OVERVIEW

First set of lectures: Basics of Brownian motion, Stochastic Differential Equations, Ito calculus, connection with jump processes, Fokker-Planck equations and related PDE’s

References:
Oksendal, Stochastic Differential Equations: An Introduction with Applications
Schuss, Introduction to Stochastic Differential Equations
Gardiner, Handbook of Stochastics Methods for Physics and Chemistry

Second set of lecturers: Asymptotic approximations, including large number of events, different time scales, small noise, and stochastic averaging

References:
Stuart and Pavliotis, Multiscale methods: Homogenization and Averaging
Bender and Orszag, Advanced Mathematical Methods for Scientists and Engineers
(Also references with specific examples)
1 From Einstein to Ito

Motivating Examples

1.1 Basic stochastic models: continuous time and state space

Characteristics of the noise:

*Example 1: Motion of a Brownian particle:*

\[ mx_{tt} = -\eta x_t + X \]  

Random

Figure 1: Realization for noise process, particle position, and velocity
Langevin gave an alternative (heuristic) derivation to that of Einstein
(physically motivated)

\[ m = \text{mass} \]
\[ \eta = \text{viscous drag} \]
\[ X = \text{irregular fluctuations}: \langle X \rangle = 0. \]

\[ T = \text{temperature} \]

From statistical mechanics: \[ \frac{1}{2} m \langle v^2 \rangle = \frac{1}{2} m \langle x_t^2 \rangle = \frac{kT}{2} \text{ as } t \to \infty \]

\( (k \text{ is Boltzmann’s constant}) \) Then averaging the equation or \( x \) gives: \( x \to x_0 \) as \( t \to \infty \).

Note: we are averaging over realizations, \( \langle f(x) \rangle = \int f(x)p(x)dx \) for \( p(x) \) the (stationary) density of \( x \). The variation about initial condition has mean zero \( E[x - x_0] = \langle (x - x_0) \rangle \)

Now let’s look at the variance \( E[x^2] \): (take \( x_0 = 0 \) for simplicity)

Multiply equation of motion by \( x \), integrate:

(assume \( X \) is "nice")

\[ \Rightarrow mxx_{tt} = -\eta xx_t + X \cdot x \]

\[ \Rightarrow \frac{m}{2} ((x^2)_{tt} - 2(x_t)^2) = -\frac{\eta}{2} (x^2)_t + X \cdot x \]

\[ \Rightarrow \frac{m}{2} \langle x^2 \rangle_{tt} = -\frac{\eta}{2} \langle x^2 \rangle_t + kT + \langle X \rangle \langle x \rangle \]

What is \( E[x^2] \)?
assuming independence of $X$ and $x$

Then, we integrate the equation for $\langle x^2 \rangle_t$ to get

$$\langle x^2 \rangle_t = \frac{2kT}{\eta} + Ce^{-\frac{m}{\eta}t}$$

(5)

$$\Rightarrow \quad \text{as} \quad t \to \infty \quad \langle x^2 \rangle_t \to \frac{2kT}{\eta}t$$

(6)

**Variance of the particle position**

Then the variance $\langle x^2 \rangle$ behaves as a constant times $t$

Einstein had derived a similar result, looking for probability density corresponding to diffusion (later).

Note that Langevin’s derivation assumes that the usual rules of calculus apply.

**Highlights for the behaviour of $X$**

We consider the equation for $v = x_t$ to look more closely at the implications for the behavior of the noisy term $X$. What are these fluctuations?

Write the equations of motion of a Brownian particle:

$$mv_t = -\eta v + X$$

(7)
Formal solution for \( v \): (assuming \( X \) is a “nice” function)

\[
v = v_0 e^{-\frac{m t}{n}} + \frac{1}{m} \int_0^t e^{-\frac{m (t-s)}{n}} X(s) ds
\]  
\[(8)\]

so we can consider the variance of \( v \) about its mean:

\[
\left< \left( v - v_0 e^{-\frac{m t}{n}} \right)^2 \right> = \left< \frac{1}{m^2} \left( \int_0^t e^{-\frac{m (t-s)}{n}} X(s) ds \right)^2 \right>
\]
\[(9)\]

Let’s now compare the average of this term \((\int \ldots ds)^2\) to the result \( \frac{1}{2} m \left< v^2 \right> = \frac{k T}{2} \)

Thinking of the integral as the limit of a sum, we write

\[
\int_0^t e^{-\frac{m (t-s)}{n}} X(s) ds = \lim_{N \to \infty} \sum_{j=1}^N e^{-\frac{m}{n} (t-j \Delta s)} \Delta b_j,
\]
\[(10)\]

taking \( X(s) \approx \frac{\Delta b}{\Delta s} \)

(we use "b" for Brownian motion, which we see later)

That is, we view \( X \) as the derivative of random fluctuations (does not exist as \( \Delta s \to 0 \)).

Assuming the \( \Delta b_j \) are independent (independent increments in the random process) and \( \langle \Delta b_j^2 \rangle = q \Delta s \) yields

\[
\left< \left( \sum_{j=1}^N e^{-\frac{m}{n} (t-j \Delta s)} \Delta b_j \right)^2 \right> = \sum_{j=1}^N e^{-\frac{2m}{n} (t-j \Delta s)} q \Delta s + \sum_{j \neq k} \langle \Delta b_j \Delta b_k \rangle
\]
\[(11)\]
For independent increments, the last term vanishes, yielding, as $\Delta s \to 0$, an integral

$$\int_0^t e^{-\frac{2\eta}{m(t-s)}q} ds$$

which can be easily evaluated as $\text{Const} \cdot (1 - e^{-\frac{2\eta}{m}t})$.

Then, $\frac{m}{2} \langle v^2 \rangle = \text{const}$, as $t \to \infty$. We can then choose $q$ appropriately so that

$$m \langle v^2 \rangle = kT.$$ 

So we have $X(s)$ as a “derivative” of the random fluctuations, such that $\langle (X(s)\Delta s)^2 \rangle = \Delta s$ and, furthermore, have assumed $x$ is independent of $X$ ($\langle Xx \rangle = \langle X \rangle \langle x \rangle$ previously). Essentially this follows from the assumption of independent increments of the random fluctuations.

These calculations illustrate several key mathematical properties of Brownian motion = standard Wiener process $w(t)$:

- $w(t+s) - w(t)$ is independent of $t$ and independent of $w(t) - w(t-u)$ ($u \geq 0, s \geq 0$)

- Paths of $w(t)$ are continuous ($X(t)$ undefined, but integrated, gives continuous paths for $w$)

- Joint probability distribution of $(w(t_1), w(t_2), ..., w(t_n))$ is mean zero Gaussian $E[w^2(t)] = t$ (normalized)

  $w(t)$ is distributed as $\text{N}(0, \text{sqrt}(t))$  \text{Var}(w) = t
Covariance:
\[ E[w(s)w(t)] = E([w(t) - w(s)]w(s)) + E[w^2(s)] = 0 + s = \min(t, s) \text{ (for } s < t) \]

**Simulation of SDE’s and Langevin-type stochastic models:**

Langevin-type models

\[
dx = f(x)dt + qdw
\]  

(12)

We write this in differential form, rather than

\[
\frac{dx}{dt} = f(x) + q\frac{dw}{dt}
\]

(13)

since \( w \) is continuous but not differentiable.

General Stochastic Differential Equations (SDE):

\[
dx = A(x, t)dt + B(x, t)dw
\]

(14)

The behavior of the increment \( dw \) suggests a simple approach for numerical simulation:

Simulate at discrete time steps: \( t_j = j\Delta t \)

Discrete approximation for the derivative:

\[
\frac{dx}{dt} \approx \frac{x(t + \Delta t) - x(t)}{\Delta t}
\]
Behavior of $dw$:

$$dw \approx \Delta w \sim N(0, \sqrt{\Delta t})$$

Then,

$$x(t + \Delta t) = x(t) + A(x(t), t) \Delta t + B(x(t), t) \sqrt{\Delta t} Z$$

where $Z \sim N(0, 1)$.

In general, writing $x_j = x(j\Delta t)$ we have an iterative procedure to get $x_j$ for all $j$, starting with an initial condition $x = x_0$ and computing to a time $T = K\Delta t$.

$$x(1) = x_0;$$

for $j = 1 : K - 1$

$$x_{j+1} = x_j + A(x_j, t_j) \Delta t + B(x_j, t_j) \sqrt{\Delta t} Z_j$$

end

This is the Euler-Maruyama method. It is generally associated with the Ito interpretation of SDE’s (later). This has implications for the dynamics via the interpretation of the noise increments.

Higher order methods: Reference: Kloeden and Platen, Numerical Solution of SDE’s, 1992. Both Weak (in distribution) and Strong (pathwise) methods.

