PDE for Finance Notes, Spring 2000 – Section 5.
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: Some administrative matters:

- I'll be out of town April 3-6. Therefore there will be *no class* on April 4, and I will not hold office hours on April 6.
- The final lecture will be May 2 (not, as the official NYU calendar would have it, April 25). The final exam will be May 9.
- The solutions to HW4 will be distributed 3/28. The next problem set, HW5, will also be distributed 3/28.

Links between stochastic differential equations and PDE. A stochastic differential equation, together with its initial condition, determines a diffusion process. We can use it to define a deterministic function of space and time in two fundamentally different ways:

(a) by considering the "expected value" of a suitable cost or utility, as a function of the initial position and time;

(b) by considering the probability of being in a certain state at a given time, given knowledge of the initial state and time.

Viewpoint (a) is closely related to stochastic control. In fact it's just the analysis of "trivial" control problems, where the system is stochastic but there is in fact no control. Thus we already know what to do – though of course we have some things to pull together. It leads to analysis of a special PDE associated with the stochastic process, the *backward Kolmogorov equation*. When we take discounting into consideration we get the *Feynman-Kac formula*.

Viewpoint (b) is different from (a), but not unrelated. It is in fact *dual* to viewpoint (a), in a sense that we will make precise. The evolving probability density solves a different PDE, the *forward Kolmogorov equation*. It is in fact the adjoint of the backward Kolmogorov equation.

If you know a little finance, you know that the value of a European option can be determined in two different ways: (i) as the expected discounted value of the payoff (with respect to the risk-neutral probability), and (ii) as the solution of the Black-Scholes partial differential equation. The backward Kolmogorov equation and the Feynman-Kac formula provide the mathematical framework for relating these two approaches.

The backward and forward Kolmogorov equations (and the relation between them) is discussed, for example, in Fleming and Rishel. My discussion is largely a more expository, less rigorous version of what's there.

****

Expected values and the backward Kolmogorov equation. Here's the most basic version of the story. Suppose $y(t)$ solves the scalar stochastic differential equation

$$dy = f(y,t)dt + g(y,t)dw,$$

and let

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

be the expected value of some payoff Φ at maturity time $T > t$, given that $y(t) = x$. Then u solves

$$u_t + f(x,t)u_x + \tfrac{1}{2}g^2(x,t)u_{xx} = 0 \text{ for } t < T, \text{ with } u(x,T) = \Phi(x). \tag{1}$$

Sounds familiar, right? It's just like our discussion of stochastic control – except that there is no control, hence no need to maximize over anything.

This is a special case of arguments we've done before. Let's review the explanation anyway. For any function $\phi(y,t)$, Ito's lemma gives

$$\begin{aligned} d(\phi(y(t),t)) &= \phi_y dy + \tfrac{1}{2}\phi_{yy}dydy + \phi_t dt \\ &= (\phi_t + f\phi_y + \tfrac{1}{2}g^2\phi_{yy})dt + g\phi_y dw. \end{aligned}$$

Choosing $\phi = u$, the solution of (1), we get

$$u(y(T),T) - u(y(t),t) = \int_t^T (u_t + fy_y + \tfrac{1}{2}g^2u_{yy})dt + \int_t^T gu_y dw.$$

Taking the expected value and using the PDE gives

$$E_{y(t)=x}\left[\Phi(y(T))\right] - u(x,t) = 0$$

which is precisely our assertion.

That was the simplest case. It can be jazzed up in many ways. We discuss some of them:

Vector-valued diffusion. Suppose y solves a vector-valued stochastic differential equation

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

where each component of w is an independent Brownian motion. Then

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

solves

$$u_t + \mathcal{L}u = 0 \text{ for } t < T, \text{ with } u(x,T) = \Phi(x),$$

where $\mathcal{L}$ is the differential operator

$$\mathcal{L}u(x,t) = \sum_i f_i \frac{\partial u}{\partial x_i} + \tfrac{1}{2}\sum_{i,j,k} g_{ik}g_{jk}\frac{\partial^2 u}{\partial x_i \partial x_j}.$$

The justification is just as in the scalar case, using the multidimensional version of Ito's lemma. The operator $\mathcal{L}$ is called the "infinitesimal generator" of the diffusion process $y(t)$.

The Feynman-Kac formula. We discuss the scalar case first, for clarity. Consider as above the solution of

$$dy = f(y,t)dt + g(y,t)dw$$

but suppose we are interested in a suitably "discounted" final-time payoff of the form:

$$u(x,t) = E_{y(t)=x}\left[e^{-\int_t^T b(y(s))ds}\Phi(y(T))\right] \tag{2}$$

for some specified function $b(y)$. Then u solves

$$u_t + f(x,t)u_x + \tfrac{1}{2}g^2(x,t)u_{xx} - b(x)u = 0 \tag{3}$$

instead of (1). (Its final-time condition is unchanged: $u(x,T) = \Phi(x)$.) If you know some finance you'll recognize that when y is log-normal and b is the interest rate, (3) is precisely the Black-Scholes partial differential equation.

