Notes for Math 450
Elements of Stochastic Calculus

Renato Feres

These notes supplement the paper by Higham and provide more information on the basic ideas of stochastic calculus and stochastic differential equations. You will need some of this material for homework assignment 12 in addition to Higham's paper. There are many places where you can find this theory developed in greater detail and better than here. See, for example, [CKO]. Also note that the information you mostly need for the homework is contained in the second part of these notes. The section about the Langevin equation won't be needed.

1 Informal introduction to stochastic ODEs

We begin by developing an intuitive understanding of what stochastic ODEs are and what kinds of situations they can be used to model.

1.1 Processes driven by noise

The processes we wish to consider can be regarded as solutions to differential equations in which noisy terms or coefficients are present. A specific type of such equation with broad applicability is

$$X' = f(t, X) + g(t, X)\eta_t$$

where a solution $X(t) \in \mathbb{R}^n$ describes the state of a physical system at time t and $X'(t)$ is the time derivative of $X(t)$. Here η_t is meant to represent a source of "noise" or external randomness imposed on the system. Thus $g(t, X)\eta_t$ is a noisy driving factor superposed to a deterministic "drift" term represented by $f(t, X)$. A general first order system of ordinary differential equations of the deterministic (i.e., standard) kind is then one for which $g(t, X)$ zero.

Before getting into the mathematics that underlies a calculus with noise (which I take here to essentially mean the Itô calculus to be introduced in the next few sections), we should form a preliminary idea of what we want the source of noise, η_t, to represent. As we will see, $t \mapsto \eta_t$ cannot (in any naive way) be described as an ordinary stochastic process, but we can still think of it as, in a sense, being the derivative of a perfectly nice stochastic process, W_t, called *Brownian motion*, or *Wiener process*.

The term *white noise* was created to describe the sound made by current fluctuations due to thermal agitation of electrons in electronic devices. Physical Brownian motion relates to the erratic motion of small particles suspended on a liquid due to the random bombardment by liquid molecules. Similar random phenomena is seen in the erratic fluctuation in the price of some financial asset during a period of little overall market change, in which case the "microscopic" factors are the large number of individual financial transactions taking place in the stock exchange. The key point in all such cases is that a large number of individual random events happening mostly independently of each other have a statistically identifiable collective behavior that is manifested as noise. (The NYSE is for sure a very noisy place, but of course in this or most other cases I don't mean the term noise to refer to sound.) Since on the most basic level these kinds of phenomena relate to counting large numbers of random events, it makes sense to start our informal description of noise by briefly going back to the Poisson process.

Let N_t denote a Poisson process with parameter λ. N_t could describe, for example, the number of electrons that have arrived at the end of a wire by time t. The difference $\Delta N_t = N_{t+\Delta t} - N_t$ represents the number of arrivals during the interval $[t, t + \Delta t]$. Recall from lecture notes 3 that the mean value of ΔN_t is $E[\Delta N_t] = \lambda \Delta t$, so on average we have $\lambda \Delta t$ arrivals during a time interval of length Δt. Also recall that the variance of ΔN_t is equal to its mean, $\mathrm{Var}(\Delta N_t) = \lambda t$. When the number of arrivals is big enough on average (which for many practical purposes we can take to mean that $\lambda \Delta t$ is about 20 or greater), ΔN_t can be approximated by a normal random variable with mean and variance $\mu = \sigma^2 = \lambda \Delta t$. So we write

$$\Delta N_t \cong \lambda \Delta t + \sqrt{\lambda \Delta t} Z,$$

where Z is a standard normal random variable (with mean 0 and variance 1).

The numbers of Poisson arrivals during non-overlapping time intervals are independent random variables. If $Z_1, Z_2, \ldots$ are a sequence of independent standard normal random variables, then the time-discretized process N_{t_k}, $t_k = k\Delta t$, $k = 0, 1, 2, \ldots$, can be approximated (for large $\lambda \Delta t$) by

$$N_{t_k} \cong N_{t_{k-1}} + \lambda \Delta t + \sqrt{\lambda \Delta t} Z_k.$$

We assume $N_0 = 0$. Notice that the process V_t defined by

$$V_t = \frac{N_t - \lambda t}{\sqrt{\lambda}}$$

(approximately) has the following properties:

1. $V_0 = 0$;
2. If $s < t < u$, then $V_u - V_t$ and $V_t - V_s$ are independent random variables;
3. $V_t - V_s$ is a normal random variable with mean 0 and variance $t - s$.

The Wiener process W_t, which we already formally introduced in class (you can find it in section 4.4 of Norris's text), is a process that satisfies these properties exactly (in addition to the assumption that sample paths $t \mapsto W_t$ are continuous functions with probability 1). Thus, it makes sense to write:

$$N_t \cong \lambda t + \sqrt{\lambda} W_t.$$

If N_t represents the number of arrivals of electrons at the end of a wire, then the time derivative of N_t, if it could legitimately be taken, would describe the electric current as a random process:

$$I_t = N_t' = \lambda + \sqrt{\lambda}\eta_t$$

where η_t is now $\sqrt{\lambda}W_t'$, the time-derivative of the Wiener process. This is what we would like to think of as "noise." It should be clear, however, that there are difficulties in interpreting the derivative of W_t as a process. In fact the variance of the quotient $\Delta W_t/\Delta t$ is given by

$$\mathrm{Var}\left(\frac{\Delta W_t}{\Delta t}\right) = \mathrm{Var}(\Delta W_t)/\Delta t^2 = \Delta t/\Delta t^2 = 1/\Delta t,$$

which does not have a limit as $\Delta t \to 0$.

It turns out that, properly interpreted, the differential dW_t still makes sense. The key idea is contained in the definition of Itô integral, introduced later. We can now write the above differential equation as a stochastic differential

$$dX_t = f(t, X_t) + g(t, X_t)dW_t$$

which is interpreted in terms of stochastic integrals:

$$X_t - X_0 = \int_0^t f(s, X_s)ds + \int_0^t g(s, X_s)dW_s.$$

The definition of a stochastic integral will be given shortly.

1.2 W_t as limit of random walks

Continuing with our informal introduction, we wish now to see how the process W_t can be interpreted as the trajectory of a particle following random (Brownian) motion. Consider a random walk on the one-dimensional grid consisting of all integer multiples of a small positive number δ. We model the motion by a continuous time Markov process with set of states given by the numbers $k\delta$, $k \in \mathbb{Z}$, and transition rate constant for jumping forward or backward equal to $\lambda/2$. On average we have λ jumps per unit time, roughly half of them forward and half backward.

Let X_t denote the position along the grid at time t. We wish to pass to the limit as $\delta \to 0$ and $\lambda \to \infty$. It turns out to be necessary to impose a relation

between these two parameters in order to obtain a reasonable limit. We assume that λ and δ satisfy the relation

$$\lambda\delta^2 = \sigma^2$$

where σ is a constant.

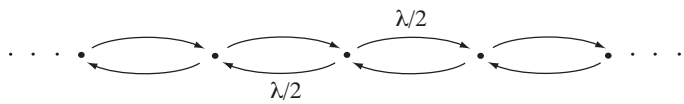

Figure 1: Random walk along the grid $\delta\mathbb{Z}$. The transition rate constants are $\lambda/2$. In particular, the holding times between consecutive jumps are exponential random variables with parameter λ.