For higher dimensional SDE:

$$dx = A(x, t) dt + B(x, t) dw$$ (15)
\( \mathbf{x}, \mathbf{A} \) are \( d \)-dimensional, \( \mathbf{w} \) is \( n \)-dimensional, \( B \) is \( d \times n \) dimensional.

Example:

\[
mx_{tt} = -\eta x_t + qX
\]  

(16)

For \( X \) as above, we write

\[
\begin{align*}
dx &= vdt \\
dv &= -\frac{\eta}{m}vdt + \frac{q}{m}dw
\end{align*}
\]

drift is linear in \( v \)

\( v \) is an Ornstein-Uhlenbeck (O-U process)

Figure 2: Realization for noise process, particle position, and velocity
\[ x(1) = x_0; v(1) = v_0 \]
for \( j = 1 : K - 1 \)
\[ x_{j+1} = v_j \Delta t \]
\[ v_{j+1} = v_j - \frac{\eta}{m} v_j \Delta t + \frac{q}{m} \sqrt{\Delta t} Z_j \]
end

Note: \( v \) is an Ornstein-Uhlenbeck process (OU process). It has a linear drift term and a constant coefficient for the diffusion term. OU processes have a Gaussian stationary density with a constant mean and variance, i.e. normally distributed. That is, as \( t \to \infty \),
\[ v \sim N(0, \frac{q}{\sqrt{2\eta}}). \]

*Dynamics driven by Jump Processes - “Shot Noise”*

\[ \frac{dI}{dt} = -\alpha I + q\mu(t) \]  
(17)

\[ \mu(t) = \sum \delta(t - t_k) \]  
(18)

\( I=\)current, \( t_k=\)arrival time of electron. Note that \( \mu(t) \) is not mean zero.

Define: \( N(t)=\)sum of Poisson arrivals,
\[ \frac{dN}{dt} = \mu \]  
(19)
formally, but \( N \) is a step function. If we want to write the current equation in terms of a mean zero process (analogous to \( X \) in previous example) then consider \( E[dN] = \lambda dt \), \( Var[dN] = \lambda dt \), where \( \lambda \) is the mean arrival rate. Then

\[
d\eta = dN - \lambda dt
\]

is a centered Poisson increment, with mean zero, and variance \( \lambda dt \).

Formally

\[
\frac{dI}{dt} = -\alpha I + \lambda q + q \frac{d\eta}{dt}
\]  

\[(20)\]

*Note on Numerical simulation*

Here \( \eta \) is not a continuous process, rather, events occur at times that are exponentially distributed. Then some time interval increments of length \( \Delta t \) will have an event, and some will not.

Averaging yields

\[
\frac{d\langle I \rangle}{dt} = (-\alpha \langle I \rangle + \lambda q), \quad \langle I \rangle \rightarrow \frac{\lambda q}{\alpha} \quad \text{as} \quad t \rightarrow \infty
\]  

\[(21)\]

What about \( \langle I^2 \rangle \)?

Suppose we consider \( d(I^2) \) and compare with equation above:

\[
d(I^2) = (I + dI)^2 - I^2 = 2IdI + (dI)^2
\]  

\[(22)\]
Figure 3: Realization for $I$

and

$$(dI)^2 = ((-\alpha I + \lambda q)dt)^2 + 2(-\alpha I + \lambda q)dt\eta + q^2d\eta^2$$ \hfill (23)

Compare with heuristic as in previous example: multiply equation for $I$ by $I$ and integrate and average):

multiply equation (20) by $I$, and integrate to get

$$\frac{1}{2} \frac{d \langle I^2 \rangle}{dt} = -\alpha \langle I^2 \rangle + \lambda q \langle I \rangle + q \langle I \frac{d\eta}{dt} \rangle$$ \hfill (24)
What about $\langle I \frac{dn}{dt} \rangle$? Is it $= 0$? If so, then we can solve for $\langle I^2 \rangle$ as $t \to \infty$.

$$\langle I^2 \rangle = \left( \frac{\lambda q}{\alpha} \right)^2$$

$$\Rightarrow Var(I) = \langle I^2 \rangle - \langle I \rangle^2 = 0$$

! No effect of noise - not correct!

This heuristic neglects the term $(dI)^2$ - it is not in equation for $\langle I^2 \rangle$. So we need to examine that equation

$$(dI)^2 = (dt)^2 + \langle d\eta \rangle + q^2 \langle d\eta^2 \rangle$$

(25)

using the properties of the centered Poisson increment.

Recall: mean = variance for Poisson random variables

$$\langle d\eta \rangle = 0 \quad \langle (d\eta)^2 \rangle = \lambda dt$$

So now we can get the correct equation for $I^2$.

$$d \langle (I)^2 \rangle = 2 \langle IdI \rangle + \langle (dI)^2 \rangle$$

(26)

$$= \langle (2I(-\alpha I + \lambda q))dt \rangle + \langle (2qI d\eta) \rangle + q^2 \lambda dt + o(dt)$$

(27)

$$\Rightarrow Var \langle I^2 \rangle = \frac{q^2 \lambda}{2\alpha} \quad t \to \infty$$

(28)

Recall that for the Wiener process $\langle w(t) \rangle = 0$ and $\langle w^2 \rangle = t$. Later we see that the if Poisson events are very frequent, then one can replace Poisson random variables with
appropriate “centering” and Brownian motion (normal random variables).

**Other types of increments: e.g. Levy Processes**

Levy Processes can be characterized by a combination of Brownian motion and jumps.

A simple class of Levy Processes are alpha-stable processes. Instead of having a density with exponentially decaying tails, like the Gaussian distribution, they have “fat” tails; that is, the tails of the density have the behavior

\[ p(x) = |x|^\alpha + 1 \text{ as } x \to \infty \text{ for } \alpha < 2 \]

To simulate an SDE with alpha-stable noise increments, e.g.

\[ dx = H(x)dt + \kappa dL^{\alpha,\beta} \quad (29) \]

the iterative steps of the Euler-type method take the form

\[ x_{n+1} = x_n + H(x_n)\Delta t + \kappa \delta L^{\alpha,\beta}_n \quad \Delta L^{\alpha,\beta}_n \sim S_{\alpha}(\beta, (\Delta t)^{1/\alpha}) \quad (30) \]

The alpha-stable distribution is given in terms of its characteristic function, rather than the density, so \( X \sim S^{\beta,\sigma}_\alpha \) has a characteristic function of the form

\[ \psi(k) = \mathcal{F}[p(x)] = \int e^{ikx} p(x) dx = \exp(-\sigma^\alpha |k|^{\alpha}(1 - i\beta \text{sgn}(k)\phi(k))) \]

\[ \phi(k) = \begin{cases} \tan \left( \frac{\pi\alpha}{2} \right) & \alpha \neq 1 \\ -\frac{2}{\pi} \log |k| & \alpha = 1 \end{cases} \quad (31) \]
Note: for $\alpha = 2$, we have a Gaussian distribution.

Example:
Linear system forced by Levy noise (alpha-stable)

\[ \alpha = 2 \]

Gaussian

\[ \alpha < 1 \]

“fat tails”

\[ x(t) \]

\[ p(x) \]

time series

probability density on logarithmic scale

Example:
Linear system forced by Levy noise (alpha-stable)
2 Ito’s Formula

Now we ask:

What are the rules for functions of $w(t)$: $f(w(t))$?

In particular, we are interested in how to treat $df(w(t))$, write stochastic differential equations including $w(t)$

Ito’s formula: find equation for $df(x)$ given $dx = a dt + b dw$

*Simple example:* $y = w^2$

\[
\int_s^t dy = w^2(t) - w^2(s)
\]

(32)

Can we write $dy = a dt + b dw$? Or,

\[
\int dy = \int a dt + \int b dw
\]

(33)

where $a, b$ may be functions of $y, t$?

We need to define $bdw$: how do we evaluate or interpret $b dw$?
Ito interpretation:

\[
\int_s^t b(w) dw = \sum_j b(w(t_j))(w(t_{j+1}) - w(t_j))
\]  

(34)

Stratonovich: evaluate \( b \) at the average of \( w(t_j) \) and \( w(t_{j+1}) \)

\[
\int_s^t b(w) dw = \sum_j b \left( \frac{w(t_j) + w(t_{j+1})}{2} \right) (w(t_{j+1}) - w(t_j))
\]  

(35)

Back to the equation for \( y = w^2 \): We compare \( \int dy \) with \( \int w dw \) - we might expect this from usual rules of calculus.

