ECON2706*

Zhengyuan Gao

Contents

*Unfinished version. This version: March 24, 2017

Part I

Foundation

1 Differential Equations in A Nutshell

Let's start with a simple problem:

$$\frac{du}{dt} = ru,\ t > 0$$

which is a typical model in population growth and resource supply. It is called the Malthus model. It is a simple model where the time rate of change of population $u = u(t)$ is proportional to the population. Here t is time and $u = u(t)$ is the population of a given system of individuals. u is often refered to state variable. The equation governs the evolution of this state variable. The real number r is a parameter that rpresents the relative growth rate and could be measured for the given population. The solution to this model is

$$u(t) = u_0 e^{rt}, \quad t > 0$$

where $u_0 = u(0)$ represents the initial population. The Malthus model is about exponential growth. The state variable is a function of a single independent variable (time t).

A paritial differential equation (PDE) differs from this model in that the state variable depends on *more than* one independent variable. Thus a PDE model models the evolution of a system with more than one variable of interests. The system could be about both time and the other variables $u = u(t, x)$ from

$$\frac{\partial u(t,x)}{\partial t} = c\frac{\partial^2 u(t,x)}{\partial x^2}$$

or it could be independent of time but depend on several variables $u(\mathbf{x}) = u(x_1, x_2, x_3)$ from

$$\frac{\partial u(\mathbf{x})}{\partial x_1} + \frac{\partial u(\mathbf{x})}{\partial x_2} + \frac{\partial u(\mathbf{x})}{\partial x_3} = 0$$

such as spatial variables $\mathbf{x}$. If x belongs to a bounded set $x \in \mathcal{X}$, it is important to give some conditions of $u(t, x)$ on the boundary of this set $\mathcal{X}$. For example $u(0, t) = 0$ and $u(x_{\max}, t) = 0$ for any $t > 0$. This type of conditions is called *boundary conditions*. On the other hand, if initially the model has a functional form, for example $u(x, 0) = \varphi(x)$ for $0 < x < x_{\max}$, then we call this condition an *initial condition*. Because it specifies the state variable at time $t = 0$.

In general, a PDE in one variable x and time t is an equation of the following form

$$G\left(x, t, u, \frac{\partial u}{\partial t}, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial t^2}, \frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial t \partial x}, \dots\right) = 0,\ x \in \mathcal{X},\ t \in \mathcal{T}.$$

Very often, the time interval $\mathcal{T}$ is just the positive time $t \geq 0$. If $G(\cdot)$ is a linear function in u and in all of its derivatives, this PDE is called a *linear PDE problem.* It means that u and its derivatives are uncorrelated and the equation only contains their first order information. A linear equation is *homogenous* if every term contains u or some derivative of u. Linear equations have an algebraic structure to their solution sets. For example, the sum of two solutions to a homogenous linear equation is again a solution, as are constant multiples of solutions. Nonlinear equations lost this property.

Examples of physical models using PDEs are the heat equation for the evolution of the temperature distribution in a body, the wave equation for the motion of a wavefront, the flow equation for the flow of fluids and Laplace's equation for an electrostatic potential or elastic strain field. Some standard linear homogeneous PDEs

$$\begin{aligned}
\text{Flow Equation: } & c\frac{\partial u}{\partial x} + \frac{\partial u}{\partial t} = 0, \\
\text{Heat Equation: } & c^2\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} = 0, \\
\text{Wave Equation: } & c^2\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = 0, \\
\text{Laplace's Equation: } & \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = 0.
\end{aligned}$$

An example of a linear but non homogeneous PDE, Poisson's equation:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y)$$

An example of a nonlinear PDE, a nonlinear heat equation:

$$c^2\frac{\partial^2 u}{\partial x^2} - u\frac{\partial u}{\partial t} = 0.$$

Sometimes we denote partial differentiation by subscripts, as in u_x for $\frac{\partial u}{\partial x}$ or u_{tt} for $\frac{\partial^2 u}{\partial t^2}$. So, for example, the heat equation can be written

$$c^2 u_{xx} - u_t = 0.$$

If we have more variables, or need to write programs for solving a PDE, then often we use numbered variables, such as, $x_1, x_2, \ldots, x_n$ and then we denote a partial derivative by the subscript number. For example, u_{22} is the second partial derivative of u with respect to x_2.

A solution to the PDE means a function $u = u(x,t)$ satisfies the equation and the conditions. It means that u must have continuous partial derivatives as required by the PDE. There could be many solutions satisfying the requirements. For example,

$$u(x,t) = x^2 + 2t, \quad u(x,t) = e^{-t}\sin x$$

both satisfy $u_t = u_{xx}$.

Remark. In some simple cases, solutions can be found in terms of sums and products of elementary functions. Find all the first and second order partial derivatives of the functions $u(x,t)$ and try to match the functions to the above linear homogeneous partial differential equations with suitable initial and boundary conditions. Solutions to practical models using these equations are usually very difficult to obtain analytically and computers are used to obtain numerical approximate solutions by standard iterative procedures. You can see examples of some numerical programs in R from the book Solving Differential Equations in R.

Many PDE models come from a basic balance law called *conservation law*. A conservation law is just a mathematical formulation of the fact that the rate of quantity changes in a given domain must equal the rate of quantity flows on the boundary plus the rate of birth and death within this domain. Let $u = u(t,x)$ be the population density of one city. Let $f(x,t)$ be the birth and death rate at location x at time t and let $\varphi(x,t)$ be the number of inhabitants. At location x at time t, the change of number of inhabitants can be measured by $\varphi_x(x,t)$. Then the dynamical population in this city follows the conservation law

$$\int_{\mathcal{X}} (u_t(x,t) - (\varphi_x(x,t) + f(x,t)))\, dx = 0.$$

If u_t, φ_x and f are continuous, the integrand must identically everywhere in $\mathcal{X}$, so we have

$$u_t(x,t) = \varphi_x(x,t) + f(x,t).$$

If φ is proportional to the density $\varphi = cu$, then above model is called an advection model. The equation is reduced to

$$u_t - cu_x = 0$$

if $f = 0$. The solution of this advection equation is $u(x,t) = F(x+ct)$ for any differentiable function F. This solution just shifts $F(x)$ to ct units. This kind of shift is called a transport. If in addition, we assume that $u(x,0) = u_0(x)$. Then the solution becomes $u(x,t) = u_0(x+ct)$. If $\mathcal{X} \subset \mathbb{R}$, We can think of the density moving along a straight line $\xi = x + ct$. This line or this parameterization ξ is called characteristic. A general transport equation has the form

$$u_t - cu_x + au = f(x,t).$$

We can solve this equation by the characteristic. Let $\xi = x + ct$. Denote $u(x,t)$ in the new variable by $U(\xi,t)$. Then by the chain rule

$$u_t = cU_\xi + U_t,\ u_x = U_\xi.$$

Substitute these into the previous transport equation, the equation becomes $U_t + aU = f(\xi,t)$, an ODE equation. This equation can be solved by multiplying an integrating factor e^{at} and integrating w.r.t. t.

If the advection equation $u_t = \varphi_x$ has two additional conditions (1) the movement is from higher concentrations to lower concentrations, and (2) the steeper the concentration gradient, the greater φ, then we can have another type of equation, heat equation or diffusion equation. Assuming a simple linear relation $\varphi = u_x$, we have $u_t = u_{xx}$ which is the heat equation. We will consider its solution later in this course.

2 Expectations and PDEs

Many quantities in economics represent as the expected value of some random variable that is moving in a stochastic model of the market. Given the model, there are two ways to compute the expected value. One method starts with values that are expected to achieve in the future and computes expected values at successively earlier times until the present expected values are found. The other method starts with given probabilities for current market conditions and works forwards in time to find probabilities for market conditions at a desired future time. These two evolution equations are similar but not identical. One of the differences is the natural direction of time change, backwards for expected values and forwards for probabilities.

