Sensitivities in Option Pricing Models

Tirtha P. Timsina

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Mathematics

Ekkehard W. Sachs , Professor of Mathematics, Chair

Jeffrey T. Borggaard, Professor of Mathematics

Michael T. Cliff, Professor of Finance

Martin V. Day, Professor of Mathematics

August 13, 2007

Blacksburg, Virginia

Keywords: Adjoint equations, Gradient method, Jump diffusion, Optimization,

Sensitivity

Sensitivities in Option Pricing Models

Tirtha Prasad Timsina

(Abstract)

The inverse problem in finance consists of determining the unknown parameters of the pricing equation from the values quoted from the market. We formulate the inverse problem as a minimization problem for an appropriate cost function to minimize the difference between the solution of the model and the market observations. Efficient gradient based optimization requires accurate gradient estimation of the cost function. In this thesis we highlight the adjoint method for computing gradients of the cost function in the context of gradient based optimization and show its importance. We derive the continuous adjoint equations with appropriate boundary conditions for three main option pricing models: the Black-Scholes model, the Heston's model and the jump diffusion model, for European type options. These adjoint equations can be used to compute the gradient of the cost function accurately for parameter estimation problems.

The adjoint method allows efficient evaluation of the gradient of a cost function $F(\sigma)$ with respect to parameters σ where F depends on σ indirectly, via an intermediate variable. Compared to the finite difference method and the sensitivity equation method, the adjoint equation method is very efficient in computing the gradient of the cost function. The sensitivity equations method requires solving a PDE corresponding to each parameter in the model to estimate the gradient of the cost function. The adjoint method requires solving a single adjoint equation once. Hence, for a large number of parameters in the model, the adjoint equation method is very efficient.

Due to its nature, the adjoint equation has to be solved backward in time. The adjoint equation derived from the jump diffusion model is harder to solve due to its non local integral term. But algorithms that can be used to solve the Partial Integro-Differential Equation (PIDE) derived from jump diffusion model can be modified to solve the adjoint equation derived from the PIDE.

DEDICATION

To my family for all their support.....

ACKNOWLEDGEMENT

I would like to express my sincere appreciation and gratitude to my advisor, Professor Ekkehard W. Sachs. Thank you for believing in me, being patient with me, supporting me and guiding me throughout this thesis. I would like to thank Professor Day, Professor Borggaard and Professor Cliff for serving on my committee.

I am very thankful to my friends and family for supporting me while I have been working on my dissertation. Thank you to all of my school teachers, undergraduate professors and the Virginia Tech Mathematics department for all their support. Of course I can not thank enough my parents, brothers and sisters without whose support I would not have come this far.

People too numerous to mention, have made my time in America and especially at Virginia Tech very productive and enjoyable. I would like to take this opportunity to thank all of them. Last, and certainly not least, I thank my wife Sarita for her support and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The Brownian motion has played an important role in describing the evolution of market prices. A market described by Brownian motion relates the stochastic problem of finding a fair price for derivatives with the solution of a deterministic partial differential equation with constant coefficient. With few more assumptions about the market, the approach proposed by Black-Scholes [11] has gained a great success among practitioners because it is simple has a closed form solution. Unfortunately, the real market data do not always follow the assumptions made in the Black-Scholes model. Hence it does not adequately describe the underlying asset price process. A key assumption of the Black-Scholes model is that the underlying asset price follows a geometric Brownian motion with constant volatility. However, the implied volatilities from the market prices of the options tend to vary across both strike prices and maturities. To reduce the model specification error, various alternative models that relax the unrealistic assumptions in the Black-Scholes model have been proposed.

Although the Brownian motion implies a continuous market evolution, discontinuous and unforeseeable abrupt movements of markets are common in reality. Hence the price process can also jump with the possibility of high or infinite jump frequencies. This gave arise to a new class of models called the jump diffusion models.

Jump diffusion models can be categorized into finite activity models and infinite activity models which are a part of Lévy models. In jump diffusion models the log-price of the underlying asset follows a diffusion process with rare jumps at some random time. These random time jumps can be modelled using a compound Poisson process. This approach can be reasoned by the observation that most of the time the price only changes marginally in a small time period. Once in a while the stock price tends to jump by a significant amount for example after a new trade deal,

winning a court case or a terrorist attack.

A first notable attempt to correct the lack of fit was made by Merton [37]. He added jumps to the continuous model, where the jumps are counted by a Poisson process and the jump amplitude are assumed to be log-normally distributed. Using most features of the Black-Scholes model, he obtained a closed form solution for the jump diffusion pricing problem. Similarly, Kou's model [29] is another example of a jump diffusion model that assumes that the jump sizes follow a double exponential distribution. The Merton and Kou's model are a class of finite activity Lévy models. The Variance Gamma process introduced by Madan and Seneta [34] is an example of an infinite activity Lévy process. Just to remain simple and consistent we will only consider finite activity models. For infinite activity models see [14, 17, 19] and references within. Most jump-diffusion models do not have a closed form solution and have to be solved numerically

As option contracts and pricing PDE's become more complex, the numerical techniques for solving nonlinear pricing PDEs become very important. Jump-diffusion models like the ones described by Merton and Kou are getting popular as they can explain observed features like the non-normal log-returns and the volatility smile. Under certain assumptions the jump-diffusion models lead to a partial integro-differential equation (PIDE) involving a non-local integral term. Because of the non-local integral term it is much harder to solve than the Black-Scholes PDE.

Stochastic volatility models are also a modification of the Black-Scholes model that assumes that the volatility is no longer a constant. Stochastic volatility models have been studied extensively by Cox-Ingersoll-Ross [20], Heston [24], Hull-White [25] and Lewis [31]. Although each of the authors have their own assumptions about the volatility process, the most popular one is the model derived by Heston as it has a closed form solution.

Parameter estimation is very important issue in option price modelling. Without

correct parameters, mispricing can still occur even if the model is correctly specified. Normally, the parameters in the underlying financial model are not directly observable from the market and have to be determined indirectly from other observable quantities. In financial modelling, calibrating a model means finding numerical values of its parameters such that the model is consistent with the market. Reconstruction of volatility surfaces from market data is an example of a calibration problem. A well calibrated models fit the observed market prices within acceptable tolerances. Hence, parameter estimation is an inverse process that consists of determining the coefficients of the differential equation governing the option pricing processes. The inverse problem is formulated as a minimization problem for an appropriate cost function. The cost function depends on the parameter vector $\mathbf{p}$ and it is minimized with respect to the set of all parameters. Parameters are found by solving the inverse problem that is formulated as minimizing a cost function. So the calibration problem is the inverse of the pricing problem. Instead of computing prices in a model with given values for its parameters, one computes the values of the model parameters that are consistent with observed prices. Pricing and hedging financial derivatives depend heavily on the knowledge of parameters such as volatility.

As pricing models become more complex calibration also becomes hard. Compared to the calibration problem for a generalized Black-Scholes model with a constant or deterministic local volatility function in [15], the calibration problem under the jump diffusion model is harder and substantially more expensive in terms of computational time.

Gradient based methods are an efficient way of solving minimization problems when gradient of the cost function F is available. However, it is not always possible to evaluate the gradient of the cost function efficiently. The easiest way of finding the gradient in minimization problem is by using finite difference method. When

the gradient of the objective function F is not computable in a closed form, or not computable numerically with the required accuracy, as an alternative to gradient methods, derivative free optimization methods are used. As opposed to gradient methods, these algorithms use only the values and not the gradient of F and typically require a lot of function evaluations.

In this thesis, we introduce an efficient way of evaluating the gradient of the objective function by using adjoint equations. This involves the derivation of an adjoint equation of the model partial differential equation and the solution of the adjoint equation. The adjoint method consists of constructing a reverse model in order to compute the derivative of objective function with respect to the model parameters. The solution of the direct problem depends critically on the choice of some of the parameters and to a lesser degree on the other parameters. This leads to the problem of determining the sensitivities of the solution with respect to each of the parameters involved. Hence we also highlight the importance of sensitivity analysis in parameter estimation.

OUTLINE OF THIS THESIS

Chapter 2 reviews the basic assumptions about the popular financial models. As a motivation, without much mathematical rigor and complexity the Black-Scholes PDE is derived. The weakness of the model is shown by an implied volatility smile and an implied volatility surface. Next, the PDE for stochastic volatility is presented and its scope and popularity is highlighted. Finally a jump diffusion model is derived. This model assumes that the stock price follows geometric Brownian motion with finite intensity jumps that are log-normally distributed. The resulting model is called a PIDE model.

Chapter 3 is concerned with the available finite difference methods to solve the PIDE model. Strengths and weaknesses of implicit, explicit and implicit-explicit

methods for solving the PIDE are discussed. When the integral term is non-local, the possibility of truncation and replacing the integral term with a linear combination of delta functions is also discussed.

Chapter 4 contains the **main findings of this thesis**. To estimate the unknown parameters in any of the financial models we first reformulate it as an optimization problem. Gradient based optimization techniques utilize the gradient information to identify search directions and are generally more efficient than non-gradient optimizations for well-behaved functions. We derive adjoint equations for all three types of option pricing models and use the adjoint method to efficiently evaluate the derivatives of the objective function depending on the solution of a partial differential equation (PDE). The adjoint method gives rise to a powerful and efficient way of estimating parameters in option pricing models. The advantage of adjoint equations method over other methods is also highlighted.

Chapter 5 gives the basics of gradient based optimization and some examples of gradient based parameter estimation. Finally in conclusion we summarize the contribution of this thesis and highlight future research directions.

CHAPTER 2

FINANCIAL DERIVATIVES

2.1 INTRODUCTION

A derivative is a financial instrument that has a value determined by the value or values of one or more underlying assets or indices of assets. Derivatives can be based on stocks, bonds, currencies, and even indices of these various things, such as the Dow Jones Industrial Average. A derivative can depend on almost any variable from the harvest of coffee in South America to the number of hurricanes striking Florida in one season. For example, the current price of a stock of a particular company is determined by the market; however, the future price of the stock typically is uncertain. A month or a year in the future, the price may increase, decrease, or remain the same. Hence, the buyers and sellers often like to hedge their assets against this uncertainty about future price by signing a contract for future trading at a specified price. Hence, derivatives can be used to reduce risk by allowing the investor to hedge an investment or exposure and function as an insurance policy against adverse market movements. Derivatives, unlike the underlying assets, are mostly synthesized by investment banks and other financial institutions. They can either be made specific to a particular client or, general enough to be traded in a financial market, just like the underlying assets. The range of possible derivatives is essentially unlimited but the most common types of derivatives are the called options.

An option is a derivative with a specified payoff function that can depend on the prices of one or more underlying assets. As the name suggests, the holder of an option has no obligation to buy or sell the underlying asset. The time when the contract ends is known as the expiration date or the maturity date of the option. The price at which the asset may be bought or sold is called the strike, or exercise

price. Most options can only be exercised once, and have a fixed expiration date, after which the option is no longer valid. There are many different schemes for prescribing when an option can be exercised. If the asset can only be bought or sold at expiration, the option is said to be European. If the option can be exercised at any time up to the expiration, the option is called American. The options of European type are also known as plain vanilla options. At a fundamental level every option is divided into two categories; **call** options and **put** options.

> **Call** option gives its holder the right, but not the obligation *to buy* an asset at an agreed expiration date ($t = T$) at an agreed strike price (K).
>
> **Put** option, gives holder the right, but not the obligation, *to sell* an asset at an agreed expiration date ($t = T$) at an agreed strike price (K).

The price at which the asset will be bought or sold is called the strike price (K). The strike price is set at the time of writing the option. Options may also be used as a hedging instrument against a possible increase or decrease in value of the underlying asset. For example, if Starbucks needs a particular commodity, such as coffee beans, on a regular basis, then they can guard against a rise in the price of coffee beans by purchasing a put option. If the price of coffee beans remains low, then the option is not exercised and coffee beans are bought at the current price in the market. If the price rises above the strike price, then the option is exercised to buy coffee beans at a below-market value.

A holder of a call option will only buy the underlying asset at strike price K if the price of the asset is higher than the strike price of the option. In the case of a put option, the asset will only be sold at strike price K if the price of the asset is lower than the strike price of the option. If $S(t)$ is the asset value at time t, then at the

time of expiration ($t = T$), the payoff for the European option is:

$$\text{pay off} = \begin{cases} \max\left[S(T) - K, 0\right] & \text{for call options} \\ \max\left[K - S(T), 0\right] & \text{for put options} \end{cases}$$

This is because if $S(T) > K$ the call option will be exercised for a profit of $S(T) - K$. If $S(T) \leq K$ the call option will not be exercised. Similarly, a put option is exercised only if the asset price falls below the strike price for a profit of $K - S(T)$. If $S(T) \geq K$, the put option will not be exercised. A typical payoff graph for a call and a put options at strike price of 10 is given in Figure 2.1.

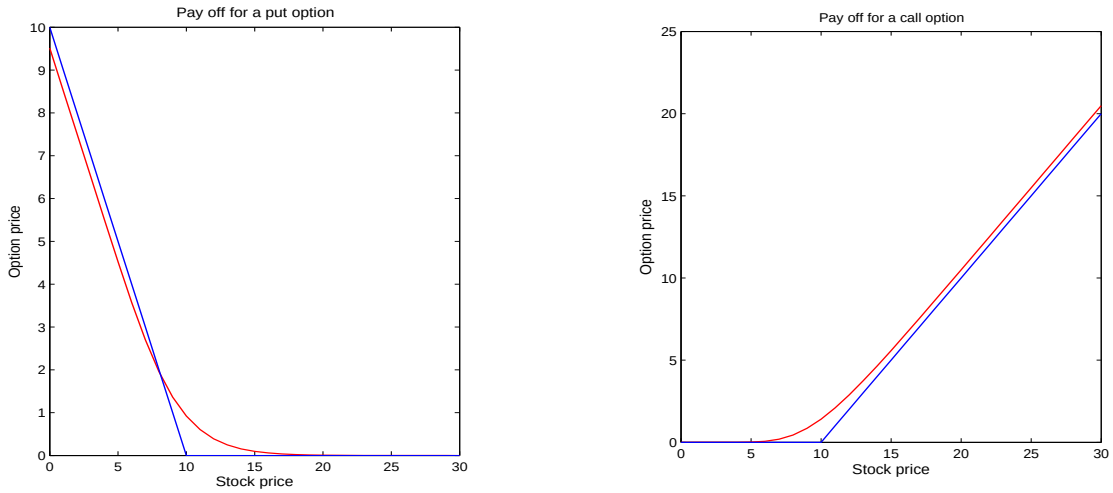


Figure 2.1. Payoff function for typical Put and Call options with strike price 10.

The blue line indicates the pay off values at the time of expiration and the red line indicates the option values when time to expiration is 1 year for both call and put options where other variables are fixed. Throughout this thesis, we will consider only the European options as they are the most popular and mathematically less complex than American options. General background to the mathematical theory including arbitrage pricing, Ito calculus and Brownian motion is not considered here. For general introduction to the theory of financial mathematics and stochastic

calculus see, among others, [9, 10, 26, 40, 42, 49].

2.2 DERIVATION OF THE BLACK-SCHOLES EQUATION

The Black-Scholes option-pricing model, first introduced in the 1973 paper, *The Pricing of Options and Corporate Liabilities*, is probably the most widely taught, and the best-known option pricing model in finance. In spite of its unrealistic assumptions about the market, the concepts behind the Black-Scholes model provides an important framework for studying the basics of option pricing. The Black-Scholes model proved the importance of mathematics in the field of finance. It also led to the growth and success of the new field of financial mathematics. Since the introduction of the Black-Scholes model, all other option pricing models have been either an extension or a generalization of this model. Prices derived from the Black-Scholes model are good approximations to the actual option prices in the market. Hence traders often use it as a base model and prefer to think of the options in terms of the implied volatility and not the actual price of the options.

Let us assume that an asset corresponds to the price of a stock traded in the market and we have a call option on the stock. An option derives its value primarily from three sources. The first source is the intrinsic value, the value for an investor who exercises the option immediately. This value would be the difference between the stock price and the exercise price. If a call option is in the money, the stock price is greater than the exercise price and the investor will benefit from exercising the option to buy the stock at the agreed exercise price and sell it immediately at the market price. Conversely, if a call option is out of the money the stock price is less than the exercise price, and the investor will not benefit from exercising the option immediately.

The second source is the time value. If an option is out of the money now, there is still a possibility that it will be in the money at expiration. The time value of the

option is derived from the time remaining to maturity. In pricing the time value, an investor needs to take account of both the probability of the option maturing in the money, as well as the degree to which it will be in the money.

The third factor contributing to the value of the option is the volatility of the underlying asset. Volatility can be defined as the degree of uncertainty with respect to the future price of the asset. The higher the volatility, the wider the range of potential future prices. This results in a proportionately wide range of possible outcomes for the shareholder. A large decrease in the stock price would result in a large loss while a large increase in the stock price would result in a large gain.

The value of financial derivatives is known at the time of expiration. The derivative pricing problem is concerned with finding, if it does exist, a fair price for a derivative at any time $t < T$. So we want to find a deterministic function that can give the fair price of the derivative any time before expiration. The original work by Black-Scholes is the basis for modern financial mathematics. Black-Scholes uses a hedging strategy to derive the Black-Scholes derivative pricing equation. It regards the possibility for the writer to minimize the risk associated to the derivative he is going to sell. In this case one looks for a deterministic hedging strategy. In this section we provide an introduction to this model, proving its completeness in the absence of arbitrage. Moreover we deduce the deterministic partial differential operator related to the hedging technique for pricing derivative. In their paper, Black and Scholes made some assumptions on the kind of market they were working in. These are listed below:

1. The asset price processes follows a Geometric Brownian Motion (GBM) with constant variance.

2. Borrowing and lending take place at a constant interest rate r, and the underlying stock does not pay cash dividends.

3. Markets are *perfect*; there are no transaction costs, taxes, or short-sale restrictions.

4. Trading takes place continuously in time.

A market described by a Brownian Motion relates the stochastic problem of finding a fair price for derivatives to the solution of a deterministic partial differential equation with constant coefficients. Let S be the price of a stock at time t. Consider a small time interval dt during which the price of the underlying asset S changes by an amount dS. Then the value of underlying asset S follows a stochastic differential equation (SDE)

$$dS = \mu S dt + \sigma S dW \tag{2.1}$$

where μ is the interest rate of risk free asset, σ the volatility of the stock and dW, the Brownian motion that follows normal distribution $N(0, \sqrt{t})$, with mean 0 and variance t. Let V be the value of the option given by $V = V(t, S)$ where $V(t, S)$ is twice differentiable in S and differentiable in t. Applying Ito's lemma we get

$$\begin{aligned} dV(S,t) &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(dS)^2 + \frac{1}{2}\frac{\partial^2 V}{\partial t^2}(dt)^2 + \frac{\partial^2 V}{\partial S \partial t}dt\ dS \\ &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\ dt \end{aligned}$$

After substituting for dS from equation (2.1), and using the fact that $lim\ dt \to 0$, $dt \cdot dS = 0$, $dt^2 = 0$, and $lim\ dS \to 0, dS^2 = dt$, the equation simplifies to

$$dV(S,t) = (\frac{\partial V}{\partial t} + \mu S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt + \sigma S \frac{\partial V}{\partial S} dW \tag{2.2}$$

Let us construct a portfolio of one option and -Δ of the underlying asset. The value of such a portfolio is $\Pi = V - \Delta S$ and Δ is held fixed between times t and $t + dt$.

Therefore the change in the value of such a portfolio during the time dt is given by:

$$\begin{aligned} d\Pi &= dV - \Delta dS \\ &= (\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2})dt + \sigma S\frac{\partial V}{\partial S}dW - \Delta(\mu S dt + \sigma S dW) \\ &= (\frac{\partial V}{\partial t} + \mu S\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} - \Delta\mu S)dt + \sigma S(\frac{\partial V}{\partial S} - \Delta)dW \end{aligned}$$

We can eliminate the randomness in the above PDE by choosing $\Delta = \partial V/\partial S$. Then this portfolio will be totally risk-free. Since this portfolio contains no risk, any risk-less portfolio must earn the risk-less interest rate r. If it earned more, then one could make a profit by selling the risk free securities and using the proceeds to buy this portfolio. If the portfolio earned less, then one could make a risk less profit by selling the portfolio and buying the risk-free securities. Hence, we must have:

$$d\Pi = (\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2})dt = r\Pi dt = r(V - \frac{\partial V}{\partial S}S)dt \tag{2.3}$$

Thus we can rewrite equation (2.3) to get:

$$\begin{aligned} &\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} = r(V - S\frac{\partial V}{\partial S}) \\ &\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 V}{\partial S^2} - rV + rS\frac{\partial V}{\partial S} = 0 \end{aligned} \tag{2.4}$$

The resulting equation (2.4) with appropriate boundary and final conditions is a second order parabolic partial differential equation, called the Black-Scholes Partial Differential Equation. Since the value of the option is known at the time of expiration (final time T), the Black-Scholes equation is a final value problem. It has a closed form solution given by:

$$V(S, K, \tau, \sigma) = \begin{cases} SN(d_1) - Ke^{-r\tau}N(d_2) & \text{for call option} \\ -SN(-d_1) + Ke^{-r\tau}N(-d_2) & \text{for put option} \end{cases}$$

where $N(\cdot)$ is the cumulative distribution function for a standard normal random

variable, given by

$$N(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} exp(-\frac{z^2}{2})dz.$$

$$\text{and} \qquad d_1 = \frac{\ln(S/K) + \tau(r + \frac{\sigma^2}{2})}{\sigma\sqrt{\tau}}, \quad d_2 = \frac{\ln(S/K) + \tau(r - \frac{\sigma^2}{2})}{\sigma\sqrt{\tau}},$$

$$\tau = T - t.$$

The derivation of the closed form solution of Black-Scholes equation is straight forward and is found in a variety of literature in finance and economics [10], [11], [48] and many others for its full derivation.

The Black-Scholes model derived from Brownian motion has gained a great success among practitioners because it has a closed form solution and the PDE analysis allows a simple numerical computation to the problem. In spite of its simplicity, the model ignores some of the features like market jumps and crashes which are rare but significant events. The Brownian motion implies continuity of the market evolution, but in reality price process can jump. Also, the real market prices do not appear to be distributed log-normally as assumed but have heavier tails and asymmetric features. In addition, the Black-Scholes model assumes a constant volatility, but the implied volatility curve does not represent constant flat lines as assumed but resembles a *smile*.

2.3 IMPLIED VOLATILITY AND VOLATILITY SMILES

Option prices obtained from the Black-Scholes model are functions of the following five parameters: the time t, the strike price K, the risk-free rate r, the current underlying price S and the market volatility σ. The only unknown parameter that is not directly observed from the market is the volatility of the market σ and it has to be estimated. Estimation of volatility from the historical data of the underlying is called the *historical volatility*. Since there is only one unknown parameter in the

Black-Scholes model namely the volatility σ, there is a one-to-one correspondence between the value of any financial derivative contract, such as an option, and the volatility of its underlying asset. In general, the more volatile the asset, the more the derivative contract is worth. Thus, when a market has set the price for a contract, it is often the case that this price corresponds to a unique implied volatility. Suppose we use the Black-Scholes model to infer the volatility used by option traders to price the option. We search for a volatility such that the model represents an option price that corresponds to the market price. The volatility obtained this way is called the *implied volatility*. The Black-Scholes model assumes that the implied volatility is

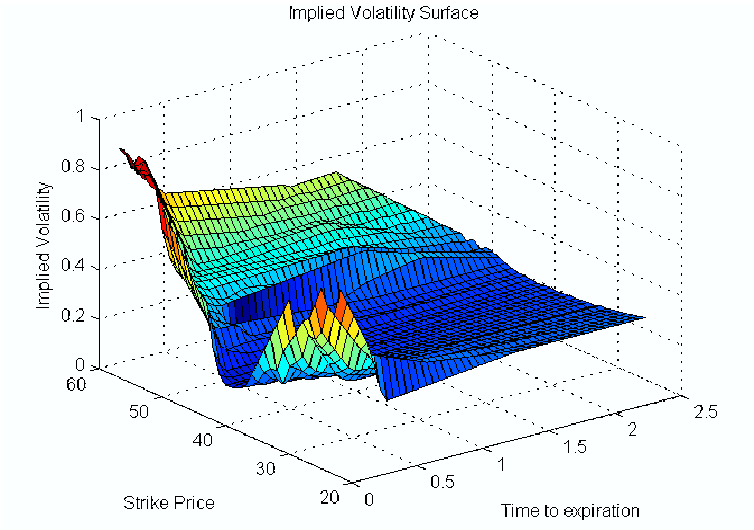


Figure 2.2. Implied volatility surface generated by Black-Scholes model for a particular stock trading at \$40.

constant and homogeneous for options on the same underlying with different strikes and maturities. However, in practice the implied volatility of call or put options at a given time t is a function of the strike price and the time to maturity $T-t$. Often deeply out of the money or deeply in the money options have significantly higher implied volatilities than options at the money. The graph of the strike price and the implied volatility forms a shape of a smile called the volatility smile. The collection of implied volatilities for different strikes and maturities is called the implied

volatility surface. Figure 2.2 shows an implied volatility surface for a particular stock obtained by using Black-Scholes formula. We can see that for options close to expiration that are either highly in the money or highly out of the money have high implied volatility and give the smile effect. As the time to expiration increases, the smile effect eventually smooths out.

Black-Scholes model being a constant volatility model is unable to explain the volatility smile and the volatility surface caused by this discrepancy. The shortcoming of the Black-Scholes model has led to a considerable amount of research and alternative models that attempt to explain the dynamics of the underlying asset in terms of alternative distributions which match a representation of the implied volatility surface. A first outstanding attempt to overcome the lack of fit was made by Merton [37]. He added a jump component to the continuous model, where the jumps are counted by a Poisson process and the jump amplitude is log normally distributed. Merton's model also has a closed form solution. In recent years different versions of jump diffusion models have been proposed by different authors and all of these models require numerical solutions. Heston [24] proposed a stochastic volatility model by adding more parameters to the model. In his model, volatility is assumed to be stochastic. In recent years there have been attempts to combine jumps with the stochastic volatility models and also to derive the jump diffusion models through Lévy process with a different jump distribution.

