Stochastic Calculus, Courant Institute, Fall 2019
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2019/StochCalc.html
Jonathan Goodman, September, 2019

Class 8, More differential equations

1 Introduction

This week we return to the subject of diffusions and partial differential equations, PDEs. One goal for the week is relations between different PDEs for the same diffusion. A highlight of this is the *duality* relation between the forward and backward equations. A related goal is qualitative properties of solutions. Qualitative properties (the solution exists, it is unique, it is differentiable, etc.) may not lead directly to solution formulas, but they are essential to understanding the nature of solutions and designing solution algorithms. Another highlight is ways to derive PDEs to evaluate things about diffusion processes and ways to find things about diffusion processes that may be used to evaluate the solution of a PDE.

2 Backward and forward equation, duality

The *backward equation*, in Section 2 (this section) is the PDE

$$\partial_t f(x,t) + Lf(x,t) = 0 \; . \tag{1}$$

Here, L is the *generator*, which is defined by

$$Lg(x) = \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \mu_{ij}(x) \partial_{x_i} \partial_{x_j} g(x) + \sum_{i=1}^{d} a_i(x) \partial_{x_i} g(x) \; . \tag{2}$$

This is related to the SDE

$$dX_t = a(X_t)dt + b(X_t)dW_t \; . \tag{3}$$

The *noise coefficient* matrix $b(x)$ is related to the infinitesimal covariance $\mu(x)$ by

$$\mu(x) = b(x)b^t(x) \; , \quad \mu_{ij}(x) = \sum_{k=1}^{m} b_{ik}(x) b_{jk}(x) \; . \tag{4}$$

Here, m is the number of sources of noise, which may be less than d, the number of components of X. The m Brownian motions (the m components of W_t) are assumed to be independent standard Brownian motions. Any correlations in the noise driving different components of X_t is built into the coefficient matrices

$b(x)$. A concrete version of the backward equation comes from putting the generator formula (2) into the abstract backward equation (1):

$$0 = \partial_t f(x,t) + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\mu_{ij}(x)\partial_{x_i}\partial_{x_j}f(x,t) + \sum_{i=1}^{d}a_i(x)\partial_{x_i}f(x,t)\ . \qquad (5)$$

The unknown function $f(x,t)$ can be the value function corresponding to a payout:

$$f(x,t) = \mathrm{E}_{x,t}[\,V(X_T)]\ . \qquad (6)$$

These facts were derived in previous classes and will be reviewed in future classes.

The *forward equation* is satisfied by the PDF of X_t. Suppose $p(\cdot,t)$ is the PDF of X_t. This satisfies (derivation below)

$$\partial_t p(x,t) = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\partial_{x_i}\partial_{x_j}[\,\mu_{ij}(x)p(x,t)] - \sum_{i=1}^{d}\partial_{x_i}[\,a_i(x)p(x,t)]\ . \qquad (7)$$

The linear operator on the right is the *adjoint* of the generator and is written L^*. The adjoint operator acts on a function $q(x)$ by

$$L^*q(x) = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\partial_{x_i}\partial_{x_j}[\,\mu_{ij}(x)q(x)] - \sum_{i=1}^{d}\partial_{x_i}[\,a_i(x)q(x)]\ . \qquad (8)$$

The abstract form of the forward equation, which corresponds to the abstract backward equation (1) is

$$\partial_t p(x,t) = L^*p(x,t)\ . \qquad (9)$$

If the *initial data*, $p(x,0)$, is given, ["Data" is the plural of "datum". A datum is one number or one piece of information. Data means a lot of information. We should say the data "are" given, but we often don't.] then the evolution equation (9) or (7) determines $p(x,t)$ for $t > 0$.

