Bergische Universität Wuppertal

Fakultät für Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 21/32

Anna Clevenhaus, Claudia Totzeck and Matthias Ehrhardt

A Gradient Descent Algorithm for the Heston Model

2021-10-25

http://www.imacm.uni-wuppertal.de

Gradient Descent Algorithm for the Heston model

Anna Clevenhaus, Claudia Totzeck, Matthias Ehrhardt

2021-10-25

Contents

Abstract

The Heston model is a well-known two-dimensional financial model. Since the Heston model contains implicit parameters that cannot be determined directly from real market data, calibrating the parameters to real market data is challenging. Moreover, some of the parameters within the model are nonlinear, which makes it difficult to find the global minimum of the optimization problem. In our paper, we present a gradient descent algorithm for parameter calibration of the Heston model. Numerical results show that our calibration of the Heston partial differential equation (PDE) works well for the various challenges in the calibration process. Since the model and algorithm are well known, this work is formulated as a proof of concept. This proof of concept will be incorporated into the space mapping approach in the future.

1 Introduction

In finance, calibrating the parameters of models to fit real market data is challenging because most model parameters are implicit in real market data [9]. Therefore, improving models by introducing additional processes increases the difficulty of the calibration process. We consider the well-known two-dimensional Heston model, which contains at least four parameters implicit in the market data. In contrast to recent research in this area, we focus on optimizing the parameters within a PDE approach instead of a stochastic differential equation (SDE) approach [7, 9]. The parameter calibration results in a constrained optimization problem to minimize a cost functional. The cost functional describes the difference between the reference data and the data obtained by numerically solving our model. We solve the optimization problem using the gradient descent algorithm. In addition to our financial model, the gradient descent algorithm requires the adjoints of the model. Therefore, we first derive the adjoints of the Heston model formally and construct the gradient using well-known techniques from PDE optimization [3, 10]. The gradient descent algorithm has been previously applied to the Heston model using neural networks, so it has not been explicitly computed [6]. In this work, we focus on the optimization of the calibration rather than the numerical solution of the model, i.e., our obtained results can be further improved by using more accurate numerical methods.
In Section 2, we introduce the Heston model and our approach to solving the resulting log-transformed Heston PDE. We then focus on the parameter calibration. In Section 3, we derive the adjoint of the log-transformed Heston model and the corresponding gradient. With the help of the gradient we calibrate the parameters of the second dimension, which are implicit in the real market data. In more detail, a gradient descent method is proposed to compute the best-approximation of the parameters. Finally, in Section 4 we focus on numerical results for the application of the gradient descent algorithm to the Heston model. We analyze the behavior of the algorithm in various problems. The paper ends with a conclusion and an outlook in Section 5.

2 Financial modelling

A financial derivative is a contract between parties whose value at maturity T is determined by the underlying assets at or before time T. Examples of financial derivatives are options. An option is a contract that gives the holder the right (but not the obligation) to exercise a specified transaction at or before time T at a specified price K (strike). A distinction is made between call and put options. A call option holder has the right to buy from the writer, and if he holds a put option, he has the right to sell it to the writer. The time of exercise is determined by the type of option. The standard options are European options, where the holder can exercise only at time T, and American options, where the holder can exercise at any time before and after T. The simplest choice of option is the European option Plain Vanilla, where Plain Vanilla means that the exercise time is limited to T. The payoff function ϕ is given by

$$\phi(S) = \begin{cases} \max(S-K,0) = (S-K)^+, & \text{for } S \geq 0 \quad \text{(call)}, \\ \max(K-S,0) = (K-S)^+, & \text{for } S \geq 0 \quad \text{(put)}, \end{cases} \tag{1}$$

where S is the price of the underlying and $K \in \mathbb{R}^+$. The dynamics of the price of the underlying can be described via a SDE which corresponds to a PDE. In the following we limit ourselves to Plain Vanilla European put options modelled with the Heston model.

2.1 The Heston model

The Heston model was developed by Heston in 1993 [2] and describes the dynamics of the underlying asset through a two-dimensional SDE that includes a stochastic process for the underlying asset S and one for the variance ν. This model is an extension of the well-known Black-Scholes-Merton model, which considers only one stochastic process for the asset

$$dS_t = (r-q)S_t\,dt + \sigma_S S_t\,dW_t^S, \quad S_0 > 0, \tag{2}$$

where r denotes the risk-free rate, q denotes the dividend rate, σ_S denotes volatility, and dW_t^S is a Brownian motion. By definition, the variance is the square of the volatility of the asset, $\nu = \sigma_S^2$. Heston considered a Cox-Ingersoll process for modeling the variance leading to the SDE system of Heston's model under risk neutral measure given by

$$\begin{cases} dS_t = (r-q)S_t\,dt + \sqrt{\nu_t}S_t\,dW_t^S, & S_0 > 0, \\ d\nu_t = \kappa_\nu(\mu_\nu - \nu_t)\,dt + \sigma_\nu\sqrt{\nu_t}\,dW_t^\nu, & \nu_0 > 0, \end{cases}, \tag{3}$$

where κ_ν is the mean reversion rate, μ_ν is the long-term mean, and σ_ν is the volatility-of-variance. The Brownian motions dW_t^S and dW_t^ν are correlated by $\rho \in [-1,1]$. For the variance process to be positive, the Feller condition $2\kappa_\nu\mu_\nu \geq \sigma_\nu^2$ must be satisfied. If the condition is violated, problems arise in the calculation of the square root, because it is complex.

