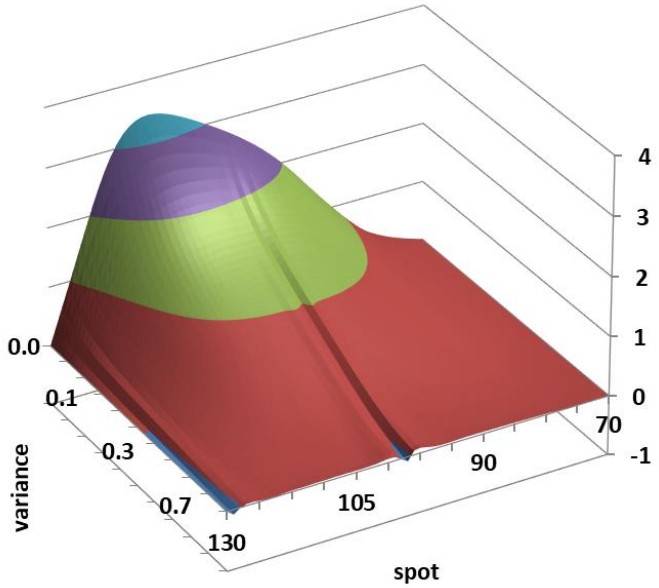

Figure 7. Solution surface of the sample barrier call ($K = 100$) with pure Crank-Nicolson scheme
PV comparison: 1.3826 (FEM) vs 1.3501 (Analytic)

3.4.3. Barrier Call with Knock-Out Rebate

Lastly, we employ the FEM solver to value a barrier call option with a rebate payment upon knock-out. The restrictions of zero correlation and equal rates have been relaxed and the model parameters used in this experiment are shown in Table 6. The solution surface is displayed in Figure 8.

Table 6. Model and PDE parameters for a double barrier with a knock-out rebate

Heston	Initial Spot Price	X_0	100
	Initial Variance	v_0	0.12
	Mean Variance/Reversion Level	θ	0.10
	Mean Reversion Rate	κ	1.25
	Volatility of Variance	ξ	0.60
	Correlation	ρ	-0.70
	Risk Free Rate/Domestic Rate	r	5%
	Dividend Rate/Foreign Rate	q	3%
Trade	Call Strike	K	90
	Option Expiry	T	1Y
	Lower Barrier		85
	Upper Barrier		130
	Rebate		1.00
PDE	Variance Resolution	N_v	50
	Log Spot Resolution	N_y	60
	Temporal Resolution	N_t	50