The number of forward jumps by time t, N_t^+, and the number of backward jumps by time t, N_t^-, are independent Poisson processes with parameter $\lambda/2$. Using the approximation of a Poisson random variable by a normal random variable as described above we have:

$$\begin{aligned} N_t^+ &\cong \lambda t/2 + \sqrt{\lambda/2}W_1(t) \\ N_t^- &\cong \lambda t/2 + \sqrt{\lambda/2}W_2(t), \end{aligned}$$

where $W_1(t)$ and $W_2(t)$ are independent Wiener processes. We can now write:

$$\begin{aligned} X_t &= \delta\left[N_t^+ - N_t^-\right] \\ &\cong \delta\sqrt{\lambda}\left[\frac{W_1(t) - W_2(t)}{\sqrt{2}}\right] \\ &= \sigma W(t). \end{aligned}$$

Notice that the combination $(W_1(t) - W_2(t))/\sqrt{2}$ of two independent Wiener processes, which we have indicated by $W(t)$, also satisfies the conditions defining a Wiener process. This shows that if we scale δ down to zero and λ up to infinity in such a way that $\sigma^2 = \lambda\delta^2$ remains constant, then the random walk should be expected to converge (in some appropriate sense of convergence) to a Wiener process.

Since W_t is a normal random variable with mean 0 and variance t, σW_t is a normal random variable with mean 0 and variance $\sigma^2 t$. Its probability density function can thus be written as

$$\phi(t,x) = \frac{1}{\sigma\sqrt{2\pi t}}\exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2 t}\right).$$

This is the probability density that W_t will be found at position x at time t.

A simple partial derivatives exercise shows that ϕ satisfies the diffusion equation:

$$\frac{\partial \phi}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 \phi}{\partial x^2}.$$

Therefore, the Brownian motion described by W_t may be viewed as the "microscopic" motion of particles diffusing with a diffusion constant $D = \sigma^2/2$. We will show later in a more direct way that the probability density function associated to more general stochastic differentials must satisfy similar diffusion equations.

1.3 The chemical Langevin equation

Stochastic Petri nets provide a general source of examples to which the above ideas can be applied. Recall that the state of a Petri net at time t is given by a (column) vector $X_t = (X_1(t), \ldots, X_m(t))'$ whose components are the population numbers of the places (or species) of the net. We denote by $N_t = (N_1(t), \ldots, N_n(t))'$ the column vector whose components are the number of reaction events up to time t for each reaction type. Let $\Delta N_j(t)$ be the number of reaction events over the time interval $[t, t + \Delta t]$ corresponding to reaction r_j, $j = 1, \ldots, n$, and ΔN_t the column vector with components $\Delta N_j(t)$. Let U be the m-by-n stoichiometric matrix of the net. Then, as we saw before (lecture notes on Petri nets),

$$\Delta X_t = U \Delta N_t.$$

To obtain dynamical equations, we introduce the reaction rate functions $h_j(X_t)$, for $j = 1, \ldots, n$. The number of events of each reaction is then a (non-homogeneous) Poisson process with rate $h_j(X_t)$. This means that for small Δt, the increment $\Delta N_j(t)$ is approximately a Poisson$(h_j(X_t)\Delta t)$ random variable and increments for different time steps are independent. (It can be shown that the increment $N_b - N_a$ is also Poisson with parameter $\int_a^b h_j(X_t)dt$.)

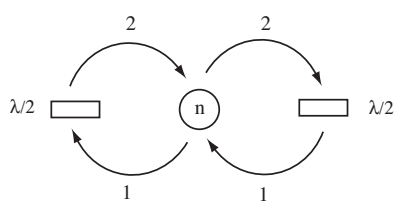

Figure 2: Petri net for the random walk with parameter $\lambda/2$ for forward and backward jump. The integer n represents the position along the grid $\mathbb{Z}\delta$,

The normal approximation of the Poisson process, $dN_t = \lambda dt + \sqrt{\lambda} dW_t$, gives a stochastic differential model for the number of reaction events:

$$dN_j(t) = h_j(X_t)dt + \sqrt{h_j(X_t)} dW_j(t).$$

Substituting into the equation for dX_t and using the differential notation gives a system of stochastic differentials:

$$dX_i(t) = \sum_{j=1}^{n} u_{ij} h_j(X_t)dt + \sum_{j=1}^{n} u_{ij}\sqrt{h_j(X_t)}dW_j(t).$$

These equations can be written more succinctly in matrix form as

$$dX_t = Uh(X_t)dt + U\sqrt{H(X_t)}dW_t$$

where h is the column vector of length n of reaction rate functions, H is the diagonal matrix with entries h_j, $\sqrt{H}$ is the diagonal matrix with entries $\sqrt{h_j}$, and $W_t = (W_1(t), \ldots, W_n(t))'$ is the column vector whose entries are independent Wiener processes.

A simple example is given by figure 2. Notice that this is simply the random walks example already discussed. The next section has a more elaborate example.

1.4 Example: the Michaelis-Menten enzyme model

The Michaelis-Menten system is a simple model for a chemical reaction in which a substrate S is converted to a product P only in the presence of a catalyst (enzyme) E. The process involves the elementary reactions:

$$\begin{aligned} r_1 &: S + E \to SE \\ r_2 &: SE \to S + E \\ r_3 &: SE \to P + E. \end{aligned}$$

A diagram is shown in figure 3.

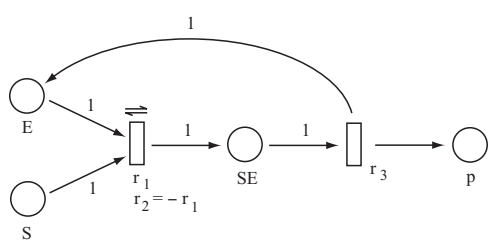

Figure 3: Petri net for the Michaelis-Menten system.

Let $X = (X_E, X_S, X_{SE}, X_P)$ be the population numbers for the different species. We order the molecular species as indicated by this vector. The stoichiometric matrix is

$$U = U_{post} - U_{pre} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 1 \\ -1 & 1 & 0 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

The rate functions are (using numerical indices rather than the molecular symbols)

$$
\begin{aligned}
h_1(X_t) &= k_1 X_1 X_2 \\
h_2(X_t) &= k_2 X_3 \\
h_3(X_t) &= k_3 X_3.
\end{aligned}
$$

This leads to the system of four stochastic differentials:

$$
\begin{pmatrix} dX_1 \\ dX_2 \\ dX_3 \\ dX_4 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 1 \\ -1 & 1 & 0 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} h_1 dt + h_1^{1/2} dW_1 \\ h_2 dt + h_2^{1/2} dW_2 \\ h_3 dt + h_3^{1/2} dW_3 \end{pmatrix}
$$

A numerical simulation of this system requires greater care than I am going to take here. The following simple program obtains sample paths for a chemical Langevin system with given stoichiometric matrices and reaction rate constants. The program assumes the mass-action kinetics. One problem with this particular implementation is that occasionally, if some of the population numbers become too small, the random fluctuations may turn them negative and render the result of the calculation completely meaningless. The program includes a parameter σ multiplying the terms dW_i, which can be adjusted to change the relative contribution of the noisy terms. (It can be thought of as a function of the system's temperature.) If $\sigma = 0$ the program reduces to the Euler method for systems of ordinary differential equations. Tinkering with σ may help evaluate whether something we are seeing is an artifact of this poor numerical implementation or something more likely to be a feature of the real solution.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [t,X]=chemlangevin(U_pre,U_post,c,X_0,T,N,sigma)
%Obtains a sample path for the solution X_t of a
%discretized chemical Langevin system
%associated to a stochastic Petri net with
%stoichiometric matrix U=U_post-U_pre, where
%U_pre is the m-by-n matrix of
%place-to-reaction multiplicities (there are
%m places and n reactions); U_post is the m-by-n
%matrix of reaction-to-place multiplicities, and
%c=[c1,...,cn] are the reaction constants. We assume
%mass-action kinetics. The initial
%state is a row vector X0 of length m, the time interval
%is [0, T] and the number of step sizes is N.
%The solution X(t) is given as a matrix of
%size m-by-N, where m is the number of species.
%sigma is a positive number to control the variance of
%the source Wiener processes. When sigma=0 the system
```