We apply the log-transformation $x = \log(S)$ to the pure Heston model (3) and obtain the SDE system

$$\begin{cases} dx_t = (r - q - \frac{1}{2}\nu_t)\,dt + \sqrt{\nu_t}\,dW_t^x, & x_0 = \log(S_0), \\ d\nu_t = \kappa_\nu(\mu_\nu - \nu_t)\,dt + \sigma_\nu\sqrt{\nu_t}\,dW_t^\nu, & \nu_0 > 0 \end{cases}. \tag{4}$$

Before deriving the Heston PDE, we invert time by introducing $\tau = T - t$ and obtain an initial condition from the payoff function. Using Kolmogorov's backward equation, we derive the log-transformed Heston PDE under risk-neutral measure

$$V_\tau = \frac{\nu}{2}V_{xx} + \frac{1}{2}\sigma_\nu^2\nu V_{\nu\nu} + (r - q - \frac{\nu}{2})V_x + \kappa_\nu(\mu_\nu - \nu)V_\nu + \sigma_\nu\nu\rho V_{x\nu} - rV, \tag{5}$$

where $V(x, \nu, \tau)$ denotes the fair price of the European Plain Vanilla Put option. The initial condition is given by the log transformed payoff-function for put options

$$V(0, \nu, x) = \tilde{\phi}(x) = \max(K - \exp(x), 0). \tag{6}$$

Further the following boundary conditions are considered

$$\begin{aligned} V(\tau, \nu, -\infty) &= K\exp(-r\tau), \\ V(\tau, \nu, \infty) &= 0, \\ rV(\tau, 0, x) &= -V_\tau(\tau, 0, x) + (r - q)V_x(\tau, 0, x) + \kappa_\nu\mu_\nu V_\nu(\tau, 0, x), \\ rV(\tau, \infty, x) &= -V_\tau(\tau, \infty, x) + \frac{\nu}{2}V_{xx}(\tau, \infty, x) + (r - q - \frac{\nu}{2})V_x(\tau, \infty, x). \end{aligned} \tag{7}$$

The boundary condition for $\nu = 0$ is derived by substituting into the equation (5) and for $\nu \to \infty$ we assume that a steady state is reached, i.e. the change within the derivative to ν is zero. Therefore, we obtain an asymptotic boundary condition for $\nu_{\max}$ instead of an analytical one as in the other cases.

2.2 Discretization

Since this is a proof of concept, simple and well-known discretization methods are used to present the approach. We consider uniform meshes in each direction and obtain for the spatial directions $x_i = x_{\min} + i\Delta_x$ for $i = 0, \ldots, N_x$ with $\Delta_x = (x_{\max} - x_{\min})/N_x$ and $\nu_j = j\Delta_\nu$ for $j = 0, \ldots, N_\nu$ with $\Delta_\nu = \frac{\nu_{\max}}{N_\nu}$, as well as $\tau_k = k \cdot \Delta_\tau$ for $k = 0, \ldots, N_\tau$ with $\Delta_\tau = \frac{T}{N_\tau}$ for the temporal direction. Second order central difference stencils are used for spatial discretization. Second-order forward or backward stencils are used on the non-zero boundaries. An exception was made for the first derivative in ν, since the prefactor $\kappa_\nu(\mu_\nu - \nu)$ induces a sign change over the domain. We use an upwind method to reduce the effect of the sign change in the drift term of the variance. In the context of the upwind method, we use the following stencil

$$-\frac{1}{2}a(u_{j+1} - u_{j+1}) + \frac{1}{2}|a|(u_{j+1} - 2u_j + u_{j-1}), \tag{8}$$

where $a = \kappa_\nu(\mu_\nu - \nu)$ [5]. Let $v_{i,j}^k \approx V(x_i, \nu_j, \tau_k)$ and simplify $v^k \approx V(x, \nu, \tau_k)$. For the time discretization, we use the well-known alternating direction implicit (ADI) method, the modified Craig-Sneyd scheme (mCS) [4]. In a first step, we split the operator of the Heston PDE into three operators. The splitting of the operator of the Heston PDE is

$$\mathcal{F}(\tau) = \mathcal{F}_0(\tau) + \mathcal{F}_x(\tau) + \mathcal{F}_\nu(\tau), \tag{9}$$

where the mixed derivative is represented by $\mathcal{F}_0$ and $\mathcal{F}_x$ contains the derivatives in the x-direction and $\mathcal{F}_\nu$ contains the derivatives in the ν-direction, respectively. In each time step, we have to solve the following system of equations

$$\begin{cases} Y_0 = v^k + \Delta\tau \mathcal{F}(\tau_k, v^k), \\ Y_x = Y_0 + \theta\Delta\tau \left(\mathcal{F}_x(\tau_{k+1}, Y_x) - \mathcal{F}_x(\tau_k, v^k)\right), \\ Y_\nu = Y_x + \theta\Delta\tau \left(\mathcal{F}_\nu(\tau_{k+1}, Y_\nu) - \mathcal{F}_\nu(\tau_k, v^k)\right), \\ \hat{Y}_0 = Y_0 + \theta\Delta\tau \left(\mathcal{F}_0(\tau_{k+1}, Y_\nu) - \mathcal{F}_0(\tau_k, v^k)\right), \\ \tilde{Y}_0 = \hat{Y}_0 + (\frac{1}{2} - \theta)\Delta\tau \left(\mathcal{F}(\tau_{k+1}, Y_\nu) - \mathcal{F}(\tau_k, v^k)\right), \\ \tilde{Y}_x = \tilde{Y}_0 + \theta\Delta\tau \left(\mathcal{F}_x(\tau_{k+1}, \tilde{Y}_x) - \mathcal{F}_x(\tau^k, v^k)\right), \\ \tilde{Y}_\nu = \tilde{Y}_x + \theta\Delta\tau \left(\mathcal{F}_\nu(\tau_{k+1}, \tilde{Y}_\nu) - \mathcal{F}_\nu(\tau_k, v^k)\right), \\ v^{k+1} = \tilde{Y}_\nu. \end{cases} \tag{10}$$

We choose $\theta = 2/3$ and improve the implementation as we use the approach in [8].

3 Parameter Calibration

Starting from some given data, we calibrate the parameters of the Heston model using a constrained optimization problem. For a given data set V_{data} and reference parameters u_{ref}, we try to minimize the cost functional

$$J(V, u) = \frac{1}{2}\int_0^T \|V - V_{\text{data}}\|^2 dt + \frac{\lambda}{2}\|u - u_{\text{ref}}\|^2.$$

In the case where no reference parameters u_{ref} are available, we set $\lambda = 0$. In the following, we derive a gradient-based algorithm that allows us to calibrate the parameters numerically.

