Diffusive and Stochastic Processes: Lecture notes

Namiko Mitarai

April 13, 2023

Preface

These lecture notes are made for the "Diffusive and Stochastic Processes" course. The course aims to understand physical systems as a stochastic process and learn the power of stochastic descriptions, as well as learn the analytical and numerical tools to study the stochastic systems so that you can use them for your research in the future.

The lecture notes start by analyzing **diffusive processes**, which is the most classical example of stochastic process. Here we learn the physics of diffusive systems, and at the same time, we learn the basic concepts and tools to deal with stochastic processes. We then extend our horizon to more general **stochastic processes**, where we learn the application of the stochastic description to the systems far from equilibrium. I also pay attention to making the connection between different topics clear. The overview is given in the figure below.

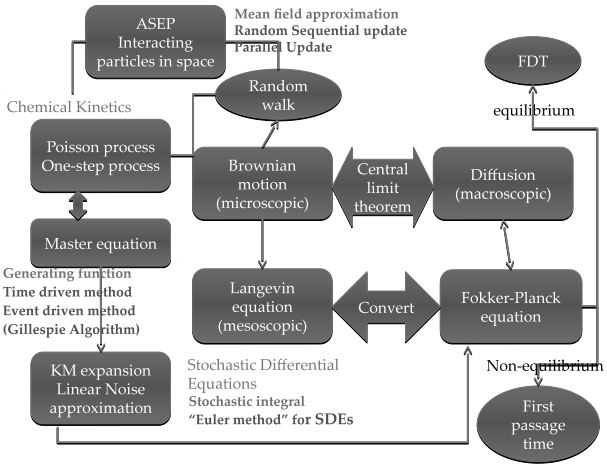
It is very important to solve exercises. Exercises are needed for you to be able to follow the lectures since things will be built upon them. It is mandatory to solve basic exercises, while "advanced exercises" are optional but useful to bring your understanding to the level competent for research.

Since this course aims to make it possible for you to use what you learn to your problems, you need to be able to *derive* relations. It is also necessary that you understand the relationship between mathematics and real phenomena. I tried to use a wide range of phenomena as examples, though it is limited to the systems I know. Computational exercises (marked by *) are aimed to give you numerical abilities as well as to "visualize" analytical calculations to help intuitive understanding. Also, note that many models used in the research frontier are not analytically solvable. Hence, the ability to simulate stochastic processes is often necessary to use what you learn for research projects. Within this course's exercise, I recommend using Python. I also use FORTRAN, golang, Matlab, and I can read c and c++ to some extent. At the start of the course, I will ask which programming language you want to use. It is strongly encouraged to submit your result of computational exercises through Absalon so that you can get feedback. For most of the exercises, you get feedback only on the resulting plots, but for the exercises marked by ** you can also submit the code for the feedback.

The curriculum (pensum) is the lecture notes and exercise materials. Exceptions can be specified during the course. 20% of the grades will be evaluated based on programming assignments specified later in the class. The final exam that covers 80% of the grade will be an oral exam.

I started this course in 2011, but it is under continuous development. Any of your feedback is welcome. Please let me know if you find any mistakes or misleading typos in the lecture note.

I hope you enjoy the course and learn something from it. Welcome to the course!



Contents

Chapter 1

Introduction: Brownian motion and diffusion

1.1 Brownian motion and diffusion

1.1.1 Brownian motion as a physical phenomenon

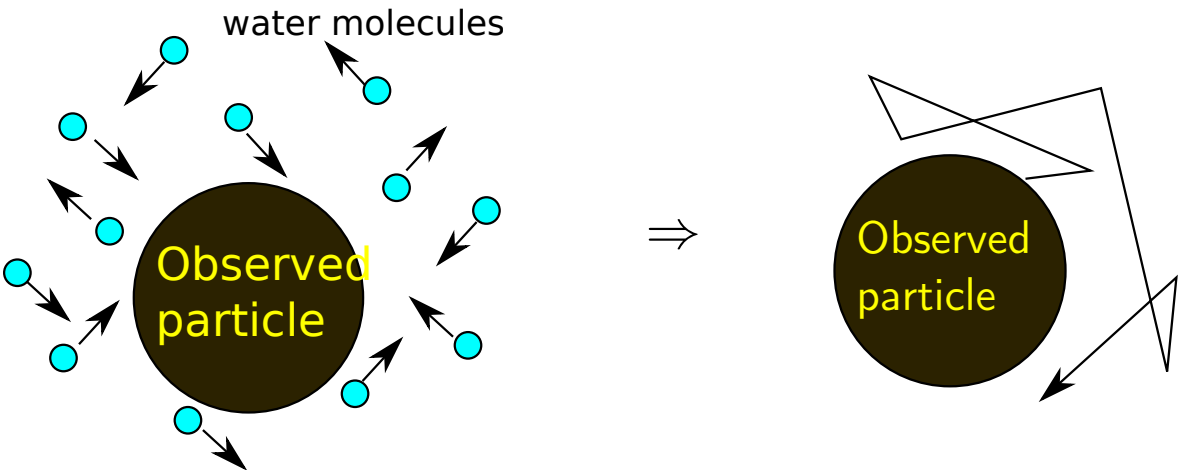


Figure 1.1: Schematic description of Brownian motion.

Brownian motion is the random motion of particles suspended in a fluid. The origin of the motion is the thermal fluctuation of molecules that consist of fluid, e.g. water molecules. They collide with the suspended particles, and if the suspended particles are small (but still a lot larger than water molecules), the collisions move the particles at a visible scale. Typically observed particles are the scale of 1 μm in diameter, visible with microscopes, while fluid molecules are not directly visible (Fig. 1.1).

Macroscopically, the Brownian motion causes **diffusion**, where the particles spread from higher concentration regions to lower concentration regions, driving the system to approach a uniform distribution of the particles (Fig. 1.2).

The Brownian motion (or diffusion) is one of the simplest and the most well-studied stochastic phenomena. Many tools to analyze stochastic processes have been developed by studying them. Therefore, let us start from the diffusive processes to learn the basics of **stochastic processes**.

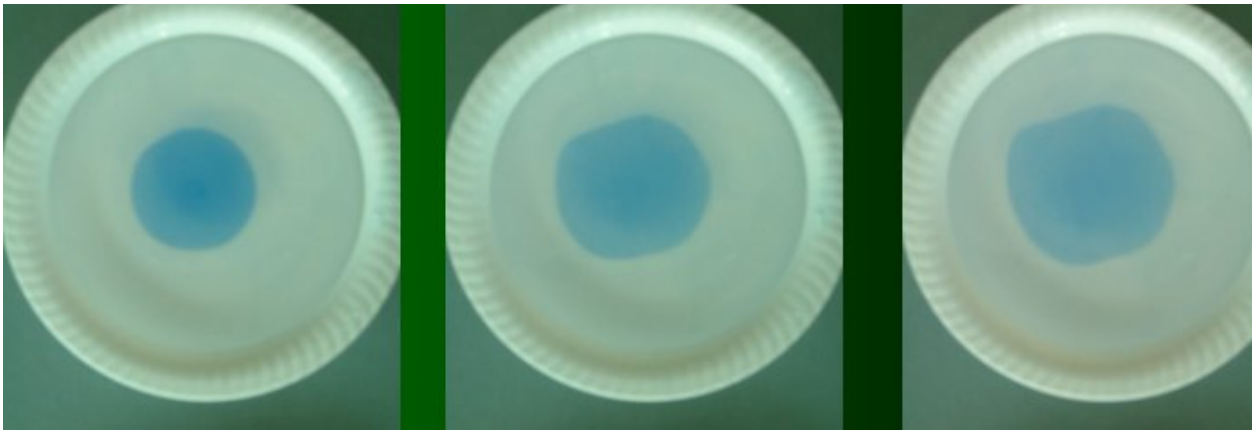


Figure 1.2: Diffusion of ink in water. The time proceeds from left to right.

Fig. 1.1. Brownian motion caused by incessant impacts of gas molecules against a mirror suspended in a dilute gas [1.5]

Figure 1.3: Experimental observation of a Brownian motion trajectory. From ref. [1].

1.1.2 The Brownian motion as a stochastic process

What is a "stochastic process"? Suppose we did an experiment to record the position of a Brownian particle $x(t)$ as a function of time t for $0 \leq t \leq T$ as shown in Fig. 1.3. We do this experiment N times, and get N records

$$x_1(t), x_2(t), \cdots, x_N(t). \tag{1.1}$$

Each $x_i(t)$ are different - each $x_i(t)$ is a sample from a *statistical ensemble*. If we make N bigger and bigger, we should be able to guess the distribution of $x(t)$.

We express this as follows: There is a **stochastic variable** $X(t)$, and at each realization, $X(t)$ takes a **value** $x(t)$ that is drawn from the distribution. A time series of a stochastic variable is a **stochastic process**.

Let us consider the probability that a observed value of $X(t)$ at time t, $x(t)$, takes the value between x and $x + dx$, $Pr(x < X(t) \leq x + dx)$. This defines the probability density $P(x, t)$ as

$$P(x, t)dx = Pr(x < X(t) \leq x + dx). \tag{1.2}$$

Here, Pr(condition) means the probability that the condition in the parentheses are satisfied. Eq. (1.2) assumes that $Pr(x < X(t) \leq x + dx)$ is proportional to dx in $dx \to 0$ limit. Note that $P(x, t)$ is normalized as

$$\int P(x, t)dx = 1, \tag{1.3}$$

where the integration range is over all possible values of x.

Discrete observation If it is possible to observe in continuous time, we get a real function of time $x(t)$ as a sample of the stochastic process $X(t)$. If the observation is done for n discrete time points,

$$0 \leq t_1 < t_2 < \cdots < t_n \leq T, \tag{1.4}$$

then our sample is a set of n real numbers

$$x(t_1), x(t_2), \cdots, x(t_n). \tag{1.5}$$

We can also consider the probability density for n time points:

$$P_n(x_1, t_1; x_2, t_2; \cdots; x_n, t_n)dx_1dx_2\cdots dx_n = Pr(x_j < X(t_j) \leq x_j + dx_j, j = 1, 2, \cdots, n). \tag{1.6}$$

The continuous-time description $x(t)$ can be regarded as the limit where we take infinite n and make the time interval infinitesimally small.

1.2 Diffusion equation: A macroscopic description of Brownian motion

When many Brownian particles are initially distributed spatially inhomogeneously, the distribution becomes uniform through diffusion (Fig. 1.2). In this section, we phenomenologically derive the equation to describe the diffusion of many Brownian particles in one-dimensional space (extension to a higher dimension is straightforward and given in the exercise), assuming that particles are identical and particle-particle interactions are negligible.

In this case, the density distribution $n(x, t)$ is related to the probability distribution $P(x, t)$ as $n(x, t) = NP(x, t)$, where N is the total number of particles in the system. $n(x, t)dx$ denotes the number of particles in the interval of the position $(x, x + dx)$ at time t. Therefore, deriving the equation for $n(x, t)$ is equivalent to deriving the equation for $P(x, t)$.

1.2.1 Phenomenological derivation of diffusion equation

Fick's law

To derive an equation that can describe diffusion, let us summarize the characteristics of the diffusion:

- There is no visible diffusion flux of particles if n is spatially homogeneous.
- Particle moves from a dense to a dilute region on average.

It is then natural to assume that the flux is proportional to the first-order spatial derivative of n as the lowest-order effect. This gives **Fick's law**: The flux of the particles by diffusion j_d is proportional to the density gradient:

$$j_d = -D\frac{\partial n}{\partial x}, \tag{1.7}$$

with a positive constant D.

Equation of continuity

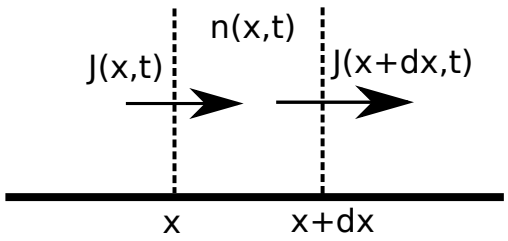


Figure 1.4: Schematic description of continuity.

When the number of particles is conserved (i.e., there is no source or sink), the change of the density in the space between x and $x + dx$ during a small time step dt is only due to the flux J (Fig. 1.4). Namely,

$$n(x, t + dt)dx - n(x, t)dx = J(x, t)dt - J(x + dx, t)dt.$$

By expanding this for small dx and dt, we get the equation of continuity

$$\frac{\partial n}{\partial t} = -\frac{\partial J}{\partial x}. \tag{1.8}$$

Diffusion equation

The density field $n(x, t)$ obeys the equation of continuity because the number of particles is conserved. For diffusion, $J = j_d$ and we have

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial x^2}. \tag{1.9}$$

This is the **diffusion equation**. D is called the **diffusion constant**.

Equation for the probability density

Since $n(x, t) = NP(x, t)$, we have from (1.9)

$$\frac{\partial P(x, t)}{\partial t} = D\frac{\partial^2 P(x, t)}{\partial x^2}. \tag{1.10}$$

Since $P(x, t)$ is the probability density, it is normalized as eq. (1.3).

1.2.2 Solution of the diffusion equation

The time derivative of the partial differential equation (1.10) is the first order. Thus, when we know the probability density at initial time $t = t_0$, $P(x, t_0)$, and the boundary condition, then $P(x, t)$ for $t > t_0$ can be obtained.

Consider the situation where the particle is located at $x = x_0$ at time $t = t_0$, namely, the initial condition is given by

$$P(x, t_0) = \delta(x - x_0). \tag{1.11}$$

The solution of eq.(1.10) with the initial condition (1.11) under the boundary condition that $P(x, t) \to 0$ and $\frac{\partial}{\partial x}P(x, t) \to 0$ for $x \to \pm\infty$ is given by

$$P(x, t) = \frac{1}{\sqrt{4\pi D(t - t_0)}} \exp\left(-\frac{(x - x_0)^2}{4D(t - t_0)}\right). \tag{1.12}$$

This is the simplest description of the Brownian motion. The derivation is left for exercise. This is a *normal distribution* or a *Gaussian distribution.*

Properties of δ function $\delta(x)$

Since we will often use the delta-function in this course, let us remember what it is. The definition is the following:

(a) $\delta(x) = 0$ for $x \neq 0$,

(b) For any function $f(x)$ that is continuous at $x = 0$,

$$\int_{-\infty}^{\infty} f(x)\delta(x)dx = f(0).$$

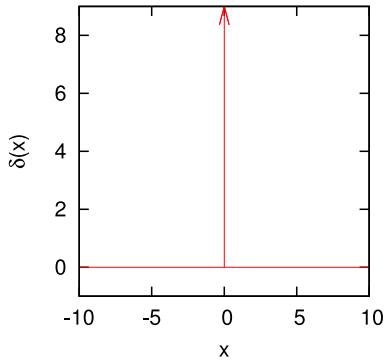


Figure 1.5: Schematic description of δ-function.

Schematically, $\delta(x)$ has "an infinitely high peak at $x = 0$ with an infinitely small width, and the integrated area is one" (Fig. 1.5).

From the definition, we can derive the following properties:

$$\delta(x) = \delta(-x), \tag{1.13}$$

$$\int_{-\infty}^{\infty} \delta(x)dx = 1, \tag{1.14}$$

$$\int_{-\infty}^{\infty} \delta(x-y)f(x)dx = f(y). \tag{1.15}$$

For $\epsilon > 0$,

$$\int_{x_0-\epsilon}^{x_0+\epsilon} f(x)\delta(x-x_0)dx = f(x_0). \tag{1.16}$$

For $y < x_0 - \epsilon$,

$$\int_{y}^{x_0-\epsilon} f(x)\delta(x-x_0)dx = 0. \tag{1.17}$$

For $y' > x_0 + \epsilon$,

$$\int_{x_0+\epsilon}^{y'} f(x)\delta(x-x_0)dx = 0. \tag{1.18}$$

1.3 Gaussian distribution

Now let us go back to diffusion and study more about what the solution (1.12) means. For simplicity, let us consider the probability distribution for Brownian particles with the initial condition $P(x, 0) = \delta(0)$, which is

$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) \tag{1.19}$$

(eq. (1.12) with $x_0 = 0, t_0 = 0$.).

1.3.1 Mean and Variance

We define $\langle f(X)\rangle$ as average of an arbitrary function of stochastic variable X, $f(X)$. When the distribution function of $X(t)$ is given by $P(x,t)$, the average at time t is given by

$$\langle f(X(t))\rangle = \int f(x)P(x,t)dx, \tag{1.20}$$

where integration runs over all possible values of x.

Now let us calculate the mean position

$$m \equiv \langle X(t) \rangle \tag{1.21}$$

and its variance

$$\sigma^2 \equiv \langle (X(t) - \langle X(t) \rangle)^2 \rangle = \langle X(t)^2 \rangle - \langle X(t) \rangle^2 \tag{1.22}$$

when the distribution function is given by

$$P(x,t) = \frac{1}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right). \tag{1.23}$$

The mean is given by

$$m = \langle X(t) \rangle = \int_{-\infty}^{\infty} x P(x,t) dx = \frac{1}{\sqrt{4\pi D t}} \int_{-\infty}^{\infty} x \exp\left(-\frac{x^2}{4Dt}\right) dx = 0. \tag{1.24}$$

The variance is given by

$$\sigma^2 = \langle X(t)^2 \rangle - \langle X(t) \rangle^2 = \frac{1}{\sqrt{4\pi D t}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{4Dt}\right) dx - 0^2 = 2Dt. \tag{1.25}$$

Namely, the mean position is at 0 (initial position), and the variance grows linearly with time t, with the coefficient proportional to D. The relation (1.25) is a very important property of diffusion, and we will encounter this expression many times from now on.

In the derivation, we used the Gauss integral

$$\int_{-\infty}^{\infty} \exp\left(-ax^2\right) dx = \sqrt{\frac{\pi}{a}}. \tag{1.26}$$

$$\int_{-\infty}^{\infty} x^2 \exp\left(-ax^2\right) dx = -\frac{\partial}{\partial a} \int_{-\infty}^{\infty} \exp\left(-ax^2\right) dx = \frac{1}{2}\sqrt{\frac{\pi}{a^3}}. \tag{1.27}$$

In general, $\langle X^p \rangle$ is called the *p-th order moment*. It is useful to note that any moment of Gaussian is finite and one can calculate them. For example, it is easy to show (exercise) for a Gaussian distribution with $m = 0$ and variance $\sigma^2 \neq 0$ that

$$\langle X^3 \rangle = 0, \quad \langle X^4 \rangle = 3\sigma^4. \tag{1.28}$$

Cases with diverging moments

It should be noted that the (higher) moments may not exist for some probability distribution. One of the famous examples that have diverging moments is the *Cauchy distribution*

$$\frac{1}{\pi}\left[\frac{1}{1+x^2}\right].$$

For the Cauchy distribution, even the first moment is not well defined because

$$\langle X \rangle = \int_{-\infty}^{\infty} \frac{x}{\pi}\left[\frac{1}{1+x^2}\right] dx.$$

For large x the function inside the integral approaches $1/x$, and the integrated function $\ln x$ diverges in $x \to \infty$ limit [1].

[1] Intuitively, this means the following: Suppose you are trying to calculate the mean of a stochastic variable X from a finite number of samples - what you will do is calculate their average by summing them up and dividing it by the number of samples. If the value of X obeys the Cauchy distribution, the probability of getting "infinitely large" $|x|$ is rather high. Every time you get a new sample, it may be a very large positive or negative number. This moves the average a lot. The average will fluctuate even if you have a very large number of samples and never converge.

1.4 Random walk: A microscopic description of Brownian motion

So far, we have been considering the diffusion equation, which is a macroscopic description of the Brownian motion. Another very famous model describes the Brownian motion is *random walk on a lattice*. Many of you may have already heard about it, but here we quickly review what it is and see how the random walk and the diffusion are related.

1.4.1 Random walk in one dimension

To define a random walk, we discretize time and space as follows:

- A Brownian particle starts at $x = 0$ and is free to move in either direction along the x-axis.
- The particle hops every time step Δt, and the displacement per step is either Δx or $-\Delta x$, with equal probabilities 1/2.

We use the following variables for the description of the random walk:

Z_i: The stochastic variable that represents the motion at the i-th step. $+\Delta x$ if the i-th step is to the plus x direction, $-\Delta x$ if the i-th step is to the minus x direction.

X_n: The position of the particle at time $t_n = n\Delta t$.

By using these variables, we can write

$$X_n = X_{n-1} + Z_n. \tag{1.29}$$

Mean position and variance

Let us first calculate the average position $\langle X_n \rangle$. From (1.29) and the initial condition, we have

$$X_n = \sum_{i=1}^{n} Z_i.$$

Since Z_i takes $+\Delta x$ or $-\Delta x$ with equal probability 1/2, we have

$$\langle Z_i \rangle = \frac{1}{2} \cdot (+\Delta x) + \frac{1}{2} \cdot (-\Delta x) = 0. \tag{1.30}$$

Thus we have

$$\langle X_n \rangle = \langle \sum_{i=1}^{n} Z_i \rangle = \sum_{i=1}^{n} \langle Z_i \rangle = 0.$$

Next, let us calculate the variance $\sigma^2 = \langle X_n^2 \rangle - \langle X_n \rangle^2$. From the definition and eq. (1.30) we get

$$\langle Z_i^2 \rangle = \frac{1}{2} \cdot (+\Delta x)^2 + \frac{1}{2} \cdot (-\Delta x)^2 = \Delta x^2.$$

For the cross term with $i \neq j$, noting that Z_i and Z_j are statistically independent [2] and using eq. (1.30), we have

$$\langle Z_i Z_j \rangle = \langle Z_i \rangle \langle Z_j \rangle = 0 \quad \text{for} \quad i \neq j.$$

[2]Note on "statistical independence": Suppose the probability that a stochastic valuable A takes the value a is $P(a)$, and the probability that another stochastic valuable B takes the value b is $P(b)$. If A and B are "statistically independent", i.e., the value of A does not affect the value of B and vice-versa, then the probability to have $A = a$ AND $B = b$, $P(a\&b)$, is given by $P(a\&b) = P(a)P(b)$. In this case, we have

$$\langle AB \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j P(a_i \& b_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j P(a_i) P(b_j) = \left(\sum_{i=1}^{n} a_i P(a_i) \right) \left(\sum_{j=1}^{n} b_j P(b_j) \right) = \langle A \rangle \langle B \rangle.$$

Since

$$X_n^2 = \left(\sum_{i=1}^n Z_i\right)^2 = \left(\sum_{i=1}^n Z_i\right)\left(\sum_{j=1}^n Z_j\right) = \sum_{i=1}^n\sum_{j=1}^n Z_i Z_j = \sum_{i=1}^n Z_i^2 + \sum_{i=1}^n\sum_{j\neq i} Z_i Z_j,$$

we have

$$\langle X_n^2\rangle = \sum_{i=1}^n \langle Z_i^2\rangle + \sum_{i=1}^n\sum_{j\neq i}\langle Z_i Z_j\rangle = \sum_{i=1}^n \Delta x^2 = n\Delta x^2. \tag{1.31}$$

1.4.2 Random walk and Diffusion

Let us compare these results with what we got for the diffusion equation.

Mean and variance

$$\langle X_n\rangle = 0, \tag{1.32}$$

i.e., the random walker does not move on average independent of the time step n, and this is the same as the diffusion case (1.24).

From (1.31), the variance of the random walk is given by

$$\sigma^2 = \langle X_n^2\rangle - \langle X_n\rangle^2 = n\Delta x^2 = \frac{\Delta x^2}{\Delta t}t. \tag{1.33}$$

Here, we used the conversion $t = n\Delta t$, i.e., we express the time at n-th step as t. Comparing (1.33) with (1.25), we see that the random walk behaves like diffusion, and the diffusion constant is given by

$$D = \frac{\Delta x^2}{2\Delta t}. \tag{1.34}$$

Probability distribution for random walk

What is the probability $P_n(m)$ to be at position $x = m\Delta x$ after n time steps ($t = n\Delta t$)? Can we say more about the correspondence between random walk and diffusion from the probability distribution?

The probability $P(l, r)$ that the walker has l left steps and r right steps are given by a binomial distribution

$$P(l,r) = \begin{pmatrix} r+l \\ l \end{pmatrix}\left(\frac{1}{2}\right)^{r+l} = \frac{(r+l)!}{r!l!}\left(\frac{1}{2}\right)^{r+l}.$$

At the same time, in order that the position after n step is $m\Delta x$, l and r should satisfy the conditions $l + r = n$ and $r - l = m$. Thus we have

$$P_n(m) = \frac{n!}{\frac{n+m}{2}!\frac{n-m}{2}!}\left(\frac{1}{2}\right)^n. \tag{1.35}$$

This is true only when $n - m$ is even: otherwise $P_n(m)$ is zero. This is because, at even (odd) step n, the walker can be only at one of the even (odd) sites. So more precisely, we have

$$P_n(m) = \begin{cases} \frac{n!}{\frac{n+m}{2}!\frac{n-m}{2}!}\left(\frac{1}{2}\right)^n & \text{if } m+n \text{ is even,} \\ 0 & \text{if } m+n \text{ is odd.} \end{cases}.$$

large n limit

For large n, the variable becomes practically continuous, namely, the time step Δt and the step size Δx is very small compared to the time passed and distance moved. We can find the distribution in this limit by using the Stirling's approximation for large n:

$$\ln n! \approx \frac{1}{2}\ln 2\pi + \left(n + \frac{1}{2}\right)\ln n - n. \tag{1.36}$$

We use this and expand $\ln P_n(m)$ around its most probable value $\langle m \rangle = 0$, by assuming m is small, up to m^2. In the end, we get:

$$P_n(m) \approx \frac{2}{\sqrt{2\pi n}} e^{-m^2/(2n)}. \tag{1.37}$$

Normalization is correct because $P_n(m)$ is zero when $n - m$ is odd.

Considering $t = n\Delta t$ and $x = m\Delta x$, and $P_n(m)$ for given n takes non-zero value every interval of $2\Delta x$, (1.37) becomes

$$P(x)(2\Delta x) \approx P_n(m) \approx \frac{2}{\sqrt{2\pi t/\Delta t}} e^{-(x/\Delta x)^2/(2t/\Delta t)},$$

which leads

$$P(x) \approx \frac{1}{\sqrt{4\pi \cdot (\Delta x^2/2\Delta t)t}} e^{-x^2/(4(\Delta x^2/2\Delta t)t}. \tag{1.38}$$

This is the Gaussian distribution (1.23) with the diffusion constant defined by (1.34)! We have confirmed that not only the mean and the variance but also the probability distribution of random walk at large t "obeys" the diffusion equation.

1.5 The central limit theorem

The fact that the random walk in large n limit gives the Gaussian distribution with variance proportional n is not just a coincidence. This is an example of the **central limit theorem**.

> **Central limit theorem.** Consider a stochastic variable Z, that obeys a probability distribution $P(z)$, and that has zero mean $\langle Z \rangle = 0$ (This can be done without loss of generality as long as the mean is finite, by just subtracting the mean from the original stochastic variable.) and finite variance σ_z^2. We draw n independent stochastic variable Z_j with $j = 1, \cdots, n$ from the distribution $P(z)$, and construct the following sum:
>
> $$Y_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j, \tag{1.39}$$
>
> Naturally $\langle Y_n \rangle = 0$. The central limit theorem states that, the distribution of Y_n, $P_n(y)$, approaches toward a Gaussian distribution as n becomes large, that is
>
> $$\lim_{n\to\infty} P_n(y) = \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-y^2/(2\sigma_z^2)}. \tag{1.40}$$

For the random walk case, Z_j is the displacement of the each time step, and $\sqrt{n}Y_n$ can be the position at n time steps, X_n. Since the variance of Z_j is Δx^2, the theorem states that in the large n limit X_n follows Gaussian with zero mean and with variance $n\Delta x^2$, as we just showed! In other words, we have just shown the central limit theorem for a special case.

The central limit theorem explains why actual Brownian motion (not only the random walk model) obeys the Gaussian distribution (1.23). For a very short time scale, the motion of a Brownian particle should not be diffusive; if a particle is moving at a velocity u, then it takes certain *correlation time* τ_c before the particle moves with a completely different velocity. However, if we divide the time into steps of Δt, and if we can take $\Delta t \gg \tau_c$, then the displacement for each time step, Z_j, can be considered as independent random variables. Then, the total displacement, which is the sum of this small displacement Z_j, obeys the central limit theorem. Therefore, it gives the normal distribution with variance proportional to $t = n\Delta t$.

The central limit theorem can be easily proven by using the Fourier transform of the distribution functions called the characteristic function. The idea is that, if you express the characteristic function for the distribution of Y_n by using the characteristic function of the distribution of Z, then we can easily take the limit of $n \to \infty$ to see that the characteristic function converges to the Gaussian form in this limit. Interested readers can find the proof in Appendix A (And I recommend checking it out if you have never seen any proof for the central limit theorem before).

1.6 Summary

- The Brownian motion can be considered a stochastic process.
- Fick's law and the equation of continuity give the diffusion equation

$$\frac{\partial}{\partial t}P(x,t) = D\frac{\partial^2}{\partial x^2}P(x,t).$$

- The solution of the diffusion equation, where particle start at $x = 0$ at time $t = 0$, is a Gaussian,

$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right).$$

- The variance of the position grows linearly with time, i.e.,

$$\sigma^2 = \langle X(t)^2\rangle - \langle X(t)\rangle^2 = 2Dt.$$

- The random walk model with step size Δx and the time step Δt can reproduce diffusion, with $D = \frac{\Delta x^2}{2\Delta t}$.
- The proportionality of the variance being proportional to time t can be explained by the central limit theorem.

1.7 Exercise

1. Using the definition of δ-function, show (1.13) to (1.18).
2. Using the definition of δ-function, calculate the following two quantities.

   (a)
   $$\int_0^5 (x+2)\delta(x-10)dx$$

   (b)
   $$\int_0^{20} (x+2)\delta(x-10)dx$$

3. Derive (1.12) by solving eq.(1.10) with eq.(1.11), under the boundary condition that $P(x,t) \to 0$ and $\frac{\partial}{\partial x}P(x,t) \to 0$ for $x \to \pm\infty$. Below the procedure and hints are given, but you are of course welcome to solve it in your own way if you have a good idea!

   (a) The Fourier Transform (FT) of $P(x,t)$ in space is defined as
   $$Q(k,t) = \int_{-\infty}^{\infty} P(x,t)e^{ikx}dx.$$
   Perform FT for eq.(1.10) and show that
   $$\frac{\partial}{\partial t}Q(k,t) = -k^2DQ(k,t).$$

   (b) Perform FT of the initial condition eq.(1.11) and show that
   $$Q(k,t_0) = e^{ikx_0}.$$

   (c) Solve the obtained equations for $Q(k,t)$.

(d) Perform the inverse FT of obtained $Q(k,t)$ as

$$P(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} Q(k,t)e^{-ikx}dk$$

to derive (1.12).

Hints: You need to use the integration by parts

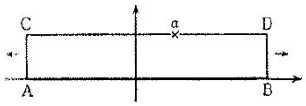
$$\int \frac{df(x)}{dx}g(x)dx = f(x)g(x) - \int f(x)\frac{dg(x)}{dx}dx.$$

You also need to use the Gauss integral:

$$\int_{-\infty}^{\infty} \exp\left(-ax^2\right)dx = \sqrt{\frac{\pi}{a}}.$$

You also need to prove

$$\int_{-\infty}^{\infty} \exp\left(-a(x+ib)^2\right)dx = \int_{-\infty}^{\infty} \exp\left(-ax^2\right)dx$$



for real $a > 0$. This can be proven by considering the integral of a function $f(z) = e^{-az^2}$ in a complex plain in the closed path shown in the figure.