The relationship between L and L^* is called *duality*. This is a generalization of the relation between a matrix A and its transpose A^t. Duality is a simple way to derive the forward equation. We already have a derivation for the backward equation. That derivation, plus duality, gives the simplest derivation I know for the forward equation. A fancier version of duality, one with boundary conditions, will help derive boundary conditions for backward equations. One can write L in "operator form" as

$$L = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\mu_{ij}(x)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d}a_i(x)\partial_{x_i}\ . \qquad (10)$$

This expresses L as a sum of "variable coefficients" $\mu_{ij}(x)$ and $a_i(x)$ multiplying "differential operators" $\partial_{x_i}\partial_{x_j}$, which are "second order", and ∂_{x_i}, which are first

order. You get the original expression (2) by applying the operator L here to the function g. There does not seem to be a clear way to write L^* in simple operator form. Instead, the expression (8) expresses the *action* of L^* on a function q.

Here is the mathematical relation between the operators L and L^*. This relation makes it easy (see below) to derive one equation from the other. The L^2 *inner product* between functions $q(\cdot)$ and $g(\cdot)$ is

$$\langle q, g \rangle = \int_{\mathbb{R}^d} q(x)g(x)\,dx \ . \tag{11}$$

If there is just one component of X, then this is

$$\langle q, g \rangle = \int_{-\infty}^{\infty} q(x)g(x)\,dx \ .$$

This inner product is a function analogue of the vector inner product. If u and v are $n-$component column vectors, then

$$\langle u, v \rangle = u^t v = \sum_{k=1}^{n} u_k v_k \ .$$

The function version (11) is similar to this, but with the sum replaced by an integral. If L is a linear operator, the adjoint is defined implicitly by

$$\langle L^* q, g \rangle = \langle q, Lg \rangle \ . \tag{12}$$

This is supposed to be satisfied for "every" pair of functions q and g. More on the word "every" below. If you have an operator L, then for every pair q and g the number $\langle q, Lg \rangle$ is defined. The adjoint operator, L^*, is an operator so that if you instead calculate L^*q, then $\langle L^*q, g \rangle$ always is the same number. The point is to find L^*.

If A is an $n \times n$ matrix, the adjoint definition A^* is that for every pair $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$, this relation is satisfied:

$$\langle u, Av \rangle = \langle A^* u, v \rangle \ .$$

We can figure this out using matrix algebra (associativity, transpose reverses order, $A = (A^t)^t$):

$$\langle u, Av \rangle = u^t \left(Av\right) = u^t \left(A^t\right)^t v = \left(A^t u\right)^t v = \langle \left(A^t u\right), v \rangle \ .$$

This proves the relationship $A^* = A^t$. The abstract adjoint of a matrix is just the matrix transpose. [Warning: The notation A^* sometimes refers to the complex conjugate transpose of a complex matrix. In this sense, if A has entries a_{ij} then A^* has entries $\overline{a}_{ji}$. Here, A^* is the abstract adjoint of A "with respect to the inner product" $\langle u, v \rangle = u^t v$.]

The calculation of L^* from L is an exercise in integration by parts. The idea is to move all derivatives from g to q, which puts the integral $\langle q, Lg \rangle$,

which involves just $q(x)$, into the integral $\langle L^* q, g \rangle$, which just involves $g(x)$. The operator L (10) is the sum of many parts, so we consider them one by one. The term $a_i(x)\partial_{x_i}$ and the integral involving it

$$\int q(x) a_i(x) \partial_{x_i} g(x)\, dx \ .$$

We integrate by parts with respect to x_i and neglect boundary terms. The result is

$$\int q(x) a_i(x) \partial_{x_i} g(x)\, dx = -\int \left[\partial_{x_i} \left(a_i(x) q(x)\right)\right] g(x)\, dx \ .$$

Instead of q and $a_i(x)\partial_{x_i} g(x)$, we have $\partial_{x_i}\left(a_i(x)q(x)\right)$ and $g(x)$. For the diffusion terms $\mu_{ij}(x)\partial) x_i \partial_{x_j}$, it takes two integrations by parts to remove both derivatives from g.

$$\begin{aligned} \int q(x) \mu_{ij}(x) \partial_{x_i} \partial_{x_j} q(x)\, dx &= -\int \left[\partial_{x_j}(\mu_{ij}(x) q(x))\right] \left[\partial_{x_i} g(x)\right]\, dx \\ &= \int \left[\partial_{x_i} \partial_{x_j}(\mu_{ij}(x) q(x))\right] \partial_{x_i} g(x)\, dx \ . \end{aligned}$$

If you apply this reasoning to all the terms in (2), you get all the terms in (8).