If we use the Ito interpretation, then

Rewrite Ito definition in a convenient way

\[
\int_s^t w dw = \frac{1}{2} \left[ \sum_j w^2(t_{j+1}) - w^2(t_j) - (w(t_{j+1}) - w(t_j))^2 \right]
\]  

(36)

\[
= \frac{1}{2} (w^2(t) - w^2(s)) - \frac{1}{2} \sum_j (w(t_{j+1}) - w(t_j))^2 \text{ dw}_j^2
\]  

(37)

This looks like some integrated quantity evaluated at the endpoints \( t \) and \( s \) plus a sum.
Considering the expected value of the sum:

\[ E \left[ \sum_{j=1}^{N} (w(t_{j+1}) - w(t_j))^2 \right] = \sum_{j=1}^{N} E [w^2(t_{j+1})] + E [w^2(t_j)] - 2 \min(t_{j+1}, t_j) \]

\[ = \sum_{j=1}^{N} t_{j+1} + t_j - 2t_j = \sum_{j=1}^{N} \Delta t = t - s \]

Furthermore, one can show that

\[ \text{Var} \left[ \sum_{j=1}^{N} (w(t_{j+1}) - w(t_j))^2 \right] \sim \frac{1}{N} \]  

(referenced: Schuss, Introduction to SDE’s)

So

\[ 2 \int_{s}^{t} w dw = w^2(t) - w^2(s) - t - s = y(t) - y(s) - t - s \]  

in probability, as \( N \to \infty \). Then for \( y = w^2 \),

\[ dy = \underbrace{2w}_{b} dw + \underbrace{1}_{a=1} dt \]

Note: The expression is not simply \( dy = 2w \, dw \), but there is also an additional term \( dt \).
Next, we compute the rule for products

Consider $dx = adt + bdw$ and $dw$

For $a, b$ constant, $x = at + bw$, for $x(0) = 0$  

Using Ito interpretation

\[
\begin{align*}
    xw &= atw + (bw^2) \\
    d(xw) &= atdw + awdt + b(2wdw) + bdt \\
    &= xdw + wdx + bdt
\end{align*}
\]

And we can use this, together with the result for $y = w^2$ to find by induction (exercise)

\[
dw^m(t) = m(w(t))^{m-1}dw + \frac{m(m - 1)}{2}w^{m-2}dt \quad \text{for} \quad m \geq 2
\]

Then, for any Polynomial $P(w)$

\[
dP(w) = P'(w)dw + \frac{1}{2}P''(w)dt
\]

From there, we can write general “noise” functions as $f = g(t)\phi(w)$, so

\[
\begin{align*}
    df &= \phi g'(t)dt + gd\phi \\
    \Rightarrow df &= \left[\phi g'(t) + \frac{1}{2}g\phi''(w)\right]dt + g\phi'(w)dw \\
    &= \left[f_t + \frac{1}{2}f_{ww}\right]dt + (f_w)dw
\end{align*}
\]
Applying this term by term to a general expression,
\[ f(w, t) = \sum g(t)\phi(w), \]  
(49)
yields the same result for \( f(w, t) \)

Finally, for \( f(x, t) \), \( x = at + bw \), \( a,b \), constant
\[ \frac{\partial f}{\partial t} = \frac{\partial f}{\partial t} + a \frac{\partial f}{\partial x} \]  
\[ \frac{\partial f}{\partial w} = b \frac{\partial f}{\partial x} \]  
\[ \frac{\partial^2 f}{\partial w^2} = b^2 \frac{\partial^2 f}{\partial x^2} \]  

Then
\[ df = \left[ \frac{\partial f}{\partial t} + a \frac{\partial f}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2 f}{\partial x^2} \right] dt + b \frac{\partial f}{\partial x} dw \]  
(51)

This is Ito’s formula, relating the equation for \( f \) to the equation for \( x \), \( dx = adt + bdw \).

It can be generalized for \( a(x, t) \) and \( b(x, t) \).

Note: Throughout this we use the Ito interpretation for \( ()dw \).

Is there a Stratonovich Formula?

(S) \( \int f(w) \, dw \) follows the usual rules of calculus
\[ f(w_i + \text{Delta } w_i/2) \]

(S) \[ \int f(w) \, dw = \sum f \left( \frac{w_i + w_{i+1}}{2}, t_i \right) \Delta w_i \sim \sum f(w_i) \Delta w_i + \frac{(\Delta w_i)^2}{2} f'(w_i) \]

Simple example:
(S) \[ w \, dw = w \, dw + \frac{(\Delta w)^2}{2} \sim w \, dw + \frac{\Delta t}{2} = \frac{1}{2} d(w^2) \]

\[ dx = adt + bdw(S) = adt + b \circ dw = (a + \frac{1}{2} bb_x) dt + bdw \]

\( b=\text{const}, \ (S) = (I) \quad \text{Show as Exercise: repeat same procedures as above, for powers of } w, \text{ polynomials in } w \]

\[ \text{products of functions of } t \text{ and } w, \text{ etc} \]

For Levy Processes, the analogy for the Stratonovich interpretation is known as the Marcus interpretation. Here, written in an SDE with increments of an alpha-stable process

\[ dx = a(x) dt + b(x) \diamond dL_{\alpha,\beta} \quad (52) \]

For \( \alpha = 2 \) it is equivalent to the Stratonovich interpretation.

To implement this for evaluation/simulation, the Marcus integral takes the form

\[ \int_0^t b(x_s) \diamond dL_{\alpha,\beta}^s = \sum_{s \leq t} \theta(1; \Delta L_s, x_s^-) - x_s^- \]

\[ \frac{d\theta}{dr} = \Delta L_x b(\theta), \ \theta(0) = z_s^- \quad (53) \]

The quantity \( \theta \) here is known as the Marcus map, and it is a time-like variable that travels infinitely fast along a curve connecting across the jumps. Then, a numerical approximation of \( dx = a(x) dt + b(x) \diamond dL_{\alpha,\beta} \) is given by

\[ x_{n+1} = x_n + a(x_n) \Delta t + (\theta(1; \Delta L_{\alpha,\beta}^x, x_n) - x_n) \quad (54) \]
Additional references on implementation:

Notes at: http://www.math.ubc.ca/~rachel/cv/lec1jul15_rw.pdf

Printed copies tomorrow, post on summer school website next week
3 Master, Forward, and Backward equations

**Random walk as a Markov process**

where $X =$ position

and $X_j =$ position after $j$ steps.

$P$(step up: $X_{n+1} = X_n + 1$) = $p$

$P$(step down: $X_{n+1} = X_n - 1$) = $1 - p$

Each step is independent of previous steps (Markovian) (depends only on present location):

(conditional probability)

$$P(X_n = k) = P(X_n = k | X_{n-1} = k - 1)P(X_{n-1} = k - 1)$$
$$+ P(X_n = k | X_{n-1} = k + 1)P(X_{n-1} = k + 1)$$

$$= pP(X_{n-1} = k - 1) + (1 - p)P(X_{n-1} = k + 1)$$

In general,

$$P(X_n = k) = P( \text{location after n steps, } m + \text{steps, } l - \text{steps,}$$

with $m - l = k, m + l = n)
Random walk: \( P(k, j) = P(X_j = k) \)

\[
P(k, j) = P(X_j = k) = pP(X_{j-1} = k - 1) + (1 - p)P(X_{j-1} = k + 1)
= pP(k - 1, j - 1) + (1 - p)P(k + 1, j - 1)
\]

For \( p = \frac{1}{2} \) (symmetric) we rewrite this as:

\[
P(k, j) - P(k, j - 1) = \frac{1}{2} [P(k - 1, j - 1) + P(k + 1, j - 1) - 2P(k, j - 1)]
\]

It doesn’t matter what the grid size is, so, take time step size \( \Delta t \), space step size \( \Delta X \),

\[
P(X_k, t_j) - P(X_k, t_j - \Delta t) = \frac{1}{2} [P(X_k - \Delta X, t_j - \Delta t) + \n P(X_k + \Delta X, t_j - \Delta t) - 2P(X_k, t_j)]
\]

Suppose

\[
\textbf{Diffusive scaling} \quad (\Delta X)^2 = C\Delta t \quad \text{(not +)} \quad t_{j - 1}
\]

Then the equation for \( P(X_k, t_j) \) becomes (for all \( k \)):

\[
\frac{P(X_k, t_j) - P(X_k, t_j - \Delta t)}{C\Delta t} = \frac{1}{2(\Delta X)^2} [P(X_k - \Delta X, t_j - \Delta t) + \n P(X_k + \Delta X, t_j - \Delta t) - 2P(X_k, t_j)]
\]

\[
\textbf{Diffusion equation: Heat equation} \quad \Rightarrow \quad P_t = \frac{C}{2} P_{XX} \quad \text{as} \quad \Delta t, (\Delta x)^2 \to 0
\]

\textbf{Exercise: Taylor expansion around Delta t=0 and (Delta X)^2=0}
The solution is a mean zero Gaussian with variance proportional to $t$! This is the density of Brownian motion.

$$P(X, t) = \frac{1}{\sqrt{\pi Ct}} e^{X^2/(Ct)} \quad \text{for } P(X, 0) = \delta(X)$$

$C/2 = \text{diffusion coefficient}$

So we get Brownian motion under a diffusive scaling for the random walk.

