# Week 11
# Backwards again, Feynman Kac, etc.

Jonathan Goodman

November 26, 2012

## 1 Introduction to the material for the week

sec:intro

This week has more about the relationship between SDE and PDE. We discuss ways to formulate the solution of a PDE in terms of an SDE and how to calculate things about an SDE using a PDE. Informally, activity of this kind is called *Feynman Kac* in certain circles, and *Fokker Planck* in other circles. Neither name is accurate historically, but this is not a history class.

One topic is the full forward equation. We have done pieces of it, but we now do it in general for general diffusions. We derive the forward equation from the backward equation using a duality argument.

Next we discuss backward equations for multiplicative function of a stochastic process. If

$$f(x,t) = \mathrm{E}_{x,t}\left[ e^{\int_t^T V(X_s)ds} \right] , \tag{1}$$

mf

and

$$dX_t = a(X_t)dt + b(X_t)dW_t , \tag{2}$$

sde

then

$$0 = \partial_t f + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} \left(b(x)b^t(x)\right)_{ij} \partial_{x_i}\partial_{x_j} f + \sum_{i=1}^{n} a_i(x)\partial_{x_i} f + V(x)f . \tag{3}$$

fepde

One of the differences between this and Girsanov's formula from last week is that here the exponent does not have an Ito integral. The relationship between (mf 26) and (fepde 3) goes both ways. You can learn about the expectation (mf 26) by solving a PDE. Numerical PDE methods are generally more accurate than direct Monte Carlo evaluation of (mf 26), that is, if $X$ does not have more than a few components. In the other direction, you can use Monte Carlo on (mf 26) to estimate the solution of the PDE (fepde 3). This can be useful if the dimension of the the PDE larger than, say, 4 or 5.

We discuss a general principle often called *splitting*. This says that if there are two or more parts of the dynamics, then you find the differential equation describing the dynamics by adding terms corresponding to each part of the dynamics. The PDE (fepde 3) illustrates this admittedly vague principle. The quantity

$f$ is determined by three "factors" (vague term not related to, say, factor analysis in statistics): diffusion, advection, and the multiplicative functional. The dynamical equation (3) has one term for each factor. The second term $\frac{1}{2}\sum\cdots$ corresponds to diffusion, $dX = b(X)dW$. The third term corresponds to advection, $dX = a(X)dt$. The last corresponds to multiplication over a $dt$ time interval by $e^{V(X_t)dt}$. Splitting applies already to the SDE (2). The right hand side has one term corresponding to advection, $adt$, and another corresponding to diffusion $bdW$.

# 2 Backward and forward

This section has much review of things we covered earlier in the course, much earlier in some cases. It serves as review and it puts the new material into context.

Let $X_t$ is a Markov process of some kind. It could be a discrete time Markov chain, or a diffusion process, or a jump process, whatever. Let $\mathcal{S}$ be the state space. For each $t$, $X_t \in \mathcal{S}$. Take times $t < T$ and define the value function $f(x,t)$ by

$$f(X_t, t) = \mathrm{E}[\, V(X_t) \mid \mathcal{F}_t \,] \ . \tag{4}$$ vf

This definition applies to all kinds of Markov processes. The $t$ variable is discrete or continuous depending on whether the process takes place in continuous or discrete time. The $x$ variable is an element of the state space $\mathcal{S}$. For a diffusion, $\mathcal{S} = \mathbb{R}^n$, so $f$ is a function of $n$ real variables $(x_1, \ldots, x_n)$, as in $f(x_1, \ldots, x_n, t)$. Here $n$ is the number of components of $X_t$.

The *generator* of a stochastic process is called $L$. The generator is a matrix or a linear operator. Either way, the generator has the action

$$g \ \xrightarrow{L}\ Lg \ ,$$

that is linear: $ag \to aLg$, and $(g_1 + g_2) \to Lg_1 + Lg_2$. Here is the definition of the generator, as it acts on a function.[1] The discrete time process starts with $X_0 = x$ and takes one discrete time step to $X_1$:

$$Lg(x) = \mathrm{E}[\, g(X_1)] \ . \tag{5}$$ dgd

To say this more explicitly, if $h = Lg$, then $h(x) = \mathrm{E}[\, g(X_1)]$. For example, for a simple random walk on the integers, $\mathcal{S} = \mathbb{Z}$. Suppose $\mathrm{P}(X \to X+1) = .5$, $\mathrm{P}(X \to X-1) = .3$ and $\mathrm{P}(X \to X) = .2$. Then

$$\begin{aligned}\mathrm{E}[\, g(X_1)] &= g(x+1) \cdot \mathrm{P}(X_1 = x+1) + g(x) \cdot \mathrm{P}(X_1 = x) + g(x+1) \cdot \mathrm{P}(X_1 = x-1) \\ &= .5 \cdot g(x+1) + .3 \cdot g(x) + .2 \cdot g(x-1) \ .\end{aligned}$$

[1] It is common to define abstract objects by their actions. There is a children's book with the line: "Don't ask me what Voom is. I never will know. But boy let me tell you it does clean up show."