We briefly discuss these evolution equations in the simple case of discrete time and discrete state space in Section 2.1. Then in Section 5 we extense to more complex situations, continuous time and continuous state space. There is a duality relationship between them which will be discussed in Section 6.[1]

2.1 Markov Chain: Backwards and Forwards

2.1.1 Forwards

Many discrete time discrete state space stochastic models are Markov chains. Such a Markov chain is characterized by its state space, $\mathcal{S}$, and its transition matrix, P. For discrete state space, we denote the probability mass of $X(t)$ at x at time t as $u(x,t) = \Pr(X(t) = x)$. These probabilities satisfy an evolution equation moving forward in time. We use similar notation for conditional probabilities, for example, $u(x,t|X(0) = x_0) = \Pr(X(t) = x|X(0) = x_0)$, $p(x,y) = \Pr(x \to y) = \Pr(X(t+1) = y|X(t) = x)$. These "transition probabilities" are the elements of the transition matrix, P. The transition probabilities have the properties:

$$0 \leq p(x,y) \leq 1 \quad \text{for all } x \in \mathcal{S} \text{and } y \in \mathcal{S}. \tag{1}$$

and $\sum_{y\in\mathcal{S}} p(x,y) = 1 \quad$ for all $x \in \mathcal{S}$. The first property comes from the fact that $p(x,y)$ are probabilities, the second comes from the fact that the state x must go somewhere in next period. It will possibly go back to x

[1] It is an extension of the relationship between a matrix and its transpose.

The Markov property is that knowledge of the state at time t is:

$$\Pr(X(t+1) = y | X(t) = x_0, X(t-1) = x_1, \ldots) = \Pr(X(t+1) = y | X(t) = x_0)$$

namely all the information to predict the future only depends on the present relevant, no matter what extra history information ($X(t-1) = x_1, \ldots$) we have for the past relevant. This may be thought of as a lack of long term memory. It may also be thought of as a completeness property of the model: the state space at time t is rich enough to characterize the possible future state of the system completely.

The evolution equation for the probabilities $u(x,t)$ is found using conditional probability:

$$\begin{aligned} u(x,t+1) &= \Pr(X(t+1) = x) \\ &= \sum_{y \in \mathcal{S}} \Pr(X(t+1) = x | X(t) = y) \cdot \Pr(X(t) = y) \\ u(x,t+1) &= \sum_{y \in \mathcal{S}} p(y,x) u(y,t) \quad . \end{aligned} \tag{2}$$

To express this in matrix form, we suppose that the state space, $\mathcal{S}$, is finite, and that the states have been numbered $x_1, \ldots, x_n$. The transition matrix, P, is $n \times n$ and has (i,j) entry $p_{ij} = p(x_i, x_j)$. Sometimes it is also natural to write $p_{xy} = p(x,y)$. With this convention, (2) can be interpreted as vector–matrix multiplication if we define a *row* vector $\underline{u}(t)$ with components $(u_1(t), \ldots, u_n(t))$, where we have written $u_i(t)$ for $u(x_i,t)$. As long as ordering is unimportant[2], we could also write $u_x(t) = u(x,t)$. Now, (2) can be rewritten

$$\underline{u}(t+1) = \underline{u}(t) P \quad . \tag{3}$$

Since $\underline{u}$ is a row vector, the expression $P\underline{u}$ does not make sense because the dimensions of the matrices are incompatible for matrix multiplication. The previous relation can be used repeatedly to yield

$$\underline{u}(t) = \underline{u}(0) P^t \quad , \tag{4}$$

where P^t means P to the power t (for transpose of P, it is $P^\top$).

[2]If you start programming in the computer, you need to order the states.

2.1.2 Backwards

There are several situations in which expected (present values of) payouts can be computed using an evolution equation that has time moving backwards from the future to the present. The basic idea comes through an undiscounted terminal payout. At the terminal time, T, we get a payout that depends on the state of the system at that time: $f_T(X(T))$. We want to compute the expected value of this payout:

$$\mathbb{E}\left[f_T(X(T))\right]. \tag{5}$$

To compute this, we compute a connected collection of expectation values, $f(x,t)$, defined as

$$f(x,t) = \mathbb{E}\left[f_T(X(T))|X(t)=x\right] \quad . \tag{6}$$

We find a relationship between these numbers by considering one step of the Markov chain. If the system is in state x at time t, then the probability for it to be at state y at the next time is $p(x \to y) = p(x,y)$. For expectation values, this implies

$$\begin{aligned} f(x,t) &= \mathbb{E}\left[f_T(X(T))|X(t)=x\right] \\ &= \sum_{y\in\mathcal{S}} \mathbb{E}\left[f_T(X(T))|X(t+1)=y\right] \cdot \Pr\left(X(t+1)=y \mid X(t)=x\right) \\ f(x,t) &= \sum_{y\in\mathcal{S}} f(y,t+1)p(x,y) \quad . \end{aligned} \tag{7}$$

The final time values, $f(x,T)$ are the given values $f_T(x)$. From these, we compute all the numbers $f(x,T-1)$. Continuing like this, we eventually get to $t=0$. We may know $X(0)$, the state of the system at the current time. Otherwise we can use

$$\begin{aligned} \mathbb{E}\left[f_T(X(T))\right] &= \sum_{x\in\mathcal{S}} \mathbb{E}\left[f_T(X(T))|X(0)=x\right] \cdot \Pr\left(X(0)=x\right) \\ &= \sum_{x\in\mathcal{S}} f(x,0)u(x,0) \quad . \end{aligned}$$

All the values on the bottom line should be known.

As with the probability evolution equation (2), the equation for the evolution of the expectation values (7) can be written in matrix form. The difference from the probability evolution equation is that here we arrange the numbers $f_j = f(x_j,t)$ into a *column* vector,

$\underline{f}(t)$. The evolution equation for the expectation values is then written in matrix form as

$$\underline{f}(t) = P\underline{f}(t+1) \quad . \tag{8}$$

This time, the vector goes on the right:

$$\underline{f}(t) = P^{T-t}\underline{f}(T) \quad . \tag{9}$$

2.1.3 Duality

The *forward evolution equation* (2) and *the backward equation* (7) are connected through a duality relation. For any time t, we compute (6) as

$$\begin{aligned} \mathbb{E}\left[f_T(X(T))\right] &= \sum_{x\in\mathcal{S}} \mathbb{E}\left[f_T(X(T))|X(t)=x\right]\cdot \Pr(X(t)=x) \\ &= \sum_{x\in\mathcal{S}} f(x,t)u(x,t) \quad . \end{aligned} \tag{10}$$

For now, the main point is that the sum on the bottom line does not depend on t. Given the constancy of this sum and the u evolution equation, we can give another derivation of the f evolution equation. Start with

$$\sum_{x\in\mathcal{S}} f(x,t+1)u(x,t+1) = \sum_{y\in\mathcal{S}} f(y,t)u(y,t) \quad .$$

Then use (2) on the left side and rearrange the sum:

$$\sum_{y\in\mathcal{S}} \left(\sum_{x\in\mathcal{S}} f(x,t+1)p(y,x) \right) u(y,t) = \sum_{y\in\mathcal{S}} f(y,t)u(y,t) \quad .$$

Now, if this is going to be true for any $u(y,t)$, the coefficients of $u(y,t)$ on the left and right sides must be equal for each y.

The evolution equations (2) and (7) have some qualitative properties in common. One is that they preserve positivity. If $u(x,t) \geq 0$ for all $x \in \mathcal{S}$, then $u(x,t+1) \geq 0$ for all $x \in \mathcal{S}$ also. Likewise, if $f(x,t+1) \geq 0$ for all x, then $f(x,t) \geq 0$ for all x. These properties are simple consequences of (2) and (7) and the positivity of the $p(x,y)$.[3]

[3]Positivity preservation does not work in reverse. It is possible, for example, that $f(x,t+1) < 0$ for some x even though $f(x,t) \geq 0$ for all x.

The probability evolution equation (2) has a *conservation law* not shared by (7). It is

$$\sum_{x\in\mathcal{S}} u(x,t) = \text{const} \ . \tag{11}$$

independent of t. This is natural if u is a probability distribution, so that the constant is 1. The expected value evolution equation (7) has a *maximum principle*

$$\max_{x\in\mathcal{S}} f(x,t) \leq \max_{x\in\mathcal{S}} f(x,t+1) \tag{12}$$

as $f(x,t)$ has a convex combination representation $\sum_{y\in\mathcal{S}} f(y,t+1)p(x,y)$. This is a natural consequence of the interpretation of f as an expectation value. The probabilities, $u(x,t)$ need not satisfy a maximum principle either forward or backward in time.

This duality relation has is particularly transparent in matrix terms. The formula 6 is expressed explicitly in terms of the probabilities at time T as

$$\sum_{x\in\mathcal{S}} f(x,T)u(x,T) \ \ ,$$

which has the matrix form[4]

$$\underline{u}(T)\underline{f}(T) \ \ .$$

We may rewrite this as

$$\underline{u}(0)P^T\underline{f}(T) \ \ .$$

Because matrix multiplication is associative, this may be rewritten

$$\left[\underline{u}(0)P^t\right]\cdot\left[P^{T-t}\underline{f}(T)\right] \tag{13}$$

for any t. This is the same as saying that $\underline{u}(t)\underline{f}(t)$ is independent of t, as we already saw.