These integration-by-parts arguments are a little informal in that we did not say which functions q and g are allowed and we did not say why it is possible to ignore the boundary terms. [Calculations with symbols that may not be justified are often called "formal" rather than the more proper "informal". That is because they are deductions using the forms of equations and symbols rather than their actual meanings.]

To be concrete, suppose we work in one dimension. Think about the informal computation

$$\int_{-\infty}^{\infty} q(x) a(x) \partial_x g(x)\, dx \ .$$

Think of $q(x)$ as a typical probability density for a diffusion process and $g(x)$ as a typical value function. In our examples, $q(x)$ goes to zero exponentially as $x \to \pm\infty$. The value function may not go to zero, but many of our examples grow only like a power of x as $x \to \pm\infty$. The integral over an infinite interval, sometimes called an "improper" integral (is something wrong with it?). It is defined as the limit of proper integrals

$$\int_{-\infty}^{\infty} q(x) a(x) \partial_x g(x)\, dx = \lim_{R \to \infty} \int_{-R}^{R} q(x) a(x) \partial_x g(x)\, dx$$

Integrate by parts in the proper integral and you get proper boundary terms

$$\int_{-R}^{R} q(x) a(x) \partial_x g(x)\, dx = q(x) a(x) g(x) \Big|_{-R}^{R} - \int_{-R}^{R} \partial_x \left[q(x) a(x)\right] g(x)\, dx$$

The integration by parts calculations over infinite intervals are essentially equivalent to the statement that the boundary terms go to zero as $R \to \infty$. For example, $q(x)a(x)g(x) \to 0$ as $R \to \infty$. In our examples, $q(x)$ goes to zero exponentially while $a(x)$ and $g(x)$ grow at most like a power of x. The exponential "beats" the power and the product goes to zero.

3 Deriving the forward equation from the backward equation

Suppose $f(x,t)$ is the value function (6) and $p(\cdot,t)$ is the PDF of X_t. Suppose we fix the distribution of X_0, which means fixing the time zero PDF $p(x,0)$. At the same time, we consider the value functions corresponding the "any" payout function. That means that there are many functions $f(x,t)$ and only one $p(x,t)$. We use the fact that every such f satisfies the backward equation (1) to show that p satisfies the forward equation (9). This is the *duality* argument.

For any payout V, and any intermediate time t between 0 and T, we have

$$\mathrm{E}[\, V(X_T)] = \int p(x,t) f(x,t)\, dx \; .$$

The left side (this is the main point) is independent of t, so the right side is also. Its derivative with respect to t is zero. The calculations are shorter in the abstract inner product notation:

$$\begin{aligned} 0 &= \frac{d}{dt} \langle p(\cdot,t), f(\cdot,t) \rangle \\ &= \langle \partial_t p(\cdot,t), f(\cdot,t) \rangle + \langle p(\cdot,t), \partial_t f(\cdot,t) \rangle \; . \end{aligned}$$

This calculation used the Leibnitz rule (product rule) for differentiation of the inner product (11). You can verify this by writing the inner product as an integral and using the "fact" that $\frac{d}{dt} \int (\cdots) = \int \partial_t (\cdots)$. The backward equation (1), in the form $\partial_t f(\cdot,t) = -Lf(\cdot,t)$, applies to the second term on the right. After that substitution, some algebra using the properties of inner products and adjoints:

$$\begin{aligned} 0 &= \langle \partial_t p(\cdot,t), f(\cdot,t) \rangle - \langle p(\cdot,t), Lf(\cdot,t) \rangle \\ &= \langle \partial_t p(\cdot,t), f(\cdot,t) \rangle - \langle L^* p(\cdot,t) f(\cdot,t) \rangle \\ 0 &= \langle \, [\, \partial_t p(\cdot,t) - L^* p(\cdot,t)] \, , f(\cdot,t) \rangle \; . \end{aligned}$$

