623 Computational Finance, Fall 2014.

Lecturer: Sergey Nadtochiy.

Chapter II. PDE methods.

Partial information can be found in Chapters 3–9 in "The Mathematics of Financial Derivatives" by Wilmott, Howison and Dewynne 1998; and in Chapters 1.1-1.3, 3.2.1, 3.2.3, 10.4 in "Interest Rate Models – Theory and Practice" by Brigo and Mercurio 2006.

1 Black-Scholes model

- there is only one risky asset: $d = 1$,
- the riskless asset yields a constant (continuously compounded) interest rate r: $B_t = \exp(rt)$.
- and the risky asset S follows a Geometric Brownian Motion (GBM):

$$S_t = S_0 \exp\left((\mu - \sigma^2/2)t + \sigma W_t\right),$$

  where μ is the drift, σ is the volatility, and W is a Brownian motion (BM).

Rem 1. *We will, typically, assume that the dividend rate is zero:* $q = 0$.

Samuelson (1965) was the first one to propose using GBM for modeling the dynamics of the price of a unit of asset (a stock share).

Motivation for using GBM to model asset prices

- Discrete time and discrete space models are convenient for computations, but have several drawbacks.
  - The resulting expressions and formulae are often too convoluted, making it hard to develop intuition about the roles of various parameters in the model, etc.
  - The output of a discrete model depends on the discretization of the time and space variables: Δt and, for example, ΔS. In practice, we often need to change the discretization to obtain a more realistic model. Then the important question is: does the discrete model converge as the discretization becomes more and more precise? This is related to **stability** of the numerical method.
- If we want the risky asset S to be **continuous**
- and its log-returns

$$R_t = \log\left(S_{t+\Delta t}/S_t\right)$$

  to be **stationary** and **independent**,

- then, $\log(S_t)$ has to be given by a Brownian Motion with drift:

$$\log(S_t) = \log(S_0) + at + \sigma W_t$$

Taking $a = \mu - \sigma^2/2$, we obtain

$$S_t = S_0 \exp\left((\mu - \sigma^2/2)t + \sigma W_t\right)$$

- There is no particular reason to give preference to any specific binomial model (i.e. choose the values of p, u and d). However, the above arguments show that there are several reasons to choose GBM to model asset prices.
- In addition, GBM can be viewed as a (unifying) limit of (various) binomial models. Recall that random walk, after appropriate normalization, converges to BM. Similarly, we will show that Black-Scholes model can be viewed as a limit of discrete time tree models.
- This is an example of how the (seemingly more abstract) continuous time models may actually be more natural, even from the practical point of view, than the discrete-time models (which may have too many "free" parameters).

Ito's lemma

Consider an arbitrary Ito process

$$X_t = X_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s dW_s,$$

or, equivalently,

$$dX_t = \alpha_t dt + \beta_t dW_t$$

Then, for any smooth enough function $(x,t) \mapsto f(x,t)$,, we have:

$$df(X_t,t) = \left(\partial_t f(X_t,t) + \alpha_t \partial_x f(X_t,t) + \frac{1}{2}\beta_t^2 \partial_{xx}^2 f(X_t,t)\right) dt + \beta_t \partial_x f(X_t,t) dW_t$$

Using Ito's lemma, it is easy to see that, in the Black-Scholes model

$$\begin{cases} dB_t = rB_t dt, \\ dS_t = \mu S_t dt + \sigma S_t dW_t \end{cases}$$

Exercise 1. *Use Ito's lemma to show that the above equations hold.*

Notice that we haven't checked that the model is arbitrage-free. In fact, it can be deduced, using the general theory (FTAP I) that the above model is always arbitrage free provided $\sigma > 0$.

1.1 Pricing and hedging in Black-Scholes model

Consider a European option with payoff $F(S_T)$ at the time of expiry T.

Q 1. *What is the arbitrage-free price of this claim?*

First, we will derive the pricing formula via the **hedging** argument.

- Due to the **Markov nature of the model**, and because the **payoff of the option is not path-dependent**, it is natural to assume that the price of the option is given by $V(S_t, t)$ with some deterministic function V.
- Assume that we sold the option at time t. We search for a hedging portfolio π_t, such that the overall change of the hedging portfolio and the option price is zero

$$0 = \pi_t dS_t + (V(S_t,t) - \pi_t S_t) dB_t/B_t - dV(S_t,t),$$

$$0 = \pi_t \mu S_t dt + \pi_t \sigma S_t dW_t + (V(S_t,t) - \pi_t S_t) r dt$$

$$- \left(\partial_t V(S_t,t) + \mu S_t \partial_S V(S_t,t) + \frac{1}{2}\sigma^2 S_t^2 \partial^2_{SS} V(S_t,t) \right) dt - \sigma S_t \partial_S V(S_t,t) dW_t,$$

  where we used Ito's formula.

- Equating the terms with dt and dW_t, we obtain:

$$\pi_t = \frac{\partial}{\partial_S} V(S_t,t), \tag{1}$$

$$\frac{\partial}{\partial t} V(S_t,t) + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2}{\partial S^2} V(S_t,t) + r S_t \frac{\partial}{\partial S} V(S_t,t) - r V(S_t,t) = 0$$

- Since S_t and t can take arbitrary values in $(0,\infty)$ and $[0,T]$ respectively, we conclude that the above equation should hold for all values of the space and time variables. Thus, we obtain the **Black-Scholes PDE (BSPDE)**:

$$\frac{\partial}{\partial t} V + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2}{\partial S^2} V + r S \frac{\partial}{\partial S} V - r V = 0 \tag{2}$$

  for $0 < S < \infty$ and $0 < t < T$.

- This is a **linear backward parabolic equation** and it is solved with the **terminal condition**

$$V(S,T) = F(S)$$

  where F is the payoff function of the option (e.g. $F(S) = (S-K)^+$).

- Assume that we can find function V solving the above PDE with the corresponding terminal condition. Then, working backwards, it is easy to show, using Ito's formula, that the hedging strategy given by π_t, in (1), has zero initial value and its terminal value is equal to $(V_T - V_0 = F(S_T) - V(S_0, 0)$. Then, due to the NA principle, $V(S_0, 0)$ has to be the initial price of the option.

Rem 2.
- *The BS model is rather flexible*

  - *the asset can move arbitrarily far in any small interval (albeit with a small probability),*
  - *it is easy to extend the setup to allow for $\mu = \mu(S_t, t)$, $\sigma = \sigma(S_t, t)$ and $r = r(t)$. The pricing algorithm does not changes – neither does the BSPDE.*

- *Binomial model can be viewed as an approximation of the BS model:*
  - *We will see that the underlying price process in binomial model converges to the GBM, as $\Delta t \to 0$.*
  - *In fact, the corresponding pricing formulas also converge. Two pricing steps in a binomial model (or a single step in a trinomial model), with appropriately chosen parameters, coincide with a single step in the* **explicit finite difference** *approximation of the solution to the BSPDE.*

Notice that **there is no μ in the pricing PDE**! This is due to the fact that we are pricing under the risk neutral measure $\mathbb{Q}$, under which the dynamics of the underlying may be different. In particular the dynamics of S (more precisely, its drift) under $\mathbb{Q}$ are determined by the condition that S/B is a martingale. Here is a useful characterization of the martingale property.

Thm 1. *Consider an Ito process*

$$X_t = X_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s dW_s$$

If X is a martingale, then, $\alpha \equiv 0$.

Rem 3. *Strictly speaking, the converse is not true: $\alpha \equiv 0$ does not imply that X is a martingale. However, this implication does hold under some additional assumptions. In this course, we will always assume that these assumptions hold, and will deduce from $\alpha \equiv 0$ that X is a mtg.*

Let us find the dynamics of S under $\mathbb{Q}$.

Main idea: all that changes when we pass from the original measure $\mathbb{P}$ to the pricing measure $\mathbb{Q}$ is that the BM W may become a BM with drift under $\mathbb{Q}$.

In fact, the converse is also true: for any given drift, we can find the associated measure $\mathbb{Q}$, under which W is a BM with the prescribed drift. This is a well-known fact from Stochastic Calculus:

Thm 2. *Assume that W is a BM under the original measure $\mathbb{P}$. For any (regular enough) stochastic process (a_t), there exists an equivalent measure $\mathbb{Q}$, such that $W_t = W_t^{\mathbb{Q}} - \int_0^t \lambda_s ds$, where $W^{\mathbb{Q}}$ is a BM under $\mathbb{Q}$.*

Then, we have:

$$d(St/B_t) = d(e^{-rt}S_t) = -re^{-rt}S_t dt + e^{-rt}dS_t = e^{-rt}(\mu - r)S_t dt + e^{-rt}\sigma S_t dW_t$$

$$= e^{-rt}(\mu - r - \sigma\lambda_t)S_t dt + e^{-rt}\sigma S_t dW_t^{\mathbb{Q}}$$

The condition $\mu - r - \sigma\lambda_t = 0$ implies $\lambda_t = (\mu - r)/\sigma$. Hence, under $\mathbb{Q}$

$$dS_t = \mu S_t dt + \sigma S_t dW_t = rS_t dt + \sigma S_t dW_t^{\mathbb{Q}}$$

Since $W^{\mathbb{Q}}$ is just another BM (and there are no other BMs in the model, under the same measure), we will ofter drop the "tilde" and use W in place of $W^{\mathbb{Q}}$, emphasizing that the dynamics are written under the pricing measure $\mathbb{Q}$.

Once the dynamics of S under $\mathbb{Q}$ are established, we can derive the BSPDE from the FTAP 1, without using the hedging argument.

Recall that, under $\mathbb{Q}$, the discounted price processes of ALL tradable assets have to be martingales!

In particular, under $\mathbb{Q}$, the discounted option price $V(S_t,t)/B_t$ has to be a mtg as well. Using Ito's lemma, we obtain

$$d(V(S_t,t)/B_t) = -re^{-rt}V(S_t,t)dt + e^{-rt}dV(S_t,t)$$
$$= e^{-rt}\left(\partial_t V(S_t,t) + rS_t\partial_S V(S_t,t) + \frac{1}{2}\sigma^2 S_t^2\partial_{SS}^2 V(S_t,t) - rV(S_t,t)\right)dt + e^{-rt}\sigma S_t\partial_S V(S_t,t)dW_t^{\mathbb{Q}}$$

And the martingale property implies

$$\partial_t V(S_t,t) + rS_t\partial_S V(S_t,t) + \frac{1}{2}\sigma^2 S_t^2\partial_{SS}^2 V(S_t,t) - rV(S_t,t) = 0,$$

which yields the BSPDE.

Rem 4. *BS model is* **complete***: any contingent claim can be hedged, and the pricing measure is unique.*

Dividends

If the risky asset S pays dividends at a constant rate q, then, under $\mathbb{Q}$, we have

$$dS_t = (r-q)S_t dt + \sigma S_t W_t^{\mathbb{Q}}$$

The above derivations can be repeated, and the BSPDE becomes:

$$\frac{\partial}{\partial t}V + \frac{1}{2}\sigma^2 S^2\frac{\partial^2}{\partial S^2}V + (r-q)S\frac{\partial}{\partial S}V - rV = 0$$

1.2 Solving the Black-Scholes PDE

Connection to the heat equation

$$\frac{\partial}{\partial t}V + \frac{\sigma^2}{2}S^2\frac{\partial^2}{\partial S^2}V + rS\frac{\partial}{\partial S}V - rV = 0, \quad t\in(0,T),\ S>0,$$
$$V(S,T) = F(S)$$

The BSPDE can be reduced to the **heat equation**.

- Introduce a change of variables

$$x = x(S) = \log(S), \quad S(x) = e^x, \quad u(x,t) = V(e^x,t), \quad V(S,t) = u(\log(S),t)$$

Then, using the chain rule, we obtain

$$\frac{\partial}{\partial S}V = \frac{1}{S}\frac{\partial}{\partial x}u(\log(S),t),$$
$$\frac{\partial^2}{\partial S^2}V = -\frac{1}{S^2}\frac{\partial}{\partial x}u(\log(S),t) + \frac{1}{S^2}\frac{\partial^2}{\partial x^2}u(\log(S),t),$$

- Plugging this in BSPDE, we obtain

$$\frac{\partial}{\partial t}u + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}u + (r-\sigma^2/2)\frac{\partial}{\partial x}u - ru = 0, \qquad t \in (0,T),\ x \in \mathbb{R}, \tag{3}$$

  with the terminal condition $u(x,T) = F(e^x)$.

- The above equation has constant coefficients, which is sufficient for the analysis we are about to carry out (in fact, simply bounded coefficients is enough). However, we can proceed further and reduce the above parabolic PDE to the classical heat equation.

  – Namely, we make the following change of variables:

$$v(x,t) = e^{-rt}u(x + (r-\sigma^2/2)t, t), \qquad u(x,t) = e^{rt}v(x - (r-\sigma^2/2)t, t),$$

  $$\frac{\partial}{\partial t}u(x,t) = e^{rt}\frac{\partial}{\partial t}v(x-(r-\sigma^2/2)t,t) - e^{rt}(r-\sigma^2/2)\frac{\partial}{\partial x}v(x-(r-\sigma^2/2)t,t) + re^{rt}v(x-(r-\sigma^2/2)t,t),$$

  which we plug into (3) to obtain

$$\frac{\partial}{\partial t}v + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}v = 0, \qquad t \in (0,T),\ x \in \mathbb{R}, \tag{4}$$

  with the terminal condition $v(x,T) = e^{-rT}F(\exp\big(x + (r-\sigma^2/2)T\big))$.

Every linear backward parabolic PDE, satisfying certain regularity conditions, has a **fundamental solution** $p(x,t;y;s)$, defined for (x,t) and (y,s) in the domain of the equation, and with $t < s$.

- The fundamental solution $p(x,t;y,s)$, roughly speaking, can be interpreted as a solution to the PDE in (x,t) variables, with a terminal condition at $t = s$ given by a **delta-function** at y, $\delta(x-y)$.

- Every continuous function $G : \mathbb{R} \to \mathbb{R}$ can be represented as

$$G(x) = \int_{\mathbb{R}} G(y)\delta(x-y)dy$$

  In fact, the above relation is used in the rigorous definition of delta-function.

- Then, we deduce from the linearity of the heat equation that its solution $v(x,t)$, with terminal condition $v(x,s) = G(x)$, has to be given by

$$v(x,t) = \int_{\mathbb{R}} G(y)p(x,t;y,s)dy \tag{5}$$

Rem 5. Rigorous definition of the fundamental solution*: function $p(x,t;y,s)$, with $(x,t),(y,s) \in \mathbb{R}\times[0,T]$ and $t < s$, is the fundamental solution of a given backward parabolic PDE if, for any continuous polynomially bounded function G and any $s \in (0,T]$, the solution u to this PDE, with the terminal condition G at time $t = s$, is given by (5).*

- In general, fundamental solutions are not available in closed form. However, the standard (backward) heat equation (4) has the following fundamental solution

$$p(x,t;y,s) = \frac{1}{\sigma\sqrt{2\pi(s-t)}} e^{-\frac{(x-y)^2}{2\sigma^2(s-t)}}$$

- Because the BSPDE (2) can be reduced to the heat equation, we can also derive its fundamental solution in a closed form. **This allows us to price any European option by simple numerical integration**:

$$v(x,t) = \frac{1}{\sigma\sqrt{2\pi(s-t)}} \int_{\mathbb{R}} e^{-rT} F(\exp\big(x + (r-\sigma^2/2)T\big)) e^{-\frac{(x-y)^2}{2\sigma^2(s-t)}}\, dy,$$

$$V(S,t) = e^{rt} v(\log S - (r-\sigma^2/2)t, t)$$

$$= \frac{e^{-r(T-t)}}{\sigma\sqrt{2\pi(s-t)}} \int_{\mathbb{R}} F(\exp\big(\log S + (r-\sigma^2/2)(T-t)\big)) e^{-\frac{(\log S - (r-\sigma^2/2)t - y)^2}{2\sigma^2(s-t)}}\, dy$$

- The above formula can be simplified for the call or put options. Namely, if $F(S) = (S-K)^+$, and in the presence of possible dividends of rate q, the BS price of a European call is given by

$$C^{BS}(S,K,T-t,\sigma,r,q) = Se^{-q(T-t)}\mathcal{N}(d_1) - Ke^{-r(T-t)}\mathcal{N}(d_2),$$

$$d_1 = \frac{\log(S/K) + (r-q+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}},$$

$$d_2 = d_1 - \sigma\sqrt{T-t},$$

where $\mathcal{N}$ is the cdf of a standard normal – this is where the numerical integration is required (but, luckily, $\mathcal{N}$ has already been implemented in MatLab).

2 Diffusion models

2.1 Implied smile

Q 2. *How do we choose the value of σ in the BS model?*

- Estimation from historical data:

$$\frac{1}{(N-1)\Delta t} \sum_{i=1}^{N} \left(\frac{S_{t_i+\Delta t} - S_{t_i}}{S_{t_i}} \right)^2.$$

This may be a natural choice if one wants to use the BS model to study the risky asset S alone, or its volatility.