4. Show that

$$\langle X^3\rangle = 0, \quad \langle X^4\rangle = 3\sigma^4 \tag{1.41}$$

for Gaussian distribution with $m = 0$ and $\sigma \neq 0$.

5. * Perform the computer simulation of 1-dimensional random walk on a lattice defined in the subsection 1.4.1, with $\Delta x = 1$ and $\Delta t = 1$. Simulate 200 trajectories from time zero to time step 20. Calculate the mean and variance at time 100 and confirm that it agrees with what we expect (see the subsection 1.4.1). Note that in this chapter we discussed ensemble averages, NOT time averages.

Advanced exercises

6. Assuming that Fick's law holds, the diffusion equation for an in 3-dimension inhomogeneous space (i.e., there is no special direction) is given by

$$\frac{\partial P(\vec{x},t)}{\partial t} = D\nabla^2 P(\vec{x},t), \tag{1.42}$$

where $P(\vec{x},t)d\vec{x}$ is the probability density for a Brownian particle to be at a position $\vec{x}$ to $\vec{x}+d\vec{x}$. Here, $\vec{x} = (x,y,z)$ is 3-dimensional vector and

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

(a) Solve eq. (1.42) with the initial condition

$$P(\vec{x},0) = \delta(x)\delta(y)\delta(z) \tag{1.43}$$

and under the boundary condition that $P(\vec{x},t) \to 0$ and $\nabla P(\vec{x},t) \to \vec{0}$ for $|\vec{x}| \to \infty$.

(b) Using the obtained solution, show that the mean at time t is

$$\langle \vec{X}(t)\rangle = \vec{0}, \tag{1.44}$$

and the variance is

$$\langle \vec{X}(t)^2\rangle - \langle \vec{X}(t)\rangle^2 = \langle X(t)^2 + Y(t)^2 + Z(t)^2\rangle = 6Dt. \tag{1.45}$$

(The first equality used the fact that the mean is zero.)

In general, a d-dimensional diffusion equation is given by (1.42) with $\vec{x}$ and ∇^2 interpreted in the corresponding dimension, and the variance obeys

$$\langle \vec{X}(t)^2 \rangle - \langle \vec{X}(t) \rangle^2 = 2dDt. \tag{1.46}$$

7. Show (1.37) by expanding the logarithm of (1.35) in the large n limit.

8. Random walk in 2-dimension.

   (a) * Perform computer simulation of random walk on a square lattice with $\Delta x = 1$ (in 2-d there are 4 neighbours), where every time step $\Delta t = 1$ the walker can hop to one of the 4 neighbours. Confirm that the average position is constant over time, and variance grows in proportion to time.

   (b) What is the relation between the diffusion constant D and Δx and Δt? Remember the relation in d-dimension (1.46).

9. Prove the central limit theorem by using the characteristic function. The answer is given in Appendix A.

1.8 Further reading

This chapter is based on [1, 2].

Chapter 2

Langevin equation

2.1 Langevin equation

We have been looking at a few descriptions of Brownian motion, but none of them was talking about the mechanics that much. See Fig. 1.1: it is in principle about the interaction between the water molecules and the observed particles, so we should be able to treat it as a mechanics problem. However, at the same time, it is too much work to consider equations of motion for all water molecules.

In this chapter, we introduce a course-grained "equation of motion" to describe the dynamics of a Brownian particle, called a **Langevin equation**.

2.1.1 Equation of motion for a Brownian particle

We consider the equation of motion for a Brownian particle, which is simply

$$m\dot{\vec{V}} = \vec{F}, \tag{2.1}$$

with the particle's mass m, the particle's velocity $\vec{V}$, and the force from the collision of the solvent molecules to the solute particle $\vec{F}$. The key part is how to model the force $\vec{F}$.

Force on a Brownian particle

We divide the force into two contributions as follows:

$$\vec{F} = \vec{R}(t) - \eta\vec{V}. \tag{2.2}$$

This division comes from the following intuition.

- When the solute particle is not moving ($\vec{V} = 0$): The collision of the solvent molecules happens spatially uniformly. The instantaneous force is irregular and unpredictable, but the net force should be zero on average. This gives rise to the **random force** $\vec{R}(t)$. This is a stochastic variable.
- When the solute particle is moving ($\vec{V} \neq 0$): The collision happens more at the leading surface of the particle, which results in drag. This is described by the **drag force** $-\eta\vec{V}$ with $\eta > 0$.

Obviously, these two contributions are related to each other. We will see it later.

Langevin equation

The equation of motion with a "stochastic force term",

$$m\dot{\vec{V}} = -\eta\vec{V} + \vec{R}(t), \tag{2.3}$$

is a **Langevin equation**, first introduced by Paul Langevin in 1908 [3]. But we have not fully defined it yet.

- By saying that the random force is "stochastic", we are considering an ensemble... What is it?
- How should we model the random force $R(t)$?

Ensemble in the Langevin equation

The Langevin equation can be interpreted in the following way:

- There are many identical Brownian particles, starting from the same initial condition, and the Langevin equation with each realization of random force time series describes the motion of each particle. The ensemble is taken over the different realizations.

Here, it should be noted that the "initial condition" is set for the macroscopic quantities, e.g. the conditions for the solute particle, the density and the temperature of the solvent, etc. But we do not set the initial condition for each solvent molecule, and that is where the variation of the trajectory comes from in each realization.

Modeling random force

We consider a one-dimensional Langevin equation for simplicity.

$$m\dot{V} = -\eta V + R(t). \tag{2.4}$$

We postulate the following physically plausible properties to $R(t)$:

(i) We assume that $R(t)$ is treated as a stochastic variable.

(ii) We assume that the average random force is zero, which reflects the fact that the force should act uniformly on average.
$$\langle R(t)\rangle = 0. \tag{2.5}$$

(iii) We assume that the correlation time of the random force is "very short", and practically the random force at different time has nothing to do with each other.
$$\langle R(t)R(t')\rangle = 2D_u\delta(t-t'). \tag{2.6}$$
Here, D_u is a constant and $\delta(t)$ is the delta function.

Gaussian process The conditions (i)-(iii) are not enough to fully determine the property of $R(t)$ since they define only up to the 2nd moment. For further definition, it is natural to assume that

(iv) $R(t)$ is a *Gaussian noise*; if you integrate it over a very short time interval of Δt, its distribution will be Gaussian. Gaussian property is physically plausible due to the central limit theorem, since the random force $R(t)$ is actually considered to be the sum of forces from water molecules over a short interval.

We do not go into the precise definitions now, but later we will see more about the property of the Gaussian noise.

Macroscopic relaxation time and meaning of correlation time for the random force

Condition (ii) enables us to consider the average motion from eq. (2.4), which obeys

$$m\langle\dot{V}\rangle = m\frac{d}{dt}\langle V\rangle = -\eta\langle V\rangle. \tag{2.7}$$

This gives us the relaxation behaviour of the average velocity:

$$\langle V(t)\rangle = \langle V(0)\rangle \exp\left(-\gamma t\right), \tag{2.8}$$

with $\gamma = \eta/m$. Thus the macroscopic relaxation time scale is $1/\gamma$. Note that this is valid only for $t > 0$ so that it obeys the causality.

The "very short" correlation time in condition (iii) means that the correlation time for the random force is much shorter than the macroscopic relaxation time $1/\gamma$.

2.2 Velocity correlation of Brownian particles

2.2.1 Solving the Langevin equation

The formal integration of

$$m\dot{V} = -\eta V + R(t).$$

gives

$$V(t) = e^{-\gamma t}V(0) + \int_0^t e^{-\gamma(t-s)}\frac{1}{m}R(s)ds \tag{2.9}$$

(exercise). By taking the average of this equation we can get our previous result eq.(2.8).

2.2.2 Velocity correlation function

Calculating correlations

Using eq.(2.9), we calculate the velocity auto-correlation function $\langle V(t)V(t')\rangle$ for $t' \geq t > 0$. The product $V(t)V(t')$ is given by

$$\begin{aligned} V(t)V(t') &= e^{-\gamma(t+t')}V(0)^2 + \int_0^t e^{-\gamma(t-s+t')}\frac{1}{m}V(0)R(s)ds + \int_0^{t'} e^{-\gamma(t'-s+t)}\frac{1}{m}V(0)R(s)ds \\ &\quad + \int_0^t ds \int_0^{t'} ds' e^{-\gamma(t-s+t'-s')}\frac{1}{m^2}R(s)R(s'). \end{aligned} \tag{2.10}$$

Now we take the average of this expression. Here, we assume that the initial velocity $V(0)$ is not correlated with the random force $R(t)$ for $t \geq 0$, based on the physical intuition that future force cannot affect the velocity at present. Then, the average of the second and the third terms are zero, because

$$\left\langle \int_0^t e^{-\gamma(t-s+t')}\frac{1}{m}V(0)R(s)ds \right\rangle = \int_0^t e^{-\gamma(t-s+t')}\frac{1}{m}\langle V(0)R(s)\rangle ds = \int_0^t e^{-\gamma(t-s+t')}\frac{1}{m}\langle V(0)\rangle\langle R(s)\rangle ds = 0,$$

using $\langle R(s)\rangle = 0$ (condition ii).

Now we get, by using the condition (iii),

$$\begin{aligned} \langle V(t)V(t')\rangle &= e^{-\gamma(t+t')}\langle V(0)^2\rangle + \int_0^t ds \int_0^{t'} ds' e^{-\gamma(t-s+t'-s')}\frac{1}{m^2}2D_u\delta(s-s') \\ &= e^{-\gamma(t+t')}\langle V(0)^2\rangle + \int_0^t ds \left[\int_0^t ds' e^{-\gamma(t-s+t'-s')}\frac{1}{m^2}2D_u\delta(s-s') + \int_t^{t'} ds' e^{-\gamma(t-s+t'-s')}\frac{1}{m^2}2D_u\delta(s-s')\right] \\ &= e^{-\gamma(t+t')}\langle V(0)^2\rangle + \int_0^t ds \left[\int_0^t ds' e^{-\gamma(t-s+t'-s')}\frac{1}{m^2}2D_u\delta(s-s')\right] \\ &= e^{-\gamma(t+t')}\langle V(0)^2\rangle + \int_0^t ds e^{-\gamma(t-s+t'-s)}\frac{1}{m^2}2D_u \\ &= e^{-\gamma(t+t')}\langle V(0)^2\rangle + \frac{D_u}{m^2\gamma}e^{-\gamma(t+t')}(e^{2\gamma t}-1) \\ &= e^{-\gamma(t+t')}\left(\langle V(0)^2\rangle - \frac{D_u}{m^2\gamma}\right) + \frac{D_u}{m^2\gamma}e^{-\gamma(t'-t)}. \end{aligned}$$

From the second line to the third line, $\delta(x) = 0$ for $x \neq 0$ was used. From the third line to the fourth line, (1.16) was used. In summary, we get

$$\langle V(t)V(t')\rangle = e^{-\gamma(t+t')}\left(\langle V(0)^2\rangle - \frac{D_u}{m^2\gamma}\right) + \frac{D_u}{m^2\gamma}e^{-\gamma|t'-t|}. \tag{2.11}$$

Here, the absolute value in the last term came from the condition $t' \geq t$, which we used in the derivation. By using the absolute value we can extend the solution to any $t', t > 0$.

Especially when $t = t'$ we have

$$\langle V(t)^2 \rangle = e^{-2\gamma t} \left(\langle V(0)^2 \rangle - \frac{D_u}{m^2\gamma} \right) + \frac{D_u}{m^2\gamma}. \tag{2.12}$$

2.2.3 An example of fluctuation-dissipation theorem

In the long time, $t \gg 1/(2\gamma)$, we get

$$\langle V(\infty)^2 \rangle = \frac{D_u}{m^2\gamma}. \tag{2.13}$$

At the same time, the system should be in the *thermal equilibrium* after such a long time. This implies that the equipartition theorem holds, namely for the one-dimensional system at temperature T,

$$\frac{m}{2}\langle V(\infty)^2 \rangle = \frac{1}{2}k_B T. \tag{2.14}$$

Here, k_B is the Boltzmann constant. Eqs. (2.13) and (2.14) give the relation

$$\frac{D_u}{m\gamma} = k_B T. \tag{2.15}$$

This relation connects the equilibrium fluctuation and the non-equilibrium, irreversible relaxation process. This is an example of the **fluctuation-dissipation theorem**, and clearly shows that the random force and the viscous drag are related.

Velocity correlation in the equilibrium

Eq. (2.11) also tells us the following: if we do not set $V(0)$ to a particular value $v(0)$ but take average over thermal equilibrium velocity distribution for the initial condition, then we will have $\langle V(0)^2 \rangle = \langle V(\infty)^2 \rangle = \frac{D_u}{m^2\gamma}$, and the velocity correlation (2.11)becomes

$$\langle V(t)V(t') \rangle = \frac{D_u}{m^2\gamma} e^{-\gamma|t'-t|}. \tag{2.16}$$

Namely, it depends on the time difference $|t' - t|$ only.

2.3 Auto-correlation function and Wiener-Khinchin theorem

2.3.1 Stationary process

The velocity distribution for the Brownian particle in the thermal equilibrium is time-independent. This indicates that the velocity of V in thermal equilibrium can be considered as a **stationary process**.

When a stochastic process is **stationary**, the n-point probability density (1.6) satisfies

$$P_n(x_1, t_1; x_2, t_2; \cdots; x_n, t_n) = P_n(x_1, t_1 + \tau; x_2, t_2 + \tau; \cdots; x_n, t_n + \tau). \tag{2.17}$$

for any time shift τ. Thus, stationarity implies time translational invariance and vice versa. Especially, the one-point probability density must therefore be independent of time;

$$P(x_1, t_1) = P(x_1, t_1 + \tau) \quad \text{for any } \tau. \tag{2.18}$$

2.3.2 Auto-correlation in stationary process

Time translational invariance holds in stationary process. Therefore, the following holds for the auto-correlation function:

$$\langle X(t)X(t') \rangle = \langle X(t - t')X(0) \rangle = \langle X(0)X(t' - t) \rangle \tag{2.19}$$

By setting $t' = 0$ we get

$$\langle X(t)X(0)\rangle = \langle X(0)X(-t)\rangle. \tag{2.20}$$

Namely, for a stationary process, autocorrelation function

$$C(t) = \langle X(t_0)X(t_0+t)\rangle \tag{2.21}$$

is independent of t_0 and an even function of t, i.e. $C(t) = C(-t)$.

2.3.3 Wiener-Khinchin theorem and Power spectrum density

By using the Fourier transform of $X(t)$

$$\tilde{X}(\omega) \equiv \int_{-\infty}^{\infty} X(t)e^{i\omega t}dt, \tag{2.22}$$

the power spectrum density $I(\omega)$ of a stationary process $X(t)$ is defined as:

$$\langle \tilde{X}(\omega)\tilde{X}^*(\omega')\rangle = 2\pi\delta(\omega-\omega')I(\omega). \tag{2.23}$$

Here, $\tilde{X}^*$ means the complex conjugate, i.e., $\tilde{X}^*(\omega) \equiv \int_{-\infty}^{\infty} X(t)e^{-i\omega t}dt$.

Wiener-Khinchin theorem. The **Wiener-Khinchin theorem** relates the auto-correlation function $C(t)$ and the power spectrum density $I(\omega)$ as:

$$I(\omega) = \int_{-\infty}^{\infty} C(t)e^{i\omega t}dt. \tag{2.24}$$

This also means that the auto-correlation function can be obtained once we know the power spectrum density by

$$C(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} I(\omega)e^{-i\omega t}d\omega.$$

Proof of the Wiener-Khinchin theorem

From the definition,

$$\langle \tilde{X}(\omega)\tilde{X}^*(\omega')\rangle = \int_{-\infty}^{\infty} e^{i\omega t}\int_{-\infty}^{\infty} e^{-i\omega' t'}\langle X(t)X(t')\rangle dtdt'. \tag{2.25}$$

For the stationary process, we have (2.21), therefore

$$\langle \tilde{X}(\omega)\tilde{X}^*(\omega')\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\omega t - i\omega' t'}C(t-t')dtdt'. \tag{2.26}$$

By changing variable t' to $s \equiv t - t'$, we have

$$\langle \tilde{X}(\omega)\tilde{X}^*(\omega')\rangle = \int_{-\infty}^{\infty} e^{i(\omega-\omega')t}dt\int_{\infty}^{-\infty} e^{i\omega' s}C(s)(-ds) \tag{2.27}$$

$$= 2\pi\delta(\omega-\omega')\int_{-\infty}^{\infty} e^{i\omega s}C(s)ds. \tag{2.28}$$

For the last equality, we used

$$f(\omega')\delta(\omega-\omega') = f(\omega)\delta(\omega-\omega').$$

Note that the Fourier transform of the delta function gives

$$\int_{-\infty}^{\infty}\delta(x)e^{i\omega x}dx = 1.$$

Therefore, the inverse Fourier transformation of unity gives the delta function, i.e.,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} 1\cdot e^{-i\omega x}d\omega = \delta(x).$$

Note: the WK theorem for finite time series

We showed the WK theorem for an infinite time series (see the time runs from $-\infty$ to ∞ in integrals). We can prove the WK theorem (2.24) also for a finite time series, where power spectrum density $I(\omega)$ is defined as

$$A(\omega) = \int_0^T X(t) \exp(i\omega t)\, dt, \tag{2.29}$$

$$I(\omega) = \lim_{T\to\infty} \frac{1}{T}|A(\omega)|^2. \tag{2.30}$$

In this case, we need the process to be not only stationary but also **ergodic**, i.e. the long-time average and ensemble average matches in the process. For example,

$$\langle X(t)X(t')\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T X(t+\tau)X(t'+\tau)d\tau. \tag{2.31}$$

For interested readers, the proof the WK theorem for a finite time series is given in the appendix B.

2.3.4 Examples

White noise The random force $R(t)$ that we consider in (2.4) is a stationary process. The power spectrum $I(\omega)$ for the random force is

$$I(\omega) = 2D_u. \tag{2.32}$$

Therefore, the power spectrum contains all the frequencies at constant power. Such a random force is called a *white noise*. The random force we are using is often called a *Gaussian white noise*.

Pink noise We sometimes see that the power spectrum of a noisy signal follows a power low, $I(\omega) \propto \omega^{-\alpha}$ often with $0 < \alpha < 2$; such a noise is called $1/f$ noise or sometimes pink noise.

Power spectrum of the velocity The velocity of a Brownian particle $V(t)$ given by (2.4) is also a stationary process. The power spectrum is given by

$$I(\omega) = \frac{D_u}{m^2\gamma} \cdot \frac{2\gamma}{\gamma^2 + \omega^2}. \tag{2.33}$$

This functional form is called a *Lorenzian*.

Experimentally, it is often easy to measure the power spectrum of a time series of a signal that contains some kind of randomness, and often people use the power spectrum to categorize different signals/noises. If the process is stationary, the Wiener-Khinchin theorem connects the power spectrum to the auto-correlation of the signal.

2.4 Langevin equation and Diffusion

We have introduced the Langevin equation as a model for Brownian motion. In chapter 1, we have also seen that Brownian motion results in diffusion. So hopefully the Langevin equation can be also related to diffusion! In this section, we see how the diffusion process is described by the Langevin equation.

2.4.1 Diffusion and velocity correlation

Position of a Brownian particle

The position of the particle $X(t)$ satisfies

$$\dot{X}(t) = V(t),$$

thus $X(t)$ is obtained by integrating the velocity with the initial condition for the position $X(0) = x(0)$. We fix the initial position, meaning that diffusing particles / random walkers start from a certain position - as in the cases we considered in Chapter 1. Namely, we have

$$X(t) = \int_0^t V(s)ds + x(0). \tag{2.34}$$

Average position

By taking average of (2.34) and remembering eq. (2.8), we have

$$\begin{aligned} \langle X(t)\rangle &= \int_0^t \langle V(s)\rangle ds + x(0) = \int_0^t \langle V(0)\rangle \exp(-\gamma s)\, ds + x(0) \\ &= \frac{\langle V(0)\rangle}{\gamma}\left[1 - e^{-\gamma t}\right] + x(0). \end{aligned} \tag{2.35}$$

In the time scale longer than the velocity relaxation time scale ($t \gg 1/\gamma$), we can ignore $e^{-\gamma t}$ term, and we have

$$\langle X(t)\rangle \approx \frac{\langle V(0)\rangle}{\gamma} + x(0).$$

Thus initial condition determines the average position, and it stays constant. Especially, if the time scale of velocity relaxation is very short, i.e., $1/\gamma \ll 1$, this will become $x(0)$.

Variance of the position

We are interested in diffusion, which is characterized when we consider the time dependence of the variance of the position. By using eq. (2.34), we can express the variance of position as follows:

$$\begin{aligned} \langle (X(t) - \langle X(t)\rangle)^2\rangle &= \left\langle \left(\int_0^t V(s)ds - \int_0^t \langle V(s)\rangle ds \right)^2 \right\rangle \\ &= \int_0^t \int_0^t \langle V(s_1)V(s_2)\rangle ds_1 ds_2 - \left(\int_0^t \langle V(s)\rangle ds \right)^2. \end{aligned} \tag{2.36}$$

This expression depicts that the diffusion behaviour is determined by the integral of velocity correlation in general.

The first term of eq. (2.36) can be calculated by using the velocity correlation (2.11) we have just calculated. We then have

$$\int_0^t \int_0^t \langle V(s_1)V(s_2)\rangle ds_1 ds_2 = \int_0^t \int_0^t \left[e^{-\gamma(s_1+s_2)} \left(\langle V(0)^2\rangle - \frac{D_u}{m^2\gamma} \right) + \frac{D_u}{m^2\gamma} e^{-\gamma|s_1 - s_2|} \right] ds_1 ds_2. \tag{2.37}$$

Noting that

$$\int_0^t \int_0^t e^{-\gamma(s_1+s_2)} ds_1 ds_2 = \frac{1}{\gamma^2}\left(1 - e^{-\gamma t}\right)^2 \tag{2.38}$$

and

$$\begin{aligned} &\int_0^t \int_0^t e^{-\gamma|s_1 - s_2|} ds_1 ds_2 = 2\int_0^t \int_0^{s_1} e^{-\gamma(s_1 - s_2)} ds_1 ds_2 \\ &= 2\int_0^t e^{-\gamma s_1} \frac{1}{\gamma}\left(e^{\gamma s_1} - 1\right) ds_1 = 2\int_0^t \frac{1}{\gamma}\left(1 - e^{-\gamma s_1}\right) ds_1 \\ &= 2\frac{1}{\gamma}\left[t - \frac{1}{\gamma}\left(1 - e^{-\gamma t}\right) \right], \end{aligned} \tag{2.39}$$

we have

$$\begin{aligned}&\int_0^t \int_0^t \langle V(s_1)V(s_2)\rangle ds_1 ds_2 \\ =\ & \left(\langle V(0)^2\rangle - \frac{D_u}{m^2\gamma}\right)\frac{1}{\gamma^2}\left(1-e^{-\gamma t}\right)^2 + 2\frac{D_u}{m^2\gamma^2}\left[t - \frac{1}{\gamma}\left(1-e^{-\gamma t}\right)\right]. \end{aligned} \tag{2.40}$$

Substituting (2.35) (Note that $\int_0^t \langle V(s)\rangle ds = \langle X(t)\rangle - x(0)$) and (2.40) to (2.36), we finally obtain

$$\begin{aligned}&\langle (X(t) - \langle X(t)\rangle)^2\rangle \\ &= \left(\langle V(0)^2\rangle - \langle V(0)\rangle^2 - \frac{D_u}{m^2\gamma}\right)\frac{1}{\gamma^2}\left(1-e^{-\gamma t}\right)^2 + 2\frac{D_u}{m^2\gamma^2}\left[t - \frac{1}{\gamma}\left(1-e^{-\gamma t}\right)\right]. \end{aligned} \tag{2.41}$$

Long-time behavior of the variance

You may think that the variance we get (2.41) does not look like diffusion motion. But if we consider the long time scale ($t \gg 1/\gamma$), we get

$$\langle (X(t) - \langle X(t)\rangle)^2\rangle \approx \frac{1}{\gamma^2}\left(\langle V(0)^2\rangle - \langle V(0)\rangle^2\right) + 2\frac{D_u}{m^2\gamma^2}\left(t - \frac{3}{2\gamma}\right) \approx 2\frac{D_u}{m^2\gamma^2}t.$$

In the last term, we took the term growing in t only, since it is dominant in the long time $t \gg 1/\gamma$.

Comparison with the Diffusion constant

Remember from chapter 1 that in the diffusion we have

$$\langle X(t)\rangle = x(0),$$

and

$$\langle (X(t) - X(0))^2\rangle = 2Dt.$$

Thus the Langevin equation eq.(2.4) describes the diffusion in the time scale longer than velocity relaxation time scale ($t \gg 1/\gamma$), and the diffusion constant is given by

$$D = \frac{D_u}{m^2\gamma^2} = \frac{k_B T}{m\gamma} = \frac{k_B T}{\eta}$$

Here, in the second last equality, we used (2.15). This relation between the diffusion constant and the drag coefficient is called **Einstein's relation**, which is also an example of the fluctuation-dissipation theorem.

2.4.2 Langevin equation in over-damped limit

Over-damped description

We started with an equation of motion:

$$m\dot{V} = -\eta V + R(t)$$

and saw that the system is diffusive when the time scale is much longer than $1/\gamma = m/\eta$. This is the time scale that the inertia effect relaxes. If our shortest timescale of interest is still much longer than $1/\gamma = m/\eta$, we can ignore the inertia effect, and then the behaviour should look diffusive.

Equation for over-damped case

In this over-damped case, we ignore the inertia term $m\dot{V}$ to have

$$0 = -\eta V + R(t) \Leftrightarrow \dot{X} = \frac{1}{\eta} R(t). \tag{2.42}$$

If we re-define random force as $\xi(t) = \frac{1}{\eta} R(t)$, we have:

$$\begin{aligned} &\dot{X}(t) = \xi(t), \langle \xi(t) \rangle = 0, \langle \xi(t)\xi(t') \rangle = 2D\delta(t-t'), \\ &D = \frac{D_u}{\eta^2} = \frac{D_u}{m^2\gamma^2} = \frac{k_B T}{m\gamma} \end{aligned} \tag{2.43}$$

The over-damped equation is often used because our time scale of interest is often much longer than $1/\gamma$. An example is given in the exercise.

Note on over-damped case

We can also see this by considering the limit of $\gamma \to \infty$ (More precisely $m \to 0$ with η constant. Note that $\gamma = \eta/m$.) in (2.11), assuming that the initial velocity distribution is in thermal equilibrium, i.e., $\langle V(0)^2 \rangle = \frac{D_u}{m^2\gamma}$. Then we have

$$\begin{aligned} \langle V(t)V(t') \rangle &= \frac{D_u}{m^2\gamma} e^{-\gamma|t'-t|} \\ &= 2\frac{D_u}{\eta^2}\frac{\gamma}{2} e^{-\gamma|t'-t|}. \end{aligned} \tag{2.44}$$

Noting that when $\gamma \to \infty$ ($m \to 0$ with constant η)

$$\lim_{\gamma\to\infty} \frac{\gamma}{2} e^{-\gamma|t-t'|} = \delta(t-t'), \tag{2.45}$$

we have

$$\langle V(t)V(t') \rangle \to \frac{2D_u}{\eta^2}\delta(t-t'). \tag{2.46}$$

Since $\dot{X} = V$, this means that

$$\dot{X} = \xi(t), \tag{2.47}$$

in the limit, with $\xi(t)$ defined in (2.43).

Over-damped equation describes the ordinary diffusion

By taking the over-damped limit $\gamma \to \infty$ ($m \to 0$ with constant η) of eq. (2.35), we get

$$\langle X(t) \rangle = x(0). \tag{2.48}$$

Noting $m\gamma = \eta$, the same limit of the variance eq. (2.41) gives

$$\langle (X(t) - \langle X(t) \rangle)^2 \rangle = 2Dt. \tag{2.49}$$

The over-damped limit describes ordinary diffusion, as it should. Of course we can show (2.48) and (2.49) by directly integrating the over-damped Langevin equation (2.43), which is left for exercise.

2.5 Generalized forms of Langevin equation

2.5.1 Langevin equation for a particle under potential

Consider a Brownian particle. If the particle is under an external force, given by a potential $U(x)$, how should the Langevin equation be modified?

The answer is straightforward. We just add the potential force to the equation of motion.

$$m\dot{V}(t) = -\eta\dot{X}(t) - \frac{dU(X)}{dX} + R(t),$$

where the LHS gives the acceleration term, while the RHS includes the friction force, the potential force, and the random force. The random force of course satisfies (2.5) and (2.6).

Langevin equation for a particle under potential, over-damped case

In the over-damped case, we ignore the inertia term:

$$0 = -\eta\dot{X}(t) - \frac{dU(X)}{dX} + R(t).$$

This can be rewritten as

$$\dot{X}(t) = -\frac{1}{\eta}\frac{dU(X)}{dX} + \xi(t), \tag{2.50}$$

$$\langle\xi(t)\rangle = 0, \langle\xi(t)\xi(t')\rangle = 2D\delta(t-t'). \tag{2.51}$$

2.5.2 Linear and Nonlinear Langevin equation

Mathematically,

$$\dot{X}(t) = LX(t) + R(t).$$

with a constant L and random force $R(t)$ satisfying (2.5) and (2.6) is called a linear Langevin equation. Equation (2.4) is a linear Langevin equation with the constant L being negative. We have seen / will see a few more of them in the exercise.

This type of stochastic process $X(t)$ with $L < 0$ belongs to so-called the *Ornstein-Uhlenbeck process*. The first Langevin equation we considered, eq. (2.4), is an Ornstein-Uhlenbeck process. When $L = 0$, as the free Brownian particle's position in the over-damped limit eq. (2.47), and if $D = 1/2$ and the initial condition is $X(0) = 0$, then the process is called the *Wiener process*. These two processes are the most studied stochastic process. We will learn more about the Wiener process later in this course.

In general, one can consider a nonlinear Langevin equation:

$$\dot{X}(t) = F(X(t)) + R(t),$$

where $F(X)$ is a nonlinear function of X.

It is also possible to consider a nonlinear Langevin equation with multiplicative noise:

$$\dot{X}(t) = F(X(t)) + G(X(t))R(t).$$

where $F(X)$ and $G(X)$ are nonlinear functions of X. The noise term is *multiplicative*, i.e., the random force is multiplied by a function that depends on a stochastic variable. This type of equation requires some care in mathematical treatment. We deal with such a case later, but for a while, we consider only an *additive* noise, i.e., the random force is just added to the equation.

2.6 Numerical simulation of Langevin equations

Numerical simulation is often a very powerful tool to understand the behaviour of the system. Here we see how we can numerically solve a Langevin equation.

2.6.1 Euler method for an ordinary differential equation

Let us first consider an ordinary differential equation (ODE)

$$\dot{x} = f(x). \tag{2.52}$$

We consider the incremental change of x for a small time step Δt.

