Please cite as J. R. Movellan (2011) Tutorial on Stochastic Differential Equations, MPLab Tutorials Version 06.1.

Tutorial on Stochastic Differential Equations

Javier R. Movellan

This document is being reorganized. Expect redundancy, inconsistencies, disorganized presentation ...

1 Motivation

There is a wide range of interesting processes in robotics, control, economics, that can be described as a differential equations with non-deterministic dynamics. Suppose the original processes is described by the following differential equation

$$\frac{dX_t}{dt} = a(X_t) \tag{1}$$

with initial condition X_0, which could be random. We wish to construct a mathematical model of how the may behave in the presence of noise. We wish for this noise source to be stationary and independent of the current state of the system. We also want for the resulting paths to be continuous.

As it turns out building such a model is tricky. An elegant mathematical solution to this problem may be found by considering a discrete time versions of the process and then taking limits in some meaningful way. Let $\pi = \{0 = t_0 \leq t_1 \cdots \leq t_n = t\}$ be a partition of the interval $[0, t]$. Let $\Delta^\pi t_k = t_{k+1} - t_k$. For each partition π we can construct a continuous time process X^π defined as follows

$$X^\pi_{t_0} = X_0 \tag{2}$$

$$X^\pi_{t_{k+1}} = X^\pi_{t_k} + a(X^\pi_{t_k})\Delta^\pi t_k + c(X^\pi_{t_k})(N_{t_{k+1}} - N_{t_k}) \tag{3}$$

where N is a noise process whose properties remain to be determined and b is a function that allows us to have the amount of noise be a function of time and of the state. To make the process be continuous in time, we make it piecewise constant between the intervals defined by the partition, i.e.

$$X^\pi_t = X^\pi_{t_k} \text{ for } t \in [t_k, t_{k+1}) \tag{4}$$

We want for the noise N_t to be continuous and for the increments $N_{t_{k+1}} - N_{t_k}$ to have zero mean, and to be independently and identically distributed. It turns out that the only noise source that satisfies these requirements is Brownian motion. Thus we get

$$X^\pi_t = X_0 + \sum_{k=0}^{n-1} a(X^\pi_{t_k})\Delta t_k + \sum_{k=0}^{n-1} c(X^\pi_{t_k})\Delta B_k \tag{5}$$

where $\Delta t_k = t_{k+1} - t_k$, and $\Delta B_k = B_{t_{k+1}} - B_{t_k}$ where B is Brownian Motion. Let $\|\pi\| = \max\{\Delta t_k\}$ be the norm of the partition π. It can be shown that as $\pi \to 0$ the X^π processes converge in probability to a stochastic process X. It follows that

$$\lim_{\|\pi\|\to 0} \sum_{k=0}^{n-1} a(X^\pi_{t_k})\Delta_k = \int_0^t a(X_s)ds \tag{6}$$

and that

$$\sum_{k=0}^{n-1} c(X^\pi_{t_k})\Delta B_k \tag{7}$$

converges to a process I_t

$$I_t = \lim_{\|\pi\|\to 0} \sum_{k=0}^{n-1} c(X^\pi_{t_k})\Delta B_k \tag{8}$$

Note I_t looks like an integral where the integrand is a random variable $c(X_s)$ and the integrator ΔB_k is also a random variable. As we will see later, I_t turns out to be an Ito Stochastic Integral. We can now express the limit process X as a process satisfying the following equation

$$X_t = X_0 + \int_0^t a(X_s)ds + I_t \tag{9}$$

Sketch of Proof of Convergence: Construct a sequence of partitions $\pi_1, \pi_2, \cdots$ each one being a refinement of the previous one. Show that the corresponding $X_t^{\pi-i}$ form a Cauchy sequence in L_2 and therefore converge to a limit. Call that process X.

□

In order to get a better understanding of the limit process X there are two things we need to do: (1) To study the properties of Brownian motion and (2) to study the properties of the Ito stochastic integral.

2 Standard Brownian motion

By Brownian motion we refer to a mathematical model of the random movement of particles suspended in a fluid. This type of motion was named after Robert Brown that observed it in pollens of grains in water. The processes was described mathematically by Norbert Wiener, and is is thus also called a Wiener Processes. Mathematically a standard Brownian motion (or Wiener Process) is defined by the following properties:

1. The process starts at zero with probability 1, i.e., $P(B_0 = 0) = 1$

2. The probability that a randomly generated Brownian path be continuous is 1.

3. The path increments are independent Gaussian, zero mean, with variance equal to the temporal extension of the increment. Specifically for $0 \leq s_1 \leq t_1 \leq s_2, \leq t_2$

$$B_{t_1} - B_{s_1} \sim \mathcal{N}(0, s_1 - t_1) \tag{10}$$
$$B_{t_2} - B_{s_2} \sim \mathcal{N}(0, s_2 - t_2) \tag{11}$$

   and $B_{t_2} - B_{s_2}$ is independent of $B_{t_1} - B_{s_1}$.

Wiener showed that such a process exists, i.e., there is a stochastic process that does not violate the axioms of probability theory and that satisfies the 3 aforementioned properties.

2.1 Properties of Brownian Motion

2.1.1 Statistics

From the properties of Gaussian random variables,

$$\mathbb{E}(B_t - B_s) = 0 \tag{12}$$

$$\text{Var}(B_t - B_s) = \mathbb{E}[(B_t - B_s)^2] = t - s \tag{13}$$

$$\mathbb{E}((B_t - B_s)^4] = 3(t - s) \tag{14}$$

$$\text{Var}[(B_t - B_s)^2] = \mathbb{E}[(B_t - B_s)^4] - \mathbb{E}[(B_t - B_s)^2]^2 = 2(t - s)^2 \tag{15}$$

$$\text{Cov}(B_s, B_t) = s, \text{ for } t > s \tag{16}$$

$$\text{Corr}(B_s, B_t) = \sqrt{\frac{s}{t}}, \quad \text{for } t > s. \tag{17}$$

Proof: For the variance of $(B_t - B_s)^2$ we used the that for a standard random variable Z

$$\mathbb{E}(Z^4) = 3 \tag{18}$$

Note

$$\text{Var}(B_T) = \text{Var}(B_T - B_0) = T \tag{19}$$

since $P(B_0 = 0)$ and for all $\Delta_t \geq 0$

$$\text{Var}(B_{t+\Delta_t} - B_t) = \Delta_t \tag{20}$$

Moreover,

$$\begin{aligned} \text{Cov}(B_s, B_t) &= \text{Cov}(B_s, B_s + (B_t - B_s)) = \text{Cov}(B_s, B_s) + \text{Cov}(B_s, (B_t - B_s)) \\ &= \text{Var}(B_s) = s \end{aligned} \tag{21}$$

since B_s and $B_t - B_s$ are uncorrelated.

2.1.2 Distributional Properties

Let B represent a standard Brownian motion (SBM) process.

- Self-similarity:
  For any $c \neq 0$, $X_t = \frac{1}{\sqrt{c}} B_{ct}$ is SBM.

  We can use this property to simulate SBM in any given interval [0,T] if we know how to simulate in the interval $[0, 1]$:
  If B is SBM in [0,1], $c = \frac{1}{T}$ then $X_t = \sqrt{T}\, B_{\frac{1}{T}t}$ is SBM in [0,T].
- Time Inversion: $X_t = tB_{\frac{1}{t}}$ is SBM
- Time Reversal: $X_t = B_T - B_{T-t}$ is SBM in the interval $[0, T]$
- Symmetry: $X_t = -B_t$ is SBM

2.1.3 Pathwise Properties

- Brownian motion sample paths are non-differentiable with probability 1

This is the basic why we need to develop a generalization of ordinary calculus to handle stochastic differential equations. If we were to define such equations simply as

$$\frac{dX_t}{dt} = a(X_t) + c(X_t)\frac{dB_t}{dt} \tag{22}$$

we would have the obvious problem that the derivative of Brownian motion does not exist.

Proof: Let X be a real valued stochastic process. For a fixed t let $\pi = \{0 = t_0 \leq t_1, \cdots \leq t_n = t\}$ be a partition of the interval $[0, t]$. Let $\|\pi\|$ be the norm of the partition. The quadratic variation of X at t is a random variable represented as $< X, X >_t^2$ and defined as follows

$$< X, X >_t^2 = \lim_{\|\pi\| \to 0} \sum_{k=1}^{n} |X_{t_{k+1}} - X_{t_k}|^2 \tag{23}$$

We will show that the quadratic variation of SBM is larger than zero with probability one, and therefore the quadratic paths are not differentiable with probability 1.

Let B be a Standard Brownian Motion. For a partition $\pi = \{0 = t_0 \leq t_1, \cdots \leq t_n = t\}$ let B_k^π be defined as follows

$$B_k^\pi = B_{t_k} \tag{24}$$

Let

$$S^\pi = \sum_{k=1}^{n} (\Delta B_k^\pi)^2 \tag{25}$$

Note

$$\mathbb{E}(S^\pi) = \sum_{k=0}^{n-1} t_{k+1} - t_k = t \tag{26}$$

and

$$\begin{aligned} 0 \leq \mathrm{Var}(S^\pi) &= \sum_{k=0}^{n-1} \mathrm{Var}\left[(\Delta B_k^\pi)^2\right] \\ &= 2 \sum_{k=0}^{n-1} (t_{k+1} - t_k)^2 \\ &\leq 2\|\pi\| \sum_{k=0}^{n-1} (t_{k+1} - t_k) = 2\|\pi\| t \end{aligned} \tag{27}$$

Thus

$$\lim_{\|\pi\| \to 0} \mathrm{Var}(S^\pi) = \lim_{\|\pi\| \to 0} \mathbb{E}\left[\left(\sum_{k=0}^{n-1} (\Delta B_k^\pi)^2 - t\right)^2\right] = 0 \tag{28}$$

This shows mean square convergence, which implies convergence in probability, of S^π to t. (I think) almost sure convergence can also be shown.

Comments:

- If we were to define the stochastic integral $\int_0^t (dB_s)^2$ as

  $$\int_0^t (dB_s)^2 = \lim_{\|\pi\| \to 0} S^\pi \tag{29}$$

  Then

  $$\int_0^t (dB_s)^2 = \int_0^t d_s = t \tag{30}$$

- If a path $X_t(\omega)$ were differentiable almost everywhere in the interval $[0,T]$ then

$$< X, X >_t^2 (\omega)) \leq \lim_{\Delta_t \to 0} \sum_{k=0}^{n-1} (\Delta_t X'_{t_k}(\omega))^2 \tag{31}$$

$$= (\max_{t\in[0,T]} X'_t(\omega)^2) \lim_{n\to\infty} \sum \Delta_t^2 \tag{32}$$

$$= (\max_{t\in[0,T]} X'_t(\omega)^2) \lim_{n\to\infty} (n)(T/n)^2 = 0 \tag{33}$$

where $X' = dX/dt$. Since Brownian paths have non-zero quadratic variation with probability one, they are also non-differentiable with probability one.

2.2 Simulating Brownian Motion

Let $\pi = \{0 = t_0 \leq t_1 \cdots \leq t_n = t\}$ be a partition of the interval $[0,t]$. Let $\{Z_1, \cdots, Z_n\}$; be i.i.d Gaussian random variables $\mathbb{E}(Z_i) = 0; \quad \text{Var}(Z_i) = 1$. Let the stochastic process B^π as follows,

$$B^\pi_{t_0} = 0 \tag{34}$$

$$B^\pi_{t_1} = B^\pi_{t_0} + \sqrt{t_1 - t_0} Z_1 \tag{35}$$

$$\vdots \tag{36}$$

$$B^\pi_{t_k} = B^\pi_{t_{k-1}} + \sqrt{t_k - t_{k-1}} Z_k \tag{37}$$

Moreover,

$$B^\pi_t = B^\pi_{t_{k-1}} \text{ for } t \in [t_{k-1}, t_k) \tag{38}$$

For each partition π this defines a continuous time process. It can be shown that as $\|\pi\| \to 0$ the process B^π converges in distribution to Standard Brownian Motion.