- However, if one trades options (especially options with multiple strikes and/or maturities), the preferred way is to deduce σ from the options' prices currently observed in the market. This gives rise to the **Implied Volatility** (IV). Given the market price of an option, find the volatility Σ^I for which the BS price coincides with the market price:

$$V_t^{mkt} = V^{BS}(\Sigma^I)$$

Implied Volatility is defined as a solution to the above equation, where the options in question are the European call (or put) options:

$$C_t^{mkt} = C^{BS}(\Sigma^I)$$

IV is well defined, provided

- C_t^{mkt} is **not impossible**: $C_t^{mkt} \in \left[\left(S - Ke^{-r(T-t)}\right)^+, S\right]$.
- And $\frac{\partial}{\partial\sigma}C^{BS}$ **does not change its sign**, as a function of σ.

The first condition is satisfied, since otherwise there is a **model-independent arbitrage**.

The second condition is satisfied, since, as the BS Vega is always nonnegative.

If the BS model was true, there would exist one value of implied volatility Σ^I for call options of all strikes and maturities.

However, **this is not true** in practice! Typically, for each pair (K, T), we have a different value of implied vol $\Sigma^I(K, T)$.

Plotted as a function of *negative log-moneyness* $x = \log(K/S)$, this function is typically *convex* around $x = 0$, and, hence, is often referred to as the **implied smile**. See Figure 1.

Rem 6. *Log-moneyness of a call or a put option is defined as* $\log(S/K)$.

In equity markets (where S is the price of a stock or stock index), the implied smile typically has a **negative skew**, assigning higher values to negative $x = \log(K/S)$ (i.e. $K < S$). See Figure 1.

A behavioral explanation for the presence of skew is based on the fact that people tend to **overestimate** the **risks of extreme negative events** (that's why insurance companies make such a nice profit).

Another, rational, explanation is related to the so called "**leverage effect**". It is based on the observation that the stock returns and its volatility are negatively correlated. We will discuss this effect further in this section.

In FX (foreign exchange) markets (where S is the exchange rate of one currency for another), on contrary, the smile is more-or-less symmetric around at-the-money value ($x = 0$). The above argument does not apply in this case, because, typically, the risks of extreme negative events (such as defaults) for large countries are believed to be negligible. (*although things have changed since 2008...*)

Recall that we need to calibrate a model, for example, in order to produce an **arbitrage-free price of an option that is not liquidly traded**.

For example, this could be a call with a strike that is not among the ones that re traded on the exchange (i.e. does not coincide with the horizontal coordinates of any of the red and green dots in Figure 1). This problem, in principle, can be solved by interpolating the IVs given by the market: i.e. by producing the blue line in Figure 1. Indeed, given the $\Sigma^I(K)$, for any K, we obtain the price of any call option via the BS formula.

Q 3. *In order to avoid mispricing the liquid options, it is clear that the blue line in Figure 1 has to go between the green and red dots. However,* **is any such interpolation acceptable***? In other words,* **can we draw the blue line in Figure 1 arbitrarily***, with the only constraint that it goes between the green and red dots?*

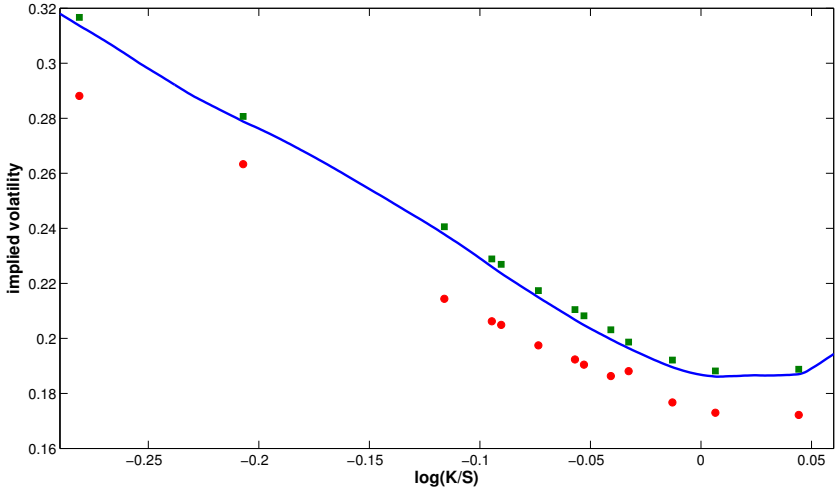

Figure 1: Interpolated implied volatility as a function of $\log(K/S)$. Computed from European call options on S&P500 on January 12, 2011

NO! There are additional conditions that make this interpolation a very complicated task. In fact, to construct the blue line in Figure 1, we had to calibrate a model to the observed call prices and then use this model to construct the interpolation (the blue line).

The main reason why interpolation of IVs (or, equivalently, call prices) is not straightforward, is that the interpolated IV $\Sigma^I(K,T)$ has to produce call prices $C(K,T) = C^{BS}(K,T,\Sigma^I(K,T))$, which satisfy the so called "static no-arbitrage" conditions:

- $C(K,T)$ is decreasing and convex in K;
- if r and q are constant, then $e^{qT}C(Ke^{(r-q)T},T)$ is increasing in T, for each K;
- $C(K,T)$ satisfies the boundary conditions:
  - $C(K,0) = (S-K)^+$;
  - $C(0,T) = S_0$;
  - $C(\infty,T) = 0$.

If the above conditions are not satisfied, then, one can construct an arbitrage portfolio using the call options. Hence, the **interpolation of implied vol (IV) has to ensure the above conditions are satisfied!**

Exercise 2. *Construct an arbitrage portfolio assuming the condition $\partial^2_{KK}C(K,T) \geq 0$ is violated.*

2.2 Diffusion-based models

We have seen that BS model cannot be calibrated to the call (or put) prices with multiple strikes and maturities. In particular, it is inconsistent with certain persistent phenomena, such as the skew of IV. To address this problem, here, we consider the class of general **diffusion** models.

Def 1. *Diffusion process X is defined as a solution to an SDE of the form*

$$dX_t = b(X_t, t)dt + a(X_t, t)dW_t, \tag{6}$$

where a and b are given functions, and W is a BM.

Rem 7. *In general, a solution to (6)* **may not exist** *or* **may not be unique**. *However, if the functions a and b are* **regular enough**, *there* **exists a unique solution to the above SDE**, *for any deterministic initial condition. For example, this is the case if a and b are* **Lipschitz continuous with at most linear growth**.

When the drift and diffusion coefficients, b and a, respectively, do not satisfy the desired continuity and growth conditions, one has to be careful with the notion of a solution to (6). For example, even if a solution exists, it may be restricted to a certain domain, and one needs to impose the appropriate boundary conditions, to determine the solution uniquely.

Rem 8. *The process X may be multi-dimensional. In this case,*

- *X is a vector of stochastic processes $X_t = (X_t^1, \dots, X_t^n)^T$,*
- *$b(x,t) = (b^1(x,t), \dots, b^n(x,t))^T$ is a vector-valued function,*
- *$W_t = (W_t^1, \dots, W_t^m)^T$ is a multidimensional BM,*
- *and $a(x,t) = (a^{i,j}(x,t))_{i=1,\dots,n,\, j=1,\dots,m}$ is a matrix-valued function.*

Rem 9. *Diffusion processes include all* **continuous strongly Markov processes** *with values in $\mathbb{R}^n$.*

Pricing and hedging in diffusion models

In a diffusion-based model, the risky asset S and the short interest rate r are modeled as functions X_t and t: $S_t = S(X_t, t)$, $r = r(X_t, t)$.

To price a European option, with payoff $F(S_T)$ at time T, in a diffusion model, we

- **determine the dynamics of X under the risk-neutral measure** $\mathbb{Q}$ (i.e. choose the drift of X so that S/B is a mtg);
- and **derive the drift of the discounted option price and set it to zero, just like it was done in the BS model**.

As a result, we obtain a PDE for the price function, which we still refer to as the BSPDE. The latter is possible because the price of the claim can still be viewed as $V_t = V(X_t, t)$.

To hedge an option with price $V(X_t, t)$, we

- derive the differential of the hedging portfolio, which has π_t units of S at time t,
- and minimize its diffusion component (coefficient in front of dW_t) over π_t:

$$\pi_t dS(X_t, t) - dV(X_t, t) = \alpha(X_t, t; \pi_t)dt + \beta(X_t, t; \pi_t)dW_t,$$

$$\pi_t = \text{argmin}_\pi \|\beta(X_t, t; \pi)\|$$

Rem 10. *In general, the BSPDE can be defined for X and W having any dimension. However, the completeness is typically lost if the dimension of W is larger than the dimension of S, hence there may not exist a perfect hedging strategy.*

2.3 Local volatility models

Local volatility (LV) model is a model in which the risky asset S is given by a one-dimensional diffusion process.

In the notation introduced earlier, $S = X$, with

- $b(S, t) = \mu(S, t)S$
- and $a(S, t) = \sigma(S, t)S$,

for some functions μ and σ, which are called the drift and local volatility, respectively.

In other words,

$$dS_t = \mu(S_t, t)S_t dt + \sigma(S_t, t)S_t dW_t$$

Intuitively,

- $\mu(S_t, t)$ is the "**instantaneous drift**" at time t;
- and $\sigma(S_t, t)$ is the "**instantaneous volatility**" at time t.

Here, for simplicity, we assume that the interest rate r is constant, although it is easy to extend this setting to the case of $r = r(S_t, t)$.

It turns out that the **flexibility provided by the choice of function σ is sufficient to fit the prices of European call (or put) options with any number of strikes and maturities!** We will discuss this fact in more detail later in the course.

Pricing and hedging in LV models

Consider a European option, with payoff $F(S_T)$ at time T, in a LV model.

- First, notice that, under the risk neutral measure $\mathbb{Q}$, S/B is a mtg. Hence,

$$dS_t = rS_t dt + \sigma(S_t, t)S_t dW_t$$

- Then, we repeat the derivations made for BS model, setting the drift of the discounted option price to zero. As a result we obtain the pricing PDE

$$\frac{\partial}{\partial t}V + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2}{\partial S^2}V + rS\frac{\partial}{\partial S}V - rV = 0, \quad t \in (0,T),\ S > 0,$$

with the terminal condition $V(S,T) = F(S)$.

We still refer to this equation as **the Black-Scholes PDE (BSPDE)**.

- Similarly, we obtain the delta-hedging formula

$$\pi_t = \partial_S V(S_t, t)$$

LV models are typically complete!

- We can also change the variables and transform the BSPDE as before:

$$x = x(S) = \log(S), \quad S(x) = e^x, \quad u(x,t) = V(e^x,t), \quad V(S,t) = u(\log(S),t),$$

$$\frac{\partial}{\partial t}u + \frac{1}{2}\sigma^2(e^x,t)\frac{\partial^2}{\partial x^2}u + (r - \frac{1}{2}\sigma^2(e^x,t))\frac{\partial}{\partial x}u - ru = 0, \quad t \in (0,T),\ x \in \mathbb{R}, \tag{7}$$

with the terminal condition $u(x,T) = F(e^x)$.

- We can continue transforming the above equation and reduce it to a heat equation. However, the resulting heat equation will have time- and space-dependent coefficient, and its fundamental solution, in general, is not available in a closed form.

- In fact, it is only beneficial to consider (7) if $\sigma(S,t)$ is bounded.

Shifted (Displaced) Lognormal model

Main idea: the firm's value consists of the equity and debt. Assuming that the value of the firm follows a GBM, its equity (i.e. stock) price has to follow a **Shifted (Displaced) Lognormal model**:

$$S_t = G_t + \theta,$$

where G is a **GBM**, θ is a constant (can also be a deterministic function of time), which, in the case of equity markets represents the **level of debt** of the company.

We assume that r is constant.

In this model, the **local volatility function** is

$$\sigma(S_t,t) = \sigma(S_t) = \sigma\frac{S_t - \theta}{S_t}$$

It is easy to see that, if $\theta < 0$, then the **volatility becomes a decreasing function of the stock value**. This model explains the connection between **leverage** and the **negative correlation between stock price and its volatility**.

$$dS_t = dG_t = \mu G_t dt + \sigma G_t dW_t = \mu G_t dt + \sigma(S_t - \theta)dW_t$$

Under the risk neutral measure $\mathbb{Q}$, S/B is a mtg, hence the drift of S will change to rS_t:

$$dS_t = rS_t dt + \sigma(S_t - \theta)dW_t^{\mathbb{Q}}$$

However, if $\theta < 0$, there is a problem with this model.

- Notice that S_t takes values in (θ, ∞). If $\theta < 0$, then S_t can take negative values, which results in **arbitrage**, as our standing assumption is "limited liability" which implies that $S_t \geq 0$.
- The static no-arbitrage conditions rely on the assumptions that $S \geq 0$.
- All this, in particular, implies that **one may not be able to calibrate this model to the given market prices of call options**.

Nevertheless, **if the range of available strikes is not too large, and if $|\theta|$ is not too large, the Shifted Lognormal model produces call prices that do not contradict the assumption** $S \geq 0$.

- Let us compute call prices in the Shifted Lognormal model.
  - In the case when $r = 0$, we can either solve the BSPDE

    $$\frac{\partial}{\partial t}V + \frac{1}{2}\sigma^2(S-\theta)^2\frac{\partial^2}{\partial S^2}V = 0, \quad t \in (0,T),\ S > 0,$$

    with the terminal condition $V(S,T) = (S-K)^+$, via the following change of variables: $S \mapsto S - \theta$;
  - or, we can simply use the risk neutral pricing formula directly:

    $$C(K,T) = \mathbb{E}^{\mathbb{Q}}(S_T - K)^+ = \mathbb{E}^{\mathbb{Q}}(G_T - (K-\theta))^+$$
    $$= C^{BS}(S_0 - \theta, K - \theta, T, \sigma),$$

    since, under $\mathbb{Q}$

    $$dG_t = dS_t = \sigma G_t dW_t^{\mathbb{Q}}$$

- Computing the IV, $\Sigma^I(K,T)$, of $C(K,T)$, we can deduce that, for any fixed maturity T,
  - $\Sigma^I(K,T)$ is increasing in K, if $\theta > 0$, and decreasing in K if $\theta < 0$.
  - $\Sigma^I(K,T) \to \sigma$, as $K \to \infty$.

Shifted Lognormal is not a very good model, as it produces negative stock prices, which is inconsistent with our basic assumptions and, hence, may result in option prices that contradict the static no-arbitrage conditions. However, it is fairly popular among practitioners due to its simplicity.

Shifted Lognormal is also used for modeling short rate r_t. In this case, it is rather unusual that r_t becomes negative but it does not lead to any arbitrage!

Constant Elasticity of Variance (CEV) model

This was the first popular extension of the BS model.
The LV function is given by

$$\sigma(S,t) = \delta S^{\beta/2-1},$$

and, under $\mathbb{Q}$

$$dS_t = rS_t dt + \delta S_t^{\beta/2} dW_t$$

Here, again, we have a LV that is a **decreasing function of S, when $\beta < 2$** – hence, the **leverage effect is reproduced by the model**.

When $\beta = 2$, we recover the BS model.

If $\beta < 2$, this model also has some issues when S_t approaches zero.

- In fact, if we start from a positive level $S_0 > 0$, there is always a **positive portability that S hits zero** before any finite time $T > 0$.
- It turns out that the **solution of the above SDE is not defined uniquely when $S_0 = 0$**. This is due to the fact that σ is **unbounded around zero**.

However defining **zero to be an absorbing point**, we obtain a well-defined arbitrage-free model (provided $\delta > 0$).

Options, in this model, can be priced by the BSPDE

$$\frac{\partial}{\partial t}V + \frac{1}{2}\delta^2 S^\beta \frac{\partial^2}{\partial S^2}V + rS\frac{\partial}{\partial S}V - rV = 0, \quad t \in (0,T),\ S > 0$$

However, there is no explicit solution to this equation, unless $\beta = 2$. Hence, we need to develop numerical approximation of its solution.

Rem 11. *Alternatively, one can reduce the CEV process to a normalized and time-changed Bessel process, and develop the pricing formulas for call and put options using the properties of Bessel processes (via the "modified Bessel functions")*

Vasicek model

It is different from what we have seen before. This is a fixed-income model: there is **no tradable risky asset** S. Instead, we **model the short interest rate** r as a stochastic process: $r = X$.

More specifically r is modeled as the **Ornstein-Uhlenbeck (OU)** process:

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t$$

Notice that, since r is not tradable, there is **no restriction on its drift** under the pricing measure $\mathbb{Q}$!
This is why, in such models, we typically postulate the dynamics under $\mathbb{Q}$ directly!

A nice feature of this model is that the short rate exhibits **mean-reverting** behavior.

The problem with this model is that r_t **can be negative** with positive probability. But this **does not lead to arbitrage**, and may not be such a bad feature, especially given the recent events (recall negative yield on German bonds in Aug 2014).

Any contingent claim in this model, whose time t price is in the form $V(r_t, t)$, satisfies the BSPDE:

$$\frac{\partial}{\partial t}V + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial r^2}V + \kappa(\theta - r)\frac{\partial}{\partial r}V - rV = 0, \quad t \in (0,T),\ r \in \mathbb{R} \tag{8}$$

A typical example of an **option on interest rate** is the **zero-coupon default-free bond**. To find the price of such a bond, with maturity T and face value 1, we need to solve the above PDE with the terminal condition

$$V(r,T) = 1.$$

In fact, Vasicek belongs to the class of the so called **affine models**. And it is well known that in such models the price of a default-free zero-coupon bond is given by

$$P(t,T) = A(t,T)e^{-B(t,T)r_t},$$

or

$$V(r,t) = A(t)e^{-B(t)r}$$

with some deterministic functions A and B. Plugging the above representation into (8), we obtain

$$A'e^{-Br} - B'rAe^{-Br} + \frac{1}{2}\sigma^2 B^2 Ae^{-Br} - \kappa(\theta - r)BAe^{-Br} - rAe^{-Br} = 0,$$

which splits into two equations

$$\begin{cases} A' + \frac{1}{2}\sigma^2 B^2 A - \kappa\theta BA = 0, \\ -B' + \kappa B - 1 = 0, \end{cases}$$

for $t \in (0,T)$, with the terminal conditions $A(T) = 1$ and $B(T) = 0$.

