PDE for Finance Notes – Section 5 Addendum
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 1999.

More about stopping times. Then discussion of transition probabilities and the forward Kolmogorov equation. My discussion of stopping times draws mainly from Oksendal. My treatment of the forward Kolmogorov equation draws mainly from Fleming and Rishel, but also from Jonathan Goodman's lecture notes (Computational Methods in Finance, Lecture 2: Diffusions and Diffusion Equations, available on the web). Arnold's book is good for further discussion of this topic. The discrete-time analogue of a diffusion process is a Markov chain; see Jonathan Goodman's lecture notes (Computational Methods in Finance, Lecture 1: Duality and Dynamic Programming) for the analogous discussion in that setting.

The material on "Elliptic boundary-value problems" at the end of Section 5 was skipped in lecture. So was the material at the beginning of this addendum concerning stopping times. These topics will naturally not be on the final exam.

****

More on the link to elliptic boundary-value problems. Our brief discussion of elliptic boundary-value problems, at the end of Section 5, assumed that

$$\tau = \text{the first time } y(s) \text{ exits from } D$$

was finite (for almost every sample path) and moreover $E[\tau] < \infty$. How can we see that this is true? We need some hypotheses, of course; let us discuss just the simplest case: D is a bounded domain in R^n, and $y(s)$ is multidimensional Brownian motion starting at x. Recall that by Ito's lemma, $t \to \phi(y(t))$ satisfies

$$d\phi = \nabla\phi\, dw + \tfrac{1}{2}\Delta\phi\, dt \tag{1}$$

for any function ϕ. Let's apply this with $\phi(y) = |y|^2$, integrating in time up to the stopping time

$$\tau_T(x) = \min\{\tau(x), T\} = \begin{cases} \text{first time } y(s) \text{ exits from } D & \text{if less than } T \\ T & \text{otherwise.} \end{cases}$$

We get

$$\begin{aligned} E\left[|y(\tau_T(x))|^2\right] - |x|^2 &= \frac{1}{2}\int_0^{\tau_T(x)} \Delta\phi(y(s))ds \\ &= nE\left[\tau_T(x)\right] \end{aligned} \tag{2}$$

since $\Delta\phi = 2n$. Now let $T \to \infty$. The left hand side of (2) stays finite, since we're considering a *bounded* domain, and by definition $y(\tau_T(x))$ is either in D or on the boundary of D. Thus we conclude that

$$\lim_{T\to\infty} E\left[\tau_T(x)\right] < \infty.$$

It follows (using the monotone convergence theorem, from real variables) that the exit time $\tau = \lim_{T\to\infty} \tau_T$ is almost surely finite, and $E[\tau] < \infty$, for any starting point $x \in D$.

I can't bear to leave this topic without using it to establish some properties of Brownian motion.

QUESTION: Consider Brownian motion in R^n, starting at a point x with $|x| = b$. Given $r < b$, what is the probability that the path ever enters the ball of radius r centered at 0?

ANSWER: For $n = 1, 2$ this probability is 1. For $n \geq 3$ it is $(b/r)^{2-n}$.

Interpretation: Brownian motion is "recurrent" in dimensions 1 and 2 (it comes arbitrarily close to any point, infinitely often, regardless of where it starts). However Brownian motion is "transient" in dimensions 3 and higher (opposite of recurrent).

Consider first the case $n \geq 3$. We use the stopping time τ_k = first exit time from the annulus

$$D_k = \{r < |x| < 2^k r\}.$$

Since D_k is bounded, $E[\tau_k] < \infty$ and we can integrate the stochastic differential equation (1) up to time τ_k. Let's do this with the special choice

$$\phi(y) = |y|^{2-n}.$$

This ϕ solves Laplace's equation $\Delta\phi = 0$ away from its singularity at $y = 0$. (The singularity does not bother us, since we only evaluate ϕ at points $y(s) \in D_k$ and 0 does not belong to D_k.) The analogue of (2) is

$$E\left[|y(\tau_k)|^{2-n}\right] - b^{2-n} = \frac{1}{2}\int_0^{\tau_k} \Delta\phi(y(s))ds = 0.$$

If p_k is the probability that y leaves the annulus D_k at radius r, and $q_k = 1 - p_k$ is the probability that it leaves the annulus at radius $2^k r$, we have

$$r^{2-n} p_k + (2^k r)^{2-n} q_k = b^{2-n}.$$

As $k \to \infty$ this gives $p_k \to (b/r)^{2-n}$, as asserted.