This last equation is true for "every" function f. A property of inner products is that if $\langle u, v \rangle = 0$ for every v, then $u = 0$. This is clearly true if you can take $v = u$, which would give $\langle u, u \rangle = 0$. A property of inner products is that $\langle u, u \rangle > 0$ unless $u = 0$. The situation with $p(\cdot,t)$ and $f(\cdot,t)$ is more complicated because we do not know that there is a payout $V(x)$ so that $f(\cdot,t) = p(\cdot,t)$. However, it is possible find a T and a V so that the corresponding f approximates p as accurately as you want. Just take T close to t and $V(x)$ close to $p(x,t)$. The

conclusion is that $[\cdots] = 0$. This is $\partial_t p(\cdot,t) - L^* p(\cdot,t) = 0$, which is the forward equation (9).

You might wish for a more direct derivation of the forward equation. The backward equation was derived using simple Taylor expansions, combined with the infinitesimal mean and covariance. The forward equation (7) involves the same quantities, but its derivation is, almost literally, backwards. It is possible to derive the forward equation directly, but the calculations are more complicated. Besides, duality is important enough that it is worthwhile to show off what you can use it for.

4 Probability flux

The *probability flux*, also called probability *current*, is a vector quantity $F(x,t) = (F_1(x,t), \ldots, F_d(x,t))$ associated to the evolution of the probability density $p(x,t)$. Suppose $D \subset \mathbb{R}^d$ is a region, such as a solid ball or solid cube or some other shape. We use Γ to denote the boundary of D. If D is a solid ball, then Γ is the sphere. [For mathematicians, the *unit ball* is the set $\|x\| \leq 1$ and the *unit sphere* is the set $\|x\| = 1$. The sphere is the boundary of the ball.] The amount of probability in D is

$$\Pr(X_t \in D) = \int_D p(x,t)\, dx \ .$$

For diffusions, probability can enter or leave D only by crossing Γ. This is related to the fact that X_t is a continuous function of t, so X_t must cross Γ if it enters or leaves D.

The probability flux (or current) describes the "flow" of probability density. Suppose x is a point in $\mathbb{R}^d$ and dS is a little piece of a surface. Suppose $n = (n_1, \ldots, n_d)$ is the normal vector to the piece of surface and dA is its surface area. Then the amount of probability crossing dS per unit time is $F \cdot n dA$, which also may be written $n^t F\, dA$ or $\left(\sum_{j=1}^d F_j(x,t) n_j \right) dA$. The rate of change of the probability of D is the integral over the boundary of D

$$\frac{d}{dt} \int_D p(x,t)\, dx = - \int_\Gamma F^t(x,t) n(x) dA(x) \ . \tag{13}$$

Explanations: The minus sign on the right side assumes that $n(x)$ is the outward facing normal to Γ. Then $F^t n > 0$ if F also points out of D. If flux (also called current) is flowing out of D, then the probability of being inside D is decreasing. In that situation (outward pointing flux, decreasing probability), the signs of the two sides agree. The integral on the right involves the element of surface area dA. In d dimensions, this is a $d-1$ dimensional "area". For $d-2$, D is two dimensional and Γ is one dimensional (a curve). In that case dA is a unit of arc length on the curve. In four or more dimensions it may be hard to visualize. The boundary of a four dimensional body is a three dimensional surface.

It might help your intuition about probability, and the flow of probability, to think about probability *density*. One meaning of "density" is the number of

particles per unit length, area, volume, etc. Imagine dust in air with a density of 6 dust particles per cubic centimeter. This means that in a region that is 10 cm on a side (volume $= 1000$ cm), there are around 6 (particles/cm^3) $\times$ 1000(cm^3) $= 6000$ particles. A probability density measures probability per unit volume. You can turn this (approximately) into a number density by imagining a large number of independent samples from a probability density. Then the number of "particles" in a region is (approximately) proportional to the probability of the region. In a diffusion process, the noise and drift keep these particles in constant motion. If you have a piece of surface dS, some particles will cross from right to left and some from left to right (some cross one way, others cross the opposite way). The net number crossing (right crossings minus left crossings) is proportional to to the probability flux across the surface, which is the product of F (flux), n (unit normal to dS), and dA (the area). If there is no noise, then all the particles near a point x move by approximately $a(x)dt$ in time dt. If there is noise and the particles move independently, then each particle near x moves to a different nearby spot in time dt. The net flux resulting from this random motion is F.

