Stochastic Calculus Notes, Lecture 5

Last modified October 26, 2004

1 Integrals involving Brownian motion

1.1. Introduction: There are two kinds of integrals involving Brownian motion, *time integrals* and *Ito integrals*. The time integral, which is discussed here, is just the ordinary Riemann integral of a continuous but random function of t with respect to t. Such integrals define stochastic processes that satisfy interesting backward equations. On the one hand, this allows us to compute the expected value of the integral by solving a partial differential equation. On the other hand, we may find the solution of the partial differential equation by computing the expected value by Monte Carlo, for example. The *Feynman Kac* formula is one of the examples in this section.

1.2. The integral of Brownian motion: Consider the random variable, where $X(t)$ continues to be standard Brownian motion,

$$Y = \int_0^T X(t)dt \; . \qquad (1)$$

We expect Y to be Gaussian because the integral is a linear functional of the (Gaussian) Brownian motion path X. Because $X(t)$ is a continuous function of t, this is a standard Riemann integral. The Riemann sum approximations converge. As usual, for $n > 0$ we define $\Delta t = T/n$ and $t_k = k\Delta t$. The Riemann sum approximation is

$$Y_n = \Delta t \sum_{k=0}^{n-1} X(t_k) \; , \qquad (2)$$

and $Y_n \to Y$ as $n \to \infty$ because $X(t)$ is a continuous function of t. The n summands in (2), $X(t_k)$, form an n dimensional multivariate normal, so each of the Y_n is normal. It would be surprising if Y, as the limit of Gaussians, were not Gaussian.

1.3. The variance of Y: We will start the hard way, computing the variance from (2) and letting $\Delta t \to 0$. The trick is to use two summation variables $Y_n = \Delta t \sum_{k=0}^{n-1} X(t_k)$ and $Y_n = \Delta t \sum_{j=0}^{n-1} X(t_j)$. It is immediate from (2) that $E[Y_n] = 0$ and $\text{var}(Y_n) = E[Y_n^2]$:

$$\begin{aligned}
E[Y_n^2] &= E[Y_n \cdot Y_n] \\
&= E\left[\left(\Delta t \sum_{k=0}^{n-1} X(t_k) \right) \cdot \left(\Delta t \sum_{j=0}^{n-1} X(t_j) \right) \right] \\
&= \Delta t^2 \sum_{jk} E[X(t_k)X(t_j)] \; .
\end{aligned}$$

If we now let $\Delta t \to 0$, the left side converges to $E[Y^2]$ and the right side converges to a double integral:

$$E[Y^2] = \int_{s=0}^{T} \int_{t=0}^{T} E[X(t)X(s)]dsdt \; . \tag{3}$$

We can find the needed $E[X(t)X(s)]$ if $s > t$ by writing $X(s) = X(t) + \Delta X$ with ΔX independent of $X(t)$, so

$$\begin{aligned} E[X(t)X(s) &= E[X(t)(X(t) + \Delta X)] \\ &= E[X(t)X(t)] \\ &= t \; . \end{aligned}$$

A variation of this argument gives $E[X_t X_s] = s$ if $s < t$. Altogether

$$E[X_t X_s] = \min(t, s) \; ,$$

which is a famous formula. This now gives

$$E[Y^2] = \int_{s=0}^{T} \int_{t=0}^{T} E[X_t X_s]dsdt = \int_{s=0}^{T} \int_{t=0}^{T} \min(s,t)dsdt = \frac{1}{3}T^3 \; .$$

There is a simpler and equally rigorous way to get this. Write $Y = \int_{s=0}^{T} X(s)ds$ and $\int_{t=0}^{T} X(t)dt$ so that again

$$\begin{aligned} E[Y^2] &= E\left[\int_{s=0}^{T} X(s)ds \cdot \int_{t=0}^{T} X(t)dt\right] \\ &= E\left[\int_{s=0}^{T} \int_{t=0}^{T} X(s)X(t)dtds\right] \qquad (4) \\ &= \int_{s=0}^{T} \int_{t=0}^{T} E[X(s)X(t)]dtds; . \qquad (5) \end{aligned}$$

Going from the (4) to (5) involves changing the order of integration[1]. After all, $E[\cdot]$ just represents integration over a probability space. The right side of (4) has the abstarct form

$$\int_{\omega \in \Omega} \left(\int_{s \in [0,T]} \int_{t \in [0,T]} F(\omega, s, t)dtds \right) dP(\omega) \; .$$

[1]The possibility of changing order of abstract integrals was established by the twentieth century mathematician Fubini. He proved it to be correct if the double (triple in our case) integral converges absolutely (a requirement even for ordinary Riemann integrals) and the function F is jointly measurable in all its arguments. Our integrand is nonnegative, so the result will be infinite if the integral does not converge absolutely. We omit a discussion of product measures and joint measurability.