In linear algebra and functional analysis, "adjoint" or "dual" is a generalization of the transpose operation of matrices. People who don't like to think of putting the vector to the left of the matrix think of $\underline{u}P$ as multiplication of (the transpose of) $\underline{u}$, on the right, by the transpose (or adjoint or dual) of P. In other words, we can do enough evolution to compute an expected value either using P its dual (or adjoint or transpose). This is the origin of the term "duality" in this context.

[4]Written in this order, the matrix multiplication is compatible; the other order, $\underline{f}(T)\underline{u}(T)$, would represent an $n\times n$ matrix instead of a single number.

Remark. (Dynamic programming) It is a method for making current decisions that effect the ultimate payout. The idea is to define the appropriate value function, $f(x,t)$, that satisfies a nonlinear version of the backwards evolution equation (7). I will explain the idea in an abstract sense. We have a Markov chain as before, but now the transition probabilities depend on a control parameter, ξ. That is

$$p(x,y,\xi) = \Pr\left(X(t+1) = y | X(t) = x, \xi\right) \quad .$$

We are allowed to choose the control parameter at time t, $\xi(t)$, knowing the value of $X(t)$ but not any more about the future than the transition probabilities. Because the system is a Markov chain, knowledge of earlier values, $X(t-1)$, ..., will not help predict or control the future. Choosing ξ as a function of $X(t)$ and t is a "decision". The point here is that the optimal control policy is a feedback control. That is, instead of trying to choose a whole control trajectory, $\xi(t)$ for $t = 0, 1, \ldots, T$, we instead try to choose the functions$\xi(X(t),t)$. We will write $\xi(X,t)$ for such a decision strategy. Any given strategy has an expected payout, the object is to compute $\mathbb{E}_\xi\left[f_T(X(T))\right]$ under the optimal decision strategy: $\max_\xi \mathbb{E}_\xi\left[f_T(X(T))\right]$. The principle of dynamic programming is: One replaces the multiperiod optimization problem with a sequence of hopefully simpler single period optimization problems. Conditional expected value is

$$f(x,t) = \max_\xi \mathbb{E}_\xi\left[f_T(X(T)) | X(t) = x\right] \quad . \tag{14}$$

Provieded that we have the condition $f(x,T) = f_T(x)$, we need to compute the values $f(x,t)$ in terms of already computed values $f(x,t+1)$. If we use control variable $\xi(t)$ at time t, and the optimal control thereafter, we maximize this expected payout over $\xi(t)$ gives the optimal expected payout at time t:

$$f(x,t) = \max_{\xi(t)} \sum_{y \in \mathcal{S}} f(y,t+1) p(x,y,\xi(t)) \quad . \tag{15}$$

which is the principle of dynamic programming.

3 Some Remarks

At the beginning, we mention the growth ODE. Now consider a similar but more general setting

$$\frac{dx}{dt} = u(x), \quad x(0) = x_0,$$

where u is a continuous function. When there is a unique solution $x(t)$ to this problem for all initial data x_0, then we can write

$$x(t) = \psi^t(x_0)$$

such that ψ^t is called the solution operator and forms a one-paramter group of operators

$$\psi^{t+s} = \psi^t \circ \psi^s, \ \forall t, s \in \mathbb{R} \text{ and } \psi^0 = \text{Id}$$

where Id is the identity operator. The inverse of ψ^t is ψ^{-t}. If h is Lipschitz continuous such that

$$|u(x) - u(x')| \leq L|x - x'|$$

for some constant L, then this ODE has a unique solution for any x_0. Moreover, we have the following inequality:

$$|\psi^t(x) - \psi^t(x')| \leq e^{tL}|x - x'|.$$

This result comes from Gronwall's lemma. Here is the lemma.

(Gronwall - differential form) Let $\eta(t)$ be a continuous function satisfying the differential inequality

$$\frac{d\eta(t)}{dt} \leq a\eta(t) + k(t), \ \eta(0) = \eta_0$$

then $\eta(t) \leq e^{at}(\eta_0 + \int_0^t e^{-as}k(s)ds)$.

(Gronwall - integral form) Let $\xi(t)$ be a continuous function satisfying the integral inequality

$$\xi(t) \leq a\int_0^t \xi(s)ds + b,$$

then $\xi(t) \leq b\exp(at)$.

This lemma is easy to prove. For the differential form, one can multiply it by $\exp(-at)$ so that

$$\left(\frac{d}{dt}\eta(t)\right)\exp(-at) \leq (k(t) + a\eta(t))\exp(-at)$$

which means

$$\left(\frac{d}{dt}\eta(t)\exp(-at)\right) \leq k(t)\exp(-at).$$

Integrating this inequality from 0 to t and then multiplying it by $\exp(at)$ gives the result. For integral form, one can use previous result by defining $\eta(t) = \int_0^t \xi(s)ds$. Then one has $\frac{d\eta(t)}{dt} \leq a\eta(t) + b$. Thus

$$\eta(t) \leq \frac{b}{a}\left(\exp(at) - 1\right)$$

by the previous result and $\eta_0 = 0$. As $\xi(t) \leq a\eta(t) + b$, the result follows.

Gronwall's lemma says that if a differential form (or a function) is bounded by the function itself (or an integral form), then the function is bounded exponentially. Lipschitz condition implies that

$$\frac{1}{2}\frac{d}{dt}|x - x'|^2 = \langle u(x) - u(x'), x - x'\rangle \leq L|x - x'|^2,$$

so $|x - x'|$ is exponentailly bounded. The solution $\varphi^t(x)$ and $\varphi^t(x')$ are also bounded exponentially.

Now suppose we are interested in some function f of x. The way of characterzing how this function f change with time is done by the generator $\mathbb{A}$:

$$\mathbb{A}f = u(x) \cdot \nabla f(x)$$

where ∇ denotes the gradient.[5] This generator comes from

$$\begin{aligned}\frac{d}{dt}f(x(t)) &= \nabla f(x(t)) \cdot \frac{dx}{dt}(t) \\ &= \nabla f(x(t)) \cdot u(x(t)) = \mathcal{L}f(x(t)).\end{aligned}$$

If f is positive then it is possible to use differential inequalities, such as the Gronwall lemma, to obtain bounds on $f(x)$. Then $f(x)$ is called a Lyapunov function.

Given an initial condition $\varphi(x) = f(x, 0)$, the generator induces a Cauchy problem

$$\frac{\partial f}{\partial t} = \mathbb{A}f, \quad f(x, 0) = \varphi(x).$$

It is also the a backward equation whose solution is $f(x, t) = (e^{t\mathcal{L}}\varphi)(x)$. [6] The generator $\mathbb{A}$

[5] If the problem is n-dimension, then $u(x) \cdot \nabla f(x) = \sum_{i=1}^n u(x_i)\frac{\partial f(x_i)}{\partial x_i}$.

[6] This is often referred to as the semigroup notation for the solution of a time-dependent linear operator equation.

can also be defined by the following limit

$$\mathbb{A}f = \lim_{t\to 0} \frac{e^{\mathbb{A}t}\varphi - \varphi}{t}.$$

The form $(e^{t\mathbb{A}}\varphi)(x)$ of the ODE equation is also the solution of the following linear PDE

$$\frac{\partial f}{\partial t} = \langle \nabla f(x,t),\ u(x) \rangle .$$

This representation is the same as the previous method of characteristics. It shows that the family of solutions of the nonlinear ODE as in the Cauchy problem can be represented via the solution of a linear PDE. Conversely soving the transport equation may induce a solution of a nonlinear system of ODEs.

Now consider $x_0 = X_0$ as a random variable. We place a probability measure on X_0 so that

$$\mathbb{E}[f(X_0)] = \int f(x)p(x,0)dx$$

is our interest. The generator $\mathbb{A}$ of f has an adjoint $\mathbb{A}^*$ for p. The time variation of the adjoint is characterized by the Liouville equation

$$\frac{\partial p}{\partial t} = \mathbb{A}^* p, \quad p(x,0) = p_0(x).$$

This is also sometimes refered to as the forward equation. Using the semigroup notation, the solution can be denoted by $p_t(x) = (e^{\mathbb{A}^* t}p_0)(x)$. Because $\mathbb{A}^*$ is the adjoint of $\mathbb{A}$, $e^{\mathbb{A}^* t}$ is the adjoint of $e^{\mathbb{A}t}$:

$$\mathbb{E}[f(X_t)] = \int f(x,t)p_0(x)dx = \int (e^{\mathbb{A}t}\varphi)(x)p_0(x)dx = \int (e^{\mathbb{A}^* t}p_0)(x)\varphi(x)dx.$$

Thus $\mathbb{E}[f(X_t)] = \int \varphi(x)p_t(x)dx = \int f(x,0)p_t(x)dx$.