2.2.1 Exercise

Simulate Brownian motion and verify numerically the following properties

$$\mathbb{E}(B_t) = 0 \tag{39}$$

$$\text{Var}(B_t) = t \tag{40}$$

$$\int_0^t dB_s^2 = \int_0^t ds = t \tag{41}$$

3 The Ito Stochastic Integral

We want to give meaning to the expression

$$\int_0^t Y_s dB_s \tag{42}$$

where B is standard Brownian Motion and Y is a process that does not anticipate the future of Brownian motion. For example, $Y_t = B_{t+2}$ would not be a valid integrand. A random process Y is simply a set of functions $f(t,\cdot)$ from an outcome space Ω to the real numbers, i.e for each $\omega \in \Omega$

$$Y_t(\omega) = f(t,\omega) \tag{43}$$

We will first study the case in which f is piece-wise constant. In such case there is a partition $\pi = \{0 = t_0 \leq t_1 \cdots \leq t_n = t\}$ of the interval $[0,.t]$ such that

$$f_n(t,\omega) = \sum_{k=0}^{n-1} C_k(\omega)\xi_k(t) \tag{44}$$

where

$$\xi_k(t) = \begin{cases} 1 & \text{if } t \in [t_k, t_{k+1}) \\ 0 & \text{else} \end{cases} \tag{45}$$

where C_k is a non-anticipatory random variable, i.e., a function of X_0 and the Brownian noise up to time t_k. For such a piece-wise constant process $Y_t(\omega) = f_n(t,\omega)$ we define the stochastic integral as follows. For each outcome $\omega \in \Omega$

$$\int_0^t Y_s(\omega)dB_s(\omega) = \sum_{k=0}^{n-1} C_k(\omega)\Big(B_{t_{k+1}}(\omega) - B_{t_k}(\omega)\Big) \tag{46}$$

More succinctly

$$\int_0^t Y_s dB_s = \sum_{k=0}^{n-1} C_k\Big(B_{t_{k+1}} - B_{t_k}\Big) \tag{47}$$

This leads us to the more general definition of the Ito integral

Definition of the Ito Integral Let $f(t,\cdot)$ be a non-anticipatory function from an outcome space Ω to the real numbers. Let $\{f_1, f_2, \cdots\}$ be a sequence of elementary non-anticipatory functions such that

$$\lim_{n\to\infty} \mathbb{E}[\int_0^t \Big(f(s,\omega) - f_n(s,\omega)\Big)^2 ds] = 0 \tag{48}$$

Let the random process Y be defined as follows: $Y_t(\omega) = f(t,\omega)$ Then the Ito integral

$$\int_0^t Y_s dB_s \tag{49}$$

is a random variable defined as follows. For each outcome $\omega \in \Omega$

$$\int_0^t f(s,\omega)dB_s(\omega) = \lim_{n\to\infty} \int_0^t f_n(t,\omega)dB_s(\omega) \tag{50}$$

where the limit is in $L_2(P)$. It can be shown that an approximating sequence $f_1, f_2 \cdots$ satisfying (48) exists. Moreover the limit in (50) also exists and is independent of the choice of the approximating sequence.

Comment Strictly speaking we need for f to be measurable, i.e., induce a proper random variable. We also need for $f(t,\cdot)$ to be $\mathcal{F}_t$ adapted. This basically means that Y_t must be a function of Y_0 and the Brownian motion up to time t It cannot be a function of future values of B. Moreover we need $\mathbb{E}[\int_0^t f(t,\cdot)^2 dt] \leq \infty$.

3.1 Properties of the Ito Integral

- $$\mathbb{E}(I_t) = 0 \tag{51}$$

- $$\mathrm{Var}(I_t) = \mathbb{E}(I_t^2) = \int_0^t \mathbb{E}(X_s^2) ds \tag{52}$$

- $$\int_0^t (X_s + Y_s)\, dB_s = \int_0^t X_s\, dB_s + \int_0^t Y_s\, dB_s \tag{53}$$

- $$\int_0^T X_s\, dB_s = \int_0^t X_s\, dB_s + \int_t^T X_s\, dB_s \text{ for } t \in (0, T) \tag{54}$$

- The Ito integral is a Martingale process

  $$\mathbb{E}(I_t \mid \mathcal{F}_s) = I_s l \text{for all } t > s \tag{55}$$

  where $\mathbb{E}(I_t \mid \mathcal{F}_s)$ is the least squares prediction of I_t based on all the information available up to time s.

4 Stochastic Differential Equations

In the introduction we defined a limit process X which was the limit process of a dynamical system expressed as a differential equation plus Brownian noise perturbation in the system dynamics. The process was a solution to the following equation

$$X_t = X_0 + \int_0^t a(X_s) ds + I_t \tag{56}$$

where

$$I_t = \lim_{\|\pi\| \to 0} c(X_{t_k}^{\pi}) \Delta B_k \tag{57}$$

It should now be clear that I_t is in fact an Ito Stochastic Integral

$$I_t = \int_0^t c(X_s) dB_s \tag{58}$$

and thus X can be expressed as the solution of the following stochastic integral equation

$$X_t = X_0 + \int_0^t a(X_s) ds + \int_0^t c(X_s) dB_s \tag{59}$$

It is convenient to express the integral equation above using differential notation

$$dX_t = a(X_t) dt + c(X_t) dB_t \tag{60}$$

with given initial condition X_0. We call this an Ito Stochastic Differential Equation (SDE). The differential notation is simply a pointer, and thus acquires its meaning from, the corresponding integral equation.

4.1 Second order differentials

The following rules are useful

$$\int_0^t X_t(dt)^2 = 0 \tag{61}$$

$$\int_0^t X_t dB_t dt = 0 \tag{62}$$

$$\int_0^t X_t dB_t dW_t = 0 \text{ if } B, W \text{ are independent Brownian Motions} \tag{63}$$

$$\int_0^t X_t(dB_t)^2 = \int_0^t X_t dt \tag{64}$$

$$\tag{65}$$

Symbolically this is commonly expressed as follows

$$dt^2 = 0 \tag{66}$$

$$dB_t dt = 0 \tag{67}$$

$$dB_t dW_t = 0 \tag{68}$$

$$(dB_t)^2 = dt \tag{69}$$

Sketch of proof:
Let $\pi = \{0 = t_0 \leq t_1 \cdots \leq t_n = t\}$ a partition of the $[0, t]$ with equal intervals, i.e. $t_{k+1} - t_k = \Delta t$.

- Regarding $dt^2 = 0$ note

$$\lim_{\Delta_t \to 0} \sum_{k=0}^{n-1} X_{t_k} \Delta t^2 = \lim_{\Delta_t \to 0} \Delta_t \int_0^t X_s ds = 0 \tag{70}$$

- Regarding $dB_t d_t = 0$ note

$$\lim_{\Delta_t \to 0} \sum_{k=0}^{n-1} X_{t_k} \Delta t \Delta B_k = \lim_{\Delta_t \to 0} \Delta_t \int_0^t X_s dB_s = 0 \tag{71}$$

- Regarding $dB_t^2 = dt$ note

$$\mathbb{E}\Big[\Big(\sum_{k=0}^{n-1} X_{t_k} \Delta B_k^2 - \sum_{k=0}^{n-1} X_{t_k} \Delta t\Big)^2\Big] = \mathbb{E}\Big[\Big(\sum_{k=0}^{n-1} X_{t_k} (\Delta B_k^2 - \Delta t)\Big)^2\Big]$$

$$= \sum_{k=0}^{n-1} \sum_{k'=0}^{n-1} \mathbb{E}[X_{t_k} X_{t'_k} (\Delta B_k^2 - \Delta t)(\Delta B_{k'}^2 - \Delta t)] \tag{72}$$

If $k > k'$ then $(\Delta B_k^2 - \Delta t)$ is independent of $X_{t_k} X_{t_{k'}} (\Delta B_{k'}^2 - \Delta t)$, and therefore

$$\mathbb{E}[X_{t_k} X_{t_{k'}} (\Delta B_k^2 - \Delta t)(\Delta B_{k'}^2 - \Delta t)]$$
$$= \mathbb{E}[X_{t_k} X_{t_{k'}} (\Delta B_{k'}^2 - \Delta t)] \mathbb{E}[\Delta B_k^2 - \Delta t)] = 0 \tag{73}$$

Equivalently, if $k' > k$ then $(\Delta B_{k'}^2 - \Delta t)$ is independent of $X_{t_k} X_{t_{k'}} (\Delta B_k^2 - \Delta t)$ and therefore

$$\mathbb{E}[X_{t_k} X_{t_{k'}} (\Delta B_k^2 - \Delta t)(\Delta B_{k'}^2 - \Delta t)]$$
$$= \mathbb{E}[X_{t_k} X_{t_{k'}} (\Delta B_k^2 - \Delta t)] \mathbb{E}[\Delta B_{k'}^2 - \Delta t)] = 0 \tag{74}$$

Thus

$$\sum_{k=0}^{n-1}\sum_{k'=0}^{n-1}\mathbb{E}[X_{t_k}X_{t_{k'}}(\Delta B_k^2-\Delta t)(\Delta B_{k'}^2-\Delta t)]=\sum_{k=0}^{n-1}\mathbb{E}[X_{t_k}^2(\Delta B_k^2-\Delta t)^2] \tag{75}$$

Note since ΔB_k is independent of X_{t_k} then

$$\mathbb{E}[X_{t_k}^2(\Delta B_k^2-\Delta t)^2=\mathbb{E}[X_{t_k}^2]\mathbb{E}[(\Delta B_k^2-\Delta t)^2] \tag{76}$$

$$=\mathbb{E}[X_{t_k}^2]\mathrm{Var}(\Delta B_k^2)=2\mathbb{E}[X_{t_k}^2]\Delta t^2 \tag{77}$$

Thus

$$\mathbb{E}\Big[\Big(\sum_{k=0}^{n-1}X_{t_k}\Delta B_k^2-\sum_{k=0}^{n-1}X_{t_k}\Delta t\Big)^2\Big]=\sum_{k=0}^{n-1}\mathbb{E}[X_{t_k}^2]\Delta_t^2 \tag{78}$$

which goes to zero as $\Delta t\to 0$. Thus, in the limit as $\Delta t\to 0$

$$\lim_{\Delta t\to 0}\sum_{k=0}^{n-1}X_{t_k}\Delta B_k^2=\lim_{\Delta t\to 0}\sum_{k=0}^{n-1}X_{t_k}\Delta t \tag{79}$$

where the limit is taken in the mean square sense. Thus

$$\int_0^t X_s dB_s^2=\int_0^t X_s ds \tag{80}$$

- Regarding $dB_t dW_t=0$ note

$$\mathbb{E}\Big[\Big(\sum_{k=0}^{n-1}X_{t_k}\Delta B_k\Delta W_k\Big)^2\Big]\quad=\sum_{k=0}^{n-1}\sum_{k'=0}^{n-1}\mathbb{E}[X_{t_k}X_{t_{k'}}\Delta B_k\Delta W_k\Delta B_{k'}\Delta W_{k'}] \tag{81}$$

If $k>k'$ then $\Delta B_k,\Delta W_k$ are independent of $X_{t_k}X_{t_{k'}}\Delta B_{k'}\Delta W_{k'}$ and therefore

$$\mathbb{E}[X_{t_k}X_{t_{k'}}\Delta B_k\Delta W_k\Delta B_{k'}\Delta W_{k'}]=\mathbb{E}[X_{t_k}X_{t_{k'}}\Delta B_{k'}\Delta W_{k'}]\mathbb{E}[\Delta B_k]\mathbb{E}[\Delta W_k]=0 \tag{82}$$

Equivalently, if $k'>k$ then $\Delta B_{k'},\Delta W_{k'}$ are independent of $X_{t_k}X_{t_{k'}}\Delta B_k\Delta W_k$ and therefore

$$\mathbb{E}[X_{t_k}X_{t_{k'}}\Delta B_k\Delta W_k\Delta B_{k'}\Delta W_{k'}]=\mathbb{E}[X_{t_k}X_{t_{k'}}\Delta B_k\Delta W_k]\mathbb{E}[\Delta B_{k'}]\mathbb{E}[\Delta W_{k'}]=0 \tag{83}$$

Finally, for $k=k'$, $\Delta B_k,\Delta W_k$ and X_{t_k} are independent, thus

$$\mathbb{E}[X_{t_k}^2\Delta B_k^2\Delta W_k^2]=\mathbb{E}[X_{t_k}^2]\mathbb{E}[\Delta B_k^2]\mathbb{E}[\Delta W_k^2]=\mathbb{E}[X_{t_k}^2]\Delta t^2 \tag{84}$$

Thus

$$\mathbb{E}\Big[\Big(\sum_{k=0}^{n-1}X_{t_k}\Delta B_k\Delta W_k\Big)^2\Big]=\sum_{k=0}^{n-1}\mathbb{E}[X_{t_k}^2]\Delta t^2 \tag{85}$$

which converges to 0 as $\Delta t\to 0$. Thus

$$\int_0^t X_s dB_s dW_s=0 \tag{86}$$

4.2 Vector Stochastic Differential Equations

The form

$$dX_t = a(X_t)dt + c(X_t)dB_t \tag{87}$$

is also used to represent multivariate equations. In this case X_t represents an n-dimensional random vector, B_t an m-dimensional vector of m independent standard Brownian motions, and $c(X_t$ is an $n \times m$ matrix. a is commonly known as the *drift* vector and b the *dispersion matrix*.

5 Ito's Rule

The main thing with Ito's calculus is that for the general case a differential carries quadratic and linear components. For example suppose that X_t is an Ito process. Let

$$Y_t = f(t, X_t) \tag{88}$$

then

$$\boxed{dY_t = \nabla f(t, X_t)^T dX_t + \frac{1}{2} dX_t^T \nabla^2 f(t, X_t) dX_t} \tag{89}$$

where ∇, ∇^2 are the gradient and Hessian with respect to (t, x). Note basically this is the second order Taylor series expansion. In ordinary calculus the second order terms are zero, but in Stochastic calculus, due to the fact that these processes have non-zero quadratic variation, the quadratic terms do not go away. This is really all you need to remember about Stochastic calculus, everything else derives from this basic fact.