Here $F = X(s)X(t)$, and ω is the random outcome (the whole path $X[0,T]$ here), and P represents Wiener measure. If we interchange the ordinary Riemann $dsdt$ integral with the abstract dP integral, we get

$$\int_{s\in[0,T]} \int_{t\in[0,T]} \left(\int_{\omega\in\Omega} F(\omega, s, t) dP(\omega) \right) dsdt \ ,$$

Which is the abstract form of (5).

1.4. Measurability of Brownian motion integrals: Suppose $t_1 < t_2$. Consider the integrals $U = \int_0^{t_1} X(t)dt$ and $V = \int_{t_1}^{t_2} (X(t) - X(t_1))dt$. We expect U to be measurable in $\mathcal{F}_{t_1}$ because all the X values defining U are measurable in $\mathcal{F}_{t_1}$. Similarly, all the differences defining V are independent of anything anything in $\mathcal{F}_{t_1}$. Therefore, we expect V to be independent of U. We omit the straightforward proofs of these facts, which depend on elementary properties of abstract integration.

1.5. The X_t^3 martingale: Many martingales are constructed from integrals involving Brownian motion. A simple one is

$$F(t) = X(t)^3 - 3 \int_0^t X(s)ds \ .$$

To check the martingale property, choose $t_2 > t_1$ and, for $t > t_1$, write $X(t) = X(t_1) + \Delta X(t)$. Then

$$\begin{aligned} E\left[\int_0^{t_2} X(t)ds \mid \mathcal{F}_{t_1} \right] &= E\left[\left(\int_0^{t_1} X(t)dt + \int_{t_1}^{t_2} X(t)dt \right) \Big| \ \mathcal{F}_{t_1} \right] \\ &= E\left[\int_0^{t} X(t)dt \mid \mathcal{F}_t \right] + E\left[\int_{t_1}^{t_2} \left(X(t_1) + \Delta X(t) \right) dt \mid \mathcal{F}_t \right] \\ &= \int_0^t X(t)dt + (t_2 - t_1)X(t_1) \ . \end{aligned}$$

In the last line we use the facts that $X(t) \in \mathcal{F}_{t_1}$ when $t < t_1$, and $X_{t_1} \in \mathcal{F}_{t_1}$, and that $E[\Delta X(t) \mid \mathcal{F}_{t_1}] = 0$ when $t > t_1$, which is part of the independent increments property. For the $X(t)^3$ part, we have,

$$\begin{aligned} &E\left[(X(t_1) + \Delta X(t_2))^3 \mid \mathcal{F}_{t_1} \right] \\ &\quad = E\left[X(t_1)^3 + 3X(t_1)^2 \Delta X(t_2) + 3X(t_1)\Delta X(t_2)^2 + \Delta X(t_2)^3 \mid \mathcal{F}_{t_1} \right] \\ &\quad = X(t_1)^3 + 3X(t_1)^2 \cdot 0 + 3X(t_1) E[\Delta X(t_2)^2 \mid \mathcal{F}_{t_1}] + 0 \\ &\quad = X(t_1)^3 + 3(t_2 - t_1)X(t_1) \ . \end{aligned}$$

In the last line we used the independent increments property to get $E[\Delta X(t_2) \mid \mathcal{F}_{t_1}] = 0$, and the formula for the variance of the increment to get $E[\Delta X(t_2)^2 \mid \mathcal{F}_{t_1}] = t_2 - t_1$. This verifies that $E[F(t_2) \mid \mathcal{F}_t] = F(t_1)$, which is the martingale property.