To explain (3), we must calculate the stochastic differential $d[z_1(s)\phi(y(s),s)]$ where $z_1(s) = e^{-\int_t^s b(y(r))dr}$. The multidimensional version of Ito's lemma gives

$$d[z_1(s)z_2(s)] = z_1 dz_2 + z_2 dz_1 + dz_1 dz_2.$$

We apply this with z_1 as defined above and $z_2(s) = \phi(y(s),s)$. Ito's lemma (or ordinary differentiation) gives

$$dz_1(s) = -z_1 b(y(s))ds$$

and we're already familiar with the fact that

$$\begin{aligned} dz_2(s) &= (\phi_s + f\phi_y + \tfrac{1}{2}g^2\phi_{yy})ds + g\phi_y dw \\ &= (\phi_s + \mathcal{L}\phi)ds + g\phi_y dw. \end{aligned}$$

Notice that $dz_1 dz_2 = 0$. Applying the above with $\phi = u$, the solution of the PDE (3), gives

$$\begin{aligned} d\left(e^{-\int_t^s b(y(r))dr}u(y(s),s)\right) &= z_1 dz_2 + z_2 dz_1 \\ &= z_1\left[(u_s + \mathcal{L}u)ds + gu_y dw\right] - z_1 ubds \\ &= z_1 gu_y dw. \end{aligned}$$

The right hand side has expected value 0, so

$$E_{y(t)=x}[z_1(T)z_2(T)] = z_1(t)z_2(t) = u(x,t)$$

as asserted.

A moment's thought reveals that vector-valued case is no different. The discounted expected payoff (2) solves the PDE

$$u_t + \mathcal{L}u - bu = 0$$

where $\mathcal{L}$ is the infinitesimal generator of the diffusion y.

Running payoff. Suppose we are interested in

$$u(x,t) = E_{y(t)=x}\left[\int_t^T \Psi(y(s),s)ds\right]$$

for some specified function Ψ. Then u solves

$$u_t + \mathcal{L}u + \Psi(x,t) = 0.$$

The final-time condition is $u(x,T) = 0$, since we have included no final-time term in the "payoff." The proof is hardly different from before: by Ito's lemma,

$$\begin{aligned} d[u(y(t),t)] &= (u_t + \mathcal{L}u)dt + \nabla u \cdot g \cdot dw \\ &= -\Psi(y(t),t)dt + \nabla u \cdot g \cdot dw. \end{aligned}$$

Integrating and taking the expectation gives

$$E_{y(t)=x}\left[u(y(T),T)\right] - u(x,t) = E_{y(t)=x}\left[-\int_t^T \Psi(y(s),s)ds\right].$$

This gives the desired assertion, since $u(y(T),T) = 0$.

Boundary value problems and exit times. The preceding examples use stochastic integration from time t to a fixed time T, and they give PDE's that must be solved for all $x \in R^n$. It's also interesting to consider integration from time t to the first time y exits from some specified region. The resulting PDE must be solved on this region, with suitable boundary data.

Let D be a region in R^n. Suppose y is an R^n-valued diffusion solving

$$dy = f(y,s)ds + g(y,s)dw \text{ for } s > t, \text{ with } y(t) = x$$

with $x \in D$. Let

$$\begin{aligned} \tau(x) &= \text{the first time } y(s) \text{ exits from } D, \text{ if} \\ &\quad\ \text{prior to } T; \text{ otherwise } \tau(x) = T. \end{aligned}$$

This is an example of a "stopping time" (key feature: the statement "$\tau(x) < t$" is $\mathcal{F}_t$-measurable; in other words, knowledge of events up to time t determines whether or not the process has exited from D before time t). Suppose we are interested in

$$u(x,t) = E_{y(t)=x}\left[\int_t^{\tau(x)} \Psi(y(s),s)ds + \Phi(y(\tau(x)),\tau(x))\right].$$

Then u solves

$$u_t + \mathcal{L}u + \Psi = 0 \text{ for } x \in D$$

with boundary condition

$$u(x,t) = \Phi(x,t) \text{ for } x \in \partial D \tag{4}$$

and final-time condition

$$u(x,T) = \Phi(x,T) \text{ for all } x \in D. \tag{5}$$

The justification is entirely parallel to our earlier examples. The only change is that we integrate, in the final step, to the stopping time τ rather than the final time T. (This is permissible for any stopping time satisfying $E[\tau] < \infty$.)

