PDE for Finance Notes, Spring 2000 – Section 6.

Notes by Robert V. Kohn, Courant Institute of Mathematical Sciences. For use only in connection with the NYU course PDE for Finance, G63.2706, Spring 2000.

Announcements:

- The exam will be Tuesday May 9, at the usual class time. It will be "closed-book" (no books, no notes), however you may bring two sheets of notes (8.5×11, both sides, write as small as you like). Last year's PDE for Finance final exam will give you some idea what to expect; it is on my web page at the end of the Spring 1999 PDE for Finance lecture notes. On April 25 I'll hand out a list of possible exam topics, which should help you to prepare (or at least to be less nervous).
- There will be one more problem set, HW 6, distributed April 18 and due May 2 (no extensions – a solution sheet will be distributed May 2).

The linear heat equation and more general parabolic equations. We've seen that linear parabolic equations arise as *backward* Kolmogorov equations, determining the expected values of various payoffs (for uncontrolled diffusion processes). They also arise as *forward* Kolmogorov equations, determinining the probability distribution of the diffusing state. The simplest special cases are the backward and forward linear heat equations $u_t + \frac{1}{2}\Delta u = 0$ and $p_s - \frac{1}{2}\Delta p = 0$, which are the backward and forward Kolmogorov equations for Brownian motion. Many features of the general case can be seen especially clearly in this special case. This section discusses some fundamental properties of the linear heat equation and more general linear parabolic equations. Standard references include section 2.3 of L.C. Evans' book (on reserve) and chapter 7 of F. John's book (on reserve). Another excellent source (more elementary than Evans or John, and cheaper too) is R.B. Guenther and J.W. Lee, Partial Differential Equations of Mathematical Physics and Integral Equations (Dover reprint, 1996), chapter 5.

Our attention is restricted to *linear* parabolic equations. This class includes the forward and backward Kolmogorov equations associated with a stochastic differential equation. However it does not include the Hamilton-Jacobi equation of a stochastic control problem, which is *nonlinear*. The analysis of such nonlinear parabolic equations lies beyond the scope of this course.

****

Parabolic differential equations. The general linear parabolic differential equation in one space dimension has the form

$$f_t = \alpha(x,t) f_{xx} + \beta(x,t) f_x + \gamma(x,t) f + \delta(x,t)$$

with $\alpha(x,t) > 0$. The initial value $f(x,t_0)$ must be specified (and also the boundary data if x is restricted to a an interval or a half-space). The equation then determines $f(x,t)$ for $t > t_0$. The analogous multidimensional problem is

$$\frac{\partial f}{\partial t} = \sum_{i,j} \alpha_{ij}(x,t) \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_i \beta_i(x,t) \frac{\partial f}{\partial x_i} + \gamma(x,t) f + \delta(x,t) \tag{1}$$

where $\alpha_{ij}(x,t)$ is a positive definite matrix. We shall always assume, without explicit mention, that f is smooth enough for all terms in the PDE to make sense and be continuous (thus f is at least C^1 in time and C^2 in space). Parabolic equations have a regularizing property (provided $\alpha > 0$), so less regular solutions can occur only if the coefficients α, β, etc. are themselves irregular.

Explicit solution formulas are available only in very special cases – for example when α, β and γ are independent of x and t. However the solution can be found numerically for any (reasonable) choices of α, β, γ, and δ. Moreover the qualitative behavior of solutions – and the behavior of numerical solution schemes – is largely captured by the simplest special case, the linear heat equation

$$f_t = \Delta f$$

with the usual notation $\Delta f = \partial^2 f/\partial x_1^2 + \ldots + \partial^2 f/\partial x_n^2$. Therefore we shall concentrate most of our attention on this special case.

A reminder why we care. For a diffusion described by the stochastic differential equation

$$dy_i = f_i(y,t)dt + \sum_j g_{ij}(y,t) dw_j$$

the backward Kolmogorov equation for $u(x,t)$ is

$$\frac{\partial u}{\partial t} + \sum_i f_i(x,t) \frac{\partial u}{\partial x_i} + \sum_{i,j} a_{ij}(x,t) \frac{\partial^2 u}{\partial x_i \partial x_j} = 0$$

and the forward Kolmogorov equation for $p(z,s)$ is

$$\frac{\partial p}{\partial s} + \sum_i \frac{\partial}{\partial z_i}(f_i(z,s)p) - \sum_{i,j} \frac{\partial}{\partial z_i \partial z_j}(a_{ij}(z,s)p) = 0$$

where

$$a_{ij} = \tfrac{1}{2} \sum_k g_{ik} g_{jk}.$$

Notice that $a = gg^T$ is always positive semidefinite, and it is positive definite exactly if the rows of the matrix g_{ij} are linearly independent.