The most important consequence of this fact is Ito's rule. Let X_t be governed by an SDE

$$dX_t = a(X_t, t)dt + c(X_t, t)dB_t \tag{90}$$

Let $Y_t = f(X_t, t)$. Ito's rule tells us that Y_t is governed by the following SDE

$$\boxed{dY_t \stackrel{\text{def}}{=} \nabla_t f(t, X_t)dt + \nabla_x f(t, x)^T dX_t + \frac{1}{2} dX_t^T \nabla_x^2 f(t, X_t) dX_t} \tag{91}$$

where

$$dB_{i,t} dB_{j,t} \stackrel{\text{def}}{=} \delta(i, j)\, dt \tag{92}$$

$$dX dt \stackrel{\text{def}}{=} 0 \tag{93}$$

$$dt^2 \stackrel{\text{def}}{=} 0 \tag{94}$$

Equivalently

$$\boxed{\begin{aligned} dY_t = {} & \nabla_t f(X_t, t) d_t + \nabla_x f(X_t, t)^T a(X_t, t) d_t + \nabla_x f(X_t, t)^T c(X_t, t) dB_t \\ & + \frac{1}{2} \text{trace}\Big(c(X_t, t) c(X_t, t)^T \nabla_x^2 f(X_t, t) \Big) dt \end{aligned}} \tag{95}$$

where

$$\nabla_x f(x,t)^T a(x,t) = \sum_i \frac{\partial f(x,t)}{\partial x_i} a_i(x,t) \tag{96}$$

$$\text{trace}\Big(c(x,t)c(x,t)^T \nabla_x^2 f(x,t)\Big) = \sum_i \sum_j (c(x,t)c(x,t)^T)_{ij} \frac{\partial^2 f(x,t)}{\partial x_i \partial x_j} \tag{97}$$

Note b is a matrix. *Sketch of Proof:* To second order

$$\begin{aligned}\Delta Y_t = f(X_{t+\Delta_t}, t+\Delta t) - f(X_t,t) = \nabla_t f(X_t,t)\Delta t + \nabla_x f(X_t,t)^T \Delta X_t \\ + \frac{1}{2}\Delta t^2 \nabla_t^2 f(X_t,t) + \frac{1}{2}\Delta X_t^T \nabla_x^2 f(X_t,t)\Delta X_t + \Delta t(\nabla_x \nabla_t f(X_t,t))^T \Delta X_t \Delta t\end{aligned} \tag{98}$$

where ∇_t, ∇_x are the gradients with respect to time and state, and ∇_t^2 is the second derivative with respect to time, ∇_x^2 the Hessian with respect to time and $\nabla_x \nabla_t$ the gradient with respect to state of the gradient with respect to time. Integrating over time

$$Y_t = Y_0 + \sum_{k=0'}^{n-1} \Delta Y_{t_k} \tag{99}$$

and taking limits

$$\begin{aligned}Y_t =& Y_0 + \int_0^t dY_s = Y_0 + \int_0^t \nabla_t f(X_s,s) d_s + \int_0^t \nabla_x f(X_s,s)^T dX_s \\ &+ \frac{1}{2}\int_0^t \nabla_t^2 f(X_s,s)(ds)^2 + \frac{1}{2}\int_0^t dX_s^T \nabla_x^2 f(X_s,s) dX_s \\ &+ \int_0^t (\nabla_x \nabla_t f(X_s,s))^T dX_s ds\end{aligned} \tag{100}$$

In differential form

$$\begin{aligned}dY_t =& \nabla_t f(X_t,t) d_t + \nabla_x f(X_t,t)^T dX_t \\ &+ \frac{1}{2}\nabla_t^2 f(X_t,t)(dt)^2 + \frac{1}{2} dX_t^T \nabla_x^2 f(X_t,t) dX_t \\ &+ (\nabla_x \nabla_t f(X_t,t))^T dX_t dt\end{aligned} \tag{101}$$

Expanding dX_t

$$\begin{aligned}(\nabla_x \nabla_t f(X_t,t))^T dX_t dt = (\nabla_x \nabla_t f(X_t,t))^T a(X_t,t)(dt)^2 \\ + (\nabla_x \nabla_t f(X_t,t))^T c(X_t,t) dB_t dt = 0\end{aligned} \tag{102}$$

where we used the standard rules for second order differentials

$$(dt)^2 = 0 \tag{103}$$

$$(dB_t)dt = 0 \tag{104}$$

$$\tag{105}$$

Moreover

$$\begin{aligned}&dX_t^T \nabla_x^2 f(X_t,t) dX_t \\ &= (a(X_t,t)dt + c(X_t,t)dB_t)^T \nabla_x^2 f(X_t,t)(a(X_t,t)dt + c(X_t,t)dB_t) \\ &= a(X_t,t)^T \nabla_x^2 f(X_t,t) a(X_t,t)(dt)^2 \\ &+ 2a(X_t,t)^T \nabla_x^2 f(X_t,t) c(X_t,t)(dB_t)dt \\ &+ dB_t^T c(X_t,t)^T \nabla_x^2 f(X_t,t) c(X_t,t)(dB_t)\end{aligned} \tag{106}$$

Using the rules for second order differentials

$$(dt)^2 = 0 \tag{107}$$
$$(dB_t)dt = 0 \tag{108}$$
$$dB_t^T K(X_t,t)dB_t = \sum_i \sum_j K_{i,j}(X_t,t)dB_{i,t}dB_{j,t} = \sum_i K_{i,i}dt \tag{109}$$

where

$$K(X_t,t) = c(X_t,t)^T \nabla_x^2 f(X_t,t)c(X_t,t) \tag{110}$$

Thus

$$\begin{aligned} dY_t =& \nabla_t f(X_t,t)d_t + \nabla_x f(X_t,t)^T a(X_t,t)dt + \nabla_x f(X_t,t)^T c(X_t,t)dB_t \\ &+ \frac{1}{2}\text{trace}\Big(c(X_t,t)c(X_t,t)^T \nabla_x^2 f(X_t,t)\Big)dt \end{aligned} \tag{111}$$

where we used the fact that

$$\begin{aligned} \sum_i K_{ii}(X_t,t)dt &= \text{trace}(K)dt \\ &= \text{trace}\Big(c(X_t,t)^T \nabla_x^2 f(X_t,t)c(X_t,t)\Big) \\ &= \text{trace}\Big(c(X_t,t)c(X_t,t)^T \nabla_x^2 f(X_t,t)\Big) \end{aligned} \tag{112}$$

□

5.1 Product Rule

Let X, Y be Ito processes then

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + dX_t dY_t \tag{113}$$

Proof: Consider X, Y as a joint Ito process and take $f(x,y,t) = xy$. Then

$$\frac{\partial f}{\partial t} = 0 \tag{114}$$
$$\frac{\partial f}{\partial x} = y \tag{115}$$
$$\frac{\partial f}{\partial y} = x \tag{116}$$
$$\frac{\partial^2 f}{\partial x \partial y} = 1 \tag{117}$$
$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial y^2} = 0 \tag{118}$$

Applying Ito's rule, the Product Rule follows. □

Exercise: Solve $\int_0^T B_t dB_t$ symbolically.
Let $a(X_t,t) = 0, c(X_t,t) = 1, f(x,t) = x^2$. Thus

$$dX_t = dB_t \tag{119}$$
$$X_t = B_t \tag{120}$$

and

$$\frac{\partial f(t,x)}{\partial t} = 0 \tag{121}$$

$$\frac{\partial f(t,x)}{\partial x} = 2x \tag{122}$$

$$\frac{\partial^2 f(t,x)}{\partial x^2} = 2 \tag{123}$$

Applying Ito's rule

$$\begin{aligned} df(X_t,t) =& \frac{\partial f(X_t,t)}{\partial t} dt + \frac{\partial f(X_t,t)}{\partial x} a(X_t,t)dt + \frac{\partial f(X_t,t)}{\partial x} c(X_t,t)dB_t \\ &+ \frac{1}{2}\text{trace}\Big(c(X_t,t)c(X_t,t)^T \frac{\partial^2 f(x,t)}{\partial x^2}\Big) \end{aligned} \tag{124}$$

we get

$$dB_t^2 = 2B_t dB_t + dt \tag{125}$$

Equivalently

$$\int_0^t dB_s^2 = 2\int_0^t B_s dB_s + \int_0^t ds \tag{126}$$

$$B_t^2 = 2\int_0^t B_s dB_s + t \tag{127}$$

Therefore

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t \tag{128}$$

NOTE: dB_t^2 is different from $(dB_t)^2$.

Exercise: Get $\mathbb{E}[e^\beta B_t]$

Let Let $a(X_t,t) = 0, c(X_t,t) = 1$, i.e., $dX_t = dB_t$. Let $Y_t = f(X_t,t) = e^{\beta B_t}$, and $dX_t = dB_t$. Using Ito's rule

$$dY_t = \beta e^{\beta B_t} dB_t + \frac{1}{2}\beta^2 e^{\beta B_t} dt \tag{129}$$

$$Y_t = Y_0 + \beta \int_0^t e^{\beta B_s} dB_s + \frac{\beta^2}{2}\int_0^t e^{\beta B_s} ds \tag{130}$$

Taking expected values

$$\mathbb{E}[Y_t] = \mathbb{E}[Y_0] + \frac{\beta^2}{2}\int_0^t \mathbb{E}[Y_s] ds \tag{131}$$

where we used the fact that $\mathbb{E}[\int_0^t e^{\beta B_s} dB_s] = 0$ because for any non anticipatory random variable Y_t, we know that $\mathbb{E}[\int_0^t Y_s dB_s] = 0$. Thus

$$\frac{d\mathbb{E}[Y_t]}{dt} = \frac{\beta^2}{2}\mathbb{E}[Y_t] \tag{132}$$

and since $\mathbb{E}[Y_0] = 1$

$$\mathbb{E}[e^{\beta B_t}] = e^{\frac{\beta^2}{2}t} \tag{133}$$

Exercise: Solve the following SDE

$$dX_t = \alpha X_t dt + \beta X_t dB_t \tag{134}$$

In this case $a(X_t, t) = \alpha X_t$, $c(X_t, t) = \beta X_t$. Using Ito's formula for $f(x,t) = \log(x)$

$$\frac{\partial f(t,x)}{\partial t} = 0 \tag{135}$$

$$\frac{\partial f(t,x)}{\partial x} = \frac{1}{x} \tag{136}$$

$$\frac{\partial^2 f(t,x)}{\partial x^2} = -\frac{1}{x^2} \tag{137}$$

Thus

$$d\log(X_t) = \frac{1}{X_t}\alpha X_t dt + \frac{1}{X_t}\beta X_t dB_t - \frac{1}{2X_t^2}\beta^2 X_t^2 dt = (\alpha - \frac{\beta^2}{2})dt + \beta dB_t \tag{138}$$

Integrating over time

$$\log(X_t) = \log(X_0) + (\alpha - \frac{\beta^2}{2})t + \beta B_t \tag{139}$$

$$X_t = X_0 \exp((\alpha - \frac{\beta^2}{2})t)\exp(\beta B_t) \tag{140}$$

Note

$$\mathbb{E}[X_t] = \mathbb{E}[X_0]e^{(\alpha - \frac{\beta^2}{2})t}\mathbb{E}[\exp(\alpha B_t)] = \mathbb{E}[X_0]e^{\alpha t} \tag{141}$$

6 Moment Equations

Consider an SDE of the form

$$dX_t = a(X_t)dt + c(X_t)dB_t \tag{142}$$

Taking expected values we get the differential equation for first order moments

$$\frac{d\mathbb{E}[X_t]}{dt} = \mathbb{E}[a(X_t)] \tag{143}$$

seems weird that c has no effect. Double check with generator of ito diffusion result

With respect to second order moments, let

$$Y_t = f(X_t) = X_{i,t}X_{j,t} \tag{144}$$

using Ito's product rule

$$\begin{aligned} dY_t = d(X_{i,t}X_{j,t}) =& X_{i,t}dX_{j,t} + X_{j,t}dX_{i,t} + dX_{i,t}dX_{j,t} \\ =& X_{i,t}(a_j(X_t)dt + (c(X_t)dB_t)_j) + X_{j,t}(a_i(X_t)dt + (c(X_t)dB_t)_i) \\ &+ (a_i(X_t)dt + (c(X_t)dB_t)_i)(a_j(X_t)dt + (c(X_t)dB_t)_j) \\ &= X_{i,t}(a_j(X_t)dt + (c(X_t)dB_t)_j) + X_{j,t}(a_i(X_t)dt + (c(X_t)dB_t)_i) \\ &+ c_i(X_t)c_j(X_t)dt \end{aligned} \tag{145}$$

Taking expected values

$$\frac{d\mathbb{E}[X_{i,t}X_{j,t}]}{dt} = \mathbb{E}[X_{i,t}a_j(X_t)] + \mathbb{E}[X_{j,t}a_i(X_t)] + \mathbb{E}[c_i(X_t)c_j(X_t)] \tag{146}$$