4 Heat Equation

When the state $\mathcal{S}$ is continuous, we denote the transition probability function as $p = p(x, A)$, $x \in \mathcal{S}$, $A \in \mathcal{F}$ where $\mathcal{F}$ is the σ-algebra generated by $\mathcal{S}$. Let $(\mathcal{S}, \mathcal{F})$ be the state space of a Markov process X_t. Markov property now becomes:

$$\Pr(X_{t+1} \in B | \sigma(X_0, \ldots, X_t)) = p(X_t, B).$$

Notice that if we define,

$$\begin{aligned}
&\Pr(X_0 \in A_0) = u(A_0) \\
&\Pr(X_0 \in A_0, X_1 \in A_1) = \int_{A_0} p(x_0, A_1) u(dx_0) \\
&\Pr(X_0 \in A_0, X_1 \in A_1, X_2 \in A_2) = \int_{A_0} u(dx_0) \int_{A_1} p(x_0, dx_1) p(x_1, A_2)
\end{aligned}$$

and so on, then we have a sequence of distributions on $(\mathcal{S}, \mathcal{F}), (\mathcal{S}^2, \mathcal{F}^2), \ldots$.

If we define a linear operator[7]

$$\mathbb{T}u(B) = \int p(x, B) u(dx),$$

then $\mathbb{T}u$ is a new probability distribution. Similarly,

$$\mathbb{T}^t u(B) = \int p_t(x, B) u(dx) = \text{distribution of } X_n$$

where $p_t(x, B) = \Pr(X_t \in B | X_0 = x)$. In fact, we can describe $p_t(x, B)$ recursively as

$$p_t(x, B) = \int p_{t-1}(x, dy) p(y, B).$$

Using this fact, it is straightforward to verify the associative property that $\mathbb{T}(\mathbb{T}u) = \mathbb{T}^2 u$. One can extend this argument to $\mathbb{T}^{s+t} = \mathbb{T}^s \circ \mathbb{T}^t$ for any $s, t > 0$. For each bounded and measurable f,

$$\mathbb{T}^{t+s} f(x) = \int p_{s+t}(x, dy) f(y)$$

[7] A detail discussion of the linear operator can be in Section 6.2. Here one can think a linear operator as the general abstract analogue of a square matrix. That is, the transition probability $p(x, B)$ evaluated at n points of x is an $n \times n$ matrix.

$$= \int p_s(x, dz) \int p_t(z, dy) f(y) \qquad \text{Chapman-Kolmogrov equation}$$
$$= \int p_s(x, dz) \mathbb{T}^t(f(z)) = (\mathbb{T}^s \circ \mathbb{T}^t) f(x)$$

This is semigroup property.

Remark. The Chapman-Kolmogorov (CK) equation for Markov process says that

$$p_{t+\tau}(x, x') = \int_{\mathcal{S}} p_t(x, dy) p_\tau(y, x').$$

If $\{p_t, t \geq 0\}$ is the family of transition kernels of a Markov process, the Markov property guarantees the Chapman-Kolmogrov relation, so the family of opertators $\mathbb{T}$ associated with a Markov process is a semigroup. There exists a generator $\mathbb{A}$ of the transition semigroup $\mathbb{T}$ defined as

$$\mathbb{A}f := \lim_{t \downarrow 0} \frac{\mathbb{T}^t f - f}{t}$$

for suitable f.

By the homogenous property of Brownian motion, we can deduce a new Chapman-Kolmogorov equation

$$\frac{1}{\sqrt{2\pi t}} e^{-\frac{(x'-x)^2}{2t}} = p_t(x - x') = \int_{\mathcal{S}} p_t(x - y) p(y - x') dy$$

By definition and change of variables,

$$\begin{aligned} \mathbb{T}^t f(x) &= \int_{\mathbb{R}} p_t(x, y) f(y) dy \\ &= \int \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} f(y) dy \\ &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} f(x + \sqrt{t} z) dz \end{aligned}$$

As for the generator, for f with two continuous derivatives f' and f'' such that f'' is bounded,

$$\begin{aligned} \mathbb{A}f(x) &= \lim_{t \downarrow 0} \frac{\mathbb{T}^t f(x) - f(x)}{t} \\ &= \lim_{t \downarrow 0} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{f(x + \sqrt{t} z) - f(x)}{t} dz \end{aligned}$$

$$
\begin{aligned}
&= \lim_{t\downarrow 0} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{f'(x)\sqrt{t}z + f''(x+\kappa\sqrt{t}z)tz^2/2}{t} dz \\
&= \lim_{t\downarrow 0} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{f''(x+\kappa\sqrt{t}z)z^2}{2} dz \\
&= \frac{1}{2} f''(x)
\end{aligned}
$$

where $\kappa \in [0,1]$ is function of x and $\sqrt{t}z$, so as $t \downarrow 0$ there is the convergence $\kappa\sqrt{t}z \to 0$, hence $f''(x+\kappa\sqrt{t}z) \to f''(x)$ by continuity of f'', and the last step comes from $\mathrm{Var}(Z) = 1$ if $Z \sim \mathcal{N}(0,1)$.

The expected vaule can be defined as

$$
u(t,x) = \int_{\mathcal{S}} p_t(x-y)u(0,y)dy.
$$

Then we have

$$
\mathbb{A}u(t,x) = \lim_{t\downarrow 0} \frac{\mathbb{T}^t u(0,x) - u(0,x)}{t} = \lim_{t\downarrow 0} \frac{u(t,x) - u(0,x)}{t} = \frac{\partial u(t,x)}{\partial t}.
$$

As we have seen $\mathbb{A}f(x)$ for Brownian motion is equivaelent to $f''(x)/2$, we have the following equation

$$
\frac{\partial u(t,x)}{\partial t} = \frac{1}{2} \frac{\partial u(t,x)}{\partial^2 x}
$$

which is called heat equation or diffusion equation. This is a PDE problem.

Remark. We can give a random walk interpretation of heat equation. Consider the lattice $\mathbf{x} = (x_1, x_2)$ with spatial steps $h \in \mathbb{Z}$ and a collection of particles moving on it. Let $u_{i,j}$ be the number of particles at (i,j). The particles only move at times $t_k = k\triangle t$. Let $u^k_{i,j}$ be the number of particles at (i,j) at time t_k. For each step, a particle at (i,j) can only hop to the four adjacent grids $(i+1,j)$, $(i-1,j)$, $(i,j+1)$, $(i-1,j)$ with equal probability (1/4 each). Then

$$
u^{k+1}_{i,j} = \frac{1}{4}\left[u^k_{i+1,j} + u^k_{i-1,j} + u^k_{i,j+1} + u^k_{i,j-1}\right].
$$

Subtracting $u^k_{i,j}$ from both sides, we have

$$
u^{k+1}_{i,j} - u^k_{i,j} = \frac{1}{4}\left[u^k_{i+1,j} + u^k_{i-1,j} + u^k_{i,j+1} + u^k_{i,j-1} - 4u^k_{i,j}\right].
$$

Let $\Delta t = h^2/4$, we obtain

$$\frac{u_{i,j}^{k+1} - u_{i,j}^k}{\Delta t} = \frac{\left[u_{i+1,j}^k + u_{i-1,j}^k + u_{i,j+1}^k + u_{i,j-1}^k - 4u_{i,j}^k\right]}{h^2}.$$

The left hand side looks like $u_t(t,x)$. Note that by Taylor's theorem

$$\begin{aligned} f(x+\epsilon) =& f(x) + \epsilon f_x(x) + \frac{\epsilon^2}{2} f_{xx}(x) + o(\epsilon^2), \\ f(x-\epsilon) =& f(x) - \epsilon f_x(x) + \frac{\epsilon^2}{2} f_{xx}(x) + o(\epsilon^2), \end{aligned}$$

where $o(\epsilon^2) \to 0$ as $\epsilon \to 0$. If we add the two equations and let $\epsilon \to 0$, we have

$$f_{xx}(x) = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) + f(x-\epsilon) - 2f(x)}{\epsilon^2}.$$

Now we can expect[8]

$$\lim_{h \to 0} \frac{\left[u_{i+1,j}^k + u_{i-1,j}^k - 2u_{i,j}^k + u_{i,j+1}^k + u_{i,j-1}^k - 2u_{i,j}^k\right]}{h^2} = u_{xx}(\mathbf{x}, t)$$

The result follows.