The case $n = 2$ is treated similarly, using

$$\phi(y) = \log y,$$

which solves $\Delta\phi = 0$ in the plane, away from $y = 0$. Arguing as before we get

$$p_k \log(r) + q_k \log(2^k r) = \log b.$$

As $k \to \infty$ this gives $q_k \to 0$. So $p_k \to 1$, as asserted.

The case $n = 1$ is similar to $n = 2$, using $\phi(y) = |y|$.

****

Transition probabilities and the forward Kolmogorov equation. We've shown that when the state evolves according to a stochastic differential equation

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

the expected final position

$$u(x,t) = E_{y(t)=x}\left[\Phi(y(T))\right]$$

solves the backward Kolmogorov equation

$$u_t + \sum_i f_i \frac{\partial u}{\partial x_i} + \frac{1}{2}\sum_{i,j,k} g_{ik}g_{jk}\frac{\partial^2 u}{\partial x_i \partial x_j} = 0 \text{ for } t < T, \text{ with } u = \Phi \text{ at } t = T. \tag{3}$$

We can write the backward Kolmogorov equation as

$$u_t + \mathcal{L}u = 0 \tag{4}$$

with

$$\mathcal{L}u = \sum_i f_i \frac{\partial u}{\partial x_i} + a_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j}, \tag{5}$$

where $a_{ij} = (1/2)\sum_k g_{ik}g_{jk} = (1/2)(gg^T)_{ij}$.

The solution of the stochastic differential equation is a *Markov process*, so it has a well-defined *transition probability*

$$p(z,s;x,t) = \text{probability of being at } z \text{ at time } s\text{, given that it started at } x \text{ at time } t.$$

More precisely: $p(\cdot,s;x,t)$ is the probability density of the state at time s, given that it started at x at time t. Of course p is only defined for $s > t$. To describe a Markov process, p must satisfy the Chapman-Kolmogorov equation

$$p(z,s;x,t) = \int_{R^n} p(z_1,s_1;x,t)p(z,s;z_1,s_1)\,dz_1$$

for any s_1 satisfying $t < s_1 < s$. Intuitively: the state can get from (x,t) to (z,s) by way of being at various intermediate states z_1 at a chosen intermediate time s_1. The Chapman-Kolmogorov equation calculates $p(z,s;x,t)$ by adding up (integrating) the probabilities of getting from (x,t) to (z,s) via (z_1,s_1), for all possible intermediate positions z_1.

How should we visualize p? Consider first the case when y is multidimensional Brownian motion. Then $p(\cdot,s;x,t)$ is the density of a Gaussian random variable with mean x and variance $s-t$. The graph of $z \to p(z,s;x,t)$ always has volume 1 below it (since p is a probability density); as $s \to \infty$ its maximum value tends to 0 (a Brownian particle diffuses further and further away, on average, as time increases); as $s \to t$ it becomes infinitely tall and thin (at time $s \approx t$ the Brownian particle is very close to its initial position x). The situation for a general stochastic differential equation is similar: p becomes infinitely tall and thin, concentrating at $z = x$, as $s \to t$; and if $gg^T > 0$ then the graph of p keeps spreading as $s \to 0$. Of course in the general case p does not describe a Gaussian distribution, and there is no simple formula for the mean or variance – they are simply the mean and variance of $y(s)$.

If the stochastic differential equation does not involve time explicitly, then the transition probability depends only on the "elapsed time":

if $dy = f(y)dt + g(y)dw$ with f, g depending only on y, then $p(z, s; x, t) = p(z, s - t; x, 0)$.

If the stochastic differential equation does not involve space explicitly, then the transition probability depends only on the "relative position":

if $dy = f(t)dt + g(t)dw$ with f, g depending only on t, then $p(z, s; x, t) = p(z - x, s; 0, t)$.

The initial position of a Markov process need not be deterministic. Even if it is (e.g. if $y(0) = x$ is fixed), we may wish to consider a later time as the "initial time" (for example in deriving the Hamilton-Jacobi-Bellman equation). The transition probability determines the evolution of the spatial distribution, no matter what its initial value: if $\rho_0(x)$ is the probability density of the state at time t then

$$\rho(z, s) = \int_{R^n} p(z, s; x, t)\rho_0(x)\, dx \tag{6}$$

gives the probability density (as a function of z) at any time $s > t$.