In matrix form

$$\frac{d\mathbb{E}[X_t X_t']}{dt} = \mathbb{E}[X_t a(X_t)'] + \mathbb{E}[a(X_t)X_t'] + \mathbb{E}[c(X_t)c(X_t)'] \tag{147}$$

The moment formulas are particularly useful when a, c are constant with respect to X_t, in such case

$$\frac{d\mathbb{E}[X_t]}{dt} = a\mathbb{E}[X_t] \tag{148}$$

$$\frac{d\mathbb{E}[X_t X_t']}{dt} = \mathbb{E}[X_t X_t']a' + a\mathbb{E}[X_t X_t'] + cc' \tag{149}$$

$$\frac{\mathrm{Var}[X_t]}{dt} = \mathbb{E}[X_t X_t']a' + a\mathbb{E}[X_t X_t'] - a\mathbb{E}[X_t]\mathbb{E}[X_t]'a' + cc' \tag{150}$$

$$\tag{151}$$

Example Calculate the equilibrium mean and variance of the following process

$$dX_t = -X_t + cdB_t \tag{152}$$

The first and second moment equations are

$$\frac{d\mathbb{E}[X_t]}{dt} = -\mathbb{E}[X_t] \tag{153}$$

$$\frac{d\mathbb{E}[X_t^2]}{dt} = -2\mathbb{E}[X_t]^2 + c^2 \tag{154}$$

Thus

$$\lim_{t\to\infty} \mathbb{E}[X_t] = 0 \tag{155}$$

$$\lim_{t\to\infty} \mathbb{E}[X_t^2] = \lim_{t\to\infty} \mathrm{Var}[X_t] = \frac{c^2}{2} \tag{156}$$

7 Generator of an Ito Diffusion

The generator G_t of the Ito diffusion

$$dX_t = a(X_t, t)dt + c(X_t, t)dB_t \tag{157}$$

is a second order partial differential operator. For any function f it provides the directional derivative of f averaged across the paths generated by the diffusion. In particular given the function f, the function $G_t[f]$ is defined as follows

$$G_t[f](x) = \frac{d\mathbb{E}[f(X_t) \mid X_t = x]}{dt} = \lim_{\Delta t \to 0} \frac{\mathbb{E}[f(X_{t+\Delta_t}) \mid X_t = x] - f(x)}{\Delta t}$$
$$= \frac{\mathbb{E}[df(X_t) \mid X_t = x]}{dt} \tag{158}$$

Note using Ito's rule

$$d\, f(X_t) = \nabla_x f(X_t, t)^T a(X_t, t)dt + \nabla_x c(X_t, t)^T dB_t$$
$$+ \frac{1}{2}\mathrm{trace}\Big(c(X_t, t)c(X_t, t)^T \nabla_x^2 f(X_t, t)\Big)dt \tag{159}$$

Taking expected values

$$G_t[f](x) = \frac{\mathbb{E}[df(X_t) \mid X_t = x]}{dt} = \nabla_x f(x)^T a(x,t) + \frac{1}{2}\text{trace}\Big(c(x,t)c(x,t)^T \nabla_x^2 f(x)\Big) \tag{160}$$

In other words

$$G_t[\cdot] = \sum_i a_i(x,t)\frac{\partial}{\partial x_i}[\cdot] + \frac{1}{2}\sum_i\sum_j (c(x,t)c(x,t)^T)_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}[\cdot] \tag{161}$$

8 Adjoints

Every linear operator G on a Hilbert space H with inner product $< \cdot, \cdot >$ has a corresponding adjoint operator G^* such that

$$< Gx, y >=< x, G^* y > \text{ for all } x, y \in H \tag{162}$$

In our case the elements of the Hilbert space are functions f, g and the inner product will be of the form

$$< f, g >= \int f(x) \cdot g(x) \, dx \tag{163}$$

Using partial integrations it can be shown that if

$$G[f](x) = \sum_i \frac{\partial f(x)}{\partial x_i} a_i(x,t) + \frac{1}{2}\text{trace}\Big(c(x,t)c(x,t)^T \nabla_x^2 f(x)\Big) \tag{164}$$

$$\tag{165}$$

then

$$G^*[f](x) = -\sum_i \frac{\partial}{\partial x_i}[f(x)a_i(x,t)] + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}[(c(x,t)c(x,t)^T)_{ij} f(x)] \tag{166}$$

9 The Feynman-Kac Formula (Terminal Condition Version)

Let X be an Ito diffusion

$$dX_t = a(X_t,t)dt + c(X_t,t)dB_t \tag{167}$$

with generator G_t

$$G_t[v](x) = \sum_i a_i(x,t)\frac{\partial v(x,t)}{\partial x_i} + \frac{1}{2}\sum_i\sum_j (c(x,t)c(x,t)^T)_{i,j}\frac{\partial^2 v(x,t)}{\partial x_i \partial x_j} \tag{168}$$

Let v be the solution to the following pde

$$-\frac{\partial v(x,t)}{\partial t} = G_t[v](x,t) - v(x,t)f(x,t) \tag{169}$$

with a known terminal condition $v(x,T)$, and function f. It can be shown that the solution to the pde (169) is as follows

$$v(x,s) = \mathbb{E}\Big[v(X_T,T) \; \exp\big(-\int_s^T f(X_t)dt\big) \mid X_s = x\Big] \tag{170}$$

We can think of $v(X_T, T)$ as a terminal reward and of $\int_s^T f(X_t)dt$ as a discount factor.

Informal Proof:
Let $s \le t \le T$ let $Y_t = v(X_t, t)$, $Z_t = \exp(-\int_s^t f(X_\tau)d\tau)$, $U_t = Y_t Z_t$. It can be shown (see Lemma below) that

$$dZ_t = -Z_t f(X_t)dt \tag{171}$$

Using Ito's product rule

$$dU_t = d(Y_t Z_t) = Z_t dY_t + Y_t dZ_t + dY_t dZ_t \tag{172}$$

Since dZ_t has a dt term, it follows that $dY_t dZ_t = 0$. Thus

$$dU_t = Z_t dv(X_t, t) - v(X_t, t) Z_t f(X_t)dt \tag{173}$$

Using Ito's rule on dv we get

$$\begin{aligned} dv(X_t, t) =& \nabla_t v(X_t, t)dt + (\nabla_x v(X_t, t))^T a(X_t, t)dt + (\nabla_x v(X_t, t))^T c(X_t, t)dB_t \\ &+ \frac{1}{2}\text{trace}\Big(c(X_t, t)c(X_t, t)^T \nabla_x^2 v(X_t, t)\Big)dt \end{aligned} \tag{174}$$

Thus

$$\begin{aligned} dU_t =& Z_t\Big[\nabla_t v(X_t, t) + (\nabla_x v(X_t, t))^T a(X_t, t) \\ &+ \frac{1}{2}\text{trace}\Big(c(X_t, t)c(X_t, t)^T \nabla_x^2 v(X_t, t)\Big) - v(X_t, t) f(X_t)\Big]dt \\ &+ Z_t(\nabla_x v(X_t, t))^T c(X_t, t)dB_t \end{aligned} \tag{175}$$

and since v is the solution to (169) then

$$dU_t = (\nabla_x v(X_t, t))^T c(X_t, t)dB_t \tag{176}$$

Integrating

$$U_T - U_s = \int_s^T Y_t(\nabla_x v(X_t, t))^T c(X_t, t)dB_t \tag{177}$$

taking expected values

$$\mathbb{E}[U_T \mid X_s = x] - \mathbb{E}[U_s \mid X_s = x] = 0 \tag{178}$$

where we used the fact that the expected values of integrals with respect to Brownian motion is zero. Thus, since $U_s = Y_0 Z_0 = v(X_s, s)$

$$\mathbb{E}[U_T \mid X_s = x] = \mathbb{E}[U_s \mid X_s = x] = v(x, s) \tag{179}$$

Using the definition of U_T we get

$$v(x, s) = \mathbb{E}[v(X_T, T)e^{-\int_s^T f(X_t)dt} \mid X_s = x] \tag{180}$$

We end the proof by showing that

$$dZ_t = -Z_t f(X_t)dt \tag{181}$$

First let $Y_t = \int_s^t f(X_\tau)d\tau$ and note

$$\Delta Y_t = \int_t^{t+\Delta_t} f(X_\tau)d\tau \approx f(X_t)\Delta t \tag{182}$$

$$dY_t = f(X_t)dt \tag{183}$$

Let $Z_t = \exp(-Y_t)$. Using Ito's rule

$$dZ_t = \nabla e^{-Y_t} dY_t + \frac{1}{2}\nabla^2 e^{-Y_t}(dY_t)^2 = -e^{-Y_t} f(X_t)dt = -Z_t f(X_t)dt \tag{184}$$

where we used the fact that

$$(dY_t)^2 = Z_t^2 f(X_t)^2 (dt)^2 = 0 \tag{185}$$

□

10 Kolmogorov Backward equation

The Kolmogorov backward equation tells us at time s whether at a future time t the system will be in the target set A. We let ξ be the indicator function of A, i.e, $\xi(x) = 1$ if $x \in A$, otherwise it is zero. We want to know for every state x at time $s < T$ what is the probability of ending up in the target set A at time T. This is call the the hit probability.

Let X be an Ito diffusion

$$dX_t = a(X_t, t)dt + c(X_t, t)dB_t \tag{186}$$

$$X_0 = x \tag{187}$$

The hit probability $p(x,t)$ satisfies the Kolmogorov backward pde

$$-\frac{\partial p(x,t)}{\partial t} = G_t[p](x,t) \tag{188}$$

i.e.,

$$\boxed{-\frac{\partial p(x,t)}{\partial t} = \sum_i a_i(x,t)\frac{\partial p(x,t)}{\partial x_i} + \frac{1}{2}\sum_{i,j}(c(x,t)c(x,t)^T)_{ij}\frac{\partial^2 p(x,t)}{\partial x_i \partial x_j}} \tag{189}$$

subject to the final condition $p(x,T) = \xi(x)$. The equation can be derived from the Feynman-Kac formula, noting that the hit probability is an expected value over paths that originate at x at time $s \le T$, and setting $f(x) = 0$, $q(x) = \xi(x)$ for all x

$$p(x,t) = p(X_T \in A \mid X_t = x) = \mathbb{E}[\xi(X_T) \mid X_t = x] = \mathbb{E}[q(X_T)e^{\int_t^T f(X_s)ds}] \tag{190}$$

11 The Kolmogorov Forward equation

Let X be an Ito diffusion

$$dX_t = a(X_t, t)dt + c(X_t, t)dB_t \tag{191}$$

$$X_0 = x_0 \tag{192}$$

with generator G. Let $p(x,t)$ represent the probability density of X_t evaluated at x given the initial state x_0. Then

$$\frac{\partial p(x,t)}{\partial t} = G^*[p](x,t) \tag{193}$$

where G^* is the adjoint of G, i.e.,

$$\boxed{\frac{\partial p(x,t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}[p(x,t)a_i(x,t)] + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}[(c(x,t)c(x,t)^T)_{ij}p(x,t)]} \tag{194}$$

It is sometimes useful to express the equation in terms of the negative divergence (inflow) of a probability current J, caused by a probability velocity V

$$\frac{\partial p(x,t)}{\partial t} = -\nabla \cdot J(x,t) = -\sum_i \frac{\partial J_i(x,t)}{\partial x_i} \tag{195}$$

$$J(x,t) = p(x,t)V(x,t) \tag{196}$$

$$V_i(x,t) = a_i(x,t) - \frac{1}{2}\sum_i k(x,t)_{i,j}\frac{\partial}{\partial x_j}\log(p(x,t)k_{i,j}(x)) \tag{197}$$

$$k(x) = c(x,t)c(x,t)^T \tag{198}$$

From this point of view the Kolmogorov forward equation is just a law of conservation of probability (the rate of accumulation of probability in a state x equals the inflow of probability due to the probability field V).