[8] This method is called the five-point stencil finite difference method. It corresponds to the matrix $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, the discrete Laplacian operator.

5 Analytical Solutions of PDE

Recall that an ordinary differential equation is an equation in which the derivative of a variable appears. Let $X(t)$ be a function of variable t whose derivative is $dX(t)/dt$. A differential equation consists of $dX(t)/dt$. The differential equation for a growth rate model is

$$\frac{dX}{dt}(t) = r.$$

The solution of this equation is $X(T) = X(0) + rT$ which can be easily found using the tools of calculus $\int_0^T dX(t) = \int_0^T r dt$ giving $X(0)$. The principle of causal relations among phenomena finds its simplest mathematical expression by means of differential equations.[9]

Remark. (Differential equation and Markovian) One can easily see that the differential equation $dX(t)/dt = f(X(t))$ is in a Markov process. The equation says that the rate of change of X at time t depends only on X at t and not on $X(\tau)$, $\tau < t$. As a result of this, for $t_1 < t_2$, the solution at t_2 is a function of x_{t_1} and does not depend on x_τ for $\tau < t_1$. The Markov property says that the probability law of the process in the future, once it is in a given state, does not depend on how the process arrived at the given state. Namely, the future can be predicted from a knowledge of the present.

5.1 Separation of Variables

A power series about x_0 is an infinite sum of the form

$$a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots = \sum_{n=0}^{\infty} a_n (x - x_0)^n,$$

where the a_n's are constants. Many functions can be represented efficiently as such an infinite series. For example

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}x^n, \tag{16}$$

and the trigonometric functions,

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots = \sum_{k=0}^{\infty} (-1)^k \frac{1}{(2k)!}x^{2k}$$

[9] The greater part of the laws of nature discovered at the time of the birth of mathematical physics are expressed in just such a manner.

and

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots = \sum_{k=0}^{\infty}(-1)^k \frac{1}{(2k+1)!}x^{2k+1}.$$

An infinite series of this type is called a power series. In order for a power series to be useful, the infinite sum must actually converge to a finite number, at least for some values of x. Let s_N be the sum of the first $(N+1)$ terms,

$$s_N = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_N(x - x_0)^N = \sum_{n=0}^{N} a_n(x - x_0)^n.$$

We say that the power series $\sum_{n=0}^{\infty} a_n(x - x_0)^n$ converges if the sum s_N approaches a finite limit as $N \to \infty$.

Power series can help us to solve differential equations. The equation of the following ordinary differential equation[10]:

$$\frac{d^2y}{dx^2} + y = 0, \tag{17}$$

Suppose that we don't know the general solution and want to find it by means of power series. We could start by assuming that

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots = \sum_{n=0}^{\infty} a_nx^n.$$

Assuming that the standard technique for differentiating polynomials also works for power series, we would expect that

$$\frac{dy}{dx} = a_1 + 2a_2x + 3a_3x^2 + \cdots = \sum_{n=1}^{\infty} na_nx^{n-1}.$$

(Note that the last summation only goes from 1 to ∞.) Differentiating again would yield

$$\frac{d^2y}{dx^2} = 2a_2 + 3 \cdot 2a_3x + 4 \cdot 3a_4x^2 + \cdots = \sum_{n=2}^{\infty} n(n-1)a_nx^{n-2}.$$

[10]It is called the simple harmonic motion that can be obtained by means of Newton's second law (and Hooke's law for a mass on a spring).

We can replace n by $m+2$ in the last summation so that

$$\frac{d^2y}{dx^2} = \sum_{m+2=2}^{\infty} (m+2)(m+2-1)a_{m+2}x^{m+2-2} = \sum_{m=0}^{\infty}(m+2)(m+1)a_{m+2}x^m.$$

The index m is a dummy variable in the summation and can be replaced by any other letter. We replace m by n and obtain the formula

$$\frac{d^2y}{dx^2} = \sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n.$$

Substitution into equation (17) yields

$$\sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}x^n + \sum_{n=0}^{\infty} a_n x^n = 0,$$

or

$$\sum_{n=0}^{\infty}[(n+2)(n+1)a_{n+2} + a_n]x^n = 0.$$

Now a polynomial is zero only if all its coefficients are zero. By analogy, we expect that a power series can be zero only if all of its coefficients are zero. Thus we must have

$$(n+2)(n+1)a_{n+2} + a_n = 0,$$

or

$$a_{n+2} = -\frac{a_n}{(n+2)(n+1)}. \tag{18}$$

This is called a recursion formula for the coefficients a_n.

The first two coefficients a_0 and a_1 in the power series can be determined from the initial conditions,

$$y(0) = a_0, \qquad \frac{dy}{dx}(0) = a_1.$$

Then the recursion formula can be used to determine the remaining coefficients in the power series by the process of induction. We will find that

$$y = a_0 + a_1x - \frac{1}{2!}a_0x^2 - -\frac{1}{3!}a_1x^3 + \frac{1}{4!}a_0x^4 + \cdots$$

$$= a_0 \left(1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots\right) + a_1 \left(x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots\right).$$

We recognize that the expressions within parentheses are power series expansions of the functions $\sin x$ and $\cos x$, and hence we obtain the solution,

$$y = a_0 \sin x + a_1 \cos x.$$

Using a similar idea, we can solve some PDE problems. First we need to transfer a PDE problem into an ordinary differential equation. We substitute $u(x,t) = p(x)g(t)$ into the heat equation and obtain

$$p(x)g'(t) = c^2 p''(x)g(t).$$

Now we separate variables, putting all the functions involving t on the left, all the functions involving x on the right:

$$\frac{g'(t)}{g(t)} = c^2 \frac{p''(x)}{p(x)}.$$

The left-hand side of this equation does not depend on x, while the right-hand side does not depend on t. Hence neither side can depend upon either x or t. In other words, the two sides must equal a constant, which we denote by λ and call the separating constant. Our equation now becomes

$$\frac{g'(t)}{c^2 g(t)} = \frac{p''(x)}{p(x)} = \lambda,$$

which separates into two ordinary differential equations,

$$\frac{g'(t)}{c^2 g(t)} = \lambda, \quad \text{or} \quad g'(t) = \lambda c^2 g(t), \tag{19}$$

and

$$\frac{p''(x)}{p(x)} = \lambda, \quad \text{or} \quad p''(x) = \lambda p(x). \tag{20}$$

The second problem is an eigenvalue problem for the differential operator

$$\triangle = \frac{d^2}{dx^2}.$$

One can apply power series to solve it.

5.2 Fourier Series

The solution of previous ordinary differential equation includes periodic functions:

$$y = a_0 \sin x + a_1 \cos x.$$

It gives us a hint that periodic properties may provide other uses in solving PDEs.

If a function $f(t)$ is periodic, then for some C,

$$f(t + C) = f(t)$$

for every t. The smallest C is the period. Hence $f(t + nC) = f(t)$ for every integer n. The most important family of periodic functions of time are sinusoids, $A\sin(\lambda t + b)$. A periodic function may be approximated by a finite sum of sinusoids, i.e.

$$f(t) = \sum_{n=0}^{N} A_n \sin(2\pi nt + \varphi_n).$$

Based on previous trigonometric relations, we have

$$\begin{aligned}\sum_{n=0}^{N} A_n \sin(2\pi nt + \varphi_n) &= \sum_{n=0}^{N} (A_n \sin\varphi_n \cos 2\pi nt + A_n \cos\varphi_n \sin 2\pi nt) \\ &= \sum_{n=0}^{N} (a_n \cos 2\pi nt + b_n \sin 2\pi nt) = \sum_{n=-N}^{N} c_n e^{2\pi int}.\end{aligned}$$

or $f(t) = \sum_{n=-N}^{N} c_n e^{2\pi int}$.[11] Then we may expect that any periodic function or even any function can be expressed as finitely or infinitely many sinusoids (complex exponentials).