You can see that the matrix $L$ is the same as the transition matrix

$$P_{ij} = \mathrm{P}(i \to j) = \mathrm{P}(X_{t+1} = j \mid X_t = i) \ .$$

This is because

$$\mathrm{E}[\, g(X_1)] = \sum_j \mathrm{P}(X_1 = j \mid X_0 = x)\, g(j) \ .$$

If $g$ is the column vector whose entries are the values $g(i)$, then this says that $Lg(j)$ is component $j$ of the vector $Pg$. The transition matrix is $P$. The generator is $L$. They are the same but have a different name.

There is a *dynamics* of value functions (conditional expectations) that involves the generator. The derivation uses the tower property. For example,

$$\mathrm{E}[\, g(X_2)] = \mathrm{E}[\, \mathrm{E}[\, g(X_2) \mid \mathcal{F}_1]] = \mathrm{E}[\, Lg(X_1)] = (L\,(Lg))\,(x) = L^2 g(x) \ .$$

The expected value after $s$ steps is

$$\mathrm{E}[\, g(X_s)] = L^s g(x) \ .$$

This looks slightly different if we let $s = T - t$ be the time remaining between $t$ and a final time $T$. Then

$$\mathrm{E}_{x,t}[\, g(X_T)] = L^{T-t} g(x) \ .$$

The time variable in the backward equation is the time between the start and the stop. When you increase this time variable, you can imagine starting further from the stopping time, or starting at the same time and running the process longer. This is the value function for a payout of $g(X_T)$ at time $T$.

Now suppose you have a continuous time Markov process in a discrete state space. At each time $t$, the state $X_t$ is one of the elements of $\mathcal{S}$. The process is described by a *transition rate matrix*, $R$, with $R_{ij}$ being the *transition rate* for $i \to j$ transitions. This means that if $j \neq i$ are two elements of $\mathcal{S}$, then

$$\mathrm{P}(X_{t+dt} = j \mid X_t = i) = R_{ij} dt \ .$$

Another way to express this is

$$\mathrm{P}(i \to j \ \text{ in time } dt \mid X_t = i) = R_{ij} dt \ .$$

The generator of a continuous time process is defined a little differently from the generator of a discrete time process. In continuous time you need to focus on the rate of change of quantities, not the quantities themselves. For this reason, we define $Lg(x)$ as (assume $X_0 = x$ as before)

$$Lg(x) = \lim_{\Delta t \to 0} \frac{\mathrm{E}[\, g(X_{\Delta t})] - g(x)}{\Delta t} \ . \tag{6}$$ cL

In time $\Delta t$ we expect $\mathrm{E}[\, g(X_{\Delta t})]$ not to be very different from $g(x)$. The definition (6) describes the rate of change. Two things about the discrete time

problem are similar to this continuous time problem. The generator is the same as the transition rate matrix. The evolution of expectation values over a longer time is given by the backward equation using the generator.

To see why $L = R$, we write approximate expressions for the probabilities $P_{xy}(\Delta t) = \mathrm{P}(X_{\Delta t} = y \mid X_0 = x)$. (The notation keeps changing, sometimes $i, j$, sometimes $x, y$. This week it's not an accident.) For $y \neq x$, the probability is approximately $R_{xy}\Delta t$. For small $\Delta t$, the same state probability $P_{xx}(\Delta t)$ is approximately equal to 1. We define the diagonal elements of $R$ to make the formula

$$P_{xy}(\Delta t) = \delta_{xy} + \Delta t R_{xy} \tag{7}$$ PR

true for all $x, y$. It is already true for $x \neq y$. To make it true for $x = y$, we need

$$R_{xx} = -\sum_{y \neq x} R_{xy} \ .$$

The off diagonal entries of the rate matrix are non-negative. The diagonal entries are negative. The sum over all landing states is zero:

$$\sum_{y \in \mathcal{S}} R_{xy} = 0 \ .$$

Assuming this, the formula (7) for $P$ gives

$$\sum_{y \in \mathcal{S}} P_{xy} = \sum_{y \in \mathcal{S}} \mathrm{P}(X_{\Delta t} = y \mid X_0 = x) = 1 \ .$$

This is supposed to be true. With the definitions above, it is.

With all this, we can evaluate the limit (6). Start with

$$\mathrm{E}[\, g(X_{\Delta t})] = \sum_{y \in S} P_{xy}(\Delta t) g(y) \ .$$

Now that (7) applies for all $x, y$ we just get

$$\mathrm{E}[\, g(X_{\Delta t})] \approx g(x) + \Delta t \sum_{y \in \mathcal{S}} R_{xy} g(y) \ .$$

We substitute this into (6), cancel the $g(x)$, and then the $\Delta t$. The result is

$$Lg(x) = \sum_{y \in \mathcal{S}} R_{xy} g(y) \ .$$

This is $Lg = Rg$, if we think of $R$ as a matrix and $g$ as the column vector made of the numbers $g(y)$.