11.1 Example: Discretizing an SDE in state/time

Consider the following SDE

$$dX_t = a(X_t)X_t dt + c(X_t)dB_t \tag{199}$$

The Kolmogorov Forward equation looks as follows

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial a(x)p(x,t)}{\partial x} + \frac{1}{2}\frac{\partial^2 c(x)^2 p(x,t)}{\partial x^2} \tag{200}$$

Discretizing in time and space, to first order

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{\Delta_t}(p(x,t+\Delta_t) - p(x-,t)) \tag{201}$$

and

$$\begin{aligned}
&\frac{\partial a(x)p(x,t)}{\partial x} = \frac{1}{2\Delta_x}(a(x+\Delta_x)p(x+\Delta_x,t) - a(x-\Delta_x)p(x-\Delta_x,t)) \\
&= \frac{1}{2\Delta_x}\Big((a(x) + \Delta_x\frac{\partial a(x)}{\partial x})p(x+\Delta_x,t) - (a(x) - \Delta_x\frac{\partial a(x)}{\partial x})p(x-\Delta_x,t)\Big) \\
&= p(x+\Delta_x,t)(\frac{a(x)}{2\Delta_x} + \frac{\partial a(x)}{2\partial x}) - p(x-\Delta_x,t)(\frac{a(x)}{2\Delta_x} - \frac{\partial a(x)}{2\partial x})
\end{aligned} \tag{202}$$

and

$$\frac{\partial^2 c^2(x)p(x,t)}{\partial x^2} = \frac{1}{\Delta_x^2}\Big(c^2(x+\Delta_x)p(x+\Delta_x,t) + c^2(x-\Delta_x)p(x-\Delta_x,t) - 2c^2(x)p(x,t)\Big) \tag{203}$$

$$\begin{aligned}
&= \frac{1}{\Delta_x^2}\Big((c^2(x) + 2\Delta_x c(x)\frac{\partial c(x)}{\partial x})p(x+\Delta_x,t) + (c^2(x) - 2\Delta_x c(x)\frac{\partial c(x)}{\partial x})p(x-\Delta_x,t) \\
&\quad - 2c^2(x)p(x,t)\Big) \\
&= p(x+\Delta_x,t)(\Big(\frac{c(x)}{\Delta_x}\Big)^2 + 2\frac{c(x)}{\Delta_x}\frac{\partial c(x)}{\partial x}) \\
&- 2p(x,t)\Big(\frac{c(x)}{\Delta_x}\Big)^2 \\
&+ p(x-\Delta_x,t)(\Big(\frac{c(x)}{\Delta_x}\Big)^2 - 2\frac{c(x)}{\Delta_x}\frac{\partial c(x)}{\partial x})
\end{aligned} \tag{204}$$

Putting it together, the Kolmogorov Forward Equation can be approximated as follows

$$\begin{aligned}
\frac{p(x,t+\Delta_t) - p(x,t)}{\Delta_t} =& p(x-\Delta_x,t)(\frac{a(x)}{2\Delta_x} - \frac{\partial a(x)}{2\partial x}) \\
&- p(x+\Delta_x,t)(\frac{a(x)}{2\Delta_x} + \frac{\partial a(x)}{2\partial x}) \\
&+ p(x+\Delta_x,t)(\frac{1}{2}\Big(\frac{c(x)}{\Delta_x}\Big)^2 + \frac{c(x)}{\Delta_x}\frac{\partial c(x)}{\partial x}) \\
&- p(x,t)\Big(\frac{c(x)}{\Delta_x}\Big)^2 \\
&+ p(x-\Delta_x,t)(\frac{1}{2}\Big(\frac{c(x)}{\Delta_x}\Big)^2 - \frac{c(x)}{\Delta_x}\frac{\partial c(x)}{\partial x})
\end{aligned} \tag{205}$$

Rearranging terms

$$
\begin{aligned}
p(x, t+\Delta_t) =& p(x,t)\Big(1 - \frac{\Delta_t c^2(x)}{\Delta_x^2}\Big) \\
&+ p(x-\Delta_x, t)\frac{\Delta_t}{2\Delta_x}\Big(\frac{c^2(x)}{\Delta_x} - 2c(x)\frac{\partial c(x)}{\partial x} + a(x) - \Delta_x \frac{\partial a(x)}{\partial x}\Big) \\
&+ p(x+\Delta_x, t)\frac{\Delta_t}{2\Delta_x}\Big(\frac{c^2(x)}{\Delta_x} + 2c(x)\frac{\partial c(x)}{\partial x} - a(x) - \Delta_x \frac{\partial a(x)}{\partial x}\Big) \qquad (206)
\end{aligned}
$$

Considering in a discrete time, discrete state system

$$
p(X_{t+\Delta_t} = x) = \sum_{x'} p(X_t = x') p(X_{t+\Delta_t} = x \mid X_t = x') \qquad (207)
$$

we make the following discrete time/discrete state approximation

$$
p(x_{t+\Delta_t} \mid x_t) = \begin{cases} \frac{\Delta_t}{2\Delta_x}\Big(\frac{c^2(x)}{\Delta_x} - a(x) + 2c(x)\frac{\partial c(x)}{\partial x} - \Delta_x \frac{\partial a(x)}{\partial x}\Big) & \text{if } x_{t+\Delta_t} = x_t - \Delta_x \\ \frac{\Delta_t}{2\Delta_x}\Big(\frac{c^2(x)}{\Delta_x} + a(x) - 2c(x)\frac{\partial c(x)}{\partial x} - \Delta_x \frac{\partial a(x)}{\partial x}\Big) & \text{if } x_{t+\Delta_t} = x_t + \Delta_x \\ 1 - \frac{\Delta_t c^2(x)}{\Delta_x^2} & \text{if } x_{t+\Delta_t} = x_t \\ 0 & \text{else} \end{cases} \qquad (208)
$$

Note if the derivative of the drift function is zero, i.e., $\partial a(x)/\partial x = 0$ the conditional probabilities add up to one. Not sure how to deal with the case in which the derivative is not zero.

11.2 Girsanov's Theorem (Version I)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let B be a standard m-dimensional Brownian motion adapted to the filtration $\mathcal{F}_t$. Let X, Y be defined by the following SDEs

$$
dX_t = a(X_t)dt + c(X_t)dB_t \qquad (209)
$$

$$
dY_t = (c(Y_t)U_t + a(Y_t))dt + c(Y_t)dB_t \qquad (210)
$$

$$
X_0 = Y_0 = x \qquad (211)
$$

where $X_t \in \mathbb{R}^n, B_t \in \mathbb{R}^m$ and a, c satisfy the necessary conditions for the SDEs to be well defined and U_t is an $\mathcal{F}_t$ adapted process such that $\mathbb{P}(\int_0^t \|c(X_s)U_s\|^2 ds < \infty) = 1$. Let

$$
Z_t = \int_0^t U_s' dB_s + \frac{1}{2}\int_0^t U_s' U_s ds \qquad (212)
$$

$$
\Lambda_t = e^{-Z_t} \qquad (213)
$$

and

$$
d\mathbb{Q}_t = \Lambda_t d\mathbb{P} \qquad (214)
$$

i.e., for all $A \in \mathcal{F}_t$

$$
\mathbb{Q}_t(A) = E^{\mathbb{P}}[\Lambda_t I_A] \qquad (215)
$$

Then

$$
W_t = \int_0^t U_s ds + B_t \qquad (216)
$$

is a standard Brownian motion with respect to $\mathbb{Q}_t$.

Informal Proof: We'll provide a heuristic argument for the discrete time case. In discrete time the equation for $W = (W_1, \cdots, W_n)$ would look as follows

$$\Delta W_k = U_k \Delta t + \sqrt{\Delta t} G_k \tag{217}$$

where $G_1, G_2, \cdots$ are independent standard Gaussian vectors under $\mathbb{P}$. Thus, under $\mathbb{P}$ the log-likelihodd of W is as follows

$$\log p(W) = h(n, \Delta t) - \frac{1}{2\Delta t} \sum_{k=1}^{n-1} (\Delta W_k - U_k \Delta t)'(\Delta W_k - U_k \Delta t) \tag{218}$$

where $h(n, \Delta t)$ is constant with respect to W, U. For W to behave as Brownian motion under $\mathbb{Q}$ we need the probability density of W under $\mathbb{Q}$ to be as follows

$$\log q(W) = h(n, \Delta t) - \frac{1}{2\Delta t} \sum_{k=1}^{n-1} \Delta W_k' \Delta W_k \tag{219}$$

Let the random variable Z be defined as follows

$$Z = \log \frac{p(W)}{q(W)} \tag{220}$$

where q, p represent the probability densities of W under $\mathbb{Q}$ and under $\mathbb{P}$ respectively. Thus

$$Z = \sum_{k=1}^{n-1} U_k' (\Delta W_k - U_k \Delta t) + \frac{1}{2} \sum_{k=1}^{n-1} U_k' U_k \Delta t \tag{221}$$

$$= \sum_{k=1}^{n-1} U_k' \sqrt{\Delta t} G_t + \frac{1}{2} \sum_{k=1}^{n-1} U_k' U_k \Delta t \tag{222}$$

$$\tag{223}$$

Note as $\Delta t \to 0$

$$Z \to \int_0^t U_s dB_s + \frac{1}{2} \int_0^t U_s' U_s ds \tag{224}$$

$$\frac{q(W)}{p(W)} \to e^{-Z} \tag{225}$$

□

Remark 11.1.

$$dY_t = (H_t + a(X_t))dt + c(Y_t)dB_t \tag{226}$$

$$= a(X_t)dt + c(Y_t)(U_t + dB_t) \tag{227}$$

$$= a(X_t)dt + c(Y_t)dW_t \tag{228}$$

Therefore the distribution of Y under $\mathbb{Q}_t$ is the same as the distribution of X under $\mathbb{P}$, i.e., for all $A \in \mathcal{F}_t$.

$$\mathbb{P}(X \in A) = \mathbb{Q}_t(Y \in A) \tag{229}$$

or more generally

$$E^{\mathbb{P}}[f(X_{0:t})] = E^{\mathbb{Q}_t}[f(Y_{0:t})] = E^{\mathbb{P}}[f(Y_{0:t})\Lambda_t] \tag{230}$$

Remark 11.2. Radon Nykodim derivative Λ_t is the Radon Nykodim derivative of $\mathbb{Q}_t$ with respect to $\mathbb{P}$. This is typically represented as follows

$$\Lambda_t = e^{-Z_t} = \frac{d\mathbb{Q}_t}{d\mathbb{P}} \tag{231}$$

We can get the derivative of $\mathbb{P}$ with respect to $\mathbb{Q}_t$ by inverting Λ_t, i.e.,

$$\frac{d\mathbb{P}}{d\mathbb{Q}_t} = e^{Z_t} \tag{232}$$

Remark 11.3. Likelihood Ratio This tells us that Λ_t is the likelihood ratio between the process $X_{0:t}$ and the process $Y_{0:t}$, i.e., the equivalent of $p_X(X)/p_Y(Y)$ where p_X, p_Y are the probability densities of X and Y.

Remark 11.4. Importance Sampling Λ_t can be used in importance sampling schemes. Suppose $(y^{[1]}, \lambda^{[1]}), \cdots, (y^{[n]}, \lambda^{[n]})$ are iid samples from $(Y_{0:t}, \Lambda_t)$, then we can estimate $E^{\mathbb{P}}[f(X_{0:t})]$ as follows

$$E^{\mathbb{P}}[f(X_{0:t})] \approx \frac{1}{n}\sum_{i=1}^{n} f(y^{[i]})\lambda^{[i]} \tag{233}$$

11.2.1 Girsanov Version II

Let

$$dX_t = c(X_t)dB_t \tag{234}$$
$$dY_t = b(Y_t)dt + c(Y_t)dB_t \tag{235}$$
$$X_0 = Y_0 = x \tag{236}$$

Let

$$\begin{aligned} Z_t = &\int_0^t b(Y_s)'(c(Y_s)^{-1})'dB_s \\ &+ \frac{1}{2}\int_0^t b(Y_s)'k(Y_s)b(Y_s)ds \end{aligned} \tag{237}$$
$$\frac{d\mathbb{P}}{d\mathbb{Q}_t} = e^{Z_t} \tag{238}$$

where

$$k(Y_s) = (c(Y_s)'c(Y_s))^{-1} \tag{239}$$

Then under $\mathbb{Q}_t$ the process Y has the same distribution as the process X under $\mathbb{P}$.

Proof. We apply Girsanov's version I with $a(X_t) = 0$, $U_t = c(Y_t)^{-1}b(X_t)$. Thus

$$\begin{aligned} Z_t &= \int_0^t U_s' dB_s + \frac{1}{2}\int_0^t U_s' U_s ds \\ &= \int_0^t b(Y_s)'(c(Y_s)^{-1})'dB_s \\ &+ \frac{1}{2}\int_0^t b(Y_s)'k(Y_s)b(Y_s)ds \end{aligned} \tag{240}$$
$$\frac{d\mathbb{P}}{d\mathbb{Q}_t} = e^{Z_t} \tag{241}$$

And under $\mathbb{Q}_t$ the process Y looks like the process X, i.e., a process with zero drift. □

11.2.2 Girsanov Version III

Let

$$dX_t = c(X_t)dB_t \tag{242}$$
$$dY_t = b(Y_t)dt + c(Y_t)dB_t \tag{243}$$
$$X_0 = Y_0 = x \tag{244}$$

Let

$$\begin{aligned} Z_t = &\int_0^t b(Y_s)'k(Y_s)dY_t \\ &- \frac{1}{2}\int_0^t b(Y_s)'k(Y_s)b(Y_s)ds \end{aligned} \tag{245}$$
$$\frac{d\mathbb{P}}{d\mathbb{Q}_t} = e^{Z_t} \tag{246}$$

where

$$k(Y_s) = (c(Y_s)'c(Y_s))^{-1} \tag{247}$$

Then under $\mathbb{Q}_t$ the process Y has the same distribution as the process X under $\mathbb{P}$.