However, none of the more complex models have been very popular among the practitioners due to the fact that they are computationally expensive, complex, and lack the market information needed to calibrate the models. We will now explore some basics of stochastic volatility and jump diffusion models.

2.4 STOCHASTIC VOLATILITY MODELS

2.4.1 Introduction

In this chapter we discuss continuous-time stochastic volatility models. By this we mean two-dimensional diffusion models where only one of the coordinates is observable and where the stochastic differential equation has a special form. The models were introduced in the financial mathematics literature in the late eighties as modifications of the classical Black-Scholes model.

The Black-Scholes model is a constant volatility model. An alternative theory to the Black-Scholes model considers the volatility to be stochastic process. Stochastic volatility models in finance have been extensively studied by Wiggins [47] , Hull and White [25], Hull [26], Stein and Stein [43], Heston [24], Bates [8], Lewis [31], Bakshi, Cao and Chen [6] giving rise to the whole family of stochastic volatility models. In this section we will consider a general stochastic volatility model and derive a partial different equation arising from the Heston stochastic volatility model. A stochastic volatility model introduces more random sources then traded assets, hence the model is incomplete. Pricing in a market with stochastic volatility is thus an incomplete market problem, which means that the derivative cannot be perfectly hedged. Heston [24] obtained an analytical formula for the option price starting from the characteristic function of the risk neutral probability and the inverse of the Fourier transform.

Stochastic Volatility Models are widely used in the financial market as a refinement of the Black-Schole's model. They explain why the options with different strikes and expirations have different Black-Scholes implied volatilities and have the *smile effect*. In particular, traders who use the Black- Scholes model to hedge must continuously change the volatility assumption in order to match market prices. Their hedge ratios change accordingly in an uncontrolled way. More interestingly for us, the prices of exotic options given by models based on Black-Scholes assumptions can be wrong

and dealers in such options are motivated to find models which can take the volatility smile into account when pricing these options. In stochastic volatility models the value of an option is usually specified by a partial differential equation. Since closed-form solutions are often unavailable we have to rely on numerical techniques to estimate the value of an option.

2.4.2 Model motivation

Let S be the value of an underlying asset such as a stock at time t. The value of underlying asset S follows a stochastic differential equation (SDE)

$$dS = \mu S dt + \sigma S dW_1 \tag{2.5}$$

where μ is the average rate of growth, σ the volatility of the stock and dW_1, the Brownian motion that follows normal distribution $N(0, \sqrt{t})$, with mean 0 and variance t. Volatility of the underlying asset σ is given by the square root of the variance ν, ie $\nu = \sigma^2$.

To get to Heston's PDE we follow the derivation done by Gatheral [22]. Let the variance be defined by another stochastic process

$$d\nu = \alpha(S, \nu, t)\ dt + \xi\ \beta(S, \nu, t)\ \nu^{\gamma}\ dW_2 \tag{2.6}$$

where α and β will be defined later. The covariance of two stochastic processes is given by $\mathrm{Cov}\langle dW_1, dW_2\rangle = \rho dt$ and ξ is defined as the volatility of volatility. If $V(S, \nu, t)$ is twice differentiable function in S and ν and differentiable in t, then applying Ito's lemma we get

$$\begin{aligned} dV = \frac{\partial V}{\partial t}\ dt + \frac{\partial V}{\partial S}\ dS + \frac{\partial V}{\partial \nu}\ d\nu + \frac{\partial^2 V}{\partial S \partial \nu}\ \mathrm{Cov}(dS, d\nu) \\ + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\ Var(dS) + \frac{1}{2}\frac{\partial^2 V}{\partial \nu^2} Var(d\nu). \end{aligned} \tag{2.7}$$

Similarly, applying Ito's lemma in equation (2.5) we get

$$\begin{aligned}(dS)^2 &= \mu^2\ S^2\ (dt)^2 + 2\mu\ S^2\ dt\sqrt{\nu}\ dW_1 + \nu\ S^2\ (dW_1)^2 \\ &= \nu S^2 (dW_1)^2 = \nu S^2 dt\end{aligned} \tag{2.8}$$

Similarly applying Ito's lemma in equation (2.6) we get

$$\begin{aligned}(d\nu)^2 &= \alpha^2\ dt^2 + \xi^2\beta^2\ \nu^{2\gamma}(dW_2)^2 + 2\alpha\ dt\ \xi\ \beta\nu^{\gamma}\ (dW_2) \\ &= \xi^2\ \beta^2\ \nu^{2\gamma}\ dt\end{aligned} \tag{2.9}$$

Since $\mu, S, \xi, \alpha, \beta$ are all known at time t,

$$\begin{aligned}\mathrm{Cov}(\mu S dt + \sqrt{\nu} S\ dW_1, \alpha dt + \xi\beta\nu^{\gamma}\ dW_2) &= \mathrm{Cov}(\sqrt{\nu} S\ dW_1, \xi\beta\nu^{\gamma}\ dW_2) \\ &= \nu^{(\gamma+\frac{1}{2})} S\xi\beta\rho dt\end{aligned} \tag{2.10}$$

Substituting equations (2.8), (2.9) and (2.10) in (2.7) we get:

$$\begin{aligned}dV &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{\partial V}{\partial \nu}d\nu + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho\frac{\partial^2 V}{\partial S \partial \nu}dt + \\ &\quad \frac{1}{2}\nu S^2\frac{\partial^2 V}{\partial S^2}dt + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma}\frac{\partial^2 V}{\partial \nu^2}dt \\ &= \left(V_t + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma}V_{\nu\nu}\right)dt + V_S dS + V_\nu d\nu\end{aligned} \tag{2.11}$$

Let Π be the portfolio containing the option whose value is $V(t, S, \nu)$, the quantity $-\Delta$ of stock and $-\Delta^1$ of another asset whose value V^1 depends on volatility i.e $V^1 = V^1(t, S, \nu)$. Therefore

$$\Pi = V - \Delta S - \Delta^1 V^1 \tag{2.12}$$

Taking the derivative on both sides we get:

$$\begin{aligned}
d\Pi &= dV - \Delta dS - \Delta^1 dV^1 \\
&= \left(V_t + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma} V_{\nu\nu}\right) dt + V_S dS + V_\nu d\nu - \Delta dS \\
&\quad -\Delta^1[\left(V_t^1 + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu}^1 + \frac{1}{2}\nu S^2 V_{SS}^1 + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma} V_{\nu\nu}^1\right) dt + V_S^1 dS + V_\nu^1 d\nu] \\
&= \left(V_t + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma} V_{\nu\nu}\right) dt \\
&\quad -\Delta^1 \left(V_t^1 + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu}^1 + \frac{1}{2}\nu S^2 V_{SS}^1 + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma} V_{\nu\nu}^1\right) dt \\
&\quad +(V_S - \Delta - \Delta^1 V_S^1) dS + (V_\nu - \Delta^1 V_\nu^1) d\nu \qquad (2.13)
\end{aligned}$$

To make this portfolio instantaneously risk free, we must eliminate the risky terms containing S and ν by setting each part to 0 separately. Hence, we have

$$V_\nu - \Delta^1 V_\nu^1 = 0 \text{ and } (V_S - \Delta - \Delta^1 V_S^1) = 0$$

So, from the first condition, we have $\Delta^1 = \frac{V_\nu}{V_\nu^1}$ and from the second condition also have

$$V_S - \Delta - \frac{V_\nu}{V_\nu^1} V_S^1 = 0, \text{ or } \Delta = V_S - \frac{V_\nu}{V_\nu^1} V_S^1$$

For the portfolio (2.12) to be risk free, its return has to be same as the return from risk free rate r. Hence,

$$d\Pi = r\Pi dt = r(V - \Delta S - \Delta^1 V^1) dt$$

From equation (2.13) and with substituting appropriate values for Δ and Δ^1 as given above we get:

$$\begin{aligned}
r(V - \Delta S - \Delta^1 V^1) dt &= (V_t + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma} V_{\nu\nu}) dt - \\
&\quad \frac{V_\nu}{V_\nu^1}(V_t^1 + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu}^1 + \frac{1}{2}\nu S^2 V_{SS}^1 + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma} V_{\nu\nu}^1) dt \qquad (2.14)
\end{aligned}$$

Collecting all the terms containing V on one side and V^1 on the other side we get:

$$\begin{aligned}&\frac{V_t + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma}V_{\nu\nu} - rV + rSV_S}{V_\nu} \\ &= \frac{V^1_t + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V^1_{S\nu} + \frac{1}{2}\nu S^2 V^1_{SS} + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma}V^1_{\nu\nu} - rV^1 + rSV^1_S}{V^1_\nu}\end{aligned} \tag{2.15}$$

We see that the left side depends only on V while the right side depends only on V^1. This is only possible if there is a function f that depends on the variables S, ν, t such that the above quantity is equal to f. So,

$$\frac{V_t + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma}V_{\nu\nu} - rV + rSV_S}{V_\nu} = -(\alpha - \phi\beta) = f(t, S, \nu)$$

where $\beta \neq 0$. Rearranging the terms we get,

$$V_t + \nu^{\gamma+\frac{1}{2}} S\xi\beta\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\beta^2\nu^{2\gamma}V_{\nu\nu} - rV + rSV_S = -(\alpha - \phi\beta)\frac{\partial V}{\partial \nu} \tag{2.16}$$

Conventionally, $\phi(t, S, \nu)$ is called the market price of volatility risk. This tells us how much of the expected return of V is explained by the risk of ν in the asset pricing.

If we replace $\alpha = -\lambda(\nu(t) - \bar{\nu})$, $\beta = 1$ and $\gamma = 1/2$ in equation (2.6) we get the ***Heston Model***

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sqrt{\nu(t)}S(t)dW_1 \\ d\nu(t) &= -\lambda(\nu(t) - \bar{\nu})dt + \xi\sqrt{\nu(t)}dW_2 dt \end{aligned}$$

where, λ is called the speed of reversion of variance $\nu(t)$ to its long term mean $\bar{\nu}$. If we replace α and β with the new values from the Heston Model in equation (2.16), we get:

$$V_t + \nu S\xi\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\nu V_{\nu\nu} - rV + rSV_S = (\lambda(\nu(t) - \bar{\nu}) - \phi)V_\nu \tag{2.17}$$

Various models suggest that the market price of volatility risk ϕ is proportional to volatility. Therefore we can set $\phi = \theta\nu$ for some constant θ.

Define risk adjusted parameters $\acute{\lambda}$ and $\acute{\bar{\nu}}$ as $\acute{\lambda} = \lambda - \theta$ and $\acute{\bar{\nu}} = \lambda\bar{\nu}/\acute{\nu}$ sothat $\lambda = \acute{\lambda} + \theta$ and $\bar{\nu} = \acute{\lambda}\acute{\bar{\nu}}/\lambda$. Therefore,

$$\begin{aligned}
\lambda(\nu - \bar{\nu}) - \phi &= \lambda(\nu - \frac{\acute{\lambda}\acute{\bar{\nu}}}{\lambda}) - \theta\nu \\
&= \lambda\nu - \acute{\lambda}\acute{\bar{\nu}} - \theta\nu \\
&= (\acute{\lambda} + \theta)\nu - \acute{\lambda}\acute{\bar{\nu}} - \theta\nu \\
&= \acute{\lambda}(\nu - \acute{\bar{\nu}})
\end{aligned}$$

Since, ϕ does not appear in the new equation, new parameters are called the risk adjusted parameters. So the above equation (2.17) can be written in terms of risk adjusted parameters as:

$$\begin{aligned}
V_t + \nu S\xi\rho V_{S\nu} + \frac{1}{2}\nu S^2 V_{SS} + \frac{1}{2}\xi^2\nu V_{\nu\nu} - rV + rSV_S &= \acute{\lambda}(\nu(t) - \acute{\bar{\nu}})V_\nu \\
&= \lambda(\nu - \bar{\nu})V_\nu \qquad (2.18)
\end{aligned}$$

Now, let $x = \ln(S)$. Then $dx/dS = 1/S$ and $d^2x/dS^2 = -1/S^2$ Then we have:

$$\begin{aligned}
\frac{\partial V}{\partial S} &= \frac{1}{S}V_x \\
\frac{\partial^2 V}{\partial S^2} &= \frac{V_{xx}}{S^2} - \frac{V_x}{S^2} = \frac{1}{S^2}(V_{xx} - V_x)
\end{aligned}$$

If we consider only the log transformed values of stock prices and write it in terms of variable x as defined above and substitute these values in equation (2.18) we get:

$$V_t + \frac{1}{2}\nu V_{xx} + (r - \frac{1}{2}\nu)V_x + \nu\xi\rho V_{x\nu} + \frac{1}{2}\xi^2\nu V_{\nu\nu} - \lambda(\nu - \bar{\nu})V_\nu - rV = 0 \qquad (2.19)$$

Using the conditions given by Apel, Winkler and Wystup [4] equation (2.19) has the final conditions for European options as

$$V(T, x, \nu) = \begin{cases} max\left[e^x - K, 0\right] & \text{for call options} \\ max\left[K - e^x, 0\right] & \text{for put options} \end{cases}$$

and the boundary conditions for put options given by

$$V(t, -\infty, \nu) = Ke^{-r\tau} \quad V(t, \infty, \nu) = 0$$

$$V(t, x, \infty) = Ke^{-r\tau}$$

By analogy to Black Scholes formula for European option with strike price K and time to maturity τ, Heston guessed the solution of the call option of the form

$$V(\tau, x, \nu) = e^x P_1 - Ke^{-r\tau} P_2 \tag{2.20}$$

where P_1 and P_2 are unknown functions such that they represent similar probabilities as in the case of Black Scholes model. P_1 and P_2 are calculated through the Fourier transformation of the characteristic function. For detail of the derivation and justification readers are referred to [22], [24] and [31].

For the Heston partial differential equation (2.19), we can find its solution via the Fourier transform method. The solution however can not be expressed in terms of elementary functions. It has to be computed by numerical integration.

Although stochastic volatility model is based on more realistic assumption about the volatility and prices derivatives better than the traditional Black-Scholes model, the model is mathematically complex. For many stochastic volatility models, closed-form solutions are not available and numerical computation is expensive. The Heston model does have a closed form solution but it still requires numerical computation to find the probabilities, P_1 and P_2.

The Black-Scholes and the Heston model assume that the evolution of stock price follows a continuous Brownian motion. In reality, the market responds drastically to some unexpected news and market jump or crash is seen once in a while. These events can not be explained solely by the Brownian motion. To explain the jump phenomenon that appears frequently in the market we have another class of models

called the jump diffusion models. We are reviewing few of the popular derivative pricing models with unknown parameters. We will use these models to derive an efficient technique to estimate gradient of the cost function that is needed for parameter estimation in gradient based optimization.

2.5 JUMP DIFFUSION MODEL

2.5.1 Introduction

To price financial derivatives under actual market conditions, more complex models are required. In this section we will present and motivate a jump diffusion model where the asset prices evolve by jumps and crashes added to the standard Black-Scholes diffusion driven by Brownian motion. The jump diffusion model looks more realistic then other previously defined models. It adds the idea of sudden price movement caused by discrete economic events such as surprised high earnings, winning a major government contract or a sudden terrorist attack to the idea of an otherwise efficient market driven by Brownian motion. This assumption leads to the pricing equation to be a parabolic, partial integro-differential equation (PIDE). The jumps in the logarithm of the prices may be distributed by any finite activity process such as the the normal distribution in the case of Merton's model and the exponential or double exponential distributions in the case of Kou's model.

2.5.2 Motivation of the integral term

Let stock S follow the geometric Brownian Motion equation given by

$$dS = \mu S dt + \sigma S dW \tag{2.21}$$

From Ito's lemma, after log transformation of the stock price we get

$$d(\log S) = (\mu - \frac{\sigma^2}{2})dt + \sigma dW \tag{2.22}$$

Suppose that we have asset price at discrete times t_i, such that $S(t_i) = S_i$ and $\Delta t = t_{i+1} - t_i$. Then from (2.22) we have

$$\log S_{i+1} - \log S_i = \log \frac{S_{i+1}}{S_i} \approx (\mu - \frac{\sigma^2}{2})\Delta t + \sigma\phi\sqrt{\Delta t} \tag{2.23}$$

where ϕ is $N(0,1)$. If Δt is very small, then Δt is much smaller than $\sqrt{\Delta t}$. After ignoring the terms with Δt equation (2.23) can be written as

$$\log\left(\frac{S_{i+1} - S_i + S_i}{S_i}\right) = \log\left(1 + \frac{S_{i+1} - S_i}{S_i}\right) \approx \sigma\phi\sqrt{\Delta t} \tag{2.24}$$

Let return in the period $[t_{i+1}, t_i]$ be R_i. Then $R_i = \dfrac{S_{i+1} - S_i}{S_i}$ and equation (2.24) becomes $\log(1 + R_i) \approx R_i = \sigma\phi\sqrt{\Delta t}$.

Definition 2.1. A random process $X(t)$ is said to be a counting process if $X(t)$ represents the total number of events that have occurred in the time interval $(0, t)$. A counting process must satisfy the following conditions:

1. $X(t) \geq 0,\ X(0) = 0$
2. $X(t)$ is integer valued
3. $X(s) < X(t)$ if $s < t$
4. $X(t) - X(s)$ equals the number of events that have occurred in the interval (s, t)

Definition 2.2. A Poisson process is a counting process with intensity $\lambda > 0$ if

1. $X(0) = 0$
2. $X(t)$ has independent and and stationary increments
3. $P[X(t + dt) - X(t) = 1] = \lambda dt + o(dt)$ where P is the probability

4. $P[X(t+dt) - X(t) \geq 2] = o(dt)$ where

$$\lim_{dt \to 0} \frac{o(dt)}{dt} = 0$$

In the contest of our problem, we can define the Poisson process dq as follows

$$dq = \begin{cases} 0 & \text{with probability } 1 - \lambda dt \\ 1 & \text{with probability } \lambda dt \end{cases}$$

where λ is called the Poisson arrival intensity. The jumps of the Poisson process are not at all deterministic. The process $P(t)$ is stochastically continuous such that there is zero probability of having a jump at any deterministic time, i.e. $\lim_{dt \to 0} P(t + dt) - P(t) = 0$. Now, the mean and variance of dq is

$$\begin{aligned} E(dq) &= \lambda dt \cdot 1 + (1 - \lambda dt) \cdot 0 = \lambda dt \\ Var(dq) &= E[(dq - E(dq))^2] = E[(dq - \lambda dt)^2] \\ &= (1 - \lambda dt)^2 \lambda dt + (0 - \lambda dt)^2 (1 - \lambda) dt = \lambda dt + O(dt)^2 \end{aligned}$$

Consider a situation where most of the time nothing happens, but occasional jumps in stock price. The jump size is independent of time interval but the probability of jump occurring depends on time. Suppose a jump occurs in $[t, t + \Delta t]$ with probability λdt such that the stock moves to a new value $S \to JS$ where J is the size of the positive jump. Then the change due to a jump in the stock price is given by $[dS]_{jump} = (J - 1)Sdq$. If a jump occurs then

$$\begin{aligned} S_{(after\ jump)} &= S_{(before\ jump)} + (dS)_{(jump)} \\ &= S_{(before\ jump)} + (J - 1) S_{(before\ jump)} \\ &= JS_{(before\ Jump)} \end{aligned}$$

So, if stock price follows a combination of Brownian Motion and rare jump events,

then change in stock price is given by

$$dS = \underbrace{\mu S dt + \sigma S dW} + \underbrace{(J-1)S dq}$$

The first part of the equation is due to Brownian motion and the second part is due to jump. Assume that the jump size has some known probability density $g(J)$. Given that a jump occurs, the probability of a jump in $[J, J+dJ]$ is $g(J)dJ$. Also, $\int_{-\infty}^{\infty} g(J)dJ = \int_{0}^{\infty} g(J)dJ = 1$, and we consider only the positive jumps such that $g(J) = 0$ if $J < 0$. If $f = f(J)$, then expected value of f is

$$E(f) = \int_0^\infty f(J)g(J)dJ \tag{2.25}$$

2.5.3 The pricing equation for finite jumps

Suppose we have one option worth V and Δ shares at price S. If Π is the value of the portfolio then $\Pi = V - \Delta S$. Consider the change in the value of portfolio

$$(d\Pi)_{total} = [d\Pi]_{brownian} + [d\Pi]_{jump} \tag{2.26}$$

From Ito's lemma

$$[d\Pi]_{brownian} = \left[V_t + \mu S V_s + \frac{\sigma^2 S^2}{2} V_{ss} - \Delta \mu S\right] dt + \sigma S \left[V_s - \Delta\right] dW \tag{2.27}$$

Noting that jump is of finite size,

$$[d\Pi]_{jump} = [V(JS,t) - V(S,t)]dq - \Delta(J-1)Sdq \tag{2.28}$$

If we now hedge Brownian Motion risk by choosing $\Delta = V_s$, then equation (2.26) gives us

$$d\Pi = \left[V_t + \frac{\sigma^2 S^2}{2} V_{ss}\right] dt + \left[V(JS,t) - V(S,t)\right] dq - V_s(J-1)S\ dq \tag{2.29}$$

The change in the value of portfolio still has a random component dq which can't be hedged away. So, we take the expectation on both sides to get

$$E[d\Pi] = E(\left[V_t + \frac{\sigma^2 S^2}{2} V_{ss}\right] dt) + E\left[V(JS,t) - V(S,t)\right] E[dq] - V_s\ S\ E[(J-1)]E[dq] \tag{2.30}$$

We have assumed that the probability of jump and probability of jump sizes are independent. Now, define $E[J-1] = \kappa$, then equation (2.30) becomes

$$E[d\Pi] = \left[V_t + \frac{\sigma^2 S^2}{2} V_{ss}\right] dt + E\left[V(JS,t) - V(S,t)\right] \lambda dt - V_s S \kappa \lambda \, dt \tag{2.31}$$

Assume that investor holds a diversified portfolio of hedging portfolios for different stocks. We assume that the jumps for these portfolios are uncorrelated and the variance of the portfolio of portfolios is small (there is little risk). The expected return should be

$$E[d\Pi] = r\Pi dt \tag{2.32}$$

Now equating equations (2.31) and (2.32) we get:

$$V_t + \frac{\sigma^2 S^2}{2} V_{ss} + (rS - S\kappa\lambda)V_s - (r+\lambda)V + E[V(JS,t)]\lambda = 0 \tag{2.33}$$

Using equation (2.25) in (2.33) gives

$$V_t + \frac{\sigma^2 S^2}{2} V_{ss} + (rS - S\kappa\lambda)V_s - (r+\lambda)V + \lambda \int_0^\infty g(J)V(JS,t)dJ = 0 \tag{2.34}$$

Equation (2.34) is called a Partial Integro Differential Equation (PIDE) that models jump diffusion in option pricing. Here, $g(J)$ can follow any appropriate probability density function. For example Merton's model assumes $g(J)$ to be log-normal whereas Kou's model assumes it to be double exponential. More careful derivation of jump models with Lévy process can be found in Cont and Tankov [17] and is not repeated here. Other sources for jump diffusion models and its derivation include Bates [8], Cont and Voltchkova [19], Kuo [29] and many others.

Since additional parameters λ, κ, J and the jump distribution $g(J)$ in equation (2.34) are all unknown in jump diffusion model the parameter estimation problem become more complex. It is important to estimate these parameters correctly and efficiently to price the financial derivatives correctly.

In this chapter we have given the basic insight to the three types of option pricing

models. After finding a good pricing model it is important to be able to find the solution of the pricing equation efficiently. Although Black-Scholes and the Heston PDE have analytical solution in a closed form, they can be solved numerically efficiently. Since the PIDE have no closed form solution, they have to be solved numerically. Next chapter highlights some of the popular numerical methods used to solve pricing problems.

CHAPTER 3

NUMERICAL METHOD FOR OPTION PRICING

FINITE DIFFERENCE METHOD

Solving partial differential equations (PDEs) is of concern to all areas of applied analysis from physics to finance. Most PDEs encountered in practice do not have analytical solutions and have to be solved numerically. Numerical solution of PDEs is very extensive and quite advanced. In this section, we try to understand the basics of one of the numerical methods, the finite difference method, used to solve the types of PDE that arises in finance.

The most common finite-difference methods for solving the diffusion type equation seen in finance are the explicit method, the fully implicit method, and the Crank-Nicolson method. We will briefly describe each of the methods in relation to solving Black-Scholes equation. We will also introduce explicit-implicit method (a combination of implicit and explict methods) to solve a Partial Integro-Differential Equation (PIDE) that arises from Jump Diffusion process. Finally we will solve the PIDE derived from Merton's Jump diffusion model using splitting technique as done in [27] .

3.1 INTRODUCTION TO FINITE DIFFERENCE

Finite Difference Method are capable of evaluating many types of linear and non-linear Partial and Ordinary Differential Equations. The finite difference method consists of transforming the partial derivatives in difference equations over a small interval. This method attempts to solve partial differential equations by approximating the partial derivatives using an explicit, an implicit or some combination of both of the methods. Finite difference method can produce accurate numerical solutions to many partial differential equations including those arising in finance.

In finite difference methods, we replace the partial derivatives by approximations based on Taylor series expansions of functions near the point of interest.

Lemma 3.1.1. *Let $u(\tau, x)$ be a function of independent variables x and τ. If $u(\tau, x)$ is continuously differentiable with Lipschitz continuous derivative then partial derivative $\frac{\partial u}{\partial \tau}$ approximated by*

$$\frac{\partial u}{\partial \tau} = \frac{u(\tau + \delta\tau, x) - u(\tau, x)}{\delta\tau} + O(\delta\tau) \tag{3.1}$$

is called the forward finite difference approximation and the partial derivative $\frac{\partial u}{\partial \tau}$ approximated by

$$\frac{\partial u}{\partial \tau} = \frac{u(\tau, x) - u(\tau - \delta\tau, x)}{\delta\tau} + O(\delta\tau) \tag{3.2}$$

is called the backward finite difference approximation.

Lemma 3.1.2. *Let $u(\tau, x)$ be a function of independent variables x and τ. If $u(\tau, x)$ is continuously differentiable with Lipschitz continuous derivative then partial derivative $\frac{\partial u}{\partial \tau}$ approximated by*

$$\frac{\partial u}{\partial \tau} = \frac{u(\tau + \delta\tau, x) - u(\tau - \delta\tau, x)}{2\delta\tau} + O((\delta\tau)^2) \tag{3.3}$$

is called the central finite difference approximation.

