PDE for Finance Notes, Spring 2003 – Section 3.

Notes by Robert V. Kohn, Courant Institute of Mathematical Sciences. For use in connection with the NYU course PDE for Finance, G63.2706.

More about linear PDE's: the heat equation on an interval; uniqueness via the maximum principle. Section 2 gave the solution formula for the heat equation in a half-space – the essential tool for pricing barrier options. The natural next topic is to consider the heat equation in an interval – the essential tool for pricing double-barrier options. Then we turn to the question of uniqueness, something we've been assuming up to now. The proof uses the maximum principle – a basic and powerful tool for obtaining qualitative information about solutions of parabolic equations.

Our treatment of the heat equation in an interval is by separation of variables; this can be found in most basic PDE books. The recent article *Exotic Spectra* by Vadim Linetsky (Risk Magazine, April 2002) reviews this topic and explores its financial applications; a xerox copy is on reserve in the CIMS library green box for this class. Uniqueness via the maximum principle is also quite standard; students of finance may find it convenient to read this in Steele's book (the relevant section is in the middle of the book but is quite self-contained).

****

The heat equation in an interval. A double-barrier option has both an upper and lower barrier. Its value satisfies the Black-Scholes PDE with appropriate boundary data at the barriers. If the underlying is lognormal then this problem can be reduced, as shown in Section 2, to solving the linear heat equation on an interval with appropriate initial and boundary data. For simplicity we focus on the case when the interval is $0 < x < 1$. Thus we wish to solve:

$$u_t = u_{xx} \quad \text{for } t > 0 \text{ and } 0 < x < 1$$

with initial data $u = g$ at $t = 0$, and boundary data $u = \phi_0$ at $x = 0$, $u = \phi_1$ at $x = 1$.

Consider first the case when the boundary condition is zero, i.e. $\phi_0 = \phi_1 = 0$. We will use the following basic fact: any function $f(x)$ defined for $0 < x < 1$ and vanishing at the endpoints $x = 0, 1$ can be expanded as a Fourier sine series:

$$f(x) = \sum_{k=1}^{\infty} a_n \sin(n\pi x); \tag{1}$$

moreover the coefficients a_n are determined by f via

$$a_n = 2\int_0^1 f(x)\sin(n\pi x)\,dx. \tag{2}$$

If you accept (1), the justification of (2) is easy. It follows from the fact that the functions $\{\sin(n\pi x)\}_{n=1}^{\infty}$ are orthogonal in the inner product $(f, g) = \int_0^1 f(x)g(x)\,dx$, and each has norm $1/\sqrt{2}$.

Now consider the solution of $u_t = u_{xx}$ with boundary data 0 and initial data $g(x)$. Applying (1) at each time, we have

$$u(x,t) = \sum_{k=1}^{\infty} a_n(t) \sin(n\pi x)$$

for some functions $a_n(t)$. To get initial data right, $a_n(0)$ must be the Fourier sine-series coefficients of g. To satisfy the PDE we must have

$$da_n/dt = -n^2\pi^2 a_n,$$

whence $a_n(t) = a_n(0)e^{-n^2\pi^2 t}$. Thus the solution is

$$u(x,t) = \sum_{n=1}^{\infty} g_n e^{-n^2\pi^2 t} \sin(n\pi x) \tag{3}$$

where

$$g_n = 2\int_0^1 g(x) \sin(n\pi x)\, dx. \tag{4}$$

Notice an important feature of this solution formula: the nth term decays like $e^{-n^2\pi^2 t}$. Thus terms with higher n (corresponding to higher-order modes of the initial data) decay faster. If you only need the solution to a certain accuracy at a fixed time T then you only need consider a limited number of modes – and the number of modes decreases as T increases! Thus valuing an option by this method gets *easier* as the maturity gets larger. This is opposite to the behavior of a time-stepping numerical scheme (such as finite-difference approximation).

It is natural to seek a solution formula in the form

$$u(x,t) = \int_0^1 G(x,y,t) g(y)\, dy \tag{5}$$

since $G(x,y,t)$ is then the probability that a random walker starting at y arrives at x at time t without first hitting the boundary. (Our random walker executes, as usual, $\sqrt{2}$ times Brownian motion.) This is just a matter of manipulation: combining (3) and (4) gives (5) with

$$G(x,y,t) = 2\sum_{n=1}^{\infty} e^{-n^2\pi^2 t} \sin(n\pi x)\sin(n\pi y). \tag{6}$$

One might think the solution formula just presented was limited to initial conditions with $g = 0$ at the boundary, i.e. "consistent" initial data. However this is not the case: our formula is correct even for inconsistent data. The reason is that *any* function on $[0,1]$ has a Fourier sine series – which equals the function almost everywhere (though perhaps not at the endpoints).