3.1 First-order optimality conditions for the Heston model

We formally compute the first-order optimality constraints using a Lagrangian approach. We define the operator e, which serves as a constraint in the Lagrangian approach and thus contains the Heston equation and the corresponding boundary conditions. In detail, we denote the Lagrange multipliers by $\psi = (\varphi, \varphi_a, \varphi_b, \varphi_c, \varphi_d)$, set $\Omega = (0, \infty) \times (-\infty, \infty)$ and split the boundary $\partial\Omega$ into

$$\Gamma_a = \partial\Omega \cap \{x = -\infty\}, \qquad \Gamma_b = \partial\Omega \cap \{x = \infty\},$$
$$\Gamma_c = \partial\Omega \cap \{\nu = 0\}, \qquad \Gamma_d = \partial\Omega \cap \{\nu = \infty\}.$$

Moreover, we define

$$A = \frac{1}{2}\nu \begin{pmatrix} \sigma_\nu^2 & \sigma_\nu \rho \\ \sigma_\nu \rho & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} \kappa_\nu(\mu_\nu - \nu) - \frac{1}{2}\sigma_\nu^2 \\ r - q - \frac{\nu}{2} - \frac{1}{2}\sigma_\nu \rho \end{pmatrix}.$$

Then the log-transformed Heston equation can be written as

$$\frac{\partial V}{\partial \tau} - \nabla \cdot A \nabla V - b \cdot \nabla V + rV = 0.$$

Next, we define the operator e which will represent the constraint in the Lagrangian. It is implicitly defined by

$$\begin{aligned}
\langle e(V,u), \psi \rangle = &\int_0^T \int_\Omega \Big[\frac{\partial V}{\partial \tau} - \nabla \cdot A \nabla V - b \cdot \nabla V + rV \Big] \varphi \, dz d\tau \\
&+ \int_0^T \int_{\Gamma_a} \Big[V - K \exp(-r\tau) \Big] \varphi_a \, ds d\tau \\
&+ \int_0^T \int_{\Gamma_b} V \varphi_b \, ds d\tau \\
&+ \int_0^T \int_{\Gamma_c} \Big[rV + \frac{\partial V}{\partial \tau} - (r-q) \frac{\partial V}{\partial x} - \kappa_\nu \mu_\nu \frac{\partial V}{\partial \nu} \Big] \varphi_c \, ds d\tau \\
&+ \int_0^T \int_{\Gamma_d} \Big[\frac{\partial V}{\partial \tau} - \frac{\nu}{2} \frac{\partial^2 V}{\partial x^2} - (r - q - \frac{\nu}{2}) \frac{\partial V}{\partial x} + rV \Big] \varphi_d \, ds d\tau \\
=: &\, T_1 + T_2 + T_3 + T_4 + T_5.
\end{aligned}$$

The Lagrangian for the constrained parameter calibration problem is then given by

$$L(V, u, \psi) = J(V, u) - \langle e(V, u), \psi \rangle.$$

We formally compute the first-order optimality conditions by setting $dL = 0$, for details on the method we refer to [3, 10]. Before computing the Gâteaux derivatives of L [3] in arbitrary directions, we note that by Green's first identity it holds

$$\int_\Omega \varphi (b \cdot \nabla V) \, dz = \int_{\partial\Omega} (b \cdot \vec{n}) V \varphi \, ds - \int_\Omega V \nabla \cdot (b\varphi) \, dz. \tag{11}$$

Therefore, we can rewrite

$$\begin{aligned}
T_1 = &\int_0^T \int_\Omega \varphi \frac{\partial V}{\partial \tau} + A \nabla V \cdot \nabla \varphi - \frac{1}{2} b \cdot \nabla V \varphi + \frac{1}{2} V b \cdot \nabla \varphi + (r + \frac{1}{2} \nabla \cdot b) V \varphi \, dz \\
&- \int_{\partial\Omega} (A \nabla V) \cdot \vec{n} \varphi \, ds - \frac{1}{2} \int_{\partial\Omega} (b \cdot \vec{n}) V \varphi \, ds d\tau \\
= &\Big[\int_\Omega \varphi V \, dz \Big]_{\tau=0}^{\tau=T} + \int_0^T \int_\Omega V \Big[-\frac{\partial \varphi}{\partial \tau} - \nabla \cdot A^\top \nabla \varphi + b \cdot \nabla \varphi + (r + \nabla \cdot b) \varphi \Big] \, dz \\
&+ \int_{\partial\Omega} [(A^\top \nabla \varphi) \cdot \vec{n} - (b \cdot \vec{n}) \varphi] V \, ds - \int_{\partial\Omega} (A \nabla V) \cdot \vec{n} \varphi \, ds d\tau.
\end{aligned}$$

As e is linear in V we obtain for some arbitrary direction h

$$\begin{aligned} d_V\langle e(V,u),\psi\rangle[h] = &\left[\int_\Omega \varphi h\,dz\right]_{\tau=0}^{\tau=T} + \int_0^T\int_\Omega h\left[-\frac{\partial\varphi}{\partial\tau} - \nabla\cdot A^\top\nabla\varphi + b\cdot\nabla\varphi + (r+\nabla\cdot b)\varphi\right]dz \\ &+ \int_{\partial\Omega}[(A^\top\nabla\varphi)\cdot\vec{n} - (b\cdot\vec{n})\varphi]h\,ds - \int_{\partial\Omega}(A\nabla h)\cdot\vec{n}\varphi\,ds\,d\tau \\ &+ \int_0^T\int_{\Gamma_a} h\varphi_a\,ds\,d\tau + \int_0^T\int_{\Gamma_b} h\varphi_b\,ds\,d\tau \\ &+ \int_0^T\int_{\Gamma_c}\left[rh + \frac{\partial h}{\partial\tau} - (r-q)\frac{\partial h}{\partial x} - \kappa_\nu\mu_\nu\frac{\partial h}{\partial\nu}\right]\varphi_c\,ds\,d\tau \\ &+ \int_0^T\int_{\Gamma_d}\left[\frac{\partial h}{\partial\tau} - \frac{\nu}{2}\frac{\partial^2 h}{\partial x^2} - (r-q-\frac{\nu}{2})\frac{\partial h}{\partial x} + rh\right]\varphi_d\,ds\,d\tau. \end{aligned}$$