Remark. Fourier transform and Fourier inverse transform:

$$\mathbb{F}f(n) = \int f(t) e^{-2\pi int} dt$$

is called a Fourier transform of $f(t)$. Under suitable conditions, f can be reconstructed from

[11] Recall the Euler's formula: $e^{ix} = \cos x + i \sin x$ and its relation to trigonometry $\cos x = (e^{ix} + e^{-ix})/2$ and $\sin x = (e^{ix} - e^{-ix})/2i$.

$\mathbb{F}f(n)$ by the inverse transform

$$f(t) = \int \mathbb{F}f(n)e^{2\pi int}dn = (\mathbb{F}^{-1}\mathbb{F})f(t)$$

for every real number t. I describe briefly about its origin. Suppose that $f(t) = \sum_{n=-N}^{N} c_n e^{2\pi int}$. We can determine the coefficients $\{c_k\}$. Note that

$$\begin{aligned}\frac{f(t)}{e^{2\pi ikt}} =& c_{-N}e^{2\pi i(-N-k)t} + \cdots + c_k + c_{k+1}e^{2\pi it} + c_N e^{2\pi i(N-k)t} \\ =& c_k + \sum_{n=-N, n\neq k} c_n e^{2\pi i(n-k)t}.\end{aligned}$$

By arranging the expression, we have

$$c_k = f(t)e^{-2\pi ikt} - \sum_{n=-N,n\neq k} c_n e^{2\pi i(n-k)t} = \int f(t)e^{-i2\pi kt}dt.$$

To obtain the last equality, we take the integrals on both sides such that $\int c_k dt = c_k$ and

$$\int_0^1 e^{2\pi i(n-k)t}dt = \begin{cases}1, & n = k, \\ 0, & n \neq k.\end{cases}$$

Thus we have a dual representation of $f(t)$, $f(t) = \sum_{n=-\infty}^{\infty}(\mathbb{F}f)(n)e^{2\pi int}$ where $(\mathbb{F}f)(n) = c_n = \int f(t)e^{-2\pi int}dt$. Normally, t represents time and the transform variable n represents frequency.

In fact, for any (periodic) function $f(t)$ with $\int |f(t)|^2 dt < \infty$, it is known that there is

$$\lim_{N\to\infty} \int \left| f(t) - \sum_{n=-N}^{N} \mathbb{F}f(n)e^{2\pi int} \right|^2 dt = 0.$$

$\{e^{2\pi int}\}_N$ is called a Fourier series with coefficients $\mathbb{F}f(n)$, $n = -N, \ldots N$. $\sum_{n=-\infty}^{\infty} \mathbb{F}f(n)e^{2\pi int}$ is the best approximation to $f(t)$.

With this result, we can apply Fourier series to solve PDE. Recall the problem of the heat equation:

$$\frac{\partial u}{\partial t}(x,t) = \frac{1}{2}u_{xx}(x,t)$$

with $u(x,0) = p(x)$ at $t = 0$. Now apply Fourier transform to $u_{xx}(x,t)$ over the variable x:

$$\frac{\partial}{\partial t}\mathbb{F}u(\lambda,t) = -\frac{1}{2}\lambda^2\mathbb{F}u(\lambda,t)$$

and $\mathbb{F}u(\lambda,0) = \mathbb{F}p(\lambda)$ by the fact that the Fourier transform of $\partial_{xx}u(x,t)$ is $(-i\lambda)^2\mathbb{F}u(\lambda,t)$. The above problem is not a partial differential equation anymore but an ordinary differential equation. Solving this differential equation we have

$$\mathbb{F}u(t,\lambda) = c(\lambda)e^{-\lambda^2 t/2}$$

where $c(\lambda) = \mathbb{F}p(\lambda)$ by the Fourier transform of the initial condition. Therefore:

$$\begin{aligned}(\mathbb{F}^{-1}\mathbb{F})u(x,t) =& \mathbb{F}^{-1}\left[\mathbb{F}u(\lambda,0)e^{-\lambda^2 t/2}\right] \\ =& \frac{1}{\sqrt{2\pi t}}\int e^{-(x-y)^2/2t}p(y)dy.\end{aligned}$$

The second equality comes from the fact that inverse Fourier transform of $e^{-\lambda^2/2}$ is the Gaussian density. If the initial condition is $u(x,0) = f(x) = \delta_x$, then the solution is the Gaussian distribution $(2\pi t)^{-\frac{1}{2}}e^{-x^2/2t}$. It is also called the fundamental solution of the heat equation.

6 Duality

In linear algebra and functional analysis, "adjoint" or "dual" is a generalization of the transpose operation of matrices. In Section 2 Markov chain has

$$\underline{u}(T) = \underline{u}(0)P^t = \left[\left[P^t\right]^\top \left[\underline{u}(0)\right]^\top \right]^\top .$$

People who don't like to think of putting the vector to the left of the matrix think of $\underline{u}P$ as multiplication of (the transpose of) $\underline{u}$, on the right, by the transpose (or adjoint or dual) of P. In other words, we can do enough evolution to compute an expected value either using P its dual (or adjoint or transpose). This is the origin of the term "duality" in this context.

6.1 Duality in Diffusion Equations

Now we turn to continuous time and continuous state space. The state at time t is a vector, $X \in \mathbf{R}^n$ consisting of n components, or "factors", $X = (X_1, \ldots, X_n)$. The dynamics are given by the Ito differential equation

$$dX(t) = a(X(t))dt + b(X(t))dW \quad . \tag{21}$$

Here $W(t)$ is a vector of m independent standard Brownian motions. For each x, there is a drift, $a(x)$, and an $n \times m$ matrix $b(x)$, that is related to the volatility. There is no reason that m, the number of noise sources, should equal n, the number of factors, but there is no reason ever to have more noises than factors. The columns of b are vectors in $\mathbb{R}^n$. Column j gives the influence of noise j on the dynamics. If these columns do not span $\mathbf{R}^n$, then the diffusion is degenerate.

There are forward and backward evolution equations that are dual to each other. The forward equation is for $u(x,t)$, the probability density for $X(t)$. This is the *diffusion* equation

$$\partial_t u = -\sum_{j=1}^{n} \left(\partial_{x_j} a_j(x) u\right) + \frac{1}{2} \sum_{j,k=1}^{n} \partial_{x_j} \partial_{x_k} \left(\mu_{jk}(x) u\right) \quad . \tag{22}$$

The matrix of diffusion coefficients, μ, is related to b by

$$\mu(x) = b(x) \cdot b^\top(x) \quad . \tag{23}$$

We write $M^\top$ for the transpose of a matrix, M. The coefficients, $a_j(x)$ are the components of $a(\cdot)$.[12] The second term on the right of (22) involves two derivatives. The actual form (22) has the martingale property that, if there is no drift ($a \equiv 0$), then the expected value of X does not change with time. To see this, use (22) with $a = 0$ and compute

$$\begin{aligned} \partial_t \mathbb{E}\left[X(t)\right] &= \partial_t \int x u(x,t) dx \\ &= \int x \frac{1}{2} \sum_{j,k=1}^{n} \partial_{x_j} \partial_{x_k} \mu_{jk}(x) u(x,t) dx \\ &= 0 \ . \end{aligned}$$

The last line follow from the one above if you integrate by parts twice to put the two derivatives on the x. The result would generally not be zero if $a \neq 0$.

The drift term, $\partial_x au$, corresponds to the drift term, $a(X)dt$. It is easy to see what the term would be if a were constant (independent of x and t) and b were zero. In that case the solution of (21) would be $X(t) = X(0) + at$. For this reason, the probability density is also simply shifted with speed a: $u(x,t) = u(x - at, 0)$. A rough idea about the relation (22) between volatility and diffusion coefficients is given in the simplest case $n = m = 1$. The number μ should depend on b in some way. Observe that the diffusion governed by (21) will be unchanged of b if it is replaced by $-b$, because $W(t)$ is indistinguishable from $-W(t)$. This implies the formula $\mu = b^2$.