1.6. Backward equations for expected values of integrals: Many integrals involving Brownian motion arise in applications and may be "solved" using backward equations. One example is $F = \int_0^T V(X(t))dt$, which represents the total accumulated $V(X)$ over a Brownian motion path. If $V(x)$ is a continuous function of x, the integral is a standard Riemann integral, because $V(X(t))$ is a continuous function of t. We can calculate $E[F]$, using the more general function

$$f(x,t) = E_{x,t}\left[\int_t^T V(X(s))ds\right] . \tag{6}$$

As before, we can describe the function $f(x,t)$ in terms of the random variable

$$F(t) = E\left[\int_t^T V(X(s))dt \mid \mathcal{F}_t\right] .$$

Since $F(t)$ is measurable in $\mathcal{F}_t$ and depends only on future values ($X(s)$ with $s > t$), $F(t)$ is measurable in $\mathcal{G}_t$. Since $\mathcal{G}_t$ is generated by $X(t)$ alone, this means that $F(t)$ is a function of $X(t)$, which we write as $F(t) = f(X(t),t)$. Of course, this definition is a big restatement of definition (6). Once we know $f(x,t)$, we can plug in $t = 0$ to get $E[F] = F(0) = f(x_0, 0)$ if $X(0) = x_0$ is known. Otherwise, $E[F] = E[f(X(0),t)]$.

The backward equation for f is

$$\partial_t f + \frac{1}{2}\partial_x^2 f + V(x,t) = 0 , \tag{7}$$

with final conditions $f(x,T) = 0$. The derivation is similar to the one we used before for the backward equation for $E_{x,t}[V(X_T)]$. We use Taylor series and the tower property to calculate how f changes over a small time increment, Δt. We start with

$$\int_t^T V(X(s))ds = \int_t^{t+\Delta t} V(X(s))ds + \int_{t+\Delta t}^T V(X(s))ds ,$$

take the x,t expectation, and use (6) to get

$$f(x,t) = E_{x,t}\left[\int_t^{t+\Delta t} V(X(s))ds \;\Big|\; \mathcal{F}_t\right] + E_{x,t}\left[\int_{t+\Delta t}^T V(X(s))ds \;\Big|\; \mathcal{F}_t\right] . \tag{8}$$

The first integral on the right has the value $V(x)\Delta t + o(\Delta t)$. We write $o(\Delta t)$ for a quantity that is smaller than Δt in the sense that $o(\Delta t)/\Delta t \to 0$ as $\Delta t \to 0$ (we will shortly divide by Δt, take the limit $\Delta t \to 0$, and neglect all $o(\Delta t)$ terms.). The second term has

$$E_{x,t}\left[\int_{t+\Delta t}^T V(X(s))ds \mid \mathcal{F}_{t+\Delta t}\right] = F(X_{t+\Delta t}) = f(X(t+\Delta t), t+\Delta t) .$$

Writing $X(t+\Delta t) = X(t) + \Delta X$, we use the tower property with $\mathcal{F}_t \subset \mathcal{F}_{t+\Delta t}$ to get

$$E\left[\int_{t+\Delta t}^{T} V(X(s))ds \mid \mathcal{F}_t\right] = E\left[f(X_t + \Delta X, t+\Delta t) \mid \mathcal{F}_t\right] \ .$$

As before, we use Taylor expansions the conditional expectation to get first

$$f(x+\Delta X, t+\Delta t) = f(x,t) + \Delta t \partial_t f(x,t) + \Delta X \partial_x f(x,t) + \frac{1}{2}\Delta X^2 \partial_x^2 f(x,t) + o(\Delta t) \ ,$$

then

$$E_{x,t}\left[f(x+\Delta X, t+\Delta t\right] = f(x,t) + \Delta t \partial_t f(x,t) + \frac{1}{2}\Delta t \partial_x^2 f(x,t) + o(\Delta t) \ .$$

Putting all this back into (8) gives

$$f(x,t) = \Delta t V(x) + f(x,t) + \Delta t \partial_t f(x,t) + \frac{1}{2}\Delta t \partial_x^2 f(x,t) + o(\Delta t) \ .$$

Now just cancel $f(x,t)$ from both sides and let $\Delta t \to 0$ to get the promised equation (7).