For the cost functional we have

$$d_V J(V,u)[h] = \int_0^T\int_\Omega h(V - V_{\text{data}})\,dz\,d\tau.$$

We identify the adjoint equation by choosing h appropriately. Note that we are not allowed to vary V at $V(0,\nu,x)$ as the initial condition is fixed. Therefore $h(0,\nu,x)\equiv 0$.
For $h\equiv 0$ on $\partial\Omega$ and $h(T,\nu,x)=0$, we find

$$\begin{aligned} 0 &= d_V L(V,u,\psi)[h] \\ &= \int_0^T h\left[\int_\Omega (V - V_{\text{data}}) + \frac{\partial\varphi}{\partial\tau} + \nabla\cdot A^\top\nabla\varphi - b\cdot\nabla\varphi - (r+\nabla\cdot b)\varphi\right]dz\,d\tau. \end{aligned}$$

Hence the Variational Lemma allows us to conclude

$$\frac{\partial\varphi}{\partial\tau} + \nabla\cdot A^\top\nabla\varphi - b\cdot\nabla\varphi - (r+\nabla\cdot b)\varphi = -(V - V_{\text{data}}) \quad \text{on } \Omega.$$

Now, choose $h(T,\nu,x)\neq 0$ then we obtain the terminal condition $\varphi(T,\nu,x)=0$. We consider the boundary conditions separately.
On Γ_A we have

$$0 = \int_{\Gamma_a}[(A^\top\nabla\varphi)\cdot\vec{n} - (b\cdot\vec{n})\varphi]h - (A\nabla h)\cdot\vec{n}\varphi + h\varphi_a\,ds. \tag{12}$$

Choosing $h\equiv\text{const}\neq 0$ yields

$$0 = h\int_{\Gamma_a}(A^\top\nabla\varphi)\cdot\vec{n} - (b\cdot\vec{n})\varphi + \varphi_a\,ds,$$

hence $(A^\top\nabla\varphi)\cdot\vec{n} - (b\cdot\vec{n})\varphi + \varphi_a = 0$. On the other hand, when choosing $\nabla h\neq 0$, (12) must still hold. This yields $\varphi = 0$ on Γ_a. Similarly, we find on Γ_b that

$$0 = \int_{\Gamma_b}\Big[(A^\top\nabla\varphi)\cdot\vec{n} - (b\cdot\vec{n})\varphi\Big]h - (A\nabla h)\cdot\vec{n}\varphi + h\varphi_b\,ds. \tag{13}$$

With the same arguments, we obtain $\varphi = 0$ on Γ_b.
On Γ_c we find

$$0 = \int_0^T \int_{\Gamma_c} -h(b \cdot \vec{n})\varphi + \Big[rh + \frac{\partial h}{\partial \tau} - (r-q)\frac{\partial h}{\partial x} - \kappa_\nu \mu_\nu \frac{\partial h}{\partial \nu}\Big]\varphi_c ds d\tau.$$

For $h \equiv c \neq 0$, we obtain -$(b \cdot \vec{n})\varphi = r\varphi_c$ and for arbitrary h we use integration by parts on Γ_c to obtain

$$0 = \int_0^T \int_{\Gamma_c} \Big[-(b \cdot \vec{n})\varphi + r\varphi_c + \frac{\partial \varphi_c}{\partial \tau} - (r-q)\frac{\partial \varphi_c}{\partial x} - \kappa_\nu \mu_\nu \frac{\partial \varphi_c}{\partial \nu}\Big] h \, ds d\tau,$$

which leads us to

$$\frac{\partial \varphi}{\partial \tau} - (r-q)\frac{\partial \varphi}{\partial x} - \kappa_\nu \mu_\nu \frac{\partial \varphi}{\partial \nu} = 0 \quad \text{on } \Gamma_c.$$

On Γ_d we find

$$0 = \int_0^T \int_{\Gamma_d} [(A^\top \nabla \varphi) \cdot \vec{n} - (b \cdot \vec{n})\varphi]h + \Big[rh + \frac{\partial h}{\partial \tau} - (r - q - \frac{\nu}{2})\frac{\partial h}{\partial x} - \frac{\nu}{2}\frac{\partial^2 h}{\partial x^2}\Big]\varphi_d ds d\tau.$$

For $h \equiv c \neq 0$, we obtain $(A^\top \nabla \varphi) \cdot \vec{n} - (b \cdot \vec{n})\varphi = r\varphi_d$ and we use integration by parts on Γ_d for arbitrary h which leads us to

$$0 = \int_0^T \int_{\Gamma_c} \Big[(A^\top \nabla \varphi) \cdot \vec{n} - (b \cdot \vec{n})\varphi + r\varphi_d + \frac{\partial \varphi_d}{\partial \tau} - (r - q - \frac{\nu}{2})\frac{\partial \varphi_d}{\partial x} - \frac{\nu}{2}\frac{\partial^2 \varphi_d}{\partial x^2}\Big] h \, ds d\tau,$$

and we obtain

$$\frac{\partial \varphi}{\partial \tau} - (r - q - \frac{\nu}{2})\frac{\partial \varphi}{\partial x} - \frac{\nu}{2}\frac{\partial^2 \varphi}{\partial x^2} = 0 \quad \text{on } \Gamma_d.$$