The proof of the above lemma comes directly from Taylor expansion of the function $u(\tau, x)$.

When applied to diffusion equation, forward difference in $\frac{\partial u}{\partial \tau}$ leads to an explicit finite difference scheme and a backward finite difference scheme leads to a fully implicit scheme.

Finite difference of $u(\tau, x)$ in terms of the variable x is also defined exactly the same way as stated by the above lemma. The central difference approximation with respect to x is give by

$$\frac{\partial u}{\partial x} \approx \frac{u(\tau, x + \delta x) - u(\tau, x - \delta x)}{2\delta x} + O((\delta x)^2). \tag{3.4}$$

Similarly, the symmetric central difference for the second derivative $\frac{\partial^2 u}{\partial x^2}$ is approximated as

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u(\tau, x+\delta x) - 2u(\tau, x) + u(\tau, x-\delta x)}{(\delta x)^2} + O((\delta x)^2). \tag{3.5}$$

Assume that u is a function of independent variables x and τ. For finite difference method, we subdivide the $x-\tau$ plane into sets of equal rectangles of sides $\delta x = h, \delta\tau = k$ by equally spaced grid lines. Suppose, x is divided into N equally spaced units of length h and τ is divided into M equally spaced units of length k, then we will have grids of the form

$$x_j = jh \text{ for } j = 0, 1, 2, ..., M \text{ and } \tau_i = ik \text{ for } i = 0, 1, 2, ..., N.$$

Let the value of u at the representative mesh point $P(ik, jh)$ be given by

$$u_P = u(ik, jh) = u_{i,j}.$$

Then we can write finite difference quotients defined in (3.1) , (3.4) and (3.5) as:

$$\begin{aligned} \frac{\partial u}{\partial \tau} &\approx \frac{u_{i+1,j} - u_{i,j}}{k} \\ \frac{\partial u}{\partial x} &\approx \frac{u_{i,j+1} - u_{i,j-1}}{2h} \\ \frac{\partial^2 u}{\partial x^2} &\approx \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h^2} \end{aligned} \tag{3.6}$$

with leading error terms of $O(h^2),\ O(h^2)$ and $O(k)$ respectively.

3.2 FINITE DIFFERENCE METHOD FOR THE BLACK-SCHOLES PDE

Consider the Black-Scholes PDE with stock price S given by

$$\begin{aligned} \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0 \qquad (3.7) \\ V(T, S) = \max\ (0, S - K), \\ V(t, 0) = 0, \\ V(t, S) \approx S - Ke^{-r(T-t)} \text{ for large S} \end{aligned}$$

for European call options where K is the strike or exercise price of the option. This is a final value parabolic PDE with variable coefficients. To make it an initial value problem with constant coefficient we need to make some transformations.

Lets introduce a time variable $\tau = T - t$ and $x = \ln(S)$. Then equation (3.7) in terms of new variables τ and x such that $V(t, S) = U(\tau, x)$ is given by:

$$\frac{\partial U}{\partial \tau} - \frac{1}{2}\sigma^2 \frac{\partial^2 U}{\partial x^2} - (r - \frac{1}{2}\sigma^2)\frac{\partial U}{\partial x} + rU = 0 \tag{3.8}$$

$$U(0, x) = \max\,(0, e^x - K),$$

$$U(\tau, x_{min}) = 0,$$

$$U(\tau, x) = e^x - Ke^{-r\tau}, \text{ for large } x.$$

x_{min} in equation (3.8) is the truncated minimum value of the log stock price. The value of European call option is given by the solution of the partial differential equation (3.8). Equation (3.8) can be solved using appropriate finite difference method.

3.2.1 Explicit finite difference method

Consider the discretization of space variable x into M equally spaced units of size $\delta x = h$ and the time variable τ into N equally spaced units of $\delta\tau = k$. The discretization of the log asset price x and time τ can be written as:

$$x_j = x_{min} + jh \text{ for } j = 0, 1, 2, ..., M \text{ and } \tau_i = ik \text{ for } i = 0, 1, 2, ..., N.$$

Note that $x_0 = x_{min}$ and $x_M = x_{max}$ are truncated approximations for minimum and maximum log stock prices. Substituting central difference for $\dfrac{\partial U}{\partial x}$ and $\dfrac{\partial^2 U}{\partial x^2}$, and forward difference for $\dfrac{\partial U}{\partial \tau}$in (3.8) , we get:

$$\frac{U_{i+1,j} - U_{i,j}}{\delta\tau} - \frac{\sigma^2}{2}\left(\frac{U_{i,j+1} - 2U_{i,j} + U_{i,j-1}}{(\delta x)^2}\right) - (r - \frac{\sigma^2}{2})\left(\frac{U_{i,j+1} - U_{i,j-1}}{2\delta x}\right) + rU_{i,j} = 0 \tag{3.9}$$

with initial and boundary conditions given by

$$U_{0,j} = \max\,(0, e^{x_j} - K) \text{ for } j = 0, ..., M,$$

$$U_{i,M} = e^{x_M} - Ke^{-r\tau_i} \text{ and } U_{i,0} = 0 \text{ for } i = 0, ..., N. \tag{3.10}$$

Since initial and boundary conditions are given for U, the above equation is explicitly solved for every value of $U(\tau_i, x_j)$ as :

$$U_{i+1,j} = aU_{i,j-1} + bU_{i,j} + cU_{i,j+1}$$

where $a = \left[\frac{\sigma^2}{2(\delta x)^2} - (r - \frac{\sigma^2}{2})\frac{1}{2\delta x}\right]\delta\tau$, $b = \left[1 - (\frac{\sigma^2}{(\delta x)^2} + r)\delta\tau\right]$
and $c = \left[\frac{\sigma^2}{2(\delta x)^2} + (r - \frac{\sigma^2}{2})\frac{1}{2\delta x}\right]\delta\tau$.

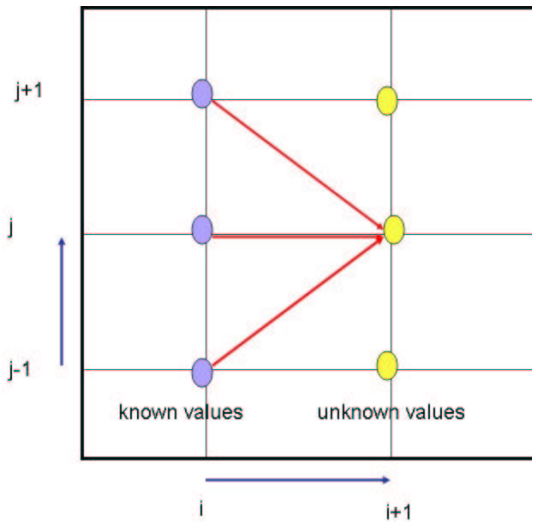


Figure 3.1. In explicit finite difference scheme, information at time step i is used to calculate the information at time step $i+1$.

3.2.2 Implicit finite difference method

Explicit finite difference method requires very small time steps $\delta\tau$ for it to be reasonably stable. The implicit finite difference method often overcomes the stability limitations of explicit methods. Implicit finite difference methods allows us to use large number of x-mesh without having to take very small time steps. Finite difference approximations for each of the derivatives in x is same as in explicit method

mentioned above but it uses the backward difference approximation for the $\frac{\partial U}{\partial \tau}$. This leads to:

$$\frac{U_{i+1,j} - U_{i,j}}{\delta\tau} - \frac{\sigma^2}{2}\left(\frac{U_{i+1,j+1} - 2U_{i+1,j} + U_{i+1,j-1}}{(\delta x)^2}\right) - (r - \frac{\sigma^2}{2})\left(\frac{U_{i+1,j+1} - U_{i+1,j-1}}{2\delta x}\right) + rU_{i+1,j} = 0 \tag{3.11}$$

with the initial and boundary conditions given by equation (3.10) and after ignoring the terms of $O(\delta\tau)$ and $O((\delta x)^2)$ as in the case of explicit method. Rearranging equation (3.11), implicit finite difference can be written in more compact form as

$$\bar{a}U_{i+1,j-1} + \bar{b}U_{i+1,j} + \bar{c}U_{i+1,j+1} = U_{i,j} \tag{3.12}$$

where $\bar{a} = \left[-\frac{\sigma^2}{2(\delta x)^2} + (r - \frac{\sigma^2}{2})\frac{1}{2\delta x}\right]\delta\tau$, $\bar{b} = \left[1 + (\frac{\sigma^2}{(\delta x)^2} + r)\delta\tau\right]$
and $\bar{c} = \left[-\frac{\sigma^2}{2(\delta x)^2} - (r - \frac{\sigma^2}{2})\frac{1}{2\delta x}\right]\delta\tau$.

From the initial condition $U_{0,j}$ is known for all j. From the boundary conditions $U_{i,M}$ and $U_{i,0}$ are also known. From equation (3.12) it is clear that the unknown values at any time index $i+1$ depends implicitly on the known values at time index i. Hence, with the given boundary and initial conditions, the unknown values are obtained by solving a linear system of equations at each time index i until we reach $\tau = T$. Therefore, for each time index $i = 1, 2, ..., N-1$, we solve the system

$$\begin{pmatrix} \bar{b} & \bar{c} & 0 & 0 & \dots & 0 \\ \bar{a} & \bar{b} & \bar{c} & 0 & \dots & 0 \\ 0 & \bar{a} & \bar{b} & \bar{c} & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \bar{a} & \bar{b} & \bar{c} \\ 0 & 0 & 0 & \dots & \bar{a} & \bar{b} \end{pmatrix} \begin{pmatrix} U_{i+1,1} \\ U_{i+1,2} \\ U_{i+1,3} \\ \vdots \\ U_{i+1,M-2} \\ U_{i+1,M-1} \end{pmatrix} = \begin{pmatrix} U_{i,1} \\ U_{i,2} \\ U_{i,3} \\ \vdots \\ U_{i,M-2} \\ U_{i,M-1} \end{pmatrix} - \begin{pmatrix} aU_{i,0} \\ 0 \\ 0 \\ \vdots \\ 0 \\ cU_{i,M} \end{pmatrix}$$

Writing the above system of equations in more compact form, we solve

$$\begin{aligned} AU_{i+1} &= U_i - y \\ \Rightarrow U_{i+1} &= A^{-1}(U_i - y) \end{aligned}$$

where A^{-1} is the inverse of matrix A. Practically inverting matrix A to find the solution requires $O(N^2)$ arithmetic operations and is not efficient for large systems. Instead, more efficient techniques like **LU** decomposition and **SOR** that requires $O(N)$ operations are preferred. For details of LU decomposition and SOR methods, readers are referred to [23], [48].

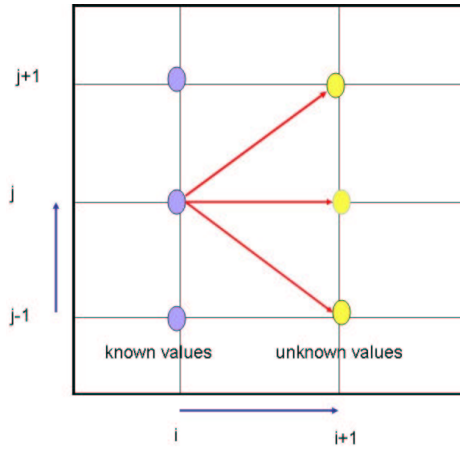


Figure 3.2. In implicit finite difference scheme, known information at time step i is expressed in terms of unknown information at time step $i+1$. A linear system of equations has to be solved to calculate the information at time step $i+1$.

3.2.3 Crank-Nicolson method

The Crank-Nicholson implicit finite difference method can be seen as the average of explicit and implicit methods. Taking the forward difference from equation (3.9)

and backward difference from equation (3.11) and averaging, we get:

$$\begin{aligned} &\bar{a}U_{i+1,j-1} + (1+\bar{b})U_{i+1,j} + \bar{c}U_{i+1,j+1} \\ &= aU_{i,j-1} + (1+b)U_{i,j} + cU_{i,j+1} \end{aligned} \tag{3.13}$$

where a, b, c and $\bar{a}, \bar{b}, \bar{c}$ are defined in the explicit and implicit finite difference methods above. Equation (3.13) can be written in matrix form as:

$$AU_{i+1} \quad = \quad BU_i$$

where U_i is the known value evaluated in the previous step and A and B are tridiagonal matrices. If we write the matrix vector product $BU_i = \mathbf{b}_i$, then we will have

$$U_{i+1} \quad = \quad A^{-1}\mathbf{b}_i$$

Both implicit and explicit finite difference methods have the rate of convergence of $O(\delta\tau)$. Averaging implicit and explicit schemes has a similar effect as making a central difference approximation at time $\frac{1}{2}(\tau_i + \tau_{i+1})$. Hence we improve the rate of convergence from $O(\delta\tau)$ to $O(\delta\tau^2)$. Crank-Nicolson method inherits the unconditional stability of the implicit method and is not more complex than the implicit method.

3.3 NUMERICAL ANALYSIS OF PIDE FOR JUMP PROCESS

Numerical methods for PIDEs occurring in Jump-diffusion models are quite new in mathematical finance. Although these PIDEs looks very similar to the Black-Scholes equation, they are harder to solve. The unknown solution $U(\tau, x)$ appears in both the differential and the integral term. Also, the PDE and the integral term is defined in an infinite interval. To solve it numerically the infinite interval has to be truncated to construct a finite difference mesh. After discretizing the Black-Scholes

PDE using implicit scheme, we get a tridiagonal system which is easy to invert. With the presence of integral term, discretizing the PIDE with implicit scheme leads to a full dense matrix which is hard to invert. Explicit methods suggested by Amin [2] and Zhang [50] suffer from instability problems and have poor convergence. Implicit methods which have better convergence and better stability properties are preferred for most finite difference methods but they involve the solution of a dense full matrix equations. Matache, Schwab, Wihler [36] use the wavelet compression technique to replace the dense matrix by a sparse matrix to solve the PIDE. The Alternating Directions Implicit (ADI) method used by Andersen and Andreasen, [3] requires computation of convolution and discrete Fourier transformation which is not straight forward to implement. More recent work by Cont and Voltchkova [19] use an explicit-implicit method and have studied the stability and convergence using the viscosity solution. La Choima [30] in her thesis suggests unconditionally stable and relatively easy to implement explicit-implicit finite difference scheme by truncating the integral domain. Recent work by Almendral and Oosterlee [1] and Ikonen and Toivanen [27] have suggested an efficient iterative techniques based on simple splitting of the dense matrix derived from Merton's and Kou's model and the use of fast Fourier transformation to help speed up the process.

In the next section, we will describe two efficient methods to solve the PIDE derived from the jump process. The method proposed by Cont and Voltchkova [19] and Lachioma [30] is called the explicit-implicit finite difference scheme. This scheme solves a system of equations without having to invert a dense matrix. The method proposed by Almendral and Oosterlee [1] uses implicit discretization of the PIDE on uniform grid and use splitting technique combined with FFT to accelerate matrix vector product to solve a system of equation.

3.3.1 Transformation of PIDE

Consider the option pricing PIDE based on jump diffusion model derived in chapter 2. For call options

$$\begin{aligned} \frac{dV}{dt} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - \kappa\lambda) S \frac{\partial V}{\partial S} - (r + \lambda) V + \\ \lambda \int_0^\infty V(t, JS) g(J)\, dJ = 0 \\ V(T, S) = \max\,(0, S - K), \\ V(t, 0) = 0, \\ V(t, S) \approx S - K e^{-r(T-t)} \text{ for large S.} \end{aligned} \tag{3.14}$$

where,

- λ is the jump intensity,
- J is the jump size,
- $g(J) \geq$ and $\int_0^\infty g(J) dJ = 1$.

Unlike the Black-Scholes equation, equation (3.14) has no closed form solution and even the numerical procedure is not straight forward. We still try to solve this equation using the finite difference method. It is a combination of implicit and explicit methods called the explicit-implicit method. Before we proceed to solve equation (3.14), we need some variable transformation to obtain a forward in time problem with constant coefficients as done in equation (3.7) to obtain equation (3.8). In addition, we also need to change the variables that appear in the integral term. Let $S = e^x$ and $J = e^y$ so that $V(S) = V(e^x) = \bar{V}(x)$ and $V(SJ) = V(e^x e^y) = V(e^{x+y}) = \bar{V}(x + y)$.

Then,

$$\frac{\partial \bar{V}}{\partial x} = S\frac{\partial V}{\partial S} \quad \text{and} \quad \frac{\partial^2 \bar{V}}{\partial x^2} = S\frac{\partial V}{\partial S} + S^2 \frac{\partial^2 V}{\partial S^2}.$$

Similarly, the integral term becomes $\int_0^\infty V(t, JS)g(J)\ dJ = \int_{-\infty}^\infty \bar{V}(t, x+y)f(y)\ dy$ where $f(y) = g(e^y)e^y$. Again, we change the variables in the new integral term, $z = x + y \Rightarrow y = z - x$ and $dy = dz$. The partial integro-differential equation (3.14) in terms of the new transformed variables x, z and $\tau = T - t$ will be:

$$\frac{d\bar{V}}{d\tau} - \frac{1}{2}\sigma^2\frac{\partial^2\bar{V}}{\partial x^2} - (r - \frac{1}{2}\sigma^2 - \kappa\lambda)\frac{\partial\bar{V}}{\partial x} + (r+\lambda)\bar{V} - \lambda\int_{-\infty}^\infty \bar{V}(\tau, z)f(z-x)\ dz = 0$$

For notational simplicity we can replace $\bar{V}(\tau, x)$ with $U(\tau, x)$ to get:

$$\frac{dU}{d\tau} - \frac{1}{2}\sigma^2\frac{\partial^2 U}{\partial x^2} - (r - \frac{1}{2}\sigma^2 - \kappa\lambda)\frac{\partial U}{\partial x} + (r+\lambda)U - \lambda\int_{-\infty}^\infty U(\tau, z)f(z-x)\ dz = 0 \tag{3.15}$$

with initial conditions given by:

$$U(0, x) = \psi(x) \text{ for } -\infty < x < \infty$$

where

$$\psi(x) = \begin{cases} \max\ [e^x - K, 0] & \text{for call options} \\ \max\ [K - e^x, 0] & \text{for put options.} \end{cases}$$

Since the log price of the underlying is denoted by $x = \ln S$, $\lim\ S \to 0$, $x \to -\infty$ and $\lim\ S \to \infty,\ x \to \infty$.

The boundary conditions are given by

$$U(\tau, x_{max}) = \begin{cases} e^{x_{max}} - Ke^{-r\tau} & \text{for call options} \\ 0 & \text{for put options} \end{cases}$$

$$U(\tau, x_{min}) = \begin{cases} 0 & \text{for call options} \\ Ke^{-r\tau} & \text{for put options} \end{cases}$$

where x_{max} and x_{min} are the truncated maximum and minimum values of x.
The original problem has been changed from a final value to an initial value problem with constant coefficient so that equation (3.15) can be solved more efficiently using finite difference method.

3.3.2 Truncating the integral

The domain of the solution of PIDE (3.15) is infinite in the space variable x. So artificial boundaries and boundary conditions will be required to define the numerical solutions on a finite domain. Consider an integral of the type $\int_{-\infty}^{\infty} g(y)\ dy$. To use finite difference to estimate the integral, first we need to truncate the integral domain. We choose A and B such that

$$| \int_{-\infty}^{\infty} g(y)\ dy - \int_{A}^{B} g(y)\ dy \ | < \epsilon$$

for some tolerance $\epsilon > 0$. For example, lets assume that jumps follow a Gaussian probability density function as in the case of Merton's model of the form

$$g(y) = \frac{1}{\sqrt{2\pi}\delta} e^{\frac{-y^2}{2\delta^2}}.$$

$g(y)$ is symmetric about $y = 0$ and also goes to zero very quickly [19]. To choose the limits of integration we look at the function when

$$g(y) \geq \epsilon \Leftrightarrow -\sqrt{-2\delta^2 \log\left(\epsilon\sqrt{2\pi}\right)} \leq y \leq \sqrt{-2\delta^2 \log\left(\epsilon\sqrt{2\pi}\right)} \tag{3.16}$$

Hence, by symmetry $A = -\sqrt{-2\delta^2 \log\left(\epsilon\sqrt{2\pi}\right)}$ and $B = -A$. This type of truncation is justified by Duffy and La Chioma [21, 30]. Next the truncated integral is approximated by some quadrature method as

$$\int_{A}^{B} g(y)\ dy \approx \frac{B-A}{N} \sum_{j=0}^{N} w_j g(y_j).$$

The transformed PIDE given by equation (3.15) with the truncated limits can now

be written in terms of derivative and integral parts as:

$$\frac{\partial U}{\partial \tau} = \mathcal{D}U + \mathcal{I}U \tag{3.17}$$

such that

$$\begin{aligned} \mathcal{D}U &= \frac{1}{2}\sigma^2\frac{\partial^2 U}{\partial x^2} + (r - \frac{1}{2}\sigma^2 - \kappa\lambda)\frac{\partial U}{\partial x} - (r+\lambda)U \\ \mathcal{I}U &= \lambda\int_A^B U(\tau, z)g(z-x)\ dz \end{aligned}$$

3.4 SOLUTION OF PIDE USING THE FINITE DIFFERENCE METHOD

3.4.1 Explicit finite difference method for PIDE

In explicit finite difference method, we try to find the next value $U(\tau + \delta\tau, x)$ in terms of previously known values explicitly. Let J and D be the discretization of the integral $\mathcal{I}$ and differential $\mathcal{D}$ in equation (3.17). Then using the explicit finite difference as described in section (3.2.1) we get:

$$\frac{U_{i+1} - U_i}{\Delta\tau} = DU_i + JU_i \Rightarrow U_{i+1} = [I + \Delta\tau(D+J)]U_i$$

for all j. Hence, unknown values at U_{i+1} can be found simply by multiplying the known values U_i by matrix $[I + \Delta\tau(D+J)]$ for each $i = 1, ..., M-1$.

Consider the PIDE (3.15) with the jump density function estimated by delta functions where α_k are constant weights whose sum is 1. Using the explicit finite difference scheme, the PIDE can be written as

$$U_{i+1,j} = aU_{i,j-1} + bU_{i,j} + cU_{i,j+1} - \delta\tau\lambda\sum_{k=1}^{m}\alpha_k U(\tau_i, x_j + y_k) \tag{3.18}$$

where $a = \left(\frac{\sigma^2}{2\delta x^2} - \frac{1}{2\delta x}(r - \frac{1}{2}\sigma^2 - \kappa\lambda)\right)\delta\tau$, $b = \left(1 - (\frac{\sigma^2}{\delta x^2} + r + \lambda)\delta\tau\right)$ and $c = \left(\frac{\sigma^2}{2\delta x^2} + \frac{1}{2\delta x}(r - \frac{\sigma^2}{2} - \kappa\lambda)\right)$. To implement the explicit scheme numerically we use the boundary and initial conditions as:

- $U(0,x) = \max\ (K - e^{(x_j)}, 0)$ where $j = 1....M$
- $U(\tau,0) = Ke^{-r(i\delta\tau)}$ where $i = 1....N$ and $U(\tau, e^{x_{max}}) = 0$

for put options and

- $U(0,x) = \max\ (e^{x_j} - K, 0)$ where $j = 1....M$
- $U(\tau, e^{x_{max}}) = e^{x_{max}} - Ke^{-r(idt)}$ where $i = 1....N$
- $U(\tau,0) = 0$

for call options. The above Partial Integro-Differential equation (PIDE) is then solved starting with time $\tau = 0$ and marching forward in time (τ). The option values at $i+1$ is a simple function of values at i and U_{i+1} can be found explicitly in terms of U_i.

Figure 3.4.1 shows an example of an explicit scheme for equation (3.14) using the jump sizes of $J = [0.8, 0.9, 1, 1.1]$, $\alpha = [1/4, 1/4, 1/4, 1/4]$ and $\lambda = 2$ in the integral term $\sum_{k=1}^{N} \alpha_k V(t_i, J_k S_j)$. First graph shows a smooth solution for $\delta S = 1,\ \delta t = 0.01$ with other parameters being fixed. For second graph we increased δt to $\delta t = 0.02$. We see that increasing the size of time step just a little causes the scheme to be unstable. Hence, explicit finite difference scheme is a simple scheme that requires restrictions on the step size and is conditionally stable. Small step size requires more computation time which can be very costly. Thus we look for unconditionally stable schemes that is relatively cheaper.

3.4.2 Implicit finite difference method for jump diffusion model

In the implicit method we represent the values $U(\tau + \delta\tau, x)$ implicitly from $U(\tau, x)$ as stated in section (3.2.2). In terms of discretized operators D for the derivative and J for the integral, the implicit scheme can be written as:

$$\frac{U_{i+1} - U_i}{\Delta\tau} = DU_{i+1} + JU_{i+1} \Rightarrow [I - \Delta\tau(D + J)]U_{i+1} = U_i$$

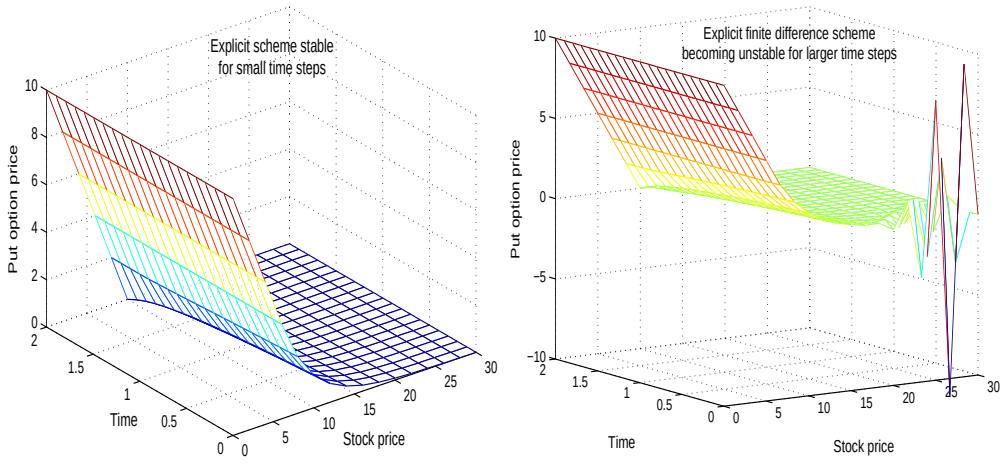


Figure 3.3. Explicit finite difference scheme is not always stable. In the first case we have $\delta t = .01, \delta S = 1$ resulting in a stable solution where as in the second case we have $(\delta t = .02, \delta S = 1)$ that resulted in instability.

for all $j = 1, ..., M$. To solve for U_{i+1} we need to invert the matrix $[I - \Delta\tau(D + J)]$. Matrix J, in this case can be a full matrix. Hence, it is computationally expensive to solve the system with a full matrix.