1.7. Application of PDE: Most commonly, we cannot evaluate either the expected value (6) or the solution of the partial differential equation (PDE) (7). How does the PDE represent progress toward evaluating f? One way is by suggesting a completely different computational procedure. If we work only from the definition (6), we would use Monte Carlo for numerical evaluation. Monte Carlo is notoriously slow and inaccurate. There are several techniques for finding the solution of a PDE that avoid Monte Carlo, including finite difference methods, finite element methods, spectral methods, and trees. When such deterministic methods are practical, they generally are more reliable, more accurate, and faster. In financial applications, we are often able to find PDEs for quantities that have no simple Monte Carlo probabilistic definition. Many such examples are related to optimization problems: maximizing an expected return or minimizing uncertainty with dynamic trading strategies in a randomly evolving market. The Black Scholes evaluation of the value of an American style option is a well known example.

1.8. The Feynman Kac formula: Consider

$$F = E\left[\exp\left(\int_0^T V(X(t)dt\right)\right] \ . \tag{9}$$

As before, we evaluate F using the related and more refined quantities

$$f(x,t) = E_{x,t}\left[e^{\int_t^T V(X_s)ds}\right] \tag{10}$$

satisfies the backward equation

$$\partial_t f + \frac{1}{2}\partial_x^2 f + V(x)f = 0 . \tag{11}$$

When someone refers to the *Feynman Kac formula*, they usually are referring to the fact that (10) is a formula for the solution of the PDE (11). In our work, the situation mostly will be reversed. We use the PDE (11) to get information about the quantity defined by (10) or even just about the process $X(t)$.

We can verify that (10) satisfies (11) more or less as in the preceding paragraph. We note that

$$\begin{aligned}
&\exp\left\{\int_t^{t+\Delta t} V(X(s))ds + \int_{t+\Delta t}^T V(X(s))ds\right\} \\
&\quad = \exp\left\{\int_t^{t+\Delta t} V(X(s))ds\right\} \cdot \exp\left\{\int_{t+\Delta t}^T V(X(s))ds\right\} \\
&\quad = (1 + \Delta t V(X(t)) + o(\Delta t)) \cdot \exp\left\{\int_{t+\Delta t}^T V(X(s))ds\right\}
\end{aligned}$$

The expectation of the rigth side with respect to $\mathcal{F}_{t+\Delta t}$ is

$$(1 + \Delta t V(X_t) + o(\Delta t)) \cdot f(X(t+\Delta X, t+\Delta t) .$$

When we now take expectation with respect to $\mathcal{F}_t$, which amounts to averaging over ΔX, using Taylor expansion of f about $f(x,t)$ as before, we get (11).

1.9. The Feynman integral: A precurser to the Feynman Kac formula, is the *Feynman integral*[2] solution to the Schrödinger equation. The Feynman integral is not an integral in the sense of measure theory. (Neither is the Ito integral, for that matter.) The colorful probabilist Marc Kac (pronounced "Katz") discovered that an actual integral over Wiener measure (10) gives the solution of (11). Feynman's reasoning will help us derive the Girsanov formula, so we pause to sketch it.

The finite difference approximation

$$\int_0^T V(X(t))dt \approx \Delta t \sum_{k=0}^{n-1} V(X(t_k)) , \tag{12}$$

(always $\Delta t = T/n$, $t_k = k\Delta t$) leads to an approximation to F of the form

$$F_n = E\left[\exp\left(\Delta t \sum_{k=0}^{n=1} V(X(t_k))\right)\right] . \tag{13}$$

[2]The American Physicist Richard Feynman was born and raised in Far Rockaway (a neighborhood of Queens, New York). He is the author of several wonderful popular books, including *Surely You're Joking, Mr. Feynman* and *The Feynman Lectures on Physics.*

The functional F_n depends only on finitely many values $X_k = X(t_k)$, so we may evaluate (13) using teh known joint density function for $\vec{X} = (X_1, \ldots, X_n)$. The density is (see "Path probabilities" from Lecture 5):

$$U^{(n)}(\vec{x}) = \frac{1}{(2\pi\Delta t^{n/2}} \exp\left(-\sum_{k=0}^{n-1}(x_{k+1}-x_k)^2/e\Delta t\right) .$$