Altogether, the adjoint equation reads

$$\frac{\partial \varphi}{\partial \tau} + \nabla \cdot A^\top \nabla \varphi - b \cdot \nabla \varphi - (r + \nabla \cdot b)\varphi = -(V - V_{\text{data}}) \quad \text{on } \Omega \tag{14}$$

which is equivalent to

$$\begin{aligned} \frac{\partial \varphi}{\partial \tau} + \frac{1}{2}\nu {\sigma_\nu}^2 \frac{\partial^2 \varphi}{\partial \nu^2} + \nu \sigma_\nu \rho \frac{\partial^2 \varphi}{\partial x \partial \nu} + \frac{1}{2}\nu \frac{\partial^2 \varphi}{\partial x^2} + (\sigma_\nu^2 - \kappa_\nu(\mu_\nu - \nu))\frac{\partial \varphi}{\partial \nu} \\ + (q - r + \frac{\nu}{2} + \sigma_\nu \rho)\frac{\partial \varphi}{\partial x} + (\kappa_\nu - r)\varphi = -(V - V_{\text{data}}) \quad \text{on } \Omega \end{aligned} \tag{15}$$

with terminal condition $\varphi(T) = 0$ and boundary conditions $\varphi = 0$ on Γ_a and Γ_b and

$$\begin{aligned} \frac{\partial \varphi}{\partial \tau} - (r-q)\frac{\partial \varphi}{\partial x} - \kappa_\nu \mu_\nu \frac{\partial \varphi}{\partial \nu} = 0 \quad \text{on } \Gamma_c, \\ \frac{\partial \varphi}{\partial \tau} - (r - q - \frac{\nu}{2})\frac{\partial \varphi}{\partial x} - \frac{\nu}{2}\frac{\partial^2 \varphi}{\partial x^2} = 0 \quad \text{on } \Gamma_d. \end{aligned} \tag{16}$$

The discretization of the adjoint is analogous to the discretization of the Heston PDE.

3.2 Derivation of the gradient

Let $u = (\sigma_\nu, \rho, \kappa_\nu, \mu_\nu)$ be the parameters to be identified, as r and q are given by the data. We compute the optimality condition by setting $d_u L(V, u, \psi) = 0$. In the following we state the derivatives w.r.t. the different parameters separately. For σ_ν we obtain

$$\begin{aligned}
d_{\sigma_\nu}\langle e(V,u),\psi\rangle[h_{\sigma_\nu}] = &\int_0^T \int_\Omega \frac{1}{2}\nu \begin{pmatrix} \sigma_\nu h_{\sigma_\nu} & h_{\sigma_\nu}\rho \\ h_{\sigma_\nu}\rho & 0 \end{pmatrix} \nabla V \cdot \nabla \varphi \\
&- \frac{1}{2}\begin{pmatrix} -\frac{1}{2}\sigma_\nu h_{\sigma_\nu} \\ -\frac{1}{2}h_{\sigma_\nu}\rho \end{pmatrix} \cdot \nabla V\varphi + \frac{1}{2}V\begin{pmatrix} -\frac{1}{2}\sigma_\nu h_{\sigma_\nu} \\ -\frac{1}{2}h_{\sigma_\nu}\rho \end{pmatrix} \cdot \nabla \varphi \\
&+ \frac{1}{2}\nabla \cdot \begin{pmatrix} -\frac{1}{2}\sigma_\nu h_{\sigma_\nu} \\ -\frac{1}{2}h_{\sigma_\nu}\rho \end{pmatrix} V\varphi \, dz - \frac{1}{2}\int_{\partial\Omega} (\begin{pmatrix} -\frac{1}{2}\sigma_\nu h_{\sigma_\nu} \\ -\frac{1}{2}h_{\sigma_\nu}\rho \end{pmatrix} \cdot \vec{n}) V\varphi \, ds \\
&- \int_{\partial\Omega} \frac{1}{2}\nu \begin{pmatrix} \sigma_\nu h_{\sigma_\nu} & h_{\sigma_\nu}\rho \\ h_{\sigma_\nu}\rho & 0 \end{pmatrix} \nabla V \vec{n}\varphi \, ds d\tau \\
= &\, h_{\sigma_\nu} \int_0^T \int_\Omega \frac{1}{2}\nu \begin{pmatrix} \sigma_\nu & \rho \\ \rho & 0 \end{pmatrix} \nabla V \cdot \nabla \varphi + \frac{1}{4}\begin{pmatrix} \sigma_\nu \\ \rho \end{pmatrix} \cdot \nabla V\varphi - \frac{1}{4}V\begin{pmatrix} \sigma_\nu \\ \rho \end{pmatrix} \cdot \nabla \varphi \\
&- \frac{1}{4}\nabla \cdot \begin{pmatrix} \sigma_\nu \\ \rho \end{pmatrix} V\varphi dz + \int_{\partial\Omega} \left(\frac{1}{4}\left(\begin{pmatrix} \sigma_\nu \\ \rho \end{pmatrix} \cdot \vec{n} \right) V\varphi - \frac{1}{2}\nu \begin{pmatrix} \sigma_\nu & \rho \\ \rho & 0 \end{pmatrix} \nabla V \vec{n}\varphi \right) ds d\tau \\
= &\, h_{\sigma_\nu} \int_0^T \int_\Omega \frac{1}{2}\nu \left(\sigma \frac{\partial V}{\partial \nu}\frac{\partial \varphi}{\partial \nu} + \rho\left(\frac{\partial V}{\partial x}\frac{\partial \varphi}{\partial \nu} + \frac{\partial V}{\partial \nu}\frac{\partial \varphi}{\partial x} \right) \right) \\
&+ \frac{1}{4}\left(\rho\left(\varphi \frac{\partial V}{\partial x} - V\frac{\partial \varphi}{\partial x} \right) + \sigma\left(\frac{\partial V}{\partial \nu}\varphi - V\frac{\partial \varphi}{\partial \nu} \right) \right) dz \\
&+ \int_{\partial\Omega} \frac{1}{4}(\sigma + \rho) V\varphi - \frac{1}{2}\nu \left((\sigma + \rho)\frac{\partial V}{\partial \nu} + \rho \frac{\partial V}{\partial x} \right) \varphi \, ds d\tau
\end{aligned}$$