```
%reduces to the determninistic system.
dt=T/N;
[m n]=size(U_pre); %n=number of reactions; m=number of species
U=U_post-U_pre; %stoichiometric matrix
t=0:dt:T-dt;
X=zeros(m,N);
X(:,1)=X_0';
for k=1:N-1
    h=c.*prod(repmat(X(:,k),1,n).^U_pre);
    H=diag(h);
    X(:,k+1)=X(:,k)+U*h'*dt+sigma*U*sqrt(H)*randn(n,1)*sqrt(dt);
    if sum(h)<=0
        return
    end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

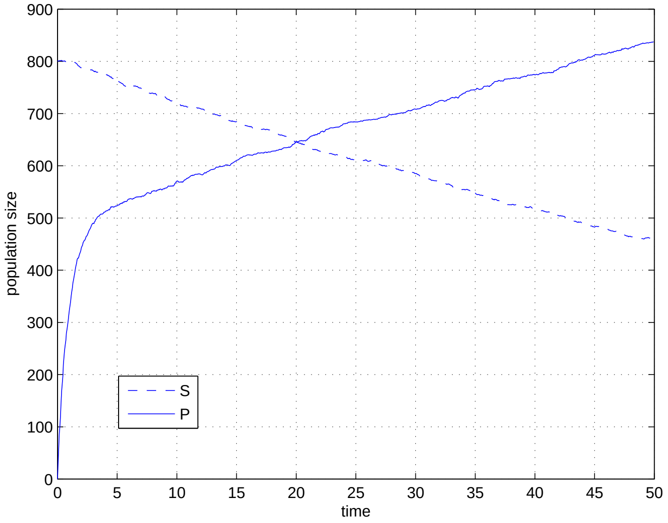

Figure 4: Population sizes of S and P as functions of time. The parameters are indicated below in the text.

We apply the program to the Michaelis-Menten system with the following parameters:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
U_pre=[1 0 0; 1 0 0; 0 1 1; 0 0 0];
```

```
U_post=[0 1 1; 0 1 0; 1 0 0; 0 0 1];
c=[10^(-5) 10^(-3) 1];
X_0=[600 800 500 0];
T=50;
N=1000;
sigma=1;
[t,X]=chemlangevin(U_pre,U_post,c,X_0,T,N,sigma);
hold off
plot(t,X(2,:),'--')
hold on
plot(t,X(4,:))
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

2 Temperature

3 The Itô integral

We begin now with a somewhat more formal development of Itô's stochastic calculus. The first and main issue to understand is how to make sense of integrals of the form

$$G = \int_0^T g(t)dW(t),$$

where $g(t)$ is a stochastic process and $W(t)$ is the Wiener process. As may be expected, we will define such an integral by a limit of Riemann sums. What will be different from ordinary calculus is the mode of convergence that will be used. (See definition of mean-square convergence below.)

The processes $g(t)$ for which a good integration theory can be obtained are those with the following two properties:

1. For each $t \geq 0$, the random variable $g(t)$ only depends on $(W_s)_{0\leq s\leq t}$, and not on any W_u for $u > t$. We say that $g(t)$ is *non-anticipating*. In other words, if t is the present time, we do not need to look into the future of the process $(W_u)_{u\geq 0}$ in order to know the value of $g(t)$.
2. The expected value of $g(t)$ is square-integrable over intervals where we wish to integrate $g(t)$ along $W(t)$. In other words,

$$\int_0^T E[g(t)^2]dt < \infty.$$

For example, $g(t) = W_t^2 - 3W_{t/2} + t^2$ satisfies the two properties for all $T > 0$.

Definition 3.1 (Mean-square convergence) A sequence $G_1, G_2, \ldots$ of random variables is said to converge in *mean-square* to a random variable G if

$$E\left[(G_n - G)^2\right] \to 0$$

as $n \to 0$. We indicate this by saying that $G_n \to G$ in mean-square.

Let now G_n be the Riemann sum

$$G_n = \sum_{j=1}^{n} g\left(t_j^{(n)}\right)\left(W\left(t_{j+1}^{(n)}\right) - W\left(t_j^{(n)}\right)\right)$$

for a partition $0 = t_1^{(n)} < t_2^{(n)} < \cdots < t_n^{(n)} = T$ of the interval $[0, T]$. It is important to notice that $g(t)$ is evaluated at the beginning of the interval. As we will see, the result could be completely different for a different choice of representative point.

It can be shown (we will not do it here) that if $g(t)$ satisfies the two properties above, then the sequence of Riemann sums G_n has a limit, G, in the sense of mean-square convergence. We call the limit the *Itô integral* of $g(t)$.

3.1 A simple example

As one example to illustrate the definition, we will show that

$$\int_0^T W_t dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}T.$$

The first term on the right-hand side is just what one would expect from our knowledge of ordinary integration. (Recall that $W_0 = 0$.) The presence of the second term, $T/2$, shows that something new is going on.

For simplicity of notation we drop the index n and write $\Delta W_{t_j} = W_{t_{j+1}} - W_{t_j}$ and $\Delta t_j = t_{j+1} - t_j$. We now write:

$$\begin{aligned}
W_T^2 &= W_T^2 - W_0^2 \\
&= \sum_{i=1}^{n}\left(W_{t_{i+1}}^2 - W_{t_i}^2\right) \\
&= \sum_{i=1}^{n}\left(W_{t_{i+1}} + W_{t_i}\right)\left(W_{t_{i+1}} - W_{t_i}\right) \\
&= \sum_{i=1}^{n}\left(W_{t_{i+1}} - W_{t_i}\right)^2 + 2\sum_{i=1}^{n} W_{t_i}\left(W_{t_{i+1}} - W_{t_i}\right) \\
&= \sum_{i=1}^{n}\left(W_{t_{i+1}} - W_{t_i}\right)^2 + 2G_n.
\end{aligned}$$

So to prove that G_n converges to $\frac{1}{2}W_T^2 - \frac{1}{2}T$ it is enough to show that

$$\sum_{i=1}^{n} \Delta W_{t_i}{}^2 \to T.$$

Write $u_i = \Delta W_{t_i}^2 - \Delta t_i$ and notice that $T = \sum_{i=1}^{n} \Delta t_i$. The limit we wish to establish amounts to $\sum_i u_i \to 0$ in mean-square. In other words, we need to

show that

$$E\left[\left(\sum_{i=1}^{n} u_i\right)^2\right] \to 0.$$

Observe that u_i and u_j are independent random variables if $i \neq j$ and have mean 0 so $E[u_i u_j] = E[u_i]E[u_j] = 0$. Also note the following facts: each ΔW_{t_j} is a normal random variable with mean 0 and variance Δt_j, so $\Delta W_{t_j} = \sqrt{\Delta t_j} Z$, where Z is a standard normal random variable. The expected value of an odd power of Z is 0 since it involves integrating an odd function over $\mathbb{R}$. Setting $c_k = E[Z^{2k}]$ (a number that we could compute explicitly if we needed), gives then

$$E[(\Delta W_{t_j})^n] = \begin{cases} 0 & \text{if } n \text{ is odd} \\ c_m(\Delta t_j)^m & \text{if } n = 2m. \end{cases}$$

Although not important for the argument, an exercise in ordinary integration gives the explicit value $c_2 = 3$. Putting all of this together yields:

$$\begin{aligned} E\left[\left(\sum_{i=1}^{n} u_i\right)^2\right] &= E\left[\sum_{i,j=1}^{n} u_i u_j\right] \\ &= \sum_{i,j=1}^{n} E[u_i u_j] \\ &= \sum_{i=1}^{n} E\left[u_i^2\right] \\ &= \sum_{i=1}^{n} E\left[(\Delta W_{t_i}^2 - \Delta t_i)^2\right] \\ &= \sum_{i=1}^{n} \left(E\left[\Delta W_{t_i}^4\right] - 2E\left[\Delta W_{t_i}^2\right]\Delta t_i + \Delta t_i^2\right) \\ &= \sum_{i=1}^{n} \left(3\Delta t_i^2 - 2\Delta t_i \Delta t_i + \Delta t_i^2\right) \\ &= 2\sum_{i=1}^{n} \Delta t_i^2 \\ &= 2T\Delta t. \end{aligned}$$

At the last step we are using the assumption that the intervals have equal step size, Δt. As $\Delta t \to 0$ (in the ordinary sense of the limit), the above expectation approaches 0. This concludes the proof.