The crucial fact about the transition probability is this: it solves the *forward Kolmogorov equation* in s and z:

$$-p_s - \sum_i \frac{\partial}{\partial z_i}(f_i(z, s)p) + \frac{1}{2}\sum_{i,j,k} \frac{\partial^2}{\partial z_i \partial z_j}(g_{ik}(z, s)g_{jk}(z, s)p) = 0 \text{ for } s > t, \tag{7}$$

with initial condition

$$p = \delta_x(z) \text{ at } s = t.$$

We can write the forward Kolmogorov equation as

$$-p_s + \mathcal{L}^* p = 0 \tag{8}$$

with

$$\mathcal{L}^* p = -\sum_i \frac{\partial}{\partial z_i}(f_i p) + \sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j}(a_{ij} p). \tag{9}$$

Here $a_{ij} = (1/2)(gg^T)_{ij}$ just as before. The initial condition $p = \delta_x(z)$ encapsulates the fact, already noted, that the graph of $p(\cdot, s; x, t)$ becomes infinitely tall and thin at x as s decreases to t. The technical meaning is that

$$\int_{R^n} p(z, s; x, t)f(z)\, dz \to f(x) \text{ as } s \text{ decreases to } t \tag{10}$$

for any continuous f.

Recall that if the initial state distribution is ρ_0 then the evolving distribution is $\rho(z,s) = \int p(z,s;x,t)\rho_0(x)\,dx$. This function $\rho(z,s)$ automatically solves the forward equation (just bring the derivatives under the integral, and use that p solves it). The initial condition on p is just what we need to have $\rho(z,s) \to \rho_0(z)$ as $s \to t$. (Demonstration: multiply (10) by $\rho_0(x)$ and integrate in x to see that

$$\int \rho(z,s) f(z)\,dz = \int p(z,s;x,t) f(z)\rho_0(x)\,dz dx \to \int f(x)\rho_0(x)\,dx$$

as $s \to t$. Since this is true for every continuous f, we conclude that $\rho(z,s)$ converges [weakly] to $\rho_0(z)$ as $s \to t$.)

Please note that the forward Kolmogorov equation describes the probability distribution by solving an initial-value problem, while the backward Kolmogorov equation describes the expected final payoff by solving a final-value problem. Students familiar with pricing options via binomial trees will find this familiar. The stock prices at various nodes of a tree are determined by working forward in time; the option values at various nodes of a tree are determined by working backward in time.

Notice that the forward and backward Kolmogorov equations are, in general, completely different. There is one case, however, when they are closely related: for Brownian motion the forward equation starting at $t = 0$ is

$$p_s - \tfrac{1}{2}\Delta p = 0 \text{ for } s > 0$$

while the backward equation with final time T is

$$u_t + \tfrac{1}{2}\Delta u = 0 \text{ for } t < T.$$

In this special case the backward equation is simply the forward equation with time reversed. More careful statement: if $u(x,t)$ solves the backward equation then $\tilde{u}(z,s) = u(z, T - s)$ solves the forward equation, and conversely.

Students with a background in physical modeling will be accustomed to equations of the form $v_t = \operatorname{div}\,(a(x)\nabla v)$. Neither the forward nor the backward Kolmogorov equation has this form. Such equations are natural in physics, but not in problems from control theory and stochastic differential equations.

Testing the plausibility of the forward equation. We will explain presently why the forward equation holds. But first let's get used to it by examining some consequences and checking some special cases. Let $\rho_0(x)$ be the probability density of the state at time 0, and and consider

$$\rho(z,s) = \int p(z,s;x,0)\rho_0(x)\,dx$$

for $s > 0$. It gives the probability density of the state at time s.

Checking the integral. Since ρ is a probability density we expect that $\int \rho(z,s)\,dz = 1$ for all s. In fact, from the forward equation

$$\begin{aligned} \frac{d}{ds}\int \rho\,dz &= \int \rho_s\,dz \\ &= \int \mathcal{L}^*\rho\,dz \\ &= 0 \end{aligned}$$

since each term of $\mathcal{L}^*\rho$ is a perfect derivative.