It is convenient to consider functions that depend explicitly on time when discussing the dynamics of expectation values. Let $f(x, t)$ be such a function. The generator discussion above implies that

$$\mathrm{E}[\, f(X_{\Delta t}, 0)] = f(x, 0) + \Delta t f(x, 0) + O(\Delta t^2) \ .$$

When you consider the explicit dependence of $f$ on $t$, this becomes

$$\mathrm{E}[\, f(X_{\Delta t}, \Delta t)] = f(x,0) + \Delta t \Big( Lf(x,0) + \partial_t f(x,0) \Big) + O(\Delta t^2) \; .$$

Now consider a sequence of time steps of size $\Delta t$, with $t_k = k\Delta t$ and $t = t_n$. Then

$$\begin{aligned} \mathrm{E}[\, f(X_{t_n}, t_n)] - f(x,0) &= \sum_{k=0}^{n-1} \mathrm{E}\big[\, f(X_{t_{k+1}}, t_{k+1}) - f(X_{t_k}, t_k)\big] \\ &\approx \sum_{k=0}^{n-1} \mathrm{E}\Big[ \Big( Lf(X_k, t_k) + \partial_t f(X_k, t_k) \Big) \Delta t \Big] \end{aligned}$$

In the limit $\Delta t \to 0$, this becomes

$$\mathrm{E}[\, f(X_t, t)] - f(x,0) = \int_0^t \mathrm{E}[\, Lf(X_s, s) + \partial_t f(X_s, s)] \, ds \; . \tag{8}$$ Df

This equation is true for any function $f(x,t)$. If $f$ satisfies the backward equation

$$Lf + \partial_t f = 0 \tag{9}$$ bec

then the right side is zero, and

$$\mathrm{E}[\, f(X_t, t)] = f(x,0) \; .$$

As for the discrete backward equation, we can replace the time interval $[0,t]$ with the interval $[t,T]$, which gives the familiar restatement

$$\mathrm{E}_{x,t}[\, f(X_T, T)] = f(x,t) \; .$$

The definition of the generator (6) is easy to work with, particularly if the process is subtle. Suppose $X_t$ satisfies the SDE (2) and $g(x)$ is a twice differentiable function. Define $\Delta X = X_{\Delta t} - x$ and make the usual approximations Then

$$g(X_{\Delta t}) - g(x) \approx \sum_i \Delta X_i \partial_{x_i} g(x) + \frac{1}{2} \sum_{ij} \Delta X_i \Delta X_j \partial_{x_i} \partial_{x_j} g(x) \; .$$

The SDE gives

$$\mathrm{E}[\, \Delta X_i] \approx a_i(x) \Delta t \;\; , \quad \mathrm{E}[\, \Delta X_i \Delta X_j] \approx \big( b(x) b^t(x) \big)_{ij} \, \Delta t$$

Therefore,

$$\mathrm{E}[\, g(X_{\Delta t}) - g(x)] \approx \left( \sum_i a_i(x) \partial_{x_i} g(x) + \frac{1}{2} \sum_{ij} \big( b(x) b^t(x) \big)_{ij} \, \partial_{x_i} \partial_{x_j} g(x) \right) \Delta t \; .$$

Therefore,

$$Lg(x) = \sum_i a_i(x)\partial_{x_i} g(x) + \frac{1}{2}\sum_{ij} \left(b(x)b^t(x)\right)_{ij} \partial_{x_i}\partial_{x_j} g(x) \ . \tag{10}$$ cgg

It is common to specify $L$ as an *operator* without putting the function $g$ into the expression. That would be

$$L = \sum_i a_i(x)\partial_{x_i} + \frac{1}{2}\sum_{ij} \left(b(x)b^t(x)\right)_{ij} \partial_{x_i}\partial_{x_j} \ . \tag{11}$$ cg

This is a *differential operator*. It is defined by how it acts on a function $g$, which is given by (10).

The general relationship between backward and forward equations may be understood using *duality*. This term has several related meanings in mathematics.[2] One of them is an abstract version of the relationship between a matrix and its transpose. In the abstract setting, a matrix becomes an *operator*, and the transpose of the matrix becomes the *adjoint* of the operator. The transpose of the matrix $L$ is $L^t$. The adjoint of the operator $L$ is $L^*$. This is important here because if $L$ is the generator of a Markov process, then $L$ is the operator that appears in the backward equation. But $L^*$, the adjoint of $L$, appears in the forward equation. This can be the easiest way to figure out the forward equation in practical examples. It will be easy to identify the forward equation for general diffusions with generator $L$ as in (11). But look back at Assignment 6 to see how hard it can be to derive the forward equation directly. When all the "bla bla" is over, this fancy duality boils down to simple integration by parts.