Consider the PIDE (3.15) with the jump density function estimated by linear combination of delta functions. Using the implicit finite difference scheme, the PIDE can be written as

$$aU_{i+1,j-1} + bU_{i+1,j} + cU_{i+1,j+1} - \delta\tau\lambda \sum_{k=1}^{m} \alpha_k U(\tau_{i+1}, x_j + y_k) = U_{i,j} \qquad (3.19)$$

where α_k are constants and $a = \left(-\frac{\sigma^2}{2\delta x^2} + \frac{1}{2\delta x}(r - \frac{1}{2}\sigma^2 - \kappa\lambda)\right)\delta\tau$,
$b = \left(1 + (\frac{\sigma^2}{\delta x^2} + r + \lambda)\delta\tau\right)$ and $c = \left(-\frac{\sigma^2}{2\delta x^2} - \frac{1}{2\delta x}(r - \frac{\sigma^2}{2} - \kappa\lambda)\right)\delta\tau$.
Using the initial and boundary conditions given in previous section, we have to solve the linear system of the type **AX=b** where **A** is a dense matrix and **b** is a known quantity.

The full implicit scheme is unconditionally stable but it requires significant computational time. Since we have to solve with a dense matrix, this method is not preferred when the size of matrix **A** is large.

3.4.3 Explicit-Implicit method

The explicit-implicit scheme can be written in terms of discretized operators D for the derivatives and J for the integral part as

$$\frac{U_{i+1,j} - U_{i,j}}{\Delta\tau} = DU_{i+1,j} + JU_{i,j} \Rightarrow [I - \Delta\tau D]U_{i+1,j} = [I + \Delta\tau J]U_{i,j}$$

To solve for $U_{i+1,j}$ we need to consider the matrix $[I - \Delta\tau D]$. The right side of the system consists of known quantities. Thus, this system can be solved using an LU decomposition or SOR with reasonable computational cost as described in section (3.2.2).

In terms of discretization, the PIDE (3.15) with the jump density function as delta functions can be written in a similar way to fully implicit scheme as,

$$aU_{i+1,j-1} + bU_{i+1,j} + cU_{i+1,j+1} - \delta\tau\lambda \underbrace{\sum_{k=1}^{m} \alpha_k U(\tau_i, x_j + y_k)}_{\mathcal{I}} = U_{i,j} \tag{3.20}$$

where a, b, and c are defined as in equation (3.19). To solve this system of equations, the integral part is evaluated exactly the way it is done in the explicit scheme. The system is then solved like a usual implicit system.

$$\begin{pmatrix} b & c & 0 & 0 & \dots & 0 \\ a & b & c & 0 & \dots & 0 \\ 0 & a & b & c & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & .. & a & b & c \\ 0 & 0 & 0 & \dots & a & b \end{pmatrix} \begin{pmatrix} U_{i+1,1} \\ U_{i+1,2} \\ U_{i+1,3} \\ \vdots \\ U_{i+1,M-2} \\ U_{i+1,M-1} \end{pmatrix} =$$

$$\begin{pmatrix} U_{i,1} + \delta\tau\lambda(\mathcal{I})_{i,1} \\ U_{i,2} + \delta\tau\lambda(\mathcal{I})_{i,2} \\ U_{i,3} + \delta\tau\lambda(\mathcal{I})_{i,3} \\ \vdots \\ U_{i,M-2} + \delta\tau\lambda(\mathcal{I})_{i,M-2} \\ U_{i,M-1} + \delta\tau\lambda(\mathcal{I})_{i,M-1} \end{pmatrix} - \begin{pmatrix} aU_{i,0} \\ 0 \\ 0 \\ 0 \\ \vdots \\ cU_{i,M} \end{pmatrix}$$

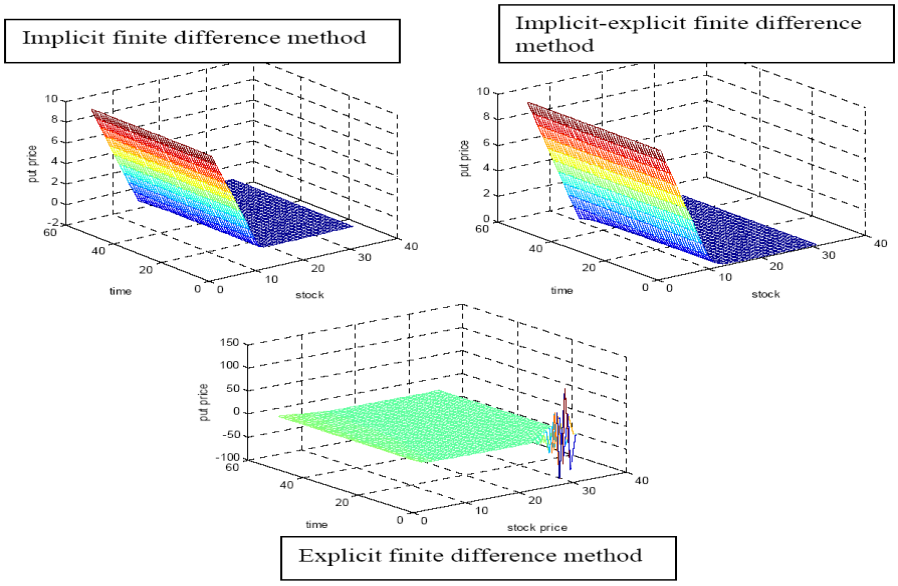
The entries $a,\ b,\ c$ form the diagonals of a trigonal matrix A. Using an LU decomposition or a conjugate gradient method the system of equations can be solved easily. The explicit-implicit scheme is stable and gives as good results as the fully implicit scheme. Since this does not require inverting a dense matrix, the system can be solved pretty fast.

To compare the stability of these three methods, ds is set sufficiently small ($ds = 1$ in this case) and dt is varied slowly. The table shows that initially all three schemes are stable for $ds = 1$ and $dt = 0.01$. Once the size of dt is increased slowly, the explicit scheme becomes unstable quickly. The implicit and explicit-implicit schemes are stable for any values of ds and dt.

Figure 3.4.3 shows graphs of implicit and explicit-implicit schemes for $J = [0.8, 0.9, 1, 1.1]$, $\alpha = [1/4, 1/4, 1/4, 1/4]$, $\lambda = 2$, strike(X)=10, interest rate(r)=.03, time to expiration(T)=1 year, market volatility (σ)=.3, Maximum stock price (Smax)=30, $\delta S = 1,\ \delta t = 0.02$.

Although implicit and explicit-implicit are both stable schemes the latter one is more feasible for computation involving large linear systems.

Figure 3.4. Even when explicit method is unstable for certain values of $(\delta t, \delta S)$, explicit-implicit method and implicit methods are still stable for the same $(\delta t, \delta S)$.



3.5 STABILITY

Theorem 3.5.1. *The explicit-implicit finite difference scheme derived from equation (3.14) is stable provided* $|r - \kappa\lambda| \leq \sigma^2$.

Proof: Consider the PIDE (3.14) with the integral term approximated by delta functions. In terms of explicit-implicit finite difference scheme it can be written as

$$a_j V_{i+1,j-1} + b_j V_{i+1,j} + c_j V_{i+1,j+1} + \delta t\lambda \sum_{k=1}^{N} \alpha_k V(t_i, J_k S_j) = V_{i,j} \tag{3.21}$$

where $a_j = \left(\frac{-\sigma^2 j^2}{2}\delta t + \frac{\delta t}{2}(rj - j\kappa\lambda)\right)$, $b_j = (1 + \sigma^2 j^2 \delta t + \delta t(r + \lambda))$ and $c_j = \left(-\frac{\sigma^2 j^2 \delta t}{2} - \frac{\delta t}{2}(rj - j\kappa\lambda)\right)$.

Let $V_i = [V_{i,0}, V_{i,1}, ..., V_{i,M}]$, be the solution vector for the above equation. Suppose the initial solution vector is perturbed so that $\hat{V_0} = V_0 + E_0$, where $E_i = [E_{i,0}, E_{i,1}, ..., E_{i,M}]$. Then, we get the following equation for the perturbation:

$$a_j E_{i+1,j-1} + b_j E_{i+1,j} + c_j E_{i+1,j+1} + \delta t\lambda \sum_{k=1}^{N} \alpha_k E(t_i, J_k S_j) = E_{i,j}$$

Table 3.1. Comparing three finite difference schemes

dt	Explicit	Explicit-Implicit	Implicit
0.01	stable	stable	stable
0.015	stable	stable	stable
0.02	unstable	stable	stable
0.05	unstable	stable	stable
0.1	unstable	stable	stable

or

$$|b_j||E_{i+1,j}| \leq |E_{i,j}| + |c_j E_{i+1,j+1}| + |a_j E_{i+1,j-1}| + \delta t\lambda \sum_{k=1}^{N} \alpha_k |E(t_i, J_k S_j)|$$

Define $\parallel E \parallel_i = \max_j |E_{i,j}|$ with $\sum_{k=1}^{N} \alpha_k = 1$, then

$$|b_j||E_{i+1,j}| \leq \parallel E \parallel_i + |c_j| \parallel E \parallel_{i+1} + |a_j| \parallel E \parallel_{i+1} + \delta t\lambda \parallel E \parallel_i \tag{3.22}$$

Equation (3.22) is valid for all $j \leq M$. So, it is true for any node j_*, where $max_j|E_{i+1,j}| = |E_{i+1,j_*}|$. So, from equation (3.22) we get

$$|b_{j_*}| \parallel E_{i+1} \parallel \leq \parallel E \parallel_i (1 + \delta t\lambda) + \parallel E \parallel_{i+1} (|a_{j_*}| + |c_{j_*}|)$$

or

$$\parallel E_{i+1} \parallel [|b_{j_*}| - |a_{j_*}| - |c_{j_*}|] \leq \parallel E \parallel_i (1 + \delta t\lambda) \tag{3.23}$$

From the given condition, if $|r - \kappa\lambda| \leq \sigma^2$ then $r - \kappa\lambda \leq \sigma^2 \leq \sigma^2 j$ and $-r + \kappa\lambda \leq \sigma^2 \leq \sigma^2 j$ for any $j > 0$.

Hence, $\pm(r - \kappa\lambda) \leq \sigma^2 j$, or $\pm(r - \kappa\lambda)j \leq \sigma^2 j^2$. Multiplying both sides of inequality

by $\frac{\delta t}{2}$, we get

$$\left(\pm(r-\kappa\lambda)j - \sigma^2 j^2\right)\frac{\delta t}{2} \leq 0$$

$$\text{or } (r-\kappa\lambda)j\frac{\delta t}{2} - \sigma^2 j^2\frac{\delta t}{2} \leq 0$$

$$\text{and } -(r-\kappa\lambda)j\frac{\delta t}{2} - \sigma^2 j^2\frac{\delta t}{2} \leq 0$$

Hence, $a_j = (r-\kappa\lambda)j\frac{\delta t}{2} - \sigma^2 j^2\frac{\delta t}{2} \leq 0$ and $c_j = -(r-\kappa\lambda)j\frac{\delta t}{2} - \sigma^2 j^2\frac{\delta t}{2} \leq 0$ for any node j.

Therefore,

$$\begin{aligned} |b_{j_*}| - |a_{j_*}| - |c_{j_*}| &= b_{j_*} + a_{j_*} + c_{j_*} \\ &= (1 + \sigma^2 j^2\delta t + \delta t(r+\lambda)) + (\frac{-\sigma^2 j^2}{2}\delta t + \frac{\delta t}{2}(rj - j\kappa\lambda)) \\ &\quad +(-\frac{\sigma^2 j^2\delta t}{2} - \frac{\delta t}{2}(rj - j\kappa\lambda)) \\ &= 1 + \delta t(r+\lambda) \end{aligned} \tag{3.24}$$

Hence $|b_{j_*}| - |a_{j_*}| - |c_{j_*}| = 1 + \delta t(r+\lambda) \geq (1+\delta t\lambda)$ for any node j. So, we have $\frac{(1+\delta t\lambda)}{1+\delta t(r+\lambda)} \leq 1$.

Now, using this information and substituting in equation (3.23), we get

$$\| E \|_{i+1} \leq \frac{(1+\delta t\lambda)}{1+\delta t(r+\lambda)} \| E \|_i \leq \| E \|_i$$

or

$$\frac{\| E \|_{i+1}}{\| E \|_i} \leq 1$$

Since error E is bounded for every time step, the finite difference scheme is stable.

Hence, proof is complete.

3.6 CONSISTENCY

A finite difference scheme is consistent if the scheme reduces to the original differential equation as increments in the independent variables tend to 0.

Theorem 3.6.1. *The explicit-implicit finite difference scheme derived from equation (3.14) is consistent.*

Proof: Consider the PIDE (3.14) with the integral term approximated by linear combination of delta functions.

Let

$$\mathcal{LV} = V_t + \frac{\sigma^2 S^2}{2} V_{ss} + V_s(rS - S\kappa\lambda) - (r+\lambda)V + \lambda \sum_{k=1}^{l} \alpha_k V(t_i, J_k S_j) = 0 \quad (3.25)$$

Now, using the Taylor series approximation for the solution V of (3.25),

$$V_{i+1,j} = V_{i,j} + \delta t (V_t)_{i,j} + \frac{(\delta t)^2}{2!}(V_{tt})_{i,j} + \frac{(\delta t)^3}{3!}(V_{ttt})_{i,j} + O(\delta t^4) \quad (3.26)$$

$$V_{i+1,j+1} = V_{i+1,j} + \delta S (V_S)_{i+1,j} + \frac{(\delta S)^2}{2!}(V_{SS})_{i+1,j} + \frac{(\delta S)^3}{3!}(V_{SSS})_{i+1,j} + O(\delta S^4) \quad (3.27)$$

$$V_{i+1,j-1} = V_{i+1,j} - \delta S (V_S)_{i+1,j} + \frac{(\delta S)^2}{2!}(V_{SS})_{i+1,j} - \frac{(\delta S)^3}{3!}(V_{SSS})_{i+1,j} + O(\delta S^4) \quad (3.28)$$

Adding (3.27) and (3.28), we get

$$V_{i+1,j+1} + V_{i+1,j-1} = 2V_{i+1,j} + (\delta S)^2 (V_{SS})_{i+1,j} + O(\delta S^4)$$

or

$$\frac{V_{i+1,j+1} - 2V_{i+1,j} + V_{i+1,j-1}}{(\delta S)^2} = (V_{SS})_{i+1,j} + O(\delta S^2) = (V_{SS})_{i,j} + O(\delta S^2)$$

Subtracting (3.28) from (3.27), we get

$$V_{i+1,j+1} - V_{i+1,j-1} = 2\delta S (V_S)_{i+1,j} + O(\delta S^3)$$

or

$$\frac{V_{i+1,j+1} - V_{i+1,j-1}}{2\delta S} = (V_S)_{i+1,j} + O(\delta S^2) = (V_S)_{i,j} + O(\delta S^2)$$

In terms of finite difference scheme, the original PIDE is written as

$$\mathcal{L}_d\mathcal{V} = \frac{V_{i+1,j} - V_{i,j}}{\delta t} + \frac{\sigma^2(S_j)^2}{2}\left(\frac{V_{i+1,j+1} - 2V_{i+1,j} + V_{i+1,j-1}}{(\delta S)^2}\right) + S_j(r - \kappa\lambda)\left(\frac{V_{i+1,j+1} - V_{i+1,j-1}}{2\delta S}\right) - (r+\lambda)V_{i+1,j} + \lambda\sum_{k=1}^{l}\alpha_k V(J_k S_j, t_i) + O(\delta t + \delta S^2) \tag{3.29}$$

Let $\mathcal{L}_d\mathcal{V}$ be the finite difference scheme for our PIDE. Replacing the derivatives in (3.29) by the Taylor series expansion derived above, we get

$$(\mathcal{L}_d\mathcal{V})_{i,j} = (V_t)_{i,j} + \frac{\sigma^2(S_j)^2}{2}(V_{SS})_{i,j} + S_j(r - \kappa\lambda)(V_S)_{i,j} - (r+\lambda)V_{i,j} + \lambda\sum_{k=1}^{l}\alpha_k V(J_k S_j, t_i) + O(\delta t + \delta S^2) = (\mathcal{L}_d\mathcal{V})_{i,j} + O(\delta t + \delta S^2) \tag{3.30}$$

This implies that for $\lim \delta S \to 0,\ \delta t \to 0,$, we obtain $(\mathcal{L}\mathcal{V})_{i,j} = (\mathcal{L}_d\mathcal{V})_{i,j}$ for each i, j. Therefore, the finite difference scheme is consistent and proof is complete.

3.7 SOLUTION OF PIDE USING ITERATIVE METHODS

We have seen that the solution of the PIDE derived from the jump diffusion process is much harder to solve if we discretize it by implicit method that results in a dense matrix. Almendral and Oosterlee [1] used the second order backward difference formula (BDF2) on a uniform grid to discretize the PIDE. They used the solver based on a splitting of the resulting dense matrix and used a fast Fourier transformation to speed up the method. It is shown that a finite difference method combined with BDF2 can be fast and efficient to solve a dense system arising from implicit discretization of the PIDE. In this section we summarize the procedure given in [1, 44] for solving a dense linear systems arising from discertizing the PIDE model.

Consider the PIDE model given by equation (3.15). We truncate the domain for $x = \ln S$ to $\Omega = (-A, A)$ where A is chosen such that the integral over the $\Omega^c = (-\infty, -A) \cup (A, \infty)$ is sufficiently small. As shown in [19], the truncation error decreases exponentially with domain size if we use the payoff function as boundary condition. However it is not sufficient to impose only boundary conditions at $x = A$ and $x = -A$ because of the nonlocal integral term. To estimate the integral on the complement of the computational domain, we use the boundary conditions $u(\tau, x) \approx 0$ when $x \in (-\infty, -A)$ and $u(\tau, x) \approx e^x - Ke^{-r\tau}$ when $x \in (A, \infty)$ for call option. So, the remainder $R(\tau, x, A)$ of the integral $\int_{\mathbb{R}} u(\tau, z) f(z - x)\, dz$ outside Ω will be estimated by

$$R(\tau, x, A) = \int_A^\infty (e^x - Ke^{-r\tau}) f(z - x)\, dz. \tag{3.31}$$

For detail explaination, see Almendral and Oosterlee [1]. In the case of Merton's model $f(x)$ is assumed to be a normal density function with mean μ and variance σ_J. Using the cumulative distribution for standard normal distribution we can simplify the remainder R in (3.31) by using the change of variables $y = (z - x)/\sigma_J$ as follows:

$$\begin{aligned}
\int_A^\infty f(z - x) dz &= \frac{1}{\sqrt{2\pi}} \int_A^\infty e^{-(z-x)^2/(2\sigma_J^2)} \frac{dz}{\sigma_J} \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{A-x}{\sigma_J}}^\infty e^{-y^2/2} dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{A-x}{\sigma_J}} e^{-y^2/2} dy \\
&= \Phi(\frac{x - A}{\sigma_J})
\end{aligned} \tag{3.32}$$

For the second integral we set $y = (z - x)/\sigma_J$ and $\bar{y} = y - \sigma_J$. So,

$$\begin{aligned}
\int_A^\infty e^z f(z-x)dz &= \frac{1}{\sqrt{2\pi}}\int_A^\infty e^{z-(z-x)^2/(2\sigma_J^2)}\frac{dz}{\sigma_J} \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{A-x}{\sigma_J}}^\infty e^{\sigma_J y+x-y^2/2}dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{A-x}{\sigma_J}}^\infty e^{-(y^2-2\sigma_J y+\sigma_J^2)/2}e^{x+\sigma_J^2/2}dy \\
&= \frac{e^{x+\sigma_J^2/2}}{\sqrt{2\pi}}\int_{\frac{A-x}{\sigma_J}-\sigma_J}^\infty e^{-\bar{y}^2/2}d\bar{y} \\
&= \Phi\left(\frac{x-A+\sigma_J^2}{\sigma_J}\right)e^{x+\sigma_J^2/2} \qquad (3.33)
\end{aligned}$$

Now using the integration of (3.32) and (3.33) in (3.31), we get

$$R(\tau,x,A) = e^{x+\sigma_J^2/2}\Phi\left(\frac{x-A+\sigma_J^2}{\sigma_J}\right) + Ke^{-r\tau}\Phi\left(\frac{x-A}{\sigma_J}\right). \qquad (3.34)$$

For more detail of the derivation see [1], [44].

3.7.1 Discretization of PIDE (fully implicit scheme using BDF2)

We denote the minimum truncated value $x = A$ as x_{min}. To discretize the integral domain we use $x_j = x_{min} + (j-1)\delta x$ for $j = 1, ..., M$ and $\tau_i = (i-1)\delta\tau$ for $i = 1, ..., N$. The integral term $\int_{\mathbb{R}} U(\tau_i, z)f(z-x_j)$ is approximated using the composite trapezoidal rule as

$$\begin{aligned}
\int_{\mathbb{R}} U(\tau_i,z)f(z-x_j)dz = \frac{h}{2}\left[u_{i,1}f_{1,k} + 2\sum_{k=2}^{m-1}U_{i,j}f_{k,j} + U_{N,M}f_{M,n}\right] \\
+R(\tau_i,x_j,x_{min}),\ \forall j \in \{2,...,M-1\}, \qquad (3.35)
\end{aligned}$$

with $f_{k,j} = f(x_k - x_j)$ and $U_{i,j} = U(\tau_i, x_j)$. The remainder term R is defined as

$$\begin{aligned}
R(\tau,x,x_{min}) &= \int_{x_{min}}^\infty (e^z - Ke^{-r\tau})f(z-x)dz \\
&= e^{x+\sigma_J^2/2}\Phi\left(\frac{x-A+\sigma_J^2}{\sigma_J}\right) + Ke^{-r\tau}\Phi\left(\frac{x-A}{\sigma_J}\right) \qquad (3.36)
\end{aligned}$$

as given by (3.34). Following the same finite difference scheme proposed in [1], we will have:

$$u_\tau(\tau_i, x_j) \approx \begin{cases} (\frac{3}{2}u_{i,j} - 2u_{i-1,j} + \frac{1}{2}u_{i-2,j})/\delta\tau & \text{for } i \geq 2, \\ (u_{i,j} - u_{i-1,j})/\delta\tau & \text{for } i = 1, \end{cases}$$

$$\begin{aligned} u_{xx}(\tau_i, x_j) &\approx \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{\delta x^2}, \\ u_x(\tau_i, x_j) &\approx \frac{u_{i,j+1} - u_{i,j-1}}{2\delta x}. \end{aligned}$$

In this finite difference scheme, the backward difference formula of second order (BDF2) is used to find the time derivative for $i \geq 2$. For Spatial derivatives a second order central difference is used. This implicit scheme results in a linear system

$$(\omega_0 I + C + D)U_{i+1} = b_i$$

where

$$\omega_0 = \begin{cases} 1 & \text{for } i = 1, \\ \frac{3}{2} & \text{for } i \geq 2, \end{cases}$$

This linear system has to be solved for every time step τ_i. If we let the matrix $A = (\omega_0 I + C + D)$, then the $M \times M$ matrix A can be written as:

$$\begin{bmatrix} \frac{3}{2} & 0 & 0 & ... & ... & 0 \\ b - \frac{c}{2}f_{2,1} & d & a - cf_{2,3} & -cf_{2,4} & ... & -\frac{c}{2}f_{2,M} \\ -\frac{c}{2}f_{3,1} & b - cf_{3,2} & d & ... & ... & -\frac{c}{2}f_{3,M} \\ \vdots & -cf_{4,2} & \ddots & ... & ... & \vdots \\ & & \ddots & ... & a - cf_{M-2,M-1} & \\ -\frac{c}{2}f_{M-1,1} & -cf_{M-1,2} & ... & b - cf_{M-1,M-2} & d & a - \frac{c}{2}f_{M-1,M} \\ 0 & 0 & ... & 0 & 0 & \frac{3}{2} \end{bmatrix}$$

where $a = (-\frac{\sigma^2}{2\delta x^2} - \frac{1}{2\delta x}(r - \frac{1}{2}\sigma^2 - \kappa\lambda))\delta\tau$, $b = (-\frac{\sigma^2}{2\delta x^2} + \frac{1}{2\delta x}(r - \frac{\sigma^2}{2} - \kappa\lambda))\delta\tau$ $c = \delta\tau\delta x\lambda$ and $d = \frac{3}{2} + (\frac{\sigma^2}{\delta x^2} + r + \lambda)\delta\tau - cf_{ii}$. From the initial and boundary conditions, we have $U_{i,1} = b_{i,1}/\omega_0$ and $U_{i,M} = b_{i,M}/\omega_0 = (e^{x_{min}} - Ke^{-r\tau_i})$ from the last and the first row respectively. Now, the remaining $(M-2) \times (M-2)$ matrix say $T = (A_{i,j})_{2 \leq i,j \leq M-1}$ is dense but structured and belongs to a class of Toeplitz matrices. When the jump is Gaussian as described by Merton's model, we have $f_{i,j} = f_{j,i}$ and the matrix T is symmetric except at the first upper and lower diagonals. Strauss [44] has suggested a way of making the matrix T symmetric by removing the convection term using transformation to the PIDE model. This way a broader range of applicable methods can be used to solve the PIDE efficiently. Strauss [44] has also suggested a fast Preconditioned Conjugate Gradient (PCG) method using the generating function for Toeplitz operator. For a non-symmetric matrix T, which arises when the jump distribution is asymmetric (eg Kou's model), [1] and [27] have used as linear solvers the splitting iterations or GMRES. These can be applied to non-symmetric linear systems and are also fast.