This also illustrates the self-similarity of this diffusion. In particular, if $w(t)$ is a standard Brownian motion, then so is

$$\frac{1}{\sqrt{\alpha}} w(\alpha t), \quad \langle (\frac{1}{\sqrt{\alpha}} w(\alpha t))^2 \rangle = t$$
4 Equation(s) for the probability density:

General continuous time/space

We derived an equation for $P(x, t)$ using conditional probability - that is conditioning on starting at $x_0$ at time $t$, then taking a step up or down.

In general, this is a useful approach to deriving and equation for $P(x, t)$

Chapman-Kolmogorov equation:

$$P(X(t) = y | X(s) = x) = p(y, t, x, s) = \int p(y, t, z, \tau)p(z, \tau, x, s)dz \quad s < \tau < t \quad (55)$$

This is the probability of going from $x$ to $y$ via $z$, “summed” over all possible intermediate $z$. While in general this doesn’t look like a simplification of the problem, it may be, given the particular problem and the choice of $z$.

This general equation is used to derive the PDE for the probability density $p(y, t|x, s)$ (often, $s = 0$) as follows:

We will show that $\frac{\partial p}{\partial t} = L^*p$ where

Fokker-Planck equation \hfill (FPE)

Forward Kolmogorov equation

$$L^*p = \frac{1}{2} \frac{\partial^2}{\partial y^2} \left( b^2(y, t)p(y, t, x, s) \right) - \frac{\partial}{\partial y} \left( a(y, t)p(y, t, x, s) \right) \quad (FPE) \quad (56)$$
where \( d\xi = adt + bdw \), \( p(\xi = y, t|\xi = x, s) = p(y, t, x, s) \)

This is accomplished by showing:

\[
\int g(y) \frac{\partial p}{\partial t} dy = \int g(y) L^* p(y, t, x, s) dy
\]

(57)

for an appropriate “nice” test function \( g \) (weak sense). To do this, we show that

\[
\int \left\{ \int p(z, t + h, y, t) [g(z) - g(y)] dz \right\} p(y, t, x, s) dy
\]

(58)

\[
= \int [p(y, t + h, x, s) - p(y, t, x, s)] g(y) dy \quad \text{(FPI)}
\]

(59)

The RHS of FPI can be approximated with

\[
h \int p_t(y, t, x, s) g(y) dy, \quad h = \Delta t
\]

(60)

The LHS of FPI can be written as:

as \( h \) goes to zero

\[
\int \left\{ \int p(z, t + h, y, t) \left[ g'(y)(z - y) + 1/2g''(y)(z - y)^2 \right] dz \right\} p(y, t, x, s) dy
\]

(61)

Taylor series expansion around \( z = y \)

Borrowing the \( \frac{1}{h} \) from (60), the LHS can be written as

\[
\int \left( \frac{1}{h} E [\xi(t + h) - \xi(t)] g'(y) + \frac{1}{2} g''(y) \frac{1}{h} E \left[ (\xi(t + h) - \xi(t))^2 \right] \right) p(y, t, x, s) dy
\]

(62)
Defining
\[
\frac{1}{h} E [\xi(t + h) - \xi(t)] = a(y, t) \quad \text{as} \quad h \to 0 \quad (63)
\]
\[
\frac{1}{h} E [(\xi(t + h) - \xi(t))^2] = b^2(y, t) \quad \text{as} \quad h \to 0 \quad (64)
\]
which is essentially the statement that the particle displacement is \(a(y, t)h\) with variance \(b^2h\) (thus a diffusion process). Together the RHS and LHS of FPI yields:

\[
\int p_t(y, t, x, s)g(y)dy = \int \left[ a(y, t)g'(y) + \frac{g''(y)}{2} b^2(y, t) \right] p(y, t, x, s)dy \quad (65)
\]

Integrating by parts yields FPE for \(p(y, t, x, s)\) (weak version - integrated with test function \(g(y)dy\)). Moves the derivatives to ap \(\text{and \ } b^{2/2} p\)

To show FPI: Write RHS:

\[
\int p(z, t + h, x, s)g(z)dz - \int p(y, t, x, s)g(y)dy \quad (66)
\]

First term
\[
\int p(z, t + h, x, s)g(z)dz = \int \left( \int p(z, t + h, y, t)p(y, t, x, s) dy \right) g(z) dz
\]
\[
= \int p(y, t, x, s) \left( \int p(z, t + h, y, t)g(z) dz \right) dy
\]
using Chapman-Kolmogorov equation. Second term:

\[
\int p(y, t, x, s) \cdot 1 \cdot g(y) dy = \int p(y, t, x, s) \left( \int p(z, t + h, y, t) dz \right) g(y) dy \\
= \int p(y, t, x, s) \left( \int p(z, t + h, y, t) g(y) dz \right) dy
\]

Subtract first and second to get the LHS of FPI.

So we can find the density by solving a PDE \( \frac{\partial p}{\partial t} = L^* p \).

Computationally, we can find \( p(x, t) \) using methods for solving a PDE. We can also obtain a numerical approximation to \( p(x, t) \) via a simulation of the SDE for \( x \).

This is accomplished by simulating the SDE for \( x \) to time \( t \) using \( N \) (independent) realizations of the SDE. Recording the \( N \) values of \( x(t) \), we construct a histogram from these values, recording the number of realizations that fall into bins of width \( \Delta x \). Then

\[
p(X, t) \Delta x \approx k_i/N \quad \text{(67)}
\]

where \( 0 \leq k_i \leq N \) is the number of realizations of \( x(t) \) that take values in the interval between \( X \) and \( X + \Delta x \). We see examples of these approximations in later applications.
4.1 Exit Time

Another key quantity is Expected Exit Time (Expected Transition Time).

Later, we will see that the equation for mean exit time has the form, $u_t + Lu = -1$ and $u(\xi \in \partial \Omega, t) = 0$, where

$$Lu = au_\xi + \frac{1}{2}b^2 u_{\xi\xi} \text{ where } d\xi = adt + bdw$$

$L$ is the adjoint of $L^*$, where $p_t = L^*p$.

See page 44 and following for exit time problems
**Asymptotic approximations**

In the next sections we cover different types of asymptotic behavior in SDE’s.

Examples:
1. Approximating systems with a large number of discrete events with a continuous process: Related to central limit theorem: normal approximation for a large number $N$ of realizations

2. Small noise asymptotics and behavior in the state space for: boundary layers in state space
   \[ dx = a \, dt + b \, dw \] where $b$ is small - small random fluctuations

3. Systems with multiple time scales: quasi-steady approximations and stochastic averaging

   *Example: Quasi-steady approximations in deterministic dynamics*

   **Michaelis-Menten Model:**

   \[
   S + E \xrightarrow{k_1} C \xrightarrow{k_2} P + E \xrightarrow{k_3}
   \]

   deterministically - nothing random

   $k_j$ are the rates of the reactions
substrate + enzyme $\rightarrow$ complex $\rightarrow$ further reaction to product + enzyme

$$S + E \xrightarrow{k_1} C$$  \hspace{1cm} (68)

$$\dot{C} = k_1 SE - k_3 C - k_2 C$$  \hspace{1cm} (69)

$$\dot{S} = -k_1 SE + k_3 C$$  \hspace{1cm} (70)

$$\dot{E} = -k_1 SE + k_2 C + k_3 C$$  \hspace{1cm} (71)

$$\dot{P} = k_2 C$$  \hspace{1cm} (72)

$E + C = \text{conserved quantity} = E_0$ \hspace{1cm} Take $(C(0) = 0)$ so we can eliminate $E$

$$\dot{C} = k_1 S(E_0 - C) - k_3 C - k_2 C$$  \hspace{1cm} linear stability shows that

$$\dot{S} = -k_1 S(E_0 - C) + k_3 C$$

$$\Rightarrow C = 0, \quad S = 0$$  \hspace{1cm} (73)

Suppose $E_0 \ll 1$ (small amount of enzyme); then also small amount of $C$.