We introduce abstract adjoints for operators by describing how things work for finite dimensional vectors and matrices. In that setting, we distinguish between the $n \times 1$ "matrix", $f$, and the $1 \times n$ matrix, $u$. As a matrix, $f$ has one column and $n$ rows. This is also called a *column vector*. As a matrix, $u$ has one row and $n$ columns, which makes $u$ a *row vector*. A row vector or a column vector have $n$ components, but they are written in different places. Suppose $A$ is and $m \times n$ matrix and $B$ is an $n \times k$ matrix. Then the matrix product $AB$ is defined but the product $BA$ is not, unless $k = m$. If $A$ is the $n$ component row vector $u$ and $B$ is the $n \times n$ matrix $L$, we have $m = 1$ and $k = n$ above. The matrix product $uL = v$ is another $1 \times n$ matrix, or row vector. The more traditional matrix vector multiplication involves $A = L$ as $n \times n$ and $f$ as $n \times 1$, so $Lf = g$ is $n \times 1$, which makes $g$ another column vector.

Suppose $\mathcal{S}$ is a finite state space with states $x_i$ for $i = 1, \ldots, n$. We can write the probability of state $x_i$ as $u(x_i)$ or $u_i$. If $f(x)$ is a function of the state, we write $f_i$ for $f(x_i)$. The expected value may be written in several ways

$$\mathrm{E}[\, f] = \sum_{i=1}^{n} u_i f_i = \sum_{x_i \in \mathcal{S}} u(x_i) f(x_i) = \sum_{x \in \mathcal{S}} u(x) f(x) \ .$$

[2] For example, the *dual* of the icosahedron is the dodecahedron, and vice versa.

Now let $u$ refer to the row vector with components $u_i$ and $f$ the column vector with components $f_i$. Then the expected value expression above may be written as the matrix product

$$\mathrm{E}[\, f] = uf \ .$$

This is the product of a $1 \times n$ matrix, $u$, with the $n \times 1$ matrix $f$. The result is a $1 \times 1$ "matrix", which is just a single number, the expected value.

Now suppose $f = f_t = \mathrm{E}_t[\, V(X_T)]$ and $u = u_t$ with $u_t(x) = \mathrm{P}(X_t = x)$. The tower property implies that the overall expectation is given by

$$\mathrm{E}[\, V(X_T)] = \mathrm{E}[\, \mathrm{E}_t[\, V(X_T)]] = \mathrm{E}[\, f_t(X_t)] = u_t f_t \ . \tag{12}$$ ext

The last form on the right is the product of the row vector $u_t$ and the column vector $f_t$. Note, and this is the main point, that the left side is independent of $t$ in the range $0 \leq t \leq T$. This implies, in particular, that

$$u_{t+1} f_{t+1} = u_t f_t \ .$$

But the relationship between $f_t$ and $f_{t+1}$ is given by (5), in the form $f_t = L f_{t+1}$. Therefore

$$u_{t+1} f_{t+1} = u_t L f_{t+1} \ ,$$

for any vector $f_{t+1}$. This may be re-written using the fact that matrix multiplication is associative as

$$(u_{t+1} - u_t L)\, f_{t+1} = 0 \ ,$$

for every $f_{t+1}$. If the set of all possible $f_{t+1}$ spans the whole space $\mathbb{R}^n$, this implies that

$$u_{t+1} = u_t L \ . \tag{13}$$ fed

To summarize: if the value function satisfies the backward equation involving the generator $L$, then the probability distribution satisfies a forward equation with that same $L$, but used in a different way – multiplying from the left rather than from the right. Recall that what we call $L$ here was called $P$ before. It is the matrix of transition probabilities – the transition matrix for the Markov chain.

You can give the relationship between the backward and forward equations in a different way if you treat all vectors as column vectors. If $u_t$ is the column vector of occupation probabilities at time $t$, then the expected value formula is $\mathrm{E}[\, V(X_T)] = u_t^t f_t$. The backward equation formula, written using the column vector convention, is $u_{t+1}^t f_{t+1} = u_t^t L f_{t+1} = (L^t u_t)\, f_{t+1}$. The reasoning we used to get (13) now gives the column vector formula

$$u_{t+1} = L^t u_t \ .$$

To summarize: $u_t f_t$, or $u_t^t f_t$ is independent of $t$. This implies a relationship between the dynamics of $f$ and the dynamics of $u$. The relationship is that the matrix $L$ that does the backward dynamics of $f$ has an adjoint (transpose) that does the forward dynamics of $u$.

We move to a continuous time version of this that applies to continuous time Markov chains on a finite state space $\mathcal{S}$. The dynamics for $f_t$ are

$$\frac{d}{dt} f_t = -L f_t \ ,$$

where $L$, the generator, is also the matrix of transition rates. If $u_t$ is the row vector of occupation probabilities, $u_t(x) = \mathrm{P}(X_t = x)$, then $u_t f_t$ is independent of $t$ for the same reason as above. Therefore

$$0 = \frac{d}{dt} \left( u_t \cdot f_t \right) = \left( \frac{d}{dt} u_t \right) f_t + u_t \left( \frac{d}{dt} f_t \right) = \left( \frac{d}{dt} u_t \right) f_t - u_t \left( L f_t \right) \ .$$

This implies, as in the discrete time case above, that

$$\left[ \left( \frac{d}{dt} u_t \right) - u_t L \right] f_t = 0 \ ,$$

for every value function vector $f_t$. If these vectors span $\mathbb{R}^n$, then the vector in brackets must be equal to zero:

$$\frac{d}{dt} u_t = u_t L \ .$$

If we use the convention of treating the occupation probabilities as a column vector, then this is

$$\frac{d}{dt} u_t = L^t u_t \ .$$

It is easy to verify all these dynamical equations directly for finite state space Markov chains in discrete or continuous time. For example, ...