If the stochastic differential equation has no drift then the expected position is independent of time. In general, $E[y(s)] - E[y(0)] = \int_0^s f(y(r), r)\, dr$ since the expected value of the integral dw vanishes. Thus when $f = 0$ the expected position $E[y(s)]$ is constant. Let's prove this again using the forward equation:

$$\begin{aligned} \frac{d}{ds}(\text{expected position}) &= \frac{d}{ds}\int z\rho(z,s)\, dz \\ &= \int z\rho_s(z,s)\, dz \\ &= \int z\mathcal{L}^*\rho(z,s)\, dz \\ &= 0 \quad \text{when } f = 0. \end{aligned}$$

The last step is the result of integration by parts; for example, if y is scalar valued ($dy = g(y,t)dw$) we have

$$\begin{aligned} \int z\mathcal{L}^*\rho\, dz &= \tfrac{1}{2}\int z\left(g^2\rho\right)_{zz}\, dz \\ &= -\tfrac{1}{2}\int \left(g^2\rho\right)_z\, dz \\ &= 0. \end{aligned}$$

The special case $f =$ constant, $g = 0$. If $g = 0$ then we're studying a deterministic motion. If in addition $f =$ constant then the solution is explicit and very simple: $y(t) = y(0) + ft$. Clearly

$$\text{Prob of being at } z \text{ at time } s = \text{Prob of being at } z - fs \text{ at time } 0,$$

whence

$$\rho(z,s) = \rho_0(z - fs).$$

In particular, $\rho_s + f\cdot\nabla\rho = 0$, which agrees with the forward equation (since f is constant).

Biting the bullet. Enough playing around; let's explain why the forward equation holds. The first main ingredient is the observation that

$$E_{y(t)=x}\left[\Phi(y(T))\right] = \int \Phi(z)p(z,T;x,t)\, dz. \tag{11}$$

We know how to determine the left hand side (by solving the backward equation, with final value Φ at $t = T$). This relation determines the integral of $p(\cdot,T;x,t)$ against any function Φ, for any value of x, t, T. This is a lot of information about p – in fact, it fully determines p. Our task is to make this algorithmic, i.e. to explain how p can actually be computed. (The answer, of course, is to solve the forward equation in z and s.)

The second main ingredient is the relation between $\mathcal{L}$ and $\mathcal{L}^*$. Briefly: $\mathcal{L}^*$ is the *adjoint* of $\mathcal{L}$ in the L^2 inner product. Explaining this: recall from linear algebra that if A is a linear operator on an inner-product space, then its adjoint A^* is defined by

$$\langle Ax, y\rangle = \langle x, A^* y\rangle.$$

When working in R^n we can represent A by a matrix, and A^* is represented by the transpose A^T. The situation is similar here, but our inner product space consists of all (square-integrable, scalar-valued) functions on R^n, with inner product

$$\langle v, w\rangle = \int_{R^n} v(x)w(x)\, dx.$$

We claim that

$$\langle \mathcal{L}v, w\rangle = \langle v, \mathcal{L}^* w\rangle. \tag{12}$$

When y is scalar-valued our claim says that

$$\int_R \left(f v_x + \tfrac{1}{2} g^2 v_{xx}\right) w\, dx = \int_R v\left(-(fw)_x + \tfrac{1}{2}(g^2 w)_{xx}\right)\, dx.$$

This is a consequence of integration by parts. For example, the first term on the left equals the first term on the right since

$$\int_R [fw] v_x\, dx = -\int_R [fw]_x v\, dx.$$

The second term on each side matches similarly, integrating by parts twice. Notice that f and g can depend on time as well as space; it doesn't change the argument. The proof of (12) when y is vector valued is essentially the same as the scalar case.

The third main ingredient is hiding in our derivation of the backward equation. We know from (1) that

$$E_{y(t)=x}\left[\phi(y(T), T)\right] - \phi(x,t) = E_{y(t)=x}\left[\int_t^T (\phi_s + \mathcal{L}\phi)(y(s), s)\, ds\right] \tag{13}$$

for any function $\phi(y, s)$. Our main use of this relation up to now was to choose ϕ so that the right hand side vanished, i.e. to choose ϕ to solve the backward equation. But we don't have to make such a restrictive choice: relation (13) holds for *any* ϕ.