It is suggestive to rewrite this as

$$U^{(n)}(\vec{x}) = \frac{1}{(2\pi\Delta t^{n/2}} \exp\left[-\frac{\Delta t}{2}\sum_{k=0}^{n-1}\left(\frac{x_{k+1}-x_k}{\Delta t}\right)^2\right] . \tag{14}$$

Using this to evaluate F_n gives

$$F_n = \frac{1}{(2\pi\Delta t^{n/2}} \int R^n \exp\left[\Delta t\sum_{k=0}^{n-1}V(x_k) - \frac{\Delta t}{2}\sum_{k=0}^{n-1}\left(\frac{x_{k+1}-x_k}{\Delta t}\right)^2\right] d\vec{x} . \tag{15}$$

It is easy to show that $F_n \to F$ as $n \to \infty$ as long as $V(x)$ is, say, continuous and bounded (see below).

Feynman proposed a view of $F = \lim_{n\to\infty} F_n$ in (15) that is not mathematically rigorous but explains "what's going on". If $x_k \approx x(t_k)$, then we should have

$$\Delta t \sum_{k=0}^{n-1} V(x_k) \to \int_{t=0}^{T} V(x(t))dt .$$

Also,

$$\left(\frac{x_{k+1}-x_k}{\Delta t}\right) \approx \frac{dx}{dt} = \dot{x}(t_k) ,$$

so we should also have

$$\frac{\Delta t}{2}\sum_{k=0}^{n-1}\left(\frac{x_{k+1}-x_k}{\Delta t}\right)^2 \to \int_0^T \dot{x}(t)^2 dt .$$

As $n \to \infty$, the integral over R_n should converge to the integral over all "paths" $x(t)$. We denote this by $\mathcal{P}$ without worring about exactly which paths are allowed (continuous, differentiable, ...?). The integration element $d\vec{x}$ has the possible formal limit

$$d\vec{x} = \prod_{k=0}^{n-1} dx_k = \prod_{k=0}^{n-1} dx(t_k) \to \prod_{t=0}^{T} dx(t) .$$

Altogether, this gives the formal expression for the limit of (15):

$$F = const \int_{\mathcal{P}} \exp\left(\int_0^T V(x(t))dt - \tfrac{1}{2}\int_0^T \dot{x}(t)^2 dt\right) \prod_{t=0}^{T} dx(t) . \tag{16}$$

1.10. Feynman and Wiener integration: Mathematicians were quick to complain about (16). For one thing, the constant $const = \lim_{n\to\infty}(2\pi\Delta t)^{n/2}$ should be infinite. More seriously, there is no abstract integral measure corresponding to $\int_{\mathcal{P}}\prod_{t=0}^{T} dx(t)$ (it is possible to prove this). Kac proposed to write (16) as

$$F = \int_{\mathcal{P}} \exp\left(\int_0^T V(x(t))dt\right)\left[const\cdot\exp\left(-\tfrac{1}{2}\int_0^T \dot{x}(t)^2 dt\right)\prod_{t=0}^{T} dx(t)\right] .$$

and then interpret the latter part as Wiener measure (dP):

$$const\cdot\exp\left(-\tfrac{1}{2}\int_0^T \dot{x}(t)^2 dt\right)\prod_{t=0}^{T} dx(t) = dP(X) \tag{17}$$

In fact, we have already implicitly argued informally (and it can be formalized) that

$$\lim_{n\to\infty} U^{(n)}(\vec{x})\prod_{k=0}^{n-1} dx_k \to dP(X) \;\text{ as }\; n\to\infty .$$

These intuitive but mathematically nonsensical formulas are a great help in understanding Brownian motion. For one thing, (17) makes clear that Wiener measure is Gaussian. Its density has the form $const\cdot\exp(-Q(x))$, where $Q(x)$ is a positive quadratic function of x. Here $Q(x) = \int \dot{x}(t)^2 dt$ (and the constant is, alas, infinite). Moreover, in many cases it is possible to approximate integrals of the form $\int \exp(\phi(\vec{x}))d\vec{x}$ by e^{ϕ_*}, where $\phi_* = \max_{\vec{x}}\phi(\vec{x})$ if the ϕ is sharply peaked around its maximum. This is particularly common in "rare event" or "large deviation" problems. In our case, this would lead us to solve the *calculus of variations* problem

$$\max_x\left(\int_0^T V(x(t)dt - \tfrac{1}{2}\int_0^T \dot{x}(t)^2 dt\right) .$$

1.11. Application of Feynman Kac: The problem of evaluating

$$f = E\left[\exp\left(\int_0^T V(X_t)dt\right)\right]$$

arises in many situations. In finance, f could represent the present value of a payment in the future subject to unknown fluctuating interest rates. The PDE (11) provides a possible way to evaluate $f = f(0,0)$, either analytically or numerically.