Proof. We apply Girsanov's version II

$$\begin{aligned} Z_t = &\int_0^t b(Y_s)'(c(Y_s)^{-1})'dB_s \\ &+ \frac{1}{2}\int_0^t b(Y_s)'(c(Y_s)^{-1})'c(Y_s)^{-1}b(Y_s)ds \end{aligned} \tag{248}$$
$$\begin{aligned} &= \int_0^t b(Y_s)'(c(Y_s)^{-1})'c(Y_s)^{-1}(dY_t - b(Y_s)ds) \\ &+ \frac{1}{2}\int_0^t b(Y_s)'(c(Y_s)^{-1})'c(Y_s)^{-1}b(Y_s)ds \end{aligned} \tag{249}$$
$$\begin{aligned} &= \int_0^t b(Y_s)'k(Y_s)dY_t \\ &- \frac{1}{2}\int_0^t b(Y_s)'k(Y_s)b(Y_s)ds \end{aligned} \tag{250}$$
$$\frac{d\mathbb{P}}{d\mathbb{Q}_t} = e^{Z_t} \tag{251}$$

□

Informal Discrete Time Based Proof: For a given path x the ratio of the probability density of x under P and Q can be approximated as follows

$$\frac{d\mathbb{P}(x)}{d\mathbb{Q}_t(x)} \approx \prod_k \frac{p(x_{t_{k+1}} - x_{t_k} \mid x_{t_k})}{q(x_{t_{k+1}} - x_{t_k} \mid x_{t_k})} \tag{252}$$

were $\pi = \{0 = t_0 < t_1 \cdots < t_n = T\}$ is a partition of $[0, T]$ and

$$p(x_{t_{k+1}} \mid x_{t_k}) = \mathcal{G}(\Delta x_{t_k} \mid a(x_{t_k})\Delta t_k, \Delta t_k k(x_t)^{-1}) \tag{253}$$
$$q(x_{t_{k+1}} \mid x_{t_k}) = \mathcal{G}(\Delta x_{t_k} \mid 0, \Delta t_k k(x_t)^{-1}) \tag{254}$$

where $\mathcal{G}(\cdot, \mu, \sigma)$ is the multivariate Gaussian distribution with mean μ and covariance matrix c. Thus

$$\log \frac{d\mathbb{P}(x)}{d\mathbb{Q}_t(x)} \approx \sum_{k=0}^{n-1} -\frac{1}{2\Delta t_k}\Big((\Delta x_{t_k} - a(x_{t_k})\Delta t_k)' k(x_t)(\Delta x_{t_k} - a(x_{t_k})\Delta t_k) - \Delta x_{t_k}' k(x_t) \Delta x_{t_k}\Big)$$

$$= \sum_{k=0}^{n-1} a(x_{t_k})' k(x_t) \Delta x_{t_k} - \frac{1}{2} a(x_{t_k})' k(x_{t_k}) a(x_{t_k}) \Delta t_k \tag{255}$$

taking limits as $|\pi| \to 0$

$$\log \frac{d\mathbb{P}(x)}{d\mathbb{Q}_t(x)} = \int_0^T a(X_t)' k(X_t)\ dX_t - \frac{1}{2}\int_0^T a(X_t)' k(X_t) a(X_t)\ dt \tag{256}$$

□

Theorem 11.1. *The $d\Lambda_t$ differential*
Let X_t be an Ito process of the form

$$dX_t = a(t, X_t)dt + c(t, X_t)dB_t \tag{257}$$

Let

$$\Lambda_t = e^{Z_t} \tag{258}$$

$$Z_t = \int_0^t a(s, X_s)' k(t, X_t) dX_s - \frac{1}{2}\int_0^t a(s, X_s)' k(s, X_s) a(s, X_s) ds \tag{259}$$

where $k(t, x) = (c(t, x)c(t, x)')^{-1}$. Then

$$d\Lambda_t = \Lambda_t a(t, X_t)' k(t, X_t) dX_t \tag{260}$$

Proof. From Ito's product rule

$$d(\Lambda_t f(X_t)) = f(X_t) d\Lambda_t + \Lambda_t f(X_t) + (d\Lambda_t)(df(X_t) \tag{261}$$

Note

$$dX_t' k(t, X) a(t, X_t) a'(t, k(t, X_t) dX_t \tag{262}$$

$$= d\Lambda_t = (\nabla_z \Lambda_t)' dZ_t + \frac{1}{2} dZ_t' (\nabla_z^2 \Lambda_t) dZ_t \tag{263}$$

$$= \Lambda_t \Big(dZ_t + \frac{1}{2} dZ_t' dZ_t\Big) \tag{264}$$

Moreover, from the definition of Z_t

$$dZ_t = a(t, X_t)' k(t, X_t) dX_t - \frac{1}{2} a(t, X_t)' k(t, X_t) a(t, X_t) dt \tag{265}$$

Thus

$$dZ_t'dZ_t = dX_t'k(t,X)a(t,X_t)a'(t,X_t)k(t,X_t)dX_t \tag{266}$$

$$=dB_t'c(t,X_t)'k(t,X)a(t,X_t)a'(t,X_t)k(t,X_t)c(t,X_t)dB_t \tag{267}$$

$$= \text{trace}\Big(c(t,X_t)'k(t,X_t)a(t,X_t)a'(t,X_t)k(t,X_t)c(t,X_t)\Big)dt \tag{268}$$

$$= \text{trace}\Big(c(t,X_t)c(t,X_t)'k(t,X_t)a(t,X_t)a'(t,X_t)k(t,X_t)\Big)dt \tag{269}$$

$$= \text{trace}\Big(a(t,X_t)a'(t,X_t)k(t,X_t)\Big)dt \tag{270}$$

$$= \text{trace}\Big(a(t,X_t)a'(t,X_t)k(t,X_t)\Big)dt \tag{271}$$

$$= \text{trace}\Big(a'(t,X_t)k(t,X_t)a(t,X_t)\Big)dt \tag{272}$$

$$= a'(t,X_t)k(t,X_t)a(t,X_t)dt \tag{273}$$

Thus

$$d\Lambda_t = \Lambda_t a(t,X_t)'k(t,X_t)dX_t \tag{274}$$

□

12 Zakai's Equation

Let

$$dX_t = a(X_t)dt + c(X_t)dB_t \tag{275}$$

$$dY_t = g(X_t)dt + h(X_t)dW_t \tag{276}$$

$$\Lambda_t = e^{Z_t} \tag{277}$$

$$Z_t = \int_0^t g(Y_t)'k(Y_t)dY_t + \frac{1}{2}\int_0^t g(Y_t)'k(Y_t)g(Y_t)dt \tag{278}$$

$$k(Y_t) = (h(Y_t)h(Y_t)')^{-1} \tag{279}$$

Using Ito's product rule

$$d(f(X_t)\Lambda_t) = \Lambda_t df(X_t) + f(X_t)d\Lambda_t + df(X_t)d\Lambda_t \tag{280}$$

where

$$df(X_t) = \nabla_x f(X_t)'dX_t + \frac{1}{2}\text{trace}\Big(c(X_t)c'(X_t)\nabla_x^2 f(X_t)\Big)dt$$

$$= G_t[f](x) + \nabla_x f(X_t)c(X_t)dB_t \tag{281}$$

$$d\Lambda_t = \Lambda_t g(Y_t)k(Y_t)dY_t \tag{282}$$

Following the rules of Ito's calculus we note $dX_t dY_t'$ is an $n \times m$ matrix of zeros. Thus

$$d(f(X_t)\Lambda_t) = \Lambda_t\Big(G_t[f](x) + \nabla_x f(X_t)c(X_t)dB_t + g(Y_t)k(Y_t)dY_t\Big) \tag{283}$$

13 Solving Stochastic Differential Equations

$$\text{Let } dX_t = a(t,X_t)dt + c(t,X_t)dB_t. \tag{284}$$

Conceptually, this is related to $\frac{dX_t}{dt} = a(t,X_t) + c(t,X_t)\frac{dB_t}{dt}$ where $\frac{dB_t}{dt}$ is white noise. However, $\frac{dB_t}{dt}$ does not exist in the usual sense, since Brownian motion is nowhere differentiable with probability one.

We interpret solving for (284), as finding a process X_t that satisfies

$$X_t = M + \int_0^t a(s, X_s)\, ds + \int_0^t c(s, X_s)\, dBs. \tag{285}$$

for a given standard Brownian process B. Here X_t is an Ito process with $a(s, X_s) = K_s$ and $c(s, X_s) = H_s$.

$a(t, X_t)$ is called the drift function.

$c(t, X_t)$ is called the dispersion function (also called diffusion or volatility function). Setting $b = 0$ gives an ordinary differential equation.

Example 1: Geometric Brownian Motion

$$dX_t = aX_t dt + bX_t dB_t \tag{286}$$
$$X_0 = \xi > 0 \tag{287}$$

Using Ito's rule on $\log X_t$ we get

$$d \log X_t = \frac{1}{X_t} dX_t + \frac{1}{2}\left(-\frac{1}{X_t^2}\right)^2 (dX_t)^2 \tag{288}$$
$$= \frac{dX_t}{X_t} - \frac{1}{2} b^2 dt \tag{289}$$
$$= \left(aX_t - \frac{1}{2} b^2\right) dt + \alpha dB_t \tag{290}$$

Thus

$$\log X_t = \log X_0 + \left(a - \frac{1}{2} b^2\right) t + bB_t \tag{291}$$

and

$$X_t = X_0 e^{\left(a - \frac{1}{2} b^2\right) t + bB_t} \tag{292}$$

Processes of the form

$$Y_t = Y_0 e^{\alpha t + \beta B_t} \tag{293}$$

where α and β are constant, are called *Geometric Brownian Motions*. Geometric Brownian motion X_t is characterized by the fact that the log of the process is Brownian motion. Thus, at each point in time, the distribution of the process is log-normal.

Let's study the dynamics of the average path. First let

$$Y_t = e^{bB_t} \tag{294}$$

Using Ito's rule

$$dY_t = be^{bB_t}dB_t + \frac{1}{2}b^2e^{bB_t}(dB_t)^2 \tag{295}$$

$$Y_t = Y_0 + b\int_0^t Y_s dB_s + \frac{1}{2}b^2 \int +0^t Y_s ds \tag{296}$$

$$\mathbb{E}(Y_t) = \mathbb{E}(Y_0) + \frac{1}{2}b^2 \int_0^t \mathbb{E}(Y_s)ds \tag{297}$$

$$\frac{d\mathbb{E}(Y_t)}{d_t} = \frac{1}{2}b^2\mathbb{E}(Y_t) \tag{298}$$

$$\mathbb{E}(Y_t) = \mathbb{E}(Y_0)e^{\frac{1}{2}b^2t} = e^{\frac{1}{2}b^2t} \tag{299}$$

Thus

$$\mathbb{E}(X_t) = \mathbb{E}(X_0)e^{\left(a-\frac{1}{2}b^2\right)t}\mathbb{E}(Y_t) = \mathbb{E}(X_0)e^{\left(a-\frac{1}{2}b^2\right)t} \tag{300}$$

Thus the average path has the same dynamics as the noiseless system. Note the result above is somewhat trivial considering

$$\mathbb{E}(dX_t) = d\mathbb{E}(X_t) = \mathbb{E}(a(X_t))dt + \mathbb{E}(c(X_t)dB_t) = \mathbb{E}(a(X_t))dt + \mathbb{E}(c(X_t))\mathbb{E}(dB_t) \tag{301}$$

$$\tag{302}$$

$$d\mathbb{E}(X_t) = \mathbb{E}(a(X_t))dt \tag{303}$$

and in the linear case

$$\mathbb{E}(a(X_t))dt = \mathbb{E}(a_tX_t + u_t)dt = a_t\mathbb{E}(X_t) + u_t)dt \tag{304}$$

These symbolic operations on differentials trace back to the corresponding integral operations they refer to.