By integrating eq.(2.52) from t to $t + \Delta t$, we get

$$x(t + \Delta t) - x(t) = \int_t^{t+\Delta t} f(x(s))ds. \tag{2.53}$$

By expanding the RHS about Δt [1], we get

$$x(t + \Delta t) - x(t) = f(x(t))\Delta t + \frac{1}{2}f'(x(t))\dot{x}(t)\Delta t^2 + O(\Delta t^3), \tag{2.54}$$

where prime indicates the derivative by its argument. Thus, for very small Δt, we get

$$x(t + \Delta t) \approx x(t) + f(x(t))\Delta t. \tag{2.55}$$

RHS of eq.(2.55) depends only on the value of x at time t, not time $t + \Delta t$. Thus, when the initial condition is given for $t = t_0$ as $x(t_0) = x_0$, we can evaluate the value at $t_0 + \Delta t$ by using eq.(2.55). By repeating this procedure we can calculate the value at any t. This is called the Euler method. The numerical error induced by the Euler method per time step is $O(\Delta t^2)$. This is the simplest method of numerical integration. Note that the error to integrate from time 0 to a *fixed time* t is proportional to Δt, since the number of steps one needs to reach t is proportional to $1/\Delta t$.

2.6.2 "Euler method" for a Langevin equation

We now consider a nonlinear Langevin equation

$$\dot{X}(t) = F(X(t)) + R(t), \tag{2.56}$$

and construct the "Euler method" for this equation.

We assume $R(t)$ is a Gaussian noise, and

$$\langle R(t)\rangle = 0, \quad \langle R(t_1)R(t_2)\rangle = 2D\delta(t_1 - t_2). \tag{2.57}$$

Discretization of time

We integrate the Langevin eq. (2.56) from t to $t + \Delta t$ as

$$\int_t^{t+\Delta t} \dot{X}(s)ds = \int_t^{t+\Delta t} F(X(s))ds + \int_t^{t+\Delta t} R(s)ds. \tag{2.58}$$

LHS of (2.58) is

$$\int_t^{t+\Delta t} \dot{X}(s)ds = X(t + \Delta t) - X(t) \equiv \Delta X(t),$$

where we define $\Delta X(t)$. For the first term in RHS of (2.58), we perform a similar expansion as in the case for ODE, i.e.,

$$\int_t^{t+\Delta t} F(X(s))ds = F(X(t))\Delta t + \frac{1}{2}\frac{dF(X(t))}{dt}\Delta t^2 + \cdots, \tag{2.59}$$

where the second term in RHS is more explicitly

$$\frac{1}{2}\frac{dF(X(t))}{dt}\Delta t^2 = \frac{1}{2}F'(X(t))\frac{dX(t)}{dt}\Delta t^2 \approx \frac{1}{2}F'(X(t))\Delta X\Delta t.$$

[1] Remember that $F(t) = \int f(t)dt$ then $\int_a^t f(s)ds = F(t) - F(a)$ and $F'(t) = f(t)$ by definition.

This is actually $O(\Delta t^{3/2})$, because ΔX is $O(\sqrt{\Delta t})$ for a nonlinear Langevin equation eq. (2.56), as we see later. The most important thing here is the first term of RHS of (2.59) is $O(\Delta t)$, and the second term is higher order in Δt.

The second term in RHS cannot be approximated like this since $R(t)$ is related to the delta function - we need to be careful. Here we just define

$$\Delta W(t) \equiv \int_t^{t+\Delta t} R(s)ds. \tag{2.60}$$

Thus we have

$$\Delta X = F(X(t))\Delta t + \Delta W + O(\Delta X \Delta t). \tag{2.61}$$

2.6.3 Property of ΔW

Since ΔW contains $R(t)$, ΔW is sort of a random number. We can calculate its property from (2.57) as follows:

- Average:

$$\langle \Delta W \rangle = \int_t^{t+\Delta t} \langle R(s) \rangle ds = 0. \tag{2.62}$$

- Squared average:

$$\begin{aligned} \langle \Delta W^2 \rangle &= \int_t^{t+\Delta t} ds_1 \int_t^{t+\Delta t} ds_2 \langle R(s_1)R(s_2) \rangle \\ &= \int_t^{t+\Delta t} ds_1 \int_t^{t+\Delta t} ds_2 2D\delta(s_1 - s_2) \\ &= 2D \int_t^{t+\Delta t} ds_1 = 2D\Delta t. \end{aligned} \tag{2.63}$$

Interestingly, (2.63) implies that ΔW^2 is order of Δt, thus ΔW is order of $\Delta t^{1/2}$.

Knowing this, we can now see that the lowest order of eq. (2.61) is actually $\Delta W \sim O(\Delta t^{1/2})$. Therefore we have

$$O(\Delta X \Delta t) = O(\Delta W \Delta t) = O(\Delta t^{3/2}).$$

Gaussian noise assumption

To define ΔW fully, we need to know its distribution. We assumed that $R(t)$ is "a Gaussian noise"; More precisely, it should mean that ΔW is a Gaussian distribution, whatever the Δt is (ΔW is the "sum" of noise over the time Δt, and the sum of a Gaussian distributed noise also follows a Gaussian).

In addition, from (2.62) and (2.63), we know the average is 0 and the variance is $2D\Delta t$. Therefore, the distribution should be

$$P(\Delta W) = \frac{1}{\sqrt{4\pi D \Delta t}} e^{-\Delta W^2/(4D\Delta t)}. \tag{2.64}$$

This gives us all the moments in terms of ΔW, e.g.,

$$\begin{aligned} \langle \Delta W \rangle = 0, \qquad & \langle \Delta W^2 \rangle = \sigma^2 = 2D\Delta t, \\ \langle \Delta W^3 \rangle = 0, \qquad & \langle \Delta W^4 \rangle = 3\sigma^4, \cdots \end{aligned} \tag{2.65}$$

Independence between different time steps

We also know that ΔW's from time step t to $t + \Delta t$ and t' to $t' + \Delta t$ do not have any correlation as long as $|t - t'| > \Delta t$:

$$\langle \Delta W(t) \Delta W(t') \rangle = \int_t^{t+\Delta t} ds \int_{t'}^{t'+\Delta t} ds' \langle R(s)R(s') \rangle = \int_t^{t+\Delta t} ds \int_{t'}^{t'+\Delta t} ds' 2D\delta(s - s') = 0 \quad (\text{for} \quad |t - t'| > \Delta t) \tag{2.66}$$

2.6.4 Summary of the Euler Method for a nonlinear Langevin equation

For the Euler method, we ignore the terms higher order than Δt. So every time step, a new position is calculated as

$$X(t+\Delta t) = X(t) + F(X(t))\Delta t + \Delta W, \tag{2.67}$$

where ΔW is drawn independently at every time step from the Gaussian distribution (2.64). Do not forget that the variance of ΔW depends on the time step Δt. Note that this method gives an error of $O(\Delta t^{3/2})$ per time step, i.e., to calculate up to a fixed time t, the error would be $O(\Delta t^{1/2})$.

2.6.5 How to obtain the Gaussian distributed random number

In some programming languages, one can call a function that gives the Gaussian distributed random number g. But in many of programming languages, such a function may not exist by default. However, we almost always have access to a random number generator that gives a random number a that is uniformly distributed between 0 and 1 [2]. Then we need to construct g out of a.

There are several different methods for this, for example, you can find Box-Muller transformation if you search. This is certainly a choice. The downside is that the operation involves taking the logarithm and square root, which is not a very fast operation.

If you prioritize speed, it may be better to make use of the central limit theorem. We know that a is average $1/2$, and we can calculate its variance to show

$$\langle (a-1/2)^2 \rangle = \frac{1}{12}. \tag{2.68}$$

We take a sum of a_i, which is independently drawn random number from the uniform distribution in the range between 0 and 1, and normalize it as follows

$$b_n = \frac{\sum_{i=1}^{n}(a_i - 1/2)}{\sqrt{n/12}}.$$

Now we have a random number b_n with an average of 0 and variance of 1, and as n gets larger, the distribution of b_n approaches to Gaussian distribution. Practically, $n \approx 6$ or so gives quite good distribution (of course the larger n the better, and then you need to decide based on the trade-off between the time and quality).

2.7 Summary

- The Langevin equation:
$$m\dot{V} = -\eta V + R(t)$$
with
$$\langle R(t) \rangle = 0, \quad \langle R(t)R(t') \rangle = 2D_u \delta(t-t').$$
- The velocity autocorrelation decays exponentially as (eq. (2.16))
$$\langle V(t)V(0) \rangle \propto \exp(-\gamma t).$$
- An example of the fluctuation-dissipation theorem:
$$\frac{D_u}{m\gamma} = k_B T.$$
- The Wiener-Khinchin theorem for a stationary process:
$$I(\omega) = \int_{-\infty}^{\infty} C(t) e^{i\omega t} dt.$$

[2] Different random number generators give different quality random numbers (e.g., how much they are correlated, or if they have periodicity). The default random number generator can be quite simple but not high quality, so when you use them seriously (e.g. in research) think about what kind of generator you should use.

- The Langevin equation with inertia describes the diffusion in the time scale much longer than the velocity relaxation time, $1/\gamma$.
- The over-damped Langevin equation is obtained in $m \to 0$ limit. For a free Brownian particle, it becomes

$$\dot{X} = \xi(t), \quad \xi(t) = R(t)/\eta,$$

and this describes diffusion. Especially

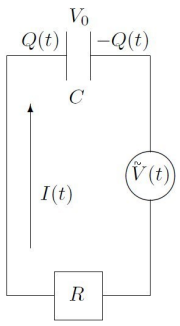
$$\langle (X(t) - X(0))^2 \rangle = 2Dt, \quad \text{with} \quad D = D_u/\eta^2.$$

- It is straightforward to include potential force in the Langevin equation.
- A nonlinear Langevin equation can be numerically integrated by using the Euler method, as

$$X(t + \Delta t) = X(t) + F(X(t))\Delta t + \Delta W.$$

2.8 Exercise

1. It has been observed that electric current can be spontaneously produced in an electric circuit with a capacitance C and a resistance R connected, as in the figure. This is because the thermal fluctuation gives a random voltage $\tilde{V}(t)$, which can be modeled as



$$\langle \tilde{V}(t) \rangle = 0, \langle \tilde{V}(t)\tilde{V}(t') \rangle = D_v \delta(t - t').$$

   Write down the equation for the charge $Q(t)$ - this is also a Langevin equation.
   (Point of this exercise is to know the wide applicability of the Langevin equation - not to drill you about electronic circuits.)

2. Consider a colloidal particle in a solvent in a one-dimensional system, that shows the Brownian motion. Now we trap the particle by optical tweezers, which exert a force centred at $x = 0$, generated by a harmonic potential $U(x) = \frac{1}{2}kx^2$. Write down the "equation of motion" for the position of the particle, $X(t)$. (Define parameters by yourself if necessary.)

3. Show (2.9) ($\gamma = \eta/m$). hint: use "variation of parameters" method.

4. Consider the random force $R(t)$ defined in (2.6). Calculate the following quantities.

   (a)
$$\left\langle \int_0^2 dt \int_0^2 dt' [R(t)R(t')] \right\rangle$$

   (b)
$$\left\langle \int_5^{10} dt \int_0^2 dt' [R(t)R(t')] \right\rangle$$

   (c)
$$\left\langle \int_0^{10} dt \int_2^5 dt' [R(t)R(t')] \right\rangle$$

5. The random force $R(t)$ that we consider in (2.4) is a stationary process. Show that the power spectrum $I(\omega)$ for the random force is

$$I(\omega) = 2D_u. \tag{2.69}$$

6. The velocity of a Brownian particle $V(t)$ given by (2.4) is also a stationary process. Calculate the power spectrum from the auto-correlation (2.16) using the Wiener-Khinchin theorem to show that

$$I(\omega) = \frac{D_u}{m^2\gamma} \cdot \frac{2\gamma}{\gamma^2 + \omega^2}. \tag{2.70}$$

How does the function look like if you plot $I(\omega)$ as a function of ω?

7. The drag coefficient η for a sphere of radius r in a liquid of viscosity η_l is given by $\eta = 6\pi\eta_l r$ (Stokes' drag). Consider a spherical colloid of radius 1μm with density 1 g/cm^3 in water. The water viscosity at room temperature is roughly $\eta_l = 1 \times 10^{-3}$ N s/m^2. What is the time scale for the velocity relaxation $1/\gamma$? How about a sphere of the same material but radius of 1mm?

8. Show that the over-damped equation (2.43) with initial condition $X(0) = x(0)$ gives

Average:

$$\langle X(t) \rangle = x(0), \tag{2.71}$$

Variance:

$$\langle (X(t) - \langle X(t) \rangle)^2 \rangle = 2Dt. \tag{2.72}$$

9. Consider a colloidal particle trapped by optical tweezers. For simplicity we consider one-dimensional system and we are interested in a rather long-time scale where the inertia effect is negligible. We assume that the potential generated by the tweezers is harmonic, $U(x) = \frac{1}{2}kx^2$.

   (a) Write down the Langevin equation for the position of the colloidal particle $X(t)$. (Define parameters by yourself if necessary.)

   (b) Calculate the average and the variance of the position under the initial condition that $X(0) = x_0$.

10. Suppose you can draw a random number r_s that obeys the standard normal distribution (namely a Gaussian distribution with the average 0 and the standard deviation 1). How can you convert r_s to a random number r_w that obeys a Gaussian distribution with the average 0 and the variance $2D\Delta t$?

11. ** Numerical integration of Langevin equation.

    (a) Simulate the Brownian motion by numerically solving

$$\dot{X}(t) = \xi(t), \quad \langle \xi(t) \rangle = 0, \quad \langle \xi(t)\xi(t') \rangle = 2D\delta(t - t').$$

    Note that the variance of ΔW depends on the time step, Δt.

    i. Set $D = 0.5$. Set the initial position to be $X(0) = 1$. Try the time step of $\Delta t = 0.1$, and plot $X(t)$ as a function of time, up to $t = 10$. Do the same with the time step of $\Delta t = 0.01$ and compare them.
       - Note: It is nice to be aware that, as long as ΔW has zero mean and correct squared average, the higher moments does not affect the long-time behaviour, thanks to the central limit theorem.

    ii. Measure the average $\langle X(t) \rangle$ and variance $\langle (X(t) - \langle X(t) \rangle)^2 \rangle$ of $X(t)$ and plot it as a function of time. Note that we must simulate many trajectories and take the ensemble average for each time point. Make sure that the slope of the variances for $\Delta t = 0.1$ and $\Delta t = 0.01$ are "the same" within reasonable deviation due to a finite number of samples and the fact that bigger Δt gives bigger errors (If not, something is wrong in the code!). Compare them with the theoretical prediction you calculated in exercise 2-8.

    (b) Now let's add the harmonic potential:

$$\dot{X}(t) = -(1/\eta)U'(X) + \xi(t), \quad U(x) = (1/2)kx^2.$$

    Simulate this with $k = 0.5, \eta = 2, D = 0.5$, with the initial position to be $X(0) = 1$. Choose an appropriate value of Δt and the simulation duration yourself. How does the particle behave? Plot $X(t)$, its average, and variance as a function of time. Compare the average and the variance with the theoretical prediction you calculated in exercise 2-9.

Advanced exercises

12. Show eq. (2.45).

13. Langevin equation for three-dimensional system can be given by

$$m\dot{\vec{V}} = -\eta\vec{V} + \vec{R}(t)$$

with the random force (now a vector) $\vec{R}(t) = (R_x(t), R_y(t), R_z(t))$ with

$$\langle R_i(t)\rangle = 0, \quad \langle R_i(t)R_j(t')\rangle = 2D_u\delta_{i,j}\delta(t-t'),$$

where i and j takes x, y, or z, and $\delta_{i,j}$ is Kronecker's delta.
Now consider the over-damped version of the equation without external force as:

$$\vec{V} = \dot{\vec{X}} = \vec{\xi}(t),$$

with $\vec{\xi}(t) = (\xi_x(t), \xi_y(t), \xi_z(t))$ with

$$\langle \xi_i(t)\rangle = 0, \quad \langle \xi_i(t)\xi_j(t')\rangle = 2D\delta_{i,j}\delta(t-t').$$

For initial condition $\vec{X}(0) = \vec{0}$, show that the variance at time t is given by

$$\left\langle \left(\vec{X}(t) - \langle\vec{X}(t)\rangle\right)^2 \right\rangle = 6Dt.$$

2.9 Further reading

This chapter is based on [1, 2, 4].

2

Chapter 3

Fokker-Planck equation

From microscopic view to macroscopic view

We have learned about the Langevin equations, analytically calculated some properties, and simulated individual trajectories. In this chapter, we learn how to derive the equation for the distribution function of the stochastic variables described by a Langevin equation. The equation is called the *Fokker-Planck* equation. We will also see that the diffusion equation can be derived as a Fokker-Planck equation corresponding to the Langevin equation of a free Brownian motion.

3.1 Derivation of the Fokker-Planck equation

3.1.1 Preparation

The system to consider

We consider a nonlinear Langevin equation

$$\dot{X}(t) = F(X(t)) + R(t), \tag{3.1}$$

and obtain the time evolution equation of the probability distribution for $X(t)$, $P(x,t)$.

We assume that $R(t)$ is a Gaussian white noise, and

$$\langle R(t) \rangle = 0, \quad \langle R(t_1)R(t_2) \rangle = 2D\delta(t_1 - t_2).$$

Assumptions to derive the FP equation

We need the following assumptions in the derivation:

a) $X(t)$ and $R(t')$ are statistically independent when $t < t'$: namely, the future random force has no effect on the present variable.

b) Consider an infinite region for x. The boundary conditions for the distribution function $P(x,t)$ are:

$$P(x,t) \to 0, \qquad \frac{\partial}{\partial x} P(x,t) \to 0,$$

in the limit of $x \to \pm\infty$.

3.1.2 Derivation

Outline of the derivation of the FP equation

The outline of the derivation is the following:

1. Consider an incremental change of $\langle f(X(t))\rangle$ from time t to $t+\Delta t$. Perform the Taylor expansion of $\langle f(X(t+\Delta t))\rangle$ with Δt up to the linear order of Δt.
2. Derive the time evolution equation for the average $\langle f(X)\rangle$ from the expansion.
3. Derive the FP equation based on this.

For the expansion, we use the results from chapter 2, in particular eq. (2.61)

$$\Delta X(t) \equiv X(t+\Delta t) - X(t), \quad \Delta X = F(X(t))\Delta t + \Delta W + O(\Delta X \Delta t)$$

and the properties of $\Delta W = \int_t^{t+\Delta t} R(s)ds$, (2.62)-(2.65).

Step 1: Taylor expansion of $f(X(t+\Delta t))$.

We will use a prime to express the derivative of a function by its argument. For example,

$$f'(X(t)) \equiv \frac{df(X)}{dX}|_{X=X(t)}, \quad f''(X(t)) = \frac{d^2 f(X)}{dX^2}|_{X=X(t)},$$

and so on.

Now, let's start the Tayler expansion.

$$\begin{aligned}
& f(X(t+\Delta t)) = f(X(t) + \Delta X(t)) \\
= \;& f(X(t)) + f'(X(t))\Delta X(t) + \frac{1}{2}f''(X(t))\Delta X(t)^2 + \cdots \\
= \;& f(X(t)) + f'(X(t))\left(F(X(t))\Delta t + \Delta W + O(\Delta X \Delta t)\right) + \frac{1}{2}f''(X(t))\left(F(X(t))\Delta t + \Delta W + O(\Delta X \Delta t)\right)^2 + O(\Delta X^3) \\
= \;& f(X(t)) + f'(X(t))F(X(t))\Delta t + f'(X(t))\Delta W \\
& + \frac{1}{2}f''(X(t))F(X(t))^2\Delta t^2 + f''(X(t))F(X(t))\Delta t \Delta W + \frac{1}{2}f''(X(t))\Delta W^2 + O(\Delta X^3) + O(\Delta X \Delta t).
\end{aligned}$$

Now we take the average of each term. By noting Δt is not a stochastic variable, clearly

$$\langle f'(X(t))F(X(t))\Delta t\rangle = \langle f'(X(t))F(X(t))\rangle \Delta t.$$

By using assumption a), we have

$$\langle f'(X(t))\Delta W\rangle = \langle f'(X(t))\rangle \langle \Delta W\rangle = 0.$$

Similarly, we can derive the following:

$$\left\langle f''(X(t))F(X(t))^2\Delta t^2\right\rangle = \left\langle f''(X(t))F(X(t))^2\right\rangle \Delta t^2.$$

$$\langle f''(X(t))F(X(t))\Delta t \Delta W\rangle \;=\; \langle f''(X(t))F(X(t))\rangle \langle \Delta W\rangle \Delta t = 0.$$

(Used the assumption a).)

$$\left\langle f''(X(t))\Delta W^2\right\rangle \;=\; \langle f''(X(t))\rangle \langle \Delta W^2\rangle = \langle f''(X(t))\rangle 2D\Delta t.$$

Here, assumption a) was used first, and then (2.65) was used.

Evaluation of remaining terms From the property of ΔW, we know $\langle \Delta W\rangle = 0$, $\langle \Delta W^2\rangle = \sigma^2 = 2D\Delta t$, $\langle \Delta W^3\rangle = 0$. Therefore, $\langle \Delta X \Delta t\rangle = \langle O(\Delta W \Delta t)\rangle + \langle O(\Delta t^2)\rangle = O(\Delta t^2)$, and similarly, since

$$\Delta X^3 = O(\Delta W^3) + O(\Delta W^2 \Delta t) + O(\Delta W \Delta t^2) + O(\Delta t^3), \tag{3.2}$$

we can see that

$$\langle \Delta X^3 \text{and higher}\rangle = \langle \Delta t^2 \text{and higher}\rangle$$

(Note that the terms with an odd power of ΔW disappear when the average is taken.) By summarizing all the results, we get

$$\langle f(X(t+\Delta t))\rangle = \langle f(X(t))\rangle + \langle f'(X(t))F(X(t))\rangle \Delta t + \langle f''(X(t))\rangle D\Delta t + O(\Delta t^2).$$

Step 2: Time evolution equation for $\langle f(X(t))\rangle$.

The equation can be modified to

$$\frac{\langle f(X(t+\Delta t))\rangle - \langle f(X(t))\rangle}{\Delta t} = \langle f'(X(t))F(X(t))\rangle + \langle f''(X(t))\rangle\, D + O(\Delta t).$$

In the limit of $\Delta t \to 0$, we get

$$\frac{d}{dt}\langle f(X(t))\rangle = \langle f'(X(t))F(X(t))\rangle + D\,\langle f''(X(t)).\rangle \tag{3.3}$$

Step 3: Convert the equation (3.3) to an equation for $P(x,t)$.

The averages can be expressed in terms of the distribution function as

$$\langle f(X(t))\rangle = \int_{-\infty}^{\infty} f(x)P(x,t)dx.$$

Therefore, for LHS of (3.3) we have

$$\frac{d}{dt}\langle f(X(t))\rangle = \int_{-\infty}^{\infty} f(x)\frac{\partial}{\partial t}P(x,t)dx.$$

The first term in RHS of (3.3) gives

$$\begin{aligned}
\langle f'(X(t))F(X(t))\rangle &= \int_{-\infty}^{\infty} \frac{df}{dx}F(x)P(x,t)dx \\
&= \left[f(x)F(x)P(x,t)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)\frac{\partial}{\partial x}\left(F(x)P(x,t)\right)dx \\
&= -\int_{-\infty}^{\infty} f(x)\frac{\partial}{\partial x}\left(F(x)P(x,t)\right)dx.
\end{aligned}$$

The second term in RHS of (3.3) gives

$$\begin{aligned}
\langle f''(X(t))\rangle &= \int_{-\infty}^{\infty} \frac{d^2f}{dx^2}P(x,t)dx \\
&= \left[\frac{df}{dx}P(x,t)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{df}{dx}\frac{\partial}{\partial x}P(x,t)dx \\
&= -\left[f(x)\frac{\partial}{\partial x}P(x,t)\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f(x)\frac{\partial^2}{\partial x^2}P(x,t)dx \\
&= \int_{-\infty}^{\infty} f(x)\frac{\partial^2}{\partial x^2}P(x,t)dx.
\end{aligned}$$

By summarizing what we get from (3.3), we have

$$\int_{-\infty}^{\infty} f(x)\frac{\partial}{\partial t}P(x,t)dx \quad = \quad -\int_{-\infty}^{\infty} f(x)\frac{\partial}{\partial x}\left(F(x)P(x,t)\right)dx + D\int_{-\infty}^{\infty} f(x)\frac{\partial^2}{\partial x^2}P(x,t)dx.$$

Since this relation should hold for any functional form of $f(x)$, we have

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}\left(F(x)P(x,t)\right) + D\frac{\partial^2}{\partial x^2}P(x,t). \tag{3.4}$$

This is the **Fokker-Planck equation** (FP equation).

The FP equation (3.4) can be rewritten as

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}\left[F(x)P(x,t) - D\frac{\partial}{\partial x}P(x,t).\right] \tag{3.5}$$

Comparing this with the equation of continuity

$$\frac{\partial P}{\partial t} = -\frac{\partial J}{\partial x},$$

we can see that the probability density is conserved when it is integrated over the x space, and the **probability flux** is given by

$$J = F(x)P(x,t) - D\frac{\partial}{\partial x}P(x,t).$$

The first term is the flux due to the "effective force" $F(x)$, and the second term is the flux due to diffusion.

3.2 Fokker-Planck equation of various systems

Fokker-Planck equation for the velocity of a Brownian particle

For a Brownian motion with inertia, we have

$$m\dot{V} = -\eta V + R(t), \quad \langle R(t)\rangle = 0, \quad \langle R(t)R(t')\rangle = 2D_u\delta(t-t'),$$

which is equivalent to

$$\dot{V} = -\gamma V + \tilde{R}(t), \quad \langle \tilde{R}(t)\rangle = 0, \quad \langle \tilde{R}(t)\tilde{R}(t')\rangle = 2\frac{D_u}{m^2}\delta(t-t').$$

The corresponding FP equation to describe the velocity distribution function is:

$$\frac{\partial}{\partial t}P(v,t) = -\frac{\partial}{\partial v}\left(-\gamma v P(v,t)\right) + \frac{D_u}{m^2}\frac{\partial^2}{\partial v^2}P(v,t). \tag{3.6}$$

Fokker-Planck equation for the position of a Brownian particle

For the over-damped case, we have

$$\dot{X} = \xi(t), \quad \langle \xi(t)\rangle = 0, \quad \langle \xi(t)\xi(t')\rangle = 2D\delta(t-t')$$

for a free Brownian particle. This gives us

$$\frac{\partial}{\partial t}P(x,t) = D\frac{\partial^2}{\partial x^2}P(x,t).$$

This is the diffusion equation! This guarantees that the Langevin equation for over-damped free Brownian particles indeed describe diffusion.

Fokker-Planck equation for Brownian motion under potential

If the particle feels the potential force, then

$$\dot{X} = -\frac{1}{\eta}U'(X) + \xi(t)$$

$$\langle \xi(t)\rangle = 0, \quad \langle \xi(t)\xi(t')\rangle = 2D\delta(t-t'),$$

and we have

$$\begin{aligned}\frac{\partial}{\partial t}P(x,t) &= \frac{\partial}{\partial x}\left(\frac{1}{\eta}U'(x)P(x,t)\right) + D\frac{\partial^2}{\partial x^2}P(x,t) \\ &= -\frac{\partial}{\partial x}\left[-\frac{1}{\eta}U'(x)P(x,t) - D\frac{\partial}{\partial x}P(x,t)\right].\end{aligned} \tag{3.7}$$

The first term in RHS can be interpreted as the flux by potential force is given by $-\frac{1}{\eta}U'(x)P(x,t)$, where $-\frac{1}{\eta}U'(x)$ is the average terminal velocity of a particle at position x.

3.3 The Fluctuation Dissipation Theorem (the 2nd kind)

Previously, by considering the Langevin equation for the Brownian motion and the energy distribution *at thermal equilibrium*, we obtained the relation

$$\frac{D_u}{m\gamma} = k_B T$$

The physical reason the diffusion constant and the viscosity are related is that the origins of the random force and the drag force are the same. Namely, they are due to the random collisions of water molecules under thermal fluctuation.

In this section, we derive this by using the Fokker-Planck equation and our knowledge of the equilibrium distribution. This enables us to generalize the **Fluctuation Dissipation Theorem (FDT)**.

3.3.1 Fluctuation Dissipation Theorem of the second kind using the Fokker-Plank equation

General case: System to consider

Consider an FP equation

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}\left(F(x)P(x,t)\right) + D\frac{\partial^2}{\partial x^2}P(x,t). \tag{3.8}$$

Because of the conservation of the probability, the distribution function $P(x,t)$ satisfies the equation of continuity

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}J(x,t), \tag{3.9}$$

where $J(x,t)$ is the probability flux. From the FP equation, we have

$$J(x,t) = -\left(-F(x) + D\frac{\partial}{\partial x}\right)P(x,t). \tag{3.10}$$

General case: Equilibrium distribution

Suppose we *know* the functional form of the equilibrium solution $P_{eq}(x)$ for the system (Typically derived from the equilibrium statistical mechanics). This distribution *should* also be the equilibrium solution of the FP equation. This means that the distribution gives zero flux:

$$\begin{aligned} 0 &= -\left(-F(x) + D\frac{\partial}{\partial x}\right)P_{eq}(x) \\ &= F(x)P_{eq}(x) - D\frac{\partial}{\partial x}P_{eq}(x). \end{aligned} \tag{3.11}$$

Equilibrium distribution and $F(x)$

Now let us define a function $S(x)$ so that

$$S(x) \equiv \ln P_{eq}(x),$$

namely $P_{eq}(x) = \exp[S(x)]$. Inserting this definition to the eq.(3.11), we get

$$\left(F(x) - D\frac{dS(x)}{dx}\right)P_{eq}(x) = 0.$$

In order that this holds for any x, we have

$$F(x) = D\frac{dS(x)}{dx} = DS'(x). \tag{3.12}$$

Especially, if $F(x) = LS'(x)$, e.g., $\dot{X} = LS'(X) + R(t)$,

$$L = D. \tag{3.13}$$

This is the **fluctuation-dissipation theorem (FDT)** of the 2nd kind.

Example: Velocity distribution for Brownian motion

From the statistical mechanics, we know that at equilibrium the velocity distribution function obeys the Boltzmann distribution:

$$P_{eq}(v) = P_{Boltzmann}(v) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{v^2}{2k_B T/m}\right),$$

giving

$$S(v) = \ln P_{eq} = -\frac{m}{2k_B T}v^2 + \ln\sqrt{\frac{m}{2\pi k_B T}}, \quad \text{thus} \quad \frac{dS(v)}{dv} = -\frac{m}{k_B T}v.$$

On the other hand, the Langevin equation for the velocity of a Brownian particle is given by

$$\dot{V} = -\gamma V + R'(t), \gamma = \eta/m, R'(t) = R(t)/m,$$
$$\langle R'(t)R'(t')\rangle = 2(D_u/m^2)\delta(t-t').$$

The corresponding FP equation to describe the velocity distribution function is then

$$\frac{\partial}{\partial t}P(v,t) \quad = \quad -\frac{\partial}{\partial v}\left(-\gamma v P(v,t)\right) + \frac{D_u}{m^2}\frac{\partial^2}{\partial v^2}P(v,t). \tag{3.14}$$

This means that

$$F(v) = -\gamma v.$$

Therefore, considering $D = D_u/m^2$ in this case, from (3.12) we have

$$-\gamma v = \frac{D_u}{m^2}\left(-\frac{m}{k_B T}v\right).$$

Since both sides are proportional to v, we, in the end, have

$$k_B T = \frac{D_u}{m\gamma}. \tag{3.15}$$

Note: FDT of the 1st kind and the 2nd kind

The fluctuation-dissipation theorem tells us that the response of the system to the external perturbation (e.g. the viscosity term determines the average response when a Brownian particle is forced to move, and this also results in *dissipation*) is related to the *fluctuation* in the system.