2 Mathematical formalism

2.1. Introduction: We examine the solution formulas for the backward and forward equation from two points of view. The first is an analogy with linear algebra, with *function spaces* playing the role of vector space and *operators* playing the role of matrices. The second is a more physical picture, interpreting $G(x,y,t)$ as the *Green's function* describing the forward diffusion of a point mass of probability or the backward diffusion of a localized unit of payout.

2.2. Solution operator As time moves forward, the probability density for X_t changes, or *evolves*. As time moves backward, the value function $f(x,t)$ also evolves[3] The backward evolution process is given by (for $s > 0$, this is a consequence of the tower property.)

$$f(x,t-s) = \int G(x,y,s)f(y,t)dy \ . \tag{18}$$

We write this abstractly as $f(t-s) = G(s)f(t)$.

This formula is anologous to the comparable Markov chain formula $f(t-s) = P^s f(t)$. In the Markov chain case, s and t are integers and $f(t)$ represents a vector in R^n whose components are $f_k(t)$. Here, $f(t)$ is a function of x whose values are $f(x,t)$. We can think of P^s as an $n \times n$ matrix or as the *linear operator* that transforms the vector f to the vector $g = P^s f$. Similarly, $G(s)$ is a linear operator, transforming a function f into g, with

$$g(x) = \int_{-\infty}^{\infty} G(x,t,s)f(y)dy \ .$$

The operation is *linear*, which means that $G(af^{(1)} + bf^{(2)}) = aGf^{(1)} + bGf^{(2)}$. The family of operators $G(s)$ for $s > 0$ produces the solution to the backward equaion, so we call $G(s)$ the *solution operator* for time s.

2.3. Duhamel's principle: The *inhomogeneous* backward equation

$$\partial_t f + \partial_x^2 f = V(x,t) \ , \tag{19}$$

with *homogeneous*[4] final condition $f(x,T) = 0$ may be solved by

$$f(x,t) = E_{x,t}\left[\int_t^T V(X(t'),t'dt')\right] \ .$$

Exchanging the order of integration, we may write

$$f(x,t) = \int_{t'=t}^T g(x,t,t')dt' \ , \tag{20}$$

where

$$g(x,t,t') = E_{x,t}\left[V(X(t'))\right] \ .$$

[3]Unlike biological evolustion, this evolution process makes the solution less complicated, not more.

[4]We often say "homogeneous" to mean zero and "inhomogeneous" to mean not zero. That may be because if $V(x,t)$ is zero then it is constant, i.e. the same everywhere, which is the usual meaning of homogeneous.

This g is the expected value (at (x,t)) of a payout ($V(\cdot,t')$ at time $t'>t$). As such, g is the solution of a homogeneous final value problem with inhomogeneous final values:

$$\left.\begin{aligned} &\partial_t g + \tfrac{1}{2}\partial_x^2 g = 0 \;\text{ for }\; t<t' \;, \\ &g(x,t') = V(x,t') \;. \end{aligned}\right\} \tag{21}$$

Duhamel's principle, which we just demonstrated, is as follows. To solve the invonogeneous final value problem (19), we solve a homogeneous final value problem (21) for each t' between t and T then we add up the results (20).

2.4. Infinitesimal generator: There are matrices of many different types that play various roles in theory and computation. And so it is with operators. In addition to the solution operator, there is the *infinitesimal generator* (or simply *generator*). For Brownian motion in one dimension, the generator is

$$L = \tfrac{1}{2}\partial_x^2 \;. \tag{22}$$

The backward equation may be written

$$\partial_t f + Lf = 0 \;. \tag{23}$$

For other diffusion processes, the generator is the operator L that puts the backward equation for process in the form (23).