It is not difficult to solve the above system, and obtain

$$B(t,T) = \frac{1}{\kappa}\left(1 - e^{-\kappa(T-t)}\right),$$

$$A(t,T) = \exp\left((\theta - \frac{\sigma^2}{2\kappa^2})(B(t,T) - T + t) - \frac{\sigma^2}{4\kappa}(B(t,T))^2\right)$$

Rem 12. *We will later derive an explicit solution for the OU equation, which will show that r is, in fact, a* **Gaussian process***. Then, the bond can be priced by via the formula*

$$P(t,T) = \mathbb{E}_t^{\mathbb{Q}} \exp(-\int_t^T r_t dt),$$

directly.

Cox-Ingersoll-Ross (CIR) model

This is also a fixed-income model: there is no tradable risky asset S. Instead, we model the short interest rate r as a stochastic process: $r = X$.

Here, r is modeled as the CIR process:

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t$$

CIR model was obtained from the General Equilibrium arguments by Cox, Ingersoll and Ross in 1985.

In this model, r also exhibits the **mean-reverting behavior**.

In addition, if $2\kappa\theta > \sigma^2$, then r_t **remains positive** at all times.

If $2\kappa\theta \leq \sigma^2$, then, we have to add a boundary condition for (r_t) at zero, making zero an **absorbing** point.

Any contingent claim in this model, whose time t price is in the form $V(r_t, t)$, satisfies the BSPDE:

$$\frac{\partial}{\partial t}V + \frac{1}{2}\sigma^2 r\frac{\partial^2}{\partial r^2}V + \kappa(\theta - r)\frac{\partial}{\partial r}V - rV = 0, \quad t \in (0,T),\ r > 0 \tag{9}$$

If r can hit zero, then an additional **boundary condition at zero** is required.

As before, we are interested in the price of a **zero-coupon default-free bond**. To find the price of such a bond, with maturity T and face value 1, we need to solve the above PDE with the terminal condition

$$V(r,T) = 1.$$

CIR also belongs to the class of **affine models** (notice that the coefficients in the above PDE are affine functions of r).

Hence, assuming $2\kappa\theta > \sigma^2$, we search for the price of a default-free zero-coupon bond in the form

$$P(t,T) = A(t,T)e^{-B(t,T)r_t},$$

or

$$V(r,t) = A(t)e^{-B(t)r}$$

with some deterministic functions A and B. Plugging the above representation into (9), we obtain

$$A'e^{-Br} - B'rAe^{-Br} + \frac{1}{2}\sigma^2 rB^2Ae^{-Br} - \kappa(\theta - r)BAe^{-Br} - rAe^{-Br} = 0,$$

which splits into two equations

$$\begin{cases} A' - \kappa\theta BA = 0, \\ -B' + \frac{1}{2}\sigma^2 B^2 + \kappa B - 1 = 0, \end{cases}$$

for $t \in (0,T)$, with the terminal conditions $A(T) = 1$ and $B(T) = 0$.

The above system is nonlinear. However, an explicit solution is still available:

$$h = \sqrt{\kappa^2 + 2\sigma^2},$$

$$B(t,T) = \frac{2(\exp((T-t)h) - 1)}{2h + (\kappa + h)(\exp((T-t)h) - 1)},$$

$$A(t,T) = \left(\frac{2h\exp((T-t)(\kappa+h)/2)}{2h + (\kappa + h)(\exp((T-t)h) - 1)}\right)^{2\kappa\theta/\sigma^2}$$

Calibration to initial term-structure and the Hull-White extensions

The Vasicek and CIR models, in particular, provide bond prices at the current time, $P(0,T)$, for various maturities T. This is called the **term structure of bond prices**.

Note that this term structure (i.e. bond prices for multiple maturities), at least partially, can be **observed from the market** at the initial time.

Rem 13. *Typically, the term structure (i.e. bond prices for multiple maturities) is given in the form of a* **yield curve***:*

$$Y(0,T) = -\partial_T \log P(0,T),$$

in the case of continuous compounding, and

$$Y(0,\Delta tn) = \frac{(P(0,\Delta tn))^{-\frac{1}{n}} - 1}{\Delta t},$$

if the compounding is discrete with the time step Δt*.*

It turns out that, very often, **neither Vasicek nor CIR can reproduce the observed initial term structure precisely** – they may not even be able to capture its qualitative behavior, such as changes in monotonicity, etc. (e.g. it is easy to see that the yield curve produced by the Vasicek model converges to a constant level exponentially fast, as $T \to \infty$).

This is why Hull and White proposed the **extensions** of the Vasicek and CIR models, which allow for **time-dependent coefficients**:

$$dr_t = \kappa(t)(\theta(t) - r_t)dt + \sigma(t)dW_t$$

or

$$dr_t = \kappa(t)(\theta(t) - r_t)dt + \sigma(t)\sqrt{r_t}dW_t,$$

where κ, θ and σ are deterministic functions of time.

As we will see later in the course, the Hull-White extended Vasicek and CIR models allow for enough flexibility in the choice of κ, θ and σ to **fit the initial term structure**.

However, in order to calibrate to the initial term structure, one needs to compute bond prices in these models. It turns out that, for the Hull-White extension of the Vasicek model, the closed-form expression is still available, but it is not available for the extended CIR model. In the latter case, one, for example, needs to solve the corresponding BSPDE,

$$\frac{\partial}{\partial t}V + \frac{1}{2}\sigma^2(t)r\frac{\partial^2}{\partial r^2}V + \kappa(t)(\theta(t) - r)\frac{\partial}{\partial r}V - rV = 0, \quad t \in (0,T),\ r > 0,$$

or the associated system of ODEs, **numerically**.

3 Finite difference methods for evolution equations

Evolution equations are the equations that prescribe the **changes** of a time-dependent object (typically, **vector** or **function**) as the time variables goes forward or backward, by specifying the differential of this object with respect to time. Here, we will study the Ordinary Differential Equations (ODEs) and the (parabolic) Partial Differential Equations (PDEs).

3.1 Ordinary Differential Equations (ODE)

A differential equation in n variables $y = (y_1, .., y_n) \in \mathbb{R}^n$ is of the form

$$\frac{dy(t)}{dt} \ = \ f(y(t), t)\ , \tag{10}$$

where $f(y,t) = (f_1(y,t), .., f_n(y,t)) \in \mathbb{R}^n$ is a vector valued function commonly referred to as a **vector field**. Thus for every $y \in \mathbb{R}^n$ it yields a time dependent vector $f(y,t)$ in $\mathbb{R}^n$.

The **fundamental theorem of ODE** tells us that the equation (10) has a unique solution both forward and backward in time when f is a regular enough function and $y(t)$ is specified at a certain time say t_0, so that $y(t_0) = y^0$ is given.

In most applications in applied mathematics we wish to solve (10) **forward in time** i.e. for $t \geq t_0$. In that case we say the condition $y(t_0) = y^0$ is the **initial condition** for the differential equation (10). An example of this occurs in fluid flow. When we place a light particle such as a leaf in a stream it will be carried along by the fluid. In that case $f(y,t)$ is the fluid velocity at the position y at time t, and y^0 is the position we put the leaf at the initial time t_0. We shall also be interested in solving differential equations **backward in time** i.e. for $t \leq t_0$. In that case the condition $y(t_0) = y^0$ is called the **terminal condition** for the equation (10). Such problems come up very often in the mathematical finance. A familiar example of such a problem for an **infinite dimensional** ODE is the problem of finding the value of a European option (e.g. a call option). The value of the option as a function of the stock price S is known at the expiration date $T > 0$ of the option, and the bank wishes to compute the value of the option today which is time $t = 0 < T$ say. In the Black-Scholes (BS) theory one finds the value of the option by solving an infinite dimensional linear ODE like (10) backward in time until time $t = 0$ with the terminal condition given by the option price at the expiration date T.

It is sometimes possible to solve the differential equation (10) explicitly. The most common situation where that occurs is when the vector field $f(y,t)$ is **linear** so that

$$f(y,t) \ = \ A(t)y + B(t) \ , \quad A(t), B(t) \text{ time dependent } n \times n \text{ matrices.} \tag{11}$$

However, in general, one cannot produce explicit formulas for the solutions to (10), or its infinite-dimensional version: recall the pricing PDEs in the CEV model and in general LV model, as well as the associated PDE and ODE in the Hull-White extended CIR model. When an explicit solution is not available we need to resort to **numerical methods**.

In this course, we discuss the numerical methods based on **finite differences**. Namely, we discretize the time variable into multiples of some small basic time increment Δt, whence time now takes the discrete values $0, \Delta t, 2\Delta t, 3\Delta t, ...$ We replace the derivative in (10) by a finite difference which approximates the derivative for small Δt. There are several ways to carry this out:

$$\text{Forward difference :} \quad \frac{dy(t)}{dt} \simeq \frac{y(t+\Delta t) - y(t)}{\Delta t} \ , \tag{12}$$

$$\text{Backward difference :} \quad \frac{dy(t)}{dt} \simeq \frac{y(t) - y(t-\Delta t)}{\Delta t} \ , \tag{13}$$

$$\text{Central difference :} \quad \frac{dy(t)}{dt} \simeq \frac{y(t+\Delta t) - y(t-\Delta t)}{2\Delta t} \ . \tag{14}$$

Note that the central difference is the average of the forward and backward difference. The error made in these approximations goes to 0 as $\Delta t \to 0$. It goes faster however to 0 in the central difference approximation (14) than in the forward or backward difference. We can see this by doing a Taylor expansion

$$y(t+\Delta t) = y(t) + y'(t)(\Delta t) + y''(t)(\Delta t)^2/2 + y'''(t)(\Delta t)^3/6 + \cdots \tag{15}$$

It follows that if we denote the error by $\varepsilon(\Delta t)$ so that

$$\varepsilon(\Delta t) \ = \ \text{difference approximation} - dy(t)/dt \ , \tag{16}$$

then $\varepsilon(\Delta t) = O(\Delta t)$ for the forward and backward differences, but $\varepsilon(\Delta t) = O(\Delta t^2)$ for the central difference approximation. Thus we say that the forward and backward differences are **first order accurate** approximations to the derivative, whereas the central difference is a **second order accurate** approximation.

We can use the difference approximations (12)–(14) to obtain a **finite difference scheme** for solving the ODE (10). Namely, we wish to solve (10) in the time interval $t \in (0, T)$. Thus, we consider a partition of the time interval into M intervals with length $\Delta t = T/M$.

Suppose first that we need to solve (10) **forward** in time, with initial condition $y(0) = y^0$ given. If we use the **forward difference** (12) to approximate the derivative, then we obtain the **explicit Euler scheme**:

$$y(t + \Delta t) \;=\; y(t) + \Delta t f(y(t), t). \tag{17}$$

On setting $y^m \sim y(m\Delta t)$, then (17) yields the recurrence relation

$$y^{m+1} = y^m + \Delta t f(y^m, m\Delta t)\;, \quad m = 0, 1, 2, \ldots, M-1. \tag{18}$$

The numerical algorithm is **explicit** in the sense that once we have computed y^m then y^{m+1} is easily computed by evaluating the RHS of (18).

Alternatively we may wish to solve (10) **backward** in time, with terminal condition $y(T) = y^M$. In this case we use the **backward difference** (13) to implement the **explicit Euler method**, whence we have

$$y(t - \Delta t) \;=\; y(t) - \Delta t f(y(t), t)\;. \tag{19}$$

Again on setting $y^m \sim y(m\Delta t)$, then (19) yields the recurrence relation

$$y^{m-1} \;=\; y^m - \Delta t f(y^m, m\Delta t)\;, \quad m = M, M-1, \ldots, 1. \tag{20}$$

Ex 1. *We consider the linear system*

$$y'(t) \;=\; ry(t)\;, \quad y(0) = 1, \tag{21}$$

which we wish to solve numerically for $t > 0$. Evidently the exact value of the solution at time $T > 0$ is $y(T) = e^{rT}$. Implementing the explicit Euler method (18) we see that

$$y^{m+1} \;=\; [1 + r\Delta t]y^m \quad m = 0, 1, 2, .., \;\; y^0 = 1. \tag{22}$$

On iterating this recurrence M times where $M\Delta t = T$, we obtain the approximate value for $y(T)$,

$$y(T) \sim y^M = [1 + r\Delta t]^M \;=\; [1 + r\Delta t]^{T/\Delta t}\;. \tag{23}$$

Since we know from calculus that

$$\lim_{\Delta t \to 0} [1 + r\Delta t]^{T/\Delta t} \;=\; e^{rT}\;, \tag{24}$$

we conclude that the numerical solution of the Euler scheme converges as $\Delta t \to 0$ to the exact solution of the ODE.

A numerical scheme for the solution of the differential equation (10) which converges as the time discretization $\Delta t \to 0$ to the solution of the ODE is called a **convergent numerical scheme**. Evidently we **should always want to use only convergent schemes**.

Thm 3. *Provided that f satisfies certain regularity conditions, the forward and backward Explicit Euler schemes are convergent.*

Another related notion is also important but is considerably more subtle. In numerical analysis we are not only interested in coming up with convergent schemes, but also in the issue of how accurate these schemes are for a given number of computations. Thus we introduce the notion of **stability** of a numerical scheme. We say the scheme has an interval of stability $\Delta t \in (0, \delta)$ if for Δt in this interval the numerical solution and the exact solution of the ODE are "reasonably close". The larger we can choose δ the more stable the scheme is. We have not of course defined what "reasonably close" is, but, intuitively, **stability means that the relative error of the approximation does not grow as the number of steps in the approximation scheme increases** (given that Δt is fixed). In other words, the errors made at each step of the scheme do not accumulate. In our applications, stability will be necessary for convergence: the numerical solution will **diverge** outside the region of stability. Stability is also very important in practice, because any real input data (such as y^0) contains noise, and one does not want the numerical scheme to amplify this noise.

We consider again the problem of **solving the ODE (10) forward in time**, for $t > 0$, with initial data $y(0) = y^0$. If instead of the forward difference as in (17), (18) we use the backward difference to approximate the derivative we now get

$$y(t - \Delta t) \ = \ y(t) - \Delta t f(y(t), t), \tag{25}$$

$$y^{m-1} \ = \ y^m - \Delta t f(y^m, (m)\Delta t) \ , \quad m = M, \ldots, 1, \tag{26}$$

but we wish to solve this **forward** in time. To compute y^m from the already computed value for y^{m-1} we need to solve the equation (26) for y^m, which requires **inverting** the function f, and so the scheme is **implicit**.

In **Example 1** it is easy to do this since the equation is

$$y^{m-1} \ = \ [1 - r\Delta t] y^m. \tag{27}$$

Thus, using the **implicit Euler method**, instead of (23), we obtain the formula

$$y(T) \sim y^M = [1 - r\Delta t]^{-M} \ = \ [1 - r\Delta t]^{-T/\Delta t} \ . \tag{28}$$

It is clear from (24), (28) that the implicit Euler method is also convergent. Its **stability properties however can be much better** than the explicit Euler method in the case when $r << 0$.

To see **why this is the case**, consider the equation

$$\frac{dy(t)}{dt} = -\lambda y(t) \ , \tag{29}$$

which we wish to solve **forward** in time $t > 0$ with initial data

$$y(0) = y^0 \ . \tag{30}$$

The **explicit Euler method** yields the following numerical approximation for the solution of (29)–(30) at time $T > 0$:

$$y(T) \sim [1 - \lambda \Delta t]^M y^0, \quad M = T/\Delta t \ . \tag{31}$$

The exact solution to (29)–(30) is of course

$$y(T) = e^{-\lambda T} y^0. \tag{32}$$

Note that if $\lambda \Delta t > 2$ then, for a fixed Δt, the numerical approximation $[1 - \lambda \Delta t]^M y^0$ is **growing exponentially** with the number of steps M, since in that case

$$1 - \lambda \Delta t < -1.$$

It also **oscillates** between large positive and large negative values. In contrast, the exact solution $y(T)$ is **exponentially smaller** in T than the initial data y^0. We conclude that if $\lambda \Delta t > 2$ the error produced by the explicit Euler method at

each step (i.e. the relative difference between the actual solution $y(m\Delta t)$ and its approximation y^m) grows with the number of steps. Hence the **stability interval** for the explicit Euler method is $\Delta t \in (0, 2/\lambda)$.

On the other hand, If we use the **backward difference** approximation for this **forward equation**, which results in the **implicit Euler scheme**, the numerical approximation of the solution to (29)–(30) at time T is given, as in (28), by

$$y(T) \sim [1 + \lambda \Delta t]^{-M} y^0, \quad M = T/\Delta t \, . \tag{33}$$

One immediately sees that (33) always gives a reasonable approximation to the exact solution (32) no matter how large Δt. The interval of stability for the implicit scheme is therefore $\Delta t < \infty$. The same conclusion holds if we use the **central difference** approximation instead of the backward difference.