Here we saw it by using a Langevin equation with the random force $R(t)$, which is a rather phenomenological quantity. How to divide the force from water molecules between the frictional force and the random force is a non-trivial problem. However, the relation can be derived without considering a Langevin equation by using the linear response theorem (Ref. [1].) To distinguish these two derivations, the FDT that includes the random force is sometimes called *the second kind*, and the other one is called the first kind.

3.4 Summary

- We have derived the Fokker-Planck equation for a non-linear Langevin equation

  $$\dot{X}(t) = F(X(t)) + R(t),$$

  which is:

  $$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}\left(F(x)P(x,t)\right) + D\frac{\partial^2}{\partial x^2}P(x,t).$$

- If the system has an equilibrium distribution $P_{eq}(x)$, then

  $$F(x) = D\frac{d}{dx}\ln P_{eq}(x).$$

  Especially, when $F(x) = L\frac{d\ln[P_{eq}(x)]}{dx}$,

  $$L = D.$$

  This is the fluctuation-dissipation theorem (FDT) of the 2nd kind.

3.5 Exercise

1. Consider an electric circuit considered in chapter 2, exercise 2-1. Write down the FP equation for this system.

2. The Langevin equation for the Brownian motion with inertia can be written as

$$\dot{X}(t) = V(t), \tag{3.16}$$
$$m\dot{V}(t) = -\eta V + R(t), \tag{3.17}$$
$$\langle R(t)\rangle = 0, \qquad \langle R(t)R(t')\rangle = 2D_u\delta(t-t'). \tag{3.18}$$

Derive the FP equation for the distribution function of the position and velocity, $P(x,v,t)$, following the derivation parallel to the text. Assume

a) $V(t)$ and $R(t')$ are statistically independent when $t < t'$: namely, the future random force has no effect on the present variable. So as $X(t)$ and $R(t')$.

b) Consider an infinite region for x and v. The boundary condition for the distribution function $P(x,v,t)$ is that

$$P(x,v,t) \to 0, \qquad \frac{\partial}{\partial v}P(x,v,t) \to 0,$$

in the limit of $x \to \pm\infty$ or $v \to \pm\infty$.

We define

$$\Delta X = X(t+\Delta t) - X(t) = \int_t^{t+\Delta t} \dot{X}(t)dt, \quad \Delta V = V(t+\Delta t) - V(t) = \int_t^{t+\Delta t} \dot{V}(t)dt,$$
$$\Delta W = \int_t^{t+\Delta t} R(t)dt. \tag{3.19}$$

Then up to the order of Δt we have

$$\Delta X = \int_t^{t+\Delta t} V(t)dt = V(t)\Delta t + O(\Delta W \Delta t), \tag{3.20}$$
$$\Delta V = \int_t^{t+\Delta t} \left[-\frac{\eta}{m}V(t) + \frac{1}{m}R(t)\right] dt = -\frac{\eta}{m}V(t)\Delta t + \frac{1}{m}\Delta W + O(\Delta W \Delta t). \tag{3.21}$$

Note that ΔW is order of $\Delta t^{1/2}$.

(a) Perform the expansion of $f(X(t+\Delta t), V(t+\Delta t))$ up to the second order of ΔX and ΔV (remember the cross term). Substitute the explicit expression eqs.(3.20) and (3.21) to ΔX and ΔV and take the average to show that

$$\langle f(X(t+\Delta t), V(t+\Delta t))\rangle = \langle f(X(t), V(t))\rangle + \langle \frac{\partial f}{\partial X}|_{X=X(t),V=V(t)} V(t)\rangle \Delta t$$
$$-\langle \frac{\partial f}{\partial V}|_{X=X(t),V=V(t)} \frac{\eta}{m} V(t)\rangle \Delta t + \frac{D_u}{m^2}\langle \frac{\partial^2 f}{\partial V^2}|_{X=X(t),V(t)}\rangle \Delta t + O(\Delta t^2), \tag{3.22}$$

by leaving the terms up to the order of Δt.
Hint: Taylor expansion of $f(a+\Delta a, b+\Delta b)$ around (a,b) is given by

$$f(a+\Delta a, b+\Delta b) = f(a,b) + \frac{\partial f(x,y)}{\partial x}|_{x=a,y=b}\Delta a + \frac{\partial f(x,y)}{\partial y}|_{x=a,y=b}\Delta b$$
$$+\frac{1}{2}\left[\frac{\partial^2 f(x,y)}{\partial x^2}|_{x=a,y=b}\Delta a^2 + 2\frac{\partial^2 f(x,y)}{\partial x \partial y}|_{x=a,y=b}\Delta a \Delta b\right.$$
$$\left.+\frac{\partial^2 f(x,y)}{\partial y^2}|_{x=a,y=b}\Delta b^2\right] + O(\Delta a^3, \Delta b^3, \Delta a^2\Delta b, \Delta a \Delta b^2).$$

(b) By taking the limit of $\Delta t \to 0$, we have

$$\frac{d}{dt}\langle f(X(t),V(t))\rangle = \langle \frac{\partial f}{\partial X}|_{X=X(t),V=V(t)} V(t)\rangle - \langle \frac{\partial f}{\partial V}|_{X=X(t),V=V(t)} \frac{\eta}{m} V(t)\rangle + \frac{D_u}{m^2}\langle \frac{\partial^2 f}{\partial V^2}|_{X=X(t),V(t)}\rangle.$$

Noting

$$\langle f(X(t),V(t))\rangle = \int_{-\infty}^{\infty} dv \int_{-\infty}^{\infty} dx f(x,v) P(x,v,t)$$

and performing integration by parts, derive

$$\frac{\partial P(x,v,t)}{\partial t} = -v\frac{\partial P(x,v,t)}{\partial x} + \frac{\eta}{m}\frac{\partial}{\partial v}\left(vP(x,v,t)\right) + \frac{D_u}{m^2}\frac{\partial^2 P(x,v,t)}{\partial v^2}.$$

This equation (plus potential term) is called the Kramers equation. (often $p = mv$ is used instead of v.)

3. By solving the Fokker-Planck equation for the velocity (3.14) for equilibrium, obtain the equilibrium velocity distribution. Confirm that it matches with the Boltzmann distribution if we assume FDT (3.15).

4. Consider a particle under potential $U(x)$ (3.7). At the thermal equilibrium at temperature T, the equilibrium distribution $P_{eq} \propto e^{-U(x)/k_B T}$. Show that the FDT gives

$$D = \frac{k_B T}{\eta}.$$

This is the *Einstein's relation.*

Advanced exercises

5. Consider an electric circuit with thermal noise (exercise 1). The equilibrium distribution of the charge $P_{eq}(q)$ for this system is $P_{eq}(q) \propto e^{-E(q)/(k_B T)}$, where $E(q) = q^2/(2C)$ is the free energy of the condenser with the charge q.

Show that the FDT gives the relation

$$2Rk_B T = D_v.$$

This is called the *Nyquist's theorem*[1].

3.6 Further reading

This chapter is based on [1, 2, 4, 5].

[1] There is another theorem called Nyquist sampling theorem, which is completely different from FDT here.

Chapter 4

First passage problems

We learned the physics of diffusive processes and tools to describe them, including the Langevin equations and the Fokker-Plank equations. However, the quantities we calculated were quite simple so far. In this chapter, we learn a few examples of *first passage problems*, which asks when a "particle" under random force starting from one place reaches the "goal" for the first time. This is a dynamic problem and has many applications in various fields of research - and the tools we have learned are very useful to solve the first passage problems.

4.1 First passage time of a free Brownian particle

As our first example of the first passage time problem, let us consider the position of an over-damped Brownian particle $X(t)$. If the particle starts at $X(0) = a > 0$, what is the time T at which the particle reaches $X(T) = 0$ for the first time? The distribution of T, $P(T)$, is called the **first passage time distribution**, and this is what we will calculate here. This is one of the simplest first passage problems, and yet it shows interesting behaviour. The $a \to +0$ limit of this problem corresponds to the first return problem since it gives the time it takes for the particle to return to the original position given that it had started. (The first return time is better defined in the discrete space.)

This can be considered as the simplest form of target search by Brownian motion in one-dimensional space [1][2].

4.1.1 Derivation of first passage time distribution

In order to calculate $P(T)$, it will be convenient if we can remove a particle as soon as it hits $X = 0$. Then, we will not double-count the particles that go back and forth around $X = 0$. All we need to calculate will then be the probability flux from $x > 0$ to $x < 0$ under this condition, which is equal to $P(T)$ [3].

There is an easy way to perform the removal of particles in the FP equation, which is just a diffusion equation in this case. Remember that the diffusion equation

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2} = -\frac{\partial}{\partial x} j_d(x,t). \tag{4.1}$$

is the equation of continuity with diffusive flux j_d from left to right, with

$$j_d(x,t) = -D\frac{\partial P(x,t)}{\partial x}. \tag{4.2}$$

[1] Three-dimensional version of this problem in a limited volume is the base for, for example, estimating how long it takes for a molecule to find its target to react in a well-mixed solution [6, 7].

[2] We can also map this to a different problem. For example, if a gambler bets a fixed money on a fair coin toss, the money that the gambler has will go up and down following a random walk - then T may be interpreted as the time at which the gambler loses all the money.

[3] There are different derivations. Another derivation that does not use the solution of the diffusion equation under the absorbing boundary is given in appendix C.

If we set an *absorbing boundary condition* at $x = 0$ by setting

$$P(0, t) = 0, \tag{4.3}$$

the equation will describe the situation where a particle disappears as soon as it hits 0. The probability flux for time duration between T and $T + dT$ from $x > 0$ to $x < 0$ is equivalent to the first passage time distribution:

$$P(T) = -j_d(0, T). \tag{4.4}$$

Here, the minus sign is because we are interested in the flux to the negative direction.

Now what we need is to solve the diffusion equation with the initial condition

$$P(x, 0) = \delta(x - a) \tag{4.5}$$

under the boundary condition (4.3) in the semi-infinite space $x \geq 0$.

We can actually solve this by an application "the method of images" that you may know from calculating the electrostatic potential when there is a charge at $x = a$ with a conductor at $x = 0$. Suppose that there is no adsorbing boundary, but instead, there is a "mirror particle" that starts at the position $-a$ at time zero. The probability distribution of the mirror particle for $x < 0$ is a mirror image of the original particle that started at the position at $+a$ and mirrored at $x = 0$. That means that, if we subtract the probability distribution of the mirror particle from the probability distribution of the original particle, then the difference is exactly zero at $x = 0$, and still it satisfies the diffusion equation!

More precisely, when $P(x, t|a, 0)$ is the solution of the diffusion equation for an infinite system with the initial condition $P(x, 0) = \delta(x - a)$, then

$$P(x, t) = P(x, t|a, 0) - P(x, t| - a, 0) \tag{4.6}$$

(i) is a solution of the diffusion equation, (ii) satisfies the absorbing boundary condition (4.3), and also (iii) satisfies the initial condition (4.5) for $x \geq 0$, because

$$P(x, 0) = \delta(x - a) - \delta(x + a) \tag{4.7}$$

but $\delta(x + a) = 0$ for $x \neq -a$. So, this is the solution!

We have already calculated $P(x, t|a, 0)$ in chapter 1 (eq. 1.12 with $t_0 = 0$ and $x_0 = a$), so the explicit form of the solution (4.6) is given by

$$P(x, t) = \frac{1}{\sqrt{4D\pi t}} \left[\exp\left(-\frac{(x - a)^2}{4Dt} \right) - \exp\left(-\frac{(x + a)^2}{4Dt} \right) \right]. \tag{4.8}$$

We can also derive this as a limit of the finite system's solution, which is left for exercise.

The rest is a straightforward calculation. From (4.2),

$$j_d(x, t) = -\frac{D}{\sqrt{4D\pi t}} \left[-\frac{x - a}{2Dt} \exp\left(-\frac{(x - a)^2}{4Dt} \right) + \frac{x + a}{2Dt} \exp\left(-\frac{(x + a)^2}{4Dt} \right) \right]. \tag{4.9}$$

Therefore, from (4.4)

$$\begin{aligned} P(T) &= \frac{D}{\sqrt{4D\pi T}} \left[-\frac{-a}{2DT} \exp\left(-\frac{(-a)^2}{4DT} \right) + \frac{a}{2DT} \exp\left(-\frac{a^2}{4DT} \right) \right] \\ &= \frac{a}{\sqrt{4D\pi} T^{3/2}} \exp\left(-\frac{a^2}{4DT} \right). \end{aligned} \tag{4.10}$$

This is plotted in Fig. 4.1 for $D = 1/2$. The red, green, and blue lines show $P(T)$ as a function of T for $a = 0.1$, 1, and 10, respectively. The probability goes to zero when $T \to 0$, while it shows power-law $T^{-3/2}$ behaviour for large T (The magenta line shows $T^{-3/2}$). As a approaches zero, the power-law region becomes bigger. Namely, the first return time distribution ($a \to +0$ limit) is proportional to $T^{-3/2}$.

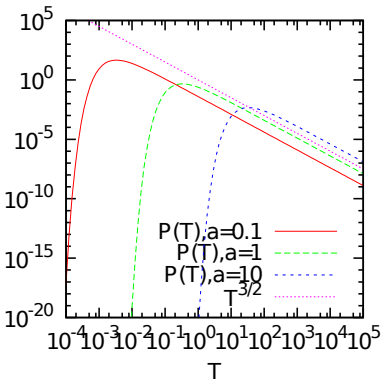


Figure 4.1: Plot of the first passage time distribution for $D = 1/2$.

4.1.2 The mean first passage time

The mean first passage time is given by

$$\langle T \rangle = \int_0^\infty TP(T)dT. \tag{4.11}$$

Since $P(T) \propto T^{-3/2}$ for large T ($T \gg a^2/6D$)[4] and

$$\int_0^\infty T \times T^{-3/2} dT = \int_0^\infty T^{-1/2} dT \propto [T^{1/2}]_0^\infty = \infty, \tag{4.12}$$

(Notice that the divergence comes from the upper limit of the integral, where $P(T) \propto T^{-3/2}$ is valid), we get

$$\langle T \rangle = \infty. \tag{4.13}$$

Namely, the mean first passage time diverges! This is one of the peculiar features of the free Brownian motion.

Note: How often does a random walker visit a particular position?

You might think that the mean first passage time diverges because some of the particles never reach 0. That is NOT true in this case. You can prove it by showing that the integral of the distribution (4.10) for all possible time is one (exercise), i.e, after infinite time, the walker will surely hit $x = 0$.

In fact, for the discrete random walk without the absorbing boundary, it is possible to show that a walker starting from $x = 0$ visits $x = 0$ infinite times after infinite time. We have calculated the probability to be at site m after n steps when $m - n$ is even in (1.37). If m is even, we can rewrite it as

$$P_{2k}(m) \approx \frac{1}{\sqrt{\pi k}} e^{-m^2/(4k)} \tag{4.14}$$

for a large enough integer $k \geq k_c$. Then we can evaluate how often a walker visits the site 0 after time step step $n_c = 2k_c$:

$$\sum_{k=k_c}^{\infty} P_{2k}(0) \approx \sum_{k=k_c}^{\infty} \frac{1}{\sqrt{\pi k}} > \frac{1}{\sqrt{\pi}} \sum_{k=k_c}^{\infty} \frac{1}{k} = \infty. \tag{4.15}$$

Intuitively this can be understood as follows. After n steps, the variance of the position grows linearly with n (the central limit theorem). The typical distance that a walker covered by then is given by the

[4]The position of the maximum of $P(T)$ can be calculated from $P'(T) = \left(-\frac{3}{2T} + \frac{a^2}{4DT^2}\right) \frac{a}{\sqrt{4D\pi T^{3/2}}} \exp\left(-\frac{a^2}{4DT}\right) = 0$, giving $T = \frac{a^2}{6D}$.

standard deviation, which scales as $\sqrt{n}$. Namely, the walker had visited n sites after n steps, but only went to distance $\sqrt{n} < n$. So it should have visited many sites multiple times. As n grows, all the sites around the origin will be surely visited multiple times.

What would happen if the walker is in higher dimensions? Consider a walker that starts from the origin, and performs a random walk in 3-dimension. Still, the central limit theorem holds, so the typical linear distance it had travelled after n steps scales with $\sqrt{n}$, which involves the order of $\sqrt{n}^3 = n\sqrt{n}$ sites. But it had visited only n sites by then. Therefore there will be some sites that had not been visited at all within the distance $\sqrt{n}$ from the origin. Indeed, the return probability (the probability that the walker starts at the origin and comes back to the origin at some point) is less than one in three and higher dimensions. This argument also tells us that 2-dimension is a marginal case and needs careful treatment. It is known that the return probability is one in the 2-dimensional case.

4.2 Kramers' escape

In this section, we consider another kind of fist passage problem, called **Kramers' escape**[5].

4.2.1 Kramers' escape problem

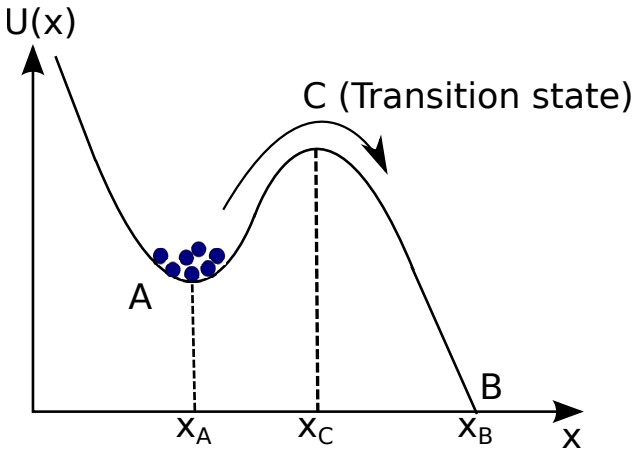


Figure 4.2: Schematic description of Kramers' escape problem

Consider a collection of independent Brownian particles in a potential $U(x)$ in one dimension, as shown in the Fig. 4.2. Point A ($x = x_A$) is the bottom of the potential well, and point B ($x = x_B$) is somewhere outside of the well. Between A and B, there is a potential barrier, whose highest point is C ($x = x_C$). Initially, all the particles are in the potential well. However, the thermal fluctuation results in a slow leak of particles across the barrier, from the well to point B. **What is the rate for a particle to escape the well?**

This can be considered a model of a chemical reaction, when the reaction coordinate is one-dimensional: There is a potential barrier to overcome so that a reaction happens, and thermal fluctuation is the only way to overcome this barrier. Once the barrier is overcome the the system is in a different state, and cannot come back.

Random force from a heat bath

The situation is not in complete equilibrium as can be seen in the existence of the flux of particles from point A to B. However, we still assume that the temperature of the system T is well defined and the strength

[5]Though the following is known as Kramers' escape, Kramers was not the first one to solve this problem in the over-damped case, which is what this section will deal with. What Kramers did was to solve this problem with inertia term. Ref. [8] contains short summary of the history.

of the random force is not affected by the particle flux, i.e., Einstein's relation (4.18) holds. This is a very plausible assumption as long as the source of the random force can be considered as a large enough heat bath.

4.2.2 FP equation for Brownian motion under potential

Let's start our calculation from the FP equation for an over-damped Brownian particle under the potential $U(x)$:

$$\begin{aligned} \frac{\partial}{\partial t}P(x,t) &= -\frac{\partial}{\partial x}\left(-\frac{1}{\eta}U'(x)P(x,t)\right) + D\frac{\partial^2}{\partial x^2}P(x,t) \\ &= -\frac{\partial}{\partial x}J(x,t), \end{aligned} \tag{4.16}$$

where the probability flux $J(x,t)$ is given by

$$J(x,t) = -\left(\frac{1}{\eta}U'(x) + D\frac{\partial}{\partial x}\right)P(x,t). \tag{4.17}$$

FDT gives us Einstein's relation

$$\frac{1}{\eta} = \frac{D}{k_B T}. \tag{4.18}$$

Overview of the calculation

The overview of the calculation is the following.

- The problem is non-equilibrium, i.e., the flux J is not zero.
- We assume that the system is quasi-stationary, namely that the flux going from A to B is small, and the probability distribution P is approximately independent of time ($\partial P/\partial t \approx 0$).
- We calculate the small flux J and the probability to be in the well p under this approximation, using the FP equation.
- The rate r for a particle to escape the well is given from the relation $J = pr$.

Calculation 1: probability flux

From the equation of continuity $\frac{\partial P(x,t)}{\partial t} = -\frac{\partial J}{\partial x}$, the quasi-stationary assumption ($\partial P/\partial t \approx 0$) means the flux J is a constant, independent of t AND x. From eq.(4.17) and FDT (4.18), we have

$$J = -\left(\frac{D}{k_B T}U'(x) + D\frac{d}{dx}\right)P(x). \tag{4.19}$$

Noticing that

$$\frac{d}{dx}\left(P(x)e^{U(x)/(k_B T)}\right) = e^{U(x)/(k_B T)}\left(\frac{U'(x)}{k_B T}P(x) + \frac{d}{dx}P(x)\right),$$

Eq.(4.19) becomes

$$J = -D\cdot e^{-U(x)/(k_B T)}\frac{d}{dx}\left(P(x)e^{U(x)/(k_B T)}\right). \tag{4.20}$$

Calculation 2: Integrate probability flux

Eq.(4.20) gives us

$$Je^{U(x)/(k_BT)} = -D \cdot \frac{d}{dx}\left(P(x)e^{U(x)/(k_BT)}\right). \tag{4.21}$$

By integrating both sides from $x = x_A$ to $x = x_B$, we have the following relation

$$\begin{aligned} J\int_{x_A}^{x_B} e^{U(x)/(k_BT)}dx &= -D\int_{x_A}^{x_B} \frac{d}{dx}\left(P(x)e^{U(x)/(k_BT)}\right)dx \\ &= -D\left[P(x_B)e^{U(x_B)/(k_BT)} - P(x_A)e^{U(x_A)/(k_BT)}\right], \end{aligned}$$

which gives

$$J = -D\frac{\left[P(x_B)e^{U(x_B)/(k_BT)} - P(x_A)e^{U(x_A)/(k_BT)}\right]}{\int_{x_A}^{x_B} e^{U(x)/(k_BT)}dx}. \tag{4.22}$$

We need to evaluate each term in eq.(4.22).

Calculation 3: Evaluation of denominator

We first evaluate the denominator of eq.(4.22),

$$I = \int_{x_A}^{x_B} e^{U(x)/(k_BT)}dx. \tag{4.23}$$

Since the largest contribution to the integration comes from the peak at point C, we evaluate this integral by around $x = x_C$ by approximating

$$U(x) = U(x_C) + \frac{1}{2}U''(x_C)(x - x_C)^2. \tag{4.24}$$

Note that $U'(x_C) = 0$ and $U''(x_C) < 0$ since $U(x)$ take the maximum at $x = x_C$. This gives us

$$\begin{aligned} I &\approx \int_{-\infty}^{\infty} \exp\left[\frac{1}{k_BT}\left(U(x_C) + \frac{1}{2}U''(x_C)(x - x_C)^2\right)\right]dx \\ &= e^{U(x_C)/(k_BT)}\sqrt{2\pi k_BT/|U''(x_C)|}. \end{aligned} \tag{4.25}$$

Calculation 4: Numerator and prob. to be in the well

- The particle is initially in the well
- The probability distribution is quasi-stationary

Thus we have

$$P(x_B) \approx 0, \tag{4.26}$$

and the probability distribution around the well ($x \approx x_A$) is close to the local equilibrium value with an infinitely high barrier at C:

$$P(x) \approx Ne^{-U(x)/(k_BT)} \quad (\text{for} \quad |x - x_A| \ll |x - x_c|). \tag{4.27}$$

Using $P(x_A) \approx Ne^{-U(x_A)/(k_BT)}$, we have

$$P(x) \approx P(x_A)e^{-(U(x)-U(x_A))/(k_BT)} \tag{4.28}$$

with

$$U(x) \approx U(x_A) + (1/2)U''(x_A)(x - x_A)^2. \tag{4.29}$$

In (4.29) we used the fact that the potential takes the minimum at $x = x_A$, i.e. $U'(x_A) = 0$ and $U''(x_A) > 0$.

With this approximation, we can evaluate the probability to be in the well ($-\Delta < x - x_A < \Delta$ with Δ being the order of the size of the well) as

$$\begin{aligned} p &= \int_{x_A-\Delta}^{x_A+\Delta} P(x)dx \\ &\approx \int_{-\infty}^{\infty} P(x_A)\exp\left[-\frac{1}{k_BT}\left(\frac{1}{2}U''(x_A)(x-x_A)^2\right)\right]dx \\ &= P(x_A)\sqrt{\frac{2\pi k_BT}{U''(x_A)}}. \end{aligned} \tag{4.30}$$

Using eqs. (4.26) and (4.30), the numerator of the eq.(4.22) becomes

$$\left[P(x_B)e^{U(x_B)/(k_BT)} - P(x_A)e^{U(x_A)/(k_BT)}\right] = -p\sqrt{\frac{U''(x_A)}{2\pi k_BT}}e^{U(x_A)/(k_BT)}. \tag{4.31}$$

Calculation 5: Value of Flux

Substituting eq.(4.25) and (4.31) to eq.(4.22), we have

$$\begin{aligned} J &\approx D\frac{p\sqrt{\frac{U''(x_A)}{2\pi k_BT}}e^{U(x_A)/(k_BT)}}{e^{U(x_C)/(k_BT)}\sqrt{\frac{2\pi k_BT}{|U''(x_C)|}}} \\ &= p\cdot\frac{D}{2\pi k_BT}\sqrt{U''(x_A)|U''(x_C)|}e^{-(U(x_C)-U(x_A))/(k_BT)} \end{aligned} \tag{4.32}$$

Knowing that $J = pr$, we have the rate for a particle to escape the well r.

4.2.3 Summary of Kramers' escape rate

The rate for a particle to escape from the potential well around point A to point B is given by

$$r = \frac{D}{2\pi k_BT}\sqrt{U''(x_A)|U''(x_C)|}\exp\left[-\frac{U(x_C)-U(x_A)}{k_BT}\right]. \tag{4.33}$$

- The escape rate r falls exponentially with the barrier height $U(x_C) - U(x_A)$.
- The flatness of bottom of the well (around A) and the saddle (around C) affects the escape rate, which comes in as the second-order derivative of the potential.
- The escape rate increase with temperature T. (From FDT we have $D/(k_BT) = 1/\eta$, the mobility, which normally depends on T rather weakly. The main dependence is from the exponential part.)
- Note that this evaluation holds only when $U(x_C) - U(x_A) >> k_BT$.
- This type of calculation can be also extended to a different class of models. A useful example is the calculation of the switching rate of a bi-stable system under noise.

Finally, let us make a rough argument to interpret this result[6]. For this purpose, note that the second-order derivative of potential can be related to the frequency of a particle oscillating under harmonic potential; a particle of mass m under a potential $\frac{1}{2}kx^2$ will oscillate with angular frequency $\omega = \sqrt{k/m}$. Also, remember Einstein's relation $D/(k_BT) = 1/\eta$, and also $1/\gamma = m/\eta$ gives the relaxation time of momentum, or in other words, after this time interval, the motion of a particle will look like a new independent random walk step. Noting these, we can rewrite (4.33) to be

$$r = \frac{\omega_A}{2\pi}\cdot\exp\left[-\frac{U(x_C)-U(x_A)}{k_BT}\right]\cdot\frac{\omega_C}{\gamma}. \tag{4.34}$$

[6] Argument from Kim Sneppen [6, 7].

with

$$\omega_A \equiv \sqrt{U''(x_A)/m}, \quad \omega_C \equiv \sqrt{|U''(x_c)|/m}.$$

The first term, $\omega_A/2\pi$, gives the frequency of the oscillation in the potential well around A, which roughly gives how often the particle tries to climb the wall. The second term is simply the Boltzmann factor, giving the chance to climb the potential difference $U(x_c)-U(x_A)$ all the way up with the aid of thermal fluctuation. The last term is giving the chance of falling to the other side of peak C. The time scale for a particle to fall from the peak is $1/\omega_C$ (Note that at C potential is repulsive, thus $|x-x_c| \propto e^{\omega_C t}$), while the correlation time of the velocity for a Brownian particle is $1/\gamma$. The situation may be viewed as follows: around the peak, the potential is almost flat, but the particle need to perform a random walk to reach $(1/\omega_C)/(1/\gamma)$ steps to the other side of the peak, with absorbing boundaries at the starting point and the endpoint: the particle falls to the B side only if it hits B side of the boundary first. The chance is proportional to the inverse of the number of the steps, ω_C/γ (cf. exercise 1, (4.40)).

4.3 Summary

- The first passage time distribution for the free Brownian motion is calculated using the solution of the diffusion equation under the absorbing boundary condition at $x = 0$, which is

$$P(x,t) = \frac{1}{\sqrt{4D\pi t}}\left[\exp\left(-\frac{(x-a)^2}{4Dt}\right) - \exp\left(-\frac{(x+a)^2}{4Dt}\right)\right].$$

- Resulting distribution obeys $P(T) \propto T^{-3/2}$ large T, which gives a diverging average for the mean first passage time.
- The Kramers' escape gives an estimate of the rate for a particle to go out from a potential well by noise. We make an approximation that system is quasi-stationary and evaluate the flux by focusing on the integral around the maximum and the minimum of the potential.
- The result is

$$r = \frac{D}{2\pi k_B T}\sqrt{U''(x_A)|U''(x_C)|}\exp\left[-\frac{U(x_C)-U(x_A)}{k_B T}\right].$$

4.4 Exercise

1. Show that integral of the first passage time distribution (4.10) from $T = 0$ to ∞ is one, i.e., the walker surely hits $x = 0$ after infinite time.
2. We consider the first passage time distribution of free Brownian particles starting at $x = a$ and absorbed either at $x = 0$ or $x = L$. Namely, the diffusion equation (4.1) needs to be solved under the boundary condition

$$P(0,t) = P(L,t) = 0, \tag{4.35}$$

and with the initial condition (4.5).

(a) Show that the solution is given by

$$P_{0,L}(x,t) = \sum_{n=1}^{\infty}\frac{2}{L}\exp\left[-\left(\frac{n\pi}{L}\right)^2 Dt\right]\cdot\sin\left(\frac{n\pi}{L}a\right)\cdot\sin\left(\frac{n\pi}{L}x\right). \tag{4.36}$$

Hint: Since the boundary condition is in a finite space, Fourier series expansion is useful to solve this problem. Especially, since the boundary condition is absorbing for both sides as (4.35), it is convenient to express the solution by the sine expansion. Remember that the sine expansion of a function $f(x)$ defined in $0 \le x \le L$ is defined as

$$f(x) = \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi}{L}x\right),$$

and the coefficients are given by

$$b_n = \frac{2}{L}\int_0^L f(x)\sin\left(\frac{n\pi}{L}x\right)dx.$$

(b) Let us consider the probability distribution that the particle is absorbed $x = 0$ or $x = L$ at time t, denoted as $P_0(t)$ and $P_L(t)$, respectively. In other words, they are the first passage time distribution at position $x = 0$ and $x = L$, respectively. Show that they are given by

$$P_0(t) = D\sum_{n=1}^{\infty}\frac{2n\pi}{L^2}\exp\left[-\left(\frac{n\pi}{L}\right)^2 Dt\right]\cdot\sin\left(\frac{n\pi}{L}a\right), \tag{4.37}$$

$$P_L(t) = -D\sum_{n=1}^{\infty}\frac{2n\pi}{L^2}\exp\left[-\left(\frac{n\pi}{L}\right)^2 Dt\right]\cdot\sin\left(\frac{n\pi}{L}a\right)\cdot(-1)^n. \tag{4.38}$$

(c) Show that the probability τ_0 (τ_L) that the particle start at $x = a$ is absorbed at $x = 0$ ($x = L$) at *any* time point is given by

$$\tau_0 = \sum_{n=1}^{\infty}\frac{2}{n\pi}\cdot\sin\left(\frac{n\pi}{L}a\right) = \frac{L-a}{L}, \tag{4.39}$$

$$\tau_L = -\sum_{n=1}^{\infty}\frac{2}{n\pi}\cdot\sin\left(\frac{n\pi}{L}a\right)\cdot(-1)^n = \frac{a}{L}. \tag{4.40}$$

Since the particle will eventually hit the boundary, this satisfies $\tau_0 + \tau_L = 1$. The τ_0 (τ_L) is the probability for a particle to hit $x = 0$ ($x = L$) before it hits $x = L$ ($x = 0$).