CHAPTER 4

PARAMETER ESTIMATION

4.1 INTRODUCTION

Unknown parameters in a our financial model cannot be measured directly from the market data. These models contain adjustable parameters which need to be determined from available market price. From the observed data from the market, our aim is to adjust the parameters in our model so that the model reproduces the market values as close to as possible. Hence, parameter identification is an inverse problem. The appropriate financial model that models the market is already chosen, and the problem is to find the parameter values that best approximate the observed market values.

It is rare that any parameter set can exactly match the given data. Hence parameter identification can be challenging for several reasons. Sometimes the existence of a solution is not guaranteed, especially when a certain phenomenon is overlooked by the model. There is also a possibility of multiple solutions. Apart from this, stability is also a concern for parameter identification problems. Some of the estimated parameters have significant influence in the model where as some have less. Hence, sensitivity analysis is done on the estimated parameters to check which parameters are the most influential and which are the least influential in the model.

To identify the unknown parameters for the given problem we formulate the problem as an optimization problem. There are several search methods for solving optimization problems. We can categorize them in terms of whether derivative information is used or not. Search methods that use only function evaluations are used for very nonlinear functions or functions with discontinuities. When the gradient of an objective function is continuous and computable, gradient methods are more efficient. Higher order methods like the Newton's method that use information of the second

derivatives are computationally expensive.

Let $\mathbf{x} \in \mathbf{R}^n$ represent the parameter values, $\mathbf{d} \in \mathbf{R}^m$ represent the observed values, and $\mathbf{y}(.) : \mathbf{R}^n \rightarrow \mathbf{R}^m$ represent the solution to the given problem for parameter values $\mathbf{x}$. Then

$$f(\mathbf{x}) = \| \, y(\mathbf{x}) - \mathbf{d} \, \|^2 \tag{4.1}$$

is a least-squares objective function. If $\mathcal{P}$ represents the set of all possible parameter values, the optimization problem for parameter identification is to find some $\mathbf{x}_* \in \mathcal{P}$ such that

$$f(\mathbf{x}_*) \leq f(\mathbf{x})$$

for all $\mathbf{x} \in \mathcal{P}$.

In parameter estimation process where the objective function has the form of equation (4.1), the number of data points m is usually much bigger than the number of variables n. If more observation points exist than parameters, the nonlinear least-squares problem can be viewed as an over determined system of equations. In this case, if the optimal set of parameters $\mathbf{x}$ exist, they may not be unique.

Parameter estimation is very important in financial models. For the simple Black-Scholes model the only parameter to be estimated is the volatility σ. As models become complex, parameter estimation becomes increasingly difficult. Various authors have used different methods to estimate parameters in more complex financial models. Avellaneda, Friedman, Holmes and Sampieri [5] use relative entropy minimization to calibrate volatility surface where as Cont and Tankov [16],[18] use similar procedure to calibrate jump diffusion option pricing models driven by Lévy process. Various authors have used gradient based optimization procedure with regularization to estimate parameters. From practical experience, no matter what algorithm we are using to solve a nonlinear optimization problem, the better the gradient is computed, the better the optimal solution of the problem is calculated.

4.2 GRADIENT ESTIMATION

In order to use gradient information in optimization, the gradient has to exist, be continuous and should be computable with reasonable expense. The latter of these requirements may present a substantial obstacle in gradient based optimization. Gradient based optimization techniques utilize the gradient information to identify search directions. As a result these methods usually require fewer steps then direct search methods to obtain an accurate approximation of a locally optimal solution. However, gradient based methods tend to get attracted to a local extremum close to the initial guess and we may never reach the global minimum that we are looking for. For a thorough overview of such techniques, and the underlying theory, we refer to the textbook [39].

Gradient free optimization techniques are direct search methods. In high dimensional parameter spaces, gradient free methods require a large number of function evaluations just to identify a downhill direction. Hence it is difficult to get close to an extremum. An advantage of non gradient based optimization is that they tend to be harder to be distracted by local extrema and more likely to find solutions that are good in a global sense. For more information on gradient free methods we refer to [38]. If gradient information is available we prefer gradient based optimization for parameter estimation.

The choice of the method by which the gradient of the cost function is determined is important. We will explain two different approaches for gradient computation in optimization problems namely sensitivity equations method and adjoint equations method. These are not the only possible choices. Finite differences and automatic differentiation are also popular ways of finding the gradient. In the context of our problem, the sensitivity equation method computes the sensitivity of the solution V of our model partial differential equation with respect to the model parameters $\mathbf{p}$ in some way. The gradient of the objective function is then evaluated by applying

the chain rule of differentiation, using the sensitivity $\frac{\partial V}{\partial \mathbf{p}}$. For simplicity, we define the objective function to be minimized as a least-squares problem given by

$$F(\mathbf{p}) = \frac{1}{2} \int \int_{\Omega} (V(t, x; \mathbf{p}) - V_{data}(t, x))^2 \, dt \, dx \tag{4.2}$$

where V is the solution of some PDE. Our goal is to find $\mathbf{p}^* \in \mathbf{R}^n$ such that, $F(\mathbf{p}^*) \leq F(\mathbf{p}), \ \forall \mathbf{p} \in \mathbf{R}^n$. Effective minimization algorithms require the gradient $F'(\mathbf{p})$ with respect to the parameters $\mathbf{p}$ at a given point $\mathbf{p}_0$. There are fast gradient based optimization methods available for problems of this type, but they require efficient computation of gradient of the objective function F with respect to each of the parameters $\mathbf{p}$. In constrained optimization even small errors in the derivative can lead to serious problems in the application of optimization algorithms [32]. Using the chain rule we differentiate the objective function (4.2) as follows:

$$\frac{dF}{d\mathbf{p}} = \frac{\partial F}{\partial V} \frac{\partial V}{\partial \mathbf{p}} \tag{4.3}$$

Since the relationship between the cost function F and the state variable V is explicit and simple, the partial derivative $\frac{\partial F}{\partial V}$ is easy to calculate exactly. However, the relationship between state variables V and parameters $\mathbf{p}$ is only implicit. Hence the partial derivatives $\partial V / \partial \mathbf{p}_i$ require differentiating the state and boundary equations and solving the resulting linearized equation at a particular set of state and parameter values.

4.2.1 Sensitivity equation method

Sensitivity is defined as the derivative of a dependent variable with respect to a model parameter. Sensitivity equation method can be used for computing the gradient of the objective function. Apart from obtaining the gradient information, we can also use this method to predict the effects of small perturbations on solutions of differential equations. In the sensitivity equation method, we differentiate the state

equation with respect to the parameter of interest leading to a linear equations on the sensitivities. In deriving sensitivities, we assume that derivative defining the sensitivity exists and the order of differentiation commutes.

As an example, lets consider the heat equation given by

$$\begin{aligned} \frac{\partial u}{\partial t} &= \kappa \frac{\partial^2 u}{\partial x^2}, \\ u(0,x) &= f(x), \\ u(t,0) = u(t,1) &= 0 \end{aligned} \tag{4.4}$$

Let $u(t,x;\kappa)$ be the solution of the heat equation (4.4). Define the sensitivity with respect to κ by $w(t,x) = \dfrac{\partial u(t,x;\kappa)}{\partial \kappa}\delta\kappa$. Then, the sensitivity equation is given by

$$\begin{aligned} \frac{\partial w}{\partial t} &= \kappa \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 u}{\partial x^2}\delta\kappa \\ w(0,x) &= 0, \\ w(t,0) = w(t,1) &= 0 \end{aligned} \tag{4.5}$$

First we solve the main problem (4.4) for $u(t,x)$. Then the linear sensitivity equation (4.5) can be solved separately after computing the solution for $u(t,x)$. Finally, the gradient of the cost functions as defined by equation (4.3) can be computed using the sensitivity equation (4.5). Sensitivity equations provide more than simply the gradient of the cost function. They can be used to predict the behavior of model parameters as well. It is important to identify highly sensitive and influential parameters in mathematical models. This is done by solving sensitivity equations to predict the model behavior under the influence of parameters. Sensitivity analysis is beyond the scope of this thesis. We refer to [12], [41], [46] for more elaborate study and use of sensitivity equations.

Sensitivity equations have a similar structure to the original governing equations and thus can often be solved by the same routine as the original PDE. In case when there are m parameters and m is a large number, we need to solve the sensitivity

equation (4.5) m times for all basis vectors $\delta\kappa$ to obtain the sensitivity for all m parameters. This could be tedious or in some cases practically not feasible. Hence it is advantageous to use the sensitivity equation approach only when the number of parameters m is relatively small. When the number of parameters to be estimated is large, we have to look for more efficient methods to estimate the gradient of the objective function. The adjoint method is efficient for problem with large number of sensitivity parameters and few objective functions [33].

4.2.2 Adjoint equation method

The adjoint equations method is an efficient way of calculating gradient of a constrained optimization problem, even for very large dimensional problems. In the adjoint equations method, one additional equation beyond the governing equation is solved. It is important to note that there is only one adjoint equation to be solved for one governing equation, no matter how many unknown parameters are involved. The adjoint equation for our problem has the characteristic of being zero at the final time and hence solved from the final time backwards to the initial time unlike the governing equation.

Adjoint equations are increasingly popular in wide area of applied mathematics for data assimilation, model calibration and sensitivity analysis. Some area with their extensive use include optimal design, meteorology and oceanography. In this section, we will describe the use of adjoint equations for gradient estimation in a simple minimization problem. Later in the chapter, we will extend this method to calibrate parameters in some popular option pricing models. The adjoint equations method is a growing area of interest to many scientists and engineers and there are even books and articles solely devoted to adjoint equations. For general derivation of adjoint equations of linear and nonlinear partial differential equations readers are referred to [35]. We will begin this section by considering a simple one dimensional

heat equation as an example to highlight the basics of the derivation of the adjoint equation. Then later in the section we will use the adjoint equation method to derive gradient information of the objective function of our interest for much more complex differential equations.

4.2.3 Derivation of Adjoint Equation

In this section we will derive an adjoint of a simple problem as an example. Consider a simple partial differential equation with initial and boundary conditions given by

$$\begin{aligned} \frac{\partial \phi}{\partial t} = \kappa \frac{\partial^2 \phi}{\partial x^2} + f, \quad t \geq 0, \\ \phi(0, x) = 0, \\ \phi(t, 0) = 0, \ \phi(t, 1) = 0. \end{aligned} \tag{4.6}$$

Assume that ϕ is a sufficiently smooth, continuously differentiable function in t on $[0, T]$ and twice continuously differentiable in x on $[0, 1]$. Let us also assume that $\phi(t, x)$ has prescribed values at the boundary and initial conditions given as in equation (4.6). Let ϕ^* be a sufficiently smooth twice differentiable function. Let us introduce an inner product of the functions $\phi(t, x)$ and $\phi^*(t, x)$ with the domain $\mathbf{Q} = (T, 0) \times (0, 1)$:

$$< \phi^*, \phi >= \int_0^T \int_0^1 \phi^* \phi \ dx \ dt$$

Let L be the differential operator such that

$$-L\phi = \frac{\partial \phi}{\partial t} - \kappa \frac{\partial^2 \phi}{\partial x^2}.$$

Then equation (4.6) becomes $L\phi = -f$. Consider the inner product of the functions ϕ^* and $L\phi$ and integrate by parts,

$$
\begin{aligned}
& < \phi^*, L\phi > \\
= & \int_0^1 \int_0^T \phi^* \left[\kappa \phi_{xx} - \phi_t\right] \, dt \, dx \\
= & \int_0^1 \int_0^T \phi^* \kappa \phi_{xx} \, dt \, dx - \int_0^1 \int_0^T \phi^* \phi_t \, dt \, dx \\
= & \int_0^T \kappa \left[\phi^* \phi_x |_0^1 \, dt - \int_0^1 \phi_x^* \phi_x \, dx \right] \, dt - \int_0^1 \left[\phi^* \phi |_0^T \, dx - \int_0^T \phi_t^* u \, dt \right] \, dx \\
= & \int_0^T \kappa \left[(\phi^* \phi_x - \phi_x^* \phi) |_0^1 + \int_0^1 \phi_{xx}^* \phi \, dx \right] \, dt - \int_0^1 \left[\phi^* \phi |_0^T \, dx - \int_0^T \phi_t^* \phi \, dt \right] \, dx \\
= & \int_0^1 \int_0^T \left[\kappa \phi_{xx}^* + \phi_t^*\right] \phi \;\; dt \, dx + \int_0^T \kappa (\phi^* \phi_x - \phi_x^* \phi) |_0^1 \, dt - \int_0^1 \phi^* \phi |_0^T \, dx
\end{aligned}
$$

If we use the given initial and boundary conditions for ϕ and choose ϕ^* such that $\phi^*(t,0) = 0, \quad \phi^*(t,1) = 0, \quad \phi^*(T,x) = 0$, the quantity outside the double integral vanishes. Hence we get

$$
\begin{aligned}
< \phi^*, L\phi > &= \int_0^1 \int_0^T \left[\kappa \phi_{xx}^* + \phi_t^*\right] \phi \; dt \, dx \\
&= < L^* \phi^*, \phi > \qquad (4.7)
\end{aligned}
$$

L^* is called the adjoint operator and ϕ^* is the solution of the adjoint equation $L^* \phi^* = g$ corresponding to the main equation (4.6). Hence the adjoint equation is given by

$$
\begin{aligned}
& \frac{\partial \phi^*}{\partial t} + \kappa \frac{\partial^2 \phi^*}{\partial x^2} = g, \quad 0 < x < 1, \; 0 \le t < T \\
& \qquad \phi^*(t,0) = \phi^*(t,1) = 0, \quad \phi^*(T,x) = 0. \qquad (4.8)
\end{aligned}
$$

Unlike the main problem (4.6) which is an initial value problem, the adjoint problem (4.8) is a final value problem with final value given at $t = T$. Hence, for time dependent problems, adjoint solutions have to be calculated backward in time, starting from the final time T. In next section we will show how the adjoint equation method is used to calculate the gradient of the objective function.

4.3 DERIVATION OF GRADIENT USING ADJOINT EQUATIONS

4.3.1 Simple Heat Equation

Consider the heat equation given by (4.4) and the corresponding sensitivity equation given by (4.5). Then we have the following theorem:

Theorem 4.3.1. *Let the objective function to be minimized be:*

$$F(\kappa) = \frac{1}{2}\int_0^T \int_0^1 \left[u(t,x;\kappa) - u_{data}(t,x)\right]^2 \; dx \; dt$$

The gradient of the objective function is given by

$$F'(\kappa) = \int_0^T \int_0^1 u_{xx} p \; dx \; dt$$

where p is the solution of the adjoint equation

$$-p_t = \kappa p_{xx} + u(t,x;\kappa) - u_{data}$$

with the final condition and the boundary conditions given by:

$$p(T,x) = 0$$

$$p(t,0) = 0, \quad p(t,1) = 0$$

Proof:

Suppose the objective function to be minimized is

$$F(\kappa) = \frac{1}{2}\int_0^T \int_0^1 \left[u(t,x;\kappa) - u_{data}(t,x)\right]^2 \; dx \; dt \tag{4.9}$$

Let $w = \dfrac{\partial u}{\partial \kappa}\delta\kappa$ be the sensitivity of u with respect to parameter κ and $w(t,x)$ solves equation (4.5). Taking the derivative of the objective function with respect to parameter κ, we get

$$\begin{aligned} F'(\kappa)\delta\kappa &= \int_0^T \int_0^1 \left[u(t,x;\kappa) - u_{data}(t,x)\right] \frac{\partial u}{\partial \kappa}\delta\kappa \; dx \; dt \\ &= \int_0^T \int_0^1 \left[u(t,x;\kappa) - u_{data}(t,x)\right] w \; dx \; dt \end{aligned} \tag{4.10}$$

Let $p(t,x)$ be a smooth function with the final condition given by $p(T,x) = 0$. Using the initial condition from equation (4.5) and the final condition on p, we will have $w(T,x)p(T,x) - w(0,x)p(0,x) = 0$. Hence,

$$\begin{aligned}
0 &= \int_0^1 w(T,x)p(T,x) - w(0,x)p(0,x)\ dx \\
&= \int_0^1 \int_0^T \frac{d}{dt}\left[w(t,x)p(t,x)\right] dt\ dx \\
&= \int_0^1 \int_0^T \left[(w_t(t,x)p(t,x) + w(t,x)p_t(t,x))\right]\ dt\ dx \\
&= \int_0^1 \int_0^T \left[\kappa w_{xx} + u_{xx}\delta\kappa\right] p + wp_t\ \ dt\ dx \\
&= \int_0^T \kappa \left[pw_x|_0^1\ dt - \int_0^1 p_x w_x\ dx\right]\ dt + \int_0^1 \int_0^T (pu_{xx}\delta\kappa + wp_t)\ dt\ dx \\
&= \int_0^T \kappa \left[(pw_x - p_x w)|_0^1 + \int_0^1 p_{xx} w\ dx\right]\ dt + \int_0^1 \int_0^T pu_{xx}\delta\kappa + wp_t\ dt\ dx \\
&= \int_0^1 \int_0^T \left(\left[\kappa p_{xx} + p_t\right] w + pu_{xx}\delta\kappa\right)\ \ dt\ dx + \int_0^T \kappa(pw_x - p_x w)|_0^1\ dt
\end{aligned}$$

Choose $p(t,x)$ such that $p(t,0) = p(t,1) = 0$. Then using the boundary conditions on $w(t,x)$ from equation (4.5), the quantity outside the double integral is 0. Hence, we have

$$\int_0^1 \int_0^T (\kappa p_{xx} + p_t)w + pu_{xx}\delta\kappa\ \ dt\ dx = 0$$

Adding and subtracting the term $(u(t,x;\kappa) - u_{data})w$ inside the integral, we get

$$\int_0^1 \int_0^T (\kappa p_{xx} + p_t + u(t,x;\kappa) - u_{data})w + pu_{xx}\delta\kappa - (u(t,x;\kappa) - u_{data})w\ \ dt\ dx = 0$$

The first part of the equation is zero. Using the previously defined conditions on p,

$$\begin{aligned}
-p_t &= \kappa p_{xx} + u(t,x;\kappa) - u_{data}, \\
p(T,x) &= 0, \quad p(t,0) = p(t,1) = 0.
\end{aligned} \tag{4.11}$$

Hence the second part of the equation is also zero,

$$\int_0^1 \int_0^T pu_{xx}\delta\kappa - (u(t,x;\kappa) - u_{data})w\ \ dt\ dx = 0.$$

Rearranging the above equation, and using equation (4.10) we get

$$F'(\kappa)\delta\kappa = \int_0^1 \int_0^T (u(t,x;\kappa) - u_{data})w \;\; dt \; dx = \int_0^1 \int_0^T p u_{xx}\delta\kappa \; dt \; dx$$

Therefore

$$F'(\kappa) = \int_0^1 \int_0^T p u_{xx} \; dt \; dx \tag{4.12}$$

Hence, the theorem is complete.

Theorem (4.3.1) states a way of finding the gradient of the cost function using the adjoint method. Although, the model equation (4.4) is simple and has only one unknown parameter, the theorem can be extended to much more complex problems with multiple unknown parameters. Without regard to the number of unknown parameters in the model, we have only one adjoint equation for the minimization problem. Finding the gradient of the cost function involves solving the model equation and the adjoint equation just once.

4.4 ADJOINT EQUATION FOR OPTION PRICING MODELS

4.4.1 Black-Scholes model

Consider the Black-Scholes equation for European put option

$$\begin{aligned} -\frac{dV}{dt} &= \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV, \\ V(T,S) &= \max\,(K-S,0), \\ V(t,0) = Ke^{-r(T-t)}, &\quad V(t,S_{max}) = 0. \end{aligned} \tag{4.13}$$

Let $x = \ln(S)$. Then the partial differential equation in terms of the new variable x will be:

$$\begin{aligned} -\frac{dV}{dt} &= \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial x^2} + (r - \frac{1}{2}\sigma^2)\frac{\partial V}{\partial x} - rV, \\ V(T,x) &= \max\,(K-e^x,0), \\ V(t,x_{min}) = Ke^{-r(T-t)}, &\quad V(t,x_{max}) = 0. \end{aligned} \tag{4.14}$$

Suppose $w = \dfrac{\partial V}{\partial \sigma}\delta\sigma$, be the sensitivity of V with respect to σ, then $\dfrac{\partial w}{\partial t} = \dfrac{\partial}{\partial t}\left(\dfrac{\partial V}{\partial \sigma}\delta\sigma\right)$. For notational purposes let $\overline{x}$ and $\underline{x}$ be x_{max} and x_{min} respectively. Hence, the sensitivity equation is given by:

$$-\frac{dw}{dt} = \frac{1}{2}\sigma^2\frac{\partial^2 w}{\partial x^2} + (r - \frac{1}{2}\sigma^2)\frac{\partial w}{\partial x} - rw + \sigma\delta\sigma\frac{\partial^2 V}{\partial x^2} - \sigma\delta\sigma\frac{\partial V}{\partial x}, \tag{4.15}$$

with final and boundary conditions

$$w(T, x) = 0,$$

$$w(t, \overline{x}) = 0, \quad w(t, \underline{x}) = 0.$$

Theorem 4.4.1. *Let the objective function to be minimized be:*

$$F(\sigma) = \frac{1}{2}\int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma, t, x) - V_{data}]^2 \; dx \; dt$$

The gradient of the the objective function is given by

$$F'(\sigma) = \int_0^T \int_{\underline{x}}^{\overline{x}} \sigma(-V_{xx} + V_x)p \; dx \; dt$$

where p is the solution of the adjoint equation

$$p_t = \frac{\sigma^2}{2}p_{xx} - (r - \frac{\sigma^2}{2})p_x - rp - V(\sigma, t, x) + V_{data}$$

with the initial and boundary conditions given by:

$$p(0, x) = 0,$$

$$p(t, \underline{x}) = 0, \quad p(t, \overline{x}) = 0$$

Proof:

Suppose the objective function to be minimized is

$$F(\sigma) = \frac{1}{2}\int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma, t, x) - V_{data}]^2 \; dx \; dt \tag{4.16}$$

Taking the derivative of the objective function with respect to parameter σ, we get

$$\begin{aligned} F'(\sigma) &= \int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}] \frac{\partial V}{\partial \sigma} \delta\sigma \;\; dx \; dt \\ &= \int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}] \, w \; dx \; dt \end{aligned} \tag{4.17}$$

Let $p(t,x)$ be a test function with initial condition $p(0,x) = 0$. Then we will have

$$\begin{aligned} 0 &= \int_{\underline{x}}^{\overline{x}} \int_0^T \frac{d}{dt}(w(t,x)p(t,x))dt \; dx \\ &= \int_{\underline{x}}^{\overline{x}} \int_0^T [(w_t(t,x)p(t,x) + w(t,x)p_t(t,x))] \, dt \; dx \\ &= \int_{\underline{x}}^{\overline{x}} \int_0^T \left[-\frac{1}{2}\sigma^2 w_{xx} - (r - \frac{1}{2}\sigma^2)w_x + rw - \sigma\delta\sigma V_{xx} + \sigma\delta\sigma V_x \right] p \; dt \; dx + \int_{\underline{x}}^{\overline{x}} \int_0^T w p_t \; dt \; dx \\ &= \int_0^T \left[-\frac{1}{2}\sigma^2 (w_x p|_{\underline{x}}^{\overline{x}} - \int_{\underline{x}}^{\overline{x}} w_x p_x \; dx) \right] dt - \int_0^T (r - \frac{1}{2}\sigma^2) \left[wp|_{\underline{x}}^{\overline{x}} - \int_{\underline{x}}^{\overline{x}} w p_x \; dx \right] \; dt + \\ &\quad \int_0^T \int_{\underline{x}}^{\overline{x}} (rwp + wp_t) + \sigma\delta\sigma(-V_{xx} + V_x)p \; dt \; dx \\ &= \int_0^T \left[-\frac{1}{2}\sigma^2 (w_x p|_{\underline{x}}^{\overline{x}} - w p_x|_{\underline{x}}^{\overline{x}} + \int_{\underline{x}}^{\overline{x}} w p_{xx} \; dx) \right] dt - \int_0^T [(r - \frac{1}{2}\sigma^2) wp|_{\underline{x}}^{\overline{x}} + \\ &\quad (r - \frac{1}{2}\sigma^2) \int_{\underline{x}}^{\overline{x}} w p_x \; dx] dt + \int_0^T \int_{\underline{x}}^{\overline{x}} (rwp + wp_t) + \sigma\delta\sigma(-V_{xx} + V_x)p \; dx \; dt \end{aligned}$$

Now, choosing the boundary conditions $p(t,\overline{x}) = 0$ and $p(t,\underline{x}) = 0$ on $p(t,x)$ and using the boundary conditions given for $w(t,x)$, the terms outside the integrals with respect to x are zero. Adding and subtracting $(V(\sigma,t,x) - V_{data})\, w$ inside the integral, we get:

$$\begin{aligned} 0 = \int_0^T \int_{\underline{x}}^{\overline{x}} \left[-\frac{1}{2}\sigma^2 p_{xx} + (r - \frac{1}{2}\sigma^2)p_x + rp + p_t + (V(\sigma,t,x) - V_{data}) \right] w \; dt \; dx \\ + \int_0^T \int_{\underline{x}}^{\overline{x}} [-(V(\sigma,t,x) - V_{data})w + \sigma\delta\sigma(-V_{xx} + V_x)] \, p \; dx \; dt \end{aligned} \tag{4.18}$$

The first part of equation (4.18) is zero due to the adjoint equation

$$p_t = \frac{1}{2}\sigma^2 p_{xx} - (r - \frac{1}{2}\sigma^2)p_x - rp - V(\sigma, t, x) + V_{data} \tag{4.19}$$

$$p(0, x) = 0, \quad p(t, \underline{x}) = 0, \quad p(t, \overline{x}) = 0.$$

Then the second part of the equation (4.18) is also zero,

$$\int_0^T \int_{\underline{x}}^{\overline{x}} [-(V(\sigma, t, x) - V_{data})w + \sigma\delta\sigma(-V_{xx} + V_x)]\, p\ dx\ dt = 0. \tag{4.20}$$

This implies that

$$\int_0^T \int_{\underline{x}}^{\overline{x}} (V(\sigma, t, x) - V_{data})w\ dx\ dt = \int_0^T \int_{\underline{x}}^{\overline{x}} \sigma\delta\sigma(-V_{xx} + V_x)p\ dx\ dt$$

From equation (4.17) we have

$$F'(\sigma)\delta\sigma = \int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma, t, x) - V_{data}]\, w\ dx = \int_0^T \int_{\underline{x}}^{\overline{x}} \sigma\delta\sigma(-V_{xx} + V_x)p\ dx\ dt$$

where p is the solution of the adjoint equation (4.19).