There's something slightly misleading about our notation in (4)-(5). We use the same notation Φ for both the boundary condition (4) and the final-time condition (5) because they come from the same term in the utility: $y(\tau(s))$ where τ is the time the curve $(y(s), s)$ exits from the cylinder $D \times [0,T]$. But Φ should be thought of as representing two *distinct* functions – one at the spatial boundary $\partial D \times [0,T]$, the other at the final time boundary $D \times \{T\}$ (see the figure). These two functions need have nothing to do with one another.

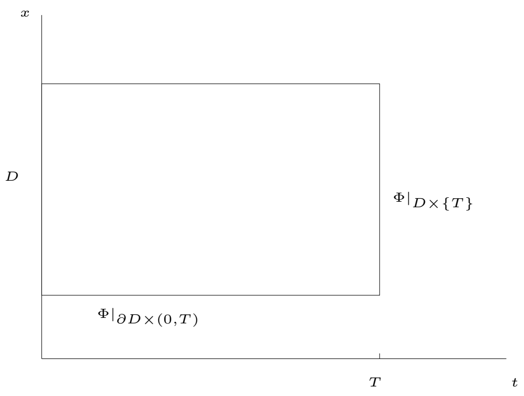

Figure 1: *Distinguishing between the two different parts of* Φ.

Often one is chosen to be zero, while the other is nontrivial. [A financial example: when one values a barrier option using the risk-neutral expectation of the payoff, Φ is zero at the knock-out price, and it equals the payoff at the maturity time.]

Elliptic boundary-value problems. Now suppose f and g in the stochastic differential equation don't depend on t, and for $x \in D$ let

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

(Unlike the previous example, we do not impose a final time T). Suppose furthermore the process does eventually exit from D, (more precisely: assume $E[\tau(x)] < \infty$, for all $x \in D$). Then

$$u(x) = E_{y(0)=x}\left[\int_0^{\tau(x)} \Psi(y(s))ds + \Phi(y(\tau(x)))\right]$$

solves

$$\mathcal{L}u + \Psi = 0 \text{ for } x \in D,$$

with boundary condition

$$u = \Phi \text{ for } x \in \partial D.$$

The justification is again entirely parallel to our earlier examples. Notice the analogy with the "least arrival time" problems of deterministic optimal control.

Application: some properties of the Brownian motion process. Let us use these results to deduce – by solving appropriate PDE's – some properties of the Brownian motion process. (This discussion is taken from Wilmott's (new) book and Oksendal's example 7.4.2. See also Oksendal's exercises 7.4 and 7.9 for related material).

QUESTION 1. Consider n-dimensional Brownian motion starting at x. What is the mean time it takes to exit from a ball of radius R, for $R > |x|$? Answer: apply the last example with $f = 0$, $g =$ identity matrix, $\Psi = 1$, $\Phi = 0$. It tells us the mean exit time is the solution $u(x)$ of

$$\tfrac{1}{2}\Delta u + 1 = 0$$

in the ball $|x| < R$, with $u = 0$ at $|x| = R$. The (unique) solution is

$$u(x) = \frac{1}{n}(R^2 - |x|^2).$$

(To do this calculation we must know in advance that the expected exit time is finite. We'll justify this as Question 3 below.)

QUESTION 2. Consider the scalar lognormal process

$$dy = \mu y dt + \sigma y dw$$

with μ and σ constant. Starting from $y(0) = x$, what is the mean exit time from a specified interval (a, b) with $a < x < b$? Answer: the mean exit time $u(x)$ solves

$$\mu x u_x + \tfrac{1}{2}\sigma^2 x^2 u_{xx} + 1 = 0 \text{ for } a < x < b$$

with boundary conditions $u(a) = u(b) = 0$. The solution is

$$u(x) = \frac{1}{\frac{1}{2}\sigma^2 - \mu}\left(\log(x/a) - \frac{1 - (x/a)^{1-2\mu/\sigma^2}}{1 - (b/a)^{1-2\mu/\sigma^2}}\log(b/a)\right)$$

(readily verified by checking the equation and boundary conditions).

QUESTION 3: Returning to the setting of Question 1, how do we know the mean exit time is finite? Answer: assume D is a bounded domain in R^n, and $y(s)$ is multidimensional Brownian motion starting at $x \in D$. Recall that by Ito's lemma, $t \to \phi(y(t))$ satisfies

$$d\phi = \nabla\phi\, dw + \tfrac{1}{2}\Delta\phi\, dt \tag{6}$$

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

$$E\left[|y(\tau_T(x))|^2\right] - |x|^2 \;=\; \frac{1}{2}\int_0^{\tau_T(x)} \Delta\phi(y(s))ds \tag{7}$$
$$\;=\; nE\left[\tau_T(x)\right]$$

since $\Delta\phi = 2n$. Now let $T \to \infty$. The left hand side of (7) 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 $x \in D$.