The backward Kolmogorov equation is almost in the form (1), as we see by rewriting it as $u_t = -\sum a_{ij} \nabla^2_{ij} u - \sum f_i \nabla_i u$. The second-order term has a minus sign, whereas the corresponding term in (1) has a plus sign, but this is easily corrected by the change of variables $t = -\tau$. So the backward Kolmogorov equation is a linear parabolic equation "running backward in time" – and its natural problems are final-value problems rather than initial-value problems.

The forward Kolmogorov equation can be put in the form (1) by carrying out the differentiations. (For example: writing $\partial(f_i p)/\partial z_i = p(\partial f_i/\partial z_i) + f_i(\partial p/\partial z_i)$.) Thus it is special case of (1) provided that the drift f_i and volatility g_{ij} are sufficiently smooth functions of space.

The preceding two paragraphs assume that $a = gg^T$ is nonsingular, i.e. that the matrix g has independent rows. This means, roughly speaking, that in the stochastic differential equation no component of y behaves deterministically. Not every financial model has this property; it is sometimes natural to treat some state variables deterministically and others stochastically. Such problems lead to *degenerate* parabolic equations, whose analysis is more subtle than the strictly parabolic case considered here.

The initial-value problem for the linear heat equation. Consider the equation

$$f_t = \Delta f \quad \text{for } x \in R^n, t > 0$$

with specified data $f(x, 0) = f_0(x)$. Here are some basic facts:

(a) There is an explicit solution formula

$$f(x,t) = (4\pi t)^{-n/2} \int e^{-|x-y|^2/4t} f_0(y)\, dy. \tag{2}$$

(b) This is the unique solution, among functions f with reasonable growth as $|x| \to \infty$.

(c) The solution satisfies a maximum principle.

(d) The solution is smooth for all $t > 0$, even if f_0 is not smooth.

(e) It is essential that we solve this equation forward (not backward) in time.

Concerning (a). We know to expect a solution formula of this type, because the PDE is the forward Kolmogorov equation associated to $\sqrt{2}w$ where w is Brownian motion. The solution formula for $f_t = \alpha\Delta f$ with α constant is easily obtained from (2) by change of variables; it is

$$f(x,t) = (4\pi\alpha t)^{-n/2} \int e^{-|x-y|^2/4\alpha t} f_0(y)\, dy.$$

Taking $\alpha = 1/2$ we obtain this interpretation of (2): the probability of a Brownian particle being at y in time t, given that it started at x at time 0, is $(2\pi t)^{-n/2} \int e^{-|x-y|^2/2t}$.

How could we have found the solution formula? It is immediately clear from the definition of Brownian motion, according to which $w(t)$ is a Gaussian random variable with mean 0 and variance t. Viewing $f_t - \frac{1}{2}\Delta f = 0$ as a forward Kolmogorov equation, we see that for given t, $f(\cdot, t)$ is the probability density of a Gaussian random variable with mean 0 and variance t. Hence the formula for f.

There are various other, non-probabilistic ways of guessing the solution formula. One uses the Fourier transform (which turns constant-coefficient PDE's into ODE's); see John section 7.1 or Evans section 4.3.

What must we assume concerning the initial data f_0? Clearly we need some restriction on the growth of f_0 at ∞, to make the integral on the right hand side of (2) converge. For example, if $f_0(x) = \exp(c|x|^2)$ with $c > 0$ then the integral diverges for $t > (4c)^{-1}$. The natural growth condition is thus

$$|f_0(x)| \leq Me^{c|x|^2} \tag{3}$$

as $|x| \to \infty$. Are there other restrictions on f_0? Basically no. The justification of this statement involves proving that the proposed solution $f(x,t)$, given by (2), does have the desired "initial value" $f_0(x)$, i.e. $\lim_{t\to 0} f(x,t) = f_0(x)$. Most textbooks prove this assuming f_0 is continuous, but the standard proof works more generally, e.g. if f_0 is just piecewise continuous. (See e.g. John or Evans for this argument.)