Rem 14. *Notice that the conclusion of the above discussion could be different if $\lambda < 0$. Indeed, in that case, the explicit scheme would be more stable than the implicit one. However, there is a reason why we are interested in the case $\lambda > 0$ much more than in the case $\lambda < 0$. Namely, in what follows, we will consider* **linear parabolic Partial Differential Equations (PDEs)***, which, in fact, can be viewed as infinite-dimensional (or, if you wish, "very high-dimensional") systems of the ODEs of the form*

$$\frac{dy^i(t)}{dt} = -\lambda^i y^i(t), \quad t > 0, \quad i = 1, \ldots, n,$$

where $n \to \infty$ and $\max_i |\lambda^i| \to \infty$. In fact, after some mathematical constructions, a parabolic PDE can be viewed as a limit of the above systems, as $n \to \infty$. Assume, now, that λ^is are negative. Then, each $y^i(t)$ increases exponentially in t with rate $|\lambda^i|$. As $\max_i |\lambda^i| \to \infty$, our solution $y(t)$ will eventually **explode***, even for very small $t > 0$, hence, the solution will not be well defined for the limiting system. These heuristic arguments, in fact, can be made precise using the tools from Functional Analysis, but the above explanation is sufficient for the present course.*

Thus, the conclusion is that, for the type of equations that appear in Financial Mathematics (i.e. parabolic PDEs), the **implicit schemes** *can be much* **more stable** *than the explicit schemes.*

As announced in the above remark, solving the BSPDE can be viewed as solving a "large" (in fact, infinite) number of **coupled** ODEs. Let us illustrate this analogy using the standard **heat equation**, which is very similar to the BSPDE:

$$\frac{\partial u(x,t)}{\partial t} \;=\; \frac{\sigma^2}{2} \frac{\partial^2 u(x,t)}{\partial x^2} \, . \tag{34}$$

A typical problem is to solve (34) for $t > 0$ in the interval $0 < x < 1$ with specified **initial condition** at $t = 0$ and **boundary conditions** at $x = 0, 1$. If we think in terms of heat flow then the interval $0 < x < 1$ represents a bar of length 1 with initial temperature distribution given by some known function $u(x, 0),\ 0 < x < 1$. The temperature at the ends of the bar $x = 0, 1$ are kept fixed in time at temperatures u_0, u_1 respectively, so that $u(0, t) = u_0,\ u(1, t) = u_1$ for $t > 0$. We can find the temperature distribution $u(x, t),\ 0 < x < 1$, within the bar at a later time t by solving (34) with the given initial and boundary conditions.

To numerically solve this problem we introduce a **space discretization** $\Delta x = 1/N$, for some large integer N. In that case we can approximate the PDE (34) by a system of $N - 1$ ODEs

$$\frac{\partial u(n\Delta x, t)}{\partial t} = \frac{\sigma^2}{2\Delta x^2}(u((n+1)\Delta x, t) + u((n-1)\Delta x, t) - 2u(n\Delta x, t)),\ n = 1, ..N - 1,$$

which we can solve numerically by introducing a **time discretization** Δt. Each equation in this ODE system is like (29) with λ proportional to $\sigma^2/\Delta x^2$, whence the **stability condition** $\Delta t < 2/\lambda$ implies that

$$\sigma^2 \Delta t/(\Delta x)^2 \tag{35}$$

has to remain **bounded**. Hence a **stable explicit Euler method** for solving the heat equation requires that the time discretization is **proportional to the square of the space discretization**. We shall see this condition come up many times during the course.

Rem 15. *Notice that the stability issue for an ODE, such as (11), can be easily resolved: by decreasing Δt, we can always make sure that the finite difference scheme is stable. And we need to decrease Δt anyway, to obtain a better accuracy of the approximation. However, as follows from the above example, the stability problem for parabolic PDEs is a non-trivial issue: one can find a decreasing sequence of $(\Delta x, \Delta t)$ for which the scheme is* **not stable***, and, hence, does not converge.*

4 Solving the BSPDE numerically via explicit Euler method

4.1 European options

We recall the BS theory for the pricing of options. Consider a European option on a stock S with expiration date $T > 0$, where $t = 0$ represents today. In the case of a call option with strike price K the payoff on the option at the expiration date is $(S - K)^+$, where S is the stock price. To find the value of the call option in a LV model, we solve the **terminal value** problem for the PDE)

$$\frac{\partial V(S,t)}{\partial t} + \frac{\sigma^2(S,t)}{2} S^2 \frac{\partial^2 V(S,t)}{\partial S^2} + rS\frac{\partial V(S,t)}{\partial S} - rV(S,t) \ = 0 \quad \text{for } S > 0,\ t < T, \tag{36}$$

$$V(S,T) \ = \ (S-K)^+ \quad \text{for } S > 0. \tag{37}$$

Here, for convenience, we assume that $\sigma^2(S,t) \leq \bar{\sigma}^2$, for all (S,t), with some constant $\bar{\sigma}$.

Observe that the Black Scholes PDE (36) has coefficients which are unbounded functions of S. Motivated by the above discussion of the standard heat equation, we can expect that the **explicit Euler scheme will not be stable if the coefficient in front of ∂^2_{SS} is unbounded** (cf. (35)). Thus, we transform the equation into a PDE with bounded coefficients by making the change of variable $S = e^x$. Hence if we write $V(S,t) = u(x,t)$ then the problem (36)–(37) becomes the terminal value problem

$$\frac{\partial u(x,t)}{\partial t} + \frac{\sigma^2(e^x,t)}{2}\frac{\partial^2 u(x,t)}{\partial x^2} + \left\{ r - \frac{\sigma^2(e^x,t)}{2} \right\} \frac{\partial u(x,t)}{\partial x} - ru(x,t) \ = \ 0 \quad \text{for } -\infty < x < \infty,\ t \in (0,T), \tag{38}$$

$$u(x,T) \ = \ (S-K)^+ \quad \text{for } -\infty < x < \infty. \tag{39}$$

To numerically solve the BS terminal value problem (36)–(37) it is sufficient to numerically solve the constant coefficient problem (38)–(39). In order to do this we need to introduce a "space" discretization Δx for the logarithm x of the stock price as well as a time discretization Δt, which we have already done for numerically solving ODEs.

We also need to reduce the **infinite interval** $-\infty < x < \infty$ to a **finite interval** $a < x < b$. First note that the finite interval should certainly include **at the money options** so we need to have $a < \log K < b$. Next we choose a sufficiently small so that the initial stock price $S_0 = e^a$ is so far below K that the probability that the option expires in the money (i.e. $S_T \geq K$) is negligible, and, hence, we can set $V(e^a,t) = u(a,t) = 0$ for $0 \leq t \leq T$.

To decide on a good value for a we, first, assume that $\sigma \equiv \bar{\sigma}$ (i.e we assume a BS model) and, assuming $S_0 = e^a$, consider the representation

$$X_T = \log(S_T) = a + (r - \bar{\sigma}^2/2)T + \bar{\sigma}W_T = a + \xi,$$

where $\xi \sim N((r - \bar{\sigma}^2/2)T, \bar{\sigma}^2 T)$.

In practice, $(r - \bar{\sigma}^2/2)T \approx 0$, for reasonable values of maturity, hence we will drop it.

The probability of the Gaussian variable ξ being larger than 3 standard deviations above its mean is 0.0014 so about $1/10$ of 1 percent. Thus if we take $a = \log K - 3\bar{\sigma}\sqrt{T}$ then,

$$\mathbb{P}(S_T \geq K \,|\, S_0 = e^a) = \mathbb{P}(a + \xi \geq \log K) = \mathbb{P}(\xi \geq 3\bar{\sigma}\sqrt{T}) = 0.0014.$$

In other words, in the BS model with volatility $\bar{\sigma}$, the probability that the option is in the money at expiration is very small, and hence **we are justified in setting** $u(a,t) = 0$ **for** $0 \leq t \leq T$.

Notice, however, that **our model is not necessarily a BS model with volatility** $\bar{\sigma}$**!**

Thm 4. *Consider two LV models: with LV functions σ_1 and σ_2 (and with the same r), satisfying certain regularity conditions. Consider a European option with a convex payoff function (e.g. a call or a put): denote its current prices (for the same maturity) in the first and the second model by V^1 and V^2 respectively. Assume that $\sigma_1(S,t) \leq \sigma_2(S,t)$ for all (S,t). Then $V^1 \leq V^2$.*

The above theorem implies that we can choose a so that $C^{BS}(a, K, T, \bar{\sigma})$ **is small enough**, and the **LV model price** $C(a, K, T)$ **will be even smaller**! In the above, we set $a = \log K - 3\bar{\sigma}\sqrt{T}$, hence it is a good choice for our model.

To find an appropriate value for b and the value we should set for $u(b,t)$ we use **put-call parity**:

$$C(S, K, T-t) - P(S, K, T-t) = S - Ke^{-r(T-t)}$$

Observe now that if we set $b = \log K + 3\sigma\sqrt{T}$ then the put option with strike K and the initial spot level $S = e^b$ is far out of the money so we can take its BS value to be 0. We conclude that in our model the price of a put with the spot level $S = e^b$ is also negligible. From the put call parity, we obtain the boundary condition at $x = b$:

$$u(b,t) \;=\; e^b - Ke^{-r(T-t)} \quad \text{for } 0 \leq t \leq T \quad \text{if } b = \log K + 3\sigma\sqrt{T}. \tag{40}$$

We can now estimate the value of the call option by solving (38) for $a < x < b,\ t < T$, with terminal data (39) and Dirichlet boundary conditions $u(a,t) = 0,\ t < T$, and $u(b,t),\ t < T$, given by (40). To construct a numerical scheme we need to approximate the first two derivatives of $u(x,t)$ with respect to x by finite differences as well as $\partial u(x,t)/\partial t$. We shall use the central difference (14) to approximate the first derivative with respect to x so we set

$$\frac{\partial u(x,t)}{\partial x} \;\sim\; \frac{u(x+\Delta x, t) - u(x - \Delta x, t)}{2\Delta x}\,. \tag{41}$$

We have already observed that the central difference approximation is second order accurate, so the error in (41) is $O[(\Delta x)^2]$. We can also easily obtain a second order accurate approximation to the second derivative, which is

$$\frac{\partial^2 u(x,t)}{\partial x^2} \;\sim\; \frac{u(x+\Delta x, t) + u(x - \Delta x, t) - 2u(x,t)}{(\Delta x)^2}\,. \tag{42}$$

We can see why (42) is second order accurate by doing a Taylor expansion as in (15) but with more terms,

$$u(x+\Delta x) = u(x) + u'(x)[\Delta x] + u''(x)[\Delta x]^2/2 + u'''(x)[\Delta x]^3/6 + u''''(x)[\Delta x]^4/24 + O[(\Delta x)^5]. \tag{43}$$

Hence we have that

$$u(x+\Delta x) + u(x - \Delta x) - 2u(x) = (\Delta x)^2\{u''(x) + u''''(x)(\Delta x)^2/12 + O[(\Delta x)^3]\}\,, \tag{44}$$

and so the difference between the LHS and RHS of (42) is $O[(\Delta x)^2]$. The finite difference approximation for the PDE (38) is completed now by taking the backward difference (13) as an approximation to the time derivative

$$\frac{\partial u(x,t)}{\partial t} \sim \frac{u(x,t) - u(x,t-\Delta t)}{\Delta t} . \tag{45}$$

The explicit backwards in time Euler method for the terminal value problem (38)–(39) is then

$$u(x,t-\Delta t) = (p - r\Delta t)u(x,t) + p^+ u(x+\Delta x,t) + p^- u(x-\Delta x,t) , \tag{46}$$

where p, p^+, p^- are given by the formulas

$$p = 1 - \sigma^2(e^x,t)\frac{\Delta t}{(\Delta x)^2}, \quad p^+ = \sigma^2(e^x,t)\frac{\Delta t}{2(\Delta x)^2} + (r - \sigma^2(e^x,t)/2)\frac{\Delta t}{2\Delta x} ,$$
$$p^- = \sigma^2(e^x,t)\frac{\Delta t}{2(\Delta x)^2} - (r - \sigma^2(e^x,t)/2)\frac{\Delta t}{2\Delta x} . \tag{47}$$

The algorithm.

- To implement the explicit Euler method we choose Δx and Δt so that
  - $M\Delta t = T$ for some integer M
  - and $N\Delta x = b - a$ for some integer N.
- We set $u_n^m \approx u(a + n\Delta x, m\Delta t)$ for $n = 0, .., N$, and $m = 0, .., M$.
- The terminal condition (39) then yields

$$u_n^M = (e^{a+n\Delta x} - K)^+ \quad \text{for } n = 0, ..N. \tag{48}$$

- The boundary conditions $u(a,t) = 0,\ 0 \le t \le T$, and (40) yield

$$u_0^m = 0, \quad u_N^m = e^b - Ke^{-r(M-m)\Delta t} \quad \text{for } m = 0, .., M. \tag{49}$$

- The recurrence (46) is then given by

$$\begin{array}{rcc} \text{for } m & = & M, \ldots, 1 \\ \text{for } n & = & 1, \ldots, N-1 \\ u_n^{m-1} & = & (p_n^m - r\Delta t)u_n^m + (p^+)_n^m u_{n+1}^m + (p^-)_n^m u_{n-1}^m, \end{array} \tag{50}$$

where p_n^m, $(p^+)_n^m$ and $(p^-)_n^m$ are given by (47), with $x = a + n\Delta x$ and $t = m\Delta t$.

Stability

We already discussed (at the end of Section 3) the issue of stability when using the explicit Euler method to numerically solve parabolic PDE such as (34). There we observed that it is necessary to take $\Delta t/(\Delta x)^2 = O(1)$ to ensure stability. We can see more clearly here why this is the case, using the fact that the parameters p, p^+, p^- of (47) satisfy $p + p^+ + p^- = 1$.

If all three parameters are positive then it follows from (46) that

$$\sup_{a<x<b} |u(x,t-\Delta t)| \ \leq |p - r\Delta t| \sup_{a<x<b} |u(x,t)| + |p^+| \sup_{a<x<b} |u(x,t)| + |p^-| \sup_{a<x<b} |u(x,t)| \tag{51}$$

$$\leq (|p - r\Delta t| + |p^+| + |p^-|) \sup_{a\leq x\leq b} |u(x,t)| = \ (1 - r\Delta t) \sup_{a\leq x\leq b} |u(x,t)|,$$

where the last equality holds because, for all small enough Δt, we have: $p - r\Delta t > 0$, $p^+ > 0$, $p^- > 0$, and hence

$$|p - r\Delta t| + |p^+| + |p^-| = p - r\Delta t + p^+ + p^- = 1 - r\Delta t$$

Since $M\Delta t = T$, we conclude that

$$\sup_{a<x<b} |u(x,0)| \ \leq \ [1 - r\Delta t]^M \sup_{a\leq x\leq b} |u(x,T)| \leq e^{-rT} \sup_{a\leq x\leq b} |u(x,T)| \,, \tag{52}$$

where we ignore the effect of the boundary conditions.

If one of the parameters p, p^+, p^- is negative then, for small enough Δt, $|p - r\Delta t| + |p^+| + |p^-| \geq 1 + \delta$ for some $\delta > 0$. In that case we could have

$$\sup_{a<x<b} |u(x,0)| \ \simeq (1+\delta)^M \ = \ (1+\delta)^{T/\Delta t} \,, \tag{53}$$

which diverges as $\Delta t \to 0$. In fact, this instability typically does occur if $|p| + |p^+| + |p^-| = 1 + \delta$ with $\delta > 0$. The numerical approximation not only **grows exponentially** in $T/\Delta t$ but also **oscillates**.

Evidently a **necessary and sufficient condition for** $p, p^+, p^- > 0$ is that

$$\frac{\Delta t}{(\Delta x)^2} \sup_{a<x<b,\, t\in(0,T)} \sigma^2(e^x, t) \ < \ 1 \,. \tag{54}$$

Connection to lattice (tree) methods

Notice that the **explicit Euler scheme** for BSPDE, described above, **resembles the pricing algorithm in the trinomial tree models**. For simplicity, let's assume that $r = 0$. Then (50) can be viewed as the recursive relation in a trinomial tree model, in which

- the risky asset, at each time $t_m = m\Delta t$, jumps from the level $e^{x_n} = e^{a+n\Delta x}$
  - either up to level $e^{x_{n+1}} = e^{a+(n+1)\Delta x}$,
  - or down to level $e^{x_{n-1}} = e^{a+(n-1)\Delta x}$,
  - or to the same level $e^{x_n} = e^{a+n\Delta x}$;
- and the "risk-neutral" probabilities are given by
  - $q_u(m,n) = p^+ = \sigma^2(e^{x_n}, t_m)\frac{\Delta t}{2(\Delta x)^2} - \sigma^2(e^{x_n}, t_m)\frac{\Delta t}{4\Delta x}$,
  - $q_d(m,n) = p^- = \sigma^2(e^{x_n}, t_m)\frac{\Delta t}{2(\Delta x)^2} + \sigma^2(e^{x_n}, t_m)\frac{\Delta t}{4\Delta x}$,
  - $1 - q_u(m,n) - q_d(m,n) = p = 1 - \sigma^2(e^{x_n}, t_m)\frac{\Delta t}{(\Delta x)^2}$;

Notice that the above quantities are, indeed, probabilities. However, they are **NOT risk-neutral**:

$$q_u(m,n)e^{\Delta x} + q_d(m,n)e^{-\Delta x} + 1 - q_u(m,n) - q_d(m,n)$$

$$= (e^{\Delta x} - 1)(\sigma^2(e^{x_n}, t_m)\frac{\Delta t}{2(\Delta x)^2} - \sigma^2(e^{x_n}, t_m)\frac{\Delta t}{4\Delta x})$$

$$+(e^{-\Delta x} - 1)(\sigma^2(e^{x_n}, t_m)\frac{\Delta t}{2(\Delta x)^2} + \sigma^2(e^{x_n}, t_m)\frac{\Delta t}{4\Delta x}) + 1 \neq 1$$

Nevertheless, when $\Delta t \to 0$, with $\Delta t/\Delta x^2 = const$, we obtain

$$q_u(m,n)e^{\Delta x} + q_d(m,n)e^{-\Delta x} + 1 - q_u(m,n) - q_d(m,n) = 1 + O(\Delta t^{3/2})$$

Over $T/\Delta t$ steps, the cumulative error (i.e. the maximum relative difference between the expected value of S_T, starting at an arbitrary time $t \in [0,T]$, and the desired value S_t) is

$$O(\Delta t^{3/2}) \cdot T/\Delta t = O(\Delta t^{1/2}) \to 0,$$

as $\Delta t \to 0$. Thus, the probabilities (q_u, q_d) are "asymptotically risk neutral".