What we have just shown justifies writing

$$\int_0^T dW_t^k = \begin{cases} W_t & \text{when } k = 1 \\ T & \text{when } k = 2 \\ 0 & \text{when } k \geq 3. \end{cases}$$

The essence of the above calculations can be summarized symbolically in the following multiplication table:

$$dW_t^2 = dt, \quad \text{and} \quad dt^2 = dt\, dW_t = dW_t\, dt = 0.$$

Recall that the standard deviation of dW_t is $\sqrt{dt}$ and that terms of order $(dt)^\alpha$, for $\alpha > 1$, can be disregarded in integration. Thus, it makes sense to write

$$(dX_t)^2 = (udt + vdW_t)^2 = v^2 dt.$$

Definition 3.2 (Itô process) Let W_t be a (one-dimensional) Wiener process. A one-dimensional Itô process, or *stochastic integral* is a stochastic process X_t of the form

$$X_t = X_0 + \int_0^t u(s)ds + \int_0^t v(s)dW_s.$$

Alternatively, we write $dX_t = u(t)dt + v(t)dW(t)$.

For example, from what we saw in the previous section, $X_t = W_t^2/2$ is an Itô process. Its stochastic differential is given by

$$d\left(\frac{1}{2}W_t^2\right) = \frac{1}{2}dt + W_t dW_t.$$

3.2 Itô's formula

Itô's formula can be thought of as the chain rule in stochastic calculus. It is used to obtain the stochastic differential of the composition $F(X_t)$, where $F(x)$ is a (twice) differentiable function and X_t is an Itô process. One difference between the deterministic and the stochastic chain rules is that the latter also involves the second derivatives of F, as we will see. We begin by stating the one-dimensional case. The higher-dimensional Itô formula will be presented later in the section.

More generally, we consider functions $F(t, x)$ that may also depend explicitly on t. All such functions will be assumed to have continuous partial derivatives up to order 2 or greater. Let $dX_t = u(t)dt + v(t)dW(t)$ be an Itô process and define the new process

$$Y_t = F(t, X_t).$$

The next theorem asserts that Y_t is also an Itô process and shows how to write it as a stochastic differential.

Theorem 3.1 (One-dimensional Itô formula) *Let $F(t, x)$ be a twice continuously differentiable function jointly in t and x. Let X_t be an Itô process given by $dX_t = udt + vdW_t$, and define $Y_t = F(t, X_t)$. Then Y_t is also an Itô process and*

$$dY_t = \frac{\partial F}{\partial t}(t, X_t)dt + \frac{\partial F}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t, X_t)(dX_t)^2.$$

Using that $(dX_t)^2 = v^2 dt$, the differential of Y_t can be written as:

$$dY_t = \left(\frac{\partial F}{\partial t}(t, X_t) + u_t \frac{\partial F}{\partial x}(t, X_t) + \frac{1}{2} v_t^2 \frac{\partial^2 F}{\partial x^2}(t, X_t) \right) dt + v_t \frac{\partial F}{\partial x}(t, X_t) dW_t.$$

The proof of Itô's formula will be sketched in the next section. We look now at a few examples.

Example 3.1 Let $F(t, x) = x^n$, $n \geq 2$, and $X_t = W_t$. Then

$$dW_t^n = nW(t)^{n-1} dW(t) + \frac{n(n-1)}{2} W(t)^{n-2} dt.$$

This shows that

$$\int_0^t W_s^{n-1} dW(s) = \frac{1}{n} W_t^n - \frac{n-1}{2} \int_0^t W_s^{n-2} ds.$$

If $n = 2$, this formula reduces to what we obtained earlier using the definition of the stochastic integral.

Example 3.2 (Integration by parts) If $F(t, x) = f(t)x$, where $f(t)$ is a differentiable function, $X_t = W_t$, and $Y_t = F(t, W_t) = f(t)W_t$, then by Itô's formula:

$$d(f(t)W_t) = f'(t)W_t dt + f(t) dW_t.$$

In integral form, this is the formula for integration by parts:

$$\int_0^t f(s) dW_s = f(t)W_t - \int_0^t f'(s) W_s ds.$$

As a special case, if $f(t) = t$, we obtain:

$$\int_0^t s dW_s = tW_t - \int_0^t W_s ds.$$

It is helpful to keep in mind that expressions like $\int f(t, W_t) dt$, for a continuous function $f(t, x)$, can be interpreted in terms of ordinary integration along individual sample paths of the process $u(t) = f(t, W_t)$. The apparatus of stochastic calculus comes in when trying to interpret integrals like $\int f(t, W_t) dW_t$. In this sense, the expression $\int W_t^2 dW_t = W_t^3/3 - \int W_t dt$ has indeed reduced the process $Y_t = \int W_t^2 dW_t$ to something we better understand.

3.3 The multi-dimensional Itô formula

Let $W_t = (W_1(t), \ldots, W_m(t)))'$ (in column vector form) represent the Wiener process in dimension m. The components $W_i(t)$ are independent standard Wiener processes. Let $u(t) = (u_1(t), \ldots, u_n(t))'$ and $v(t) = (v_{ij}(t))$, an n-by-m matrix valued process, and suppose that u and v have the properties required

for stochastic integration. An n-dimensional process X_t (in column form) is called *n-dimensional Itô process* if its stochastic differential is given by

$$dX_t = udt + vdW_t.$$

This is the matrix form of the following system:

$$\begin{cases} dX_1 = u_1 dt + v_{11} dW_1 + \cdots + v_{1m} dW_m \\ \vdots \\ dX_n = u_1 dt + v_{n1} dW_1 + \cdots + v_{nm} dW_m. \end{cases}$$

For the manipulation of stochastic differentials, it is useful to note the following symbolic multiplication table:

$$dW_i dW_j = \delta_{ij} dt, \quad dt\, dW_i = dW_i\, dt = 0, \quad dt^2 = 0,$$

where δ_{ij} is 1 if $i = j$ and 0 otherwise. The justification for these products is containing in the proof of Itô's formula.

Theorem 3.2 (The general Itô formula) *Let $dX(t) = udt + vdW(t)$ be an n-dimensional Itô process as defined above. Let $F(t,x) = (F_1(t,x), \ldots, F_p(t,x))$ be a twice continuously differentiable map from $[0,\infty) \times \mathbb{R}^n$ to $\mathbb{R}^p$. Then $Y_t = F(t, X_t)$ is also an Itô process, whose component $Y_j(t)$ is given by*

$$dY_k = \frac{\partial F_k}{\partial t}(t,X)dt + \sum_{i=1}^{n} \frac{\partial F_k}{\partial x_i}(t,X)dX_i + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 F}{\partial x_i \partial x_j}(t,X)dX_i dX_j.$$

We consider a few examples.