14 Linear SDEs

14.1 The Deterministic Case (Linear ODEs)

Constant Coefficients Let $x_t \in \Re^n$ be defined by the following ode

$$\frac{dx_t}{dt} = ax_t + u \tag{305}$$

The solution takes the following form:

$$x_t = e^{at}x_0 + a^{-1}(e^{at} - I)u \tag{306}$$

To see why note

$$\frac{dx_t}{dt} = ae^{at}x_0 + e^{at}u \tag{307}$$

and

$$ax_t + u = ae^{at}x_0 + e^{at}u - u + u = dx_t dt \tag{308}$$

Example: Let x_t be a scalar such that

$$\frac{dx_t}{dt} = \alpha\left(u - x_t\right) \tag{309}$$

Thus

$$\begin{aligned} x_t &= e^{-\alpha t}x_0 - \frac{1}{\alpha}(e^{-\alpha t} - 1)\alpha u \\ &= e^{-\alpha t}x_0 + (1 - e^{-\alpha_t})u \end{aligned} \tag{310}$$

Time variant coefficients Let $x_t \in \Re^n$ be defined by the following ode

$$\frac{dx_t}{dt} = a_t x_t + u_t \tag{311}$$

$$x_o = \xi \tag{312}$$

where u_t is known as the driving, or input, signal. The solution takes the following form:

$$x_t = \Phi_t \left(x_0 + \int_0^t \Phi_s^{-1} u_s ds \right) \tag{313}$$

where Φ_t is an $n \times n$ matrix, known as the fundamental solution, defined by the following ODE

$$\frac{d\Phi_t}{dt} = a_t \Phi_t \tag{314}$$

$$\Phi_0 = I_n \tag{315}$$

14.2 The Stochastic Case

Linear SDEs have the following form

$$dX_t = (a_t X_t + u_t)dt + \sum_{i=1}^{m} (c_{i,t} X_t + v_{i,t})\, dB_{i,t} \tag{316}$$

$$= (a_t X_t + u_t)dt + v_t dB_t + \sum_{i=1}^{m} c_{i,t} X_t dB_{i,t} \tag{317}$$

$$X_0 = \xi \tag{318}$$

where X_t is an n dimensional random vector, $B_t = (B_1, \cdots, B_m)$, $b_{i,t}$ are $n \times n$ matrices, and $v_{i,t}$ are the n-dimensional column vectors of the $n \times m$ matrix v_t. If $b_{i,t}t = 0$ for all i, t we say that the SDE is *linear in the narrow sense*. If $v_t = 0$ for all t we say that the SDE is *homogeneous*. The solution has the following form

$$X_t = \Phi_t \left(X_0 + \int_0^t \Phi_s^{-1} \left(u_s - \sum_{i=1}^{m} b_{i,s} v_{i,s} \right) ds + \int_0^t \Phi_s^{-1} \sum_{i=1}^{m} v_{i,s} dB_{i,s} \right) \tag{319}$$

where Φ_t is an $n \times n$ matrix satisfying the following matrix differential equation

$$d\Phi_t = a_t \Phi_t dt + \sum_{i=1}^{m} b_{i,s} \Phi_s dB_{i,s} \tag{320}$$

$$\Phi_0 = I_n \tag{321}$$

One property of the linear Ito SDEs is that the trajectory of the expected value equals the trajectory of the associated deterministic system with zero noise. This is due to the fact that in the Ito integral the integrand is independent of the integrator dB_t:

$$\mathbb{E}(dX_t) = d\mathbb{E}(X_t) = \mathbb{E}(a(X_t))dt + \mathbb{E}(c(X_t)dB_t) = \mathbb{E}(a(X_t))dt + \mathbb{E}(c(X_t))\mathbb{E}(dB_t) \tag{322}$$

$$\tag{323}$$

$$d\mathbb{E}(X_t) = \mathbb{E}(a(X_t))dt \tag{324}$$

and in the linear case

$$\mathbb{E}(a(X_t))dt = \mathbb{E}(a_t X_t + u_t)dt = a_t(\mathbb{E}(X_t) + u_t)dt \tag{325}$$

14.3 Solution to the Linear-in-Narrow-Sense SDEs

In this case

$$dX_t = (a_t X_t + u_t)\, dt + v_t dB_t \tag{326}$$

$$X_0 = \xi \tag{327}$$

where $v_1, \cdots, v_m$ are the columns of the $n \times m$ matrix v, and dB_t is an m-dimensional Brownian motion. In this case the solution has the following form

$$X_t = \Phi_t \left(X_0 + \int_0^t \Phi_s^{-1} u_s ds + \int_0^t \Phi_s^{-1} v_s dB_s \right) \tag{328}$$

where Φ is defined as in the ODE case

$$\frac{d\Phi_t}{dt} = a_t \tag{329}$$

$$\Phi_0 = I_n \tag{330}$$

White Noise Interpretation This solution can be interpreted using a "symbolic" view of white noise as

$$W_t = \frac{dB_t}{d_t} \tag{331}$$

and thinking of the SDE as an ordinary ODE with a driving term given by $u_t + v_t W_t$. We will see later that this interpretation breaks down for the more general linear case with $b_t \neq 0$.

Moment Equations Let

$$\rho_{r,s} \stackrel{\text{def}}{=} E\left((X_r - m_r)(X_s - m_s)\right) \tag{332}$$

$$\rho_t^2 \stackrel{\text{def}}{=} \rho_{t,t} = \text{Var}(X_t) \tag{333}$$

Then

$$\frac{d\mathbb{E}(X_t)}{dt} = a_t d\mathbb{E}(X_t) + u_t \tag{334}$$

$$\mathbb{E}(X_t) = \Phi_t \left(\mathbb{E}(X_0) + \int_0^t \Phi_s^{-1} u_s ds \right) \tag{335}$$

$$\rho_t^2 = \Phi_t \left(\rho_0^2 + \int_0^t \Phi_s^{-1} v_s \left(\Phi_s^{-1} v_s\right)^T ds \right) \Phi_t^T \tag{336}$$

$$\frac{d\rho_t^2}{dt} = a_t \rho_t^2 + \rho_t^2 a_t^T + v_t v^T \tag{337}$$

$$\rho_{r,s} = \Phi_r \left(\rho_0^2 + \int_0^{r \wedge s} \Phi_t^{-1} v_t \left(\Phi_t^{-1} v_t\right)^T dt \right) \Phi_s^T \tag{338}$$

where $r \wedge s = \min\{r, s\}$. Note the mean evolves according to the equivalent ODE with no driving noise.

Constant Coefficients:

$$dX_t = aX_t + u + vdB_t \tag{339}$$

In this case

$$\Phi_t = e^{at} \tag{340}$$

$$\mathbb{E}(X_t) = e^{at} \Big(\mathbb{E}(X_0) + \int_0^t e^{-as} u ds \Big) \tag{341}$$

$$\text{Var}(X_t) = \rho_t^2 = e^{at} \left\{ \rho_0^2 + \int_0^t e^{-as} v v^T \left(e^{-as}\right)^T ds \right\} \left(e^{at}\right)^T \tag{342}$$

Example: Linear Boltzmann Machines (Multidimensional OU Process)

$$dX_t = \theta(\gamma - X_t)dt + \sqrt{2\tau}dB_t \tag{343}$$

where θ is symmetric and $\tau > 0$. Thus in this case

$$a = -\theta, \tag{344}$$
$$u = \theta\gamma, \tag{345}$$
$$\Phi_t = e^{-t\theta} \tag{346}$$

and

$$X_t = e^{-t\theta}\left(X_0 + \int_0^t e^{s\theta} ds\theta\gamma + \sqrt{2\tau}\int_0^t e^{s\theta} dB_s\right) \tag{347}$$
$$= e^{-t\theta}\left(X_0 + \theta^{-1}\left[e^{s\theta}\right]_0^t \theta\gamma + \sqrt{2\tau}\int_0^t e^{s\theta} dB_s\right) \tag{348}$$
$$= e^{-t\theta}\left(X_0 + (e^{t\theta} - I)\gamma + \sqrt{2\tau}\int_0^t e^{-as} dB_s\right) \tag{349}$$

Thus

$$\mathbb{E}(X_t) = e^{-t\theta}\mathbb{E}(X_0) + \left(I - e^{-t\theta}\right)\gamma \tag{350}$$
$$\lim_{t\to\infty} \mathbb{E}(X_t) = \gamma \tag{351}$$

$$\text{Var}(X_t) = \rho_t^2 = e^{-t\theta}\left\{\rho_0^2 + 2\tau\int_0^t e^{2s\theta} ds\right\} e^{-t\theta} \tag{352}$$
$$= e^{-2t\theta}\left\{\rho_0^2 + 2\tau\frac{\theta^{-1}}{2}\left[e^{2s\theta}\right]_0^t\right\} \tag{353}$$
$$= e^{-2ts}\left\{\rho_0^2 + \tau\theta^{-1}\left(e^{2t\theta} - I\right)\right\} \tag{354}$$
$$= e^{-2t\theta}\rho_0^2 + \tau\theta^{-1}\left(I - e^{-2t\theta}\right) \tag{355}$$
$$= \tau\theta^{-1} + e^{-2t\theta}\left(\rho_0^2 - \tau\theta^{-1}\right) \tag{356}$$
$$\lim_{t\to\infty} \text{Var}(X_t) = \tau\theta^{-1} \tag{357}$$

where we used the fact that a, e^{at} are symmetric matrices, and $\int e^{at}dt = a^{-1}e^{at}$. If the distribution of X_0 is Gaussian, the distribution continues being Gaussian at all times.

Example: Harmonic Oscillator Let $X_t = (X_{1,t}, X_{2,t})^T$ represent the location and velocity of an oscillator

$$dX_t = aX_t + \begin{pmatrix} 0 \\ b \end{pmatrix} dB_t = \begin{pmatrix} 0 & 1 \\ -\alpha & -\beta \end{pmatrix} dX_t + \begin{pmatrix} 0 \\ b \end{pmatrix} dB_t \tag{358}$$

thus

$$X_t = e^{at}\left(X_0 + \int_0^t e^{-as}\begin{pmatrix} 0 \\ b \end{pmatrix} dB_s\right) \tag{359}$$

14.4 Solution to the General Scalar Linear Case

Here we will sketch the proof for the general solution to the scalar linear case. We have $X_t \in R$ defined by the following SDE

$$dX_t = (a_tX_t + u_t)\,dt + (b_tX_t + v_t)\,dB_t \tag{360}$$

In this case the solution takes the following form

$$X_t = \Phi_t \left(X_0 + \int_0^t \Phi_s^{-1}(u_s - v_s b_s)ds + \int_0^t \Phi_s^{-1} v_s dBs \right) \tag{361}$$

$$d\Phi_t = a_t \Phi_t dt + b_t \Phi_t dB_t \tag{362}$$

$$\Phi_0 = 1 \tag{363}$$

Sketch of the proof

- Use Ito's rule to show

$$d\Phi_t^{-1} = (b^2 - a_t)\Phi_t^{-1} dt - b_t \Phi_t^{-1} dB_t \tag{364}$$

- Use Ito's rule to show that if $Y_{1,t}, Y_{2,t}$ are scalar processes defined as follows

$$dY_{1,t} = a_1(t, Y_{1,t})dt + b_1(t, Y_{1,t})dB_t, \quad \text{for } i = 1, 2 \tag{365}$$

$$\tag{366}$$

  then

$$d(Y_{1,t}Y_{2,t}) = Y_{1,t}dY_{2,t} + Y_{2,t}dY_{1,t} + b_1(t, Y_{1,t})b_2(t, Y_{2,t})dt \tag{367}$$

- Use the property above to show that

$$d(X_t \Phi_t^{-1}) = \Phi_t^{-1} \left((u_t - v_t b_t)dt + v_t dB_t \right) \tag{368}$$

- Integrate above to get (361)
- Use Ito's rule to get

$$d \log \Phi_t^{-1} = \left(\frac{1}{2} b_t^2 - a_t \right) d_t - b_t dB_t \tag{369}$$

$$\log \Phi_t = -\log \Phi_t^{-1} = \int_0^t (a_s - \frac{1}{2} b_s) ds + \int_0^t b_s dB_s. \tag{370}$$

White Noise Interpretation Does not Work The white noise interpretation of the general linear case would be

$$dX_t = (a_t X_t + u_t)d_t + (b_t X_t + v_t)W_t dt \tag{371}$$

$$= (a_t + b_t W_t)X_t d_t + (u_t + v_t W_t)dt \tag{372}$$

If we interpret this as an ODE with noisy driving terms and coefficients, we would have a solution of the form

$$X_t = \Phi_t \left(X_0 + \int_0^t \Phi_s^{-1}(u_s + v_t W_t) \right) dt \tag{373}$$

$$= \Phi_t \left(X_0 + \int_0^t \Phi_s^{-1}(u_s - v_s b_s)ds + \int_0^t \Phi_s^{-1} v_s dBs \right) \tag{374}$$

$$\tag{375}$$

with

$$d\Phi_t = (a_t + b_t W_t)\Phi_t dt = a_t \Phi_t dt + b_t \Phi_t dB_t \tag{376}$$

$$\Phi_0 = 1 \tag{377}$$

The ODE solution to Φ would be of the form

$$\log \Phi_t = \int_0^t (a_s + b_s W_s) ds = \int_0^t a_s ds + \int_0^t b_s dB_s \tag{378}$$

which differs from the Ito SDE solution in (370) by the term $-\int_0^t b_s d_s/2$.