Accuracy and interpolation away from grid points

Now the Euler method (46) is first order accurate in Δt and second order accurate in Δx. Hence the cumulative error is given by

$$u(a + n\Delta x, 0) - u_n^0 \;=\; O[\Delta t] + O((\Delta x)^2) \quad \text{for } 0 \leq n \leq N. \tag{55}$$

It follows from (55) that for given Δx we should take Δt as large as possible consistent with (54) to minimize the number of computations without substantially increasing the cumulative error. Thus to implement explicit Euler in an optimal way we should choose Δx so that $O((\Delta x)^2)$ is an acceptable error. Then we choose Δt such that there is almost equality in (54).

If S_0 is today's stock price then the BS value of the call option is $u(x_0, 0)$ where $S_0 = e^{x_0}$. It may happen that x_0 is not a grid point for the numerical method, so for example there is a non-negative integer $n_0 < N$ such that $a + n_0\Delta x = x_0^- < x_0 < a + (n_0+1)\Delta x = x_0^+$. In that case we can approximate the value of $u(x_0, 0)$ by **interpolation** from the numerically computed values of the function on the grid points. If we use **linear interpolation** then we set

$$u(x_0, 0) \;\simeq\; \{[x_0^+ - x_0]u(x_0^-, 0) + [x_0 - x_0^-]u(x_0^+, 0)\}/\Delta x\,. \tag{56}$$

Linear interpolation has the advantage that it is completely **local** in the sense that the value taken for $u(x_0, 0)$ is just a weighted average of the values of the function $u(\cdot, 0)$ on the two grid points closest to x_0.

There is also a disadvantage in that it is only a **first order accurate** interpolation. This means that if $u(x, 0),\ a < x < b$, is a differentiable function then the difference between the LHS and RHS of (56) is $O(\Delta x)$. In our situation $u(x, 0)$ is computed at grid points x correct to second order, so the error at a grid point between the numerically computed solution and the solution to the continuous problem is $O((\Delta x)^2)$. Evidently we wish to use an interpolation scheme which preserves the second order accuracy away from grid points. Linear interpolation cannot achieve this but **spline interpolation** will. It does this by approximating the function between grid points by a polynomial -say of degree 3 as in the case of **cubic splines**. Because approximation by polynomials give some extra degrees of freedom we can make the approximating function differentiable across grid points. Note that in the case of linear interpolation there are no extra degrees of freedom and so in general the approximating function has a jump in its derivative across a grid point.

There is a cost to this extra accuracy of spline interpolation in that spline interpolation is **non-local** in the sense that the interpolated value for $u(x_0, 0)$ depends now on the values of $u(\cdot, 0)$ at **all** the grid points, not just at the two neighboring grid points of x_0. The spline interpolation is however **almost local** in the sense that the dependence of $u(x_0, 0)$ on a grid point far from x_0 is very small.

Nevertheless, the biggest problem with the interpolation is that it may increase (or produce) arbitrage opportunities caused by the approximation. As discussed above, the numerical approximation of option's price $V(S_0, 0)$, obtained via the Explicit Euler scheme, does not correspond to an arbitrage-free model (although it is asymptotically arbitrage-free), hence, it may lead to arbitrage in the market. However, choosing the discretization parameters $(\Delta x, \Delta t)$, we can make this price be very close to the LV price, which is arbitrage-free. Due to the presence of transaction costs in the market, we conclude that, in reality, our approximation becomes arbitrage-free, for small enough $(\Delta x, \Delta t)$. Thus, if one is pricing a single option written on S, the interpolation will improve the performance of the method. However, if, somehow, one decides to price several options written on S, using a solution to the same terminal value problem (it does not make sense in the present context but it will be relevant later, when we discuss the Dupire's equation), then, the interpolation may violate the static no-arbitrage conditions. In such cases, one has to be very careful with the polynomial interpolation (and, typically, we will not use it in such context).

4.2 American Options

Recall that American options have the early exercise feature, which means that one can exercise the option at any time up to the expiration date. We shall consider here the simple American put option with strike price K and expiration date T.

Assuming a BS model with volatility σ and interest rate r, we let $V(S, t)$ be the value function of the American put. In particular,

$$V(S, T) \;=\; (K - S)^+ \quad \text{for } S > 0. \tag{57}$$

Since the **value of the option cannot drop below the early exercise price** we also have that

$$V(S, t) \;\geq\; (K - S)^+ \quad \text{for } S > 0,\; t \leq T. \tag{58}$$

It is also clear that, **at the exercise time τ, the value of the option must coincide with the early exercise price** (if the value is above the early exercise price, why would the holder exercise the option?). Therefore,

$$V_{\tau \wedge T} = V(S_{\tau \wedge T}, \tau \wedge T) = (K - S_{\tau \wedge T})$$

Notice that, before the option is exercised, i.e. for $t < \tau \wedge T$, it is a tradable asset in the market. Therefore, FTAP 1 implies that (V_t/B_t) **is a $\mathbb{Q}$-mtg for** $\in [0, \tau \wedge T)$.

Thus, we can **repeat the derivation of the BSPDE** for the American option, obtaining

$$\frac{\partial}{\partial t}V + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2}{\partial S^2}V + rS\frac{\partial}{\partial S}V - rV = 0, \tag{59}$$

which, now, has to hold only **for all** (S, t) **such that** $V(S, t) > (K - S)^+$. The terminal condition remains the same: $V(S, T) = (K - S)^+$.

Evidently the value of the American option is at least the value of the corresponding European option, but it generally has only slightly greater value. In fact, we have already shown by a simple no arbitrage argument that, **if $r = 0$, then the value of the American put equals the value of the European put**.

For $r > 0$, it can be shown that there exists a curve $t \to S_{\text{exer}}(t),\ 0 \le t \le T$, with the property that $0 < S_{\text{exer}}(t) \le K$ for $0 \le t \le T$, such that

$$\begin{cases} V(S,t) \text{ satisfies (59) for } S > S_{\text{exer}}(t), \\ V(S,t) = (K-S)^+ \text{ for } S < S_{\text{exer}}(t), \\ V(S,T) = (K-S)^+ \end{cases} \tag{60}$$

The graph $\{(t,S) : S = S_{\text{exer}}(t),\ 0 \le t \le T\}$ is known as the **early exercise boundary** since it divides the half infinite region $\{(t,S) : 0 \le t \le T, S > 0\}$ into two subregions. In the lower subregion it is optimal always to immediately exercise the option and in the upper subregion to wait until either the expiration date of the option or until the stock price hits the early exercise boundary.

Typically, $S_{\text{exer}}(t)$ is an **increasing and convex** function of t with $\lim_{t\to T} S_{\text{exer}}(t) = K$. It is clear that $S_{\text{exer}}(t) < K$, becaue it does not make sense to exercise an out-of-the money option. It is also intuitively clear that $S_{\text{exer}}(t)$ is increasing and $\lim_{t\to T} S_{\text{exer}}(t) = K$, but the convexity of the graph is a subtle property. In fact even the increasing property cannot be proved in a straightforward way although the following intuitive argument provides a partial explanation. Notice that, if the option is in-the-money at time t and we expect S_T to stay around the same level as S_t, then, we should exercise the option immediately – otherwise we loose money due to discounting. When maturity approaches, clearly, the expected deviation of S_T from its current level S_t becomes smaller and smaller, hence, the holder becomes eager to exercise the option sooner, which, in turn, implies that the exercise boundary increases.

We consider some properties of the value $V(S,t)$ of the option on the early exercise boundary. In particular we have that

$$V(S,t) \ = \ K - S, \quad \frac{\partial V(S,t)}{\partial S} \ = \ -1, \qquad \text{for } S = S_{\text{exer}}(t)\ . \tag{61}$$

The first identity of (61) states that the **value of the option is continuous** across the early exercise boundary, while the second identity states that the **hedging ratio is continuous** across the boundary. The first identity is then just a consequence of the fact that the value of the option is a continuous function of (S,t). We can give an intuitive argument for the second identity based on the **no arbitrage** argument.

First assume that

$$\lim_{S \to S_{\text{exer}}(t)^+} \frac{\partial V(S,t)}{\partial S} \ < \ -1\ . \tag{62}$$

Then, for $S > S_{\text{exer}}(t)$, with $S - S_{\text{exer}}(t)$ small enough, we have: $V(S,t) < K - S$, which contradicts the inequality (58).

Alternatively let us suppose that

$$\lim_{S \to S_{\text{exer}}(t)^+} \frac{\partial V(S,t)}{\partial S} \ > \ -1\ . \tag{63}$$

Assume that the BS theory applies and we can hedge the option perfectly by trading in S and B, according to the delta-hedging rule (one can prove that this is the case, but we will skip this part). Then, for $S_t = S_{\text{exer}}(t)$ we can buy an option and construct a hedging portfolio, whose total price is zero:

$$V_t - \partial_S V S_t + (\partial_S V S_t - V_t) = 0. \tag{64}$$

Consider now the change in this portfolio over a small interval Δt, keeping the hedging ratio $\alpha = \partial_S V$ fixed. First, assume that $S_{t+\Delta t} \le S_t$, and notice that, in this case $V_t = K - S_t$, $V_{t+\Delta t} = K - S_{t+\Delta t}$ and the change of the above portfolio is:

$$(S_t - S_{t+\Delta t})(1+\alpha) + (\alpha S_t - V_t)(e^{r\Delta t} - 1) \approx (1+\alpha)\Delta S_t + (\cdots)\Delta t. \tag{65}$$

Notice that

$$|\Delta S_t| \approx rS_t|\Delta t| + \sigma S_t|\Delta W_t| \approx rS_t|\Delta t| + \sigma S_t\sqrt{|\Delta t|},$$

which has a **lower order of decay** than Δt, as $\Delta t \to 0$. Due to this observation and the assumption that $\alpha > -1$, we conclude that the right hand side of (65) is positive, for small enough Δt. Hence, the hedging portfolio generates a profit in this case.

When $S_{t+\Delta t} > S_t$, then the option is not exercised and we continue the BS delta-hedging, which produces a zero overall return. We conclude that if (63) holds then the portfolio (64) is an arbitrage since it can make a profit (with positive probability) but cannot make a loss. This is a contradiction to the no arbitrage assumption. We have therefore shown that neither of (62), (63) can hold, which implies the second identity of (61).

The mathematical analysis of (58), (59) is complicated since the early exercise boundary is a so called **free boundary**. This means it is not known **a priori** but comes as part of the solution to the problem. It is then somewhat surprising that the numerical algorithm for solving the American option problem is virtually the same as the algorithm for solving the European option problem. In fact all we do is to modify the European algorithm to insure that (58) holds. Thus we have instead of (46),

$$u(x, t-\Delta t) \;=\; \max\{K - e^x,\; (p - r\Delta t)u(x,t) + p^+u(x+\Delta x, t) + p^-u(x-\Delta x, t)\}\,, \tag{66}$$

where again $V(S,t) = u(x,t)$ with $S = e^x$, and (p, p^+, p^-) are as for the European option. The terminal and boundary conditions are given by

$$u(x,T) \;=\; (K - e^x)^+, \quad a < x < b, \tag{67}$$

$$u(a,t) \;=\; K - e^a, \quad u(b,t) \;= 0, \qquad 0 \le t < T, \tag{68}$$

where a, b are as for the European option. The boundary condition (68) at a comes from the fact that a lies below the early exercise boundary for all $0 \le t \le T$. Note that cash K is **not discounted** in the boundary condition (68), whereas in the European case (40) it is discounted.

Finally, once we have obtained the function $V(S,t),\; S > 0, 0 \le t \le T$, we approximate the early exercise boundary via

$$S^m_{\text{exer}} = \min\left\{e^{a+n\Delta x} \,:\, n = 1, \ldots, N,\, u^m_n > K - e^{a+n\Delta x}\right\}, \quad m = 0, \ldots, M.$$

4.3 Asian options

The payoff of an Asian option depends on the average of the underlying price S during the lifetime of the option. For the continuously sampled Asian call, with floating strike, the payoff at expiration is given by the formula

$$V_T = \left[S_T - \frac{1}{T}\int_0^T S_t\, dt\right]^+. \tag{69}$$

We can price this Asian option in a BS model by introducing a new variable I, in addition to S, t,

$$I_t \;=\; \int_0^t S_u\, du\,, \tag{70}$$

and consider the value of the option as being a **function** $V(S, I, t)$ **of** 3 **variables** S, I, t. The value of the option today, at $t = 0$, when the stock price is S_0, is then $V(S_0, 0, 0)$. Notice that, under the risk neutral measure, (S_t, I_t) solve the system of SDEs

$$dS_t \;=\; S_t(r dt + \sigma dB_t), \qquad dI_t \;=\; S_t\, dt\,. \tag{71}$$

Rem 16. *Notice that, now, we are in the setting of a general diffusion model, described in Subsection 2.2, with* $X_t = (S_t, I_t)^T$.

To price the option we derive the BSPDE for $V(S, I, t)$ by computing the drift of V/B under the risk neutral measure, and setting it to zero. As a result, we obtain:

$$\frac{\partial V(S,I,t)}{\partial t} + \frac{\sigma^2}{2} S^2 \frac{\partial^2 V(S,I,t)}{\partial S^2} + S \frac{\partial V(S,I,t)}{\partial I} + rS \frac{\partial V(S,I,t)}{\partial S} - rV(S,I,t) \; = 0 \, . \tag{72}$$

The equation (72) holds for $S, I > 0$ and $t < T$. From (69) we see that the terminal condition for the problem is

$$V(S,I,T) \; = \; \max[S - I/T, 0] \quad \text{for } S, I > 0. \tag{73}$$

Just as for (36), (37) we also need boundary conditions to uniquely determine the solution. One boundary condition is obvious, namely that

$$\lim_{S \to 0} V(S,I,t) \; = \; 0, \tag{74}$$

since this merely states that if the price of the stock is small then not much can be gained by betting on the difference between average stock price and the terminal stock price. We also need to impose three other boundary conditions corresponding to $S \to \infty$ and also $I \to 0, \; I \to \infty$. However, it is **not so easy** to see what these boundary conditions should be. In fact, there are few additional difficulties we will have to face if we choose to analyze the PDE (72) directly.

- **Curse of dimensionality**: it is much more computationally expensive to solve higher dimensional PDEs numerically. In the present case, we have an extra space variable, I, which will increase the number of computations and the use of memory by a factor of $N = I_{max}/\Delta I$. Generally, speaking, if the number of variables in a PDE is more than 2, one should be looking for different methods (the complexity of finite difference methods grows exponentially with the dimension).
- The PDE (72) is **degenerate** in I: the second order derivative with respect to I is missing. This creates problems on the theoretical level (with existence, uniqueness and regularity of the solution) and, potentially, may cause problems with the convergence of a numerical scheme.

Due to above observations, we will not analyze (72) directly. Instead, we will **reduce the number of variables** (S, I, t) **to** 2 **variables** by observing that the function $V(S, I, t)$ is homogeneous. To see this let $\lambda > 0$ and consider the function $V_\lambda(S, I, t) = \lambda^{-1} V(\lambda S, \lambda I, t)$. It is easy to see that $V_\lambda(S, I, t)$ satisfies (72),(73) and so we conclude from general uniqueness considerations for the terminal value problem (72),(73) that V_λ is independent of $\lambda > 0$. Choosing $\lambda = S^{-1}$ we conclude that

$$V(S,I,t) \; = \; Sw(\xi, t), \quad \xi = I/S \, , \qquad \text{for } I, S > 0, \; t < T, \tag{75}$$

where $w(\xi, t)$ is a function of two variables and

$$w(\xi, T) \; = \; \max[1 - \xi/T, 0] \quad \xi > 0. \tag{76}$$

We can obtain from (72) a PDE which the function $w(\xi, t)$ satisfies. Using the chain rule we have that

$$\frac{\partial V(S,I,t)}{\partial t} \; = \; S \frac{\partial w(\xi,t)}{\partial t} \, , \quad \frac{\partial V(S,I,t)}{\partial S} \; = \; w(\xi,t) - \xi \frac{\partial w(\xi,t)}{\partial \xi} \, ,$$
$$\frac{\partial V(S,I,t)}{\partial I} \; = \; \frac{\partial w(\xi,t)}{\partial \xi} \, , \quad S \frac{\partial^2 V(S,I,t)}{\partial S^2} \; = \; \xi^2 \frac{\partial^2 w(\xi,t)}{\partial \xi^2} \, . \tag{77}$$

Hence (72) implies that $w(\xi, t)$ is a solution to the PDE

$$\frac{\partial w(\xi,t)}{\partial t} + \frac{1}{2}\sigma^2\xi^2\frac{\partial^2 w(\xi,t)}{\partial \xi^2} + (1 - r\xi)\frac{\partial w(\xi,t)}{\partial \xi} \;=\; 0\,. \tag{78}$$

The above method looks somewhat random: it is not clear a priori why we managed to find the right change of variables to reduce the dimensionality of the problem. In fact, there is a deeper principle behind the above transformation, which is called the **change of numeraire**.