Let's turn now to the boundary value problem, i.e. the solution of $u_t - u_{xx} = 0$ with initial condition 0 and boundary data $u = \phi_0$ at $x = 0$, $u = \phi_1$ at $x = 1$. We did all the work already in Section 2: the argument given there shows that the solution is

$$u(x,t) = \int_0^t \frac{\partial G}{\partial y}(x,0,t-s)\phi_0(s)\, ds - \int_0^t \frac{\partial G}{\partial y}(x,1,t-s)\phi_1(s)\, ds$$

where G is given by (6). In particular, $\frac{\partial G}{\partial y}(x,0,t)$ is the probability that the random walker, starting from x, hits the boundary first at 0 and arrives there at time t. Similarly, $-\frac{\partial G}{\partial y}(x,1,t)$ is the probability that the random walker, starting from x, hits the boundary first at 1 and arrives there at time t.

The "separation of variables" solution method just presented is not limited to the constant-coefficient heat equation. It can be applied to any equation of the form

$$u_t - \tfrac{1}{2}a^2(x)u_{xx} - b(x)u_x + r(x)u = 0$$

on an interval, provided a is strictly positive on this interval. Thus it can be used to price double-barrier options on a fairly general class of underlyings (provided that the volatility is independent of time). The key is to recognize that Fourier sine series (and the associated decay rates $n^2\pi^2$) must be replaced by the eigenfunctions (and eigenvalues) of

$$-\tfrac{1}{2}a^2(x)v_{xx} - b(x)v_x + r(x)v = \lambda v$$

with $v = 0$ at the boundary. These eigenfunctions are orthogonal not in the L^2 inner product, but rather in a different one determined by the coefficients a and b. See Linetsky's article *Exotic Spectra* for details and applications.

****

Why does time have a preferred direction? This is a convenient time to ask: why is it so important that we solve $u_t - u_{xx} = 0$ forward rather than backward in time? The answer is that solving in the correct time direction is stable (moreover the solution is instantly smooth, regardless how rough the initial data may be) whereas solving in the wrong time direction is extremely unstable. Indeed, consider the separation of variables technique presented above. We can, in principle, apply it in the wrong time direction, since the ODE $da_n/dt = -n^2\pi^2 a_n$ is reversible. Thus the solution of $u_t - u_{xx} = 0$ for $t < T$ with final-time data $u = g$ at $t = T$ is

$$u(x,t) = \sum_{n=1}^{\infty} g_n e^{n^2\pi^2(T-t)} \sin(n\pi x)$$

But watch out! The nth mode grows extremely fast: as a constant times $e^{n^2\pi^2(T-t)}$ as $T-t$ increases. The tiniest high-frequency ripple in g will become huge as you proceed backward in time.

The whole-space problem behaves similarly: solving backward in time is extremely unstable. This is, in fact, a consequence of the fact that solving forward in time is so stable and smoothing.

Might there still be some interest in solving the heat equation the "wrong way" in time? Sure. This is the simplest example of "deblurring," a typical task in image enhancement.

Consider a photograph taken with an out-of-focus camera. Its image is (roughly speaking) the convolution of the true image with a Gaussian of known variance. Finding the original image amounts to backsolving the heat equation with the blurry photo as final-time data. Backsolving the heat equation is a typical example of an *ill-posed problem* – one whose answer depends in an unreasonably sensitive way on the data.

****

The maximum principle and uniqueness. Are our solution formulas the *only* solution of the heat equation with the specified initial and/or boundary conditions? By linearity, this amounts to asking whether $u = 0$ is the only solution with data 0? The answer is yes. We shall prove this using the maximum principle.

The maximum principle. This is an elementary yet far-reaching fact about solutions of linear parabolic equations. Here is the simplest version:

> Let D be a bounded domain. Suppose $f_t - \Delta f \leq 0$ for all $x \in D$ and $0 < t < T$. Then the maximum of f in the closed cylinder $\bar{D} \times [0, T]$ is achieved either at the "initial boundary" $t = 0$ or at the "spatial boundary" $x \in \partial D$.

Notice the asymmetry between the initial boundary $t = 0$ (where f can easily achieve its maximum) and the final boundary $t = T$ (where f does not achieve its maximum – except in the trivial case when f is constant). This asymmetry reflects once again the fact that time has a "preferred direction" when solving a parabolic PDE.

If $f_t - \Delta f$ were *strictly* negative, the principle just enunciated would be a calculus exercise. Indeed, f must achieve its maximum *somewhere* in the cylinder or on its boundary (we use here that D is bounded). Our task is to show this doesn't occur in the interior or at the "final boundary" $t = T$. At an interior maximum all first derivatives would vanish and $\partial^2 f / \partial x_i^2 \leq 0$ for each i; but then $f_t - \Delta f \geq 0$, contradicting the hypothesis that $f_t - \Delta f < 0$. At a final-time maximum (in the interior of D) all first derivatives in x would still vanish, and we would still have $\partial^2 f / \partial x_i^2 \leq 0$; we would only know $f_t \geq 0$, but this would still give $f_t - \Delta f \geq 0$, again contradicting the hypothesis of strict negativity.

If all we know is $f_t - \Delta f \leq 0$ then the preceding argument doesn't quite apply. But the fix is simple: we can apply it to $f_\epsilon(x, t) = f(x, t) - \epsilon t$ for any $\epsilon > 0$. As $\epsilon \to 0$ this gives the desired result.