14.5 Solution to the General Vectorial Linear Case

Linear SDEs have the following form

$$dX_t = (a_t X_t + u_t)dt + \sum_{i=1}^{m} (b_{i,t} X_t + v_{i,t})\, dB_{i,t} \tag{379}$$

$$= (a_t X_t + u_t)dt + v dB_t + \sum_{i=1}^{m} b_{i,t} X_t dB_{i,t} \tag{380}$$

$$X_0 = \xi \tag{381}$$

where X_t is an n dimensional random vector, $B_t = (B_1, \cdots, B_m)$, $b_{i,t}$ are $n \times n$ matrices, and $v_{i,t}$ are the n-dimensional column vectors of the $n \times m$ matrix v_t. The solution has the following form

$$X_t = \Phi_t \left(X_0 + \int_0^t \Phi_s^{-1} \left(u_s - \sum_{i=1}^{m} b_{i,s} v_{i,s} \right) ds + \int_0^t \Phi_s^{-1} \sum_{i=1}^{m} v_{i,s} dB_{i,s} \right) \tag{382}$$

where Φ_t is an $n \times n$ matrix satisfying the following matrix differential equation

$$d\Phi_t = a_t \Phi_t dt + \sum_{i=1}^{m} b_{i,s} \Phi_s dB_{i,s} \qquad \Phi_0 = I_n \tag{383}$$

An explicit solution for Φ cannot be found in general even when a, b_i are constant. However if in addition of being constant they pairwise commute, $ag_i = g_i a$, $g_i g_j = g_j g_i$ for all i, j then

$$\Phi_t = \exp \left\{ \left(a - \frac{1}{2} \sum_{i=1}^{m} b_i^2 \right) t + \sum_{i=1}^{m} b_i B_{i,t} \right\} \tag{384}$$

Moment Equations Let $m_t = \mathbb{E}(X_t)$, $s_t = \mathbb{E}(X_t X_t^T)$, then

$$\frac{dm_t}{dt} = a_t m_t + u_t, \tag{385}$$

$$\frac{ds_t}{dt} = a_t s_t + s_t a_t^T + \sum_{i=1}^{m} b_{i,t} m_t b_{i,t}^T + u_t m_t^T + m_t u_t^T \tag{386}$$

$$+ \sum_{i=1}^{m} \left(b_{i,t} m_t v_{i,t}^T + v_{i,t} m_t^T b_{i,t}^T + v_{i,t} v_{i,t}^T \right), \tag{387}$$

The first moment equation can be obtained by taking expected values on (379). Note it is equivalent to the differential equation one would obtain for the deterministic part of the original system.

For the second moment equation apply Ito's product rule

$$dX_t X^T = X_t dX_t^T + (dX_t) X_t^T + (dX_t) dX_t^T$$

$$= X_t dX_t^T + (dX_t)\, X_t^T + \sum_{i=1}^{m} \left[(b_{i,t} X_t + v_{i,t}) \left(X_t^T b_{i,t}^T + v_{i,t} \right) \right] dt \tag{388}$$

substitute the dX_t for its value in (379) and take expected values.

Example: Multidimensional Geometric Brownian Motion

$$dX_{i,t} = a_i X_{i,t} dt + X_{i,t} \sum_{j=1}^{n} b_{i,j} dB_{j,t} \tag{389}$$

for $i = 1, \cdots, n$. Using Ito's rule

$$d \log X_{i,t} = \frac{dX_{i,t}}{X_{i,t}} + \frac{1}{2} \left(\frac{-1}{X_i, t}^{2} \right) (dX_{i,t})^2 = \tag{390}$$

$$\left(a_i - \sum_j \frac{1}{2} b_{i,j}^2 \right) + \sum_j b_{i,j} dB_{i,j} \tag{391}$$

Thus

$$X_{i,t} = X_{i,0} \exp \left\{ \left(a_i - \frac{1}{2} \sum_{j=1}^{n} b_{i,j}^2 \right) t + \sum_{j=1}^{n} b_{i,j} B_{j,t} \right\} \tag{392}$$

$$\log X_{i,t} = \log(X_{i,0}) + \left(a_i - \frac{1}{2} \sum_{j=1}^{n} b_{i,j}^2 \right) t + \log \left(\sum_{j=1}^{n} b_{i,j} B_{j,t} \right) \tag{393}$$

and thus $X_{i,t}$ has a log-normal distribution.

15 Important SDE Models

- Stock Prices: Exponential Brownian Motion with Drift

$$dX_t = aX_t dt + bX_t dB_t \tag{394}$$

- Vasiceck(1977) Interest rate model: OU Process

$$dX_t = \alpha(\theta - X_t)dt + bdB_t \tag{395}$$

- Cox-ingersol-Ross (1985) Interest rate model

$$dX_t = \alpha(\theta - X_t)dt + b\sqrt{X_t} dB_t \tag{396}$$

- Constant Elasticity of Variance process

$$dX_= \mu X_t + \sigma X_t^{\gamma} dB_t, \quad \gamma \geq 0 \tag{397}$$

- Generalized Cox Ingersoll Ross Model for short term interest rate, proposed by Chan et al (1992).

$$dX_t = (\theta_0 - \theta_1 X_t)dt + \gamma X_t^{\Psi} dB_t, \text{for } \Psi, \gamma > 0 \tag{398}$$

Let

$$\tilde{X}_s = \frac{X_s^{1-\Psi}}{\gamma(1-\Psi)} \tag{399}$$

Thus

$$d\tilde{X}_s = a(\tilde{X}_s)ds + dB_s \tag{400}$$

where

$$a(x) = \frac{(\theta_0 + \theta_1 \hat{x})\hat{x}^{-\Psi}}{\gamma} - \frac{\Psi\gamma}{2} \hat{x}^{\Psi - 1} \tag{401}$$

where

$$\hat{x} = (\gamma(1-\Psi)x)^{(1-\Psi)^{-1}} \tag{402}$$

A special case is when $\Psi = 0.5$. In this case the process increments are known to have a non-central chi-squared distribution (Cox, Ingersoll, Ross, 1985)

- Logistic Growth
  - Model I

$$dX_t = aX_t(1 - X_t/k)d_t + bX_t dB_t \tag{403}$$

The solution is

$$X_t = \frac{X_0 \exp\left\{(a - b^2/2)t + bB_t\right\}}{1 + \frac{X_0}{k} a \int_0^t \exp\left\{(a - b^2/2)s + bB_s\right\} ds} \tag{404}$$

  - Model II

$$dX_t = aX_t(1 - X_t/k)d_t + bX_t(1 - X_t/k)dB_t \tag{405}$$

  - Model III

$$dX_t = rX_t(1 - X_t/k)d_t - srX_t^2 dB_t \tag{406}$$

In all the models r is the Maltusian growth constant and k the carrying capacity of the environment. In model II, k is unattainable. In the other models X_t can have arbitrarily large values with nonzero probability.

- Gompertz Growth

$$dX_t = \alpha X_t d_t r X_t \log\left(\frac{k}{X_t}\right) d_t + bX_t dB_t \tag{407}$$

where r is the Maltusian growth constant and k the carrying capacity. For $\alpha = 0$ we get the Skiadas version of the model, for $r = 0$ we get the log-normal model. Using Ito's rule on $Y_t = e^{\beta t} \log X_t$ we can get expected value follows the following equation

$$\mathbb{E}(X_t) = \exp\left\{\log(x_0)e^{-rt}\right\} \exp\left\{\frac{\gamma}{r}(1 - e^{-rt})\right\} \exp\left\{\frac{b^2}{4r}(1 - e^{-2rt}\right\} \tag{408}$$

where

$$\gamma = \alpha - \frac{b^2}{2} \tag{409}$$

Something fishy in expected value formula. Try $\alpha = b = 0$!

16 Stratonovitch and Ito SDEs

Stochastic differential equations are convenient pointers to their corresponding stochastic integral equations. The two most popular stochastic integrals are the Ito and the Stratonovitch versions. The advantage of the Ito integral is that the integrand is independent of the integrator and thus the integral is a Martingale. The advantage of the Stratonovitch definition is that it does not require changing the rules of standard calculus. The Ito interpretation of

$$dX_t = f(t, X_t)dt + \sum_{j=1}^{m} g_j(t, X_t) dB_{j,t} \tag{410}$$

is equivalent to the Stratonovitch equation

$$dX_t = \left(f(t, X_t)dt - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left[\frac{\partial g_{i,j}}{\partial x_i} g_i j \right] (t, X_t) \right) dt + \sum_{j=1}^{m} g_j(t, X_t) dB_{j,t} \quad (411)$$

and the Stratonovitch interpreation of

$$dX_t = f(t, X_t)dt + \sum_{j=1}^{m} g_j(t, X_t) dB_{j,t} \quad (412)$$

is equivalent to the Ito equation

$$dX_t = \left(f(t, X_t)dt + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left[\frac{\partial g_{i,j}}{\partial x_i} g_i j \right] (t, X_t) \right) dt + \sum_{j=1}^{m} g_j(t, X_t) dB_{j,t} \quad (413)$$

17 SDEs and Diffusions

- Diffusions are rocesses governed by the Fokker-Planck-Kolmogorov equation.
- All Ito SDEs are diffusions, i.e., they follow the FPK equation.
- There are diffusions that are not Ito diffusions, i.e., they cannot be described by an Ito SDE. Example: diffusions with reflection boundaries.

18 Appendix I: Numerical methods

18.1 Simulating Brownian Motion

18.1.1 Infinitesimal "piecewise linear" path segments

Get n independent standard Gaussian variables $\{Z_1, \cdots, Z_n\}$;
$\mathbb{E}(Z_i) = 0$; $\mathrm{Var}(Z_i) = 1$. Define the stochastic process $\hat{B}$ as follows,

$$\hat{B}_{t_0} = 0 \quad (414)$$

$$\hat{B}_{t_1} = \hat{B}_{t_0} + \sqrt{t_1 - t_0} Z_1 \quad (415)$$

$$\vdots \quad (416)$$

$$\hat{B}_{t_k} = \hat{B}_{t_{k-1}} + \sqrt{t_k - t_{k-1}} Z_k \quad (417)$$

Moreover,

$$\hat{B}_t = \hat{B}_{t_{k-1}} \text{ for } t \in [t_{k-1}, t_k) \quad (418)$$

This defines a continuous time process that converges in distribution to Brownian motion as $n \to \infty$.

18.1.2 Linear Interpolation

Same as above but linearly interpolating the starting points of path segments.

$$\hat{B}_t = \hat{B}_{t_{k-1}} + (t - t_k)(\hat{B}_{t_{k-1}} - \hat{B}_{t_k})/(t_k - t_{k-1}) \text{ for } t \in [t_{k-1}, t_k) \quad (419)$$

Note this approach is non-causal, in that it looks into the future. I believe it is inconsistent with Ito's interpretation and converges to Stratonovitch solutions

18.1.3 Fourier sine synthesis

$\hat{B}_t(\omega) = \sum_{k=0}^{n-1} Z_k(\omega)\phi_k(t)$

where $Z_k(\omega)$ are same random variables as in previous approach, and $\phi_k(t) = \frac{2\sqrt{2T}}{(2k+1)R} \sin\left(\frac{(2k+1)Rt}{2T}\right)$
As $n \to \infty$ B converges to BM in distribution. Note this approach is non-causal, in that it looks into the future. I believe it is inconsistent with Ito's interpretation and converges to Stratonovitch solutions

18.2 Simulating SDEs

Our goal is to simulate

$$dX_t = a(X_t)dt + c(X_t)dB_t, \quad 0 \le t \le T X_0 = \xi \tag{420}$$

Order of Convergencce Let $0 = t_1 < t_2 \cdots < t_k = T$

A method is said to have *strong oder of convergence* α if there is aconstant K such that

$$\sup_{t_k} \mathbb{E}\left|X_{t_k} - \hat{X}_k\right| < K(\Delta_{t_k})^\alpha \tag{421}$$

A method is said to have *week oder of convergence* α if there is a constant K such that

$$\sup_{t_k} \left|\mathbb{E}[X_{t_k}] - \mathbb{E}[\hat{X}_k]\right| < K(\Delta_{t_k})^\alpha \tag{422}$$

Euler-Maruyama Method

$$\hat{X}_k = \hat{X}_{k-1} + a(\hat{X}_{k-1})(t_k - t_{k-1}) + c(\hat{X}_{k-1})(B_k - B_{k-1}) \tag{423}$$

$$B_k = B_{k-1} + \sqrt{t_k - t_{k-1}} Z_k \tag{424}$$

where $Z_1, \cdots, Z_n$ are independent standard Gaussian random variables.

The Euler-Maruyama method has strong convergence of order $\alpha = 1/2$, which is poorer of the convergence for the Euler method in the deterministic case, which is order $\alpha = 1$. The Euler-Maruyama method has week convergence of order $\alpha = 1$.

Milstein's Higher Order Method: It is based on a higher order truncation of the Ito-Taylor expansion

$$\hat{X}_k = \hat{X}_{k-1} + a(\hat{X}_{k-1})(t_k - t_{k-1}) + c(\hat{X}_{k-1})(B_k - B_{k-1}) \tag{425}$$

$$+ \frac{1}{2}c(X_{k-1})\nabla_x c(X_{k-1})\Big((B_k - B_{k-1})^2 - (t_k - t_{k-1})\Big) \tag{426}$$

$$B_k = B_{k-1} + \sqrt{t_k - t_{k-1}} Z_k \tag{427}$$

where $Z_1, \cdots, Z_n$ are independent standard Gaussian random variables. This method has strong convergence of order $\alpha = 1$.

19 History

- The first version of this document, which was 17 pages long, was written by Javier R. Movellan in 1999.

- The document was made open source under the GNU Free Documentation License 1.1 on August 12 2002, as part of the Kolmogorov Project.
- October 9, 2003. Javier R. Movellan changed the license to GFDL 1.2 and included an endorsement section.
- March 8, 2004: Javier added Optimal Stochastic Control section based on Oksendals book.
- September 18, 2005: Javier. Some cleaning up of the Ito Rule Section. Got rid of the solving symbolically section.
- January 15, 2006: Javier added new stuff to the Ito rule. Added the Linear SDE sections. Added Boltzmann Machine Sections.
- January/Feb 2011. Major reorganization of the tutorial.