Change of numeraire.

Notice that the FTAP does not specify which of the traded assets must be used for discounting (i.e. which asset must be used as the **numeraire**). In particular, we can use the risky asset S itself: this corresponds to measuring the price of the option in the units of stock price. Then, FTAP implies that, under a risk neutral measure (which depends on the choice of numeraire), V/S is a martingale. Therefore,

$$V_t/S_t = \mathbb{E}^{\mathbb{Q}^S}(V_T/S_T) = \mathbb{E}^{\mathbb{Q}^S}(1 - \frac{1}{T}(I_T/S_T))^+.$$

This implies that the option price can be computed as

$$V_t = S_t\mathbb{E}^{\mathbb{Q}^S}(1 - \frac{1}{T}\xi_T)^+.$$

If, under $\mathbb{Q}^S$, (ξ_t) happens to be a Markov process, then V_t can be written as

$$V_t = V(S_t, I_t, t) = S_T w(\xi_t, t),$$

with

$$w(\xi, t) = \mathbb{E}_t^{\mathbb{Q}^S}((1 - \frac{1}{T}\xi_T)^+ \,|\, \xi_t = \xi).$$

It turns out that, in the BS model, (ξ_t) is, indeed, **Markov** under $\mathbb{Q}^S$. This is why the above transformation of the BSPDE allowed us to cancel the S variable.

The boundary condition (74) turns into the boundary condition

$$\lim_{\xi\to\infty} w(\xi, t) \;=\; 0, \quad \text{for } t < T. \tag{79}$$

Now (74) and (79) are not equivalent since (74) actually only implies that $\lim_{\xi\to\infty} w(\xi, t)/\xi = 0$, so (79) is a stronger assumption than (74). The assumption (79) however makes sense if we think of $w(\xi, t)$ as being the value of a type of put option. Comparing (78), (76) to (36), (57) respectively, we see that $w(\xi, t)$ is like the value of a put option with strike price T. With that analogy then (79) states that the value of the option is roughly zero if it is far out of the money.

Rem 17. *Of course, this is not a precise analogy, as the PDE (78) does not actually correspond to a pricing PDE in any arbitrage-free model (e.g. the first order term is wrong). However, we will clarify this argument below, by making a* **connection between the solutions to parabolic PDEs and expectations**.

In order to have a unique solution we need to specify a boundary condition for $w(\xi, t)$ as $\xi \to 0$. Note that the value of the Asian option today is $V(S_0, 0, 0) = S_0 w(0, 0)$ so the boundary condition at $\xi = 0$ **cannot be explicit** as

in the case of the standard European put option. We can obtain an **implicit** boundary condition by setting $\xi = 0$ in the PDE (78). Thus we have that

$$\frac{\partial w(\xi,t)}{\partial t} + \frac{\partial w(\xi,t)}{\partial \xi} \;=\; 0 \quad \text{for } \xi = 0,\ t < T. \tag{80}$$

We can understand the conditions (79) and (80) better if we write the function $w(\xi, t)$ as an expectation.

Let us now recall the **Feynman-Kac formula**, which justifies the derivation of the above boundary conditions.

Feynman-Kac formula.

Consider the following $\xi(t),\ t > 0$ as a solution to the following SDE

$$d\xi_t \;=\; (1 - r\xi_t)\, dt + \sigma \xi_t dB_t\ , \tag{81}$$

which is associated with (78).

Thm 5. *Assume that w solves (78) and satisfies certain regularity conditions. Then, we have:*

$$w(\xi,t) \;=\; E\left(\, w(\xi_T, T) \,\middle|\, \xi_t = \xi \,\right) \qquad \text{for } \xi > 0,\ t < T. \tag{82}$$

Proof:

Applying the Ito's lemma, we obtain:

$$dw(\xi_t,t) = \left(\frac{\partial w(\xi,t)}{\partial t} + \frac{1}{2}\sigma^2 \xi^2 \frac{\partial^2 w(\xi,t)}{\partial \xi^2} + (1 - r\xi)\frac{\partial w(\xi,t)}{\partial \xi} \right) dt + (\cdots) dW_t = (\cdots) dW_t$$

Thus, we conclude (under certain technical conditions) that $(w(\xi_t, t))_{t\in[0,T]}$ is a mtg. This implies (82). ∎

The Feynman-Kac formula justifies our interpretation of the solution of (78) as an expectation, even if it is not a price of anything. Thus, it provides a justification of the boundary condition (79).

We can also obtain an alternative justification of the second boundary condition, (80). Notice that

$$w(\xi,t) \;=\; E\left[\, w(\xi_{t+\Delta t}, t + \Delta t) \,\middle|\, \xi_t = \xi \,\right] \qquad \text{for all } \xi > 0,\ \Delta t > 0, \tag{83}$$

which follows from (82) upon setting $\Delta t = T - t$. Now, if ξ is small and Δt is also small the SDE (81) becomes approximately

$$d\xi(s) \approx ds, \ \text{ for } t < s < t + \Delta t,$$

and

$$dw(\xi_s, s) \approx (\frac{\partial w(\xi_s, s))}{\partial t} + \frac{\partial w(\xi_s, s)}{\partial})ds + (\cdots)dW_s.$$

Then, the martingale (i.e. zero-drift) condition for $(w(\xi_s, s))_{s\in[t,t+\Delta t]}$ implies (80).

Rem 18. *Note that the process (ξ_t), with $\xi_0 > 0$, never hits zero. The reason is that the drift term $1 - r\xi$ in (81) pulls a path $\xi(t)$ away from 0, while the diffusion coefficient $\sigma\xi_t$ decays sufficiently fast as $\xi \to 0$, so the solution to (81) never actually hits 0.*

Explicit Euler scheme for Asian option

Notice that the coefficients in (78) are not bounded. However, the usual logarithmic change of variables $x = \log S$ will not help, as the first order term will still be unbounded. That is why we will work with (78) directly.

To numerically solve the terminal value problem (76), (78), with the boundary conditions (79), (80), we need to first replace the infinite interval $\{0 < \xi < \infty\}$ by a finite interval $\{0 < \xi < \xi_{\max}\}$. We choose $\xi_{\max}$ depending on volatility similarly to the way we did for pricing the European option. We already know that $w(\xi, t)$ is given by an expectation of a function of ξ_T. This function is nonzero only in the interval $[0, T]$. Thus we wish to choose $\xi_{\max}$ sufficiently large so that

$$\mathbb{Q}(\xi_T \leq T \,|\, \xi_0 = \xi_{max}) \approx 0$$

Then, we may set $w(\xi_{\max}, t) = 0$ for $0 \leq t < T$.

Notice, however, that we cannot compute the above probability, as we do not know the distribution of the process ξ explicitly. Nevertheless, we can consider a similar process

$$d\tilde{\xi}_t = -r\tilde{\xi}_t dt + \sigma\tilde{\xi}_t dW_t$$

As the drift of $\tilde{\xi}$ is smaller than the drift of ξ, while the diffusion coefficient is the same (as a function of the process), it is clear intuitively that

$$\mathbb{Q}(\xi_t \text{ hits } T \text{ for some } t \in [0, T] \,|\, \xi_0 = \xi) \leq \mathbb{Q}(\tilde{\xi}_t \text{ hits } T \text{ for some } t \in [0, T] \,|\, \tilde{\xi}_0 = \xi) \tag{84}$$

The advantage of the above estimate is that $\tilde{\xi}$ is GBM, and we know its distribution. In particular, we can choose $\xi_{\max}$ as before:

$$\xi_{\max} = Te^{3\sigma\sqrt{T}} \,. \tag{85}$$

Then, the right hand side of (84) is sufficiently small, which makes the desired probability be small as well.

Now, we are ready to construct the explicit Euler scheme for (76), (78). Similarly to (46), we can use

- the **backward difference** approximation for $\partial_t w$;
- the **central difference** approximation for $\partial_\xi w$;
- and the **central difference** approximation for $\partial^2_{\xi\xi} w$;

to obtain the recurrence equation

$$w(\xi, t - \Delta t) \;=\; p(\xi)w(\xi, t) + p^+(\xi)w(\xi + \Delta\xi, t) + p^-(\xi)w(\xi - \Delta\xi, t) \,, \tag{86}$$

where $p(\xi), p^+(\xi), p^-(\xi)$ are given by the formulas

$$p(\xi) = 1 - \sigma^2\xi^2\Delta t/(\Delta\xi)^2, \quad p^+(\xi) = \sigma^2\xi^2\Delta t/2(\Delta\xi)^2 + (1 - r\xi)\Delta t/2\Delta\xi \,,$$
$$p^-(\xi) = \sigma^2\xi^2\Delta t/2(\Delta\xi)^2 - (1 - r\xi)\Delta t/2\Delta\xi \,. \tag{87}$$

Note that $p(\xi) + p^+(\xi) + p^-(\xi) = 1$ for $0 < \xi < \xi_{\max}$.

To implement the boundary condition (80) we approximate the derivative $\partial w(\xi, t)/\partial\xi$ at $\xi = 0$ by the **forward difference**, so (80) becomes

$$\frac{w(0, t) - w(0, t - \Delta t)}{\Delta t} + \frac{w(\Delta\xi, t) - w(0, t)}{\Delta\xi} \;=\; 0, \tag{88}$$

which implies that

$$w(0, t - \Delta t) \;=\; [1 - \Delta t/\Delta\xi]w(0, t) + (\Delta t/\Delta\xi)w(\Delta\xi, t) \,. \tag{89}$$

Similarly to (54) we need to impose the stability condition $p(\xi) > 0$ for $0 < \xi < \xi_{\max}$, which implies

$$\Delta t/(\Delta\xi)^2 \ < \ 1/\sigma^2\xi_{\max}^2 \, . \tag{90}$$

Recall that in (86), (87) we used the **central difference** to approximate $\partial_\xi w$. The reason for doing this was to preserve the **second order accuracy** of the numerical scheme in the space difference. In contrast to this we are forced to use the forward difference to implement the boundary condition (80) as in (89). Since forward difference is just first order accurate, this will tend to make the scheme only first order accurate in the space difference, so the error would be $O(\Delta\xi)$ and not $O[(\Delta\xi)^2]$.

In addition, the scheme (86), (87) is **not stable for all** ξ in the interval $[0, \xi_{\max}]$ even when (90) is satisfied! We can see that $p^-(\xi) < 0$ if $\xi < \sqrt{\Delta\xi}/\sigma$, and so we have in fact a **boundary layer** at $\xi = 0$ of width $\sqrt{\Delta\xi}/\sigma$. This is analogous to boundary layers in fluid flows where the flow can change radically over a small length scale, such as in the case of air close to the wing of an aircraft.

We can have fix this problem and obtain a numerical scheme which is stable for all ξ, in the interval $[0, \xi_{\max}]$, by simply approximating the derivative $\partial w(\xi, t)/\partial\xi$ in (78) with a **forward difference**

$$\frac{\partial w(\xi, t)}{\partial \xi} \ \sim \ \frac{w(\xi + \Delta\xi, t) - w(\xi, t)}{\Delta\xi} \, . \tag{91}$$

In that case the numerical scheme becomes (86) with $p(\xi), p^+(\xi), p^-(\xi)$ given by the formulas

$$p(\xi) = 1 - \sigma^2\xi^2\Delta t/(\Delta\xi)^2 - (1 - r\xi)\Delta t/\Delta\xi, \quad p^+(\xi) = \sigma^2\xi^2\Delta t/2(\Delta\xi)^2 + (1 - r\xi)\Delta t/\Delta\xi \, ,$$
$$p^-(\xi) = \sigma^2\xi^2\Delta t/2(\Delta\xi)^2 \, . \tag{92}$$

We still have $p(\xi) + p^+(\xi) + p^-(\xi) = 1$ but now all of $p(\xi), p^+(\xi), p^-(\xi)$ remain positive for $0 < \xi < \xi_{\max}$ provided we are away from the threshold in (90).

The error in the scheme (86), (92) is now $O(\Delta t) + O(\Delta\xi) = O[(\Delta\xi)^2] + O(\Delta\xi) = O(\Delta\xi)$, so we have **lost the second order accuracy** in order to maintain stability.

5 The Crank-Nicolson Algorithm

We already introduced the implicit second order accurate methods for solving ODEs. Here, we discuss a specific implicit method for solving linear parabolic PDEs, known as the Crank-Nicolson (CN) method. We shall use the CN method here to obtain an algorithm for pricing a **barrier** option in a **BS model**.

Recall that the value of a **down-and-out call (DOC)** option becomes zero if the price of the underlying S falls below the barrier U before the expiry time T. Otherwise, at time T, the option pays as a call with strike K.

Assuming $U = e^a$, the value of DOC satisfies the BSPDE (36), (37) in the region $S > e^a, \ t < T$, with the terminal condition $V(S, T) = (S - K)^+$ and the boundary condition $V(e^a, t) = 0, \ t < T$.

Defining $u(x, t) = V(e^x, t)$, we see that we need to solve the PDE

$$\frac{\partial u(x,t)}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 u(x,t)}{\partial x^2} + \left\{ r - \frac{\sigma^2}{2} \right\} \frac{\partial u(x,t)}{\partial x} - ru(x,t) \ = \ 0 \quad \text{for } a < x < \infty, \ t \in (0, T), \tag{93}$$

with the terminal condition $u(x, T) = (e^x - K)^+$ and the boundary condition $u(a, t) = 0, \ t < T$.

As opposed to the explicit Euler method, here, we will use the CN scheme to approximate the solution to the above problem. The main features of CN scheme are:

- it is **unconditionally stable**;
- it has **lower computational complexity** than explicit Euler scheme, for the **same precision**.

The main difference in the construction of CN scheme is that **each space derivative of u (and u itself) is approximated by an average of the finite differences at the present and the previous time levels**. To see the rationale behind it, consider the standard backward approximation:

$$\frac{u(x,t)-u(x,t-\Delta t)}{\Delta t}$$

Recall that this is a first order approximation for $\partial_t u(x,t)$:

$$\frac{u(x,t)-u(x,t-\Delta t)}{\Delta t}=\partial_t u(x,t)+O(\Delta t)$$

However, it provides a second order approximation for $\partial_t u(x,t-\Delta t/2)$:

$$\frac{u(x,t)-u(x,t-\Delta t)}{\Delta t}=\partial_t u(x,t-\Delta t/2)+O(\Delta t^2),$$

simply because it can be viewed as a **central difference** for this derivative.

It turns out that, to approximate $\partial_x u(x,t-\Delta t/2)$ and $\partial^2_{xx}u(x,t-\Delta t/2)$ with the same order of accuracy, we simply need to take an average of our usual central difference approximations of these derivatives at the time levels t and $t-\Delta t$:

$$\frac{1}{2}\left(\frac{u(x+\Delta x,t)-u(x-\Delta x,t)}{2\Delta x}+\frac{u(x+\Delta x,t-\Delta t)-u(x-\Delta x,t-\Delta t)}{2\Delta x}\right)=\partial_x u(x,t-\Delta t/2)+O(\Delta x^2)+O(\Delta t^2),$$

$$\frac{1}{2}\left(\frac{u(x+\Delta x,t)+u(x-\Delta x,t)-2u(x,t)}{\Delta x^2}+\frac{u(x+\Delta x,t-\Delta t)+u(x-\Delta x,t-\Delta t)-2u(x,t-\Delta t)}{\Delta x^2}\right)$$
$$=\partial^2_{xx}u(x,t-\Delta t/2)+O(\Delta x^2)+O(\Delta t^2),$$
$$\frac{1}{2}(u(x,t)+u(x,t-\Delta t))=u(x,t-\Delta t/2)+O(\Delta t^2)$$

Plugging the above finite difference approximations into (93), we obtain:

$$\begin{aligned}\frac{u(x,t)-u(x,t-\Delta t)}{\Delta t}+\frac{\sigma^2}{4}\frac{u(x+\Delta x,t)+u(x-\Delta x,t)-2u(x,t)}{(\Delta x)^2}+\\ \frac{\sigma^2}{4}\frac{u(x+\Delta x,t-\Delta t)+u(x-\Delta x,t-\Delta t)-2u(x,t-\Delta t)}{(\Delta x)^2}+\\ \frac{1}{2}\left\{r-\frac{\sigma^2}{2}\right\}\left[\frac{u(x+\Delta x,t)-u(x-\Delta x,t)}{2\Delta x}+\frac{u(x+\Delta x,t-\Delta t)-u(x-\Delta x,t-\Delta t)}{2\Delta x}\right]\\ -\frac{r}{2}[u(x,t)+u(x,t-\Delta t)]\;=\;0.\end{aligned}\quad(94)$$

Rewriting (94), we obtain the following recursive relation:

$$(p+r\Delta t/2)u(x,t-\Delta t)-p^+u(x+\Delta x,t-\Delta t)-p^-u(x-\Delta x,t-\Delta t)\;=\;f(x,t)\,,\quad(95)$$

where p, p^+, p^- are given by the formulas

$$p = 1 + \sigma^2 \Delta t/2(\Delta x)^2, \tag{96}$$

$$p^+ = \sigma^2 \Delta t/4(\Delta x)^2 + (r - \sigma^2/2)\Delta t/4\Delta x \,,$$

$$p^- = \sigma^2 \Delta t/4(\Delta x)^2 - (r - \sigma^2/2)\Delta t/4\Delta x \,,$$

and

$$f(x,t) = (1 - r\Delta t/2)u(x,t) + \frac{\Delta t}{4\Delta x}\left\{r - \frac{\sigma^2}{2}\right\}(u(x+\Delta x,t) - u(x-\Delta x,t))$$

$$+\frac{\Delta t \sigma^2}{4(\Delta x)^2}(u(x+\Delta x,t) + u(x-\Delta x,t) - 2u(x,t))$$

Next, we consider a grid: $t_m = \Delta t m$, with $m = 0, \ldots, M$, and $x_n = \Delta x n$, with $n = 0, \ldots, N$, where $\Delta t = T/M$, $\Delta x = (b-a)/N$, and b is chosen to be large enough ($b = \log K + 3\sqrt{T}\sigma$). We introduce the discrete approximations $\{u_n^m\}$ of u: $u_n^m \approx u(a + \Delta x n, \Delta t m)$, for $n = 0, \ldots, N$ and $m = 0, \ldots, M$, with

- the terminal condition: $u_n^M = (e^{a+n\Delta x} - K)^+$, for $n = 0, \ldots, N$;
- and the boundary conditions: $u_0^m = 0$ and $u_N^m = e^b - Ke^{-r(M-m)\Delta t}$, for $m = 0, \ldots, M$.