Just as a matrix has a transpose, an operator has an *adjoint*, written L^*. The forward equation takes the form

$$\partial_t u = L^* u \;.$$

The operator (22) for Brownian motion is *self adjoint*, which means that $L^*=L$, which is why the operator $\frac{1}{2}\partial_x^2$ is the appears in both. We will return to these points later.

2.5. Composing (multiplying) operators: If A and B are matrices, then there are two ways to form the matrix AB. One way is to multiply the matrices. The other is to *compose* the linear transformations: $f \to Bf \to ABf$. In this way, AB is the composite linear transformation formed by first applying B then applying A. We also can compose operators, even if we sometimes lack a good explicit representation for the composite AB. As with matrices, composition of operators is associative: $A(Bf) = (AB)f$.

2.6. Composing solution operators: The solution operator $G(s_1$ moves the value function backward in time by the amount s_1, which is written $f(t-s_1) = G(s_1)f(t)$. The operator $G(s_2)$ moves it back an additional s_2, i.e. $f(t-(s_1+s_2)) = G(s_2)f(t-s_1) = G(s_2)G(s_1)f(t)$. The result is to move f back by s_1+s_2 in total, which is the same as applying $G(s_1+s_2)$. This shows that for every (allowed) f, $G(s_2)G(s_1)f = G(s_2+s_1)f$,. which means that

$$G(s_2)G(s_1) = G(s_2+s_1) \;. \tag{24}$$

This is called the *semigroup property*. It is a basic property of the solution operator for any problem. The matrix anologue for Markov chains is $P^{s_2+s+1} = P^{s_2}P^{s_1}$, which is a basic fact about powers of matrices having nothing to do with Markov chains. The property (24) would be called the *group* property if we were to allow negative s_2 or s_1, which we do not. Negative s is allowed in the matrix version if P is nonsingular. There is no particular physical reason for the transition matrix of a Markov chain to be non singular.

2.7. Operator kernels: If matrix A has elements A_{jk}, we can compute $g = Af$ by doing the sum $g_j = \sum_k A_{jk} f_k$. Similarly, operator A may or may not have a *kernel*[5], which is a function $A(x,y)$ so that $g = Af$ is represented by

$$g(x) = \int A(x,y)f(y)dy \ .$$

If operators A and B both have kernels, then the composite operator has the kernel

$$(AB)(x,y) = \int A(x,z)B(z,y)dz \ . \tag{25}$$

To derive this formula, set $g = Bf$ and $h = Ag$. Then $h(x) = \int A(x,z)g(z)dz$ and $g(z) = \int B(z,y)f(y)dy$ implies that

$$h(x) = \int \left(\int A(x,z)B(z,y)dz \right) f(y)dy \ .$$

This shows that (25) is the kernel of AB. The formula is anologous to the formula for matrix multiplication.

2.8. The semigroup property: When we defined (18) the solution operators $G(s)$, we did so by specifying the kernels

$$G(x,t,s) = \frac{1}{\sqrt{2\pi s}} e^{-(x-y)^2/2s} \ .$$

According to (25). the semigroup property should be an integral identity involving G. The identity is

$$G(x,y,s_2+s_1) = \int G(x,z,s_2)G(z,y,s_1)dz \ .$$

More concretely:

$$\begin{aligned} &\frac{1}{\sqrt{2\pi(s_2+s_1)}} e^{-(x-y)^2/2(s_2+s_1)} \\ &\qquad = \frac{1}{\sqrt{2\pi(s_2)}} \frac{1}{\sqrt{2\pi(s_1)}} \int e^{-(x-z)^2/2s_2} e^{-(z-y)^2/2s_1} dz \ . \end{aligned}$$

[5]The term kernel also describes vectors f with $Af = 0$, it is unfortunate that the same word is used for these different objects.

The reader is encouraged to verify this by direct integration. It also can be verified by recognizing it as the statement that adding independent mean zero Gaussian random variables with variance s_2 and s_1 respectively gives a Gaussian with variance $s_2 + s_1$.