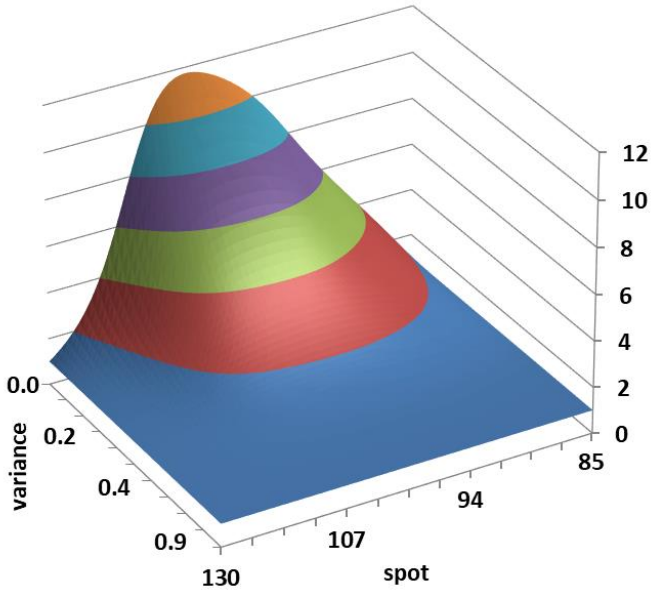

Figure 8. Solution surface for the double barrier with a knock-out rebate

REFERENCES

1. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley-Finance, 2011, pp.82
2. Shreve, S., *Stochastic Calculus for Finance II Continuous-Time Models*, Springer Finance, 2004
3. Wikipage: https://en.wikipedia.org/wiki/Cox–Ingersoll–Ross_model
4. Janek, A.; Kluge, T.; Weron, R. and Wystup, U., "*FX smile in the Heston model*," Chapter 4 in *Statistical Tools for Finance and Insurance* (2nd Ed), Cizek et al. (eds.), Springer-Verlag, 2011.
5. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley-Finance, 2011, pp.98
6. Dragulescu, A. and Yakovenko, V., *Probability distribution of returns in the Heston model with stochastic volatility*, Quantitative Finance 2: 443–453. 2002
7. del Bano Rollin, S.; Ferreiro-Castilla, A.; Utzet, F., *On the density of log–spot in the Heston volatility model*, Stoch. Process. Appl. 120, 2037–2063 (2010)
8. Wikipage: https://en.wikipedia.org/wiki/Fourier_transform
9. Wikipage: http://en.wikipedia.org/wiki/Characteristic_function_(probability_theory)
10. Duffie, D., *Dynamic Asset Pricing Theory*, 3rd Ed, Princeton University Press, 2001, pp.180
11. Wu, L., *Chapter 3 Modeling Financial Security Returns Using Lévy Processes*, pp.150, in "Handbooks in Operations Research and Management Science, Vol 15, Financial Engineering" edited by Birge, J. and Linetsky, V., Elsevier Science, 2008
    Online resource: http://faculty.baruch.cuny.edu/lwu/papers/handbooklevy.pdf
12. Online resource: http://www.math.nus.edu.sg/~matsr/ProbI/Lecture6.pdf
13. Schmelzle, M., *Option Pricing Formulae using Fourier Transform: Theory and Application*, 2010, pp.10, Online resource: http://pfadintegral.com/docs/Schmelzle2010%20Fourier%20Pricing.pdf
14. Elices, A., *Models with Time-Dependent Parameters using Transform Methods: Application to Heston's Model*, "Appendix A: Deriving the Characteristic Function", arXiv.org, Oct. 20, 2008
    Online resource: http://arxiv.org/ftp/arxiv/papers/0708/0708.2020.pdf
15. Albrecher, H.; Mayer, P.; Schoutens, W. and Tistaert, J., *The little Heston trap*, Wilmott Magazine, January 2006: 83–92. 231–262.
16. Wu, L., *From Characteristic Functions and Fourier Transforms to PDF's-CDF's*, online resource: http://faculty.baruch.cuny.edu/lwu/890/ADP_Transform.pdf
17. Wu, L., *Chapter 3 Modeling Financial Security Returns Using Lévy Processes*, pp.149, in "Handbooks in Operations Research and Management Science, Vol 15, Financial Engineering" edited by Birge, J. and Linetsky, V., Elsevier Science, 2008
    Online resource: http://faculty.baruch.cuny.edu/lwu/papers/handbooklevy.pdf
18. Wu, L., *Chapter 3 Modeling Financial Security Returns Using Lévy Processes*, pp.150, in "Handbooks in Operations Research and Management Science, Vol 15, Financial Engineering" edited by Birge, J. and Linetsky, V., Elsevier Science, 2008
    Online resource: http://faculty.baruch.cuny.edu/lwu/papers/handbooklevy.pdf

19. Wu, L., *Chapter 3 Modeling Financial Security Returns Using Lévy Processes*, pp.152, in "Handbooks in Operations Research and Management Science, Vol 15, Financial Engineering" edited by Birge, J. and Linetsky, V., Elsevier Science, 2008
Online resource: http://faculty.baruch.cuny.edu/lwu/papers/handbooklevy.pdf

20. Schmelzle, M., *Option Pricing Formulae using Fourier Transform: Theory and Application*, 2010, pp.24, Online resource: http://pfadintegral.com/docs/Schmelzle2010%20Fourier%20Pricing.pdf

21. Schmelzle, M., *Option Pricing Formulae using Fourier Transform: Theory and Application*, 2010, pp.17, Online resource: http://pfadintegral.com/docs/Schmelzle2010%20Fourier%20Pricing.pdf

22. Elices, A., *Models with Time-Dependent Parameters using Transform Methods: Application to Heston's Model*, arXiv.org, Oct. 20, 2008
Online resource: http://arxiv.org/ftp/arxiv/papers/0708/0708.2020.pdf

23. F. Black and M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy, pp. 637-654, 1973

24. Heston, S., *A closed-form solutions for options with stochastic volatility*, Review of Financial Studies, 6, 327–343, 1993

25. Topper, J. *Financial Engineering with Finite Elements*, Wiley Finance Series, 2005

26. Winkler, G.; Apel, T.; Wystup, U., *Valuation of options in Heston's stochastic volatility model using finite element methods*, in J. Hakala, U. Wystup (eds.) Foreign Exchange Risk, Risk Books, London, 2002

27. Achdou, Y.; Tchou, N., *Variational analysis for the Black-Scholes equation with stochastic volatility*, ESAIM: Mat. Mod. and Num. Analysis, 36/3, 373-395, 2002

28. Miglio, E.; Sgarra, C., *A Finite Element Framework for Option Pricing with the Bates Model*, Computing and Visualization in Science, in publish, 2009

29. Andersen and Piterbarg 2010 - *Interest Rate Modeling*, ch. 2.5.1 page 58

30. Gockenbach, M., *Understanding and Implementing the Finite Element Method*, SIAM, 2006, pp.27

31. Gockenbach, M., *Understanding and Implementing the Finite Element Method*, SIAM, 2006, pp.29

32. Gockenbach, M., *Understanding and Implementing the Finite Element Method*, SIAM, 2006, pp.32 and pp.74

33. Gockenbach, M., *Understanding and Implementing the Finite Element Method*, SIAM, 2006, pp.32

34. Wikipage: https://en.wikipedia.org/wiki/Galerkin_method

35. Gockenbach, M., *Understanding and Implementing the Finite Element Method*, SIAM, 2006, pp.57

36. Gockenbach, M., *Understanding and Implementing the Finite Element Method*, SIAM, 2006, pp.91

37. Gockenbach, M., *Understanding and Implementing the Finite Element Method*, SIAM, 2006, pp.74

38. Online resource: https://www.gnu.org/software/gsl/doc/html/spmatrix.html

39. Online resource: http://www.netlib.org/templates/matlab/cgs.m

40. Online resource: http://www.netlib.org/templates/matlab/gmres.m
Online resource: http://www.netlib.org/templates/matlab/rotmat.m

41. Kelley, T., *Iterative Methods for Linear and Nonlinear Equations*, SIAM, 1987

42. Wikipage: https://en.wikipedia.org/wiki/Preconditioner

43. Online resource: https://www.gnu.org/software/gsl/doc/html/splinalg.html

44. Lipton, A., *Mathematical Methods for Foreign Exchange*, World Scientific, 1st Edition, 2001.

45. Faulhaber, O., *Analytic methods for pricing double barrier options in the presence of stochastic volatility*, Master's thesis, Mathematics Department, University of Kaiserslautern, Germany, 2002