The divergence theorem gives the formula

$$\int_\Gamma F^t(x)n(x)\,dA(x) = \int_D \mathrm{div}(F)(x)dx = \int_D \sum_{j=1}^{d} \partial_{x_j} F_j(x,t)\,dx\ .$$

Therefore, the integral probability flux formula (13) is equivalent to

$$\int_D \partial_t p(x,t)\,dx = -\int_D \sum_{j=1}^{d} \partial_{x_j} F_j(x,t)\,dx\ .$$

This formula is true for every domain D. Therefore, it must be true *pointwise*, which means that the integrands must be equal for every x and t. Thus, the integral formula (13) is equivalent to the local flux formula

$$\partial_t p(x,t) = -\sum_{j=1}^{d} \partial_{x_j} F_j(x,t)\ . \tag{14}$$

But we already have a formula for $\partial_t p(x,t)$, the forward equation (7). These are consistent only with the flux formula

$$F_j(x,t) = -\frac{1}{2}\sum_{i=1}^{d} \partial_{x_j}[\,\mu_{ij}(x)p(x,t)] + a_j(x)p(x,t)\ . \tag{15}$$

Thus, the forward equation (7) may be formulated as saying that the probability density $p(x,t)$ is the density of a quantity that satisfies a local *conservation law* (14) with a flux (or current) given by a generalized *Fick's law* (or *Fourier law*) (15).

The two terms in the flux formula (15) have different interpretations. For one dimensional Brownian motion, the first term simplifies to $-\frac{1}{2}\partial_x p(x,t)$. We

used this formula in an earlier class when we were talking about hitting times. It says that probability flows "downhill" from regions where p is large to regions where it is smaller. If there is a high density of particles on one side of Γ and a low density of particles on the other side, then more particles will cross from the high density side to the low density side than will go the other way. The net flux will be from high density to low density. If the infinitesimal variance μ is constant, then the rate of flux is proportional to the infinitesimal variance – faster diffusing particles move faster downhill the probability gradient. If μ is not a constant, you must put $\mu(x)$ inside the derivative. The flux involves $\partial_x(\mu p)$ rather than $\mu(x)\partial_x p$. This may be understood from the martingale property of diffusions (see below), but I don't have a simple explanation here. The part of the flux that comes from the noise term in the SDE is the *diffusive flux* (related to diffusion) and is

$$F_j^{\rm dif}(x,t) = -\frac{1}{2}\sum_{i=1}^{d} \partial_{x_j}[\,\mu_{ij}(x)p(x,t)] \ .$$

The other term in the flux is the *advective* flux

$$F_j^{\rm adv}(x,t) = a_j(x)p(x,t) \ .$$

This is the flux that would remain if the noise coefficient in the SDE were zero. In that case, a particle would be carried by the drift velocity $a(x)$. The flux would be equal the drift velocity (which is $a(x)$) multiplied by the probability density. *Advection* and *convection* (I'm not sure what the difference is) refer to heat or a chemical being carried along by a moving fluid (air or water). Without noise, the the "diffusing" particles (with diffusion coefficient equal to zero) are simply advected by the drift velocity. The full probability flux is the sum of the diffusive flux, coming from the noise in the SDE, and the advective flux.

The probability flux is helpful for defining boundary conditions. If a diffusion process is confined to a region of space, then the probability flux across the boundary must equal to zero. [Warning: this statement is true in a simple way in one dimension and for many multi-dimensional problems. For other multi-dimensional problems it is possible to have a "boundary flux" that carries probability along the boundary. This flux does not act in the interior but does play a role in the boundary condition for p. It is sometimes called an *oblique* boundary condition.]