With all this practice, the PDE argument for diffusion processes is quick. Start with the one dimensional case with drift $a(x)$ and noise $b(x)$. The backward equation is

$$\partial_t f(x,t) + L f = \partial_t f + \frac{1}{2} \partial_x^2 f(x,t) + a(x) \partial_x f(x) = 0 \ .$$

Let $u(x,t)$ be the probability density for $X_t$. Then as before the time $t$ formula for the expected payout is true for all $t$ between 0 and $T$

$$\mathrm{E}[\, V(X_T)] = \int_{-\infty}^{\infty} u(x,t) f(x,t) \, dx \ .$$

We differentiate this with respect to $t$ and use the backward equation for $f$:

$$\begin{aligned} 0 &= \partial_t \int_{-\infty}^{\infty} u(x,t) f(x,t) \, dx \\ &= \int_{-\infty}^{\infty} \left( \partial_t u(x,t) \right) f(x,t) \, dx + \int_{-\infty}^{\infty} u(x,t) \left( \partial_t f(x,t) \right) \, dx \\ &= \int_{-\infty}^{\infty} \left( \partial_t u(x,t) \right) f(x,t) \, dx - \int_{-\infty}^{\infty} u(x,t) \left( L f(x,t) \right) \, dx \end{aligned}$$

The new trick here is to move $L$ onto $u$ by integration by parts. This is the continuous state space analogue of writing $u^t\,(Lf)$ as $(L^t u)^t\,f$ in the discrete state space case. In the integrations by parts we assume that there are no boundary terms at $\pm\infty$. The reason for this is that the probability density $u(x,t)$ goes to zero very rapidly as $x \to \pm\infty$. In typical examples, such as the Gaussian case of Brownian motion, $u(x,t)$ goes to zero exponentially as $x \to \pm\infty$. Therefore, even if $f(x,t)$ does not go to zero, or even goes to infinity, the boundary terms vanish in the limit $x \to \pm\infty$. Here is the algebra:

$$\begin{aligned}\int_{-\infty}^{\infty} & u(x,t)\,(Lf(x,t))\,dx \\ &= \int_{-\infty}^{\infty} u(x,t)\left(\frac{1}{2}b(x)^2\partial_x^2 f(x,t) + a(x)\partial_x f(x,t)\right)dx \\ &= \int_{-\infty}^{\infty}\left[\frac{1}{2}\partial_x^2\left(b^2(x)u(x,t)\right) - \partial_x\left(a(x)u(x,t)\right)\right]f(x,t)\,dx\ .\end{aligned}$$

The quantity in square brackets is

$$L^*u(x,t) = \frac{1}{2}\partial_x^2\Big(b^2(x)u(x,t)\Big) - \partial_x\Big(a(x)u(x,t)\Big)\ . \tag{14}$$ L*

This defines the operator $L^*$, which is the *adjoint* of the generator $L$. The integration by parts above shows that

$$\int_{-\infty}^{\infty}\left(\partial_t u(x,t) - L^*u(x,t)\right)f(x,t)\,dx = 0\ ,$$

for every value function $f(x,t)$. If there are enough value functions, the only way for all these integrals to vanish is for the $u$ part to vanish, which is

$$\partial_t u(x,t) = L^*u(x,t) = \frac{1}{2}\partial_x^2\Big(b^2(x)u(x,t)\Big) - \partial_x\Big(a(x)u(x,t)\Big)\ . \tag{15}$$ fecc

This is the *forward Kolmogorov equation* for the evolution of the probability density $u(x,t)$.

There many features of the forward equation to keep in mind. There are important differences between the forward and backward equations. One difference is that in the backward equation the noise and drift coefficients are outside the differentiation, but they are inside in the forward equation. In both cases it has to be like this. For the backward equation, constants are solutions, obviously, because if the payout is $V(X_T) = c$ independent of $X_T$, then the value function is $f(X_t,t) = c$ independent of $X_t$. You can see $f(x,t) = c$ satisfies the backward equation because $Lc = 0$. If $L$ were to have, say $\partial_x\Big(a(x)f(x,t)\Big)$, then we would have $Lc = \big(a(x)\big)c \neq 0$ if $a$ is not a constant. That would be bad. The forward equation, on the other hand, is required to preserve the integral of $u$, not constants. If we had $-a(x)\partial_x u(x,t)$ instead of $-\partial_x\big(a(x)u(x,t)\big)$, then we

would have (if $b$ is constant)

$$\begin{aligned}\frac{d}{dt}\int_{-\infty}^{\infty} u(x,t)\,dx &= \int_{-\infty}^{\infty} \partial_t u(x,t)\,dx \\ &= -\int_{-\infty}^{\infty} a(x)\partial_x u(x,t)\,dx \\ &= \int_{-\infty}^{\infty} \left(\partial_x a(x)\right) u(x,t)\,dx\ ,\end{aligned}$$

which is not equal to zero in general if $\partial_x a \neq 0$.