<< "is much less than"

$$C = E_0 c$$  \hspace{1cm} (74)

$$\Rightarrow E_0 \dot{c} = k_1 S(1 - c)E_0 - k_3 E_0 c - k_2 E_0$$  \hspace{1cm} (75)

$$\dot{S} = -k_1 S(1 - c)E_0 + k_3 E_0 c$$  \hspace{1cm} (76)

To see leading order behavior, $T = E_0 t$ \hspace{1cm} (short time)

Leading order = what are the largest terms for $E_0$ small

$$\Rightarrow E_0 c T = k_1 S_1 (1 - c) - k_3 c - k_2 c$$  \hspace{1cm} (77)

$$S_T = -k_1 S(1 - c) + k_3 c$$  \hspace{1cm} (78)
Leading order: set $E_0 \dot{C} = 0$

\[ \Rightarrow c = \frac{k_1 S_1}{k_1 S_1 + k_2 + k_3} \]

like a steady–state, if $S$ was a constant

\[ \dot{S} = -k_1 S(1 - c) + k_3 c \]

Figure 4: Realization M-M process, with $E_0 \ll 1$.

Note: Here $E_0 = 0.2$ - not even that small, but small compared to other parameters
5 Approximations of interacting individuals

The SIR model: Simple epidemiological model

Susceptible population = $S$
Infected population = $I$
Population of recovered individuals $R$

For either probabilistic or deterministic models, we think in terms of rates


transition rate
$S \rightarrow S + 1$ $\mu N$
$S \rightarrow S - 1$ $\beta SI/N + \mu S$
$I \rightarrow I + 1$ $\beta SI/N$
$I \rightarrow I - 1$ $(\gamma + \mu)I$
$R \rightarrow R + 1$ $\gamma I$
$R \rightarrow R - 1$ $\mu R$

$N$ = total population size
$\mu$ = birth/death rate $\mu$
$\gamma$ = recovery rate
$\beta I/N$ = average number of contacts with infectives per susceptible per unit time.

$\mu N$ = the average number of new contacts per person per unit time.

Asymptotic result for large $N = $ population size

$\beta SI/N$ = the probability of infection given contact with an infected individual.
\( N = \) total population size

Stochastic model: the rates (per unit time) are the conditional transition rates of the stochastic (Poisson) process \((S, I, R)\)

Can write this as a Continuous time Markov process: \(\{(S_t, I_t, R_t) : t \in [0, \infty)\}\), state space \(Z_+^3\). \(N = \) expected population size \(N = E[S + I + R]\)

We need to break this down into Poisson events - births, deaths, infections, recoveries

For example: \(P(S_{t+\Delta t} = s + 1 | S_t = s) = \mu N \Delta t + o(\Delta t)\)
A birth process is a Poisson event with probability \(\mu N \Delta t\) for a birth in time interval of length \(\Delta t\)

Other events:
\(P(S_{t+\Delta t} = s + 1 | S_t = s) = \mu N \Delta t + o(\Delta t),\)
\(P(S_{t+\Delta t} = s - 1 | S_t = s) = \mu S \Delta t + o(\Delta t)\) decrease in susceptibles due to death
\(P(S_{t+\Delta t} = s - 1 | S_t = s) = \beta SI/N \Delta t + o(\Delta t)\) decrease in susceptibles due to infections
\(P(I_{t+\Delta t} = i + 1 | I_t = i) = \beta SI/N \Delta t + o(\Delta t),\) increase in infectives (only by infections)

\textit{Approximations}
Figure 5: A realization of a stochastic SIR model and its deterministic counterpart for \( N = 2000000 \), \( \mu = 1/55 \), \( R_0 = 15 \) and \( \gamma = 20 \).

Deterministic (mean field) model

\[
\begin{align*}
\frac{dS}{dt} &= \mu(N - S) - \beta \frac{SI}{N}, \\
\frac{dI}{dt} &= \beta \frac{SI}{N} - (\gamma + \mu)I,
\end{align*}
\]  

(82)

Note: \( R \) equation is redundant, \( S + I + R = N \).

Basic reproductive number, \( R_0 = \frac{\beta}{\mu + \gamma} \)

If \( R_0 > 1 \) there is a stable endemic \((I \neq 0)\) equilibrium. (ODE exercise)

We express the stochastic equations of the process in a form easily compared with the equations (82) of the deterministic model.

Approximation of the stochastic process with a continuous time, continuous state space:
Typically for larger populations, with short time increments, \( N \gg 1 \)

Recall diffusion processes - approximating a random walk with a Brownian motion for \((\Delta x)^2 = \Delta t\).

Ultimately, approximate Poisson process increments with Brownian motion increments earlier - centered Poisson increments (subtracted off the mean of the original Poisson process, they have mean zero and variance proportional to \( \Delta t \)).

Viewing our Poisson processes as increments: To each increment we add and subtract its conditional expectation, conditioned on the value of the process at the beginning of the time increment of length \( \Delta t \). Each increment of the process is then the sum of the expected value of the increment of the process and the centered increment,

For a Poisson process \( U \), composed of two independent types of events with probability \( P_+, P_- \) for increasing or decreasing by one, we have

\[
P(U_{t+\Delta t} = u + 1|U_t = u) = P_+ \Delta t + o(\Delta t)
\]
\[
P(U_{t+\Delta t} = u - 1|U_t = u) = P_- \Delta t + o(\Delta t)
\]

Then the equation for an increment of \( U \) is: average change + Poisson increment

\[
\Delta U = (P_+ - P_-) \Delta t + Z_+ + Z_-,
\]
\[
Z_{+/-} = U^{+/-} - E[U^{+/-}]
\]

\( Z_+ Z_- \) are centered Poisson increments, that is \( E[Z_{+/-}] = 0 \) with variances \( P_{+/-} \Delta t \)
Note: Can simplify, for two processes independent: \[ \Delta U = (P_+ - P_-) \Delta t + Z \]

\( Z \) is a centered Poisson increment, that is \( E[Z] = 0 \) with variance that is the sum of the two variances \( (P_+ + P_-) \Delta t \)

Recall: \( \sigma dW \) has distribution \( N(0, \sigma \sqrt{\Delta t}) \), so \( Z \) has zero mean and standard deviation proportional to \( \sqrt{\Delta t} \), as does the standard Brownian motion.

When is a Poisson random variable well-approximated by a Gaussian random variable? Typically when there are enough events so that the Poisson parameter is large (e.g. > 10)

So we can write:

\[ \Delta S = (\mu(N - S) - \beta \frac{SI}{N}) \Delta t + \Delta Z_1 - \Delta Z_2; \]

\[ \Delta I = (\beta \frac{SI}{N} - (\gamma + \mu) I) \Delta t + \Delta Z_2 - \Delta Z_3. \]

The increments \( \Delta Z_1, \Delta Z_2, \Delta Z_3 \) are independent, centered Poisson variables with variances \( \mu(N + S) \Delta t \), \( \beta \frac{SI}{N} \Delta t \) and \( (\gamma + \mu) I \Delta t \), respectively.

Note: \( \Delta Z_2 \) is the increment of the process that influences both \( S \) and \( I \) (infection), so it must appear in both equations. Thus the noise in the equations is correlated.

We replace \( \Delta Z_i \) by increments of Brownian motion, \( dW_i \), with the same standard devia-
Diffusion approximation for the stochastic equations

dS = \left( \mu(N - S) - \frac{\beta}{N}SI \right) dt + G_1 dW_1 - G_2 dW_2,

dI = \left( \frac{\beta}{N}SI - (\gamma + \mu)I \right) dt + G_2 dW_2 - G_3 dW_3, \quad (84)

G_1 = \sqrt{\mu(N + S)}, \quad G_2 = \sqrt{\frac{\beta}{N}SI}, \quad G_3 = \sqrt{\gamma + \mu}.

Note - these have the Ito interpretation.
Note - some noise coefficients may be large or small? If the noise is small, why not just use the deterministic equations (82)

References:

\begin{align*}
    dX &= f(t, X) + G(t, X)dW \\
    dY &= f(t, Y) + B(t, Y)dV
\end{align*}

X, Y are d-dimensional, W is m-dimensional, V is d-dimensional, m > d. For GG^T = H and B = H^{1/2}, for f, G, B satisfying certain continuity conditions, giving pathwise unique solutions. Then X and Y have the same distribution. (weak sense)

So we can express the noise terms in different forms, can choose for our convenience. In the example above, it is easy to derive, easy to see influence of different biological processes.
in the equations.

*Approximating Poisson with Normal*

Expect $N$ is relatively large (larger rates for Poisson processes in a time interval)

Are stochastic terms significant? i.e. if $N$ is large, are the fluctuations significant?