Hint: What is the sine expansion of the function $f(x) = (L - x)/L$ and $f(x) = x/L$?

(d) Show that the solution (4.36) converge to (4.8) in $L \to \infty$ limit.

Hint 1: The sum will become integral in this limit.

Hint 2: You can use

$$\sin\alpha\sin\eta = -\frac{1}{2}\left[\cos(\alpha+\eta) - \cos(\alpha-\eta)\right].$$

Hint 3:

$$\int_0^\infty e^{-ay^2}\cos(by)dy = \Re\int_0^\infty e^{-ay^2}e^{iby}dy,$$

where $\Re$ means the real part. Remember the calculation for exercise 3 in chapter 1. Pay attention to the integration range.

3. Consider a Brownian particle in a potential described by

$$\dot{X} = -(1/\eta)U'(X) + R(t), \quad \langle R(t)\rangle = 0, \quad \langle R(t)R(t')\rangle = 2D\delta(t-t'),$$

with a potential

$$U(x) = -\frac{1}{3}ax^3 + \frac{1}{2}bx^2$$

($a > 0$, $b > 0$). The shape of this potential around $x = 0$ is $U(x) \approx (1/2)bx^2$.

(a) How does the potential look like? What is the expected escape rate for the particles initially in the well around $x = 0$ based on the Kramers' escape calculation? Here we define the destination of the escape x_B as the position where $U(x_B) = 0$ with $x_B \neq 0$.

(b) * Numerically integrate the Langevin equation with $\eta = 1, D = 1$ (this means $k_BT = 1$), $a = 0.2$, and $b = 1$, and starting from $X(0) = 0$. Choose an appropriate size of the time step for integration so that the numerical error is not too big.

i. Plot a few trajectories ($X(t)$ as a function of t) until the particle reaches x_B.

ii. Measure the time for a particle to reach the position of $X = x_B$. Take enough samples and plot the probability density of the escape time. You need to bin the data to get a histogram. Choose the size of the bin properly; the narrower bin requires more samples. Guess an approximate functional form of the distribution. (We will analyze it analytically later in the course.)
hint: Note that, if you set the upper limit of the time to simulate and if some samples did not reach the target within the upper limit, then the collected samples of first passage time do not show real distribution (biased towards shorter time). For the normalization, note that the definition of a probability density $P(x)$ is so that $\int P(x)dx = 1$.

iii. Calculate the average escape time from the samples. The inverse of this quantity is the escape rate. Compare the result with the analytical estimate in exercise (a).

Advanced exercises

4. Let us consider the target search time in a three-dimensional space, following Ref. [7]. The situation we want to consider is the following: There is a particle diffusing in a three-dimensional space of a finite volume V, with the diffusion constant D. We want to find the average time τ that this particle hits its target, which is a sphere of radius ϵ.

(a) Try to guess how τ depends on the quantities in the problem (V, D , and ϵ) based on dimensional analysis.

(b) To calculate this, we will slightly simplify this problem so that calculation becomes easier. Instead of considering the trajectory of one particle in a finite volume V with a reflecting boundary, we assume that the density of the particle at the positions infinitely far away from the target is $1/V$, and consider the diffusion equation. We place the target at the origin and apply the absorbing boundary condition at the surface of the target.
Then what we need to solve is the diffusion equation for the particle density $\rho(\vec{r}, t)$

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = D\nabla^2 \rho(\vec{r}, t) \tag{4.41}$$

under the boundary condition

$$\rho(|\vec{r}| = \epsilon, t) = 0, \tag{4.42}$$

$$\lim_{|\vec{r}|\to\infty} \rho(\vec{r}, t) = \frac{1}{V}. \tag{4.43}$$

This will give a constant integrated flux to the target since the density of the particles far away from the target is kept constant. The integral of the diffusion flux at the target surface corresponds to the inverse of the average target search time τ.

i. Write the diffusion equation in the spherical coordinate.

ii. Considering the symmetry, $\rho(\vec{r}, t)$ should depends only on the radial coordinate. Solve the diffusion equation in the steady state.

iii. Integrate the diffusion flux at the target surface and show that the average target search time is given by

$$\tau = \frac{V}{4\pi\epsilon D} \tag{4.44}$$

(c) * Numerically simulate the target search by numerically solving a three-dimensional Langevin equation

$$\dot{\vec{X}} = \vec{\xi}(t),$$

with $\vec{\xi}(t) = (\xi_x(t), \xi_y(t), \xi_z(t))$ with

$$\langle \xi_i(t) \rangle = 0, \quad \langle \xi_i(t)\xi_j(t') \rangle = 2D\delta_{i,j}\delta(t - t'),$$

where i and j takes x, y, or z, and $\delta_{i,j}$ is Kronecker's delta. The finite volume can be simulated by using the reflective boundary condition (i.e., if the particle goes "out" from the boundary at a time step, put it back inside at the position mirrored by the boundary), and the target hitting is when the particle hop "inside" the target. Sample the target search time over the randomly chosen initial condition uniformly outside of the target and calculate the average search time. Try this for various volumes V, and confirm that the target search time matches with what you calculated (4.44) when V is sufficiently large.

4.5 Further reading

This chapter is based on [2, 4, 9, 6, 7].

Chapter 5

Stochastic Integrals

We have learned about the Langevin equation based on a physical interpretation. Mathematically, however, the Langevin equation with the random force with δ-function correlation is not very well defined, because δ-function is too strange. This strangeness becomes apparent especially when the noise is *multiplicative*, i.e. when we consider an equation like

$$\dot{X}(t) = F(X(t)) + G(X(t))R(t).$$

Such an equation appears in several places, both in physical situations and in other fields like *finance* (we see it in the next chapter in detail). In this chapter we learn how to deal with differential equations with stochastic variables, so that we can face systems with multiplicative noise.

5.1 Langevin equation and Wiener process

5.1.1 Examples of the multiplicative noise

Example 1: Kinetic energy of a Brownian particle Remember the Langevin equation in one dimension (2.4): Let me rewrite it as

$$m\dot{V}(t) = -\eta V(t) + \sqrt{2D_u}\hat{R}(t) \tag{5.1}$$

with

$$\langle \hat{R}(t) \rangle = 0, \quad \langle \hat{R}(t)\hat{R}(t') \rangle = \delta(t - t'). \tag{5.2}$$

Now, if you want to know the time evolution of the kinetic energy $E = \frac{1}{2}mV^2$, then you would probably want to write

$$\dot{E}(t) \quad = \quad \frac{d}{dt}\left(\frac{1}{2}mV(t)^2\right) = mV\dot{V} = -\eta V(t)^2 + \sqrt{2D_u}V(t)\hat{R}(t). \tag{5.3}$$

This equation contains a multiplicative noise.

Example 2: Decay process Consider radioactive particles. They decay with a rate (probability per unit time) γ. The average number of radioactive particles will obey the equation

$$\frac{d\langle N(t) \rangle}{dt} = -\gamma \langle N(t) \rangle. \tag{5.4}$$

The actual number $N(t)$ fluctuates, because the process is stochastic. When the number is large, the width of the fluctuation is expected to be proportional to the square root of the number (c.f. the central limit theorem). Then we might be able to describe the fluctuation of the number using the following Langevin equation:

$$\frac{dN(t)}{dt} = -\gamma N(t) + \sqrt{\Gamma N(t)}\hat{R}(t), \tag{5.5}$$

where $\hat{R}(t)$ is the Gaussian white noise with variance one and Γ is a constant. This also has a multiplicative noise.

What we learn in this chapter is that, the multiplication between a function of a stochastic, time-dependent variable ($V(t)$ or $N(t)$) and the Gaussian white noise $\hat{R}(t)$ in the last terms in (5.3) and (5.4) should be interpreted in a mathematically different way! You will soon see what I mean, but for that, we need to start with some preparations.

5.1.2 Standard Brownian motion (Wiener process)

Free Brownian motion is parametrized by the diffusion constant D. The Brownian motion with $D = 1/2$ and starting from 0 at $t = 0$ is called the *standard Brownian motion*, or the *Wiener process*. In the following, we use the Wiener process as the basic process to define the multiplicative noise. Let us call the Wiener process $B(t)$. This satisfies

$$\frac{dB(t)}{dt} = \hat{R}(t), \quad \langle \hat{R}(t) \rangle = 0, \quad \langle \hat{R}(t)\hat{R}(t') \rangle = \delta(t - t'), \tag{5.6}$$

and $B(0) = 0$, with $\hat{R}(t)$ being a Gaussian white noise.

5.1.3 Langevin equation and Wiener process

We consider a Langevin equation with a multiplicative noise

$$\dot{X}(t) = F(X(t)) + G(X(t))\hat{R}(t), \tag{5.7}$$

$$\langle \hat{R}(t) \rangle = 0, \quad \langle \hat{R}(t)\hat{R}(t') \rangle = \delta(t - t'). \tag{5.8}$$

This can be re-written by using the Wiener process as follows:

$$\frac{dX(t)}{dt} = F(X(t)) + G(X(t))\frac{dB(t)}{dt}, \tag{5.9}$$

simply because

$$\frac{dB(t)}{dt} = \hat{R}(t). \tag{5.10}$$

When we want to integrate this over time, we are tempted to write

$$\int dX(t) = \int F(X(t))dt + \int G(X(t))dB(t). \tag{5.11}$$

We need to think about what this means, especially the last term. To see that it has some special properties, we start by looking at how $B(t)$ behaves.

5.1.4 Properties of the Wiener Process

Change over small time

Let us define

$$\Delta B(t) \equiv B(t + \Delta t) - B(t).$$

This is of course

$$\Delta B(t) = \int_t^{t+\Delta t} \dot{B}(s)ds = \int_t^{t+\Delta t} \hat{R}(s)ds,$$

and we already know that $\Delta B(t)$ is a random number that obeys the Gaussian distribution

$$P(\Delta B) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{\Delta B^2}{2\Delta t}\right]. \tag{5.12}$$

Especially (See 2.63)

$$\langle \Delta B(t) \rangle = 0, \quad \langle \Delta B(t)^2 \rangle = \Delta t. \tag{5.13}$$

Non-differentiable!

The Wiener process $B(t)$ is not smooth and hence *not differentiable*, in the sense that a derivative can take any large value with the probability one, as shown in the following.

The probability that $|\Delta B(t)/\Delta t|$ takes a value bigger than $A > 0$ in the limit of $\Delta t \to 0$ is given by:

$$\begin{aligned} & \lim_{\Delta t\to 0} 2\int_{A\Delta t}^{\infty} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{(\Delta B)^2}{2\Delta t}\right] d\Delta B \\ = & \lim_{\Delta t\to 0} 2\int_{A\sqrt{\Delta t}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] du \quad (\text{we set } u = \Delta B/\sqrt{\Delta t}) \\ = & 2\int_0^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] du = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] du = 1. \end{aligned}$$

This holds for any $A > 0$. Namely, the derivative is bigger than any $A > 0$ with probability 1.

So, mathematically speaking, we *cannot* write the equation like (5.10) since it is not well-defined. However, as we have seen in the previous chapters, $R(t)$ has physical interpretations and it has been used a lot. We did not have problems so far, because we have been only considering the integrated equations, and $\Delta B(t)$, integral of $\hat{R}(t)$, is well defined. In a sense, $dB/dt = \hat{R}(t)$ is not good, but $dB(t) = \hat{R}(t)dt$ is okay.

With this spirit, we will soon redefine a Langevin equation as a better-defined stochastic differential equation with a proper notation. I here just note that (especially in physics) we still use the notation such as $dB/dt = R(t)$, but when we see it we are supposed to interpret it as the corresponding stochastic differential equation.

What dB^2 means

Where does this strange behaviour come from? It becomes clearer when we think about an integral $\int_0^t (dB)^2$, which we define by using the quadrature rule as

$$\int_0^t (dB)^2 = \lim_{n\to\infty} \sum_{k=1}^{n} \left(B\left(\frac{k}{n}t\right) - B\left(\frac{k-1}{n}t\right)\right)^2. \tag{5.14}$$

It is possible to show

$$\int_0^t (dB)^2 = \lim_{n\to\infty} \sum_{k=1}^{n} \Delta B_k^2 = t = \int_0^t dt, \tag{5.15}$$

where we define $\Delta B_k \equiv B\left(\frac{k}{n}t\right) - B\left(\frac{k-1}{n}t\right)$. (5.15) can be shown in the sense of

$$\lim_{n\to\infty} \left\langle \left[\left(\sum_{k=1}^{n} \Delta B_k^2 \right) - t \right]^2 \right\rangle = 0. \tag{5.16}$$

Namely the variance in the difference between the two is zero, and hence eq. (5.15) holds with probability one. The proof is left for the exercise.

The relation (5.15) indicates

$$dB^2 = dt, \tag{5.17}$$

which is a very important relation to understand the stochastic integrals. Also, this tells us that $(dB/dt)^2 = dt/dt^2 = 1/dt$, which would diverge with $dt \to 0$, thus clearly we cannot take the derivative of $B(t)$.

Note that, from (5.13), we have

$$\langle \Delta B_k^2 \rangle = \frac{t}{n}, \tag{5.18}$$

therefore

$$\left\langle \int_0^t (dB)^2 \right\rangle = \left\langle \lim_{n\to\infty} \sum_{k=1}^{n} \Delta B_k^2 \right\rangle = \lim_{n\to\infty} \sum_{k=1}^{n} \langle \Delta B_k^2 \rangle = \lim_{n\to\infty} n \cdot \frac{t}{n} = t. \tag{5.19}$$

(5.15) is stronger than this, because no average is taken in (5.15).

5.2 Ito integral, Stratonovich integral, and Stochastic Differential Equation (SDE)

Having these strange properties of $dB(t)$ in mind, let us think about what integral means.

5.2.1 Integral of the form $\int f(s)dg(s)$ (Stieltjes integral)

For a smooth $g(s)$, an integral $\int f(s)dg(s)$ is well defined by the limit of the quadrature rule. For example, one can do as shown in Fig. 5.1 to the left, which corresponds to

$$\int_{s=0}^{s=t} f(s)\cdot dg(s) \equiv \lim_{n\to\infty}\sum_{k=1}^{n} f\left(\frac{k-1}{n}t\right)\times\left[g\left(\frac{k}{n}t\right)-g\left(\frac{k-1}{n}t\right)\right], \tag{5.20}$$

or one can do as shown in Fig. 5.1 to the right, which corresponds to

$$\int_{s=0}^{s=t} f(s)\circ dg(s) \equiv \lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{2}\left[f\left(\frac{k-1}{n}t\right)+f\left(\frac{k}{n}t\right)\right]\times\left[g\left(\frac{k}{n}t\right)-g\left(\frac{k-1}{n}t\right)\right]. \tag{5.21}$$

Of course, (5.20) and (5.21) converge to the same value for smooth functions.

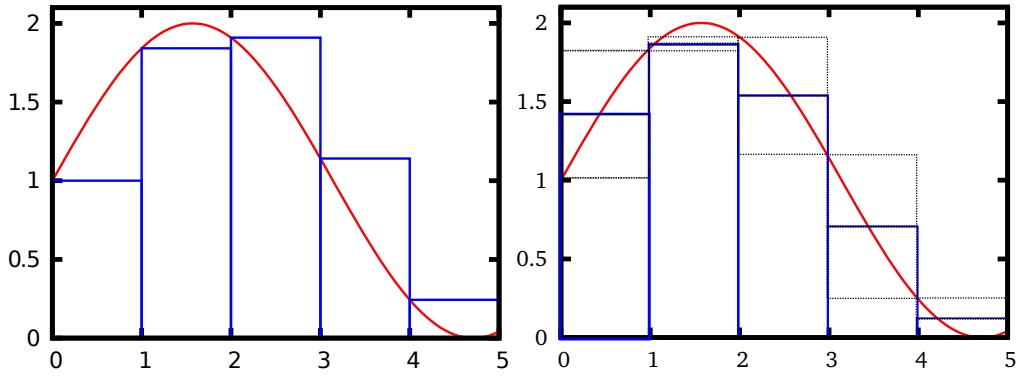


Figure 5.1: Quadrature rules.

Difference between the two discretizations

Let us see how the difference between (5.20) and (5.21) goes to zero as $n\to\infty$. For simplicity, let us look at an example where $f(s)=g(s)$.

$$\int_{s=0}^{s=t} g(s)\cdot dg(s) - \int_{s=0}^{s=t} g(s)\circ dg(s) = -\frac{1}{2}\lim_{n\to\infty}\sum_{k=1}^{n}\left[g\left(\frac{k}{n}t\right)-g\left(\frac{k-1}{n}t\right)\right]^2. \tag{5.22}$$

For a smooth function,

$$\Delta g_k \equiv \left[g\left(\frac{k}{n}t\right)-g\left(\frac{k-1}{n}t\right)\right] = g'\left(\frac{k-1}{n}t\right)\frac{t}{n}+O\left(\left(\frac{t}{n}\right)^2\right). \tag{5.23}$$

$$(5.22) = -\frac{1}{2}\lim_{n\to\infty}\sum_{k=1}^{n}\left[g'\left(\frac{k-1}{n}t\right)^2\left(\frac{t}{n}\right)^2+O\left(\left(\frac{t}{n}\right)^3\right)\right] \tag{5.24}$$

Sum of n terms that are proportional to $(t/n)^2$ gives $O(t^2/n)$ contribution, which goes to zero for $n\to\infty$.

5.2.2 Stieltjes integral for the Wiener process

The Stieltjes integral with the Wiener process has properties very different from those for smooth functions. As an example, consider the two ways of defining $\int_0^t B(s)dB(s)$:

Case 1 (Following the *Ito integral*):

$$\int_{s=0}^{s=t} B(s)\cdot dB(s) \equiv \lim_{n\to\infty}\sum_{k=1}^{n} B\left(\frac{k-1}{n}t\right)\left[B\left(\frac{k}{n}t\right)-B\left(\frac{k-1}{n}t\right)\right] \tag{5.25}$$

Case 2 (Following the *Stratonovich integral*):

$$\int_{s=0}^{s=t} B(s)\circ dB(s) \equiv \lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{2}\left[B\left(\frac{k}{n}t\right)+B\left(\frac{k-1}{n}t\right)\right]\left[B\left(\frac{k}{n}t\right)-B\left(\frac{k-1}{n}t\right)\right]. \tag{5.26}$$

Difference of the two integrals for the Wiener process

Actually (5.25) and (5.26) give different values!

$$\int_{s=0}^{s=t} B(s)\cdot dB(s)-\int_{s=0}^{s=t} B(s)\circ dB(s) = -\frac{1}{2}\lim_{n\to\infty}\sum_{k=1}^{n}\left[B\left(\frac{k}{n}t\right)-B\left(\frac{k-1}{n}t\right)\right]^2 = -\frac{1}{2}t, \tag{5.27}$$

where in the last equality we used (5.15). The fact that $dB^2 = dt$ gives finite difference between different discretizations, because the 2nd order contribution of dB is comparable with the first-order contribution of dt.

Interestingly, by calculating the sum in (5.26) explicitly, we can show that

$$\int_{s=0}^{s=t} B(s)\circ dB(s) = \lim_{n\to\infty}\frac{1}{2}\left[B\left(\frac{1}{n}t\right)^2 - B(0)^2 + \cdots + B\left(\frac{n}{n}t\right)^2 - B\left(\frac{n-1}{n}t\right)^2\right] = \frac{1}{2}\left[B(t)^2 - B(0)^2\right]. \tag{5.28}$$

The Stratonovich integral matches with the results that we would get for smooth functions.

5.2.3 The Ito integral and the Stratonovich integral

$\int f(s)dB(s)$ is not uniquely defined, as we just saw in the examples of two famous interpretations. In principle, one can consider other interpretations, but these two are the most useful and widely used.

For Ito's definition, we write

$$f(s)[B(s+\Delta s)-B(s)] \to f(s)\cdot dB(s), \tag{5.29}$$

and for Stratonovich's definition

$$\frac{1}{2}\left[f(s+\Delta s)+f(s)\right]\left[B(s+\Delta s)-B(s)\right] \to f(s)\circ dB(s). \tag{5.30}$$

Accordingly, the Ito integral is defined as

$$\int_{s=0}^{s=t} f(s)\cdot dB(s) \equiv \lim_{n\to\infty}\sum_{k=1}^{n} f\left(\frac{k-1}{n}t\right)\left[B\left(\frac{k}{n}t\right)-B\left(\frac{k-1}{n}t\right)\right] \tag{5.31}$$

and the Stratonovich integral is defined as

$$\int_{s=0}^{s=t} f(s)\circ dB(s) \equiv \lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{2}\left[f\left(\frac{k}{n}t\right)+f\left(\frac{k-1}{n}t\right)\right]\left[B\left(\frac{k}{n}t\right)-B\left(\frac{k-1}{n}t\right)\right]. \tag{5.32}$$

Throughout this lecture note, I will distinguish the two interpretations with $\cdot$ and $\circ$, and if there are no symbols it means an ordinary product (e.g. I would write $f(s)\circ dB(s) = \frac{1}{2}[f(s+ds)+f(s)]dB(s)$). Note that the distinction of $\cdot$ and $\circ$ is not a standardized notation; In literature, if a stochastic integral appears, which interpretation they take will normally be explicitly stated.

5.2.4 Stochastic differential equation (SDE)

Consider a stochastic process constructed from the Wiener process $B(t)$:

$$dX(t) = F(X(t))dt + G(X(t))dB(t). \tag{5.33}$$

Now we know that this does not give the full description of the stochastic process. If the last term is interpreted as the Ito type, we have

$$dX(t) = F(X(t))dt + G(X(t)) \cdot dB(t). \tag{5.34}$$

If it is interpreted as the Stratonovich type, we have

$$dX(t) = F(X(t))dt + G(X(t)) \circ dB(t). \tag{5.35}$$

And as mentioned, other interpretations are also possible, and a different interpretation describes a different stochastic process.

Such equations are called stochastic differential equations (SDEs).

SDE and Langevin equation

If we divide both sides of the SDEs formally with dt, the equation will look like the Langevin equation. (Historically, it is the other way around - physicists invented and studied the Langevin equations and mathematicians established the meaning of it by giving a precise definition as SDEs.)

Mathematically, the Langevin equation should be multiplied by dt and interpreted as an SDE:

$$\dot{X}(t) = F(X(t)) + G(X(t))\hat{R}(t).$$

means

$$dX(t) = F(X(t))dt + G(X(t))dB(t),$$

where it should be decided whether we interpret the last term as the Ito type or the Stratonovich type (or something else), *depending on the phenomenon that the equation is supposed to describe.*

SDE and Langevin equation with additive noise

In the last chapters, we have been considering

$$\dot{X}(t) = F(X(t)) + R(t), \tag{5.36}$$

where $R(t)$ is a Gaussian white noise with zero means, satisfying

$$\langle R(t_1)R(t_2)\rangle = 2D\delta(t - t_2). \tag{5.37}$$

This is equivalent to set $R(t) = \sqrt{2D}\hat{R}(t)$, namely

$$\dot{X}(t) = F(X(t)) + \sqrt{2D}\hat{R}(t), \tag{5.38}$$

which can be interpreted as either

$$dX(t) = F(X(t))dt + \sqrt{2D} \cdot dB(t), \tag{5.39}$$

or

$$dX(t) = F(X(t))dt + \sqrt{2D} \circ dB(t), \tag{5.40}$$

but they are the same as long as $\sqrt{2D}$ is a constant independent of $X(t)$. This is why we did not care about interpretations.

5.3 Useful relations for SDE calculus

5.3.1 Average for Ito and Stratonovich product

Consider an average of the Ito-type product

$$\langle G(X(t)) \cdot dB(t) \rangle = \langle G(X(t))(B(t+dt) - B(t)) \rangle \tag{5.41}$$

From the causality, the value of $G(X(t))$ does not depend on $dB(t)$, which is the sum of "random force" in the future (between t and $t+dt$). This means

$$(5.41) = \langle G(X(t)) \rangle \langle (B(t+dt) - B(t) \rangle = 0. \tag{5.42}$$

On the other hand, we cannot factorize the Stratonovich product.

$$\langle G(X(t)) \circ dB(t) \rangle = \left\langle \frac{1}{2}(G(X(t+dt) + G(X(t)))(B(t+dt) - B(t)) \right\rangle, \tag{5.43}$$

because the value of X at the time $t+dt$ is clearly correlated with $dB(t)$.

5.3.2 Difference of a function of stochastic variable

solving ODE Consider an ordinary differential equation (ODE)

$$\frac{dx(t)}{dt} = x(t). \tag{5.44}$$

We all know how to solve (i.e., express $x(t)$ as an explicit function of t). One way is to use the relation

$$\frac{dx}{dt} = x \quad \Rightarrow \quad \frac{1}{x} dx = dt. \tag{5.45}$$

This gives you the solution as

$$\int \frac{1}{x} dx = \int dt \quad \Rightarrow \quad \int d \ln x = \int dt \quad \Rightarrow \quad \ln x = t + c, \tag{5.46}$$

where c is a integration constant. Hence the solution is

$$x(t) = Ce^t, \tag{5.47}$$

where $C = e^c$ is a constant.

Note that in the integral in eq. (5.46), we have used

$$d \ln x(t) = \frac{1}{x(t)} dx(t). \tag{5.48}$$

If we define $f(x) = \ln x$, then $f'(x) = 1/x$, so this relation can be rewritten as

$$df(x(t)) = f'(x(t)) dx(t). \tag{5.49}$$

We use this relation (5.49) all the time when we solve ODEs, though sometimes we may not be fully aware of it. But, does this relation still hold for a function of a stochastic variable?

Define a difference of a function Consider a function $f(X(t))$ of a stochastic variable $X(t)$ that obeys an SDE. How does the difference

$$df(X(t)) \equiv f(X(t) + dX(t)) - f(X(t)) \tag{5.50}$$

look like up to order dt?

Note that we can always Taylor expand $df(X(t))$ as

$$\begin{aligned} df(X(t)) &\equiv f(X(t)+dX(t)) - f(X(t)) \\ &= f'(X(t))dX(t) + \frac{1}{2!}f''(X(t))(dX(t))^2 + O(dX^3). \end{aligned} \tag{5.51}$$

We will use (5.17)

$$dB(t)^2 = dt \tag{5.52}$$

without giving a more rigorous proof[1] - interested readers can find it in ref. [10]. This means that, to take $df(X(t))$ up to the order dt, we cannot ignore dX^2 term because it contains a dB^2 term.

Stratonovich expression of df

It is tempting to say something like $df(X(t)) = f'(X(t)) \circ dX$ or $df(X(t)) = f'(X(t)) \cdot dX$. Are these relations correct?

Let's take the difference between LHS and RHS. In the Stratonovich case:

$$\begin{aligned} & df(X(t)) - f'(X(t)) \circ dX(t) \\ = \; & f'(X(t))dX(t) + \frac{1}{2!}f''(X(t))(dX(t))^2 - \frac{1}{2}[f'(X(t+dt)) + f'(X(t))]dX(t) + O(dX(t)^3) \\ = \; & f'(X(t))dX(t) + \frac{1}{2!}f''(X(t))(dX(t))^2 - \frac{1}{2}[f'(X(t)) + f''(X(t))dX + f'(X)(t)]dX(t) + O(dX(t)^3) = O(dX(t)^3). \end{aligned}$$

Where the first to second line the definition of the Stratonovich interpretation was used, and from the second to the third line $f'(X(t+dt)) = f'(X+dX)$ the Taylor expansion was used. Therefore, up to order dt, we can say

$$df(X(t)) = f'(X(t)) \circ dX(t). \tag{5.53}$$

Note that this is the "same" as (5.49), just the product is the Stratonovich Stratonovitsh product. So, this means that the Stratonovich integrals can be performed as if they are "ordinary" integrals.

Ito case has an extra term

In the Ito case:

$$\begin{aligned} & df(X(t)) - f'(X(t)) \cdot dX(t) \\ = \; & f'(X)dX + \frac{1}{2!}f''(X)(dX)^2 - f'(X(t))dX + O(dX^3) \qquad (5.54) \\ = \; & \frac{1}{2}f''(X)(dX)^2 + O(dX^3), \qquad (5.55) \end{aligned}$$

where the definition of the Ito interpretation is used from the first line to the second line. So in the Ito case, we have an extra term up to order dX^2 as

$$df(X(t)) = f'(X(t)) \cdot dX(t) + \frac{1}{2}f''(X(t))(dX(t)^2). \tag{5.56}$$

This means Ito integrals need extra care.

Ito's formula When $X(t)$ is a stochastic variable that obeys the SDE of the Ito type (5.34)

$$dX(t) = F(X(t))dt + G(X(t)) \cdot dB(t), \tag{5.57}$$

[1] We have shown $\int dB^2 = \int dt$ but to use this for any integrals $\int f(t)dB^2 = \int f(t)dt$ needs to be shown.

we have with (5.52)

$$\begin{aligned}(dX(t))^2 &= (F(X(t))dt + G(X(t)) \cdot dB(t))^2 \\ &= G(X(t))^2 dt + \text{higher order in } dt. \end{aligned} \tag{5.58}$$

Substituting (5.57) and (5.58) into (5.56), we get *Ito's formula*

$$df(X(t)) = \left[f'(X(t))F(X(t)) + \frac{1}{2}f''(X(t))G(X(t))^2\right]dt + f'(X(t))G(X(t)) \cdot dB(t). \tag{5.59}$$

We use this property to solve (i.e., to express $X(t)$ as an explicit function of t and $B(t)$) Ito SDE in exercise.

5.3.3 Reformulation of a Stratonovich type SDE into an Ito type SDE

Consider a SDE of the Stratonovich type

$$dX(t) = F(X(t))dt + G(X(t)) \circ dB(t).$$

We can express the noise term in the Stratonovich SDE, $G(X(t)) \circ dB(t)$, with the Ito type product as follows.

$$\begin{aligned} G(X(t)) \circ dB(t) &= \frac{1}{2}\left[G(X(t+dt)) + G(X(t))\right] dB(t) \\ &= \frac{1}{2}\left[G(X(t)) + G'(X(t))dX(t) + G(X(t))\right] dB(t) \\ &= G(X(t))dB(t) + \frac{1}{2}G'(X(t))dX(t)dB(t). \end{aligned} \tag{5.60}$$

Here $dX^2 dB$ term was ignored since it is obviously higher order than dt. Considering

$$\begin{aligned} dX(t)dB(t) &= \left[F(X(t))dt + G(X(t)) \circ dB(t)\right] dB(t) \\ &= \frac{1}{2}\left[G(X(t+dt)) + G(X(t))\right] dB(t)^2 + O(dtdB) \\ &= G(X(t))dt + O(dtdB), \end{aligned}$$

we finally get

$$G(X(t)) \circ dB(t) = G(X(t)) \cdot dB(t) + \frac{1}{2}G'(X(t))G(X(t))dt. \tag{5.61}$$

Using (5.61), SDE of Stratonovich type and Ito type can be reformulated as follows:

$$\begin{aligned} dX(t) &= F(X(t))dt + G(X(t)) \circ dB(t) \\ \Leftrightarrow dX(t) &= \left[F(X(t)) + \frac{1}{2}G(X(t))\frac{\partial G(X)}{\partial X}|_{X=X(t)}\right] dt + G(X(t)) \cdot dB(t) \end{aligned} \tag{5.62}$$

5.3.4 The Fokker-Planck equations

The SDE has the subtlety of definition in integral, but the FP equations, which are partial differential equations, do not have any singular terms, and we can work with the mathematical tools that we are used to. In this subsection, we see a conversion between an SDE with multiplicative noise and an FP equation.