A feature of the forward equation that takes scientists by surprise is that the second derivative term is not

$$\frac{1}{2}\partial_x\left(b^2(x)\partial_x u(x,t)\right)\ .$$

This is because of the martingale property. In the no drift case, $a = 0$, we have supposed not to change the expected value of $X_t$. Therefore

$$0 = \frac{d}{dt}\mathrm{E}[\,X_t] = \frac{d}{dt}\int x u(x,t)\,dx = \int x\partial_t u(x,t)\,dx\ .$$

If we use the correct form, this works out:

$$\begin{aligned}\int x\partial_t u(x,t)\,dx &= \frac{1}{2}\int x\partial_x^2\left(b^2(x)u(x,t)\right)\,dx \\ &= -\frac{1}{2}\int \left(\partial_x x\right)\partial_x\left(b^2(x)u(x,t)\right)\,dx \\ &= -\frac{1}{2}\int \left(\partial_x^2 x\right)\left(b^2(x)u(x,t)\right)\,dx \\ &= 0\ .\end{aligned}$$

But the incorrect form above would give

$$\int x\partial_x\left(b^2(x)\partial_x u(x,t)\right)\,dx = -\int \left(b^2(x)\partial_x u(x,t)\right)\,dx = \int \left(\partial_x b^2(x)\right)u(x,t)\,dx\ .$$

This is not equal to zero in general.

Another crucial distinction between forward and backward is the relation between the signs of the $\partial_t$ and $\partial_x^2$ terms. This is clearest in the simplest example, which is Brownian motion: $dX = dW$, which has $a = 0$ and $b = 1$. Then the backward equation is $\partial_t f + \frac{1}{2}\partial_x^2 f = 0$, and the forward equation is $\partial_t u = \frac{1}{2}\partial_x^2 u$. You can appreciate the difference by writing the backward equation in the form $\partial_t f = -\frac{1}{2}\partial_x^2 f$. The forward equation has $+\partial_x^2$ and the backward equation has $-\partial_x^2$. This is related to the fact that the forward equation is "intended" for evolving $u(x,t)$ forward in time. If you specify $u(x,t)$, you the forward equation determines $u(x,t)$ for $t > 0$. The backward is for evolving the

value function backward in time. If you specify $f(x,T)$, the backward equation determines $f(x,t)$ for $t < T$.

One way to remember the signs is to think what should happen to a local maximum. The probability density at a local maximum should go down as you move forward in time. A local maximum represents a region of high concentration of "particles" of probability. Moving forward in time these particles will disperse. This causes the density of particles, the probability, to go down. Mathematically, a local maximum of $u$ at $x_0$ at time $t_0$ is represented by $\partial_x u(x_0,t_0) = 0$ and $\partial_x^2 u(x_0,t_0) < 0$. At such a point, we want $u$ to be decreasing, which is to say $\partial_t u(x_0,t_0) < 0$. The forward equation with $+\partial_x^2$ does this, but with $-\partial_x^2$ would get it wrong. The backward equation should lower a local maximum in the value function $f$ moving backward in time. Suppose $\partial_x f(x_0,t_0) = 0$ and $\partial_x^2 f(x_0,t_0) < 0$ (a local maximum). Then suppose you start at a time $t < t_0$ with $X_t = x_0$. There is a chance that $X_{t_0}$ will be close to $x_0$, but it probably will miss at least a little. (Actually, it misses almost surely.). Therefore $f(X_{t_0},t_0) < f(x_0,t_0)$, so $f(x_0,t) = \mathrm{E}_{x_0,t}[\, f(X_{t_0},t_0)] < f(x_0,t_0)$. This suggests that $\partial_t f(x_0,t_0) > 0$, which makes $f$ decrease as you move backward in time. The backward equation with $-\partial_x^2$ does this.

The multi-dimensional version of these calculations is similar. It is easier if you express the calculations above in a somewhat more abstract way using the generator $L$ and an *inner product* appropriate for the situation. in finite dimensions and for vectors with real components, the "standard" inner product is

$$\langle u, f \rangle = \sum_{i=1}^{n} u_i f_i \ .$$

Clearly, this is just a different notation for $u^t f$ or $uf$ depending on whether you think of $u$ as a column vector or a row vector. If $L$ is an $n \times n$ matrix, the definition of the adjoint of $L$ is that matrix $L^*$ so that

$$\langle u, Lf \rangle = \langle L^* u, f \rangle \ , \tag{16}$$ adj

for all vectors $u$ and $f$. We already did the matrix calculations to show that $L^* = L^t$ for matrices, and with the standard inner product. We derive the forward equation from the backward equation, in this notation, as follows. Start with

$$\langle u_t, f_t \rangle = \langle u_{t+1}, f_{t+1} \rangle \ .$$

Then use the backward equation and the adjoint relation:

$$\langle u_t, f_t \rangle = \langle u_t, Lf_{t+1} \rangle = \langle L^* u_t, f_{t+1} \rangle \ .$$

The additivity property of inner products allows us to write this in the form

$$\langle (L^* u_t - u_{t+1}) \, , f_{t+1} \rangle = 0 \ .$$

This is supposed to hold for all $f_{t+1}$, which forces the vector in parentheses to be zero (another property of inner products).