Solutions growing at infinity are uncommon in physics but common in finance, where the heat equation arises by a logarithmic change of variables from the Black-Scholes PDE (see e.g. Wilmott-Howison-Dewynne). The payoff of a call is linear in the stock price s as $s \to \infty$. This leads under the change of variable $x = \log s$ to a choice of f_0 which behaves like e^x as $x \to \infty$. Of course this lies well within what is permitted by (3). Discontinuous solutions are also uncommon in physics, but common in finance. A digital option pays a specified value if the stock price at maturity is greater than a specified value, and nothing otherwise. This corresponds to a discontinuous choice of f_0.

Concerning (b). Thus far we have only really argued that (2) gives *a* solution of the heat equation. To show it gives *the* solution we must demonstrate uniqueness. By linearity this amounts to showing that

$$\text{if } f_t = \Delta f \text{ for } t > 0, \text{ and } f(x,0) = 0, \text{ then } f(x,t) = 0 \text{ for all } x, t.$$

We shall show this under the additional assumption that f satisfies the natural growth condition (3). (It is false without some such hypothesis; see John for a counterexample.) The argument rests on the maximum principle, so we postpone it till a bit later.

Concerning (c). The maximum principle 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" $x \in \partial D$.

If $f_t - \Delta f$ were *strictly* negative this 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 asymmetry between the initial and final boundaries is one piece of evidence that time has a "preferred direction" for a parabolic differential equation.

Our proof of 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.

Returning to (b). 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 $x \to \infty$. Subtracting two possible solutions, our task is 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.

Concerning (d). The smoothness of solutions is immediately evident by differentiating under the integral in the solution formula (2). With slightly more work one can show that f is in fact real-analytic for $t > 0$. The point, of course, is that the fundamental solution $K(x, y; t) = (4\pi t)^{-n/2} e^{-|x-y|^2/4t}$ is a smooth (even analytic) function of x, y, t – though it gets more and more singular as $t \to 0$. Smoothness of the fundamental solution is a general feature of (uniformly) parabolic operators; of course the proof is more difficult when we don't have an explicit solution formula to point to.

Concerning (e). Is it possible to solve the heat equation with time running "the wrong way"? Clearly no, in general: by (d), the "wrong way" problem

$$\text{WRONG WAY} \quad f_t - \Delta f = 0 \text{ for } t < T, \text{ with } f(x,T) = f_T(x) \quad \text{WRONG WAY}$$

has no solution unless f_T is smooth. Of course it *can* have a solution for special choices of f_T – for example we may choose f_T by solving an initial-value problem up to time T. In such a case, the solution may exist for some interval $t \in (t_{\min}, T)$ but it will cease to exist at some time $t_{\min}$. (A bounded solution of the heat equation that exists for all negative time must be constant.)

Here's another way to see that solving the heat equation forward in time is good, while solving it backward in time is bad. Consider the initial-boundary-value problem on the unit interval $0 < x < 1$, with $f = 0$ at the spatial boundary ($f(0,t) = f(1,t) = 0$). It is natural to restrict attention to initial data f_0 satisfying the same boundary conditions. Such f_0 can be represented as a Fourier sine series:

$$f_0(x) = \sum_{k=1}^{\infty} a_k \sin(k\pi x).$$

The solution of $f_t - f_{xx} = 0$ with this initial data is

$$f(x,t) = \sum_{k=1}^{\infty} a_k e^{-k^2\pi^2 t} \sin(k\pi x). \tag{4}$$

It clearly exists for all positive time, and decays to 0 as $t \to \infty$. Moreover $f(x,t)$ is smooth as soon as $t > 0$, since its $2i$th spatial derivative has a Fourier series

$$D^{2i} f(x,t) = \sum_{k=1}^{\infty} a_k e^{-k^2\pi^2 t} (-1)^i k^{2i} \pi^{2i} \sin(k\pi x)$$

and the sum $\sum_k a_k^2 k^{4i} e^{-2k^2\pi^2 t}$ is finite for any $t > 0$. (We need assume only that $\sum a_k^2 < \infty$, i.e. f_0 is in L^2. For any fixed $t > 0$ the weight $k^{4i} e^{-2k^2\pi^2 t}$ is less than 1 once k is sufficiently large. This argument shows that $D^{2i} f$ is in L^2 for all i; this implies that f is smooth in the conventional sense.)