Then (94) allows us to express a linear combination of u_{n-1}^{m-1}, u_n^{m-1} and u_{n+1}^{m-1} via $f(x_n, t_m)$, which is given in terms of u_{n-1}^m, u_n^m and u_{n+1}^m, computed at the previous step of the algorithm. Thus, we obtain a **recursive relation** for $m = M, \ldots, 1$.

In fact, each step of the above recursive algorithm requires a solution to the following **system of linear equations**:

$$A \begin{pmatrix} u_1^{m-1} \\ \vdots \\ u_{N-1}^{m-1} \end{pmatrix} = \begin{pmatrix} f(x_1, t_m) \\ \vdots \\ f(x_{N-2}, t_m) \\ f(x_{N-1}, t_m) + p^+(e^b - Ke^{-r(M-m+1)\Delta t})^+ \end{pmatrix}, \tag{97}$$

for $m = M, \ldots, 1$, where A is a tri-diagonal $(N-1) \times (N-1)$ matrix, with the entries:

$$A_{ij} = \begin{cases} -p^-, & j = i-1 \\ -p^+, & j = i+1, \\ p + r\Delta t/2, & j = i. \end{cases}$$

Numerical Linear Algebra

To implement implicit methods for solving PDE, such as the CN scheme, one needs to be able to efficiently solve large **sparse** systems of linear equations

$$Au \;=\; b, \quad \text{where } u, b \text{ are } n \text{ dimensional vectors and } A \text{ is a } N \times N \text{ matrix.} \tag{98}$$

The standard way to solve this problem is by **Gauss elimination**, which requires $O(N^3)$ computations for large N. In the cases we will be interested in, where A is derived from the discretization of a PDE, the matrix A is sparse – that means it has $O(n)$ non-zero entries. In fact, in our cases, A has only 3 non-zero diagonals and, hence, is tridiagonal. Evidently a diagonal matrix has at most n non-zero entries and a tri-diagonal matrix has at most $3n$ non-zero entries.

Matrices such as these are therefore sparse. If we consider implementing the Gauss elimination method for solving (98) when A is sparse, we observe that most of the computations involved are trivial since we are simply adding zeros. Hence these computations- which we could have predicted the result of before any actual computing- yield no new **information**. A basic principle in determining whether a given numerical method is **efficient** is that each new computation gives more information about the quantity which we wish to estimate.

It turns out that the above system can be solved using $O(N)$ computations via the **LU factorization**. Namely, for any matrix A, there exist matrices P, L and U, such that

- P is invertible (in fact, it corresponds to a change of the basis vectors by permutation);
- L is **lower diagonal** (has only zero entries above the man diagonal);
- U is **upper diagonal** (has only zero entries above the man diagonal);
- $PA = LU$.

The decomposition $A = PLU$ can be constructed in $O(N)$ computations, if A is tridiagonal. Then, the system (98) can be written as

$$LUx = Pb$$

Next, we solve two systems of equations

- $Ly = Pb$ for y,
- $Ux = y$ for x,

and notice that x is the desired solution. The main advantage of this method is that, due to the triangular structure of L and U, the above systems (for x and y) can be solved in $O(N)$ computations. Thus, we obtain the solution in the total of $3O(N) = O(N)$ computations.

A more general way to solve equations of the form (98), even when A is not a linear transformation, is to use the **iterative techniques**. First, we rewrite (98) as a **fixed point** equation

$$u \ = \ Bu + c \quad \text{where the matrix } B \text{ and vector } c \text{ are determined by } A, b. \tag{99}$$

We make an initial guess u^0 for the solution to (99) and then define a recurrence $u^k,\ k = 1, 2, ...$, by $u^{k+1} = Bu^k + c$. If the sequence $u^k,\ k = 1, 2, ...$, converges $\lim_{k\to\infty} u^k = u^\infty$, then $u = u^\infty$ is the solution to (98). Note that the evaluation of each u^k requires just $O(N)$ computations since B is sparse. There are two aspects to the implementation of this method:

(a) **Convergence:** We need to show that the matrix B has properties which imply that sequences $u^k,\ k = 1, 2, ...$, converge.
(b) **Rate of Convergence:** We need to know how large k needs to be so that the error $u^\infty - u^k$ is small.

The number of computations required for the evaluation of u^k is $\simeq kN$, so (b) is very important in understanding how efficient the method really is. Obviously if we need to take $k = N^2$ to get a good approximation to the solution of (99), we might as well have used the straightforward Gauss elimination method instead.

Let us consider an example of matrix B for which (98), (99) are equivalent. This example is known as the **Jacobi method**. We write $A = D - E - F$, where D is a diagonal matrix consisting of just the diagonal entries of the matrix A. The matrix $-E$ consists of the **lower triangular matrix** with entries equal to the entries of A **below** the diagonal. The matrix $-F$ consists of the **upper triangular matrix** with entries equal to the entries of A **above** the diagonal. Hence (98) is equivalent to the equation $Du = (E + F)u + b$. Assuming D is invertible we see that (99) holds with

$B = D^{-1}(E+F)$ and $c = D^{-1}b$, whence the matrix B is the Jacobi iteration matrix. The corresponding iteration sequence $u^k,\ k = 1, 2, ...$, therefore satisfies the recurrence

$$Du^{k+1} \ = \ Eu^k + Fu^k + b, \quad k = 1, 2, ..., \tag{100}$$

The iteration matrix B for (100) is evidently $B = D^{-1}(E+F)$ and $c = D^{-1}b$. Since D is a diagonal and E and F are sparse, matrix the evaluation of u^{k+1} in (100) from u^k requires $O(N)$ computations.

Finally we wish to find properties of the matrix B in (99) which guarantees convergence of the sequence $u^k,\ k = 1, 2, ..$, to the solution of (99). Let $u = u^\infty$ be the solution to (99) and $\varepsilon^k = u^k - u^\infty$ be the error of the kth iteration. Since $u^\infty = Bu^\infty + c$ and $u^{k+1} = Bu^k + c$ it follows that

$$\varepsilon^{k+1} \ = \ B\varepsilon^k, \quad \text{which implies } \varepsilon^k \ = \ B^k\varepsilon^0 \ . \tag{101}$$

We measure the size of ε^k by introducing a **norm** $\|\cdot\|$ on the space $\mathbb{R}^N$ of N dimensional vectors.

Rem 19. *The most familiar norm is the* **Euclidean distance** *norm $\|\cdot\|_2$ defined by*

$$\|v\|_2 \ = \ \left\{ \sum_{j=1}^{N} v_j^2 \right\}^{1/2} \qquad v = (v_1, .., v_N)^T \in \mathbb{R}^N \ . \tag{102}$$

*Another, ∞-**norm**, is the analogue of the* **sup-norm** *for functions, and it is defined by*

$$\|v\|_\infty \ = \ \max_{1 \le j \le N} |v_j| \ , \quad v = (v_1, .., v_N)^T \in \mathbb{R}^N \ . \tag{103}$$

A norm $\|\cdot\|$ on $\mathbb{R}^N$ induces a special norm on the linear space of $N \times N$ matrices as follows:

$$\|B\| \ = \ \max_{\|v\|=1} \|Bv\| \ . \tag{104}$$

It follows from the definition (104) that $\|Bv\| \le \|B\|\|v\|$ for all $v \in \mathbb{R}^N$. Hence if B_1, B_2 are two $N \times N$ matrices then

$$\|B_1B_2v\| \ \le \ \|B_1\|\|B_2v\| \ \le \ \|B_1\|\|B_2\|\|v\| \quad \text{for } v \in \mathbb{R}^N. \tag{105}$$

In particular we have that $\|B^k\| \le \|B\|^k,\ k = 1, 2, ..$, whence it follows that if $\|B\| < 1$ then $\lim_{k\to\infty} \|B^k\| = 0$. Comparing with (101), we have therefore shown that if $\|B\| < 1$ for some induced matrix norm then the sequence $u^k,\ k = 1, 2, ..$, converges to the solution of (99).

In our cases, each iteration of the Jacobi method requires $O(N)$ computations, since the matrix B is sparse. However, the problem is that the number of iterations (i.e. the upper limit of k) necessary to achieve the same precision of the approximation, grows as the matrix size N increases. It turns out that, in the end, the total number of computations is $O(N^3)$, and it appears that our iteration method is not better than the simple Gauss elimination method. It is true that the pure Jacobi method does not do better than Gauss elimination. However, there exist more sophisticated iteration algorithms (e.g. **multi-grid algorithms**) based on the Jacobi method described above, which require $O(N \log N)$ computations. These algorithms are closely linked to **fast Fourier transform** algorithms which enable one to compute the discrete Fourier transform of an N dimensional vector in $\simeq N \log N$ computations.

The accuracy of the CN method is $O[(\Delta t)^2] + O[(\Delta x)^2]$. Thus, in the implementation of CN we take $\Delta t \simeq \Delta x$. Then, assuming we are able to solve the linear system of equations (97) in $O(N)$ computations, the total number of computations required to get a solution with error $O[(\Delta x)^2]$ is $O(M)O(N) = O((1/\Delta t)(1/\Delta x)) = O(1/\Delta x^2)$.

In the explicit Euler method, $\Delta t \simeq (\Delta x)^2$ and we need to perform $O(N) = O(1/\Delta x)$ operations at each time step. Hence the total number of computations required to get a solution with error $O[(\Delta x)^2]$ is $O((1/\Delta t)(1/\Delta x)) = O(1/\Delta x^3)$.

Thus, **for a given accuracy, the CN scheme is significantly more efficient than explicit Euler scheme, provided we have an efficient linear equation solver**. When the dimension of the space variable x is higher than one, such linear solvers are not available, and the two schemes have comparable complexity.

Rem 20. *In fact, the accuracy of any finite difference scheme also depends on the regularity of the terminal condition. In the above analysis, we assume that the terminal condition is continuous with respect to x (or S), which, in the case of a down-and-out call, implies that $K > U$. When the payoff is discontinuous (e.g. $K < U$), the convergence of the scheme is much slower than predicted (if it holds at all) around the discontinuity point.*

Stability

However, the main advantage of the CN method (as an implicit finite difference method) is its **unconditional stability**! To see this, assume that Δt and Δx are small enough. Consider, first, the explicit Euler scheme and notice that every time step of this scheme, approximately, corresponds to

$$u^{m-1} \approx (I - \sigma^2 \frac{\Delta t}{\Delta x^2} \tilde{A}) u^m,$$

where

$$\tilde{A} = \begin{pmatrix} 1 & -1/2 & 0 & \cdots & 0 \\ -1/2 & 1 & -1/2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & -1/2 & 1 & -1/2 \\ 0 & 0 & 0 & -1/2 & 1 \end{pmatrix}$$

On the other hand, every time step of CN scheme is

$$(I + \sigma^2 \frac{\Delta t}{2\Delta x^2} \tilde{A}) u^{m-1} \approx (I - \sigma^2 \frac{\Delta t}{2\Delta x^2} \tilde{A}) u^m,$$

$$u^{m-1} \approx (I + \sigma^2 \frac{\Delta t}{2\Delta x^2} \tilde{A})^{-1} (I - \sigma^2 \frac{\Delta t}{2\Delta x^2} \tilde{A}) u^m.$$

Assuming that $\tilde{A}$ is a number (i.e. matrix of dimension one), we conclude that **in the CN scheme**

$$\|u^{m-1}\| \leq \left| \frac{1 - \sigma^2 \frac{\Delta t}{2\Delta x^2} \tilde{A}}{1 + \sigma^2 \frac{\Delta t}{2\Delta x^2} \tilde{A}} \right| \|u^m\| < \|u^m\|,$$

where $\|\cdot\|$ is the "sup" norm: $\|u^m\| = \sup_n |u_n^m|$. It turns out that this conclusion is true even when $\tilde{A}$ is a matrix, as opposed to a number. This is what causes **stability of the CN scheme**.

On the other hand, when Δt is large enough, we may have

$$\|u^{m-1}\| = \left| 1 - \sigma^2 \frac{\Delta t}{2\Delta x^2} \tilde{A} \right| \|u^m\| > \|u^m\|,$$

which causes the **instability of the explicit Euler scheme**. Again, the above is easy to see when $\tilde{A}$ is a number, and we simply note here that this conclusion remains true in higher dimension.

6 Calibration of LV models

6.1 Dupire's formula

Q 4. *Can we find a LV model that reproduces the observed market prices (of European options, in the present case)? As we know, the search for such model is called* ***calibration****.*

Part of the reason why Local Volatility models became popular is the fact that they are the **simplest volatility models which can be calibrated to the market of European options** (can match the implied smile).

Assume we are in a LV model with constant interest rate r and zero dividend rate ($q = 0$). Note that the risky asset S, in a LV model, is a **Markov process**, and, if started at time t from the value S, at time $T > t$ it has a density $p(S, t; \cdot, T)$.

As a **Markov transition density**, $p(S, t; S', T)$ satisfies the **backward Kolmogorov equation**

$$\frac{\partial}{\partial t}p + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2}{\partial S^2}p + rS\frac{\partial}{\partial S}p = 0$$

Recall that $p(S, t; S', T)$ also satisfies the **forward Kolmogorov equation**

$$\frac{\partial}{\partial T}p - \frac{1}{2}\frac{\partial^2}{\partial S'^2}\left(\sigma^2(S',T)S'^2 p\right) + \frac{\partial}{\partial S'}\left(rS'p\right) = 0$$

We can connect the call prices and the transition density:

$$\begin{aligned} C(S,t;K,T) &= e^{-r(T-t)}\int_0^\infty p(S,t;S',T)(S'-K)^+ dS' \\ &= e^{-r(T-t)}\int_K^\infty p(S,t;S',T)(S'-K)dS' \end{aligned}$$

The above yields the **Shimko's formula**:

$$\begin{aligned} \frac{\partial}{\partial K}C(S,t;K,T) &= -e^{-r(T-t)}\int_K^\infty p(S,t;S',T)dS', \\ \frac{\partial^2}{\partial K^2}C(S,t;K,T) &= e^{-r(T-t)}p(S,t;K,T) \end{aligned}$$

Multiplying the forward Kolmogorov equation by $(S'-K)^+$ and integrating over S', we obtain

$$\begin{aligned} &\frac{\partial}{\partial T}\int_K^\infty p(S,t;S',T)(S'-K)dS' = \\ &\qquad \frac{1}{2}\int_K^\infty (S'-K)\frac{\partial^2}{\partial S'^2}\left(\sigma^2(S',T)S'^2 p\right)dS' - \int_K^\infty (S'-K)\frac{\partial}{\partial S'}\left(rS'p\right)dS' \\ &\qquad\qquad = \frac{1}{2}\sigma^2(K,T)K^2 p(S,t;K,T) + rK\int_K^\infty p(S,t;S',T)dS' + r\int_K^\infty p(S,t;S',T)(S'-K)dS', \end{aligned}$$

where we integrated by parts several times.

Using the relation between K-derivatives of C and p, we obtain the **Dupire's formula**:

$$\sigma^2(K,T) = 2\frac{\partial_T C(S,t;K,T) + rK\partial_K C(S,t;K,T)}{K^2\partial_K^2 C(S,t;K,T)} \tag{106}$$

The Dupire's equation

$$\partial_T C - \frac{1}{2}\sigma^2(K,T)K^2\partial_K^2 C + rK\partial_K C = 0, \quad K > 0,\ T > t \tag{107}$$

should not be confused with the Black-Scholes equation:

$$\partial_t C + \frac{1}{2}\sigma^2(S,t)S^2\partial_S^2 C + rS\partial_S C - rC = 0, \quad S > 0,\ t < T$$

The main advantage of **Dupire's equation** is that **it treats call price as a function of strike and maturity**, and such a function is **observed at a given moment in time**!