The continuous time version of this is about the same. It is a property of inner products that

$$\frac{d}{dt}\langle u_t, f_t\rangle = \langle \frac{d}{dt}u_t, f_t\rangle + \langle u_t, \frac{d}{dt}f_t\rangle \ .$$

If the inner product is independent of $t$ and $f$ satisfies the backward equation, this gives

$$\begin{aligned} 0 &= \langle \frac{d}{dt}u_t, f_t\rangle - \langle u_t, Lf_t\rangle \\ &= \langle \frac{d}{dt}u_t, f_t\rangle - \langle L^* u_t, f_t\rangle \\ &= \langle \left(\frac{d}{dt}u_t - L^* u\right), f_t\rangle \end{aligned}$$

As we argued before, if this holds for enough vectors $f_t$, it implies that the quantity in parentheses must vanish, which leads to:

$$\frac{d}{dt}u_t = L^* u_t \ . \tag{17}$$ feg

This says that to find the forward equation from the backward equation, you have to find the adjoint of the generator.

The generator of the multi-dimensional diffusion process is given by (10) or (11). We simplify the notation by defining the *diffusion coefficient* matrix $\mu(x) = b(x)b^t(x)$. The components of $\mu$ are the diffusion coefficients $\mu_{jk} = \sum_l b_{jl}b_{kl}$. The quantity that is constant in time is

$$\int_{\mathbb{R}^n} u(x_1, \ldots, x_n, t) f(x_1, \ldots . x_n, t)\, dx_1 \cdots dx_n \ . \tag{18}$$ ndi

The adjoint of $L$ is the operator $L^*$ so that

$$\int \left(L^* u(x,t)\right) f(x,t)\, dx = \int u(x,t)\left(Lf(x,t)\right)\, dx$$

The adjoint is found by integration by parts. The generator is the sum of many terms. We do the integration by parts separately for each term and add the results. A typical first derivative part of $L$ is $a_j(x)\partial_{x_j}$. The adjoint of this term is found by integrating by parts in the $x_j$ variable, which is just one of the variables in the $n$ dimensional integration (18). That is

$$\int_{x_j=-\infty}^{\infty} u(x,t)a_j(x)\partial_{x_j} f(x,t)\, dx_j = -\int_{x_j=-\infty}^{\infty} \left[\partial_{x_j}\left(a_j(x)u(x,t)\right)\right] f(x,t)\, dx_j \ .$$

The integrations over the other variables does nothing to this. The result is

$$\int_{\mathbb{R}^n} u(x,t)a_j(x)\partial_{x_j} f(x,t)\, dx = -\int_{\mathbb{R}^n} \left[\partial_{x_j}\left(a_j(x)u(x,t)\right)\right] f(x,t)\, dx \ .$$

A typical second derivative term in $L$ is $\mu_{jk}(x)\partial_{x_j}\partial_{x_k}$. If $j \neq k$, you move these derivatives from $f$ onto $u$ by integration by parts in $x_j$ and $x_k$. The overall sign comes out + because you do two integrations by parts. The result is

$$\int_{x_j,x_k} u(x,t)\mu_{jk}(x)\partial_{x_j}\partial_{x_k} f(x,t)\, dx_j dx_k = \int_{x_j,x_k} \left[\partial_{x_j}\partial_{x_k}\left(\mu_{jk}(x)u(x,t)\right)\right] f(x,t] dx_j dx_k \ .$$

You can check that this result is still true if $j = k$. Altogether, the adjoint of $L$ is

$$L^* u(x,t) = \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \partial_{x_j}\partial_{x_k}\left(\mu_{jk}(x)u(x,t)\right) - \sum_{j=1}^{n}\partial_{x_j}\left(a_j(x)u(x,t)\right) \ . \qquad (19) \quad \boxed{\texttt{L*c}}$$

The forward equation for the probability density is

$$\partial_t u(x,t) = \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \partial_{x_j}\partial_{x_k}\left(\mu_{jk}(x)u(x,t)\right) - \sum_{j=1}^{n}\partial_{x_j}\left(a_j(x)u(x,t)\right) \ . \qquad (20) \quad \boxed{\texttt{fec}}$$

A common rookie mistake is to get the factor of $\frac{1}{2}$ wrong in the second derivative terms. Remember that the off diagonal terms, the ones with $j \neq k$, are given twice in (19), once with $(j,k)$ and again with $(k,j)$. For example, in two dimensions, the second derivative expression is

$$\frac{1}{2}\partial_{x_1}^2\left(\mu_{11}(x)u(x,t)\right) + \frac{1}{2}\partial_{x_2}^2\left(\mu_{22}(x)u(x,t)\right) + \partial_{x_1}\partial_{x_2}\left(\mu_{12}(x)u(x,t)\right) \ .$$

The matrix $\mu = bb^t$ is symmetric, even when $b$ is not symmetric. Therefore, it does not matter whether you write $\mu_{12}$ or $\mu_{21}$.