$u$ is proportion of $N$ that is in S, $v$ is proportion of $N$ in I

Consider rescaled system $S = Nu, I = N v$

$u, v$ are between zero and 1

$$
du = \left( \mu(1 - u) - \frac{\beta}{u} v \right) dt + g_1 dW_1 - g_2 dW_2,
$$

$$
dv = \left( \frac{\beta}{N} SI - (\gamma + \mu) I \right) dt + g_2 dW_2 - g_3 dW_3, \tag{85}
$$

$$
g_1 = \sqrt{\frac{\mu(1 + u)}{N}}, \quad g_2 = \sqrt{\frac{\beta}{N} uv}, \quad g_3 = \sqrt{(\gamma + \mu) v/N}.
$$

Note, coefficients $g_j$ of noise term scale as $N^{-1/2}$, vanish as $N \to \infty$. Then we recover the mean field (deterministic) model (82).

What if $N$ is just large, rather than infinite?

We consider the system with the basic reproductive number, $R_0 = \frac{\beta}{\mu + \gamma} > 1$ Then the
A deterministic system has a unique nontrivial stable equilibrium point \((S_{eq}, I_{eq})\) at

\[
S_{eq} = \frac{N}{R_0}, \quad I_{eq} = \frac{N\mu}{\beta}(R_0 - 1).
\]

(86)

Consider simulations for different parameter values: Why is the noise sometimes significant?

We introduce the dimensionless variables

\[
u = \frac{S - S_{eq}}{S_{eq}}, \quad v = \frac{I - I_{eq}}{I_{eq}},
\]

Note \(S_{eq}\) and \(I_{eq}\) both proportional to \(N\)

\[t \to \Omega t, \quad \Omega = \sqrt{\frac{\beta\mu}{R_0}(R_0 - 1)} \quad \text{(for convenience)}\]

to get the equations (exercise)

\[
du = \frac{1}{\Omega} [(-\mu - \frac{\beta I_{eq}}{N})u - \frac{\beta I_{eq}}{N}v - \frac{\beta I_{eq}}{N}uv]dt + \sqrt{\frac{\mu}{\Omega S_{eq}^2} (N + S_{eq}(u + 1))dW_1(t)} - \sqrt{\frac{\beta I_{eq}}{\Omega NS_{eq}} (v + 1)(u + 1)dW_2(t)},
\]

\[
dv = \frac{1}{\Omega} \left[\frac{\beta S_{eq}}{N}(u + v) + \frac{\beta S_{eq}}{N}uv - (\gamma + \mu)v\right]dt + \sqrt{\frac{\beta S_{eq}}{\Omega NI_{eq}} (v + 1)(u + 1)dW_2(t)} - \sqrt{\frac{\gamma + \mu}{\Omega I_{eq}} (v + 1)dW_3(t)}.
\]

(87)

Noise coefficients: still \(N^{-1/2}\)
Consider equations linearized about $u = v = 0$ (near equilibrium)

$u, v$ are fluctuations around the equilibrium
Figure 7: Stationary density $p(I)$ (2000 realizations for a value of $t > 100$) Left: (a) (*'s and dash-dotted line) and (f) (diamonds and solid line) Right, (c) (*'s and dash-dotted line), (d) (dotted and solid line)

linear approximation around $u=v=0$ (Ornstein-Uhlenbeck process as an approximation)

$$
\begin{align*}
\mathbf{d} \begin{pmatrix} u \\ v \end{pmatrix} &= \mathbf{M} \begin{pmatrix} u \\ v \end{pmatrix} dt + \mathbf{G} \begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \end{pmatrix}, \\
\mathbf{M} &= \begin{pmatrix} -\frac{\mu R_0}{\Omega} & -\mu \frac{R_0-1}{\Omega} \\ \frac{\beta}{\Omega R_0} & 0 \end{pmatrix}, \\
\mathbf{G} &= \begin{pmatrix} \sqrt{\frac{\mu}{\Omega S_{eq}} (N + S_{eq})} & -\sqrt{\frac{\beta I_{eq}}{\Omega N S_{eq}}} & 0 \\ 0 & \sqrt{\frac{\beta S_{eq}}{\Omega N I_{eq}}} & -\sqrt{\frac{\gamma + \mu}{\Omega I_{eq}}} \end{pmatrix} = \begin{pmatrix} g_1 & -b^2 g_2 & 0 \\ 0 & g_2 & -g_2 \end{pmatrix}.
\end{align*}
$$

43
Solutions of the deterministic version of (88) are given in terms of the eigenvalues of \( \mathbf{M} \), which are

\[
\lambda = -\epsilon^2 \pm \sqrt{\epsilon^4 - 1},
\]

where

\[
\epsilon^2 = \frac{\mu R_0}{2\Omega}.
\]

\[
\begin{pmatrix}
  u \\
  v
\end{pmatrix}
\sim C_1 e^{-\epsilon^2 t} \begin{pmatrix} b \cos t \\ \sin t \end{pmatrix} + C_2 e^{-\epsilon^2 t} \begin{pmatrix} b \sin t \\ -\cos t \end{pmatrix},
\]

for certain values of rate constants and \( N \) large, epsilon is small

6. Lambda is complex with small negative real part

dropping \( O(\epsilon^2) \) corrections.

For \( N \) large and finite - drift and diffusion coefficients both small Oscillations have weak (slow) decay, noise coefficients may play a role

6. Mean exit time

Another key quantity is Expected Exit Time (Expected Transition Time).

If we define \( \tau_x = \) time to reach certain state, \( \partial \Omega \), given initial state \( \xi(s) = x \in \Omega \),

\[
E[\tau_x] = \text{Expected time to exit } \Omega \text{ and reach new state (} \partial \Omega \text{) given } \xi(s) = x.
\]
We can write this as
\[ E[\tau_x] = \int_0^\infty \int_\Omega p(y, t, x, s) dy dt = \int_0^\infty P(\tau_x > t) dt \] (90)
since \( P(\tau_x < t) \) is the probability that \( y \in \Omega \) at time \( t \). Can also compute this, noting that for \( u(\xi, t) \) with \( d\xi = adt + bw \),
\[ du(\xi(t), t) = \left( \frac{\partial u}{\partial t} + \frac{\partial u}{\partial \xi} a + \frac{1}{2} b^2 \frac{\partial^2 u}{\partial \xi^2} \right) dt + \frac{\partial u}{\partial \xi} bdw \] (91)
using Ito’s formula, so that
\[ u(\xi(t), t) = u(\xi(s), s) + \int_s^t [u_t + Lu] dt + \int_s^t \frac{\partial u}{\partial \xi} dw \] (92)
Note \( Lu = au_\xi + \frac{1}{2} b^2 u_{\xi\xi} \), is the adjoint operator for \( L^* u = -(au)_\xi + \frac{1}{2} (bu)_{\xi\xi} \).

Suppose, \( u_t + Lu = -1 \) and \( u(\xi \in \partial \Omega, t) = 0 \). Then substitute \( t = \tau_x \) above and take expected values,
\[ E[u(\xi(\tau_x), \tau_x)] = u(x, s) + E \left[ \int_0^{\tau_x} (-1) dt \right] + 0 \] (93)
\[ \Rightarrow 0 = u(x, s) - E[\tau_x] \] (94)

Example: Exit time of a particle from potential \( U(x) \).
Simple example: \( U(x) = \frac{x^2}{2} \), on \(-1 < x < 1\)

Particle position: \( d\xi = -\xi dt + \sqrt{2\epsilon}dw = -U'(\xi)dt + \sqrt{2\epsilon}dw \)

\[ \text{Realizations of } \xi \text{ vs } t \]

- \( \epsilon \) small
- \( \epsilon \) smaller
- \( \epsilon \) even smaller

![Realizations of \( \xi \) vs \( t \)](image)

Figure 8: Realizations for \( \epsilon = .05, .04, .03 \). The uniform approximation for \( q \), compared to the potential \( U(x) \)

Note: this is a time homogeneous case \( \Rightarrow \)
\[ \epsilon u'' - xu' = -1 \quad \text{for } u= \text{ mean exit time from interval } [-1, 1] \]

\[ u(-1) = u(1) = 0 \]

This is an example which we could solve exactly, but we see how the asymptotic solution gives physical insight: Consider for \( \epsilon \ll 1 \)

For small \( \epsilon \) (small noise), expect \( u \to \infty \) as \( \epsilon \to 0 \) (at least for \( x \neq \pm1 \))

**brief introduction to boundary layer theory**
so take \( u = C'(\epsilon)q(x) \) for \( C'(\epsilon) \to \infty \) as \( \epsilon \to 0 \)

Then, substituting into equation for \( u \),

\[
\epsilon q'' - xq' \sim 0 \quad q(-1) = q(1) = 0
\]

(95)

For \( \epsilon = 0 \), we have, to leading order: \( \Rightarrow \, q_0 = \text{constant} \quad \text{constant} \neq 0 \)

However, the boundary conditions are not satisfied, so we must find the boundary layer behaviour. The solution \( q_0 \) is the “outer” solution (away from the boundaries).

look near the boundary, and expand that region

Then, taking \( \epsilon z = 1 - x \), as a local variable near \( x = 1 \), yields

\[
\frac{\epsilon}{\epsilon^2} q_{zz} - \frac{1 - \epsilon z}{\epsilon} q_z = 0 \quad , \quad q \sim q_0 + \epsilon q_1 + \ldots
\]

(96)

\[
\Rightarrow \, q_{0zz} - q_{0z} = 0 \quad , \quad q_0(z = 0) = 0
\]

(97)

\[
q_0 = (1 - e^{-z}) = \left(1 - e^{-\frac{(1-x)}{\epsilon}}\right)
\]

(98)

Similarly for \( x \) near \(-1\).