The preceding explicit solution can also be used backward in time – if the series converges. Evidently as t decreases the kth frequency blows up exponentially fast – and higher frequencies blow up faster. Thus solving the heat equation backward in time is very unstable: the high-frequency component of the final-time data is amplified very rapidly, though it may contribute negligibly to the final-time data in any standard norm.

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. (The task of fixing the Hubble telescope's pictures was more complicated – and more nonlinear – but not entirely unlike this.)

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, and which may not even have a solution except for very special data.

Boundary value problems and a numerical solution scheme. When a parabolic equation is solved in a bounded spatial domain one must supply boundary data as well as initial data. In view of Section 5 it is natural to consider specifying $f(x,t)$ for x on the spatial boundary. (This is one acceptable type of boundary condition, but by no means the only one.) The maximum principle assures uniqueness in this setting, but some other argument is needed to see existence. Let us sketch how the solution can be constructed using a simple (explicit, finite-difference) numerical approximation scheme. We focus for simplicity on the linear heat equation $f_t = f_{xx}$ with the unit interval $0 < x < 1$ as our spatial interval. If the timestep is Δt and the spatial length scale is Δx then the numerical f is defined at $(x,t) = (j\Delta x, k\Delta t)$. The explicit finite difference scheme determines f at time $(j+1)\Delta t$ given f at time $j\Delta t$ by reading it off from

$$\frac{f((j+1)\Delta t, k\Delta x) - f(j\Delta t, k\Delta x)}{\Delta t} = \frac{f(j\Delta t, (k+1)\Delta x) - 2f(j\Delta t, k\Delta x) + f(j\Delta t, (k-1)\Delta x)}{(\Delta x)^2}.$$

Notice that we use the initial data to get started, and we use the boundary data when $k\Delta x$ is next to the boundary.

This method has the stability restriction

$$\Delta t < \frac{1}{2}(\Delta x)^2. \tag{5}$$

To see why, observe that the numerical scheme can be rewritten as

$$f((j+1)\Delta t, k\Delta x) = \tfrac{\Delta t}{(\Delta x)^2} f(j\Delta t, (k+1)\Delta x) + \tfrac{\Delta t}{(\Delta x)^2} f(j\Delta t, (k-1)\Delta x) + (1 - 2\tfrac{\Delta t}{(\Delta x)^2}) f(j\Delta t, k\Delta x).$$

If $1 - 2\frac{\Delta t}{(\Delta x)^2} > 0$ then the scheme has a discrete maximum principle: if $f \le C$ initially and at the boundary then $f \le C$ for all time; similarly if $f \ge C$ initially and at the boundary then $f \ge C$ for all time. The proof is easy, arguing inductively one timestep at a time. (If the stability restriction is violated then the scheme is unstable, and the discrete solution can grow exponentially.)

One can use this numerical scheme to prove existence (see e.g. John). But let's be less ambitious: let's just show that the numerical solution converges to the solution of the PDE as Δx and Δt tend to 0 while obeying the stability restriction (5). The main point is that the scheme is consistent, i.e.

$$\frac{g(t+\Delta t, x) - g(t,x)}{\Delta t} \to g_t \quad \text{as } \Delta t \to 0$$

and

$$\frac{g(t, x+\Delta x) - 2g(t,x) + g(t, x-\Delta x)}{(\Delta x)^2} \to g_{xx} \quad \text{as } \Delta x \to 0$$

if g is smooth enough. Let f be the numerical solution, g the PDE solution, and consider $h = f - g$ evaluated at gridpoints. Consistency gives

$$\begin{aligned} h((j+1)\Delta t, k\Delta x) \quad &= \quad \tfrac{\Delta t}{(\Delta x)^2} h(j\Delta t, (k+1)\Delta x) + \tfrac{\Delta t}{(\Delta x)^2} h(j\Delta t, (k-1)\Delta x) \\ &\quad + (1 - 2\tfrac{\Delta t}{(\Delta x)^2}) h(j\Delta t, k\Delta x) + \Delta t e(j\Delta t, k\Delta x) \end{aligned}$$

with $|e|$ uniformly small as Δx and Δt tend to zero. Stability – together with the fact that $h = 0$ initially and at the spatial boundary – gives

$$|h(j\Delta t, k\Delta x)| \leq j\Delta t \max |e|.$$

It follows that $h(t, x) \to 0$, uniformly for bounded $t = j\Delta t$, as Δt and Δx tend to 0.

Our argument captures, in this special case, a general fact about numerical schemes: that stability plus consistency implies convergence.