The FP equation for the SDE of Ito type

By using (5.59) and the property of the Ito type product, it is easy to derive the corresponding FP equation for the Ito type SDE (5.34). The resulting FP equation is

$$\frac{\partial}{\partial t}P(x,t) = \left[-\frac{\partial}{\partial x}\left(F(x)P(x,t)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(G(x)^2P(x,t)\right)\right] \tag{5.63}$$

This is easy to show by taking average of (5.59) using the property (5.42). We then get

$$\langle df(X(t))\rangle = \langle f'(X(t))F(X(t)) + \frac{1}{2}f''(X(t))G(X(t))^2\rangle dt, \tag{5.64}$$

and it is straightforward to convert this to an FP equation. The detailed derivation of (5.63) is left for exercise.

The FP equation for the SDE of Stratonovich type

Comparing the Ito type SDE and eq.(5.62), and seeing the FP equation for an Ito type SDE (5.63), it is straightforward to derive the FP equation for an SDE of Stratonovich type (5.35):

$$\frac{\partial}{\partial t}P(x,t) = \left[-\frac{\partial}{\partial x}\left(F(x) + \frac{1}{2}G(x)\frac{\partial G(x)}{\partial x}\right)P(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(G(x)^2P(x,t)\right)\right] \tag{5.65}$$

Note that (5.63) and (5.65) are the same when $G(x) = const.$. This again confirms that the difference does not matter as long as the noise is not multiplicative.

It is worth noting that the FP equation does not have "ambiguity" of interpretation of stochastic integrals. Different interpretations give different FP equations, but for a given interpretation there is a unique FP equation that fully describe the stochastic process.

5.4 Numerical integration of SDE

5.4.1 Numerical integration of SDE: review

First, recall the Euler method in section 2.6 to solve a Langevin equation

$$\dot{X}(t) = F(X(t)) + \sqrt{2D}\hat{R}(t), \tag{5.66}$$

with the Euler method, every time step Δt we calculate

$$\Delta X(t) = X(t+\Delta t) - X(t) = F(X(t))\Delta t + \sqrt{2D}\Delta B, \tag{5.67}$$

where ΔB is a random number that obeys the Gaussian distribution

$$P(\Delta B) = \frac{1}{\sqrt{2\pi\Delta t}}e^{-\Delta B^2/(2\Delta t)}. \tag{5.68}$$

Note the correspondence between SDE expression and the Euler method.

Numerical integration of SDE: Ito case

Now consider the Ito-type SDE with multiplicative noise

$$dX(t) = F(X(t))dt + G(X(t)) \cdot dB(t). \tag{5.69}$$

Since the Ito interpretation means

$$G(X(t))[B(t+\Delta t) - B(t)] = G(X(t))\Delta B \to G(X(t)) \cdot dB(t), \tag{5.70}$$

the Euler method should be interpreted as

$$\Delta X(t) \approx F(X(t))\Delta t + G(X(t))\Delta B(t). \tag{5.71}$$

This is called the Euler-Maruyama method.

Numerical integration of SDE: Stratonovich case

The Euler-Maruyama method cannot be used for the Stratonovich case straightforwardly because Stratonovich integral corresponds to

$$\Delta X(t) \approx F(X(t))\Delta t + \frac{1}{2}[G(X(t+\Delta t)) + G(X(t))]\Delta B(t), \tag{5.72}$$

where the RHS contains the value in the future. However, we can reformulate a Stratonovich SDE to Ito SDE and then use the Euler-Maruyama method.

Another way to integrate Stratonovich-type SDE is to use the Euler-Heun method, where the procedure is given by

$$\tilde{X}(t) \equiv X(t) + G(X(t))\Delta B(t), \tag{5.73}$$

$$\Delta X(t) = F(X(t))\Delta t + \frac{1}{2}[G(\tilde{X}(t)) + G(X(t))]\Delta B(t). \tag{5.74}$$

(Note that $\Delta B(t)$ in (5.73) and (5.74) are the same.) We can show this corresponds to (5.72) as follows. Using (5.73),

$$\begin{aligned} G(\tilde{X}(t)) &= G[X(t) + G(X(t))\Delta B] \\ &= G(X(t)) + \frac{\partial G(X)}{\partial X}|_{X=X(t)} G(X(t))\Delta B + O(\Delta B^2), \end{aligned}$$

therefore (5.74) gives, up to the order of $\Delta t \sim \Delta B^2$,

$$\begin{aligned} \Delta X(t) &= F(X(t))\Delta t + G(X(t))\Delta B \\ &\quad + \frac{1}{2}\frac{\partial G(X)}{\partial X}|_{X=X(t)} G(X(t))\Delta B^2 + \text{higher order}. \end{aligned} \tag{5.75}$$

We can see that this corresponds to (5.72) up to the order Δt by considering the small time step limit $\Delta t \to dt$ and $\Delta B \to dB$, where we have $\Delta B^2 \to dB^2 = dt$.

5.5 Ito or Stratonovich?

Now we know that an SDE needs to be carefully interpreted mathematically when it contains multiplicative noise. But how should *we* decide which one to use for a given problem?

It, of course, depends on the problem. Intuitively, I would say that if we consider a Langevin equation based on a real physical system with continuous variables, then noise with delta-function correlation is always an approximation of the noise with finite time correlation - i.e., it should be a smooth function of time if the time resolution is taken to be comparable with the noise correlation time scale. And if the functions are smooth, we deal with all the integral as normal functions, and that will give the result corresponding to the Stratonovich interpretation. Therefore, in such a case the Stratonovich interpretation should be used[2].

If the original variables are discrete (e.g. number of molecules), the Ito interpretation is often more appropriate. Later in the course, we learn how to deal with such a case more systematically.

In this section, we take a few examples to see which interpretation is appropriate in different cases.

5.5.1 Decay process

Let me first take our second example at the beginning of this chapter, the decay process. Intuitively one thinks that (5.5) should be interpreted as the Ito type

$$dN(t) = -\gamma N(t)dt + \sqrt{\Gamma N(t)} \cdot dB, \tag{5.76}$$

[2]However, we need to be careful if we assume *more than one time scales are much shorter than the problem of interest.* Then the appropriate calculus may depend on which time scale is shorter (i.e., the results may depend on which time scale one takes the zero limit first: known as the Ito-versus-Stratonovich problem [11]). This is already a subtle problem when we ignore the inertia to obtain the over-damped equations for the system with multiplicative noise; here, the two-time scales are the noise correlation time and the momentum relaxation time.

because the number of particles that decays in the time interval dt is proportional to the number of particles BEFORE the decay; therefore, noise should be related to the number of particles at t, not at $t + dt$. This fits Ito's interpretation. This is also supported by checking that the average of (5.76) will recover (5.4), by using (5.42).

5.5.2 Kinetic energy of a Brownian particle

Now consider our first example in this chapter, the kinetic energy of a Brownian particle, where we wrote (5.3) AS IF there is nothing strange about differentiation.

Since we are modelling a real physical system with continuous variables in continuous time, we expect that we should use the Stratonovich integrals to treat this quantity. Then it also means that all the calculus looks the same as usual integrals and differentiation. Namely, (5.3) should be interpreted as

$$dE(t) = -\eta V(t)^2 dt + \sqrt{2D_u} V(t) \circ dB \tag{5.77}$$

where the last term represents the Stratonovich interpretation.

If instead we used the Ito interpretation $\sqrt{2D_u}V(t) \cdot dB(t)$ in the last term, then taking an average of both sides would give us $\langle dE(t) \rangle = -\eta \langle V(t)^2 \rangle dt$. The RHS of this equation is zero or negative, so this equation would wrongly predict that energy will decay to zero.

It is left for exercise to use the correct expression (5.77) and show that

$$\langle dE(t) \rangle = \left(-\eta \left\langle V(t)^2 \right\rangle + \frac{D_u}{m} \right) dt, \tag{5.78}$$

which gives the correct relation eq. (2.13) in the steady state.

5.5.3 Stochastic Thermodynamics / Energetics

When we learn thermodynamics, we normally consider the *thermodynamic limit*, where the system size is so large that it is enough to consider macroscopic behaviours. But how big the system needs to be? What if the system scale is very small that we can see thermal fluctuation? Do we still have thermodynamics in such a fluctuating world? Therefore we pose the following question: How does thermodynamics look in the world described by the Langevin equation of an over-damped Brownian particle?

The first law of thermodynamics

The first law of thermodynamics says that the change of the *internal energy* of the system ΔE is the sum of the *work* done to the system ΔW and the heat transferred from the environment to the system ΔQ:

$$\Delta E = \Delta W + \Delta Q. \tag{5.79}$$

How should this look in the Langevin world?

Setup: Potential controllable from the external world

Our "system" now is a colloidal particle, whose mass is negligible. We also want to do "work" to the system, and an easy way of doing it is changing the potential applied to the particle - e.g., by moving optical tweezers. Then the potential has a “controller”, that we can move to change the parameters in the potential. We symbolically denote the effect of the controller with a parameter a. Now our potential is the function of the controller parameter a and the position; $U(x; a)$.

The Langevin equation that the particle obeys would look like this:

$$0 = -\eta V(t) - \frac{\partial U(X; a)}{\partial X} + R(t). \tag{5.80}$$

Temperature of the system

We do some operations on the system, but we assume that the environment (e.g. water) is large enough that the nature of the random force is unaffected by the operations, and the environment is characterized by a temperature T.

This means that we assume that Einstein's relation holds all the time:

$$\langle R(t)R(t')\rangle = 2\eta k_B T\delta(t-t'). \tag{5.81}$$

The RHS of the (5.80) has three force terms. Which terms will contribute to the heat exchange between the system and the environment?

Work and heat

If we think that the work done to the system is only through the controller, i.e., by changing the parameter a, it is natural to say

$$dW \equiv \frac{\partial U}{\partial a}\circ da. \tag{5.82}$$

How about the heat from the environment to the system? Considering $(-\eta V(t)+R(t))$ is the force from the environment to the system via thermal fluctuation, it would physically make sense to define

$$dQ \equiv (-\eta V(t)+R(t))\circ dX = \frac{\partial U}{\partial X}\circ dX, \tag{5.83}$$

because dQ is "the work done from the environment to the system through drag force and the random force". Note that we use the Stratonovich interpretation, because we are dealing with a physical system, where the random force is an approximation of the forces with short-time correlations.

Energy balance

When the inertia is negligible, the internal energy of the system is only the potential energy. The change of the internal energy is then given by

$$dU = \frac{\partial U}{\partial a}\circ da + \frac{\partial U}{\partial X}\circ dX \tag{5.84}$$

Here I use the fact that we can expand functions as if they are smooth by using the Stratonovich interpretation.

Compared with (5.82) and (5.83), we recover the first low of thermodynamics

$$dU = dW + dQ, \tag{5.85}$$

where dU is the change of the potential energy, dW is work done from outside, and dQ is heat received by the system from the environment. Note that this is defined for one fluctuating trajectory!

Stratonovich expression matches with physical insight

Consider the situation where we do not move the controller – i.e. $da=0$, and we do not do any work to the system. In this case, we have

$$dU = dQ = \frac{\partial U}{\partial X}\circ dX. \tag{5.86}$$

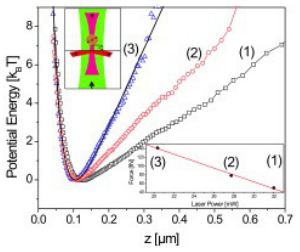
At the same time, by definition

$$dU = U(X+dX;a) - U(X;a). \tag{5.87}$$

Note that this is equal to $\frac{\partial U}{\partial X}\circ dX$ but NOT $\frac{\partial U}{\partial X}\cdot dX$ (Remember Ito's formula). So that the first equation, which physically makes sense, holds, the product in the definition of heat (5.83) should be treated as Stratonovich type.

Experimental verification

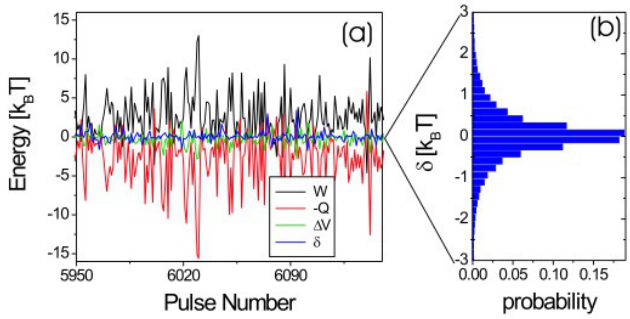
The first law of thermodynamics for fluctuating system has been verified experimentally in 2006 by using a colloidal particle and optical tweezers that applies an externally controlled potential force [12].



$$\gamma\dot{z} = -\frac{\partial V}{\partial z} + \xi,$$

$$dW = \frac{\partial V}{\partial \tau}d\tau, \qquad dQ = -\frac{\partial V}{\partial z}dz.$$

$$W[z(\tau)] - Q[z(\tau)] - \Delta V = 0$$



.

5.6 Summary

- Mathematically, a Langevin equation
$$\dot{X}(t) = F(X(t)) + G(X(t))\hat{R}(t)$$
should be interpreted as a stochastic differential equation
$$dX(t) = F(X(t))dt + G(X(t))dB(t),$$
with $B(t)$ being the Wiener process.
- $dB^2 = dt$. This makes integrals involving dB different from integrals with smooth functions because we cannot ignore dB^2 terms.
- $f(s)dB(s)$ needs to be defined carefully. Often-used definitions are
  - Ito interpretation:
$$f(s)[B(s+\Delta s) - B(s)] \to f(s)\cdot dB(s),$$
  - Stratonovich interpretation:
$$\frac{1}{2}\left[f(s+\Delta s) + f(s)\right]\left[B(s+\Delta s) - B(s)\right] \to f(s)\circ dB(s).$$
- For the Stratonovich interpretation, we have a "usual" relation
$$df(X(t)) = f'(X(t))\circ dX(t).$$
- For an Ito-type SDE, Ito's formula holds:
$$df(X(t)) = \left[f'(X(t))F(X(t)) + \frac{1}{2}f''(X(t))G(X(t))^2\right]dt + f'(X(t))G(X(t))\cdot dB(t).$$
- Reformulation of a Stratonovich-type SDE to an Ito-type SDE is possible.
- The Fokker-Planck equations for both types of SDEs derived.

5.7 Exercise

1. Show (5.16) following the instruction below.

   (a) Confirm

$$\left\langle \left(\sum_{k=1}^{n} \Delta B_k^2 - t \right)^2 \right\rangle = \left\langle \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} \Delta B_k^2 \Delta B_l^2 \right\rangle + \left\langle \sum_{k=2}^{n} \sum_{l=1}^{k-1} \Delta B_k^2 \Delta B_l^2 \right\rangle + \left\langle \sum_{k=1}^{n} \Delta B_k^4 \right\rangle - 2t \left\langle \sum_{k=1}^{n} \Delta B_k^2 \right\rangle + t^2. \tag{5.88}$$

   (b) ΔB_k and ΔB_l and are independent for $k \neq l$. For $k \neq l$, show

$$\langle \Delta B_k^2 \Delta B_l^2 \rangle = \left(\frac{t}{n} \right)^2. \tag{5.89}$$

   (c) Combining these, show that

$$\left\langle \left(\sum_{k=1}^{n} \Delta B_k^2 - t \right)^2 \right\rangle = 2\frac{t^2}{n}. \tag{5.90}$$

   You can use the fact that $\langle X^4 \rangle = 3\langle X^2 \rangle^2$ for a Gaussian distributed variable X with average zero.

   The RHS goes to zero when $n \to \infty$, which is the relation (5.16).

2. Consider a Langevin equation with multiplicative noise.

$$\frac{dX}{dt} = X\hat{R}(t)$$

   It can be interpreted as a Stratonovich type SDE

$$dX = X \circ dB, \tag{5.91}$$

   or an Ito-type SDE

$$dX = X \cdot dB. \tag{5.92}$$

   (a) Solve the SDE (5.91) to derive

$$X(t) = Ce^{B(t)},$$

   where C is an integration constant.
   Hint. Show that given eq. (5.91), $\frac{1}{X} \circ dX = dB$ holds up to dt. We have eq. (5.53).

   (b) Solve the SDE (5.92) to derive

$$X(t) = Ce^{-t/2+B(t)},$$

   where C is an integration constant.
   Hint. Define $f(X) = \ln X$ and use Ito's formula to convert eq. (5.92) into an SDE for $f(X)$.

3. Derive (5.63). Assume necessary boundary conditions, e.g., $P(x,t) \to 0$ and $\partial P(x,t)/\partial x \to 0$ for $x \to \pm\infty$ etc.

4. Decay process

   (a) Write down the corresponding SDE with the Stratonovich interpretation for the model (5.5).

   (b) Take the average of SDE from (a) and derive an equation for the average number $\langle N(t) \rangle$. Confirm that it does not recover (5.4).

   (c) Write down the FP equation that corresponds to the Ito interpretation SDE (5.76). We will come back to this equation later when we consider the decay process more systematically.

5. Derive (5.78) from (5.77). Hint: Note that the LHS of (5.77) is $dE(t)$, while the stochastic variable in the RHS is $V(t)$, not $E(t)$.

Advanced exercises

6. Numerical comparison of an Ito SDE and a Stratonovitch SDE

   (a) Numerically simulate an Ito SDE
   $$dX = X \cdot dB$$
   using the Euler-Maruyama method with the initial condition $X(0) = 1$, up to time 10. Average over 100 samples and plot the average trajectory $\langle X(t) \rangle$.

   (b) Numerically simulate a Storatonovitch SDE
   $$dX = X \circ dB$$
   using the Euler-Heun method with the initial condition $X(0) = 1$, up to time 10. Average over 100 samples and plot the average trajectory $\langle X(t) \rangle$.

   (c) Consider if the results in (a) and (b) are consistent with what you expect from the result of exercise 5-2.

5.8 Further reading

This chapter is based on [2, 13, 10]. In my experience, this level of description is good enough for most of the physics problems, but if one wants to learn more modern probability theory, which makes things more rigorous, then one can refer to [4] chapter 10 (which touches the modern terminologies without going into it) and references therein (e.g. [14]).

Chapter 6

Geometric Brownian motion as a model of a financial market

6.1 Geometric Brownian motion

Stochastic differential equations are used widely in the modelling of financial markets. One of the most famous equations is the *Black-Scholes equation*, used to determine the price of a *European call option*, which is to buy one share of the stock S at a price p at future time T.

We do not go into the Black-Scholes equation itself (You can find derivation in many places, including ref. [4]), but it is easy to imagine that the central part of the modelling is to guess the stock price S fluctuation. The often-used model is very simple as follows.

6.1.1 Model of the Stock price

Every year the price will go up at a certain rate, but it will also fluctuate. One may expect that the fluctuation would be larger when the price is high - when a share is worth 1 dollar, the price may move by 10 cents over a week, but when a share is worth 10 dollars, the price move by 1 dollar over the same period. Then one may write down a stochastic differential equation

$$dS(t) = \mu S(t)dt + \sigma S(t) \cdot dB(t), \tag{6.1}$$

where $B(t)$ is the Wiener process. μ is the expected rate of return and σ is called volatility, and they are both constants. Conventionally this equation is interpreted by the Ito interpretation, which makes sense because individual trades happen at discrete timing.

Note that $dB(t)$ is a Gaussian noise, therefore there is no extreme events/fluctuation in this model (only proportional to the current price). In reality, huge fluctuations can happen, which causes financial crises. Active research is going on to understand the real fluctuations and model them properly.

The type of the equation (6.1) is also called the **geometric Brownian motion**. It can also be found in a different context, for example, to model an exponentially growing living population with growth rate μ with some fluctuation in the growth rate due to environmental noise[1]. In this case, again, the Ito interpretation is appropriate because the growth of the population happens based on how many of the population are there at the beginning.

6.1.2 Solving the geometric Brownian motion

It is easy to solve this equation by using a new variable

$$Y = \ln S(t). \tag{6.2}$$

[1]Imagine, bacterial cells that divide on average 1/30 min but the food available in the environment fluctuates and affects the cell division rate.

Using Taylor expansion and remembering $dB^2 = dt$, we have

$$dY = \ln S(t+dt) - \ln S(t) = \ln(S+dS) - \ln S = \frac{1}{S}dS - \frac{1}{2S^2}dS^2 = \frac{1}{S}\left[\mu S dt + \sigma S \cdot dB\right] - \frac{1}{2S^2}(\sigma S dB)^2,$$

giving

$$\begin{aligned} dY &= \left[\frac{1}{S}\mu S - \frac{1}{2S^2}\sigma^2 S^2\right] dt + \frac{1}{S}\sigma S \cdot dB(t) \\ &= \left[\mu - \frac{\sigma^2}{2}\right] dt + \sigma \cdot dB(t). \end{aligned} \tag{6.3}$$

(This is equivalent to using Ito's formula (5.59).) Conveniently, the last equation does not include a multiplicative noise. By integrating both sides from time 0 to t, we have

$$\begin{aligned} \int_{t=0}^{t=t} dY &= \int_{t=0}^{t=t} \left[\mu - \frac{\sigma^2}{2}\right] dt + \sigma \int_{t=0}^{t=t} dB(t) \\ \Leftrightarrow Y(t) - Y(0) &= \left[\mu - \frac{\sigma^2}{2}\right] t + \sigma \left[B(t) - B(0)\right]. \end{aligned} \tag{6.4}$$

By using $B(0) = 0$ (part of the definition of the Wiener process) and $S(t) = e^{Y(t)}$, we get

$$S(t) = S(0) \exp\left[(\mu - \sigma^2/2)t + \sigma B(t)\right]. \tag{6.5}$$

6.2 Invest or not?

Suppose you should decide whether you should buy a stock or not. Assume that the market is actually following the SDE (6.1), and you somehow know the values of μ and σ. Which stock should you buy?

6.2.1 Two growth rates

Average stock price

One thing you can notice is that, by taking the average of eq. (6.1), we have

$$d\langle S\rangle = \mu \langle S\rangle dt, \tag{6.6}$$

namely

$$\langle S(t)\rangle = e^{\mu t}. \tag{6.7}$$

Here, for simplicity, we assumed the initial value $S(0) = 1$. So the average expected value of your stock grows exponentially, with the ensemble average growth rate

$$g_{ensemble} \equiv \frac{d}{dt}\left[\ln\langle S(t)\rangle\right] = \mu. \tag{6.8}$$

This is positive as long as $\mu > 0$, *independent of the volatility* σ. Should you always invest all your money in a stock as long as $g_{ensemble} > 0$, then? But isn't it strange that σ does not matter? Intuitively a more fluctuating market sounds riskier.

Expected growth rate

We actually have another useful equation, which is (6.4). If we take the average of this equation with $S(0) = 1$ (hence $Y(0) = 0$ from 6.2), we get

$$\langle Y(t)\rangle = \langle \ln S(t)\rangle = \left[\mu - \frac{\sigma^2}{2}\right] t. \tag{6.9}$$

Namely, if we take an average of the logarithm of the stock price, it can actually decrease over time if σ is too large. We define another growth rate here:

$$g_{log} \equiv \frac{d}{dt} \langle \ln S(t) \rangle = \mu - \frac{\sigma^2}{2}. \tag{6.10}$$

We call this the expected growth rate. This suggests that we should invest only when $g_{log} > 0$, i.e., $\mu > \sigma^2/2$.

Why do we get very different growth rates in (6.8) and (6.10), and which quantity should we use to decide our strategy?

6.2.2 Ensemble average vs. Time average and what to do personally

The difference comes from the exponential nature of the geometric Brownian motion (6.1), where the fluctuation due to $\sigma \neq 0$ is strongly amplified. As a result, the distribution of $S(t)$ becomes extremely wide as time passes. If one takes an ensemble average - i.e., average over many parallel trials, then the average value $\langle S(t) \rangle$ is dominated by the very small number of samples that became extremely rich, just by being extremely lucky and never being hit by the negative fluctuation. Of course, such luck will get rare and rare as time goes on. If you follow one particular sample, it will experience some negative fluctuation in the long run and will lose a lot.

Therefore, unless you have infinitely a lot of money to invest in parallel, or unless you get a source of money that you can restart the investment as many times as you want, you should not use the growth rate of the ensemble average $g_{ensemble}$ as your criteria for betting strategy.

The second quantity, the expected growth rate g_{log}, is more reasonable if you have a fixed amount of money and bet once and follow the long time growth. By taking the logarithm before taking the ensemble average, it is telling the typical growth rate of each sample. Therefore, it feels the effect of the fluctuation. Using g_{log} to decide the betting strategy was proposed by an economist Kelly [15], who considered gambling strategies. We can justify this argument by considering the difference between an ensemble average and a time average.

Ensemble average vs. Time average

The usage of g_{log} is better because it corresponds to the growth rate of the *long time average of one sample*, which is different from the ensemble average, i.e., the average over the infinite number of samples [16, 17]. Let's see this by following the calculation in [16].

Let's define the average growth rate of the stock price over N sample over time T as

$$g_{est}(T, N) = \frac{1}{T} \ln \left[\frac{1}{N} \sum_{i=1}^{N} \frac{S_i(T)}{S_i(0)} \right]. \tag{6.11}$$

By definition,

$$\lim_{N \to \infty} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{S_i(T)}{S_i(0)} \right] = \left\langle \frac{S(T)}{S(0)} \right\rangle, \tag{6.12}$$

so

$$\lim_{N \to \infty} g_{est}(T, N) = \mu. \tag{6.13}$$

If we want the long time average of one sample, then we should calculate $\lim_{T \to \infty} g_{est}(T, 1)$. By using the solution (6.5), this gives us

$$\begin{aligned} \lim_{T \to \infty} g_{est}(T, 1) &= \lim_{T \to \infty} \frac{1}{T} \left[\left(\mu - \frac{\sigma^2}{2} \right) T + \sigma B(T) \right] \\ &= \mu - \frac{\sigma^2}{2} + \sigma \lim_{T \to \infty} \frac{B(T)}{T}, \end{aligned} \tag{6.14}$$

where $B(T)$ is a Wiener process at time T. Since statistically $B(T) = \sqrt{T}B(1)$, the last term will approach to zero as $T \to \infty$ [2]. Namely, we obtain

$$\lim_{T\to\infty} g_{est}(T,1) = \mu - \frac{\sigma^2}{2} = g_{log}, \tag{6.15}$$

i.e., the long time average of the growth rate of one trajectory becomes g_{log}.

If one person invests to a market and wait for a long time hoping that the market goes up, then clearly the situation is represented by g_{log} - one should use g_{log} to decide one's investment, not $g_{ensemble}$.

6.3 Can you do better? Bet-hedging strategy

It is clear that we should not invest at all if $\mu < 0$. If we invest the money we have when $\mu - \frac{\sigma^2}{2} > 0$, we expect that the invested money will increase over time in the long run. Now, can we do better than just invest everything we have? Is it possible to gain something when $\frac{\sigma^2}{2} > \mu > 0$?

6.3.1 Bet-hedging: invest-or-not

One possible strategy is the *bet-hedging strategy*. This was investigated by Kelly [15], in a simple gambling game and extended to finance by Maslov and Zhang [18]. We invest only a fraction r of the money M we have to the market and keep the rest untouched. Assume that we adjust the investment with a small time interval dt to keep the ratio r.

Then, the change in the total money comes from the invested part only. This gives

$$dM = \mu r M dt + \sigma r M \cdot dB. \tag{6.16}$$

Since this has a parallel form to (6.1), we know that (see eq. (6.3))

$$d\ln M = \left[r\mu - r^2\frac{\sigma^2}{2}\right]dt + r\sigma \cdot dB(t),$$

giving the expected growth rate for total money to be

$$g_M(r) = \frac{\langle d\ln M(t)\rangle}{dt} = r\mu - r^2\frac{\sigma^2}{2}. \tag{6.17}$$

We want to set the investment fraction r so that $g_M(r)$ is maximized.

It is again clear that if $\mu < 0$, $g_M(r)$ is always negative for $0 < r < 1$, so no need to invest. When $\mu > 0$,

$$g'_M(r) = \mu - \sigma^2 r,$$

(prime indicates the derivative by r) is positive when $r = 0$ and zero when $r = \mu/\sigma^2 > 0$. This indicates that $g_M(r)$ takes its maximum value for $r = \min(\mu/\sigma^2, 1)$.

To summarize, if we use the bet-hedging strategy, the recommended action is the following:

(a) If $\mu < 0$, never invest ($r = 0$).

(b) If $0 < \mu < \sigma^2$, then invest only fraction $r = \mu/\sigma^2$ of your money. Then expected growth rate is $g_M(\mu/\sigma^2) = \frac{\mu^2}{2\sigma^2}$.

(c) If $\sigma^2 < \mu$, invest all of your money ($r = 1$).

The intermediate case (b) is the most interesting, where bet-hedging gives a better return than investing all. Especially, for $0 < \mu < \sigma^2/2$, if we invest all the money we tend to lose it (cf. eq. (6.10)), but one can get benefit from this risky market by having the bet-hedging strategy if we know the optimal investment ratio. This is because the effect of the expected return μ on the expected growth rate is linear, while the effect of volatility σ is squared. Thanks to this, one can feel the positive return more than the volatility by tuning $r < 1$.

[2] The distribution of $B(T)$, $P(B(T)) = \frac{1}{\sqrt{2\pi T}}\exp\left[-\frac{B(T)^2}{2T}\right]$ will be recovered by scaling the distribution of $B(1)$ in this conversion.

6.3.2 Bet-hedging in two markets

We just considered the case where we either invest or do not. But we can think about investing in multiple markets at different fractions, which was also analyzed in ref. [18]. To get a flavour of it, we consider investing in two markets.

We assume that markets 1 and 2 have the expected rate of return μ_1 and μ_2, respectively. Without loss of generality, we assume $\mu_1 > \mu_2$; i.e., the market 1 is better if we do not consider the fluctuation. We also need to know if the fluctuations of the two markets are correlated or not. It is probably reasonable to assume that some part is correlated and some part is not. Let's assume that there are three kinds of noise, represented by three independent Wiener processes, $B_1(t)$, $B_2(t)$, and $B_3(t)$. Then we parametrize the markets so that $\sigma_1 dB_1$ and $\sigma_2 dB_2$ represent independent fluctuations of the growth rate in each market, respectively. The common noise felt by the market 1 and 2 is represented by $\tilde{\sigma}_1 dB_3$ and $\tilde{\sigma}_2 dB_3$, respectively.