Then we can construct the uniform solution:

\[
q \sim 1 \quad \text{away from } x=1 \text{ or } x=-1
\]

\[
q \sim 1 - e^{-\frac{(1-x)}{\epsilon}} \quad \text{near } x = 1
\]

\[
q \sim 1 - e^{-\frac{(x+1)}{\epsilon}} \quad \text{near } x = -1
\]
Boundary layer approximations → constant, as matched to “outer” approximation

Uniform approximation: Add boundary layer expansions + outer approximation - parts matched

\[ u = C'(\epsilon) \left( 1 - e^{-\frac{(1-x)}{\epsilon}} - e^{-\frac{(x+1)}{\epsilon}} \right) \]  \hspace{1cm} (99)

Still, we don’t know \( C'(\epsilon) \), but can it determine using our knowledge of \( p(x) \) (invariant density). That is, \( p \) satisfies equation with the adjoint operator: \( L^*p = 0 \).

We have the equation for \( u \)

\[ Lu = -1 \]  \hspace{1cm} (100)

and

\[ L^*p = 0 = \epsilon p_{xx} + (xp)_x \Rightarrow p(x) = \frac{1}{\sqrt{2\pi\epsilon}} e^{-\frac{x^2}{2\epsilon}} \]  \hspace{1cm} (101)

So,

\[ \int pLu \, dx = p\epsilon u'|_1^0 + \int (L^*p)u \, dx = -\int_{-1}^1 1 \cdot p \, dx \]  \hspace{1cm} (102)

If we use the uniform expansion for \( u \) we can evaluate (exercise), then we get the equation for \( C'(\epsilon) \) from the calculation of \( p(x) \)

\[ C'(\epsilon) \sim K \sqrt{2\pi\epsilon} e^{-\frac{U^{(1)}}{\epsilon}} + \text{higher order corrections} \]  \hspace{1cm} (103)

Additional exercise: Could there be a longer time to escape from the bottom of the potential? Show that there is no internal layer in the construction of the uniform solution.
Use local variable $v = \epsilon^{\alpha}x$, find appropriate $\alpha$ and equation for $q(v)$. A contradiction in matching to the outer solution results, which shows no layer at $x = 0$.

**Asymptotic methods for determining $p(x, t)$ on whole domain**

Another asymptotic method that is commonly used to determine $p(x, t)$ in the small noise case is the **WKB method**.

**another asymptotic method for solving PDE's**

Let’s see how this works for the simple example above. The equation for $p(x, t)$ is

$$p_t = (x p)_x + \epsilon p_{xx} = L^* p$$ (104)

The WKB method is based on the observation that is we take $\epsilon \to 0$ in the equation for $p$, we eliminate the highest order derivative. This doesn’t make sense, as it eliminates the diffusion from the equation.

So, instead we should look for a form of the solution that better balances the effect of the diffusion with the other contributions.

$$p(x, t) = K(x, t) \exp(-\psi/\epsilon)$$ (105)

This makes intuitive sense, particularly if we are expecting densities that decay exponentially for large values.

Substituting into the equation for $p(x, t)$, we get:

$$K_t - \frac{1}{\epsilon} K \psi_t = x \left[ K_x - \frac{\psi_x}{\epsilon} K \right] + K + \epsilon \left[ -\frac{\psi_{xx}}{\epsilon} K + \frac{(\psi_x)^2}{\epsilon^2} - 2K \frac{\psi_x}{\epsilon} + K_{xx} \right]$$ (106)
For simplicity, we consider the case of the invariant density - no $t$ dependence.

Then, to leading order (terms with coefficient of $\epsilon^{-1}$) we have

$$-x\psi_x + \psi_x^2 = 0$$

(107)

from which we obtain $\psi = \text{constant}$ or $\psi = x^2/2$. The constant solution is the trivial solution, and is also not normalizable for $p$ density.

(condition that $p$ must be a probability density, has to be normalizable)

Then, the equation for $K$ (terms with coefficient $\epsilon^0$) is

$$(1 - \psi_{xx})K + (x - 2\psi_x)K_x + K_{xx} = -xK_x + K_{xx} = 0$$

(108)

The solution for $K$ that is normalizable, i.e. $\int K(x)e^{-\psi(x)/\epsilon}dx$ is $K = \text{constant}$. Then we obtain the result given above in (101).
Stochastic systems with multiple time scales:

\[
\frac{dx}{dt} = f(x, y) \\
\frac{dy}{dt} = \epsilon^{-1} g(x, y)
\]

for epsilon small:

Can view time scale of \( y \) as “fast”, i.e. for \( T = t/\epsilon \), \( y'(T) = g(x, y) \).

sometimes called "separation of time scales" for small epsilon

“Slow” time scale \( t \) and “Fast” time scale \( T \)

If fluctuations in \( y \) are fast, then they may appear as fast random fluctuations in the equation for \( x \). Then it may be possible to approximate \( x \) over the slow time scale of \( t \) with

Here fluctuations in \( y \) replaced by some kind of average

\[
\frac{dX}{dt} = \hat{f}(X, \zeta)
\]  

(109)

where \( \zeta \) is a quantity used to approximation the behavior of \( y \). Then \( X \) may approximate the behavior of \( x \) in the weak sense, e.g. in terms of moments or distribution.

How do we determine a reasonable approximation? Under what conditions does this approximation hold?

References:
Mathematical (and applications): Stuart and Pavliotis, Multiscale methods

We show here results for the following system of SDE’s:

\[
\begin{align*}
    dx &= f(x, y) \, dt \\
    dy &= \epsilon^{-1} g(x, y) \, dt + \epsilon^{-1/2} \sigma \, dW
\end{align*}
\]  

(110)

Specific example: \((x, y)\) is an Ornstein-Uhlenbeck process for this specific example

\[
\begin{align*}
    f(x, y) &= -x - ay \\
    g(x, y) &= cx - y
\end{align*}
\]  

(111)

Deterministic approximation: Averaged approximation = (A) approximation

Assuming \(y\) is “well-mixed”: that is, on the \(t\) time scale, the density of \(y\) is well-sampled, approaching the stationary density

Then the approximation is based on using the stationary conditional density, \(p(y|x)\). That is, on the fast time scale on which \(y\) fluctuates, \(x\) appears as a constant. Then the proposed (A) approximation is

\[
\frac{d\bar{x}}{dt} = \bar{f}(\bar{x}) = \int f(x, y)p(y|x) \, dy = E_{y|x}[f(x, y)]
\]  

(112)
For our example, what is \( \bar{f} \)?

Need to determine what \( p(\mathbf{y} | \mathbf{x}) \) is

If \( x \) is treated as a constant in the equation for \( y \), then \( y \) is an Ornstein-Uhlenbeck (OU) process, with a Gaussian stationary density \( N(cx, \sigma^2/2) \). So

\[
E_{y|\mathbf{x}}[f(x, y)] = -x - E[y|\mathbf{x}] = -(1 + ac)x \quad \Rightarrow \quad \frac{d\bar{x}}{dt} = -(1 + ac)\bar{x} \quad (113)
\]

What are the circumstances for which (A) is a reasonable approximation? Can we do better? Consider our example: (neglecting stochastic fluctuations, just averaging)

Taking \( x = \bar{x} + \xi \), assuming \( \xi \ll 1 \) is a small correction

substitute \( x = x\text{-bar} + xi \)

\[
\begin{aligned}
\frac{d\bar{x}}{dt} + \frac{d\xi}{dt} &= f(x, y) - \bar{f}(x) + \bar{f}(x) \quad \Rightarrow \\
\frac{d\xi}{dt} &\sim \left( f(x, y) - \bar{f}(x) \right) + \bar{f}(\bar{x}) + \bar{f}'(\bar{x})\xi - \bar{f}(\bar{x}) \\
\frac{d\xi}{dt} &\sim \bar{f}'(\bar{x})\xi + \hat{f}(x, y) \quad \text{neglect terms of higher powers of } \xi
\end{aligned}
\]

where we have used a Taylor series about \( \xi = 0 \), and \( \hat{f}(x, y) = (f(x, y) - \bar{f}(x)) \).