Example 3.3 Let $F(t,x) = f(x) \in \mathbb{R}$, for $x \in \mathbb{R}^m$, $X_t = W_t$ and define $Y_t = f(W_t)$. Then

$$dY_t = (\nabla f)(W_t) \cdot dW_t + \frac{1}{2}(\nabla^2 f)(W_t)dt.$$

Here ∇f denotes the gradient of f and $\nabla^2 f = \sum_i \frac{\partial^2 f}{\partial x_i^2}$ the Laplacian. In particular, if f is a harmonic function, i.e., $\nabla^2 f = 0$, Itô's formula reduces to the ordinary total differential formula:

$$d(f(W_t)) = \sum_i \frac{\partial f}{\partial x_i}(W_t)dW_i(t).$$

For example, let $f(x_1, x_2) = x_1^2 - x_2^2$. Then

$$d(W_1^2 - W_2^2) = 2W_1 dW_1 - 2W_2 dW_2.$$

As a general remark, if $X_t = \int_0^t u_s ds + \int_0^t v_s dW_s$ then $E[X_t] = \int_0^t E[u_s]ds$. To show that the second integral has zero expectation, note that each term $v_{t_j}(W_{t_{j+1}} - W_{t_j})$ in a Riemann sum approximation of $\int_0^t v_s dW_s$ satisfies

$$\begin{aligned} E[v_{t_j}(W_{t_{j+1}} - W_{t_j})] &= E[E[v_{t_j}(W_{t_{j+1}} - W_{t_j})|W_{t_j}]] \\ &= E[v_{t_j} E[(W_{t_{j+1}} - W_{t_j})|W_{t_j}]] \\ &= E[v_{t_j} 0] \\ &= 0. \end{aligned}$$

Example 3.4 We used before that $E[W_t^4] = 3t^2$. Here is a simple way to prove this fact. Set $c_k(t) = E[W_t^k]$. By Itô's formula,

$$dW_t^k = kW_t^{k-1}dW_t + (1/2)k(k-1)W_t^{k-2}dt.$$

Taking expected value and using the above remark about the expectation of an Itô differential gives:

$$c_k(t) = \frac{1}{2}k(k-1)\int_0^t c_{k-2}(s)ds.$$

It is easy to obtain explicitly that $c_1(t) = 0$, $c_2(t) = t$, and $c_k(0) = 0$ for all k. It follows that all odd numbered terms $c_{2j+1}(t)$ are 0 and

$$c_4(t) = 6\int_0^t sds = 3t^2.$$

Other terms can be found by induction.

Example 3.5 Let $W = (W_1, \ldots, W_n)$ be the standard Wiener process in $\mathbb{R}^n$, $n \geq 2$. The distance from $W(t)$ to the origin is given by

$$R(t) = |W(t)| = (W_1^2 + \cdots + W_n^2)^{1/2}.$$

A simple application of the Itô formula shows that

$$dR = \frac{1}{R}W \cdot dW + \frac{n-1}{2R}dt.$$

Example 3.6 Let $g_t \in \mathbb{R}^n$ be a process that it satisfies the required conditions for the Itô integral $\int_0^t g_s dW_s$ to make sense. Let $g_t \cdot g_t$ represent the standard inner product (dot product) of g_t with itself. Now define the process

$$Y_t = \exp\left(\int_0^t g_s dW_s - \frac{1}{2}\int_0^t g_s \cdot g_s ds\right).$$

I claim that Y_t is a process without drift (a Martingale). In other words, the differential dY_t does not contain the term in dt. In fact, set

$$X_t = \int_0^t g_s dW_s - \frac{1}{2}\int_0^t g_s \cdot g_s ds,$$

$F(x) = e^x$, and $Y_t = F(X_t)$. Then

$$\begin{aligned} dY_t &= F'(X_t)dX_t + \frac{1}{2}F''(X_t)(dX_t)^2 \\ &= e^{X_t}(dX_t + \frac{1}{2}(dX_t)^2) \\ &= Y_t(g_t dW_t - \frac{1}{2}g_t \cdot g_t dt + \frac{1}{2}g_t \cdot g_t dt) \\ &= Y_t g_t dW_t. \end{aligned}$$

3.4 Sketch of proof of the Itô formula

I will return to this later. Some indications of the proof will be given in class probably before I get a chance to finish this write-up. For now, the following exercise should give you a good indication of what is going on: write $G(t) = F(t, X_t)$ and develop $G(t + \Delta t) - G(t)$ in Taylor approximation up to order 2. Then form the Riemann sum and show that the equivalent integrated form the theorem holds. Limits should be taken in the sense of mean-square convergence. The new point is the need to keep terms of order 2 due to the fact that $dW_t^2 = dt$.

3.5 Stochastic ODEs

We now return to the study of stochastic differential equations. As a warm-up, we discuss a couple of examples.

Example 3.7 (Geometric Brownian motion) Consider the equation

$$dX_t = rX_t dt + \alpha X_t dW_t.$$

This can be viewed as a model for population growth (or a compound interest model), with a noisy term proportional to X_t. The case $\alpha = 0$ reduces to the equation $x' = rx$, which has solution $x(t) = x_0 e^{rt}$. We assume the initial condition $X_0 > 0$.

Due to the term dX_t/X_t, it is natural to first examine the Itô differential form of $d \ln X_t$. The Itô formula gives:

$$\begin{aligned} d \ln X_t &= \frac{dX_t}{X_t} - \frac{1}{2X_t^2}(dX_t)^2 \\ &= rdt + \alpha dW_t - \frac{1}{2}\left(\frac{dX_t}{X_t}\right)^2 \\ &= rdt + \alpha dW_t - \frac{1}{2}\left(rdt + \alpha dW_t\right)^2 \\ &= rdt + \alpha dW_t - \frac{\alpha^2}{2}dt \\ &= d\left(r - \frac{\alpha^2}{2}\right) + \alpha dW_t \end{aligned}$$

Integrating from 0 to t results in

$$\ln X_t = \ln X_0 + \left(r - \frac{\alpha^2}{2}\right) t + \alpha W_t.$$

By exponentiating both sides of the equation we obtain the solution

$$X_t = X_0 \exp\left((r - \alpha^2/2)t + \alpha W_t\right).$$

A process of the form $X_t = X_0 \exp(\mu t + \alpha W_t)$ is called a *geometric Brownian motion.*

Example 3.8 (Merton's Portfolio Selection) Let $X(t)$ denote the wealth of an investor at time t. The investor allocates a fraction w of the wealth in a risky asset and the remainder in a sure asset. the sure asset produces a rate of return s. The risky asset yields a rate of return μ (greater than s) with a variance σ^2 per unit time. In other words, the risky asset earns a return $dr(t)$ over time interval $[t, t + dt]$, where

$$dr = \mu dt + \sigma dW_t.$$

The change in investor's wealth over the same time interval is then

$$\begin{aligned} X(t + dt) &= X(t) + s(1 - w)X(t)dt + dr(t)wX(t) \\ &= X(t) + s(1 - w)X(t)dt + (\mu dt + \sigma dW(t))wX(t) \\ &= X(t) + [s(1 - w) + \mu w]X(t)dt + \sigma wX(t)dW(t). \end{aligned}$$

Consequently, the investor's wealth is described by an Ito differential equation of the form

$$dX = f(t, X)dt + \sigma(t, X)dW$$

where $f(t, x) = (s(1 - w) + \mu w)x$ and $\sigma(t, x) = \sigma wx$.

Assuming that s, w, μ, σ are all constants, we obtain that $X(t)$ is a geometric Brownian motion.

Example 3.9 (Forward contracts on a non-dividend-paying security) This discussion is taken from [Kao]. A forward contract is an agreement between two parties to buy or sell a security at a certain future time for a certain price, called the delivery price. The party to buy the security is said to assume a *long* position and the party to sell the security is said to assume a *short* position. Let t denote the current time and T the maturity date of the contract. Let $S(t)$ denote the price of the security at time t. Assume that $S(t)$ follows a geometric Brownian motion with expected return μ and volatility σ, i.e.

$$dS = \mu S dt + \sigma S dW.$$

Let $Y(t)$ denote the forward price at time t. The forward price at any time is the delivery price that would make the contract have a zero value. Let r be the

risk-free interest rate. We now consider the case in which the security does not yield any dividends.