Since (107) determines option prices uniquely, the Dupire's formula, in principle, can be used to **calibrate a LV model to call prices**.

- Indeed, we have a one-to-one mapping $\sigma(.,.) \leftrightarrow C(S,0;.,.)$ given by the two equations

$$\sigma^2(T,K) = 2\frac{\partial_T C + rK\partial_K C}{K^2\partial_K^2 C} \tag{108}$$

and

$$\begin{cases} \partial_T C - \frac{1}{2}\sigma^2(T,K)K^2\partial_K^2 C + rK\partial_K C = 0, \quad K > 0,\ T > 0 \\ C(0,K) = (S-K)^+ \end{cases} \tag{109}$$

Q 5. Works in theory, but how do we implement it?

In fact, the Dupire's formula

$$\sigma^2(T,K) = 2\frac{\partial_T C + rK\partial_K C}{K^2\partial_K^2 C}$$

has serious practical drawbacks.

- **It requires continuum of strikes and maturities**.
- **Numerical differentiation is ill-conditioned**.
- As $T \downarrow t$, the denominator converges to ∞ for at-the-money options, and to zero for out- and in-the-money options.
- The denominator tends to zero as $K \to \infty$ or $K \to 0$.

6.2 Ill-posed problems

Def 2. *A problem is called **well-posed** (according to* **J. Hadamard***) if*

1. *a solution* **exists***,*
2. *the solution is* **unique***, and*
3. *the* **solution depends "continuously" on the input data**

*A problem that is not well-posed is called **ill-posed**.*

Typically, the first two properties are verified at the stage when the problem is being formulated (e.g. it doesn't make sense to consider a problem that does not have a solution). But the derived formulas often ignore the issue of **continuity**. This issue needs to be addressed when a computation method (or algorithm) is developed, and this is what we focus on in this section.

In the above context, it is important to agree on what we understand by continuity! We will typically have **functions (or vectors)** as the input and output of a problem. Then, if a function is observable, a **natural notion of continuity** is with respect to the **sup-norm**:

$$\|f\| = \sup_x |f(x)|$$

Ex 2. *A problem may or may not be continuous with respect to the sup-norm.*

- **Integration**. *Assume that the input is a continuous function $f : [0,1] \mapsto \mathbb{R}$. The output is $F : [0,1] \mapsto \mathbb{R}$, given by*

$$F(x) = \int_0^x f(z)dz$$

If f is observed with an error, that is we observe f_ε with

$$\|f - f_\varepsilon\| = \sup_{x\in[0,1]} |f(x) - f_\varepsilon(x)| \le \varepsilon,$$

then

$$|F(x) - F_\varepsilon(x)| \le \int_0^x |f(z) - f_\varepsilon(z)|\, dz \le \varepsilon x,$$

hence $\|F - F_\varepsilon\| \le \varepsilon$, and the output is continuous with respect to the input. Thus, integration is a **well-posed** *problem.*

- **Differentiation**. *Assume that the input is a continuously differentiable function $f : [0,1] \mapsto \mathbb{R}$. And the output $F : [0,1] \mapsto \mathbb{R}$ is its derivative:*

$$F(x) = \frac{d}{dx} f(x)$$

As before, the function f is observed as f_ε, which contains a small error. Let us choose the error so that

$$f_\varepsilon(x) = f(x) + \varepsilon \sin(x/\varepsilon^2)$$

Then $\|f - f_\varepsilon\| \le \varepsilon$, but

$$F_\varepsilon = \frac{d}{dx} f_\varepsilon(x) = F(x) + \frac{1}{\varepsilon}\cos(x/\varepsilon^2),$$

and for all small enough $\varepsilon > 0$, there exists $x \in [0,1]$, such that $|F(x) - F_\varepsilon(x)| = 1/\varepsilon$. Therefore, $\|F - F_\varepsilon\| \to \infty$, as $\varepsilon \to 0$, even though $\|f - f_\varepsilon\| \to 0$. Thus, differentiation is an **ill-posed** *problem.*

Ill-posed problems may possess a well-defined (i.e. unique) solution in theory, but, in practice, it may be **impossible to compute** the solution directly, as any tiny error in the input (and there are always errors present in any actual computation) will get amplified and may distort the output enormously.

Notice that, even if we observed the call price function $C(K,T)$ for all (K,T), the **Dupire's formula** (106) involves differentiation, hence, we expect it to be **ill-posed**.

However, the situation is even worse: we don't actually observe the call prices $C(K,T)$ for all (K,T), so, it is not even clear how to interpret the right hand side of (106). This is why, in fact, we need to treat the problem of calibrating a LV model as an **inverse problem**.

6.3 Inverse problems and Tikhonov regularization

Def 3. *The problem is called an* **inverse problem** *if it is defined as an inverse of some other problem stated "more explicitly".*

Consider an abstract **inverse** problem,

$$\text{Given } y \in \mathbb{Y},\ \text{find } x \in \mathbb{X},\ \text{such that:}\quad F(x) = y,$$

where $\mathbb{Y}$ is the space of admissible input values (e.g. call prices functions $\{C(K,T)\}$) and $\mathbb{X}$ is the space of admissible output values (e.g. LV functions $\{\sigma(K,T)\}$).

Rem 21. *Notice that, for the* **direct** *problem,*

$$\textit{Given } x \in \mathbb{X},\ \textit{find } y \in \mathbb{Y},\ \textit{such that:}\quad y = F(x),$$

we have: $\mathbb{X}$ *is the space of admissible input values and* $\mathbb{Y}$ *is the space of admissible output values. To understand what is the input and what is the output, you need to find out from the context whether we refer to the direct of the inverse problem (in this subsection, we will typically discuss the inverse problem).*

Of course, an inverse problem is only interesting if the inverse transformation F^{-1} **cannot be computed explicitly**. Then, the inverse problem can formulated as the following **optimization problem**:

$$\min_{x \in \mathbb{X}} \|F(x) - y\|_{\mathbb{Y}}^2, \tag{110}$$

with an appropriate choice of norm $\|\cdot\|_{\mathbb{Y}}$.

The above formulation, in principle, allows us to use the existing optimization algorithms (such as the **gradient search** method) to solve inverse problems. However, if the **solution to the inverse problem** $F^{-1}(y)$ is not continuous with respect to the input y, we still face the same problem of numerical **instability**, discussed in the previous subsection: a small error in the input may result in a huge error in the solution.

Q 6. What features of function F cause the ill-posedness of the inverse problem?

The discontinuity of the inverse problem typically means that the **direct problem** $x \mapsto F(x)$ **is not very sensitive to changes in** x **along certain directions**.

For example, if $F : [0,1] \to \mathbb{R}$ is a one-dimensional monotone function, it is clear that a **discontinuity of the inverse problem** $F^{-1}(y)$ at $y = y_0$ corresponds to a **constant interval for the direct problem**: $F(x) = y_0$ for all $x \in [a,b]$.

Similarly, the **differentiation can be viewed as an inverse problem of integration**. As shown in the example in the previous subsection, **differentiation becomes unstable** when the error has **large and frequent oscillations, staying close to zero on average** (e.g. $\varepsilon \sin(x/\varepsilon^2)$). Note that this is exactly the type of error that will cause (almost) **no change in the direct problem** of integration!

Thus, the **discontinuity of $F^{-1}(y)$ is caused by the lack of sensitivity of $F(x)$ to the changes in x** (in certain directions).

Rem 22. *There is another, more technical, problem caused by the aforementioned lack of sensitivity. Most of the algorithms for solving the minimization problem (110) are based on the so called* **gradient search** *method. This method does not perform well if the function $F(x)$ is not sensitive to changes in x.*

The above issue can often be fixed by introducing the **Tikhonov regularization** of (110):

$$\min_{x\in\mathbb{X}} \|F(x)-y\|_{\mathbb{Y}}^2 + \varepsilon\|x-x_0\|_{\mathbb{X}}^2, \tag{111}$$

which contains the additional "**penalization**" term $\varepsilon\|x-x_0\|_{\mathbb{X}}^2$, with an appropriately chosen norm $\|\cdot\|_{\mathbb{X}}$.

The main role of the penalization term is to**increase the sensitivity of the objective function to the changes in** x, especially in those directions in which $F(x)$ is not very sensitive to the changes.

In addition, x_0 represents the "**prior beliefs**" about what the solution to the problem should look like (or what it should not be too far away from). This is, however, not so important, and x_0 is often dropped.

In the context of **calibrating a LV model** to call prices, the Tikhonov regularization of the problem looks as follows:

$$\min_{\sigma\in\mathbb{X}} \sum_{i=1}^{N} w_i \left(C^{\sigma}(K_i,T_i) - C^{mkt}(K_i,T_i)\right)^2 + \varepsilon\|\sigma-\sigma_0\|^2, \tag{112}$$

where $C^{\sigma}(K,T)$ are the call prices produced by the LV function σ via (109), $\left\{C^{mkt}(K_i,T_i)\right\}_{i=1,\dots,N}$ are the market call prices, and $\mathbb{X}$ is a subspace of LV functions.

Methods differ in:

- Choice of the **weights** w_i (more value should be given to liquid options). The standard choice is to take each weight to be proportional to the bid-ask spread.
- Choice of the **minimization algorithm**. This is a subject of a course in optimization.
- Choice of the **norm** $\|\cdot\|_{\mathbb{X}}$: for example, Lagnado, Osher, Jackson, Süli, Howison, Crépey consider

$$\|\sigma\|^2 = \int_0^{T_{\max}} \int_0^{\infty} \left(\frac{\partial\sigma}{\partial K}\right)^2 + \left(\frac{\partial\sigma}{\partial T}\right)^2 dK\,dT$$

- Choice of the **space of LV functions** $\mathbb{X}$. One can narrow down the space $\mathbb{X}$ of possible Local Volatility functions to increase the computational accuracy and speed, and, sometimes even to achieve well-posedness of the problem. Egger, Hein and Hofmann (2005) suggest to search for LV in the form

$$\sigma(T,K) = \sigma(K)\rho(T)$$

Rem 23. *Of course, the most convenient case is when $\mathbb{X}$ is a finite-dimensional family of LV functions. However, it is a very difficult task to construct an explicit* **finite-dimensional family of LV functions that can fit any given number of options' prices***. In their recent paper, Carr and Nadtochiy do construct such a family but this construction is rather complicated and goes beyond the scope of this course.*

Once the above choices are made, one solves the regularized problem (112), or, more generally, (111), and iterates over various values of $\varepsilon \to 0$, until the convergence is observed. In general, one expects to have

$$\sigma^\varepsilon \to \sigma,$$

as $\varepsilon \to 0$, where σ^ε is the solution of (112) and σ is the LV function that matches the market prices precisely. However, it is often difficult to prove such a convergence, and even to define in which sense (i.e. in which metric) the solutions converge: the convergence may not hold in the usual "pointwise" sense, and even a weaker form of convergence requires an appropriate choice of the penalization term. All these issues present a very challenging task in the theory of Inverse Problems, and they go well beyond the scope of this course.

6.4 Mixture models

Brigo an Mercurio (2000-2002) have introduced a class of simple models for the risky asset S, in which, under the risk neutral measure $\mathbb{Q}$,

$$dS_t = rS_t dt + \sigma_t S_t dW_t,$$

and the future path of the volatility (σ_t) is chosen randomly at the very beginning ($t = 0$), out of a finite family of possible deterministic paths, say, $\{\sigma_i(t)\}$, independent of the Brownian motion W.

- For example, assume that $\sigma_i^2(t) = \theta_i + (\sigma^2 - \theta_i)e^{-\eta_i t}$, and it is chosen with probability p_i, for $i = 1, 2, 3$.
- The price of a European call option given by such a local volatility surface is

$$C^\Theta(K,T) = \sum_{i=1}^m p_i C^{BS}(K,T; \sqrt{\frac{1}{T}\int_0^T \sigma_i^2(u)du}, r),$$

  where

$$\Theta = \left\{\theta_i, \sigma^2, \eta_i, p_i\right\}_{i=1,2,3}$$

- Then the Local Volatility implied by the call prices is given by

$$a^2(\tau, x) = 2\frac{\partial_T C^\Theta(K,T) + rK\partial_K C^\Theta(K,T)}{K^2 \partial^2_{KK} C^\Theta(K,T)}$$

$$= \frac{\sum_{i=1}^3 p_i(\theta_i + (\sigma^2 - \theta_i)e^{-\eta_i T})\exp(-d_i^2(K,T)/2)/v_i(T)}{\sum_{i=1}^3 p_i \exp(-d_i^2(K,T)/2)/v_i(T)},$$

  where

$$v_i(T) = \sqrt{\frac{1}{T}\int_0^T \sigma_i^2(u)du},$$

$$d_i(T,x) = \frac{\log(S/K) + (r + \frac{1}{2}v_i^2(T))T}{\sqrt{T}v_i(T)}.$$

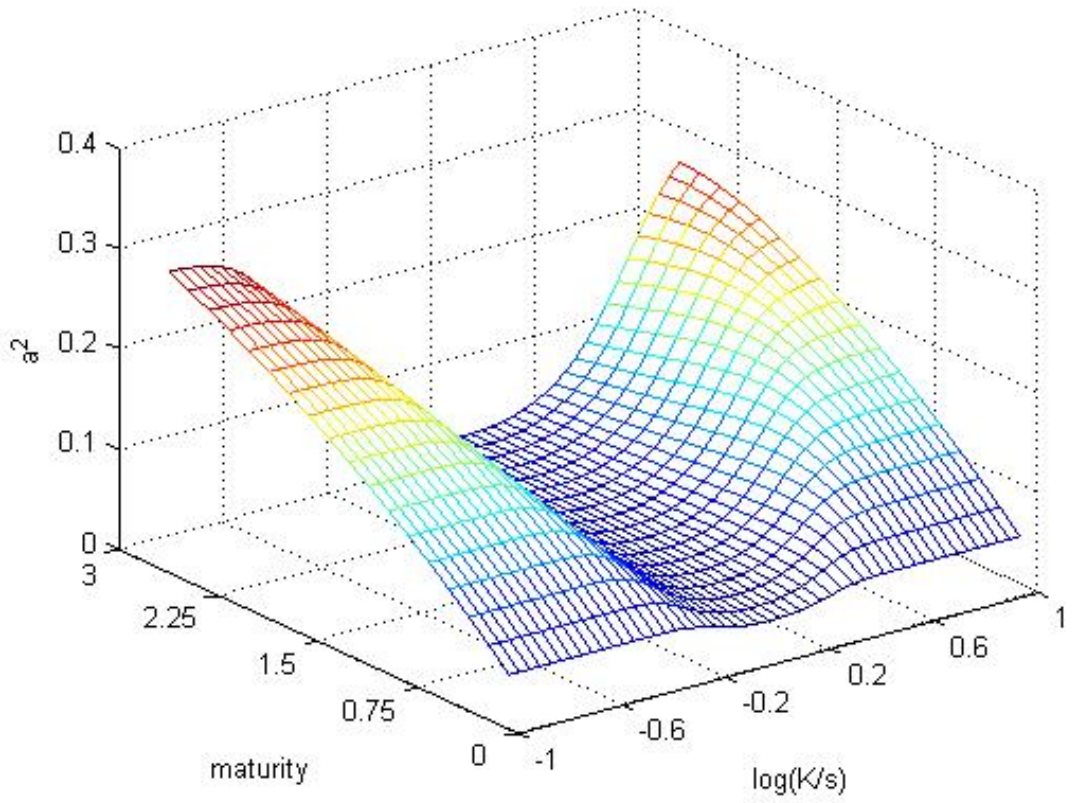

Figure 2: This figure gives the graph of a local volatility surface from the above family. It was obtained by least squares fitting to the European call option prices quoted on the SP500 index on April 3rd, 2006. The parameter values produced by our least squares optimization are: $\sigma = 0.16$, $\eta_1 = 0.26$, $\eta_2 = 0.21$, $\theta_1 = 0.52$, $\theta_2 = 0$, $p_1 = 0.26$, $p_2 = 0.31$ and $\mu = 0.045$.

Short-term implied volatility in LV models

Following Beresticky, Busca and Florent (2002), we can establish a direct relation between short-maturity Implied and Local volatilities.

- Assume that the risky asset S follows a LV model

$$dS_t = rS_t dt + \sigma(S_t, t)S_t dW_t,$$

  with a constant rate r.

- Denote by $\Sigma(K, T)$ the Implied volatility of a European call option with strike K and maturity T, given by this model.

- Introduce the change of variables

$$\begin{aligned} x &= \log(S/K) + rT, \\ \tilde{\sigma}(x, T) &= \sigma(S\exp(rT - x), T), \\ \tilde{\Sigma}(x, T) &= \Sigma(S\exp(rT - x), T) \end{aligned}$$

- Then, we have:

$$\frac{1}{\tilde{\Sigma}(x, 0)} = \frac{1}{x}\int_0^x \frac{dy}{\tilde{\sigma}(y, 0)} \tag{113}$$

- The left hand side of (113), in principle, can be observed in the market from the prices of short-maturity European options. Then, the calibration to such options can be done by simply finding $\tilde{\sigma}$ which makes the equality in (113) hold. This is a significant simplification of the calibration problem (which only works for short maturities).