Now, the gradient of the cost function (4.16) is given by:

$$F'(\sigma) = \int_0^T \int_{\underline{x}}^{\overline{x}} \sigma(-V_{xx} + V_x)p\ dx\ dt.$$

Hence the proof is complete.

Suppose that the volatility is a function of time, $\sigma = \sigma(t)$. Then we will have:

$$\begin{aligned} F'(\sigma)\delta\sigma &= \int_0^T (\int_{\underline{x}}^{\overline{x}} (-V_{xx} + V_x)p\ dx)\sigma(t)\delta\sigma(t)\ dt \\ &= < \int_{\underline{x}}^{\overline{x}} (-V_{xx} + V_x)p\ dx\ \sigma(t), \delta\sigma(t) > \end{aligned}$$

Hence, even for time dependent volatility, obtaining the gradient of the objective function through the adjoint method does not need much modification.

4.4.2 Heston's Stochastic volatility model

In this section we follow the procedure as we did with the Black-Scholes model while deriving the the adjoint equation for gradient information. Consider the stochastic volatility model given by Heston:

$$-\frac{\partial V}{\partial t} = \frac{\nu}{2}\frac{\partial^2 V}{\partial x^2} + (r - \frac{\nu}{2})\frac{\partial V}{\partial x} + \nu\eta\rho\frac{\partial^2 V}{\partial x \partial \nu} + \frac{\eta^2}{2}\nu\frac{\partial^2 V}{\partial^2 \nu} - \lambda(\nu - \bar{\theta})\frac{\partial V}{\partial \nu} - rV \tag{4.21}$$

with the following final and boundary conditions for a put option:

$$V(T, x, \nu) = max(K - Ke^x, 0), \quad V(t, x_{min}, \nu) = max(K - Ke^{-r(T-t)}),$$
$$V(t, x, \nu_{min}) = max(K - Ke^x, 0), \quad V(t, x_{max}, \nu) = 0, \quad V(t, x, \nu_{max}) = 0$$

where $x = \ln(S/K)$. For convenience and notational simplicity $x_{min}, x_{max}, \nu_{min}, \nu_{max}$ will be replaced by $\underline{x}, \overline{x}, \underline{\nu}, \overline{\nu}$ respectively.

Suppose $w = \dfrac{\partial V}{\partial \eta}\delta\eta$, be the sensitivity of V with respect to η, then $\dfrac{\partial w}{\partial t} = \dfrac{\partial}{\partial t}\left(\dfrac{\partial V}{\partial \eta}\delta\eta\right)$. The sensitivity equation is given by:

$$\begin{aligned} -\frac{\partial w}{\partial t} = \frac{\nu}{2}\frac{\partial^2 w}{\partial x^2} + (r - \frac{\nu}{2})\frac{\partial w}{\partial x} + \nu\eta\rho\frac{\partial^2 w}{\partial x \partial \nu} + \frac{\eta^2}{2}\nu\frac{\partial^2 w}{\partial \nu^2} \\ -\lambda(\nu - \bar{\theta})\frac{\partial w}{\partial \nu} - rw + \nu\rho\frac{\partial^2 V}{\partial x \partial \nu}\delta\eta + \nu\frac{\partial^2 V}{\partial \nu^2}(\eta\delta\eta) \end{aligned} \tag{4.22}$$

with final and boundary conditions given by:

$$w(T, x, \nu) = 0, \quad w(t, \underline{x}, \nu) = 0, \quad w(t, x, \underline{\nu}) = 0,$$
$$w(t, \overline{x}, \nu) = 0, \quad w(t, x, \overline{\nu}) = 0$$

Theorem 4.4.2. *Let the objective function to be minimized be:*

$$F(\eta) = \frac{1}{2}\int_0^T \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \left[V(\eta, t, x, \nu) - V_{data}\right]^2 \, dx \, d\nu \, dt \tag{4.23}$$

The gradient of the the objective function corresponding to equation (4.23) is given by

$$F'(\eta) = \int_0^T \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} [\nu\rho V_{x\nu} + \nu\eta V_{\nu\nu}]p \; dx \; d\nu \; dt = 0$$

where p is the solution of the adjoint equation

$$\begin{aligned} p_t = \frac{\nu}{2}p_{xx} - (r - \frac{\nu}{2} - \eta\rho)p_x + \rho\eta\nu p_{x\nu} + \frac{\eta^2}{2}\nu p_{\nu\nu} \\ +(\eta^2 + \lambda(\nu - \bar{\theta}))p_\nu - (\lambda - r)p + V(\eta, \nu, t, x) - V_{data} \end{aligned} \tag{4.24}$$

with the initial and boundary conditions given by:

$$p(0, x, \nu) = 0, \quad p(t, \overline{x}, \nu) = 0, \quad p(t, \underline{x}, \nu) = 0,$$

$$p(t, x, \overline{\nu}) = 0, \; p(t, x, \underline{\nu}) = 0$$

Proof: Suppose the function to be minimized is given by equation (4.23).

Taking the derivative of the function with respect to η we get

$$\begin{aligned} F'(\eta)\delta\eta &= \int_0^T \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} [V(\eta, t, x, \nu) - V_{data}] \frac{\partial V}{\partial \eta}\delta\eta \;\; dx \; d\nu \; dt \\ &= \int_0^T \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} [V(\eta, t, x, \nu) - V_{data}] \, w \; dx \; d\nu \; dt \end{aligned} \tag{4.25}$$

where $w = \dfrac{\partial V}{\partial \eta}\delta\eta$. Let $p(t, x, \nu)$ be the solution of the adjoint equation that satisfies the final condition $p(T, x, \nu) = 0$. Following the same procedure as in the case of Black-Scholes model, we will have

$$\begin{aligned} 0 &= \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \int_0^T \frac{d}{dt}(w(t, x, \nu)p(t, x, \nu))dt \; dx \; d\nu \\ &= \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \int_0^T [(w_t(t, x, \nu)p(t, x, \nu) + w(t, x, \nu)p_t(t, x, \nu))] \, dt \; dx \; d\nu \\ &= \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \int_0^T [-\frac{1}{2}\nu w_{xx}p - (r - \frac{1}{2}\nu)w_x p - \rho\eta\nu w_{x\nu}p - \frac{\eta^2}{2}\nu w_{\nu\nu}p + \lambda(\nu - \bar{\theta})w_\nu p + rwp \\ &\quad -[\nu\rho V_{x\nu} + \eta\nu V_{\nu\nu}](\delta\eta)p \; dt \; dx \; d\nu + \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \int_0^T wp_t \; dt \; dx \; d\nu \end{aligned}$$

Now, integrating each part above with the proper initial and boundary conditions.

- For the first part we have,

$$\begin{aligned} & -\int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \int_0^T \frac{1}{2}\nu w_{xx} p \; dx \; d\nu \; dt \\ = \; & -\frac{\nu}{2} \int_0^T \int_{\underline{\nu}}^{\overline{\nu}} \left[w_x p - w p_x \right] |_{\underline{x}}^{\overline{x}} + \int_{\underline{x}}^{\overline{x}} w p_{xx} \; dx \; d\nu \; dt \\ = \; & -\int_0^T \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \frac{\nu}{2} w p_{xx} \; dx \; d\nu \; dt \end{aligned}$$

  The quantity outside the triple integral will be zero after choosing $p(t, \underline{x}, \nu) = 0, \; p(t, \overline{x}, \nu) = 0$ and using the conditions on w from equation (4.22).

- For the second part we have,

$$\begin{aligned} & -\int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \int_0^T \left[r - \frac{1}{2}\nu \right] w_x p \; dx \; d\nu \; dt \\ = \; & -(r - \frac{\nu}{2}) \int_0^T \int_{\underline{\nu}}^{\overline{\nu}} w p |_{\underline{x}}^{\overline{x}} - \int_{\underline{x}}^{\overline{x}} w p_x \; dx \; d\nu \; dt \\ = \; & \int_0^T \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} (r - \frac{\nu}{2}) w p_x \; dx \; d\nu \; dt \end{aligned}$$

  The quantity outside the triple integral is zero after choosing $p(t, \overline{x}, \nu) = 0$, using the conditions defined for p above and the conditions on w from equation (4.22).

- For the third part we have,

$$\begin{aligned} & -\rho\eta \int_{\underline{\nu}}^{\overline{\nu}} \int_{\underline{x}}^{\overline{x}} \int_0^T \left[w_{x\nu} \nu p \right] \; dx \; d\nu \; dt \\ = \; & -\rho\eta \int_0^T \int_{\underline{\nu}}^{\overline{\nu}} \left[w_\nu \nu p |_{\underline{x}}^{\overline{x}} - \int_{\underline{x}}^{\overline{x}} w_\nu \nu p_x \right] \; dx \; d\nu \; dt \\ = \; & -\rho\eta \int_0^T \int_{\underline{x}}^{\overline{x}} w(\nu p_x) |_{\underline{\nu}}^{\overline{\nu}} + \int_{\underline{\nu}}^{\overline{\nu}} w(\nu p_{x\nu} + p_x) \; d\nu \; dx \; dt \\ = \; & \int_0^T \int_{\underline{x}}^{\overline{x}} \int_{\underline{\nu}}^{\overline{\nu}} -\rho\eta(\nu p_{x\nu} + p_x) w \; d\nu \; dx \; dt \end{aligned}$$

The quantity outside the triple integral is zero using the conditions defined for p above and the conditions on w from equation (4.22).

- For the fourth part we have,

$$\begin{aligned}
& -\frac{1}{2}\eta^2 \int_0^T \int_{\underline{x}}^{\overline{x}} \int_{\underline{\nu}}^{\overline{\nu}} w_{\nu\nu}\nu p \; d\nu \; dx \; dt \\
= \; & -\frac{1}{2}\eta^2 \int_0^T \int_{\underline{x}}^{\overline{x}} \left(w_\nu(\nu p)|_{\underline{\nu}}^{\overline{\nu}} - \int_{\underline{\nu}}^{\overline{\nu}} w_\nu(\nu p)_\nu \; d\nu \right) dx \; dt \\
= \; & -\frac{\eta^2}{2} \int_0^T \int_{\underline{x}}^{\overline{x}} \nu p w_\nu - (\nu p)_\nu w|_{\underline{\nu}}^{\overline{\nu}} + \int_{\underline{\nu}}^{\overline{\nu}} (\nu p)_{\nu\nu} w \; d\nu \; dx \; dt \\
= \; & -\frac{\eta^2}{2} \int_0^T \int_{\underline{x}}^{\overline{x}} \int_{\underline{\nu}}^{\overline{\nu}} (\nu p_\nu + p)_\nu w \; d\nu \; dx \; dt \\
= \; & \int_0^T \int_{\underline{x}}^{\overline{x}} \int_{\underline{\nu}}^{\overline{\nu}} -\frac{\eta^2}{2}(\nu p_{\nu\nu} + 2p_\nu) w \; d\nu \; dx \; dt
\end{aligned}$$

The integral outside the triple integral is zero after choosing $p(t, x, \overline{\nu}) = 0,\ p(t, x, \underline{\nu}) = 0$, using the conditions defined for p above and using the conditions on w from equation (4.22).

- For the fifth part we have,

$$\begin{aligned}
& \lambda \int_0^T \int_{\underline{x}}^{\overline{x}} \int_{\underline{\nu}}^{\overline{\nu}} (\nu - \bar{\theta}) w_\nu p \; d\nu \; dx \; dt \\
= \; & \lambda \int_0^T \int_{\underline{x}}^{\overline{x}} \left(w(\nu - \bar{\theta}) p|_{\underline{\nu}}^{\overline{\nu}} - \int_{\underline{\nu}}^{\overline{\nu}} w((\nu - \bar{\theta}) p_\nu + p) \; d\nu \right) \; dx \; dt \\
= \; & -\lambda \int_0^T \int_{\underline{x}}^{\overline{x}} \int_{\underline{\nu}}^{\overline{\nu}} ((\nu - \bar{\theta}) p_\nu w + pw) \; d\nu \; dx \; dt
\end{aligned}$$

The integral outside the triple integral is zero using the conditions on p as defined above and the conditions on w from equation (4.22).

Now, combining all these parts we get:

$$\begin{aligned}
&\int_{\underline{\nu}}^{\overline{\nu}}\int_{\underline{x}}^{\overline{x}}\int_0^T [w_t p + w p_t]\ dt\ dx\ d\nu \\
= &\int_{\underline{\nu}}^{\overline{\nu}}\int_{\underline{x}}^{\overline{x}}\int_0^T -\frac{\nu}{2}wp_{xx} + (r-\frac{\nu}{2})wp_x - \rho\eta(\nu p_{x\nu} + p_x)w - \frac{\eta^2}{2}(\nu p_{\nu\nu} + 2p_\nu)w - \\
&\lambda((\nu-\bar{\theta})p_\nu + p)w + rpw - wp_t - [\nu\rho V_{x\nu} + \eta\nu V_{\nu\nu}](\delta\eta)\ dt\ dx\ d\nu \\
= &\int_0^T\int_{\underline{\nu}}^{\overline{\nu}}\int_{\underline{x}}^{\overline{x}} [-\frac{\nu}{2}p_{xx} + (r-\frac{\nu}{2})p_x - \rho\eta(\nu p_{x\nu} + p_x) - \frac{\eta^2}{2}(\nu p_{\nu\nu} + 2p_\nu) - \\
&\lambda(\nu-\bar{\theta})p_\nu - \lambda p + rp + p_t]w - [\nu\rho V_{x\nu} + \nu\eta V_{\nu\nu}](\delta\eta)p\ dx\ d\nu\ dt \qquad (4.26)
\end{aligned}$$

Now, adding and subtracting $(V(\eta,t,x,\nu) - V_{data})\,w$ inside the integral, and setting the first part of equation (4.26) to zero, we arrive at the adjoint equation

$$\begin{aligned}
p_t = \frac{\nu}{2}p_{xx} - (r - \frac{\nu}{2} - \eta\rho)p_x + \rho\eta\nu p_{x\nu} + \frac{\eta^2}{2}\nu p_{\nu\nu} + (\eta^2 + \lambda(\nu-\bar{\theta}))p_\nu \\
-(r-\lambda)p + V(\eta,t,x,\nu) - V_{data}. \qquad (4.27)
\end{aligned}$$

If the first part of the equation (4.26) is zero then the remaining part of equation (4.26) should also be zero. Therefore, we get

$$\int_0^T\int_{\underline{\nu}}^{\overline{\nu}}\int_{\underline{x}}^{\overline{x}} [\nu\rho V_{x\nu} + \nu\eta V_{\nu\nu}](\delta\eta)p - [V(\eta,t,x,\nu) - V_{data}]w\ dx\ d\nu\ dt = 0$$

or

$$\begin{aligned}
&\int_0^T\int_{\underline{\nu}}^{\overline{\nu}}\int_{\underline{x}}^{\overline{x}} [\nu\rho V_{x\nu} + \nu\eta V_{\nu\nu}](\delta\eta)p\ dx\ d\nu\ dt \\
= &\int_0^T\int_{\underline{\nu}}^{\overline{\nu}}\int_{\underline{x}}^{\overline{x}} [V(\eta,t,x,\nu) - V_{data}]w\ dx\ d\nu\ dt
\end{aligned}$$

From equation (4.25) the gradient of the objective function is given by

$$F'(\eta)\delta\eta = \int_0^T\int_{\underline{\nu}}^{\overline{\nu}}\int_{\underline{x}}^{\overline{x}} [V(\eta,t,x,\nu) - V_{data}]w\ dx\ d\nu\ dt.$$

Therefore, we get

$$F'(\eta) = \int_0^T\int_{\underline{\nu}}^{\overline{\nu}}\int_{\underline{x}}^{\overline{x}} [\nu\rho V_{x\nu} + \nu\eta V_{\nu\nu}]p\ dx\ d\nu\ dt$$

where, p is the solution of the adjoint equation (4.27).

Hence the proof is complete.

4.4.3 Adjoint equation for the PIDE model

Consider the PIDE model with boundary and final conditions for a put option given by the equation

$$-\frac{dV}{dt} = \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - \kappa\lambda) S \frac{\partial V}{\partial S} - (r+\lambda)V + \lambda \int_0^\infty V(t, JS) g(J)\ dJ \quad (4.28)$$
$$V(T,S) = \max(K - S, 0), \quad V(t,0) = Ke^{-r(T-t)}, \quad V(t, S_{max}) = 0.$$

Let $S = e^x$ and $J = e^y$ so that $V(S) = V(e^x) = \bar{V}(x)$ and $V(SJ) = V(e^x e^y) = V(e^{x+y}) = \bar{V}(x+y)$.

Then,

$$\frac{\partial \bar{V}}{\partial x} = S\frac{\partial V}{\partial S} \quad \text{and} \quad \frac{\partial^2 \bar{V}}{\partial x^2} = S\frac{\partial V}{\partial S} + S^2 \frac{\partial^2 V}{\partial S^2}.$$

Similarly, the integral $\displaystyle\int_0^\infty V(JS) g(J)\ dJ = \int_{-\infty}^\infty \bar{V}(x+y) f(y)\ dy$ where $f(y) = g(e^y)e^y$. The partial integro-differential equation (4.29) in terms of the new variables x and y will be:

$$-\frac{d\bar{V}}{dt} = \frac{1}{2}\sigma^2 \frac{\partial^2 \bar{V}}{\partial x^2} + (r - \frac{1}{2}\sigma^2 - \kappa\lambda)\frac{\partial \bar{V}}{\partial x} - (r+\lambda)\bar{V} + \lambda \int_{-\infty}^\infty \bar{V}(t, x+y) f(y)\ dy.$$

For notational simplicity we can replace $\bar{V}(t,x)$ with $V(t,x)$ to get:

$$-\frac{dV}{dt} = \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial x^2} + (r - \frac{1}{2}\sigma^2 - \kappa\lambda)\frac{\partial V}{\partial x} - (r+\lambda)V + \lambda \int_{-\infty}^\infty V(t, x+y) f(y)\ dy. \quad (4.29)$$

From now onwards for notational purposes let $\overline{x}$ and $\underline{x}$ be x_{max} and x_{min} respectively. Suppose $w = \dfrac{\partial V}{\partial \sigma}\delta\sigma$, be the sensitivity of V with respect to σ, then $\dfrac{\partial w}{\partial t} = \dfrac{\partial}{\partial t}\left(\dfrac{\partial V}{\partial \sigma}\delta\sigma\right)$. The sensitivity equation corresponding to the equation (4.29) is given

by:

$$-\frac{dw}{dt} = \frac{1}{2}\sigma^2\frac{\partial^2 w}{\partial x^2} + (r - \frac{1}{2}\sigma^2 - \kappa\lambda)\frac{\partial w}{\partial x} - (r+\lambda)w + \sigma\delta\sigma(\frac{\partial^2 V}{\partial x^2} - \frac{\partial V}{\partial x}) + \lambda\int_{-\infty}^{\infty} w(t, x+y)f(y)\ dy \quad (4.30)$$

with the conditions

$$w(T,x) = 0, \quad w(t,x) = 0 \text{ for } x \in [-\infty, \underline{x}] \text{ and } x \in [\overline{x}, \infty]. \quad (4.31)$$

Theorem 4.4.3. *Let the objective function to be minimized be:*

$$F(\sigma) = \frac{1}{2}\int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}]^2\ dx\ dt \quad (4.32)$$

The gradient of the the objective function corresponding to equation (4.32) is given by

$$F'(\sigma) = \int_0^T \int_{\underline{x}}^{\overline{x}} \sigma(-V_{xx} + V_x)p\ dx\ dt$$

where p is the solution of the adjoint equation

$$p_t = \frac{1}{2}\sigma^2 p_{xx} - (r - \frac{1}{2}\sigma^2 - \kappa\lambda)p_x - (r+\lambda)p + \lambda(\int_{-\infty}^{\infty} p(t, x-y)f(y)\ dy\ ) - (V(\sigma,t,x) - V_{data}) \quad (4.33)$$

With initial and boundary conditions

$p(0,x) = 0$, $p(t,x) = 0$ *for* $x \in [-\infty, \underline{x}]$ *and* $x \in [\overline{x}, \infty]$

Proof

Suppose the objective function to be minimized is

$$F(\sigma) = \frac{1}{2}\int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}]^2\ dx\ dt \quad (4.34)$$

Taking the derivative of the objective function with respect to parameter σ, we get

$$\begin{aligned} F'(\sigma) &= \int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}]\frac{\partial V}{\partial \sigma}\delta\sigma\ \ dx\ dt \\ &= \int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}]\, w\ dx\ dt \end{aligned} \quad (4.35)$$

Let $p(t,x)$ be the solution of the adjoint equation with the initial conditions $p(0,x) = 0$.

Following the same procedure as we did with the Black-Scholes PDE to find the adjoint equation we will have

$$\begin{aligned}
0 &= \int_{\underline{x}}^{\overline{x}} \int_0^T \frac{d}{dt}(w(t,x)p(t,x))dt\ dx \\
&= \int_{\underline{x}}^{\overline{x}} \int_0^T \left[\frac{\partial w}{\partial t}p + w\frac{\partial p}{\partial t}\right]\ dt\ dx \\
&= \int_0^T \int_{\underline{x}}^{\overline{x}} [-\frac{1}{2}\sigma^2 w_{xx}p - (r - \frac{1}{2}\sigma^2 - \kappa\lambda)w_x p + (r+\lambda)wp - \\
&\quad \lambda(\int_{-\infty}^{\infty} w(t,x+y)f(y)\ dy\ )p]\ dt\ dx + \int_{\underline{x}}^{\overline{x}} \int_0^T [-\sigma\delta\sigma(V_{xx} - V_x)p + wp_t]\ dt\ dx
\end{aligned}$$

Let us consider each part of the integral separately. Integrating by parts with respect to the variable x we get:

- The first term becomes

$$\begin{aligned}
\int_{\underline{x}}^{\overline{x}} -\frac{1}{2}\sigma^2 w_{xx}p\ dx &= -\frac{1}{2}\sigma^2([w_x p - wp_x]_{\underline{x}}^{\overline{x}} + \int_{\underline{x}}^{\overline{x}} wp_{xx}\ dx) \\
&= -\frac{1}{2}\sigma^2 \int_{\underline{x}}^{\overline{x}} wp_{xx}\ dx
\end{aligned}$$

  after using the boundary conditions $w(t,\underline{x}) = w(t,\overline{x}) = 0$ and choosing the adjoint boundary conditions as $p(t,\overline{x}) = p(t,\underline{x}) = 0$.

- The second term becomes

$$\begin{aligned}
&\int_{\underline{x}}^{\overline{x}} -(r - \frac{1}{2}\sigma^2 - \kappa\lambda)w_x p\ dx = -(r - \frac{1}{2}\sigma^2 - \kappa\lambda)[wp]_{\underline{x}}^{\overline{x}} - \int_{\underline{x}}^{\overline{x}} wp_x\ dx] \\
= &\int_{\underline{x}}^{\overline{x}} (r - \frac{1}{2}\sigma^2 - \kappa\lambda)wp_x\ dx.
\end{aligned}$$

- The integral term becomes

$$\int_{\underline{x}}^{\overline{x}} \left(\int_{-\infty}^{\infty} w(t, x+y) f(y)\ dy \right)\ p(t,x)\ dx$$

$$= \int_{-\infty}^{\infty} \left(\int_{\underline{x}}^{\overline{x}} w(t, x+y) p(t,x)\ dx \right) f(y)\ dy$$

$$= \int_{-\infty}^{\infty} \left(\int_{\underline{x}}^{\overline{x}} w(t,x) p(t, x-y)\ dx \right) f(y)\ dy \quad (4.36)$$

$$= \int_{\underline{x}}^{\overline{x}} \left(\int_{-\infty}^{\infty} p(t, x-y) f(y)\ dy \right) w(t,x)\ dx.$$

See the appendix for detail.

Now, combining all of the above separate parts, we get:

$$\begin{aligned} & \int_{\underline{x}}^{\overline{x}} \int_0^T \frac{d}{dt}(w(t,x)p(t,x))dt\ dx \\ = & \int_{\underline{x}}^{\overline{x}} \int_0^T [-\frac{1}{2}\sigma^2 p_{xx} w + (r - \frac{1}{2}\sigma^2 - \kappa\lambda) p_x w + (r+\lambda) p w + w p_t - \\ & \lambda(\int_{-\infty}^{\infty} p(t, x-y) f(y)\ dy\ ) w]\ dt\ dx + \int_{\underline{x}}^{\overline{x}} \int_0^T [-\sigma\delta\sigma(V_{xx} - V_x) p]\ dt\ dx. \end{aligned} \quad (4.37)$$

Adding and subtracting $(V(\sigma, t, x) - V_{data})\, w$ inside the integral and setting the first part to zero, we get:

$$\begin{aligned} p_t = \frac{1}{2}\sigma^2 p_{xx} - (r - \frac{1}{2}\sigma^2 - \kappa\lambda) p_x - (r+\lambda) p + \\ \lambda(\int_{-\infty}^{\infty} p(t, x-y) f(y)\ dy\ ) - (V(\sigma, t, x) - V_{data}). \end{aligned} \quad (4.38)$$

With initial and boundary conditions for p

$p(0, x) = 0$, $p(t, x) = 0$ for $x \in [-\infty, \underline{x}]$ and $x \in [\overline{x}, \infty]$.