The simplest backward equation is for the expected payout starting at x at time t:

$$f(x,t) = \mathbb{E}\left[f_T(X(T)) | X(t) = x\right] \ .$$

More complicated expectations satisfy more complicated but related equations. The backward equation for f is

$$\partial_t f + \sum_j a_j(x) \partial_{x_j} f + \frac{1}{2} \sum_{jk} \mu_{jk}(x) \partial_{x_j} \partial_{x_k} f = 0 \ . \tag{24}$$

This is supplemented with initial data given at the final time, T, $f(x,T) = f_T(x)$, and determines $f(x,t)$ for $t < T$. Again, we can express unconditional expectation in terms of

[12] Because u is a probability density, the integral of $u(x,t)$ over x should be independent of t. That will happen if all the terms on the right of (22) are derivatives of something (i.e. $\partial_x(a(x)u)$ rather than $a(x)\partial_x u$). This is sometimes called conservation form.

conditional expectation starting from time t and the probability density for $X(t)$:

$$\mathbb{E}\left[f_T(X(T))\right] = \int f(x,t)u(x,t)dx \quad . \tag{25}$$

The fact that the right side is independent of t. Finally, f satisfies a maximum principle:

$$\min_y f(y,T) \leq f(x,t) \leq \max_y f(y,T) \quad \text{if } t < T.$$

The probability interpretation of f makes this obvious; the expected reward cannot be less than the least possible reward nor larger than the largest.

We can motivate a way in which (22) and (24) are consistent with each other using a PDE version of the argument. We just compute the expected value, at time 0, of the payout at time T by averaging over the possible states at some intermediate time, t. Using the probability density for $X(t)$, this gives

$$\begin{aligned} \mathbb{E}\left[f_T(X(t))\right] &= \int \mathbb{E}\left[f_T(X(T)) \mid X(t) = x\right] u(x,t)dx \\ &= \int f(x,t)u(x,t)dx \quad . \end{aligned}$$

Since the left side does not depend on t, the right side also must be independent of t. This leads to

$$\begin{aligned} 0 &= \frac{d}{dt}\int f(x,t)u(x,t)dx \\ &= \int \left\{(\partial_t f(x,t))u(x,t) + f(x,t)\partial_t u(x,t)\right\} dx \quad . \end{aligned}$$

The evolution equation (22) transforms this into

$$\begin{aligned} 0 = \int \Big\{ &\left(\partial_t f(x,t)\right) u(x,t) \quad - \quad f(x,t)\sum_{j=1}^{n} \partial_{x_j}\left(a_j(x)u(x,t)\right) \\ &+ \quad f(x,t)\frac{1}{2}\sum_{jk} \partial_{x_j}\partial_{x_k}\left(\mu_{jk}(x)u(x,t)\right) \Big\} dx \quad . \end{aligned}$$

Now we want to integrate by parts. Normally we would get terms with the derivatives on f and boundary terms. Here, there are no boundary terms and[13] $u \to 0$ as $x \to \infty$. Upon

[13]For example, if a and b are bounded it is impossible for $X(t)$ to "escape to ∞" in finite time. This

integrating by parts and grouping terms, we come to:

$$0 = \int \left(\partial_t f + \sum_j a_j(x) \partial_{x_j} f + \frac{1}{2} \sum_{jk} \mu_{jk}(x) \partial_{x_j} \partial_{x_k} f \right) u(x,t) dx \quad .$$

The simplest way for this integral to be zero automatically is for f to satisfy (24).

6.2 Abstract Duality

The relation between (22) and (24) may be stated more abstractly using the language of adjoint operators. This has the advantage of clarifying the relationship between the continuous time, continuous X version we are discussing for Markov processes and the discrete time and discrete X version for Markov chains. Recall that we distinguished between row and column vectors. The abstract version of this distinction is the distinction between a vector space, $\mathbb{V}$, and its dual space, $\mathbb{V}^*$. For example, if $\mathbb{V}$ is the space of row vectors, then $\mathbb{V}^*$ is the space of column vectors. The abstract relationship between a vector space and its dual is that elements of $\mathbb{V}^*$ are linear functionals on $\mathbb{V}$. A linear functional is a linear function that produces a number from a vector. In the row and column vector setting, a column vector, f produces a function on row vectors by taking row vector u to the matrix product $u \cdot f$. Any such abstract pairing is written (u, f).[14] Conversely, any such linear functional corresponds to a column vector. Summarizing, the spaces of row vectors and column vectors of a given dimension have a natural duality pairing given by the matrix product: $(u, f) = u \cdot f$. The matrix product $u \cdot f$ is the definition of (u, f) in this case.

In our current setting of functions of a continuous variable (or set of variables), x, the duality relation is defined by integration. If $f(x)$ is a (payout) function, then we may define a linear functional on (probability) functions by taking a function $u(x)$ to $\int u(x) f(x) dx$. That is, the duality relation given by

$$(u, f) = \int u(x) f(x) dx \quad .$$

Here again, the right side is the definition of the left side. We may be possibly able to think of functions u as infinite continuous rows and functions f as corresponding columns.

implies that $u(x)$ goes to zero as x goes to infinity.

[14] In quantum physics, the notations are often adapted to Dirac's type. (u, f) is written as $< u \mid f >$. The left part, $< u \mid$, would have been called a bra vector, and the right, $\mid f >$, a ket. Putting them together forms the Dirac bracket....

The general abstract analogue of a square matrix is a linear operator, which is a map from $\mathbb{V}$ to $\mathbb{V}$, or from $\mathbb{V}^*$ to $\mathbb{V}^*$. Now suppose we have a dual pair of vector spaces and a linear operator, $\mathbb{A}$, on $\mathbb{V}$, then there is a dual operator on V^*. The dual of $\mathbb{A}$ is written $\mathbb{A}^*$. If $f \in \mathbb{V}^*$, then $\mathbb{A}^* f$ is another element of $\mathbb{V}^*$. The duality relation for operators is that $(\mathbb{A}u, f) = (u, \mathbb{A}^* f)$ for every $u \in \mathbb{V}$ and $f \in \mathbb{V}^*$.

In the case of row and column vectors, a matrix, M, acts as a linear operator on row vectors by matrix multiplication from the right. That is $\mathbb{A}u$ is the row vector given by the matrix product $u \cdot M$. You should not think of $\mathbb{A}$ as a matrix, or $\mathbb{A}u$ as matrix vector multiplication, because u has the wrong shape multiplication from the left be a matrix. The dual of $\mathbb{A}$ is also given by the matrix M, this time acting on column vectors, f by matrix multiplication from the left. That is, $\mathbb{A}^* f$ is given by the matrix product $M \cdot f$, which is another column vector. The duality relation, $(\mathbb{A}u, f) = (u, \mathbb{A}^* f)$, in this case boils down to the associativity of matrix multiplication. First, $(\mathbb{A}u, f) = (u \cdot M, f) = (u \cdot M) \cdot f$, also $(u, \mathbb{A}^* f) = (u, M \cdot f) = u \cdot (M \cdot f)$. Because matrix multiplication is associative, $(u \cdot M) \cdot f = u \cdot (M \cdot f)$. Note, in this last formula, the parentheses refer to groupings in matrix multiplication rather than the duality pairing. This is a flaw in accepted mathematical notation that I am powerless to correct.

Now we come to the point of this subsection, the duality relation connecting the equations (2) and (5). If $u = u(x)$ is a function of the continuous variable x, then we can define the linear operator that gives the function $v(x)$ by

$$v(x) = -\sum_j^n \partial_{x_j} a_j(x) u(x) + \frac{1}{2} \sum_{jk} \partial_{x_j} \partial_{x_k} \mu_{jk}(x) u(x) \quad .$$

In operator notation, we might write

$$(\mathbb{A}u)(x) \text{ or } \mathbb{A}u(x) = -\sum_j \partial_{x_j} a_j(x) u(x) + \frac{1}{2} \sum_{jk} \partial_{x_j} \partial_{x_k} \mu_{jk}(x) u(x) \quad .$$

The evolution equation (22) is then $\partial_t u = \mathbb{A}u$.