QUESTION 4: 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. (Interpretation: Brownian motion is "recurrent in dimensions 1 and 2 – it comes arbitrarily close to any point, infinitely often, regardless of where it starts.) In higher dimensions the situation is different: in dimension $n \geq 3$ the probability of entering the ball of radius r is $(b/r)^{2-n}$. (Interpretation: Brownian motion is "transient" in dimension $n \geq 3$.)

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 (6) 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 (7) 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

$$u_t + \sum_i f_i \frac{\partial u}{\partial x_i} + \tfrac{1}{2}\sum_{i,j,k} g_{ik}g_{jk}\frac{\partial^2 u}{\partial x_i \partial x_j} = 0 \text{ for } t < T, \text{ with } u = \Phi \text{ at } t = T. \tag{8}$$

We can write the backward Kolmogorov equation as

$$u_t + \mathcal{L}u = 0 \tag{9}$$

with

$$\mathcal{L}u = \sum_i f_i \frac{\partial u}{\partial x_i} + \sum_{i,j} a_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j}, \tag{10}$$

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

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 $s \to \infty$. 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

$$\rho(z, s) = \int_{R^n} p(z, s; x, t)\rho_0(x)\, dx \tag{11}$$

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:

$$-p_s - \sum_i \frac{\partial}{\partial z_i}(f_i(z, s)p) + \tfrac{1}{2} \sum_{i,j,k} \frac{\partial^2}{\partial z_i \partial z_j}(g_{ik}(z, s)g_{jk}(z, s)p) = 0 \text{ for } s > t, \tag{12}$$

with initial condition

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

We can write the forward Kolmogorov equation as

$$-p_s + \mathcal{L}^* p = 0 \tag{13}$$

with

$$\mathcal{L}^* p = -\sum_i \frac{\partial}{\partial z_i}(f_i p) + \sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j}(a_{ij} p). \tag{14}$$

Here $a_{ij} = \frac{1}{2}(gg^T)_{ij}$ 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

$$\int_{R^n} p(z, s; x, t) f(z)\, dz \to f(x) \text{ as } s \text{ decreases to } t \tag{15}$$

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 (15) by $\rho_0(x)$ and integrate in x to see that

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

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. This is an *accident,* associated with the the self-adjointness of the Laplacian. The situation is different even for Brownian motion with constant drift f: then the forward equation is $p_s + f \cdot \nabla p - \frac{1}{2}\Delta p = 0$, while the backward equation is $u_t + f \cdot \nabla u + \frac{1}{2}\Delta u = 0$, and the two are not equivalent under time-reversal.

Students with a background in physical modeling will be accustomed to equations of the form $v_t = \text{div } (a(x)\nabla v)$. 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 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. (Here and below, we repeatedly integrate by parts, with no "boundary terms" at $\pm\infty$. We are implicitly assuming that ρ and its derivatives decay rapidly as $z \to \pm\infty$. This is true, provided the initial distribution ρ_0 has this property.)

If the stochastic differential equation has no drift then the expected position is independent of time. In general, $E[y(s)] - E[y(0)] = E\int_0^s f(y(r),r)\,dr$ 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

$$E_{y(t)=x}\left[\Phi(y(T))\right] = \int \Phi(z) p(z,T;x,t)\, dz. \tag{16}$$

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, will be 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

$$\langle \mathcal{L}v, w\rangle = \langle v, \mathcal{L}^* w\rangle. \tag{17}$$

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 (17) 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 this derivation that

$$E_{y(t)=x}\left[\phi(y(T),T)\right] - \phi(x,t) = E_{y(t)=x}\left[\int_t^T (\phi_s + \mathcal{L}\phi)(y(s), s)\, ds\right] \tag{18}$$

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 (18) holds for *any* ϕ.

Let's put these ingredients together. Rewriting (18) using the transition probabilities gives

$$\int_{R^n} \phi(z,T)p(z,T;x,t)\,dz - \phi(x,t) = \int_t^T \int_{R^n} (\phi_s + \mathcal{L}\phi)(z,s)p(z,s;x,t)\,dzds. \tag{19}$$

Using (17) and doing the obvious integration by parts in time, the right hand side becomes

$$\int_t^T \int_{R^n} -\phi p_s + \phi \mathcal{L}^* p\,dzds + \int_{R^n} \phi(z,s)p(z,s;x,t)\,dz\bigg|_{s=t}^{s=T}. \tag{20}$$

This is true for *all* ϕ. Since the left hand side of (19) involves only the initial and final times (t and T) we conclude that

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

Therefore (19)-(20) 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

$q(z,s;x,t) =$ probability that the state arrives at z at time s,
starting from x at time t, without hitting ∂D 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

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

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 $q \ge 0$ in D we have $\partial q/\partial n \le 0$ 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.)