If the first part of the above integral is zero, the remaining part of the same integral should also be zero, i.e.

$$\int_{\underline{x}}^{\overline{x}} \int_0^T [-\sigma\delta\sigma(V_{xx} - V_x)p] - [V(\sigma, t, x) - V_{data}]w \ dt \ dx = 0,$$

or

$$\int_{\underline{x}}^{\overline{x}} \int_0^T [-\sigma\delta\sigma(V_{xx} - V_x)p] \ dt \ dx = \int_{\underline{x}}^{\overline{x}} \int_0^T [V(\sigma, t, x) - V_{data}]w \ dt \ dx.$$

But, from equation (4.35)

$$F'(\sigma)\delta\sigma = \int_{\underline{x}}^{\overline{x}} \int_0^T [V(\sigma, t, x) - V_{data}]w \ dt \ dx$$

Therefore

$$F'(\sigma) = \int_{\underline{x}}^{\overline{x}} \int_0^T [-\sigma(V_{xx} - V_x)p] \ dt \ dx. \tag{4.39}$$

Hence the theorem is complete.

4.4.4 Adjoint of the PIDE model with respect to the jump distribution

Consider the PIDE model given by (4.29). Suppose $w = \dfrac{\partial V}{\partial f}\phi$, is the sensitivity of V with respect to the Lévy measure f, then the sensitivity equation corresponding to the equation (4.29) is given by:

$$\begin{aligned} -\frac{dw}{dt} = \frac{1}{2}\sigma^2\frac{\partial^2 w}{\partial x^2} + (r - \frac{1}{2}\sigma^2 - \kappa\lambda)\frac{\partial w}{\partial x} - (r + \lambda)w + \\ \lambda \int_{-\infty}^{\infty} w(t, x + y)f(y) \ dy + \int_{-\infty}^{\infty} V(t, x + y)\phi(y) \ dy \end{aligned} \tag{4.40}$$

with the final and boundary conditions $w(T, x) = 0, \ w(t, x) = 0$ for $x \in (-\infty, \underline{x}]$ and $x \in [\overline{x}, \infty)$.

Theorem 4.4.4. *Let the objective function to be minimized be:*

$$\min_f F(f) = \frac{1}{2}\int_0^T \int_{\underline{x}}^{\overline{x}} [V(f, t, x) - V_{data}]^2 \ dx \ dt. \tag{4.41}$$

The gradient of the the objective function (4.41) with respect to Lévy measure $f(y)$ corresponding to equation (4.29) is given by

$$F'(f) = \lambda \int_{\mathbf{R}} \int_0^T \left(\int_{\underline{x}}^{\overline{x}} V(t, x+y) p(t,x)\ dx \right)\ dt \phi(y)\ dy$$

where p is the solution of the adjoint equation

$$\begin{aligned} p_t = \frac{1}{2}\sigma^2 p_{xx} - (r - \frac{1}{2}\sigma^2 - \kappa\lambda) p_x - (r+\lambda)p \\ +\lambda \int_{-\infty}^{\infty} p(t, x-y) f(y)\ dy - (V(t,x) - V_{data}). \end{aligned} \tag{4.42}$$

The initial and boundary conditions given by $p(0,x) = 0$, and

$$p(t,x) = \begin{cases} 0 & \text{if } x \in (-\infty, \underline{x}] \\ \\ 0 & \text{if } x \in [\overline{x}, \infty) \end{cases}$$

Proof

Suppose the objective function to be minimized is given by equation (4.41). Then the derivative of the objective function with respect to the Lévy measure $f(y)$ is:

$$F'(f)\phi = \int_0^T \int_{\underline{x}}^{\overline{x}} (V(t,x) - V_{data}) \left(\frac{\partial V}{\partial f} \phi \right)\ dx\ dt \tag{4.43}$$

Lets choose a test function $p(t,x)$ with the following boundary and initial conditions: $p(0,x) = 0$, and

$$p(t,x) = \begin{cases} 0 & \text{if } x \in (-\infty, \underline{x}] \\ \\ 0 & \text{if } x \in [\overline{x}, \infty) \end{cases}$$

Following the same procedure as we did in proving the previous theorem, we have

$$\begin{aligned} 0 &= \int_{\underline{x}}^{\overline{x}} (w(T,x)p(T,x) - w(0,x)p(0,x))\ dx = \int_{\underline{x}}^{\overline{x}} \int_0^T \frac{d}{dt}(w(t,x)p(t,x)) dt\ dx \\ &= \int_{\underline{x}}^{\overline{x}} \int_0^T \left[\frac{\partial w}{\partial t} p + w \frac{\partial p}{\partial t} \right]\ dt\ dx \\ &= \int_{\underline{x}}^{\overline{x}} \int_0^T \Big[-\frac{1}{2}\sigma^2 w_{xx} p - (r - \frac{1}{2}\sigma^2 - \kappa\lambda) w_x p + (r+\lambda) wp - \\ &\quad \lambda \int_{-\infty}^{\infty} w(t, x+y) f(y)\ dy\ p(t,x) + \int_{-\infty}^{\infty} V(t, x+y)\phi(y) dy\ p(t,x) + w p_t \Big]\ dt\ dx \end{aligned}$$

Considering each term separately and integrating by parts we get,

$$\begin{aligned} 0 \;=\; & \int_{\underline{x}}^{\overline{x}} \int_0^T [-\frac{1}{2}\sigma^2 p_{xx} w + (r - \frac{1}{2}\sigma^2 - \kappa\lambda) p_x w + (r+\lambda) p w + w p_t - \\ & \lambda(\int_{-\infty}^{\infty} p(t, x-y) f(y)\ dy\ ) w]\ dt\ dx + \lambda \int_{\underline{x}}^{\overline{x}} \int_0^T \int_{-\infty}^{\infty} V(t, x+y)\phi(y)\ dy\ p(t,x)\ dt\ dx \end{aligned} \tag{4.44}$$

Adding and subtracting $(V(f,t,x) - V_{data})\, w$ inside the integral we get:

$$\begin{aligned} & \int_{\underline{x}}^{\overline{x}} \int_0^T \Big(-\frac{1}{2}\sigma^2 p_{xx} + (r - \frac{1}{2}\sigma^2 - \kappa\lambda) p_x + (r+\lambda) p + p_t - \\ & \lambda \int_{-\infty}^{\infty} p(t, x-y) f(y)\ dy\ + (V(f,t,x) - V_{data}) \Big)\, w(t,x) + \\ & \lambda \int_{\underline{x}}^{\overline{x}} \int_0^T \int_{-\infty}^{\infty} V(t, x+y)\phi(y)\ dy\ p(t,x)\ dx\ dt - \int_{\underline{x}}^{\overline{x}} \int_0^T (V(f,t,x) - V_{data}) w(t,x)\ dt\ dx = 0 \end{aligned}$$

The first part of the equation is zero due to the adjoint equation

$$\begin{aligned} p_t = \frac{1}{2}\sigma^2 p_{xx} - (r - \frac{1}{2}\sigma^2 - \kappa\lambda) p_x - (r+\lambda) p + \\ \lambda \int_{-\infty}^{\infty} p(t, x-y) f(y)\ dy\ - (V(f,t,x) - V_{data}) = 0 \end{aligned} \tag{4.45}$$

The remaining part to also zero,

$$\begin{aligned} & \int_0^T \int_{\underline{x}}^{\overline{x}} (V(f,t,x) - V_{data}) w(t,x)\ dt\ dx \\ & = \lambda \int_0^T \int_{\underline{x}}^{\overline{x}} \int_{-\infty}^{\infty} V(t, x+y)\phi(y)\ dy\ p(t,x)\ dt\ dx \end{aligned}$$

But from equation (4.43), we have

$$F'(f)\phi = \int_0^T \int_{\underline{x}}^{\overline{x}} (V(f,t,x) - V_{data}) w(t,x)\ dt\ dx \tag{4.46}$$

Therefore, we get

$$F'(f)\phi = \lambda \int_{-\infty}^{\infty} \int_0^T \left(\int_{\underline{x}}^{\overline{x}} V(t, x+y) p(t,x)\ dx \right)\ dt\ \phi(y)\ dy \tag{4.47}$$

Hence, the proof is complete.

4.5 FINITE DIFFERENCE FOR ADJOINT EQUATIONS

Consider the objective function

$$F(\sigma) = \frac{1}{2}\int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}]^2 \; dx \; dt \tag{4.48}$$

and its derivative with respect to σ as

$$F'(\sigma) = \int_0^T \int_{\underline{x}}^{\overline{x}} \sigma(-V_{xx} + V_x)p \; dx \; dt \tag{4.49}$$

where p is the solution of the adjoint equation

$$-\frac{1}{2}\sigma^2 p_{xx} + (r - \frac{1}{2}\sigma^2)p_x + rp + p_t + V(\sigma,t,x) - V_{data} = 0 \tag{4.50}$$
$$p(0,x) = 0, \quad p(t,\underline{x}) = 0, \quad p(t,\overline{x}) = 0.$$

Discretizing the corresponding derivatives of p, we get:

- $\dfrac{\partial p}{\partial t} = \dfrac{p_{i+1,j} - p_{i,j}}{\delta t}$
- $\dfrac{\partial p}{\partial x} = \dfrac{p_{i,j+1} - p_{i,j-1}}{2\delta x}$
- $\dfrac{\partial^2 p}{\partial x^2} = \dfrac{p_{i,j+1} - 2p_{i,j} + p_{i,j-1}}{(\delta x)^2}$

The discretized equation (4.50) becomes:

$$\frac{p_{i+1,j} - p_{i,j}}{\delta t} - \frac{\sigma^2}{2}\left(\frac{p_{i,j+1} - 2p_{i,j} + p_{i,j-1}}{(\delta x)^2}\right) +$$
$$(r - \frac{\sigma^2}{2})\left(\frac{p_{i,j+1} - p_{i,j-1}}{2\delta x}\right) + rp_{i,j} + V_{i,j} - V_{data} = 0. \tag{4.51}$$

The discretized adjoint equation is then explicitly solved as :

$$p_{i+1,j} = p_{i,j} + \frac{\sigma^2}{2}\delta t\left(\frac{p_{i,j+1} - 2p_{i,j} + p_{i,j-1}}{(\delta x)^2}\right) -$$
$$\delta t(r - \frac{\sigma^2}{2})\left(\frac{p_{i,j+1} - p_{i,j-1}}{2\delta x}\right) - \delta t \; r \; p_{i,j} - \delta t(V_{i,j} - V_{data}). \tag{4.52}$$

The derivative equation (4.49) in terms of discretization of adjoint equation can be written as:

$$F'(\sigma) = \sum_{i=1}^{N}\sum_{j=1}^{M} \sigma \left[\frac{-V_{i,j+1} + 2V_{i,j} - V_{i,j-1}}{(\delta x)^2} + \frac{V_{i,j+1} - V_{i,j-1}}{2\delta x}\right] p_{i,j}\delta x\ \delta t \qquad (4.53)$$

where V is the solution of the Black-Scholes PDE and p the solution of the adjoint equation.

The second way to compute the same gradient is through the sensitivity equation method. Consider the objective function given by equation (4.48). The derivative of the objective function with respective to the parameter σ is

$$\begin{aligned} F'(\sigma) &= \int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}] \frac{\partial V}{\partial \sigma}\delta\sigma \ \ dx\ dt \\ &= \int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma,t,x) - V_{data}]\, W\ dx\ dt \end{aligned} \qquad (4.54)$$

where W is the solution of the sensitivity equation given by equation (4.15). We see that both equations (4.49) and (4.54) give the derivative of the same objective function (4.48) is a different way.

The approximate derivative of (4.48) can also be found using finite difference as follows:

$$F'(\sigma) = \frac{\int_0^T \int_{\underline{x}}^{\overline{x}} [V(\sigma+\delta\sigma,t,x) - V_{data}] \ \ - [V(\sigma,t,x) - V_{data}]\ dx\ dt}{\delta\sigma}.$$

However, the gradient obtained from finite difference is an approximate gradient and does not match exactly with the gradient obtained from the adjoint or sensitivity equation methods.

Table (4.1) shows the computed derivative values using different methods. When Δt and Δx are small all the derivative values match quite well except the one obtained from the exact Black-Scholes formula. Since the sensitivity equation method and the adjoint equation method compute the derivative of the discretized Black-Scholes equation, it is going to be different as shown in the table. As we solve one sensitivity

Table 4.1. Calculating $\frac{\partial V}{\partial \sigma}$, the derivative with respect to σ when $\sigma = 0.3$ for Black-Scholes equation three different methods. Other parameters in the model are fixed.

Δt	Δx	Finite Diff	Exact Black-Scholes Diff	Sensitivity Eqn	Adjoint
0.0200	0.1638	-3.1145	-3.0755	-3.1146	-3.1146
0.0133	0.1092	-3.0841	-3.1271	-3.0842	-3.0842
0.0100	0.0819	-3.0772	-3.1169	-3.0774	-3.0774
0.0067	0.0546	-3.0751	-3.0836	-3.0752	-3.0752
0.0050	0.0409	-3.0756	-3.0585	-3.0757	-3.0757
0.0033	0.0273	-3.0621	-3.0676	-3.0623	-3.0623
0.0025	0.0205	-3.0598	-3.0592	-3.0599	-3.0599

and one adjoint equation to compute the gradient in this example, computational time is comparable in this case.

4.6 ADJOINT OF FINITE DIFFERENCE

Let the gradient of the cost function (4.48) in discrete form be

$$F'(\sigma) \approx \sum_{i=1}^{N} \sum_{j=1}^{M} (V_{ij} - V_{data})\, W_{ij}\ \delta t\ \delta x \tag{4.55}$$

where W_{ij} solves the descretized version of the sensitivity equation

$$-\frac{dW}{dt} = \frac{1}{2}\sigma^2 \frac{\partial^2 W}{\partial x^2} + (r - \frac{1}{2}\sigma^2)\frac{\partial W}{\partial x} - rW + \sigma\delta\sigma\frac{\partial^2 V}{\partial x^2} - \sigma\delta\sigma\frac{\partial V}{\partial x} \tag{4.56}$$

with the conditions $W(T,x) = 0, \quad W(t, x_{max}) = W(t, x_{min}) = 0.$

Writing equation (4.56) in discrete form

$$\begin{aligned} W_{i,j} = W_{i+1,j} + \frac{\sigma^2}{2}\frac{\delta t}{\delta x^2}\left(W_{i+1,j+1} - 2W_{i+1,j} + W_{i+1,j-1}\right) + \\ (r - \frac{\sigma^2}{2})\frac{\delta t}{\delta x}\left(W_{i+1,j+1} - W_{i+1,j-1}\right) + r\ \delta t W_{i+1,j} + \\ \sigma\frac{\delta t}{\delta x^2}\left(V_{i+1,j+1} - 2V_{i+1,j} + V_{i+1,j-1}\right) - \frac{\sigma}{2}\frac{\delta t}{\delta x}\left(V_{i+1,j+1} - V_{i+1,j-1}\right) \end{aligned} \tag{4.57}$$

Let $W_i,\ G_i\ , H_i\ \in \mathbb{R}^M$. Define $H_i = (V_i - V_{data_i})$. Then we can rewrite equation (4.57) as, $W_i = AW_{i+1} + G_{i+1}$ for $i = N-1, ..., 0$ and $A \in \mathbb{R}^{M\times M}$ with $W_N = 0$. Therefore, $F'(\sigma) = \sum_{i=0}^{N} {H_i}^T W_i$ and $W_i - AW_{i+1} = G_{i+1}$ for $i = N-1, ..., 0$.

Theorem 4.6.1. *Let $W_{i,j}$ solve the discretized sensitivity equation (4.57) such that $W_i = AW_{i+1} + G_{i+1}$. Then the gradient of the objective function (4.48) is given by*

$$F'(\sigma) = \sum_{i=0}^{N} p_{i+1}^T G_{i+1}$$

where p solves the discrete adjoint equation

$$p_{i+1} = A^T p_i + H_i,$$

for $i = 1, ..., N$.

Proof: Let $p_i \in \mathbb{R}^M$. We have

$$W_i - AW_{i+1} = G_{i+1}$$

For $i = N-1, ..., 0$. Multiplying by p_{i+1}^T both sides, we get:

$$\begin{aligned} p_{i+1}^T(W_i - AW_{i+1}) &= p_{i+1}^T G_{i+1} \\ p_{i+1}^T W_i - (A^T p_{i+1})^T W_{i+1} &= p_{i+1}^T G_{i+1} \end{aligned}$$

Summing over all $i = 0, ..., N-1$,

$$\begin{aligned}\sum_{i=0}^{N-1} p_{i+1}^T G_{i+1} &= \sum_{i=0}^{N-1} p_{i+1}^T W_i - \sum_{i=0}^{N-1} (A^T p_{i+1})^T W_{i+1} \\ &= \sum_{i=0}^{N-1} p_{i+1}^T W_i - \sum_{i=1}^{N} (A^T p_i)^T W_i \\ &= \sum_{i=1}^{N-1} (p_{i+1} - A^T p_i)^T W_i + {p_1}^T W_0 - (A^T p_N)^T W_N\end{aligned}$$

Since $W_N = 0$, $(A^T p_N)^T W_N$=0. Now adding and subtracting $\sum_{i=0}^{N-1} {H_i}^T W_i$ and setting the first part to 0, we get $p_{i+1} = A^T p_i + H_i$ for $i = 1, ..., N$ with $p_0 = H_0$. From the second part we get

$$\sum_{i=0}^{N-1} p_{i+1}^T G_{i+1} = \sum_{i=1}^{N} p_i^T G_i = \sum_{i=1}^{N} {H_i}^T W_i = F'(\sigma).$$

Hence proof is complete.

4.7 FINITE DIFFERENCE FOR ADJOINT OF PIDE MODEL

Let us consider the PIDE model given by equation (4.29) and the corresponding sensitivity equation with respect to the parameter σ given by equation (4.56). The adjoint equation corresponding to the PIDE (4.29) with the time variable t is

$$\begin{aligned}p_t = \frac{1}{2}\sigma^2 p_{xx} - (r - \frac{1}{2}\sigma^2 - \kappa\lambda)p_x - (r + \lambda)p + \\ \lambda \int_{\infty}^{\infty} p(t, x - y) f(y)\ dy + V(\sigma, t, x) - V_{data}.\end{aligned} \tag{4.58}$$

Before we solve the adjoint equation (4.58) we perform a variable transformation as done with the PIDE model. Let $z = x - y$, then $dz = -dy$. $f(y)dy = f(x - z)(-dz) = -f(x - z)dz$. In the case of normally distributed jump distributions (as in Merton's model), f is symmetric and we can write $f(x - z) = f(z - x)$. Hence the new adjoint equation with a variable transformation is:

$$\begin{aligned}p_t = \frac{1}{2}\sigma^2 p_{xx} - (r - \frac{1}{2}\sigma^2 - \kappa\lambda)p_x - (r + \lambda)p + \\ \lambda \int_{\infty}^{\infty} p(t, z) f(z - x)\ dz - [V(\sigma, t, x) - V_{data}]\end{aligned} \tag{4.59}$$

with vanishing initial and boundary conditions. We would like to solve the adjoint equation (4.59) efficiently using the similar procedure as solving the pricing equation given in Section 3.7

Any of the numerical methods for solving the PIDE model proposed by [1, 3, 19, 13, 27, 44] and many others can be used to solve the adjoint of the PIDE. Because of its similarity between the given pricing PIDE and the adjoint equation, we can use a similar finite difference scheme for both the pricing and the adjoint equation. Here we propose the iterative method used by [1, 44] to solve the adjoint equation. Pricing PIDE is solved forward in time variable τ where as the adjoint equation is solved backward in τ. If we change the time variable from τ to t, then equation (4.59) becomes a forward problem. Hence we can use the forward solver with minor adjustments to solve the adjoint problem.

Using the discretization scheme used by [1, 44] in the PIDE model we have the following similar discretization for the adjoint equation (4.59):

$$p_t(t_m, x_i) \approx \begin{cases} (\frac{3}{2}p_i^m - 2p_i^{m-1} + \frac{1}{2}p_i^{m-2})/k & \text{for } m \geq 2 \\ (p_i^m - p_i^{m-1})/k & \text{for } m = 1 \end{cases}$$

$$\begin{aligned} p_{xx}(t_m, x_i) &\approx \frac{p_{i+1}^m - 2p_i^m + p_{i-1}^m}{h^2} \\ p_x(t_m, x_i) &\approx \frac{p_{i+1}^m - p_{i-1}^m}{2h} \end{aligned}$$

where $k = \delta t$ and $h = \delta x$. The term inside the integration sign is discretized using a simple trapezoidal rule:

$$\int_{\mathbb{R}} p(t_m, x) f(z - x_i)\, dz \approx \frac{h}{2}\left[p_1^m f_{i,1} + 2\sum_{j=2}^{n-1} p_j^m f_{i,j} + p_n^m f_{i,n}\right] \tag{4.60}$$

Since the boundary conditions are zero at the both ends, the remainder term $R(t_m, x_i, \hat{x})$ will also be zero at the boundary. For details see [1, 44].

The main equation (4.29) is solved forward for time variable τ. But we can use the same code with minor adjustments to solve the adjoint equation by working with

time variable t instead of τ. In this case the adjoint equation will also be a forward problem on the variable t. The solution procedure is similar to the one described in section 3.7. Hence, solving the adjoint equation is similar in complexity to solving the main pricing equation.

In this section, we discussed the derivation of adjoint equations for some of the option pricing models. This technique allows efficient evaluation of the derivative of a function $F(\mathbf{p})$ with respect to parameters $\mathbf{p}$ in situations where F depends on $\mathbf{p}$ indirectly, via an intermediate variable $V(\mathbf{p})$. The derivative of $V(\mathbf{p})$ could be computationally expensive to evaluate. Adjoint method has been used in the case of parameter estimation for Black-Scholes model and our theory has been supported by numerical experiments. Application of adjoint method for more complex model like the PIDE model is yet to be verified by numerical experiments.
The use of the adjoint model has two advantages over other methods for gradient computation:

1. In the case of large number of parameters n, the adjoint model saves a lot of computation time.

2. The gradient obtained through adjoint model is exact.

Adjoint method can be used to calculate the sensitivity of the cost function with respect to the parameters. The knowledge of the sensitivities provides a better understanding about the model parameters. The benefit of using optimization theory has been utilized more with the discovery of faster methods to obtain the gradient. Gradient-based methods depend on the smoothness of the variation of the cost function with the model parameters. This requires the existence of continuous first derivatives of the cost function; namely the gradient, and possibly higher derivatives. Simple gradient-based methods only require the gradient of the objective function

with respect to the model parameters. Gradient-based methods generally require a much smaller number of iterations to converge to an optimum than non-gradient based methods. However, only convergence to a local minimum is guaranteed.

CHAPTER 5

OPTIMIZATION

Optimization is a method of minimizing or maximizing the objective function of several variables over a feasible region. Normally, the objective function is set up as a least-square minimization problem and optimization is done by numerical optimization algorithms. For the solution of least-squares problems it is efficient to use the gradient-based algorithms. There are many variations of gradient-based optimization methods, but their common feature is the utilization of first and second derivatives of the objective function. Gradient method use information about the slope of the function to decide the direction in which minimum is supposed to lie. Optimization methods that use gradient information often converge much faster than those that do not. Gradient information of the objective function can be obtained efficiently using the adjoint methods described in chapter 4.

5.1 OPTIMALITY CONDITIONS FOR UNCONSTRAINED MINIMIZATION

Definition 5.1. Suppose $f : \mathbf{R}^n \to \mathbf{R}^m$. The **derivative** of f at x is the matrix $Df(x) \in \mathbf{R}^{m\times n}$, given by

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j};\ i = 1, ..., m,\ j = 1, ...n, \tag{5.1}$$

provided the partial derivatives exist.

If the partial derivatives exist, we say f is differentiable at x. When f is a real-valued function (i.e., $f : \mathbf{R}^n \to \mathbf{R}$), the derivative $Df(x)$ is a $1 \times n$ matrix, i.e., it is a row vector. Its transpose is called the **gradient** of the function f given by:

$$\bigtriangledown f(x) = Df(x)^T, \tag{5.2}$$

which is a column vector, in $\mathbf{R}^n$. Its components are the partial derivatives of f:

$$\bigtriangledown f(x)_i = \frac{\partial f(x)}{\partial x_i},\ i = 1, ...n. \tag{5.3}$$

Definition 5.2. Consider a real valued function $f : \mathbf{R}^n \to \mathbf{R}$. **The second derivative** of f at x, denoted by $\bigtriangledown^2 f(x)$, is given by

$$\bigtriangledown^2 f(x)_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j},\ i = 1, ..., n,\ j = 1, ...n, \tag{5.4}$$

provided f is twice differentiable at x where the partial derivatives are evaluated at x.

Consider an unconstrained minimization problem

$$min_{x \in \mathbf{R}^n} f(x)$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is a smooth function.

Definition 5.3. Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is given. Then $x^* \in \mathbf{R}^n$ is a local minimizer of f if there exists $\delta > 0$ such that

$$\parallel x - x^* \parallel < \delta \Rightarrow f(x^*) \leq f(x)$$

The point x^* is a strict local minimizer if

$$0 < \parallel x - x^* \parallel < \delta \Rightarrow f(x^*) < f(x)$$

Thus, x^* is a local minimizer of f if locally, f does not attain a smaller value.

Definition 5.4. Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is given. Then $x^* \in \mathbf{R}^n$ is a global minimizer of f if

$$f(x^*) \leq f(x)\ for\ all\ x \in \mathbf{R}^n$$

Therefore, a global minimizer yields the absolute minimum of f. The definition of local minimizer is of little use in recognizing a solution. Algorithms for nonlinear problems are usually iterative that require an initial estimate of the solution $x^{(0)}$. The definition of local minimizer requires that $f(x^{(0)})$ be compared with $f(x)$ for every x in the neighborhood of $x^{(0)}$. So, even if the starting point is the solution the algorithm still tries to improve upon $x^{(0)}$. Optimality conditions are essential in numerical optimization. A necessary condition must be satisfied by any solution but necessary condition can be satisfied by non solution as well. But if a point satisfies a sufficient condition, then it is guaranteed to be a solution.