Similarly, we obtain for the other derivatives

$$\begin{aligned}
d_\rho\langle e(V,u),\psi\rangle[h_\rho] = &\, h_\rho \int_0^T \int_\Omega \frac{1}{2}\nu\sigma \left(\frac{\partial V}{\partial x}\frac{\partial \varphi}{\partial \nu} + \frac{\partial V}{\partial \nu}\frac{\partial \varphi}{\partial x} \right) + \frac{1}{4}\sigma\left(\frac{\partial V}{\partial x}\varphi - V\frac{\partial \varphi}{\partial x} \right) dz \\
&+ \int_{\partial\Omega} \frac{1}{4}\sigma V\varphi - \frac{1}{2}\nu\sigma\varphi\left(\frac{\partial V}{\partial x} + \frac{\partial V}{\partial \nu} \right) ds d\tau \\
d_{\kappa_\nu}\langle e(V,u),\psi\rangle[h_{\kappa_\nu}] = &\quad h_{\kappa_\nu} \int_0^T \int_\Omega \frac{1}{2}(\mu_\nu - \nu)\left(V\frac{\partial \varphi}{\partial \nu} - \frac{\partial V}{\partial \nu}\varphi \right) - \frac{1}{2}V\varphi dz \\
&- \frac{1}{2}\int_{\partial\Omega} (\mu_\nu - \nu)\varphi V ds - \int_{\Gamma_c} \mu_\nu \frac{\partial V}{\partial \nu}\varphi \, ds d\tau \\
d_{\mu_\nu}\langle e(V,u),\psi\rangle[h_{\mu_\nu}] = &\, h_{\mu_\nu} \int_0^T \int_\Omega \frac{1}{2}\kappa\left(V\frac{\partial \varphi}{\partial \nu} - \frac{\partial V}{\partial \nu}\varphi \right) dz \\
&- \frac{1}{2}\int_{\partial\Omega} \kappa V\varphi ds - \int_{\Gamma_c} \kappa_\nu \frac{\partial V}{\partial \nu}\varphi \, ds d\tau
\end{aligned}$$

Note that $d_u L(V, u, \psi)[h_u] = 0$ needs to hold for arbitrary directions h_u. Therefore, we can read off the gradient from the above expressions.

3.3 Gradient descent algorithm for the parameter calibration

Solving the first-order optimality condition all at once will be difficult due to the forward-backward structure. Therefore, we propose a gradient descent algorithm in the following. For a given initial parameter set u_0, we can solve the state equation. With the state solution at hand, we can compute the adjoint equation backwards in time. Then we have all the information available to compute the gradient and update the parameter set using a gradient step.

Algorithm 1: Gradient descent method for Heston parameter calibration

```
Result: calibrated parameters for Heston model
initialize parameters u_0;
while ||gradient|| > ε do
    solve (5);
    solve (14);
    compute the gradient;
    line search for step size;
    update the parameter set;
end
```

For the line search we use the *Armijo rule* [10], which is described in the following. Given a descent direction d_k of f at u_k, for example $d_k = -\nabla f(u_k)$, choose the maximum $\sigma_k \in \{1, 1/2, 1/4, \ldots\}$ for which

$$f(u_k + \sigma_k d_k) - f(u_k) \leq \gamma \sigma_k \langle f'(u_k), d_k \rangle. \tag{17}$$

Here $\gamma \in (0, 1)$ is a numerical constant, which is problem dependent and typically chosen as $\gamma = 10^{-4}$, as we have. As we have restrictions for the parameter domain for κ_ν, μ_ν, σ_ν and ρ, as well as the condition that the Feller condition has to be fulfilled, we further use the *projected Armijo rule* [10]. Choose the maximum $\sigma_k \in \{1, 1/2, 1/4, \ldots\}$ for which

$$f(\mathcal{P}(u_k - \sigma_k \nabla f(u_k))) - f(u_k) \leq -\frac{\gamma}{\sigma_k} \|\mathcal{P}(u_k - \sigma_k \nabla f(u_k)) - u_k\|_W^2.$$

As before, $\gamma \in (0, 1)$ is a numerical constant.

4 Numerical Results

In this section, we report the behavior of the gradient descent algorithm for the Heston model, Algorithm 1. We focus on the behavior when using random initial values, short and long run times, and different discretization meshes. For all test sets, we generate V_{data} using the following parameters for the Heston model

$$K = 10,\ r = 0.1,\ q = 0.05,\ \kappa_\nu = 5,\ \mu_\nu = 0.16,\ \sigma_\nu = 0.9,\ \rho = 0.1. \tag{18}$$

If nothing else is set due to the test setting, the discretization mesh is $N_x = N_\nu = 80$, and $N_\tau = 40$. The same discretization is used for the test set and the reference set.
For the different test cases, the percentage improvement of the cost function given by

$$\frac{J(V, u_0) - J(V, u_{\text{opt}})}{J(V, u_0)}, \tag{19}$$

the percentage change of the optimized parameter from the initial value, calculated similarly to (19), and the number of iterations are reported. For the first two tests, we choose $\kappa_\nu = 5.2$, $\mu_\nu = 0.18$, $\sigma_\nu = 0.92$ and $\rho = 0.05$ as initial values. The first test focuses on different discretization meshes, where we choose $N_x = [80, 90, 100, 110, 120, 130, 140]$ and $N_\nu = \frac{N_x}{2}$ and $N_\nu = N_x$. Therefore, we show only the N_ν values in the plots, since their range is unique. The results are shown in the Figures 1, 2, and 3.
In the second test, the behavior using different maturities is examined. The results are shown in the Figures 4, 5 and 6.
In the last test, we set a maximum percentage deviation where the initial estimate for the parameters is calculated randomly. For each maximum deviation, we compute 100 initial values. The results can be found in the Figures 7, 8 and 9.
In comparison, we find that the iteration count is less than 10 iterations in most cases, as Figures 3, 8 and 6 show. Furthermore, the iteration count increases with larger deviation of the initial values from the global optimum, see Figure 8. Figures 3 and 4 show a significant improvement of the cost functional independent of the maturity or the chosen mesh size. The small effect of the maturity for the optimized parameters is shown in Figure 2.
The importance of the initial parameter is shown in Figure 7, as we observe that the best optimized cost functional deteriorates with larger deviations from the reference values, and in Figure 9, where the optimal parameter deviates from the global optimum, indicating that the optimal values are local minima. Finally, we conclude that the presented gradient descent algorithm gives very good results in finding the nearest local minimum that sufficiently improves the cost functional, as verified by Figures 7 and 9.

5 Conclusion and Outlook

The numerical results show that the application of the gradient descent algorithm is feasible even if we only reach a local minimum. Since this work is a proof-of-concept, it is a first step towards (real) data-based calibration of the model. Indeed, as future work we plan to extend this method using the space-mapping approach [1]. This will allow us to fit the parameters of the Heston model to real market data rather than to a constructed example, using the SDE solver as the fine solver and the PDE approach as the coarse solver.