There is an analogous minimum principle:

> Let D be a bounded domain. Suppose $f_t - \Delta f \geq 0$ for all $x \in D$ and $0 < t < T$. Then the minimum of f in the closed cylinder $\bar{D} \times [0, T]$ is achieved either at the "initial boundary" $t = 0$ or at the "spatial boundary" $x \in \partial D$.

It follows from the maximum principle applied to $-f$. In particular, if $f_t - \Delta f = 0$ in the cylinder then f assumes its maximum and minimum values at the spatial boundary or the initial boundary.

The proof just given for the maximum principle generalizes straightforwardly to more general linear parabolic equations, provided there is no zeroth-order term. For example: if $f_t - \sum_{i,j} \alpha_{ij}(x,t)\nabla^2_{ij} f - \sum_i \beta_i(x,t)\nabla_i f \leq 0$ then f achieves its maximum in $\bar{D} \times [0,T]$ at the initial or spatial boundary.

Uniqueness. Uniqueness of the initial-boundary-value problem in a bounded domain follows immediately from the maximum principle. Since the equation is linear, if there were two solutions with the same data then their difference would be a solution with data 0. So the main point is this:

> Suppose $f_t = \Delta f$ for $t > 0$ and $x \in D$. Assume moreover f has initial data 0 ($f(x,0) = 0$ for $x \in D$) and boundary data 0 ($f(x,t) = 0$ for $x \in \partial D$).
> Then $f(x,t) = 0$ for all $x \in D$, $t > 0$.

Indeed: the maximum and minimum of f are 0, by the maximum (and minimum) principles. So f is identically 0 in the cylinder.

To show uniqueness for the initial-value problem in all space one must work a bit harder. The problem is that we no longer have a spatial boundary – and we mean to allow solutions that grow at ∞, so the maximum of $f(x,t)$ over all $0 < t < T$ and $x \in R^n$ might well occur as $x \to \infty$. We already know, however, that it's natural to assume a growth condition of the form $|f(x,t) \leq Me^{c|x|^2}$ for some M and c. Subtracting two possible solutions, our task is thus to show the following:

> Suppose $f_t = \Delta f$ for $t > 0$ and $x \in R^n$. Assume moreover f has initial data 0 and $|f(x,t) \leq Me^{c|x|^2}$ for some M and c. Then $f(x,t) = 0$ for all $x \in R^n$, $t > 0$.

A brief simplification: we need only show that $f = 0$ for $0 < t \leq t_0$ for some $t_0 > 0$; then applying this statement k times gives $f = 0$ for $t \leq kt_0$ and we can let $k \to \infty$. Another simplification: we need only show $f \leq 0$; then applying this statement to $-f$ we conclude $f = 0$.

Here's the idea: we'll show $f \leq 0$ by applying the maximum principle not to f, but rather to

$$g(x,t) = f(x,t) - \frac{\delta}{(t_1 - t)^{n/2}} e^{\frac{|x|^2}{4(t_1 - t)}}.$$

for suitable choices of the constants δ and t_1. The space-time cylinder will be of the form $D \times [0, t_0]$ where D is a large ball and $t_0 < t_1$.

Step 1. Observe that $g_t - \Delta g = 0$. This can be checked by direct calculation. But a more conceptual reason is this: the term we've subtracted from f is a constant times the fundamental solution evaluated at ix and $t_1 - t$. The heat equation is invariant under this change of variables.

Step 2. Let D be a ball of radius r. We know from the maximum principle that the maximum of g on $D \times [0, t_0]$ is achieved at the initial boundary or spatial boundary. At the initial boundary clearly

$$g(x,0) < f(x,0) = 0.$$

At the spatial boundary we have $|x| = r$ so

$$\begin{aligned} g(x,t) &= f(x,t) - \frac{\delta}{(t_1-t)^{n/2}} e^{\frac{r^2}{4(t_1-t)}} \\ &\leq Me^{c|x|^2} - \frac{\delta}{(t_1-t)^{n/2}} e^{\frac{r^2}{4(t_1-t)}} \\ &\leq Me^{cr^2} - \frac{\delta}{t_1^{n/2}} e^{\frac{r^2}{4t_1}} \end{aligned}$$

We may choose t_1 so that $1/(4t_1) > c$. Then when r is large enough the second term dominates the first one, giving

$$g(x,t) \leq 0 \quad \text{at the spatial boundary } |x| = r.$$

We conclude from the maximum principle that $g(x,t) \leq 0$ on the entire space-time cylinder. This argument works for any sufficiently large r, so we have shown that

$$f(x,t) \leq \frac{\delta}{(t_1-t)^{n/2}} e^{\frac{|x|^2}{4(t_1-t)}}$$

for all $x \in R^n$ and all $t < t_1$. Restricting attention to $t < t_0$ for some fixed $t_0 < t_1$, we pass to the limit $\delta \to 0$ to deduce that $f \leq 0$ as desired. This completes the proof of uniqueness.