5.2 FIRST ORDER NECESSARY CONDITION

Theorem 5.2.1. *Suppose $f : \mathbf{R}^n \to \mathbf{R}$ has a local minimum at $x = x^*$. If f is differentiable at x^*, then*

$$\nabla f(x^*) = 0.$$

A point x^* satisfying $\nabla f(x^*) = 0$ is called a *stationary point* of f. If $\nabla f(x^*) \neq 0$ then the vector $-\nabla f(x^*)$ defines the direction in which f decreases. Let $p = -\nabla f(x^*)$. Then $f(x^* + hp) < f(x^*)$ for all h sufficiently small. The first order condition is only a necessary condition. If x^* is a local minimizer of f then x^* is a stationary point. But, x^* is a stationary point does not guarantee that it is also a local minimizer. For general nonlinear problems, there is no first order sufficient condition. To obtain the sufficient condition, we have to examine the second derivatives as well

5.3 SECOND-ORDER NECESSARY CONDITION

Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is twice differentiable at x^* and x^* is a local minimizer of f, then

$$f(x) = f(x^*) + \nabla f(x^*)^T(x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*) + o(\| x - x^* \|^2).$$

Let $x = x^* + hy$, for some scalar h and some vector y. Applying the first-order necessary condition, we get:

$$f(x^* + hy) = f(x^*) + \frac{h^2}{2} y^T \nabla^2 f(x^*) y + o(h^2)$$

Since the error term $o(h^2)$ is small compared to $\frac{h^2}{2} y^T \nabla^2 f(x^*) y$ for small h, $f(x^* + hy) \geq f(x^*)$ for all h sufficiently small. This implies that

$$\begin{aligned} &\frac{h^2}{2} y^T \nabla^2 f(x^*) y \geq 0 \\ &\text{or } y^T \nabla^2 f(x^*) y \geq 0 \end{aligned} \tag{5.5}$$

because $\nabla^2 f(x^*)$ is a symmetric matrix. This must hold for all $y \in \mathbf{R}^n$ if x^* is a local minimizer of f. This is called the second order necessary condition for a local minimizer.

Definition 5.5. Suppose $A \in \mathbf{R}^{n \times n}$ is symmetric. If $y^T A y \geq 0 \ \forall y \in \mathbf{R}^n$, then A is said to be positive semidefinite. If $y^T A y > 0 \ \forall y \in \mathbf{R}^n, y \neq 0$, then A is said to be positive definite.

Theorem 5.3.1. *Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is twice differentiable at x^* and x^* is a local minimizer of f. Then $\nabla^2 f(x^*)$ is positive semidefinite.*

5.4 THE SECOND-ORDER SUFFICIENT CONDITION

Theorem 5.4.1. *Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is twice differentiable at x^*. If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite then x^* is a strict local minimizer of f.*

Proof: The function $y^T \nabla^2 f(x^*) y$ is continuous and positive over the compact set $\{y \in \mathbf{R}^n : \| y \| = 1\}$. Therefore, there exists $\alpha > 0$ such that

$$\| y \| = 1 \Rightarrow y^T \nabla^2 f(x^*) y \geq \alpha \tag{5.6}$$

Since f is twice differentiable at x^* and $\nabla f(x^*)$=0 by assumption,

$$f(x^* + hy) = f(x^*) + \frac{h^2}{2} y^T \nabla^2 f(x^*) y + o(h^2) \tag{5.7}$$

By definition,$|o(h^2)| < \frac{h^2}{2}$ for all h sufficiently small. Hence there exists $h_0 > 0$ such that $0 < h < h_0$ and $\| y \| = 1$, implies

$$\frac{h^2}{2} y^T \nabla^2 f(x^*) y + o(h^2) > 0$$

Together with equation (5.7), this shows that $f(x^* + hy) > f(x^*)$ for all h, y with $0 < h < h_0, \| y \| = 1$. Since

$$x^* + hy : 0 < h < h_0, \| y \| = 1 = x : 0 < \| x - x^* \| < h_0$$

it follows that $0 < \| x - x^* \| < h_0 \Rightarrow f(x^*) < f(x)$ and hence that x^* is a strict local minimizer of f. This ends the proof.

Note that x^* can be a local minimizer of f and yet $\nabla^2 f(x^*)$ may not be positive definite. So, The sufficient condition is not necessary for x^* to be a minimizer.

5.5 PARAMETER IDENTIFICATION

Some parameters in financial models have to be determined from the available market price. Parameter estimation is done through least-square minimization. The technique of finding the minimum of an objective function is iterative. Iteration starts from an initial value x_0, and then in each iteration search direction p_n and step length α_n are computed to give $x_{k+1} = x_k + \alpha_k p_k$. A natural choice for the search direction is the negative gradient $-\nabla f(x)$. The step length must be chosen so that x_{k+1} reduces sufficiently. The iteration stops when x_k is sufficiently close to minimum. The details of numerical optimization is found in [39, 28, 45] and many others.

Once the parameters are estimated through optimization, it is important to know the stability of the estimated parameters. Sensitivity analysis is done on the estimated parameters to check the stability of the parameters.

Gradient descent method The resulting algorithm is called the gradient algorithm or gradient descent method.

5.5.1 One variable unconstrained optimization

Consider the Black-Scholes model

$$V_t + \frac{\sigma^2 S^2}{2} V_{ss} + rSV_s - rV = 0. \tag{5.8}$$

In this model, we assume that the actual market price $V(t, S)$ and all other variables are known except the market volatility. We consider an inverse problem of finding the market volatility σ using the procedure of numerical optimization. Suppose $V(t, S)$ is the current option price of a stock trading at S. Let $V_c(t, S, T, K, r, \sigma)$ be the function of the unknown variable σ only. We want to find σ such that

$$|V(t, S) - V_c(t, S, T, K, r, \sigma)|^2$$

is minimized. Our goal is to compute the zero of the function

$$f(\sigma) = |V(t, S) - V_c(t, S, T, K, r, \sigma)|^2. \tag{5.9}$$

The minimization problem is set as follows:

$$min_\sigma f(\sigma) = \sum [V(t, S) - V_c(t, S, T, K, r, \sigma)]^2 \tag{5.10}$$

For this single variable unconstrained optimization, the only parameter to be estimated is σ. For the Black-Scholes model the estimate is called the implied volatility of the market.

Here is an example from the market. Various put option prices for different expiration dates is given for a particular product we name it XXXX. This product is currently trading at $38.00 per share.

Table 5.1. Put price for some product XXXX quoted on a certain day

Strike price	V(38,0.4)	V(38,0.6)	V(38,1.2)	V(38,2.2)
25	0.05	0.05	0.2	0.5
30	0.15	0.25	0.65	1.1
35	0.7	0.9	1.6	2.15
40	2.9	2.65	3.3	4.3
45	7.9	6.7	6.9	7

If we use the Black-Scholes pricing model, the only unknown parameter is the volatility of the market. Therefore, the objective function to be minimized for this problem is given by equation (5.10). From the information given in the table above, we want to find the optimal σ that minimizes the objective function.

Using the standard optimization subroutine in matlab and implicit finite difference method for the Black- Scholes equation, the optimal volatility σ was estimated to be 0.2027 which gave the residual norm of 2.0234.

5.5.2 Multi-variable constrained optimization in PIDE model

Consider the PIDE model

$$V_t + \frac{\sigma^2 S^2}{2} V_{ss} + (rS - S\kappa\lambda)V_s - (r+\lambda)V + \lambda E[V(t, JS)] = 0 \tag{5.11}$$

where $\kappa = E[J-1]$.

In equation (5.11), we replace the expectation by a linear combination of the delta functions as follows

$$E[V(t, JS)] = \sum_{k=1}^{N} \alpha_k V(t, J_k S)$$

Hence our simplified model is given by

$$V_t + \frac{\sigma^2 S^2}{2} V_{ss} + (rS - S\kappa\lambda)V_s - (r+\lambda)V + \lambda \sum_{k=1}^{N} \alpha_k V(t, J_k S) = 0. \qquad (5.12)$$

In this model, apart from the market volatility σ, jump sizes J_k, weights α_k and the parameter λ also have to be estimated from the market data. In addition we also have certain conditions on the weights, i.e. $\sum_{k=1}^{N} \alpha_k = 1$ and J_k are all positive. Hence, our optimization problem is

$$min_{\sigma,\lambda,\alpha_k,J_k} f(\sigma, \alpha_k, J_k) = \sum [V(t,S) - V_c(t,S,T,K,r,\sigma,\alpha_k,J_k)]^2 \qquad (5.13)$$

subject to the constraints,$\sum_{k=1}^{N} \alpha_k = 1$ and $J_k, \sigma, \lambda > 0$. The problem becomes a multi-variable optimization with constraints. Here, equality constraints are written as:

$$\alpha_1 + \alpha_2 + ... + \alpha_N = 1$$

and inequality constraints as:

$$\lambda, \sigma, \ J_1, \ J_2, ..., \ J_N > 0$$

For this problem, a reasonably few jumps J_k and weights α_k are chosen to save the computational time. Initially three of each J_k and α_k are chosen in addition to unknown σ and λ. So, we have total of eight parameters to be estimated in equation (5.13). Also, certain restrictions were imposed on the the parameters. For example, it is reasonable to assume that $0.5 \leq J_k \leq 1.5$, $0 < \sigma < 1$ and also $0 \leq \alpha_k \leq 1$. With these additional bounds on the parameters, the estimated parameters were as follows: $\sigma = 0.2069$, $J = [0.8653, 1, 0.6]$, $\alpha = [0.7403, 0, 0.2597]$ and $\lambda = 0.2662$. With these estimated values the residual norm was 1.8195. This is better than the one given by Black-Scholes model. The estimated parameters were not always unique but depended on the starting values. If the constraints were relaxed to some extent, the residual was smaller than the one obtained from Black-Scholes model.

5.5.3 Sensitivity analysis

The reliability of the parameters found by optimization can often be examined by sensitivity analysis. The derivative of the solution with respect to a parameter of interest is a measure of the sensitivity of the solution with respect to small change. Continuous sensitivity equations method examines the change in solutions to the problem with respect to small changes in the model parameters. To find this sensitivity, the equations are formally differentiated with respect to the parameter leading to a linear equation for the sensitivity (see chapter 4 for tetails). For our purpose we assume that the sensitivities exist and satisfy the sensitivity equation. Examining the sensitivity of solutions to small change is important because small changes in the parameter can play a large role in the solution process. Sensitivity also provides valuable information showing where the solution is sensitive and which ones are more sensitive then the others. Since the sensitivity equation is very similar to the original problem, one can couple the sensitivity equation to the original system and simply solve them simultaneously.

Sensitivity analysis is a vast area of research and is beyond the scope of this thesis. However, we will present a simple example to show its importance.

5.5.4 Sensitivity with respect to the weights, α_i

Lets consider the PIDE model given by equation (5.11). Let the derivative of V with respect to α_i be $\dfrac{\partial V}{\partial \alpha_i} = W$. Then,

$$\frac{\partial}{\partial \alpha_i}\left(V_t + \frac{\sigma^2 S^2}{2}V_{ss} + (rS - S\kappa\lambda)V_s - (r+\lambda)V + \lambda\sum_{k=1}^{N}\alpha_k V(t, J_k S)\right) = 0$$

or

$$W_t + \frac{\sigma^2 S^2}{2}W_{ss} + (rS - S\kappa\lambda)W_s - (r+\lambda)W + \lambda\sum_{k=1}^{N}\alpha_k W(t, J_k S) = \lambda S J_i V_S - \lambda V(t, J_i S) \qquad (5.14)$$

Now, the final and boundary conditions for W is given by:
$W(t,0) = 0,\ W(t,S_{max}) = 0,\ W(T,S) = 0$. This equation is similar to the original pricing equation with non homogeneous right hand side. This can be solved the same way the original pricing equation is solved.

5.5.5 Sensitivity with respect to the Jump sizes, J_i

Let the derivative of V with respect to J_i in our PIDE model (5.11) be $\frac{\partial V}{\partial J_i} = Z$. Then we get the following PDE which looks similar to equation (5.14):

$$Z_t + \frac{\sigma^2 S^2}{2} Z_{ss} + (rS - S\kappa\lambda)Z_s - (r+\lambda)Z + \lambda\alpha_i S Z(t, J_i S) - S\lambda\alpha_i V_s = 0 \tag{5.15}$$

with the final and boundary conditions:
$Z(t,0) = 0,\ Z(t,S_{max}) = 0,\ Z(T,S) = 0$. This can be solved the same way as the original equation with very little modification.

Figure 5.1 shows how sensitivity evolves across both time and space horizons for a model problem with strike price at 10. Since this problem is solved backwards in time sensitivity increases as time decreases. It is important to know which parameters are the most influential and which are least influential in the given time horizon. This can be done through sensitivity analysis of the parameters in a model.

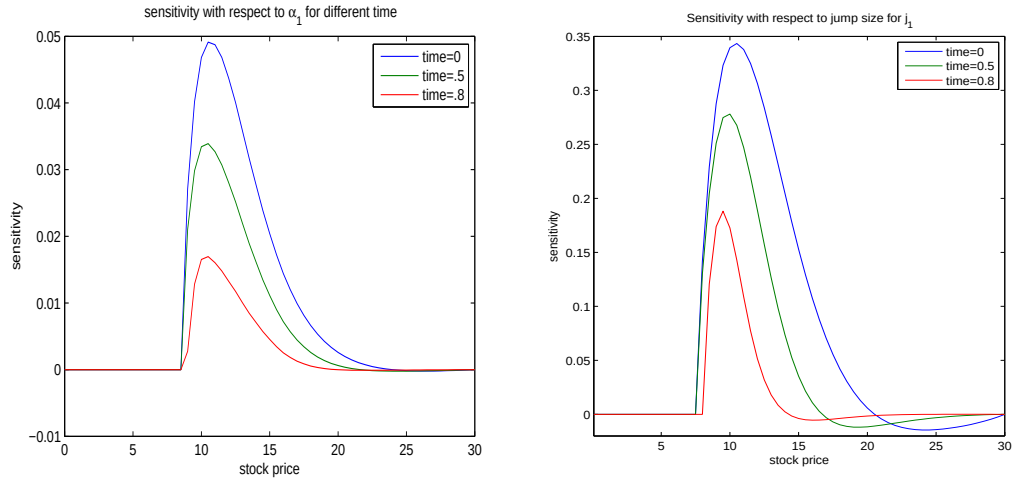


Figure 5.1. Sensitivity with respect to jump size J_i and weight α_i for a sample problem when $J_i = .8$ and $\alpha_i = .74$ where strike price is 10.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTION

Parameter estimation is a very important issue in option price modelling. As the financial contracts become more complex, the need to price the derivatives and estimate the model parameters accurately becomes extremely important. Even if the model is correctly specified, mispricing can still occur if model parameters are not correct. In financial modelling, calibrating a model means finding numerical values of its parameters such that the model is consistent with the market. Parameter estimation is done through optimization.

For problems with a large number of parameters, the most efficient algorithms are gradient based algorithm. Gradient based optimization techniques utilize the gradient information to identify search directions and require far fewer steps to obtain an accurate approximation of a locally optimal solution then the non-gradient based optimization. The main concern with the gradient method is to effectively compute the gradient of the objective function that has to be minimized.

In this thesis I have developed an improved computational methodology for estimating parameters in the popular financial models like the Heston stochastic volatilty model and a special case of the Merton jump diffusion model. I have done so by adapting methodology known as the **adjoint** approach from control theory, which allows for efficient computation of gradients in least-squares estimation algorithms. An alternative to an adjoint approach to estimate the gradient would be the sensitivity equation approach. This approach first computes the sensitivity of the solution V of the PDE/PIDE with respect to each of the parameters P_i by solving sensitivity equations. When the number of parameters n in the model is really large, n number of sensitivity equations corresponding to each parameter has to be solved. This could be computationally expensive process.

The difference between the two approaches is that the solution of the adjoint equation is specific to the minimizing function i.e. one adjoint equation is computed per cost function and used in the computation of all components of the gradient. On the other hand, the sensitivity equations are specific to the parameter i.e. one sensitivity $\partial V/\partial P_i$ is computed for each component of the parameter vector P and can be used to calculate the derivatives of the required objective function. The costs for solving the adjoint or one of the sensitivity equations are comparable, and are the most significant costs for either method. Thus, it is advantageous to use the sensitivity equation approach if the number of parameters m is fairly small while it is more efficient to use the adjoint approach if a large number of parameters is present. Hence gradients can be efficiently evaluated using the adjoint approach as this approach is independent of number of parameters. Once the gradient is computed, any of the gradient based optimization can be used.

I have not been able to verify the theoretical results with the numerical results for the financial models in this thesis. However, the adjoint approach has been used extensively in optimal control problems for parameter estimation and has been very successful. A similar approach taken by Barth [7] using adjoints for discretized PIDE models have shown promising results. Numerical verification of the adjoint approach for the financial models described in this thesis will be a continuing part of my future work. Future work could go in a number of directions, including but not limited to the extension of jump-diffusion models to wider class of Lévy models. Derivation of adjoints for such models and estimating Lévy measure through the adjoint process will be challenging. As the pricing models become more complex, efficient parameter estimation and sensitivity analysis will be essential. The adjoint approach will gain popularity and lead the way for parameter estimation and sensitivity analysis in complex financial models in the future.

APPENDIX

INTEGRAL ADJUSTMENT

Consider the integral:

$$\int_{\underline{x}}^{\overline{x}} w(t, x+y)p(t,x)\ dx$$

where x_{min} and x_{max} is written as $\underline{x}$ and $\overline{x}$ for notational simplicity.

Let $w(t,x) = 0$ for $x \in [-\infty, \underline{x}]$ and $x \in [\overline{x}, \infty]$. Also, $p(t,x) = 0$ for $x \in [-\infty, \underline{x}]$ and $x \in [\overline{x}, \infty]$.

Suppose $z = x + y$. Then $x = z - y$ and $dx = dz$. Therefore

$$\int_{\underline{x}}^{\overline{x}} w(t, x+y)p(t,x)\ dx = \int_{\underline{x}+y}^{\overline{x}+y} w(t,z)p(t,z-y)\ dz \tag{1}$$

Then, $w(t,z)p(t,z-y) = 0$ if $w(t,z) = 0$ or $p(t,z-y) = 0$

$w(t,z) = 0$ when $z \geq \overline{x}$ or $z \leq \underline{x}$.

$p(t,z-y) = 0$ when $z - y \geq \overline{x} \Leftrightarrow z \geq \overline{x} + y$

or $z - y \leq \underline{x} \Leftrightarrow z \leq \underline{x} + y$.

For the upper limit

- If $y \geq 0$ then $\overline{x} \leq z \leq \overline{x} + y \Rightarrow w(z,t) = 0$.
- If $y \leq 0$, then $\overline{x} + y \leq z \leq \overline{x} \Rightarrow p(z-y,t) = 0$.

For lower limit

- If $y \geq 0$, then for $\underline{x} \leq z \leq \underline{x} + y \Rightarrow p(z-y,t) = 0$.

- If $y \leq 0$, then for $\underline{x} + y \leq z \leq \underline{x} \Rightarrow w(z,t) = 0$.

Therefore, $w(z,t)p(z-y,t) = 0$ in all cases mentioned above. Hence,

$$\int_{\underline{x}+y}^{\overline{x}+y} w(z,t)p(z-y,t)\ dz = \int_{\underline{x}}^{\overline{x}} w(z,t)p(z-y,t)\ dz. \tag{2}$$

REFERENCES

[1] A. Almendral and C.W. Oosterlee. Numerical valuation of options with jumps in the underlying. *Applied Numerical Mathematics*, 53:1–18, 2005.

[2] K. I. Amin. Jump diffusion option valuation in discrete time. *Journal of Finance*, 48:1833–1863, 1993.

[3] Leif Andersen and Jesper Andreasen. Jump diffusion process: Volatility smile fitting and numerical methods for pricing. *Review of Derivative Research*, 4:231–262, 2000.

[4] T. Apel, G. Winkler, and U. Wystup. Valuation of options in hestons stochastic volatility model using finite element methods. *Foreign Exchange Risk, Risk Publications*, 2001.

[5] M. Avellaneda, C. Friedman, R. Holmes, and L. Sampieri. Calibrating volatility surfaces via relative entropy minimization. *Applied Mathematical Finance*, 4:37–64, 1997.

[6] Gurdip Bakshi, Charles Cao, and Zhiwu Chen. Empirical performance of alternative option pricing models. *Journal of Finance*, pages 2003–2049, January 1997.

[7] Michaela Barth. Calibration with levy models. Diploma thesis, University of Trier, Trier, Germany, July 2007.

[8] David S Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, 9(1):69–107, 1996.

[9] M. Baxter and A. Rennie. *Financial Calculus: An Introduction to Derivative Pricing*. Cambridge University Press, 1996.

[10] T. Bjork. *Arbitrage theory in continuous time*. Oxford University Press, 1998.

[11] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–659, 1973.

[12] J.T. Borggaard and J.A Burns. A pde sensitivity equation method for optimal aerodynamic design. *The Journal of Computational Physics*, 136:366–384, 1997.

[13] M. Briani, C. La Chioma, and R Natalini. Convergence of numerical schemes for viscosity solutions to integro-differential degenerate parabolic problems arising in financial theory. *Numer. Math.*, 98:607–646, 2004.

[14] M. Briani, R. Natalini, and G. Russo. Implicit-explicit numerical schemes for jumpdiffusion processes. IAC Report 38, April 2004.

[15] Thomas F. Coleman, Yuying Li, and Arun Verma. Reconstructing the unknown volatility function. Technical Report TR98-1706, 1998.

[16] Rama Cont and Peter Tankov. Calibration of jump-diffusion option-pricing models: a robust non-parametric approach. *Journal of Computational Finance*, 7(3):1–49, Sept 2002.

[17] Rama Cont and Peter Tankov. *Financial Modelling with jump processes.* 2004.

[18] Rama Cont and Peter Tankov. Retrieving lévy processes from option prices: regularization of an ill-posed inverse problem. *Siam Journal of Control and Optimization*, 45(1):1–25, 2006.

[19] Rama Cont and Ekaterina Voltchkova. A finite difference scheme for option pricing in jump diffusion and exponential levy models. *SIAM Journal on Numerical Analysis*, 43(4):1596–1626, 2005.

[20] J. C. Cox and Ross S.A. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3, 1976.

[21] Daniel. J. Duffy. *Finite Difference Methods in Financial Engineering.*

[22] Jim Gatheral. Stochastic volatility and local volatility: Case studies in financial modelling course notes, 2002.

[23] Gene H. Golub and Charles F. Van Loan. *Matrix Computations third edition.* The Johns Hopkins University Press, Baltimore, 1996.

[24] S. Heston. A closed-form solution for options with stochastic volatility with ap-

plications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.

[25] J. Hull and A White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300, 1987.

[26] J. C. Hull. *Options, Futures, and Other Derivatives*. Prentice Hall, 3rd edition edition, 1997.

[27] S. Ikonen and J Toivanen. Numerical valuation of european and american options with kou's jump-diffusion model. Inria-Rocquencourt, France, 2006. Amamef Conference on Numerical Methods in Finance.

[28] C. T Kelley. *Iterative methods for optimization*. Society for Industrial and Applied Mathematics, 1999.

[29] S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101, August 2002.

[30] Claudia La Chioma. *IntegroDifferential Problems Arising in Pricing Derivatives in JumpDiffusion Markets*. PhD thesis, Universita degli Studi di Roma, December 2003.

[31] Alan L. Lewis. *Option Valuation under Stochastic Volatility*. Finance Press, 2000.

[32] R.M. Lewis. Numerical computation of sensitivities and the adjoint approach. *ICASE Report No.97-61*, 1997.

[33] Shengtai Li and Linda Petzold. Adjoint sensitivity analysis for time-dependent partial differential equations with adeptive mess refinement. *Journal of computational physics*, 198:310–325, 2004.

[34] D. B. Madan and E. Seneta. The variance-gamma (v. g.) model for share market returns. *J. Business*, 63:511–524.

[35] G. I. Marchuk. *Adjoint Equations and Analysis of Complex Systems*, volume 295 of *Mathematics and its applications*. Kluwer Academic Publishers.

[36] A.M. Matache, Petersdorff T. Von, and C. Schwab. Fast deterministic pricing of options on lêvy driven assets. *Mathematical Modelling and Numerical Analysis*, 38(1):37–72, 2004.

[37] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3, 1976.

[38] J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 26, 1965.

[39] J. Nocedal and S.J Wright. *Numerical Optimization*. Springer, 1999.

[40] Steven E. Shreve. *Stochastic Calculus Models for Finance: Continuous Time Models*. Springer Verlag, 2006.

[41] John R. Singler. *Sensitivity Analysis of Partial Differential Equations With Applications to Fluid Flow*. PhD thesis, Virginia Tech, June, 2005.

[42] Dieter Sondermann. *Introduction to stochastic calculus for finance : a new didactic approach*. Springer, 2006.

[43] E. Stein and J Stein. Stock-price distributions with stochastic volatility - an analytic approach. *Review of Financial Studies*, 4:727–752, 1991.

[44] Arne K. Strauss. Numerical analysis of jump diffusion models for option pricing. Master's thesis, Virginia Tech, Blacksburg, Virginia, July 2006.

[45] Wenyu Sun and Ya-Xiang Yuan. *Optimization theory and methods: Nonlinear Programming*. Springer, 2006.

[46] Kay E Vugrin. *On the Effects of Noise on Parameter Identification Optimization Problems*. PhD thesis, Virginia Tech, 2005.

[47] J. Wiggins. Option values under stochastic volatility - theory and empirical estimates. *Journal of Financial Economics*, 19:351–372, 1987.

[48] Paul Wilmott. *Paul Wilmott on Quantitative Finance*, volume 1, 2 and 3. Wiley, 2006.

[49] Paul Wilmott, Sam Howison, and Jeff Dewynne. *The Mathematics of Financial*

Derivatives. 1995.

[50] X. L Zhang. Numerical analysis of american option pricing in a jump-diffusion model. *Mathematics of Operations Research*, 22:668–690, 1997.

VITA

Tirtha Prasad Timsina was born in Tehrathum, Nepal on April 28, 1973. He finished his High School in 1993 from Budhanilkantha School, Kathmandu. He completed his Bachelors degree in Mathematics/Computer Science from Concord College, WV in 2000. In 2002 Tirtha completed his Masters degree in Mathematics from Virginia Polytechnic Institute and State University and started working on his Ph.D under the guidance of Professor Ekkehard Sachs. He accepted a teaching position in Mathematics at Ferrum College, Ferrum, Virginia starting in the fall of 2007.