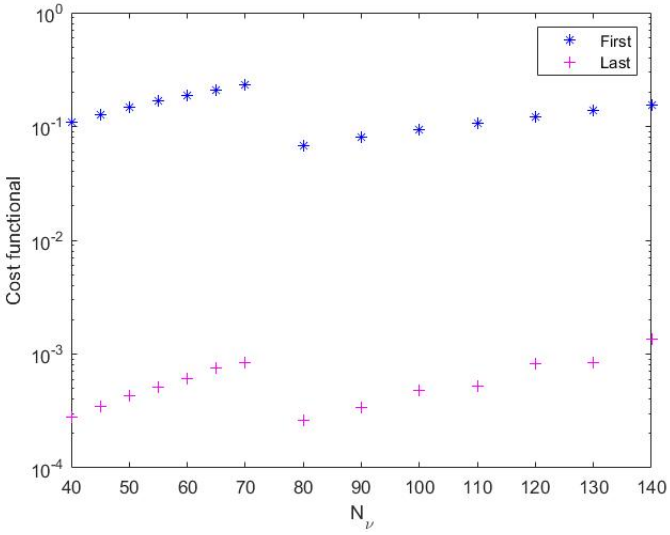

Figure 1

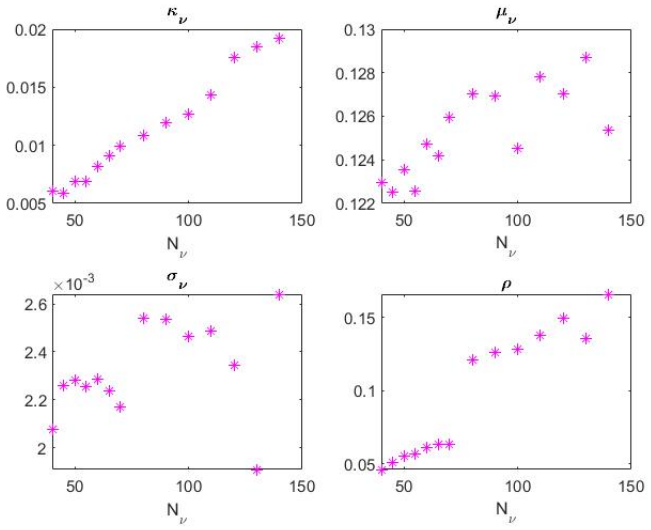

Figure 2: The percentage deviation of the initial guess to the optimized parameter for the different mesh sizes.

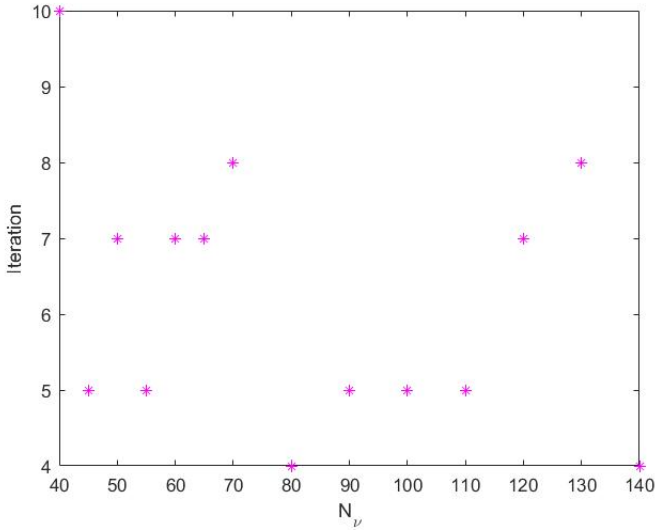

Figure 3: The number of iterations for the different mesh sizes.

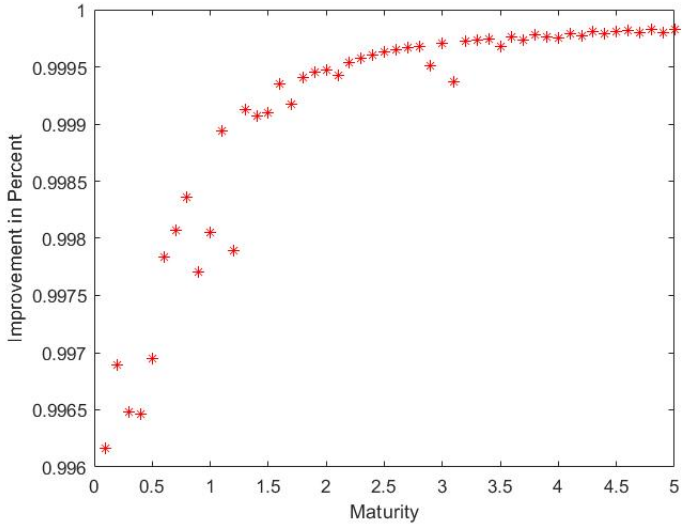

Figure 4: The percentage optimization of the cost functional computed with 19.

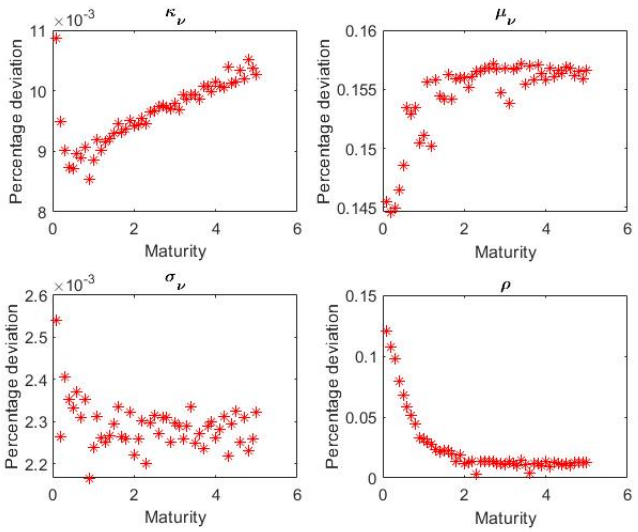

Figure 5: The percentage deviation of the initial guess to the optimized parameter for the short and long time maturities.