For arbitrage opportunities to be absent, the forward and security prices must be related as follows:

$$Y(t) = S(t)e^{r(T-t)}.$$

Otherwise, if $Y(t) > S(t)\exp(r(T-t))$, an arbitrageur can borrow S dollars for a period of$T-t$ at the risk-free interest rate r, buy the security, and take a short position in the forward contract. At time T, the security is sold for $Y(t)$. After paying the loan of $S(t)\exp(r(T-r))$ the arbitrageur nets a profit of $Y(t) - S(t)\exp(r(T-t))$. If the inequality goes the other direction, a similar scenario can be constructed.

A simple application of Itô's lemma now gives

$$dY = (\mu - r)Y\,dt + \sigma Y\,dW.$$

Therefore, Y is also a geometric Brownian motion with an expected growth rate of $\mu - r$ and volatility σ.

Example 3.10 (Stochastic harmonic oscillator) First recall the deterministic case. We assume a spring-mass system with mass m, spring constant k and friction coefficient c.

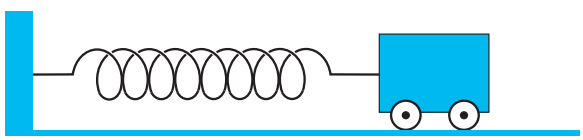

Figure 5: Spring-mass system. We assume that the mass attached to the spring is acted on by a random force term given by white noise.

Recall that the differential equation for this system (obtained from Newton's second law) has the form:

$$mx'' = -kx - cx' + f(t)$$

where $f(t)$ is a forcing term. Let $\kappa = k/m$, $\gamma = c/m$, and $\eta(t) = f(t)/m$. Let $v = x'$. The second order differential equation can be written as a system of two first order equations

$$\begin{aligned} x' &= v \\ v' &= -\kappa x - \gamma v + \eta(t). \end{aligned}$$

We assume that the forcing term is random, of the form $\eta(t)dt = \sigma dW_t$. This leads to the system of stochastic differential equations written in matrix form as:

$$\begin{pmatrix} dX_t \\ dV_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\kappa & -\gamma \end{pmatrix} \begin{pmatrix} X_t \\ V_t \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma \end{pmatrix} dW_t.$$

More generally, consider the system

$$dU_t = AU_t dt + BdW_t$$

where A is a constant square matrix, U_t is a vector-valued process of dimension n written in column form, B is a constant column vector of dimension n and W_t is Wiener process in $\mathbb{R}$.

Using Itô's formula, we can show

$$d\left(e^{-At}U_t\right) = e^{-At}BdW_t.$$

Some standard manipulation gives

$$U_t = e^{At}U_0 + \int_0^t e^{(t-s)A}BdW_s.$$

As a simple example, let $\gamma = 0$ and $\kappa = 1$. Also assume $U_0 = 0$. This gives

$$X_t = \sigma \int_0^t \sin(t-s)dW_s = \sigma \int_0^t \cos(t-s)W_s ds.$$

There is a general existence and uniqueness theorem for stochastic differential equations, which I'm not going to discuss in any detail at this point. (I may come back later and add a fuller discussion of stochastic ODEs here.) For now, I will only point out that the following conditions are enough to prove existence and uniqueness of solutions. Let $f(x)$ and $g(x)$ be differentiable function on $\mathbb{R}$ and suppose that the absolute values of their derivatives are bounded above by a finite constant. Suppose that $X(0) = x_0$ with probability 1. Then the Itô equation

$$X(t) = x_0 + \int_0^t f(X(s))ds + \int_0^t g(X(s))dW(s)$$

has a solution $X(t)$ defined for all t and the solution is unique.

Example 3.11 (Double-well potential) We consider a stochastic mechanical system with potential $V(x) = (x^2-1)^2$ and dissipative term $-cv$, where v is velocity and c a positive constant. The graph of $V(x)$ is shown in figure 6. The deterministic system has the equation

$$mx'' = -V'(x) - cx' + f(t)$$

where $f(t)$ is an external force. We assume that $f(t)$ is pure noise. We assume the stochastic model the velocity process):

$$\begin{aligned} dX_t &= V_t \\ dV_t &= [-4X_t(X_t^2-1) - cV_t]dt + \sigma dW_t. \end{aligned}$$

There are two stable equilibrium points for the deterministic motion: $x = 1$ and $x = -1$. The presence of noise can impart enough energy for the particle

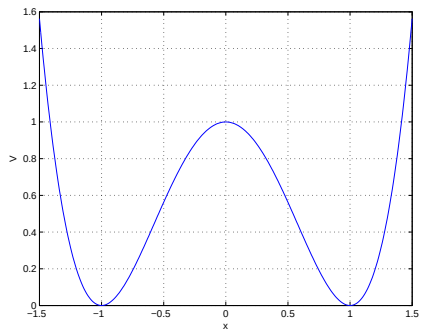

Figure 6: A double-well potential.

to cross the potential barrier from time to time. How often this will happen depends on the value of σ.

Figure 7 shows the particle motion over a period of time. A few transitions between equilibrium points are clearly shown.

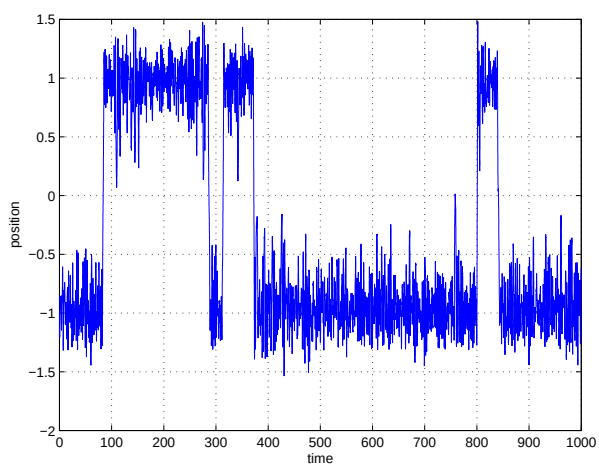

Figure 7: Occasionally the stochastic term yields enough energy for the particle to cross the potential barrier.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
s=.7;
```

```
c=1;
N=100000;
T=600;
dt=T/N;
Z=zeros(2,N);
Z(:,1)=[-1;0];
for i=1:N-1;
    x=Z(1,i);
    v=Z(2,i);
    Z(:,i+1)=Z(:,i)+[v;-4*x*(x^2-1)-c*v]*dt+[0;s*sqrt(dt)*randn];
end
t=0:dt:T-dt;
plot(t,Z(1,:))
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

4 Diffusion processes and PDEs

There is a direct relationship between stochastic differential equations and boundary value problems for parabolic partial differential equations (diffusion equations). This section provides a quick overview of this topic.

4.1 Kolmogorov's backward equation

We consider a time homogeneous stochastic differential equation

$$dX = f(X)dt + \sigma(X)dW.$$

Fix a twice continuously differentiable function $h(x)$. We think of $h(x)$ as a *test function* for observing the behavior of $X(t)$. In particular, we wish to study the expected value:

$$u(x,t) = E_x[h(X(t))].$$

The symbol E_x denotes expectation over all sample paths of X that satisfy $X(0) = x$.

Assume that the conditions for existence and uniqueness of solutions hold. We call the resulting process X_t an Itô diffusion. Itô diffusions satisfiy the by now familiar *Markov property*: the future behavior of the process given its history up to time a is the same as its behavior if it had started at X_a. More precisely, the following theorem holds. (See [Oks].)