2.9. Fundamental solution: The operators $G(t)$ form a *fundamental solution*[6] for the problem $f_t + Lf = 0$ if

$$\partial_t G = LG \ , \quad \text{for } \ t > 0 \ , \tag{26}$$

$$G(0) = I \ . \tag{27}$$

The property (26) really means that $\partial_t\big(G(t)f\big) = L\big(Gf\big)$ for any f. If $G(t)$ has a kernel $G(x,y,t)$, this in turn means (as the reader should ckeck) that

$$\partial_t G(x,y,t) = L_x G(x,y,t) \ , \tag{28}$$

where L_x means that the derivatives on L are with respect to the x variables in G. In our case with G being the *heat kernel*, this is

$$\partial_t \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t} = \tfrac{1}{2}\partial_x^2 \tfrac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t} \ ,$$

which we have checked and rechecked.

Without matrices, we still have the identity operator: $If = f$ for all f. The property (27) really means that $G(t)f \to f$ as $t \to 0$. It is easy to verify this for our heat kernel provided that f is continuous.

2.10. Duhamel with fundamental solution operator: The g appearing in (20) may be expresses as $g(t,t') = G(t'-t)V(t')$, where $V(t')$ is the function with values $V(x,t')$. This puts (20) in the form

$$f(t) = \int_t^T G(t'-t)V(t')dt' \ . \tag{29}$$

We illustrate the properties of the fundamental solution operator by verifying (29) directly. We want to show taht (29) implies that $\partial_t f + Lf = V(t)$ and $f(T) = 0$. The latter is clear. For the former we compute $\partial_t f(t)$ by differentiating the right side of (29):

$$\partial_t \int_t^T G(t'-t)V(t')dt' = -G(t-t)V(t) - \int_t^T G'(t'-t)V(t')dt' \ ,$$

We write $G'(t)$ to represent $\partial_t G(t)$. This allows us to write $\partial_t G(t'-t) = -G'(t'-t) = -LG(t'-t)$. Continuing, the left side is

$$-V(t) - \int_t^T LG(t'-t)V(t')dt' = -V(t) - \int_t^T LG(t'-t)V(t')dt' \ .$$

[6] We have adjusted this definition from its original form in books on ordinary differential equations to accomodate the backward evolution of the backward equation. This amounts to reversing the sign of L.

If we take L outside the integral on the right, we recognize what is left in the integral as $f(t)$. Altogether, we have $\partial_t f = -V(t) - Lf(t)$. This is almost right, I just have to fix the minus sign somehow.

2.11. Green's function: Consider the solution formula for the homogeneous final value problem $\partial_t f + Lf = 0$, $f(T) = V$:

$$f(x,t) = \int G(x,y,T-t)V(y)dy \; . \tag{30}$$

Consider a special "jackpot" payout $V(y) = \delta(y - x_0)$. If you like, you can think of $V(y) = \frac{1}{\epsilon}$ when $|y = x_0| < \epsilon$ then let $\epsilon \to 0$. We then get $f(x,t) = G(x,x_0,T-t)$. The function that satisfies $\partial_t G + L_x G = 0$, $G(x,T = \delta(x - x_0)$ is called the *Greens's function*[7]. The Green's function represents the result of a *point mass* payout. A general payout can be expressed as a sum (integral) of point mass payouts as x_0 with weight $V(x_0)$:

$$V(y) = \int V(x_0)\delta(y - x_0)dx_0 \; .$$

Since the backward equation is linear, the general value function will be the weighted sum (integral) of the point mass value functions, which is the formula (30).

2.12. More generally: Brownian motion is special in that $G(x,y,t)$ is a function of $x - y$. This is because Brownian motion is translation invariant: a Brownian motion starting from any point looks like a Brownian motion starting from any other point. Brownian motion is also special in that the forward equation and backward equations are nearly the same, having the same spatial operator $L = \frac{1}{2}\partial_x^2$.

More general diffusion processes loose both these properties. The solution operator depends in a more complicated way on x and y. The backward equation is $\partial_t f + Lf = 0$ but the forward equation is $\partial_t u = L^* u$. The Green's function, $G(x,y,t)$ is the fundamental solution for the backward equation in the x,t variables with y as a parameter. It also is the fundamental solution to the forward equation in the y,t variables with x as a parameter. This material will be in a future lecture.

[7]This is in honor of a 19^{th} century Englishman named Green.