This stuff is closely related to Ito's lemma. Suppose $f(x,t)$ is some function. The Ito differential is

$$df(X_t,t) = \partial_t f dt + Lf(X_t,t)dt + \nabla f(X_t,t)b(X_t,t)dW_t \ . \qquad (21) \quad \boxed{\texttt{ilg}}$$

This is not a form we have used before, but it is easy to check. The expectation from this is

$$\mathrm{E}[\, df(X_t,t) \mid \mathcal{F}_t] = \left(\partial_t f(X_t,t) + Lf(X_t,t)\right) dt \ . \qquad (22) \quad \boxed{\texttt{ilge}}$$

This is more or less our definition of generator above.

# 3 Feynman Kac and other equations

There are backward equations for lots of other functions of a process. One example is the *running cost* or *running payout* problem. (Engineers talk about costs. Finance people talk about payouts.)

$$A = \int_0^T V(X_s)ds \ . \qquad (23) \quad \boxed{\texttt{af}}$$

The corresponding value function is

$$f(x,t) = \mathrm{E}_{x,t}\left[ \int_t^T V(X_s)ds \right] . \tag{24}$$ afv

The Ito differential (ilg 21) is the easy to find the backward equation. On one hand you have

$$d\left[ \int_t^T V(X_s)ds \mid \mathcal{F}_t \right] = -V(X_t)dt .$$

On the other hand we have (ilg 21) or (ilge 22). Together, this give

$$-V(x)dt = \partial_t f(x,t) + Lf(x,t) .$$

This gives the backward equation

$$0 = \partial_t f + Lf + V(x)f . \tag{25}$$ afbe

If you want to know the value of $A$ in (af 23), one way to find it is to solve the backward equation with final condition $f(x,T) = 0$. Then $A = f(x_0, 0)$. In this approach you have to solve the whole PDE and compute the whole value function just to find the single number $A$.

The correspondence between the additive function (af 23) and the backward equation (afbe 25) can go both ways. You can use the PDE to find the value of $A$. You can use the definition (afv 24) to find the solution of the PDE (afbe 25). This is useful in situations where the backward equation was not derived as a probability model. It is most important when the dimension of the problem is more than a few, maybe more than 4 or 5, where PDE solution methods are impractical.

Another connection of this kind concerns the multiplicative function

$$A = \exp\left( \int_0^T V(X_s)ds \right) . \tag{26}$$ mf

We study this using the corresponding value function

$$f(x,t) = \mathrm{E}_{x,t}\left[ \exp\left( \int_t^T V(X_s)ds \right) \mid \mathcal{F}_t \right] . \tag{27}$$ mfv

We find the backward equation for this value function following the reasoning we used for the previous one. On one hand, we have

$$d\,\exp\left( \int_t^T V(X_s)ds \right) = -V(X_t)\,\exp\left( \int_t^T V(X_s)ds \right) dt .$$

(If $Y_t$ is the additive functional $Y_t = \int_t^T V(X_s)ds$, then $dY$ has no Ito part, so $d\,e^{Y_t} = e^{Y_t}dY_t$. We just did $dY_t$.) So we equate this with the Ito version of $df$ given in (ilg 21) and (ilge 22) to get

$$\mathrm{E}[\,df(X_t,t) \mid \mathcal{F}_t] = \partial_t f(X_t, tdt) + Lf(X_t,t)dt = -V(X_t)f(X_t,t)dt .$$

Rearranging this leads to

$$\partial_t f(x,t) + Lf(x,t) + V(x)f(x,t) = 0\ . \tag{28}$$ mfbe

The final condition for this one is $f(x,T) = 1$.

Again, the relationship between the multiplicative value function (mfv 27) and the backward equation (mfbe 28) can be useful either way. Solving the PDE allows you to calculate the expectation of the multiplicative function. Using the multiplicative function (mfv 27) allows you to use Monte Carlo to compute the solution of the PDE (mfbe 28). It might happen that we want the PDE solution with final condition $f(x,T) = g(x)$ that is not identically equal to 1. In this case, the solution formula is clearly (think this through)

$$f(x,t) = \mathrm{E}_{x,t}\left[ g(X_t) \exp\left( \int_t^T V(X_s) ds \right) \mid \mathcal{F}_t \right]\ .$$

This solution formula for the PDE is called the *Feynman Kac* formula. A version of this was proposed by the physicist Feynman in the 1940's for the related PDE that has $iL$ instead of $L$. Feynman's formula was criticized by mathematicians for not being rigorous, in my opinion somewhat unfairly. The mathematician Kac[3] discovered the present completely rigorous version of Feynman's formula. Modern probabalists, particularly applied probabilists working in finance or operations research, call the PDE (mfbe 28) the Feynman Kac formula instead. This reverses the original intention. Some go even further, using the term "Feynman Kac" for any backward equation of any kind.

[3]Pronounced "cats". Another spelling of the same Polish name is "Katz".