Theorem 4.1 (Markov property for Itô diffusions) *Let X_t be an Itô diffusion in $\mathbb{R}^n$ and h a bounded Borel function from $\mathbb{R}^n$ to $\mathbb{R}$. Then for $a, t \geq 0$*

$$E_x[h(X_{t+a})|(X_s)_{s\leq a}] = E_{X_a}[h(X_t)].$$

We now show that $u(x,t)$ must satisfy a parabolic partial differential equation of the form $\frac{\partial u}{\partial t} = Lu$ where L is a second order differential operator which

we call the *infinitesimal generator* of X. But first, define this infinitesimal generator as follows:

Definition 4.1 (Generator of Itô diffusion) let X be an Itô diffusion in $\mathbb{R}^n$. The *infinitesimal generator* of X is the operation on functions defined by

$$(Lg)(x) = \lim_{a \to 0} \frac{E_x[g(X_a)] - g(x)}{a}$$

for all functions $g(x)$ for which this operation makes sense. (This will include all functions with compact support that are twice continuously differentiable.)

An explicit expression for L is obtained by applying Itô's formula (we denote by g_i, g_{ij}, etc., the partial derivatives of $g(x)$):

$$\begin{aligned}
\frac{E_x[g(X_a)] - g(x)}{a} &= E_x\left[\frac{1}{a}\int_0^a dg(X_s)\right] \\
&= \frac{1}{a}\int_0^a E_x\left[\sum_i g_i(X_s)dX_i + \sum_{i,j}\frac{1}{2}g_{ij}(X_s)dX_i dX_j\right] \\
&= \frac{1}{a}\int_0^a E_x\left[\sum_i f_i(X_s)g_i(X_s) + \sum_{i,j}\frac{1}{2}(\sigma\sigma^T)_{ij}(X_s)g_{ij}(X_s)\right] ds \\
&\to \sum_i f_i(x)g_i(x) + \sum_{i,j}\frac{1}{2}(\sigma\sigma^T)_{ij}(x)g_{ij}(x) \\
&= (Lg)(x) \text{ as } a \to 0.
\end{aligned}$$

The expression $\sigma\sigma^T$ corresponds to the product of the matrix σ and its transpose.

Therefore,

$$Lg = f \cdot \nabla g + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{ij}g_{ij}.$$

In particular, if $X = W$ is n-dimensional Wiener process

$$Lg = \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 g}{\partial x_i^2}$$

is (one-half times) the Laplacian. In dimension one, $Lg = f(x)g'(x) + \frac{1}{2}\sigma^2(x)g''(x)$.

Theorem 4.2 (Kolmogorov's backward equation) *Let X be an Itô diffusion satisfying the stochastic differential equation*

$$dX = f(X)dt + \sigma(X)dW$$

and define $u(x,t) = E_x[h(X(t)]$. Then

$$\frac{\partial u}{\partial t} = Lu(x,t)$$

and

$$u(x,0) = h(x).$$

for any twice continuously differentiable test function $h(x)$.

Proof. Fix $t \geq 0$. An application of the Markov property gives:

$$\begin{aligned}
\frac{E_x[u(X_a,t)] - u(x,t)}{a} &= \frac{1}{a} E_x\left[E_{X_a}[h(X_t)] - E_x[h(X_t)]\right] \\
&= \frac{1}{a} E_x\left[E_x[h(X_{t+a})|(X_s)_{s\leq a}] - E_x[h(X_t)|(X_s)_{s\leq a}]\right] \\
&= \frac{1}{a} E_x\left[E_x[h(X_{t+a}) - h(X_t)|(X_s)_{s\leq a}]\right] \\
&= \frac{1}{a} E_x[h(X_{t+a}) - h(X_t)] \\
&= \frac{E_x[h(X_{t+a})] - E_x[h(X_t)]}{a} \\
&= \frac{u(x,t+a) - u(x,t)}{a}.
\end{aligned}$$

This gives time derivative of $u(x,t)$:

$$\begin{aligned}
\frac{\partial u}{\partial t}(x,t) &= \lim_{a\to 0} \frac{E_x[u(X_a,t)] - u(x,t)}{a} \\
&= Lu(x,t).
\end{aligned}$$

This proves Kolmogorov's equation. □

4.2 The Dynkin formula

Essentially the same argument used above also proves:

Theorem 4.3 (Dynkin's formula) *Let g be a twice continuously differentiable function on $\mathbb{R}^n$ with compact support. Let τ be a stopping time with finite expectation: $E_x[\tau] < \infty$. Then*

$$E_x[g(X_\tau)] = g(x) + E_x\left[\int_0^\tau Lg(X_s)ds\right].$$

We give a couple of applications of this formula.

Example 4.1 (Expected hitting time I) Let X_t be standard Wiener process in $\mathbb{R}^n$ starting at x_0 and let $B(0,r)$ denote the ball of radius r and center 0. Suppose that $|x_0| = a < r$. Applying Dynkin's formula to $g(x) = |x|^2 = \sum_i x_i^2$ and τ the hitting time at the boundary of $B(0,r)$ gives (note that Lg is the one-half the Laplacian of g, which is the constant n):

$$r^2 = E_{x_0}[|X_\tau|^2] = a^2 + E_{x_0}\left[\int_0^\tau Lg(X_s)ds\right] = a^2 + nE_{x_0}[\tau].$$

Therefore, the expected hitting time at the boundary of the ball is

$$E_{x_0}[\tau] = \frac{r^2 - a^2}{n}.$$

If $X_t = W_t$ (starting at 0), the expected hitting time is r^2/n.

Example 4.2 (Expected hitting time II) Suppose that K is a bounded open connected region on $\mathbb{R}^n$ and let ∂K denote the boundary of K. Let τ be the hitting time of Brownian motion at ∂K. Let $g(x)$ be the solution of $Lg = -1$, where L is one-half the Laplacian in $\mathbb{R}^n$ with boundary value $g(x) = 0$ for $x \in \partial K$. Then $E_x[g(X_\tau)] = 0$ for all $x \in K$, and $0 = E_x[g(X_\tau)] = g(x) - E_x[\tau]$. This shows that the solution of this boundary value problem can be interpreted as the expected hitting time at ∂K for Brownian motion starting at $x \in K$, i.e.,

$$g(x) = E_x[\tau].$$

4.3 The Feynman-Kac formula

A similar argument to the one used to prove theorem 4.2 also gives the next theorem.

Theorem 4.4 (Feynman-Kac formula) *Assume the same notation of theorem 4.2, except that we now consider the expression*

$$v(x,t) = E_{x,t}\left[h(X(T))\exp\left(\int_t^T V(X(s),s)ds\right)\right].$$

Here $V(x,t)$ is a continuous function. (It may be thought of as a physical potential.*) Then $v(x,t)$ satisfies the partial differential equation*

$$\frac{\partial v}{\partial t} + Lv + Vv = 0$$

and the condition

$$v(x,t) \to h(x) \text{ as } t \to T.$$

Exercise 4.1 (Backward Feynman-Kac) Assume the same notation as of theorem 4.4, except that we now define

$$v(x,t) = E_{x,0}\left[h(X(t))\exp\left(\int_0^t V(X(s))ds\right)\right].$$

Show that $v(x,t)$ satisfies the equation

$$\frac{\partial v}{\partial t} = Lv + Vv$$

and $v(x,t) = h(x)$ for $t > 0$, $x \in \mathbb{R}$. (See section on Fokker-Plank equation.)

Exercise 4.2 Show that the function

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

satisfies

$$\frac{\partial u}{\partial t} + Lu + g = 0, \quad u(x,T) = 0$$

for $t < T$ and $x \in \mathbb{R}$.