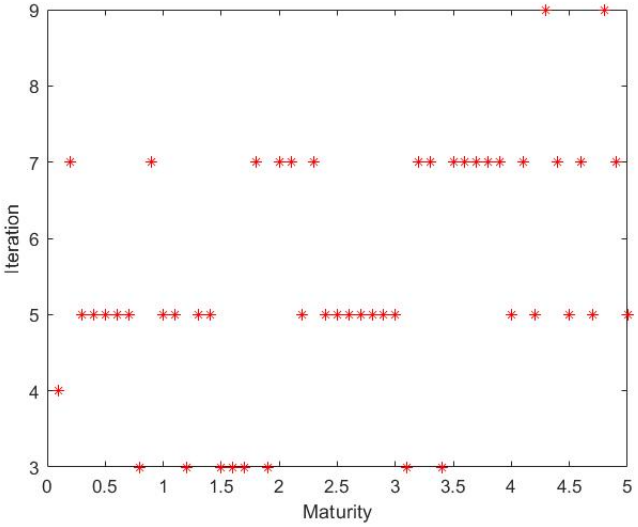

Figure 6: The iteration count for the different maturities is shown.

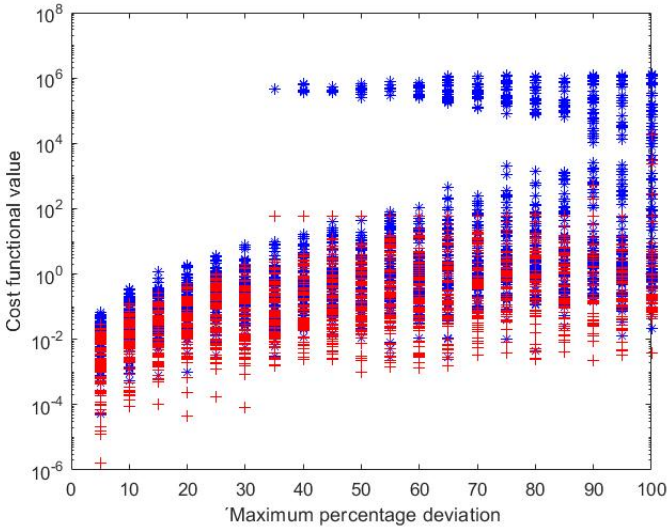

Figure 7: The first cost functional values are given in blue stars and the optimized cost functional values are red crosses.

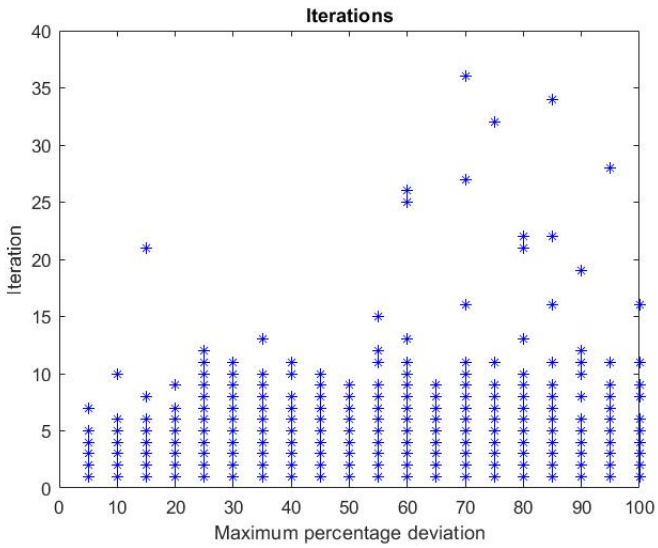

Figure 8: The number of iterations for the optimization of different initial values is presented.

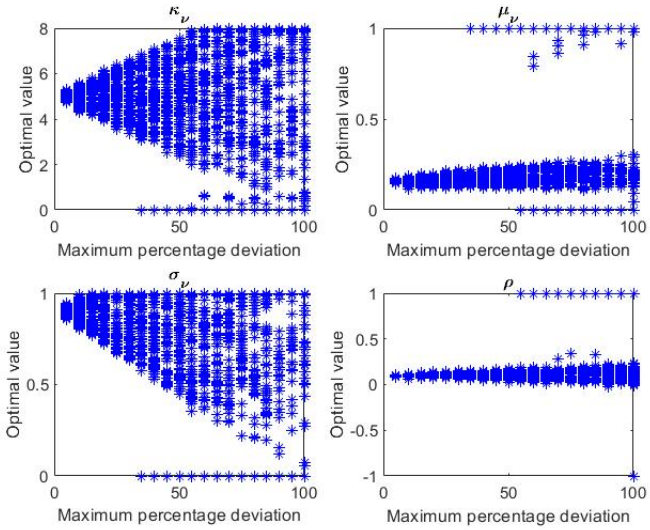

Figure 9: The optimized parameter for the random initial values.

References

[1] D. Echeverría, D. Lahaye, and P.W. Hemker. Space mapping and defect correction. In *Model Order Reduction: Theory, Research Aspects and Applications*. Springer, Berlin, Heidelberg, 2008.

[2] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Fin. Stud.*, 6(2):327–343, 1993.

[3] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*. Springer, 2009.

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

[5] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser, Basel, 1992.

[6] S. Liu, C.W. Oosterlee, and S. M. Bohte. Pricing options and computing implied volatilities using neural networks. *Risks*, 7(1):16, 2019.

[7] M. Mrázek and J. Pospíšil. Calibration and simulation of Heston model. *Open Mathematics*, 15, 2017.

[8] L. Teng and A. Clevenhaus. Accelerated implementation of the ADI schemes for the Heston model with stochastic correlation. *J. Comput. Sci.*, 36:101022, 2019.

[9] L. Teng, M. Ehrhardt, and M. Günther. On the Heston model with stochastic correlation. *Int. J. Theor. Appl. Fin.*, 6, 2016.

[10] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen*. Springer, 2009.