We find the dual operator for $\mathbb{A}$ through the definition of the duality pairing and integration by parts. Using the notation$g = \mathbb{A}^* f$, we have, using integration by parts,

$$\begin{aligned} (\mathbb{A}u, f) &= (u, \mathbb{A}^* f) = (u, g) \\ \int \mathbb{A}u(x) f(x) dx &= (u, g) \end{aligned}$$

$$\int \left\{ -\sum_j \partial_{x_j} a_j(x) u(x) + \frac{1}{2} \sum_{jk} \partial_{x_j} \partial_{x_k} \mu_{jk}(x) u(x) \right\} f(x) dx = (u, g)$$

$$\int u(x) \left\{ \sum_j a_j(x) \partial_{x_j} f(x) + \frac{1}{2} \mu_{jk}(x) \sum_{jk} \partial_{x_j} \partial_{x_k} f(x) \right\} = \int u(x) g(x) dx \ .$$

Look at the last line here. If we want this to be true for every function $u(x)$, we should set the left parts of each side equal. That is

$$g(x) = \mathbb{A}^* f(x) = \sum_j a_j(x) \partial_{x_j} f(x) + \frac{1}{2} \mu_{jk}(x) \sum_{jk} \partial_{x_j} \partial_{x_k} f(x) \ .$$

The backward evolution equation (24) may now be written

$$\partial_t f = -\mathbb{A}^* f \ .$$

The derivation of (22) from (24) may be written abstractly too. For each t we have a vector u, which we write $u(t)$. Now $u(t)$ is a vector function of t rather than an ordinary function. That is, $u(t)$ is an element of $\mathbb{V}$ rather than being a single number. We similarly define $f(t) \in \mathbb{V}^*$. The expectation that is independent of t is $(u(t), f(t)) = \int u(x,t) f(x,t) dx$. Differentiating and using the abstract form of the u evolution equation, we get[15]

$$0 = \partial_t (u(t), f(t)) = (\partial_t u(t), f(t)) + (u(t), \partial_t f(t)) \ .$$

Using $\partial_t u = \mathbb{A} u$ and taking the dual, this gives

$$(u(t), \mathbb{A}^* f(t)) = (u(t), -\partial_t f(t)) \ .$$

The simplest way to make this true is to have $\partial_t f(t) = -\mathbb{A}^* f(t)$. This is the abstract form of the f evolution equation (24).

6.3 Conditions in PDE and Green's Function

Diffusion equations often come with boundary or initial conditions. In this case, the domain will be a subset of all possible x values. The boundary of the domain will be called B. Suppose that there will be no payout if $X(t) \in B$ for any $0 \leq t \leq T$. To value such an

[15] The product rule for differentiation also works for duality pairings as it does for ordinary products.

instrument, we use the backwards equation with Dirichlet boundary condition $f(x,t) = 0$ for $x \in B$. If the forward equation is applicable, we also apply the Dirichlet boundary condition to u. In the latter case, we usually have $\int u dx < 1$. This $u(x,t)$ represents the probability density for those paths that have never touched the boundary. The complimentary probability is the probability of touching the boundary at some time:

$$\int u(x,t)dx + \Pr\left(X(t') \in B \text{ for some } t' \leq T\right) = 1 \quad .$$

Most diffusions do not live in all of $\mathbb{R}^n$, but in a natural subset. For example, stock prices and interest rates are usually positive. In these cases, the diffusion coefficients may go to zero, as X gets close to the edge, in such a way that $X(t)$ can never leave the set. For example, in the model used in the Black-Scholes model, $dS = rSdt + \sigma SdW$, the $S(t)$ can never become negative if it starts positive. In these cases, no extra boundary conditions need to be specified.

We do not give a boundary condition at $S = 0$ when solving the Black-Scholes equation. Often the initial data, $f(x,T)$, or $u(x,0)$, are singular. A singularity is an x value where the function is not smooth. For example, if $X(0) = x_0$ is known, then the initial probability density is a delta function: $u(x,0) = \delta(x - x_0)$. The payout for a stock option has a jump in its derivative at the strike price.[16]

Although not essential to the modeling motivation, it is difficult to go past the derivation of the Green's function without some reference to the delta function $\delta(x)$. The delta function is like an administrative tool; it makes things easier to treat, once you have the confidence to use it, but at first sight it is very strange.

For a start it is not a function. It is called a generalized function. It has the property that it is zero everywhere except at $x = 0$, where it is infinite enough that $\int_{-\infty}^{\infty} \delta(x)\, dx = 1$. Obviously this is not really satisfactory. There are two formal ways of thinking about generalized functions. The first is that we think of a generalized function as being defined as a "limit" (more precisely, an equivalence class) of a family of smooth functions; for example, the delta function is associated with the family of functions $\sqrt{\frac{1}{\pi t}} e^{-x^2/t}$ as $t \to 0$.[17]

Remark. (Generalized Function) The main way of manipulating generalized functions is not to manipulate them directly at all, but to define their action by means of integrals. If $\{g_n(x)\}$ is a family of functions in the equivalence class of a generalized function $g(x)$, then we define

[16] If the diffusion is nondegenerate (the coefficient matrix, μ, is positive definite), such singularities quickly smooth out.

[17] Note that there are many such families.

the integral

$$\int_{-\infty}^{\infty} \varphi(x)k(x)\,dx = \lim_{n\to\infty} \int_{-\infty}^{\infty} \varphi(x)k_n(x)\,dx.$$

Using this definition, it is possible to prove that one can manipulate generalized functions directly by using this integral property, so that for example, one can define the delta function through the requirement that

$$\int_{-\infty}^{\infty} \varphi(x)\delta(x)\,dx = v(0)$$

for all smooth functions $\varphi(x)$.

Green's function is the impulse response of an inhomogeneous ordinary differential equation defined on a domain with specified initial conditions or boundary conditions. The convolution of a Green's function with the smooth function $\varphi(x)$ on that domain is the solution to the inhomogeneous differential equation for $\varphi(x)$. A Green's function, $G(x)$, of a linear differential operator $\mathbb{A}$ acting over a subset of the Euclidean space is any solution of

$$\mathbb{A}G(x) = \delta(x) \tag{26}$$

This property of a Green's function can be exploited to solve differential equations of the form. If the kernel of $\mathbb{A}$ is non-trivial, then the Green's function is not unique. Green's functions in general are distributions, not necessarily proper functions. It is a generalized function.

If a Green's function G can be found for the operator $\mathbb{A}$, then, if we multiply the equation (26) for the Green's function by f, and then integrate:

$$\int \mathbb{A}G(x-s)\varphi(s)\,ds = \int \delta(x-s)\varphi(s)\,ds = \varphi(x)$$

Suppose that for a Markov system, we know $\mathbb{A}u(x) = \varphi(x)$, a general linear equation. We can use Green's function to represent the system:

$$\mathbb{A}u(x) = \int \mathbb{A}G(x-s)\varphi(s)ds = \varphi(x)$$

which means, if the operator $\mathbb{A}$ and the integral $\int$ can interchange, $u(x) = \int G(x-s)\varphi(s)ds$.

Similarly, we can introduce time dimension to the Green's function. We can solve $\mathbb{A}u(x) =$

$\varphi(x)$ using the dynamical system

$$\frac{dG(x,t)}{dt} = -\mathbb{A}G(x,t) \quad , \quad G(x,0) = \varphi(x) \quad . \tag{27}$$

If $G(x,t) \to 0$ quickly enough as $t \to \infty$, then

$$u(x) = \int_{t=0}^{\infty} G(x,t)dt \tag{28}$$

satisfies $\mathbb{A}u(x) = \varphi(x)$.

Let's consider an example of deriving the expression of a Green's function. If we want to solve an initial value problem for Heat equation

$$u_t = \Delta u, \quad u(x,0) = \varphi(x).$$

Then the Green's function $G(x,t)$ should satisfy $G_t = \Delta G, \quad G(x,0) = \delta(x)$. As in Section 5.2, we know that by Fourier transfer on x, one has $\mathbb{F}G(\lambda,t) = e^{-\lambda^2 t}$ and hence

$$\left[\mathbb{F}^{-1}\mathbb{F}G\right](x,t) = G(x,t) = \frac{1}{(4\pi t)^{d/2}} e^{-|x|^2/4t}.$$

We can consider a more difficult problem for $G(x)$. Modifying the Heat equation wiath a decay term $m^2 G$, we have a diffusion deletion equation:

$$\partial_t G = \triangle G - m^2 G \quad , \quad G(x,0) = \delta(x) \quad . \tag{29}$$

The decay term, $-m^2 G$, is handled by including an additional exponential decay factor. The solution to (29) is

$$G(x,t) = \frac{1}{(4\pi t)^{d/2}} e^{-|x|^2/4t} e^{-m^2 t} \quad .$$

We can take the integral to get the desired integral representation for the stationary distribution $u(x)$:

$$u(x) = \frac{1}{m^2} \int_{t=0}^{\infty} m^2 e^{-m^2 t} dt \cdot \frac{1}{(4\pi t)^{d/2}} e^{-|x|^2/4t}.$$

Remark. This integral representation is useful in Monte Carlo simulation. It suggests a strategy for sampling stationary distribution. Notice that the first factor on the right is an exponential density with mean $1/m^2$ while the second is a Gaussian density in d-dimensions

with each component having variance $2t$. We have turned a seemingly hard d dimensional sampling problem into a much easier $d+1$ dimensional sampling problem. Some of the most effective innovative Monte Carlo methods developed in recent years are based on clever enlargements of the sampling space.

Part II
Applications