4.4 The Fokker-Planck equation

Let $X(t)$ be the solution of the stochastic differential equation $dX = fdt + \sigma dW$, where f and σ satisfy the conditions for existence and uniqueness of solutions. Define

$$F(t,x,s,y) = P(X(s) \leq y | X(t) = x).$$

It can be shown that $F(t,x,s,y)$ is obtained from a transition probability density function $p(t,x,s,y)$, i.e.,

$$F(t,x,s,y) = \int_{-\infty}^{y} p(t,x,s,u)du.$$

The Markov property implies the property

$$p(t,x,s,y) = \int_{-\infty}^{\infty} p(t,x,t',u)p(t',u,s,y)dt'.$$

This is interpreted as saying that the probability that X goes from x to y in the time interval $[t,s]$ is the probability that X goes to any point u at any time t' and then, independently of the way it reached u, it goes to y in the time interval $[t',s]$

Theorem 4.5 (Fokker-Plank equation) *Let $X(t)$ be the solution to $dX = fdt + \sigma dW$ and let $p(t,x,s,y)$ be the transition probabilities density for X. Then $p = p(t,x,s,y)$ satisfies the partial differential equation*

$$\frac{\partial p}{\partial s} + \frac{\partial}{\partial y}[f(s,y)p] - \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(s,y)p] = 0$$

and $p(t,x,s,y) \to \delta(x-y)$ as $s \to t$.

In the above, $\delta(x-y)$ is the distribution (generalized function) defined by

$$\int_{-\infty}^{\infty} h(x)\delta(x-y)dx = h(y)$$

for any continuous function h.

An analogous formula can be derived by taking derivatives of p in t and x. The following is known as the *backward Kolmogorov equation*:

$$\frac{\partial p}{\partial t} + f(t,x)\frac{\partial p}{\partial y} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 p}{\partial y^2} = 0.$$

If f and σ are independent of t, we say that the equation $dx = fdt + \sigma dW$ is *time homogeneous*. This implies that

$$p(t,x,s,y) = p(0,x,s-t,y).$$

In particular,

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial s}.$$

Therefore, the backward Kolmogorov equation can be written in this case as

$$\frac{\partial p}{\partial t} = Lp$$

and the Fokker-Plank equation has the form

$$\frac{\partial p}{\partial t} = L^* p$$

where L^* is the adjoint operator of L.

We can use the backward Kolmogorov equation to derive a partial differential equation for

$$u(x,t) = E_{x,0}[h(X(t)] = \int_{-\infty}^{\infty} h(y)p(x,t,y)dy$$

where $p(x,t,y) = p(0,x,t,y)$ and $X(t)$ is solution to the time-homogenous stochastic differential equation.

4.5 The multi-dimensional case

The formulas derived in the previous sections have counterparts for processes in $\mathbb{R}^n$. The main change lies in the definition of the derivative operator L. If $X(t)$ is solution to the stochastic differential equation

$$dX = fdt + \sigma dW$$

where $f(t,x) \in \mathbb{R}^n$ is a vector-valued function of $t \geq 0$ and $x \in \mathbb{R}^n$ W is the Wiener process in $\mathbb{R}^p$, and $\sigma(t,x)$ is a n-by-p matrix, then

$$Lu = f(t,x)\cdot\nabla_x u + \frac{1}{2}\sum_{i,j}(\sigma\sigma^t)_{ij}(t,x)\frac{\partial^2 u}{\partial x_i \partial x_j}.$$

The symbol $\sigma\sigma^t$ represents the matrix product of σ and its transpose. It is an n-by-n matrix.

The operator L plays a similar role in this theory as the Q-matrix does for the theory of Markov chains. We call it the *infinitesimal generator* of $X(t)$. Note that if $X(t) = W(t)$, L reduces to (one-half times) the Laplace operator. In this sense, we say that the Laplace operator is the infinitesimal generator of Brownian motion. In general, the matrix $\frac{1}{2}\sigma\sigma^t$ is called the *diffusion matrix*. If this matrix is of the form DI where I is the identity matrix and D is a constant, then D is called the diffusion constant of the process.

4.6 Boundary value problems

Let $X(t)$ be the solution of the vector-valued stochastic differential equation $dX = fdt + \sigma dW$ in a bounded domain (an open and connected set contained in some big enough ball) $\mathcal{R} \subset \mathbb{R}^n$ and define the exit time from $\mathcal{R}$, starting at x at time s, as

$$\tau_{x,s} = \inf\{t \geq s : X(t) \in \partial\mathcal{R} \text{ and } X(s) = x\}.$$

We wish to find the expected value of the exit time: $E_{x,0}[\tau_{x,s}]$.

Let $u(x,t)$ be the solution of the problem:

$$\frac{\partial u}{\partial t} + Lu = -1, \quad t \geq s, x \in \mathcal{R}$$

and $u(x,t) = 0$ if $x \in \partial\mathcal{R}$. (The operator L is defined in the previous section.) By Itô's formula

$$u(X(t),t) = u(x,s) + \int_s^t \left(\frac{\partial u}{\partial t} + Lu\right) dt' + \int_s^t \nabla u \cdot (\sigma dW_{t'})$$

for all $s \leq t \leq \tau_{x,s}$. Setting $t = \tau_{x,s}$ and taking expectations,

$$0 = E[u(X(\tau_{x,s}), \tau_{x,s})] = u(x,s) + E\left[\int_s^{\tau_{x,s}} (-1)dt'\right] = u(x,s) - E[\tau_{x,s}] + s.$$

We have used the fact that $u(X(\tau_{x,s}), \tau_{x,s}) = 0$ since $X(\tau_{x,s})$ is a point on the boundary of $\mathcal{R}$. Therefore,

$$E[\tau_{x,s}] = s + u(x,s).$$

If the stochastic equation is time-homogenous (f and σ do not depend explicitly on t), $u(x,s) = u(x)$ is independent of s. In this case we write τ_x, with the understanding the the initial time is 0. Then we have

$$E[\tau_x] = u(x).$$

This proves the following theorem.

Theorem 4.6 (Dynkin's formula) *Let $u(x)$ be the solution to the boundary value problem $Lu = -1$ on a bounded domain $\mathcal{R}$, with boundary values 0, where L is the infinitesimal generator of a time homogeneous stochastic differential equation $dX = fdt + \sigma dW$. Then $u(x) = E_{x,0}[\tau_x]$ is the expected value of the first exit time from $\mathcal{R}$, starting at $x \in \mathcal{R}$.*

5 Distribution of exit points

Let $g(x)$ be any smooth function on $\partial\mathcal{R}$ and let $u(x,t)$ be the solution of the problem

$$\frac{\partial u}{\partial t} + Lu = 0$$

for $t \geq s$ and $x \in \mathcal{R}$, with boundary value $u(x,t) = g(x)$ for $x \in \partial\mathcal{R}$. Using Itô's formula, we obtain *Kolmogorov's formula*

$$E[u(X(\tau_{x,s}), \tau_{x,s})] = u(x,s).$$

Therefore,

$$u(x,s) = E[g(X(\tau_{x,s}))].$$

If the stochastic equation is time-homogeneous, we obtain the following theorem.

Theorem 5.1 (Solution to the Dirichlet problem) *Let L be the infinitesimal generator of a time homogeneous stochastic differential equation. Let $u(x)$ be the solution to the Dirichlet boundary value problem $Lu(x) = 0$ for x in a boundary domain $\mathcal{R}$ and $u(x) = g(x)$ for $x \in \partial\mathcal{R}$. Then $u(x) = E[g(X(\tau_x))]$ is the expected value of $g(X(\tau_x))$, where $X(t)$ is the solution to the stochastic differential equation with initial value (at time 0) x.*

6 Some ideas from mathematical finance

References

[CKO] S. Cyganowski, P. Kloeden, and J. Ombach. *From Elementary Probability to Stochastic Differential Equations with MAPLE*, Springer, 2002.

[Kao] E.P.C. Kao. *An Introduction to Stochastic Processes*,

[Oks] B. Øksendal. *Stochastic Differential Equations - An Introduction with Applications*, Springer, 1998.

[Sch] Z. Schuss. *Theory and Applications of Stochastic Differential Equations*, John Wiley & Sons, 1980.