Then, to complete the approximation, we need to approximate \( \hat{f}(x, y) \). This term represents the fluctuations of \( f \) about its average, due to fast fluctuations in \( y \). As we look for a weak approximation to \( x \) on the time scale \( t \) (slow time scale), then we look for an
approximation to \( \hat{f} \) which has the same properties - in the weak sense.

Specifically, can \( \hat{f} \) be approximated by Brownian motion, or some other random process?

What is the motivation for this approximation?

We expect the properties of \( \hat{f} \) gives the fast fluctuations of \( f(x, y) \) about its average \( \bar{f} \). So if varying on a fast time scale, these can appear as fluctuations of a random process with mean zero. So if we can approximate with something known, then we have a lower dimensional approximation for \( \xi \) and thus for \( x \sim X = \bar{x} + \xi \).

If \( \hat{f} \) has the properties of a Brownian motion \( DdW \), then we consider the integrated behavior of the correlation \( C \):

\[
\int C_{\hat{f}\hat{f}}(\tau) d\tau = \int E[\hat{f}(x, y(t+\tau))\hat{f}(x, y(t))]d\tau = D^2 \int \delta(\tau)d\tau \quad (114)
\]

Here \( \tau \) is the lag variable. Usually \( C(\tau) \to 0 \) as \( \tau \to \infty \). That is, correlation between different points in a realization decreases with increasing time intervals between the points.

Note: for simplicity we take the stationary result for \( C_{\hat{f}\hat{f}} \) - that is, as \( \epsilon \to 0 \), on the fast time scale \( t/\epsilon \to \infty \). So \( C \) is independent of \( t \).

This result can be expressed more generally in the case that the behavior is not identically a Brownian motion, but rather one that in the limit as \( \epsilon \to 0 \) (time scales are separated) has the behavior of a \( \delta \)-function.
\[
\int C_{\hat{f}\hat{f}}(\tau)\,d\tau = \beta \int \frac{1}{\epsilon} h(u/\epsilon)du \quad \text{where}
\]
\[
\lim_{\epsilon \to 0} \int \frac{1}{\epsilon} h(u/\epsilon)du = \int \delta(u)du = 1 \tag{115}
\]

For example, if \( \hat{f} \) has the properties of an Ornstein-Uhlenbeck process, on the fast scale, e.g.

\[
dZ = -\frac{\mu}{\epsilon}Z\,dt + \frac{\Sigma}{\sqrt{\epsilon}}\,dW(t) \tag{116}
\]

Then \( C_{ZZ}(\tau) = \epsilon\Sigma^2/(2\mu)\frac{1}{\epsilon}e^{-\mu|\tau|/\epsilon} \), and

\[
\lim_{\epsilon \to 0} \int \int C_{ZZ}(\tau)d\tau = \epsilon \frac{\Sigma^2}{\mu^2} \tag{117}
\]

Now let’s see how this works for our example:

Recall, \( \bar{f}(x) = -(1 + ac)x \)

Then \( \hat{f} = f(x, y) - \bar{f}(x) = -x - ay + x + acx = -ay + cx \)

**condition on** \( x = \text{const. for x slow, y fast} \)

and \( C_{\hat{f}\hat{f}} = a^2 E_{y|x}[\{(y(t) - cx)(y(t + \tau) - cx)\}] = a^2 C_{yy}(\tau) \). Considering the stationary behavior of \( y \) (for large time), we have
\[ \int C_{\hat{f}\hat{f}}(\tau) d\tau = \epsilon^2 a^2 \sigma^2 \int_0^\infty \frac{1}{\epsilon} e^{(-|\tau|/\epsilon)} d\tau = a^2 \sigma^2 \epsilon \]

Then we conclude that \( \hat{f} \) can be weakly approximated by increments of a Brownian motion \( \epsilon ab dW(t) \), and the slow dynamics for \( x \) in

\[
\begin{align*}
    dx &= -x - ay \\
    dy &= -\frac{cx + y}{\epsilon} + \frac{\sigma}{\sqrt{\epsilon}} dW(t)
\end{align*}
\]

(118)

can be weakly approximated as \( x = \overline{x} + \xi \) where

\[
\begin{align*}
    d\overline{x} &= -(1 - ac) \overline{x} \\
    d\xi &= \hat{f}'(x) \xi + \sqrt{\epsilon} a \sigma dW(t) = -(1 - ac) \xi + \sqrt{\epsilon} a \sigma dW(t)
\end{align*}
\]

(119)

This is known as the (L) approximation. It is a linear SDE approximation for the slow dynamics \( x \). We can see that for certain values of the parameters \( a \) and \( b \), namely where \( ab = \epsilon^\gamma \), for \( \gamma > -\frac{1}{2} \) the stochastic fluctuations are negligible as \( \epsilon \to 0 \). Then \( \xi \) decays to its mean 0, and we are left with the (A) approximation.

This form is relatively simple, given that the original system is linear. Furthermore, terms of the form \( \hat{f}'(\overline{x}) \xi^2 / 2 \) vanish. But in a more general nonlinear system, these terms would not vanish and could be a source of error in the approximation.

*General multiple scale averaging approximation* (nonlinear):
The (N+) approximation:

$$dx = \overline{f}(x) dt + D(x) \circ dW \quad \text{(Stratonovich)} \quad (120)$$

$$= \overline{f}(x) dt + G(x) + D dW \quad \text{(Ito)} \quad (121)$$

Note that \( \circ \) indicates the Stratonovich interpretation for the noise increment, and \( G(x) = D(x)D'(x)/2 \) is the correction between the Ito and Stratonovich interpretations.

Derivation of this result:
Reference: Stuart and Pavliotis, Multiscale Methods, Averaging and Homogenization (Springer)

The approach is based on looking at the operator \( L \) corresponding to the generator for

Main ingredients: PDE theory - operator \( L \) associated with the SDE : generator for the process
the full process. Recall, for $u = E(f(x, y))$,

$$
\begin{align*}
\frac{\partial u}{\partial t} &= Lu = a \cdot \nabla u + B \nabla^2 u, \quad B = bb^T/2 \\
\frac{dz}{dt} &= a(z) \, dt + b(z) \, dW
\end{align*}
$$

and then looking at the asymptotic behavior of that operator as $\epsilon \to 0$, and identifying the process that corresponds to the asymptotic approximation to $L$.

Example:

$$
\begin{align*}
\frac{dx}{dt} &= \frac{v(x)y}{\sqrt{\epsilon}} \\
\frac{dy}{dt} &= -\frac{\alpha y}{\epsilon} + \frac{\sqrt{2\alpha}}{\sqrt{\epsilon}} \, dW(t) 
\end{align*}
$$

The equation for $u$ has the form:

$$
\frac{\partial u}{\partial t} = Lu = \left( L_0 + \frac{1}{\sqrt{\epsilon}} L_1 + \frac{1}{\epsilon} L_2 \right)
$$

Using an asymptotic expansion for $u = u_0 + \sqrt{\epsilon} u_1 + \epsilon u_2$, then one can consider the sequence of equations obtained at each order of $\epsilon$.

different orders correspond to different powers of epsilon

After a number of calculations, the equation has the form:

$$
L_0 u_2 = M(u_0)
$$
From the Fredholm Alternative Theorem, for there to be a solution we must have

\[ \int M(u_0)\rho(y) \, dy = 0 \quad \text{since} \quad L_0^*\rho(y) = 0 \]  \hspace{1cm} (126)

where \( \rho(y) \) is the invariant density for the fast variable \( y \). Note that this condition is essentially an averaging over \( y \). From that condition we find

\[ \int M(u_0)\rho(y) \, dy = L_x u_0 \]  \hspace{1cm} (127)

where \( L_x \) is the operator corresponding to the reduced equation in (120), which is the N+ approximation.

Other recent results: Stochastic averaging for systems forces by Levy (alpha stable processes), and other noise processes with "fat tails"

(Thompson, Kuske, and Monahan, arXiv)

instead of Stratonovich, we have Marcus interpretation of stochastic increments
More research on interactions of noise and nonlinear dynamics, stochastic bifurcations, noise sensitivity - in applications: physics, (bio)-mechanics, neuroscience, epidemiology, pattern dynamics, environment

http://www.math.ubc.ca/~rachel

Recruiting PhD students and postdoctoral researchers in 2016:

http://www.math.ubc.ca: Information on PhD programs

Applications for postdocs: Math Jobs (by November 2016)