Now, let's consider a strategy where we invest a fraction of r to market 1, and the rest $1-r$ to market 2. Then the total money will obey the following SDE:

$$dM = [\mu_1 r + \mu_2(1-r)]Mdt + \sigma_1 rM \cdot dB_1 + \sigma_2(1-r)M \cdot dB_2 + (\tilde{\sigma}_1 r + \tilde{\sigma}_2(1-r))M \cdot dB_3. \tag{6.18}$$

By using the property of the Ito calculus and the independence of the B_1, B_2, and B_3, we obtain the expected growth of the total money as follows (derivation left for exercise):

$$g_{M,2}(r) \equiv \frac{d\langle \log M \rangle}{dt} = \mu_1 r + \mu_2(1-r) - \frac{1}{2}\left(\sigma_1^2 r^2 + \sigma_2^2(1-r)^2 + (\tilde{\sigma}_1 r + \tilde{\sigma}_2(1-r))^2\right). \tag{6.19}$$

Notice that if $\mu_2 = 0$, $\sigma_2 = 0$, and $\tilde{\sigma}_1 = \tilde{\sigma}_2 = 0$, we recover the previous case of investing one market or not, (6.17). Finding the optimal r to maximize this growth rate gives the optimal investment ratio to be

$$r_{opt} = \frac{\mu_1 - \mu_2 + \sigma_2^2 + \tilde{\sigma}_2(\tilde{\sigma}_2 - \tilde{\sigma}_1)}{\sigma_1^2 + \sigma_2^2 + (\tilde{\sigma}_1 - \tilde{\sigma}_2)^2}. \tag{6.20}$$

if this is between 0 and 1.

This is a rather complex expression, but it tells us something when we consider simple cases as follows.

When common noise are the same $\tilde{\sigma}_1 = \tilde{\sigma}_2$

Then we have

$$r_{opt} = \frac{\sigma_2^2 + (\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2}. \tag{6.21}$$

This does not include any common noise term; if we anyway invest, then there is no way to optimize against the common noise with the same amplitude.

If in addition $\sigma_2 \to 0$: In this case, we have

$$r_{opt} = \frac{(\mu_1 - \mu_2)}{\sigma_1^2}. \tag{6.22}$$

$\mu_2 = 0$ goes back to the simplest bet-hedging case of either invest or not (no investment gives no growth, no fluctuation) we considered before.

When there is only common noise ($\sigma_1 = \sigma_2 = 0$)

Then we have

$$r_{opt} = \frac{\tilde{\sigma}_2(\tilde{\sigma}_2 - \tilde{\sigma}_1) + (\mu_1 - \mu_2)}{(\tilde{\sigma}_1 - \tilde{\sigma}_2)^2}. \tag{6.23}$$

The condition $0 < r_{opt} < 1$ can be rewritten as

$$\tilde{\sigma}_2(\tilde{\sigma}_1 - \tilde{\sigma}_2) < \mu_1 - \mu_2 < \tilde{\sigma}_1(\tilde{\sigma}_1 - \tilde{\sigma}_2). \tag{6.24}$$

If $\mu_1 = \mu_2$, then there is no hedging because r_{opt} cannot satisfy $0 < r_{opt} < 1$; one should hedge to lower $\tilde{\sigma}$ strategy. Hedging is advantageous only when $\mu_1 > \mu_2$ and $\tilde{\sigma}_1 > \tilde{\sigma}_2$ (or vice versa), i.e. if the faster-growing market is feeling the fluctuation of the world economy stronger.

6.4 Summary

- Geometric Brownian motion
$$dS = \mu S dt + \sigma S \cdot dB$$
is used as a basic model for a stock price.

- Two kinds of "growth rate" is obtained, the ensemble average growth rate
$$g_{ensemble} \equiv \frac{d}{dt}\left[\ln\langle S(t)\rangle\right] = \mu$$
and the expected growth rate
$$g_{log} \equiv \frac{d}{dt}\langle \ln S(t)\rangle = \mu - \frac{\sigma^2}{2}.$$

- The difference between the two growth rates comes from the fact that geometric Brownian motion has an extremely wide distribution and ensemble average does not give a "typical" value. The expected growth rate represents a long-time average of one trajectory.

- If $0 < \mu < \sigma^2$, investing fraction $r = \mu/\sigma^2$ of your money (bet-hedging strategy) will enable one to gain optimally in a risky market.

- When multiple markets are considered, possible correlation of noise should be taken into account.

6.5 Exercise

1. Consider a Langevin equation with multiplicative noise.
$$\frac{dX}{dt} = X\hat{R}(t)$$
Solve this equation analytically (express $X(t)$ as a function of t and $B(t)$) as an Ito SDE. Take the average to show
$$\langle X(t)\rangle = \langle X(0)\rangle.$$

2. * Numerical simulation of geometric Brownian motion.

   Let's simulate the investment with the bet-hedging strategy for the case of invest-or-not (subsection 6.3.1) for a case with $0 < \mu < \sigma^2/2$. Let's set $\mu = 0.3$ and $\sigma = 1$ (This satisfies $0 < \mu < \sigma^2/2$). We set the initial amount of money as $M(0) = 1$.

   (a) Set r to be 0.01, and numerically integrate the SDE (6.16) to obtain the $M(t)$ until $t = 5$. Plot 3 realizations of trajectories as a function of time.

   (b) For a given r, calculate the ensemble growth rate $\ln\langle M(t_f)\rangle/t_f$ and the log growth rate $\langle \ln M(t_f)\rangle/t_f$ at time $t_f = 5$ for 100 samples or more. Also, calculate the long-time growth rate of one sample $\ln(M(T))/T$ with $T = 500$ or longer. Do this by changing r from 0.01 to 1 with a step size of 0.01, and plot these three growth rates as a function of r. Is there a maximum at a certain r? Do the results agree with the theoretical prediction?

3. Invest to two markets

   (a) Derive eq. (6.19) from eq. (6.18).

   (b) Using eq. (6.19), show that the optimal investment ratio is (6.20).

Advanced exercises

4. Instead of using Euler-Maruyama method, you can numerically calculate the solution of SDE (6.1) by numerically calculating the standard Brownian motion and using eq. (6.5). Plot $S(t)$ and compare the result from Euler-Maruyama method, using the same random noise sequence- They should be "the same" except for the numerical error.

5. Generalize the two markets case in the previous exercise to N markets case with an integer $N \geq 2$.

   (a) Define a SDE with investment fraction for each markets by generalizing eq. (6.18).

   (b) Calculate the expected growth rate.

   (c) Express the optimization problem of the investment fraction by using a Lagrange multiplier. Solve it for no common noise case.

6.6 Further reading

This chapter is based on [4, 17, 16, 18].

Chapter 7

Poisson process and one-step process

From chapter 2 to 5, we have been focusing on the stochastic process driven by the Wiener process. This is, of course, not the only class of stochastic process. For example, remember the random walk in chapter 1: It was defined so that a particle jumps to its one of neighbour positions with probability 1/2 after a time interval Δt.

We may consider a continuous-time version of the random walk by defining the jump **rate** r per unit time. The definition of a rate is simple: it is the probability for an event happens per unit of time, therefore its dimension is 1/time. Naturally, the probability of an event happening per small time interval of Δt is $r\Delta t$. But now we assume that the time is continuous, i.e., Δt can be arbitrarily small, still keeping r constant. Stochastic processes defined by using rates are very common. In this chapter, we will learn how to deal with such processes.

7.1 Poisson process

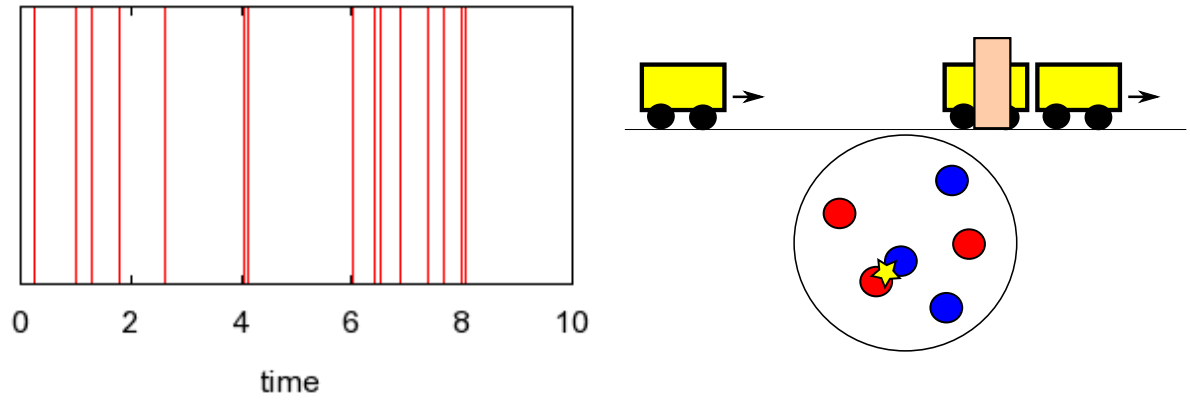


Figure 7.1: Left: Event sequence of Poisson process with $r = 1$. Right: Schematic examples of Poisson process.

The most basic process is the so-called **Poisson process**, the process where events happen at a constant rate of r. An example event sequence of a Poisson process is shown in Fig. 7.1 left. Here, we consider only one kind of event. A good real-life example can be a time sequence of a signal from a sensor, which detects the arrival of a particle (e.g., photons or electrons), and the particles hit the sensor at a constant rate r. We see fluctuations in the signal because particle arrival looks like Fig. 7.1 left, and this becomes the dominant source of noise when the rate r is low - this is called the "Poisson noise" or the "shot noise". Or it may be a model of the arrival of buses on a snowy day (Fig. 7.1 right top) - Buses come one per X min on average as written on the schedule (As long as they started), but it is hard to tell when the next bus comes to your

stop when the traffic is chaotic. Then to assume that the bus arrival event happens with a constant rate of 1/X [1/min] stochastically may be an okay approximation. Or it can be a model of a chemical reaction in a well-mixed solvent: Two kinds of molecules in a solvent collide stochastically at a certain rate and a reaction happens (Fig. 7.1 right bottom).

Let us start by describing the Poisson process in detail, and then proceed to more general cases.

7.1.1 The inter-event interval distribution for Poisson process

As shown in Fig. 7.1 left, the time interval between events is fluctuating. The first question we consider is: what is the probability density $P_{int}(t)$ that the interval time between two events is t?

This can be calculated easily by dividing the time into M segments with equal interval $\Delta t = t/M$ and then later taking the small Δt limit. The probability that no event happens between time 0 and $M\Delta t$ AND a event happens between $M\Delta t$ and $(M+1)\Delta t$ is given by

$$P_{int}(t)\Delta t = (1 - r\Delta t)^M \times (r\Delta t) \qquad (\text{in the limit of} \quad M \to \infty.) \tag{7.1}$$

By dividing both side with Δt, we get

$$P_{int}(t) = \lim_{M\to\infty} r\left(1 - r\frac{t}{M}\right)^M = re^{-rt}. \tag{7.2}$$

Namely, the interval time distribution obeys exponential distribution. This will be later important when we simulate processes described by rates.

We can show that the mean time interval is

$$\langle t \rangle = \frac{1}{r} \quad (\text{as expected}), \tag{7.3}$$

and the variance is

$$\langle (t - \langle t \rangle)^2 \rangle = \frac{1}{r^2}. \tag{7.4}$$

for exponential distribution (Exercise).

7.1.2 Master equation for Poisson process

Next, let us consider the number of events that have happened between time 0 and t, $N(t)$. What is the probability $P(n,t)$ that $N(t) = n$ at time t?

This can also be calculated by dividing time into M intervals, expressing it as a binomial distribution and then taking $M \to \infty$ limit (left for advanced exercise). However, here we take another way to calculate this. We first derive a differential equation for $P(n,t)$

$$\frac{\partial P(n,t)}{\partial t} = ?$$

and then we solve it to get $P(n,t)$. It is easy to generalize this approach to more complex stochastic processes.

To derive such an equation, we need to be aware of the following things:

- The probability that one event happens for a small time interval dt is rdt. The probability that two events happen in this interval is $O(dt^2)$, which can be ignored in $dt \to 0$ limit.
- The probability that n events have happened by time t is $P(n,t)$.

Therefore, the change in the probability for a small time interval, dt, is given by

$$\begin{aligned} P(0, t+dt) &= P(0,t) - rP(0,t)dt, \\ P(n, t+dt) &= P(n,t) + rP(n-1,t)dt - rP(n,t)dt \quad (\text{for } n > 0). \end{aligned}$$

Here, the second term denotes that the n-th event occurred between t and $t+dt$, and the third term denotes that $n+1$-th event occurred between t and $t+dt$. The case with $n=0$ is special since there is no negative number of events.

Moving the first term in RHS to LHS and dividing both sides with dt, we get

$$\dot{P}(0,t) = -rP(0,t), \tag{7.5}$$
$$\dot{P}(n,t) = rP(n-1,t) - rP(n,t) \quad (\text{for } n>0). \tag{7.6}$$

This is the **master equation** for the Poisson process. Now all we need to do is solve this equation!

7.1.3 Solving Master equation by Generating function method

Generating function

To solve master equations for discrete variables like (7.6), it is sometimes convenient to use the **generating function** method [1].

Consider the probability $P_n(t)$ for a stochastic variable N to take an integer n at time t. The *generating function* is defined as

$$G(s,t) = \sum_{\text{all possible } n} s^n P(n,t). \tag{7.7}$$

The inverse transformation from the generating function to the probability is given by

$$P(n,t) = \frac{1}{2\pi i}\oint_C z^{-n-1}G(z,t)dz, \tag{7.8}$$

where the integration is on the complex plain z along a loop C that goes around zero counterclockwise. This can be easily seen by substituting the definition (7.7) to $G(z,t)$, and using the residue theorem to evaluate the complex integral.

Though it is nice to know the formal inverse transformation, there are many quantities that we can calculate using a generating function without using the inverse transformation. In this lecture notes, we mainly consider such examples.

Moments and generating function

The generating function helps to calculate the k-th moment

$$\langle N(t)^k\rangle = \sum_n n^k P(n,t). \tag{7.9}$$

Normalization First, the normalization condition gives

$$G(1,t) = \sum_n P(n,t) = 1. \tag{7.10}$$

Mean Next, we can calculate the mean from the generating function by using

$$\frac{\partial}{\partial s}G(s,t) = \sum_n \frac{ds^n}{ds}P(n,t) = \sum_n n s^{n-1}P(n,t). \tag{7.11}$$

Thus by setting $s=1$, we have

$$\frac{\partial}{\partial s}G(s,t)|_{s=1} = \sum_n nP(n,t) = \langle N(t)\rangle. \tag{7.12}$$

[1] Generating function below is defined for the discrete state variable. For the continuous state variable, the Fourier transform of the probability density called the characteristic function plays a similar role. You can find them in appendix A.1

The 2nd moment and Variance The second derivative of $G(s,t)$ by s gives

$$\frac{\partial^2}{\partial s^2}G(s,t) = \sum_n \frac{d^2 s^n}{ds^2}P(n,t) = \sum_n n(n-1)s^{n-2}P(n,t). \tag{7.13}$$

Thus by setting $s=1$, we have

$$\frac{\partial^2}{\partial s^2}G(s,t)|_{s=1} = \sum_n n(n-1)P(n,t) = \langle N(t)^2\rangle - \langle N(t)\rangle. \tag{7.14}$$

Thus, the variance of $N(t)$ is calculated as

$$\begin{aligned}\sigma^2(t) &= \langle N(t)^2\rangle - \langle N(t)\rangle^2 \\ &= \langle N(t)^2\rangle - \langle N(t)\rangle + \langle N(t)\rangle - \langle N(t)\rangle^2 \\ &= \frac{\partial^2 G}{\partial s^2}|_{s=1} + \frac{\partial G}{\partial s}|_{s=1} - \left(\frac{\partial G}{\partial s}|_{s=1}\right)^2.\end{aligned} \tag{7.15}$$

One can similarly calculate the higher moments.

7.1.4 Time evolution equation for $G(s,t)$: Poisson process

These relations are useful only if we can calculate the generating function for the process we are interested in. Often we can derive the time evolution equation for $G(s,t)$ from the master equations for $P_n(t)$. The equation for $G(s,t)$ is often easier to solve than master equations.

Let us see this first from the Poisson process eqs. (7.5)-(7.6). In this case, n runs from 0 to ∞, therefore the generating function is defined as

$$G(s,t) = \sum_{n=0}^{\infty} s^n P(n,t). \tag{7.16}$$

Convenient relations to know:

- The normalization condition
$$G(1,t) = \sum_{n=0}^{\infty} P(n,t) = 1 \tag{7.17}$$
- For a integer $0 \le m \le n$,
$$\sum_{n=m}^{\infty} s^n P(n-m,t) = s^m \sum_{n=m}^{\infty} s^{n-m} P(n-m,t) = s^m \sum_{k=0}^{\infty} s^k P(k,t) = s^m G(s,t) \tag{7.18}$$
- For time derivative:
$$\frac{\partial}{\partial t}G(s,t) = \frac{\partial}{\partial t}\sum_{n=0}^{\infty} s^n P(n,t) = \sum_{n=0}^{\infty} s^n \left[\frac{\partial}{\partial t}P(n,t)\right]. \tag{7.19}$$

Equation for generating function for Poisson process

The master equations for the Poisson process are eqs.(7.5)-(7.6). By using eqs.(7.19), (7.5), and (7.6), we have

$$\begin{aligned}\frac{\partial G(s,t)}{\partial t} &= s^0 \frac{\partial}{\partial t}P(0,t) + \sum_{n=1}^{\infty} s^n \frac{\partial}{\partial t}P(n,t) \\ &= -rP(0,t) + \sum_{n=1}^{\infty} s^n \left[rP(n-1,t) - rP(n,t)\right] \\ &= r\sum_{n=1}^{\infty} s^n P(n-1,t) - r\sum_{n=0}^{\infty} s^n P(n,t) \\ &= rsG(s,t) - rG(s,t) = r(s-1)G(s,t).\end{aligned} \tag{7.20}$$

Eq.(7.18) was used at the 4th equality. This transforms the master equations for $P_n(t)$ to one partial differential equation for $G(s,t)$, as we wanted.

Calculate the generating function for the Poisson process

Now we have

$$\frac{\partial}{\partial t}G(s,t) = r(s-1)G(s,t).$$

Integrating the equation over t (with noting that s can be considered as a constant when we are doing this), we get

$$G(s,t) = G(s,0)e^{r(s-1)t}. \tag{7.21}$$

The initial value $G(s,0)$ is a function of s in general.

If we consider the initial condition that $N(0) = 0$, i.e.,

$$P(n,0) = \delta_{n,0}, \tag{7.22}$$

where $\delta_{i,j}$ is the Kronecker delta ($\delta_{i,j} = 1$ when $i = j$, otherwise $\delta_{i,j} = 0$), we have

$$G(s,0) = \sum_{n=0}^{\infty} s^n P(n,0) = \sum_{n=0}^{\infty} s^n \delta_{n,0} = 1. \tag{7.23}$$

Thus we get

$$G(s,t) = e^{r(s-1)t}. \tag{7.24}$$

Calculate the distribution

We have

$$G(s,t) = e^{r(s-1)t} = e^{-rt}e^{rst} = e^{-rt}\sum_{n=0}^{\infty}\frac{(rt)^n s^n}{n!}. \tag{7.25}$$

Comparing this with the definition (7.16), we get

$$P(n,t) = \frac{(rt)^n}{n!}e^{-rt}, \tag{7.26}$$

which is the Poisson distribution.

The Poisson distribution is known to have a variance equal to the mean, which can be easily shown by using the obtained generating function (exercise). There are many examples of stochastic processes where the basic process is the Poisson process. Therefore, the ratio (variance/mean) is called the Fano factor and it is sometimes used as a quantity to quickly check if the process could be related to the Poisson process.

7.2 One-step process (Birth-death process)

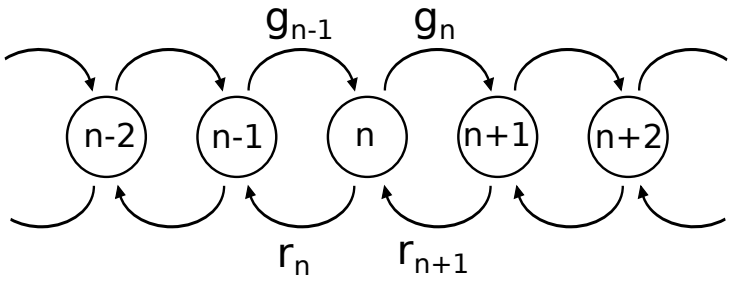


Figure 7.2: Schematic description of the one-step process (Birth-death process).

Let us generalize the Poisson process slightly. A class of models where states are denoted by discrete integers n and the transition is restricted to the jumps between adjacent states (Fig. 7.2) is called **one-step process** or **birth-death process**. This results in the following master equation:

$$\dot{P}(n,t) = g(n-1)P(n-1,t) - g(n)P(n,t) + r(n+1)P(n+1,t) - r(n)P(n,t). \tag{7.27}$$

Poisson process described by eq. (7.6) is the simplest example of a one-step process, with $g(n) = r$ for $n \geq 0$ and all other rates are zero. Continuous-time random walk in one dimension is another example of the one-step process, with n denoting position on a lattice, and setting $r(n) = g(n) = c$ with c being a constant.

It should be noted that, in general, the transition rate does not have to be a constant, but can depend on n. In the argument in Fig. 7.2, we are assuming that the rate is a function of the state it is starting from - the rate to increase from n to $n+1$ is $g(n)$, and the rate to decrease from n to $n-1$ is $r(n)$. Some chemical reactions, where n is the number of particles/elements which can increase or decrease one by one with some rate, provide good examples of the one-step process with n-dependent reaction rate. We consider examples later.

7.2.1 Example: Continuous-time (biased) random walk

Let us consider a particle that hops to the right with a rate p, and to the left with a rate q. This is the simplest continuous-time biased random walk [2]. $N(t)$ denotes the position on a one-dimensional lattice, and $P_n(t)$ is the probability that the particle is at position n at time t. This corresponds to the one-step process with $r(n) = q$, $g(n) = p$, and the master equation is given by:

$$\dot{P}(n,t) = qP(n+1,t) - qP(n,t) + pP(n-1,t) - pP(n,t). \tag{7.28}$$

When $p = q$, this is a symmetric random walk. When $p \neq q$, the random walk is *biased*, namely there is a finite mean velocity of the particle. Here we consider general case (Namely p and q can be different), with infinite space ($-\infty < n < \infty$).

We can use the generating function method in subsection 7.1.3 to analyze the behaviour of the particle. The generating function for this system is defined as

$$G(s,t) = \sum_{n=-\infty}^{\infty} s^n P(n,t). \tag{7.29}$$

For this example, we can also explicitly write down the equation for $G(s,t)$ from the master equation and obtain the solution for the initial condition $N(0) = 0$ to show (exercise) that

$$\langle N(t) \rangle = (p-q)t,$$

and

$$\sigma^2(t) = \langle N(t)^2 \rangle - \langle N(t) \rangle^2 = (p+q)t.$$

Namely, for $p = q$, the model describes ordinary Brownian motion, while in the biased ($p \neq q$) case, there is an mean motion whose velocity is proportional to the rate difference, but the fluctuation around this mean motion is still diffusive, i.e., variance grows linearly with time.

7.2.2 Example of the linear one-step process: Simple chemical reaction

The one-step processes with the rate $r(n)$ and/or $g(n)$ being a linear function of n is called a *linear one-step process*. Here we see an example, a very simple chemical reaction scheme as follows.

Consider a chemical reaction where molecule X is produced at a constant production rate k, and degraded with a constant degradation rate *per molecule* Γ. $N(t)$ denotes the number of molecule X at time t, and $P(n,t)$ is the probability that there are n molecule X at time t.

[2] Here, we consider the case where the time interval between the jump is determined by a constant rate of r, hence the interval time is exponentially distributed. The word "Continuous Time Random Walk", often abbreviated as "CTRW", is used for a wider range of processes, where waiting time distribution can be much more complicated, and also the jump distance can vary. But in this lecture, we only consider constant rate, constant jump distance case.

This corresponds to the one-step process with $g(n) = k$, $r(n) = \Gamma n$. The master equation is:

$$\dot{P}(n,t) = kP(n-1,t) - kP(n,t) + \Gamma(n+1)P(n+1,t) - \Gamma nP(n,t). \tag{7.30}$$

Since $N(t)$ is the number of particles, we consider half-infinite range ($0 \leq n < \infty$). Thus for $n = 0$, we have

$$\dot{P}(0,t) = -kP(0,t) + \Gamma P(1,t). \tag{7.31}$$

Generating function for simple chemical reaction

We define the generating function for this case as

$$G(s,t) = \sum_{n=0}^{\infty} s^n P(n,t). \tag{7.32}$$

The time evolution is given by

$$\begin{aligned}
\frac{\partial G(s,t)}{\partial t} &= k\sum_{n=1} s^n P(n-1,t) - k\sum_{n=0} s^n P(n,t) + \Gamma\sum_{n=0} s^n(n+1)P(n+1,t) - \Gamma\sum_{n=0} s^n nP(n,t) \\
&= ks\sum_{n=1} s^{n-1}P(n-1,t) - k\sum_{n=0} s^n P(n,t) \\
&\quad + s^{-1}\Gamma\sum_{n=0} s^{n+1}(n+1)P(n+1,t) - \Gamma\sum_{n=0} s^n nP(n,t).
\end{aligned} \tag{7.33}$$

Noticing

$$\sum_{n=0}^{\infty} s^n nP(n,t) = \sum_{n=1}^{\infty} s^n nP(n,t) \quad (\text{because } s^0 \cdot 0 \cdot P(0,t) = 0), \tag{7.34}$$

$$\sum_{n=0}^{\infty} s^n nP(n,t) = s\frac{\partial}{\partial s}\sum_{n=0}^{\infty} s^n P(n,t) = s\frac{\partial G}{\partial s}, \tag{7.35}$$

we have

$$\frac{\partial G(s,t)}{\partial t} = ksG - kG + \Gamma\frac{\partial G}{\partial s} - s\Gamma\frac{\partial G}{\partial s} = (s-1)\left[kG - \Gamma\frac{\partial G}{\partial s}\right]. \tag{7.36}$$

It is not as easy as before to get the explicit solution $G(s,t)$ [3], but here we focus on the followings:

- Steady-state distribution.
- The time evolution equations for moments.

These are easy to calculate by using the generating function.

Steady state distribution

At the steady state, the probability does not change with time, and thus $G(s,t)$ either, so we have $\frac{\partial G}{\partial t} = 0$. Thus the generating function for the steady state $G_{st}(s)$ satisfies

$$kG_{st}(s) - \Gamma\frac{dG_{st}}{ds} = 0. \tag{7.37}$$

This with the normalization condition $G_{st}(1) = 1$, we get

$$G_{st}(s) = \exp\left[\frac{k}{\Gamma}(s-1)\right] = e^{-k/\Gamma}\sum_{n=0}^{\infty}\frac{(k/\Gamma)^n s^n}{n!}. \tag{7.38}$$

Comparing this with the definition of the generating function, we get the steady state distribution to be

$$P_{st}(n) = \frac{(k/\Gamma)^n}{n!}e^{-k/\Gamma}. \tag{7.39}$$

This is the Poisson distribution with a mean k/Γ.

[3] It is possible, though. You can find the solution in the van Kampen book [2].

Time evolution equation for moments

We can also derive the time evolution equation for the moments. First we partially differentiate eq.(7.36) with respect to s:

$$\frac{\partial^2 G(s,t)}{\partial t \partial s} = \left[kG - \Gamma \frac{\partial G}{\partial s} \right] + (s-1) \left[k \frac{\partial G}{\partial s} - \Gamma \frac{\partial^2 G}{\partial s^2} \right]. \tag{7.40}$$

Noting that

$$\begin{aligned} \frac{\partial^2 G(s,t)}{\partial t \partial s}|_{s=1} &= \left(\sum_{n=0}^{\infty} n s^{n-1} \dot{P}(n,t) \right)_{s=1} = \frac{d}{dt} \left(\sum_{n=0}^{\infty} n s^{n-1} P(n,t) \right)|_{s=1} \\ &= \frac{d}{dt} \left[\frac{\partial G(s,t)}{\partial s}|_{s=1} \right] = \frac{d}{dt} \langle N(t) \rangle, \end{aligned}$$

we have

$$\frac{d}{dt} \langle N(t) \rangle = k - \Gamma \langle N(t) \rangle, \tag{7.41}$$

by substituting $s = 1$ to (7.40). This can be easily solved to give

$$\langle N(t) \rangle = \frac{k}{\Gamma}(1 - e^{-\Gamma t}) + \langle N(0) \rangle e^{-\Gamma t}. \tag{7.42}$$

Rate equations

For chemical reactions, one often considers in terms of the concentration of molecules. If the volume of the system is Ω then (7.41) becomes

$$\frac{d}{dt} \tilde{\phi}(t) = \alpha - \Gamma \tilde{\phi}(t), \tag{7.43}$$

where $\tilde{\phi}(t) = \langle N(t) \rangle / \Omega$ is the mean concentration of molecules at time t, $\alpha \equiv k/\Omega$ is production rate of molecules in the unit of [density/time]. These types of equations are called rate equations. We will see more detail on volume dependence in a later chapter 9.

Equation for the higher moments

If we take derivative of (7.40) once more by s and then set $s = 1$, we will get an equation that contains the second moment (see eq. (7.14)). In general, by taking higher derivatives of a generating function by s, one can obtain the equations for higher moments.

7.2.3 Example of the nonlinear one-step process: Coalescence

Next, let us see an example with nonlinear rates. Consider a situation where two particles collide to be one particle (Coalescence). It can be expressed as a reaction of $A + A \to A$. If there are N particles in a system of volume Ω, the chance that two-particle collide should be proportional to $N(N-1)/\Omega$, because it is $N\times$ (rate of one particular particle meeting another), and this rate is proportional to the density of the targets $(N-1)/\Omega$. When the reaction happens the number of particles will decrease by one. Therefore this is one-step process with $g(n) = 0$, $r(n) = (\lambda/\Omega)n(n-1)$ with a constant λ. This gives the master equation

$$\frac{\partial}{\partial t} P(n,t) = \frac{\lambda}{\Omega}(n+1)nP(n+1,t) - \frac{\lambda}{\Omega} n(n-1) P(n,t) \tag{7.44}$$

for $n \geq 0$. The corresponding generating function satisfies (exercise)

$$\frac{\partial}{\partial t} G(s,t) = -\frac{\lambda}{\Omega} s(s-1) \frac{\partial^2 G}{\partial s^2}. \tag{7.45}$$

This is not simple to solve either. Besides, the steady state is not interesting since it is obviously $N = 1$ if we start from $N > 0$. And if $N = 1$ or $N = 0$, nothing can happen since the rate is completely zero. (Such

a state is called an *absorbing state*.) We will come back to this example later as we learn approximation methods.

Interestingly, from (7.45), we can show that the equation for the mean is given by (exercise)

$$\frac{d}{dt}\langle N(t)\rangle = -\frac{\lambda}{\Omega}(\langle N(t)^2\rangle - \langle N(t)\rangle). \tag{7.46}$$

This equation is not closed, since it depends on $\langle N(t)^2\rangle$. Typically nonlinear rates introduce this kind of correlation between moments.

7.3 Master equation in general

We can also consider a stochastic process where the transition is not limited to the adjacent states. The master equation is a quite general tool to describe such a situation. We will talk about it in a later chapter, but here we briefly review master equations in general.