Let's put these ingredients together. Rewriting (13) using the transition probabilities gives

$$\int_{R^n} \phi(z,T)p(z,T;x,t)\, dz - \phi(x,t) = \int_t^T \int_{R^n} (\phi_s + \mathcal{L}\phi)(z,s)p(z,s;x,t)\, dz ds. \tag{14}$$

Using (12) and doing the obvious integration by parts in time, the right hand side becomes

$$\int_t^T \int_{R^n} -\phi p_s + \phi \mathcal{L}^* p\, dz ds + \int_{R^n} \phi(z,s)p(z,s;x,t)\, dz \Big|_{s=t}^{s=T}. \tag{15}$$

This is true for all ϕ. Since the left hand side of (14) involves only the initial and final times (t and T) we conclude that

$$-p_s + \mathcal{L}^* p = 0.$$

(Rough explanation: choose $\phi(z,s) = 0$ so that $\phi = -p_s + \mathcal{L}^* p$ for $t < s < T$ but $\phi = 0$ at $s = t$ and $s = T$.) Therefore (14)-(15) reduce to

$$\int_{R^n} \phi(z,t) p(z,t;x,t)\, dz = \phi(x,t)$$

for all ϕ, which is what we mean by the initial condition "$p = \delta_x$ when $s = t$". Done!

The argument is simple; but maybe it's hard to encompass. To recapitulate its essence, let's give a new proof (using the forward equation) of the fact (known via Ito calculus) that

$$u \text{ solves the backward equation} \Longrightarrow \frac{d}{ds} E\left[u(y(s),s)\right] = 0.$$

In fact: if $\rho(z,s)$ is the probability distribution of the state at time s,

$$\begin{aligned}
\frac{d}{ds} E\left[u(y(s),s)\right] &= \frac{d}{ds} \int u(z,s)\rho(z,s)\, dz \\
&= \int u_s \rho + u \rho_s \, dz \\
&= \int u_s \rho + u \mathcal{L}^* \rho \, dz \\
&= \int u_s \rho + (\mathcal{L} u) \rho \, dz \\
&= 0
\end{aligned}$$

using in the last step our hypothesis that u solves the backward equation.

Boundary value problems. The preceding discussion concerned the backward and forward Kolmogorov equations in all space. We also considered the backward Kolmogorov equation in a bounded domain. Let's consider just the specific case when the boundary condition at ∂D is $u = 0$:

$$\begin{gathered}
u_t + \mathcal{L} u = 0 \text{ for } x \in D, t < T \\
u(x,T) = \phi(x) \text{ at } t = T \\
u(x,t) = 0 \text{ for } x \in \partial D.
\end{gathered}$$

We know that

$$u(x,t) = E_{y(t)=x}\left[\Phi(y(\tau),\tau)\right]$$

where $\tau = \tau(x)$ is the exit time from D (or T, if the path doesn't exit by time T) and

$$\Phi = 0 \text{ for } x \in \partial D;\ \Phi = \phi \text{ at the final time } T.$$

The formula for u can be written as

$$u(x,t) = \int_{R^n} \phi(z) q(z,T;x,t)\, dz$$

where

$$q(z, s; x, t) = \text{probability that the state arrives at } z \text{ at time } s,$$
$$\text{starting from } x \text{ at time } t, \text{ without hitting } \partial D \text{ first.}$$

The function $q(z, s; x, t)$ solves the forward Kolmogorov equation for $z \in D$ and $s > t$, with boundary condition $q = 0$ for $z \in \partial D$, and initial condition $q = \delta_x$. The justification is very much like the argument given above for R^n.

One thing changes significantly when we work in a bounded domain: $\int_D q(z, s; x, t)\, dz < 1$. The reason is that q gives the probability of arriving at z at time s *without hitting the boundary first.* Thus

$$1 - \int_D q(z, s; x, t)\, dz = \text{prob of hitting } \partial D \text{ by time } s, \text{ starting from } x \text{ at time } t.$$

Evidently $\int q(z, s; x, t)\, dz$ is decreasing in time. Let's check this for Brownian motion, for which $q_s - \frac{1}{2}\Delta q = 0$. We have

$$\begin{aligned} \frac{d}{ds}\int_D q(z, s; x, t)\, dz &= \int_D q_s\, dz \\ &= \tfrac{1}{2}\int_D \Delta q\, dz \\ &= \tfrac{1}{2}\int_{\partial D} \frac{\partial q}{\partial n} \\ &\leq 0. \end{aligned}$$

The inequality in the last step is a consequence of the maximum principle (to be discussed in Section 6): since $q = 0$ at ∂D and $q \geq 0$ in D we have $\partial q/\partial n \leq 0$ at ∂D, where n is the outward unit normal. (In fact $\partial q/\partial n < 0$; this is a "strong version" of the maximum principle.)