5 Properties of the forward and backward equations

The forward equation (7) and the backward equation (5) have differences that may be easier to remember if you understand some of their implications. One difference is the sign of the diffusion term. For one dimensional Brownian motion, the backward equation becomes $\partial_t f = -\frac{1}{2}\partial_x^2 f$ and the forward equation

is $\partial_t p = \frac{1}{2} p$. The minus sign in the backward equation makes it "well posed" (discussed next week) for propagating the value function backward in time. The plus sign in the forward equation makes that one well posed for propagating the probability density forward in time. The signs of the advection terms also have interpretations, which are explored in next week's assignment.

Another difference arises only when the infinitesimal mean and covariance are not constant These are the "coefficients" in the forward and backward equations. They are differentiated in the forward equation but not the backward equation. You can see why the coefficients are not differentiated in the backward equation by asking what the the value function $f(x,t)$ must be if the payout $V(x) = v$ is constant. In that case, obviously, $f = \mathrm{E}[v] = v$, no matter what $\mu_{ij}(x)$ and $a_j(x)$ are. The backward equation handles this by first differentiating f and then multiplying by coefficients. If $\partial_{x_j} f = 0$ for all j (f is constant "in space"), then $\partial_t f = 0$ also. Constant "functions" satisfy the backward equation.

The forward equation has its derivatives on the "outside", which is seen clearly in the flux formula (14). This enables the probability flux interpretation of the forward equation. If you integrate the forward equation in any of its forms (7) or (14), you get

$$\frac{d}{dt} \int_{\mathbb{R}^d} p(x,t) dx = 0 \; .$$

This is consistent with the fact that the integral on the left side is equal to 1 for all t.

But why are there two derivatives on the outside in the diffusion term in the forward equation? This is consistent with the fact that the diffusion process is a martingale. Without drift, the SDE for a diffusion process is

$$dX_t = b(X_t) dW_t \; .$$

With no dt part, this describes a martingale (more on this point in coming weeks). Therefore

$$\frac{d}{dt} \mathrm{E}[\, X_t] = \frac{d}{dt} \int_{\mathbb{R}^d} x p(x,t) dx = 0 \; .$$

In more than one dimension x is a vector, so this is supposed to hold for each component x_k. Using the forward equation (with no advection term), we can calculate

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^d} x_k p(x,t) dx &= \int_{\mathbb{R}^d} x_k \partial_t p(x,t) dx \\ &= \frac{1}{2} \int_{\mathbb{R}^d} x_k \sum_{i=1}^{d} \sum_{j=1}^{d} \partial_{x_i} \partial_{x_j} [\, \mu_{ij}(x) p(x,t)] \; . \end{aligned}$$

As before, you integrate by parts to put the derivatives on the x_k. First move

∂_{x_j}, and you get

$$-\frac{1}{2}\int_{\mathbb{R}^d}\sum_{i=1}^{d}\sum_{j=1}^{d}(\partial_{x_j}x_k)\partial_{x_i}[\,\mu_{ij}(x)p(x,t)]\;dx\;.$$

Clearly, $(\partial_{x_j}x_k)=1$ if $j=k$ and $(\partial_{x_j}x_k)=0$ otherwise. Therefore, the sum over j just gives the $j=k$ term and the integral becomes

$$-\frac{1}{2}\int_{\mathbb{R}^d}\sum_{i=1}^{d}\partial_{x_i}[\,\mu_{ik}(x)p(x,t)]\;dx\;.$$

This integral is equal to zero because it is the integral of a derivative. If the diffusive flux involved $\mu_{ij}(x)\partial_{x_i}p(x,t)$, then after the integration by parts we would have

$$-\frac{1}{2}\int_{\mathbb{R}^d}\sum_{i=1}^{d}\mu_{ik}(x)\partial_{x_i}p(x,t)\,dx\;.$$

This is not the integral of a derivative. In fact, another integration by parts makes this

$$\frac{1}{2}\int_{\mathbb{R}^d}\sum_{i=1}^{d}[\,\partial_{x_i}\mu_{ik}(x)]\,p(x,t)\,dx\;.$$

If $\mu(x)$ is not constant, then this will not be zero in general.