Discrete variable

Suppose a system can take discrete states, which are distinguished by a discrete index n. The probability to be in the state at n time t is denoted as $P(n,t)$. The *transition rate* from a state n to another state m is given by $W(m|n)$. In this case, the master equation is given by:

$$\frac{\partial}{\partial t}P(n,t) = \sum_{\text{all } m} \left[W(n|m)P(m,t) - W(m|n)P(n,t)\right]. \tag{7.47}$$

For example, if n is a number of particles, and the only possible event is to create 2 particles at a time, then the master equation will be

$$\frac{\partial}{\partial t}P(n,t) = \left[W(n|n-2)P(n-2,t) - W(n+2|n)P(n,t)\right]. \tag{7.48}$$

One-step process in Fig. 7.2 corresponds to

$$W(n|m) = r(m)\delta_{m,n+1} + g(m)\delta_{m,n-1}. \tag{7.49}$$

Continuous variable

It is also possible to write down a master equation for the probability distribution function of a system where a state is characterized by a continuous variable x. The master equation will then have the form

$$\frac{\partial}{\partial t}P(x,t) = \int \left[W(x|y)P(y,t) - W(y|x)P(x,t)\right] dy. \tag{7.50}$$

As an example, consider x as volume of water in a container, where droplets that obey a certain volume distribution $\nu(x)$ (normalized as $\int_0^\infty \nu(x)dx = 1$) can be added to it. If one droplet is added at a time stochastically with a constant rate k, then $W(x|y) = k\nu(x-y)$ if $x > y$, and otherwise $W(x|y) = 0$. In this case, the master equation for the probability density of the water volume $P(x,t)$ is given by

$$\frac{\partial}{\partial t}P(x,t) = \int_0^x \left[k\nu(x-y)P(y,t)\right] dy - kP(x,t), \tag{7.51}$$

where the second term represents that any events will increase the volume, and it is equivalent to $\int_x^\infty [-k\nu(y-x)P(x,t)]dy$.

7.4 Numerically simulate Poisson and related processes

Let us end this chapter by learning how to numerically simulate the Poisson process to get a plot like Fig. 7.1 as well as to simulate chemical reactions etc. The method will also be useful for the examples that appear later. There are two famous approaches: a **time-driven method** or an **event-driven method**. The latter is often called the **Gillespie algorithm** in the context of chemistry.

7.4.1 Time-driven method

Consider a Poisson process where events happen at a rate r. The idea of the time-driven method is very similar to the Euler method - divide the time to very small steps, Δt, and use the fact that $r\Delta t$ is the probability for an event to happen. The algorithm will be the following.

1. Suppose the current time is t. For each time step, draw a random number a, from a uniform distribution on the unit interval [0,1]. Make an event happen with probability $r\Delta t$ (i.e. when $a < r\Delta t$).
2. Proceed the time by Δt (t should be updated to $t + \Delta t$).
3. Repeat 1 and 2.

Obviously, the time step needs to be small enough so that $r\Delta t < 1$. Note that this method ignores the possibility that two and more event happens within a time step Δt. This also means that the method is not exact, as long as Δt is finite. The accuracy improves for smaller Δt, but of course, we need finite Δt to do simulation.

7.4.2 Event-driven method (Gillespie algorithm)

Remember that we know the *exact* distribution of interval-time, (7.2). If we use this, all you have to do is:

1. Suppose the current time is t. Draw a random number $\tau > 0$ from the distribution
$$P(\tau) = re^{-r\tau}.$$
2. Proceed the time by τ (i.e., $t \to t + \tau$) and make an event happen.
3. Repeat 1 and 2.

The event-driven method is ideally an exact simulation of a Poisson process in the sense that it reproduces the exact statistics of the time intervals. (Of course, there can be "numerical errors" depending on how good the random number generator is etc.) As mentioned, in (bio)chemistry and related field, this particular method is often called the Gillespie algorithm, who formalized this for chemical reactions. The word "event-driven" has a wider meaning: It just says that time proceeds by event, not by time step. The term can be used for a wider class of algorithms.

How to generate a random number that obeys the distribution you want

Sometimes the numerical language/simulation program you use may not provide the random number generator for the distribution that you want. However, it is normally possible to find a way of generating a random number a that uniformly distributed between 0 and 1.

$$P^{(u)}(a) = 1 \quad \text{for} \quad 0 < (\text{or} \leq) a < (\text{or} \leq) 1, \quad \text{otherwise} \quad P^{(u)}(a) = 0. \tag{7.52}$$

How can we convert a to the random number τ that obeys the exponential distribution $P(\tau)$, eq.(7.2)?

Converting random numbers We want to realize

$$P^{(u)}(a)da = P(\tau)d\tau. \tag{7.53}$$

By substituting the distributions for $0 < a < 1$, we get

$$da = re^{-r\tau}d\tau. \tag{7.54}$$

By integrating this equation, we have

$$a = 1 - e^{-r\tau}. \tag{7.55}$$

Note that RHS is between 0 and 1 for $0 < \tau$. Thus, we can get τ from a by $\tau = -\frac{\ln(1-a)}{r}$. Since $1 - a$ is also a uniformly distributed random number between 0 and 1, this is equivalent to

$$\tau = -(\ln a)/r. \tag{7.56}$$

So, by drawing a, we can calculate a random number τ that follows the distribution $P(\tau) = re^{-r\tau}$ via (7.56).

7.4.3 The case with m kinds of possible events

The above method was only considering one possible event at a time. However, for example in the one-step process, a particle can move from state n to state $n+1$ or to state $n-1$; there are two kinds of events that can happen. So we need to generalize our method, and we do this here for the event-driven case.

Consider a case that the rate for an event of i-kind happening is r_i, and there are m kinds of possible events (i.e. $i = 1, 2, \cdots m$). r_i can depend on the system's status, but *we assume that the rates do not change between events*. Then the inter-event interval distribution of each kind of event follows an exponential function characterized by r_i. How can we extend the Gillespie method to such a case?

Method to deal with multiple possible events (Method 1: IMPORTANT!)

The rate that ANY event occurs is $K = \sum_{i=1}^{m} r_i$, and the probability that the event occur is the event i is r_i/K. Therefore:

1. Calculate $K = \sum_{i=1}^{m} r_i$.
2. Draw a random number a, from the uniform distribution with the interval [0,1]. Calculate the duration until the next event τ by $\tau = -\frac{\ln a}{K}$.
3. Draw another random number a' from a uniform distribution on the unit interval [0,1]. If $(\sum_{j=1}^{i-1} r_j)/K < a' < (\sum_{j=1}^{i} r_j)/K$, then the event that occur is the event i. Proceed the time from t to $t+\tau$ and let the event happen.
4. Repeat 1 to 3.

This way we need only 2 random numbers per event, independent of n.

Method to deal with multiple possible events (Method 2)

There is another method/implementation, which is less efficient when there are more than two choices, which is as follows.

For a given moment, the plausible duration for the event i to happen is given by τ_i, which obeys the exponential distribution $P_i(\tau_i) = r_i e^{-r_i \tau_i}$.

1. Draw a random number a_i from the uniform distribution with the interval [0,1]. Calculate the duration until the event i by $\tau_i = -\frac{\ln a_i}{r_i}$. Do this for $i = 1, \cdots, m$. Namely, we need m random numbers.
2. The closest event is the one that should happen next. So choose smallest τ_i out of $\tau_1, \cdots \tau_m$, proceed the time by τ_i, and make the event i happen.
3. Repeat 1 to 2.

Note that the system's state changes in step 2, so to be safe we need to draw new random numbers for all τ_i's in every step 1, i.e., we need to draw m random numbers per step[4]. This makes method 2 slower than method 1 when $m > 2$.

We can prove that the two methods give statistically the same result (advanced exercise).

A concrete example: Simple chemical reaction using Method 1

Suppose there are N particles in the system at time t. The new particle will be created at a constant rate k, and each particle is degraded at a rate Γ par particle - so the rate that one of N particles is degraded is ΓN.

The rate that ANY event occurs is $K = k + \Gamma N$, and the probability that the event to occur is the creation of one particle is k/K, while the probability that the event to occur is the degradation of one particle is $\Gamma N/K$. Therefore:

[4]If the change of the system did not affect some of the rates, there are ways of re-using some of τ_i.

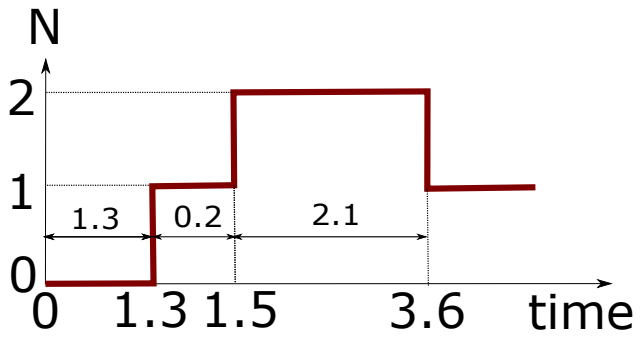


Figure 7.3: An example of time series from a Gillespie simulation.

1. Calculate $K = k + \Gamma N$.
2. Draw a random number a, from the uniform distribution with the interval [0,1]. Calculate the duration until the next event τ by $\tau = -\frac{\ln a}{K}$.
3. Draw another random number a' from a uniform distribution on the unit interval [0,1]. If $0 < a' < k/K$, then proceed the time from t to $t+\tau$ and increase N by one. If $k/K < a' < (k+\Gamma N)/K$, then proceed the time from t to $t+\tau$ and decrease N by one.
4. Repeat 1 to 3.

7.4.4 Note on time and ensemble averages with Gillespie algorithm

In the exercises, you will be asked to calculate the mean and the variance in the steady state. They are of course ensemble averages. You need to be careful to do it properly when you are using the Gillespie algorithm.

- If you take the ensemble average, then pay attention that taking a sample after a fixed number of events in the Gillespie algorithm will cause biased sampling and give a wrong result. For example, if you are doing the simple chemical reaction as the example above, then if your N is larger, you will have more degradation event - so if you randomly choose an event, it is likely that your N is larger than the typical value.

  To collect a proper ensemble, you need to run the simulation long enough to reach a steady state, and take statistics of N at a specific time T; if you have $N(t)$ at time $t < T$, and if your next event happens at τ with $t + \tau > T$, then $N(T) = N(t)$.

- In the steady state, the ensemble average and the time average match in the simple models - including what you do in the exercise. If you do the time average over the time series obtained from Gillespie simulation, then you need to pay attention that the intervals of events are not equal. For example, if you have $N = 0$ at time 0, and then the time to the next event is $\tau = 1.3$ and the N becomes 1, and then the time to the next event is $\tau = 0.2$ and N becomes 2, and then the time to the next event is $\tau = 2.1$ and N becomes 1 again - then the time average until time $1.3 + 0.2 + 2.1 = 3.6$ is $(0 \times 1.3 + 1 \times 0.2 + 2 \times 2.1)/3.6$ (See Fig. 7.3). In general, when the event interval between the i-th and the (i+1)th event is given by τ_i, the time average of, for example, a variable N, should be taken as

$$\sum_{i=1}^{all\ events} N_i \tau_i \Big/ \sum_{i=1}^{all\ events} \tau_i,$$

  where N_i is the value of N just after the event i. When you code this, be careful that how the index of N and interval τ are chosen; choosing the wrong indices is a very common mistake. You also need to have a similar consideration when calculating variance.

7.4.5 Time-driven or Event-driven?

There are some pros and cons of the time-driven and event-driven methods. A few points one should note are:

- Time-driven method
  - Algorithm is often simple.
  - When different time scales are involved in the process (e.g., sometimes many events happen for a short time, and sometimes nothing happens for a long time), one needs to adjust the time step to the shortest time scale, hence time-consuming.
  - If many independent events can happen in the system at the same time, the time-driven method can be easier to program.
  - Some models are defined in discrete time.
- Event-driven (Gillespie) method
  - "Exact" to deal with stochastic events in continuous time.
  - The time scales involved in the process does not matter since the algorithm jumps between events.
  - When there are many different possible events to happen, the algorithm can be complicated and slow.

7.5 Summary

- In the Poisson process where an event occurs at a constant rate r, the inter-event interval distribution is given by $P(t) = re^{-rt}$.
- Master equation is useful to describe stochastic processes. For one-step process, it is given by

$$\dot{P}(n,t) = r(n+1)P(n+1,t) + g(n-1)P(n-1,t) - (r(n)+g(n))P(n,t).$$

- Generating function

$$G(s,t) = \sum_{\text{all possible } n} s^n P(n,t)$$

  is useful to solve master equations.
- One can perform a numerical simulation of the Poisson process and stochastic processes defined by rates for an event to happen in two ways:
  - Time-driven method.
  - Event-driven (Gillespie) method. Especially the one for multiple possible events is useful to know.

7.6 Exercise

1. Show (7.3) and (7.4).
2. Calculate the mean and variance for the number of the events happened up to the time t for the Poisson process using the generating function (7.24).
3. One of the first applications of the Poisson distribution was to the statistics of accidents by horse-kicks in the Prussian army. Wikipedia entry for Ladislaus Bortkiewicz (1868-1931) states "The data give the number of soldiers killed by being kicked by a horse each year in each of 14 cavalry corps over 20 years. Bortkiewicz showed that those numbers follow a Poisson distribution. " What is the simplest model to reproduce these statistics? What kind of distribution do you expect if you plot the distribution of the number of horse-kick-accidents per cavalry corp per 2 years, instead of per year?

4. Consider the continuous-time biased random walk defined in subsection 7.2.1.

   (a) Show that
   $$\frac{\partial G(s,t)}{\partial t} = \left[qs^{-1} + ps - (p+q)\right] G(s,t). \tag{7.57}$$

   (b) Solve (7.57) under the initial condition that particle always starts at $n = 0$, i.e. $P(n,0) = \delta_{n,0}$, and show that
   $$G(s,t) = \exp\left[\left(qs^{-1} + ps - (p+q)\right) t\right]. \tag{7.58}$$

   (c) By using (7.58), show that $\langle N(t)\rangle = (p-q)t$, and $\sigma^2(t) = \langle N(t)^2\rangle - \langle N(t)\rangle^2 = (p+q)t$.

5. The generating function of a coalescence process.

   (a) Derive the equation for the generating function (7.45) from the master equation.

   (b) From (7.45), derive the equation for the mean (7.46).

6. It is also possible to obtain the equation for the moments by direct summation of the master equation, by knowing
   $$\frac{d}{dt}\langle N(t)\rangle = \frac{d}{dt}\sum_{all\ n} nP(n,t) = \sum_{all\ n} n\dot{P}(n,t)$$
   Substitute the RHS of the master equation for the one-step process (7.27) to $\dot{P}(n,t)$, and assuming the sum over n runs from $-\infty$ to ∞, show that
   $$\frac{d}{dt}\langle N(t)\rangle = \langle g(N(t))\rangle - \langle r(N(t))\rangle$$
   holds for a one-step process.

7. We consider the biochemical reactions in a cell, where the information on DNA is copied to mRNA, and it is translated to proteins, as schematically shown in Fig. 7.4. Assume

   - The cell volume is Ω.
   - mRNA is made at the rate $\alpha\Omega$.
   - mRNA is degraded at the rate Γ_1 per mRNA molecule.
   - Protein is made at the rate β per mRNA molecule. (mRNA is NOT consumed when one protein is made. In reality, machinery called ribosome binds to mRNA and reads the information on it, and connects amino acids according to the information on the mRNA. The same mRNA can be read again with another ribosome.)
   - Protein is degraded at the rate Γ_2 per protein molecule.

   Stochastic variables are the number of mRNA molecules N_m, and the number of proteins N_p.

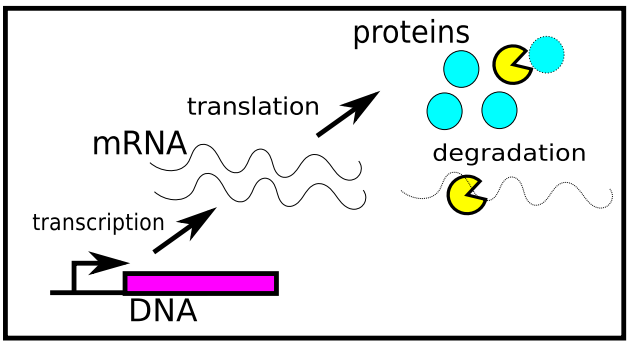


Figure 7.4: Schematic description of gene expression.

(a) First, consider the mRNA production process and ignore translation part. Then this is the same process as the simple chemical reaction considered in the text.

i. Write down the master equation for the probability that there are m mRNA molecules at a time t, $P(m,t)$.

ii. What is the mean and variance of N_m in the steady state? Confirm that with this definition of rate, the density $\tilde{\phi} \equiv \langle N_m \rangle / \Omega$ is independent of the volume Ω. How about the variance of "density", $\langle (N_m/\Omega - \tilde{\phi})^2 \rangle$?

(b) Now we consider the proteins also, and consider the probability that there is m mRNA molecules and p proteins in the cell at time t, $P(m,p,t)$.

i. Write down the master equation for $P(m,p,t)$.

ii. How $P(m,p,t)$ should be normalized?

iii. Define the generating function for this system as

$$G(u,v,t) = \sum_{m=0}^{\infty} \sum_{p=0}^{\infty} u^m v^p P(m,p,t). \tag{7.59}$$

For the number of mRNAs N_m and the number of proteins N_p, How can we calculate $\langle N_m \rangle$, $\langle N_p \rangle$, and $\langle N_m N_p \rangle$, from $G(u,v,t)$?

iv. Derive the time evolution equation for $G(u,v,t)$.

v. Show

$$\frac{d}{dt}\langle N_p \rangle = \beta \langle N_m \rangle - \Gamma_2 \langle N_p \rangle.$$

8. ** Gillespie simulation of the gene expression process: mRNA.

Consider the mRNA production in exercise 7-a. Perform the Gillespie simulation for the production and degradation of mRNA with $\alpha = 3$ [molecules/minute], $\Omega = 1$ [volume], $\Gamma_1 = 1/2$[1/minute], and plot the number of mRNAs as a function of time. Calculate the mean and the variance of the number of mRNA N_m in the steady state, as well as the Fano factor (variance/mean). Confirm that the numbers agree with what you expect for this process.

Note: Since the system reaches the steady/stationary state, you can calculate the averages (to calculate the mean and variance) by either ensemble averages in the steady state or by long-time averages. You can pick one of them as you like. However, consult the subsection 7.4.4 so that you do a proper average - otherwise, you will get the wrong number.

9. * Gillespie simulation: mRNA and proteins. Add the protein production on top of exercise 8 as considered in exercise 7-b. Keep parameters for mRNA production and degradation as $\alpha = 3$ [molecules/minute], $\Gamma_1 = 1/2$[1/minute], with $\Omega = 1$ [volume], and add protein production rate per mRNA as $\beta = 10$ [molecules/minute], and protein degradation rate as $\Gamma_2 = 1/30$ [1/minute].

(a) Plot the mRNA number and the protein number over time.

(b) Calculate the mean and the variance of the protein number in the steady state from the simulation data. Confirm that the mean obeys what is expected from the equation you obtained in 7-b-v. Calculate the Fano factor (variance/mean) for the protein number, and compare the value with the Fano factor of the mRNA number. Any thoughts on why one of them is bigger than the other?

Advanced exercise

10. Consider the Poisson process. Derive the Poisson distribution (7.26) by dividing t to M intervals and express $P(n,t)$ as a binomial distribution, and then taking $M \to \infty$ limit.

11. Using the generating function (7.58) and the inverse transformation (7.8), show that probability $P_n(t)$ for the continuous time biased random walk is given by

$$P_n(t) = e^{-(p+q)t} \sum_{k=0}^{\infty} \frac{(qt)^k (pt)^{n+k}}{k!(k+n)!}.$$

for $n \geq 0$, and

$$P_n(t) = e^{-(p+q)t} \sum_{k=0}^{\infty} \frac{(qt)^{k-n} (pt)^k}{k!(k-n)!}$$

for $n < 0$.

12. Confirm that the multiple-event Gillespie algorithm the method 1 and method 2 are equivalent by calculating the following quantities:

    (a) The probability that event of kind i is chosen in step b of method 2 should be r_i/K to agree with method 1. Show this.
    Hint: The probability $P_{i,l}(T)$ that τ_i is larger than T is $P_{i,l}(T) = \int_T^\infty r_i e^{-r_i \tau} d\tau$. Calculate the probability that τ_i is the smallest by using this.

    (b) Calculate the probability density of the duration between events, $P_{all}(\tau)$, in method 2. This should be $Ke^{-K\tau}$ to agree with method 1.
    Hint: The probability that a chosen event is i AND the duration is between τ and $\tau + d\tau$ is given by (prob. that τ_i is between τ and $\tau + d\tau$) *(prob. that the rest of τ_j's are longer than τ), i.e.,

$$r_i e^{-r_i \tau} d\tau \cdot P_{1,l}(\tau) \cdot P_{2,l}(\tau) \cdots P_{i-1,l}(\tau) \cdot P_{i+1,l}(\tau) \cdots P_{n,l}(\tau).$$

13. We continue with considering the protein number in exercise 7-b.

    (a) Calculate the variance of the protein number in the steady state, by using the equations for $\langle N_m \rangle$, $\langle N_p \rangle$, $\langle N_m N_p \rangle$, $\langle N_m^2 \rangle - \langle N_m \rangle$, and $\langle N_p^2 \rangle - \langle N_p \rangle$ which can be obtained from the generating function we have calculated. The calculation is straightforward but rather long. In the end, we get

$$\langle N_p^2 \rangle - \langle N_p \rangle^2 = \frac{\alpha \Omega \beta}{\Gamma_1 \Gamma_2} \left[1 + \frac{\beta}{\Gamma_1 + \Gamma_2} \right] = \langle N_p \rangle \left[1 + \frac{\beta}{\Gamma_1} \cdot \frac{1}{1 + \Gamma_2/\Gamma_1} \right]. \tag{7.60}$$

    This expression shows the burstiness of the protein production: If mRNAs are short-lived compared to the proteins, i.e., if $1 \gg \Gamma_2/\Gamma_1$, then the number of proteins "jumps" every time there is an mRNA, and the jump height is the protein production per mRNA, β/Γ_1.

    (b) * Perform event-driven simulation of the full system with parameters of your choice, and confirm what you calculated was correct.

7.7 Further reading

This chapter is based on [2, 4, 10]

Chapter 8

Markov Process, Master equation, and Kramers-Moyal expansion

We have introduced master equations phenomenologically in the previous section, mainly for the one-step process. In this chapter, we learn the somewhat formal formulation of stochastic processes and then derive a general master equation from the formal definition.

First, we define a class of stochastic processes, called **Markov process**. It will turn out that all the stochastic processes we have considered in this course are Markov processes. We then show that we can derive a master equation for any given **homogeneous Markov process**. We do this for the discrete stochastic variable case (such as the number of particles), while the continuous stochastic variable case (such as the position of a Brownian particle) is mostly given in the appendix.

After that, we continue to learn an expansion method called **Kramers-Moyal expansion**, which, after some approximation, allows us to relate the master equation to the Fokker-Planck equation. This is a very useful method since there are many situations where master equations are hard to solve analytically while the approximated Fokker-Planck equations are tractable. We see some practical examples later in this course.

8.1 Markov Process

8.1.1 Memoryless property

Let us recall the Poisson process, where an event happens at a constant rate of r. We considered the number of events that happened up to time t, $N(t)$, and its probability distribution $P(n,t)$. Importantly, for this process, we could calculate the probability distribution in the future $t > t_0$ as soon as we know the probability distribution at time t_0. In other words, the behaviour for $t > t_0$ does not dependent on the history before t_0.

This was the same for a diffusive particle. The probability density $P(x,t)$ for being at position x at time $t > t_0$ could be calculated as soon as we know the probability distribution at time t_0; the process did not have a memory about what happened before t_0.

All the processes we considered so far, it was enough to know a "snapshot" of the systems state at time t_0 to describe the behaviour in the future time $t > t_0$. This is a soft definition of a "Markov process"; *history does not matter / the system does not have memory.* But we need to define it more precisely in a mathematical manner.

Transition probability

The Markov process is defined more precisely by using the conditional probability, expressed as $P(\text{state 1}|\text{state 2})$, meaning that probability to be in "state 1" given that it also takes the "state 2". As an example, consider a discrete stochastic variable $N(t)$ that can take the discrete values $n(t)$. We consider the transition probability $P(n_1,t_1|n_0,t_0)$, defined as the probability that $N(t_1) = n_1$ under the condition that $N(t_0) = n_0$, for $t_0 < t_1$.

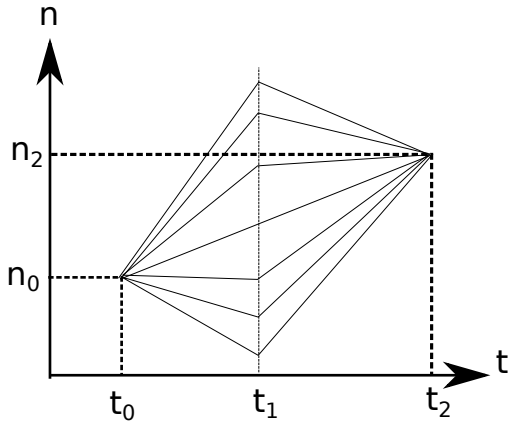


Figure 8.1: Possible path that a system can take upon transition from state $N(t_0) = n_0$ to $N(t_2) = n_2$.

This is defined as

$$P(n_1, t_1|n_0, t_0) = \frac{P_2(n_0, t_0; n_1, t_1)}{P(n_0, t_0)}, \tag{8.1}$$

where P_2 is defined as the probability to be $N(t_0) = n_0$ and $N(t_1) = n_1$:

$$P_2(n_0, t_0; x_1, t_1) = Pr(N(t_0) = n_0, N(t_1) = n_1). \tag{8.2}$$

The Markov process is defined as a *memoryless* stochastic process, i.e., all the history does not matter in the future. This is formally expressed as, for $t_0 < t_1 < \cdots < t_k$,

$$P(n_k, t_k|n_0, t_0; n_1, t_1; \cdots; n_{k-1}, t_{k-1}) = P(n_k, t_k|n_{k-1}, t_{k-1}). \tag{8.3}$$

The left-hand side means the probability of having $N(t_k) = n_k$ given that $N(t_0) = n_0$, $N(t_1) = n_1$, $\cdots$, $N(t_{k-1}) = n_{k-1}$. The right-hand side means that the probability of having $N(t_k) = n_k$ given that $N(t_{k-1}) = n_{k-1}$, *without any constraint on the history before* t_{k-1}. Hence this condition expresses the memory-less property.

8.1.2 Chapman-Kolmogorov equation

The memory-less property can also be written in a different way, for example

$$P(n_1, t_1; n_2, t_2|n_0, t_0) = P(n_2, t_2|n_1, t_1)P(n_1, t_1|n_0, t_0) \quad (t_0 < t_1 < t_2), \tag{8.4}$$

which means that the probability to be $N(t_1) = n_1$ and $N(t_2) = n_2$ given that $N(t_0) = n_0$ is equivalent to the product of the transition probability from $N(t_0) = n_0$ to $N(t_1) = n_1$ and the transition probability from $N(t_1) = n_1$ to $N(t_0) = n_0$, because the latter is independent of the former. Formally we can derive this from the memory-less property, which is left for the exercise.

This also means that a Markov process should satisfy **the Chapman-Kolmogorov equation** (CK equation)

$$P(n_2, t_2|n_0, t_0) = \sum_{\text{all possible } n_1} P(n_2, t_2|n_1, t_1)P(n_1, t_1|n_0, t_0); \tag{8.5}$$

Fig. 8.1 should help the meaning of this equation; The transition from $N(t_0) = n_0$ to $N(t_2) = n_2$ can be separated as a transition from $N(t_0) = n_0$ to $N(t_1) = n_1$ and $N(t_1) = n_1$ to $N(t_2) = n_2$. To get the full transition probability from $N(t_0) = n_0$ to $N(t_2) = n_2$, you need to multiply the transition probability from $N(t_0) = n_0$ to $N(t_1) = n_1$ and from $N(t_1) = n_1$ to $N(t_2) = n_2$, and sum the middle point n_1 up. The equality holds because how the system go to n_1 at time t_1 does not affect the future probability.

We will use the CK equation to derive the master equation.

8.2 From the Chapman-Kolmogorov equation to the master equation

We now show that a **homogeneous Markov process** is described by a master equation.

8.2.1 Homogeneous process

For many physical processes, the transition probability depends on the time difference only, i.e.,

$$P(n_1, t_1|n_0, t_0) = T(n_1|n_0, t_1 - t_0). \tag{8.6}$$

Such a process is called a **homogeneous process**. A Markov process is homogeneous if the transition rate is a constant or depends only on the state variable, while it becomes inhomogeneous if the transition rate depends explicitly on time. All the examples we saw in this lecture are homogeneous.

From eq. (8.6) and the CK eq.(8.5), for a **homogeneous Markov process** we have

$$T(n_2|n_0, \tau + \tau') = \sum_{\mathrm{a}ll possible n_1} T(n_2|n_1, \tau')T(n_1|n_0, \tau). \tag{8.7}$$

8.2.2 Derivation of the master equation

Now we are ready to derive a master equation for a homogeneous Markov process. Note that the transition probability satisfies the normalization condition

$$\sum_{all\ n_2} T(n_2|n_1, t) = 1. \tag{8.8}$$

This means that, if the system is in state n_1 at a time point, it will appear somewhere in the system after time t with probability one.

Now, it is easy to imagine that, for a very short time interval Δt, the transition probability $T(n_2|n_1, \Delta t)$ can be expanded as follows:

$$T(n_2|n_1, \Delta t) = F(\Delta t)\delta_{n_2,n_1} + \Delta t \cdot W(n_2|n_1) + O(\Delta t^2), \tag{8.9}$$

where the first term is the probability to stay in the same state and the second term is the transition probability for a very short time, with $W(n_2|n_1)$ being the transition rate from state n_1 to n_2. Substituting (8.9) to (8.8), we get

$$1 = F(\Delta t) + \Delta t \sum_{all\ n_2} W(n_2|n_1) + O(\Delta t^2), \tag{8.10}$$

which gives

$$F(\Delta t) = 1 - a_0(n_1)\Delta t + O(\Delta t^2), \tag{8.11}$$

$$a_0(n_1) = \sum_{all\ n_2} W(n_2|n_1), \tag{8.12}$$

i.e., the probability to move out from the state n_1 after a small time Δt is determined by the sum over all the transition rates going out from the state n_1. By substituting eq.(8.9) with (8.12) to (8.7) with $\tau' = \Delta t$, we get

$$\begin{aligned} T(n_2|n_0, \tau + \Delta t) &= \sum_{all\ n_1} \left[(1 - a_0(n_1)\Delta t)\delta_{n_2,n_1} + \Delta t \cdot W(n_2|n_1) + O(\Delta t^2)\right] \times T(n_1|n_0, \tau) \\ &= (1 - a_0(n_2)\Delta t)T(n_2|n_0, \tau) + \Delta t \sum_{all\ n_1} W(n_2|n_1)T(n_1|n_0, \tau) + O(\Delta t^2), \end{aligned}$$

which leads

$$T(n_2|n_0,\tau+\Delta t)-T(n_2|n_0,\tau) = \Delta t\left[\sum_{all\ n_1} W(n_2|n_1)T(n_1|n_0,\tau)-a_0(n_2)T(n_2|n_0,\tau)\right] + O(\Delta t^2). \tag{8.13}$$

By dividing both side of (8.13) by Δt and taking the limit of $\Delta t \to 0$, we get

$$\frac{\partial}{\partial\tau}T(n_2|n_0,\tau) = \sum_{all\ n_1}\left[W(n_2|n_1)T(n_1|n_0,\tau)-W(n_1|n_2)T(n_2|n_0,\tau)\right]. \tag{8.14}$$

Here, the definition of $a_0(n)$, eq. (8.12), is also used.

A more familiar form is

$$\frac{\partial}{\partial t}P(n,t) = \sum_{all\ m}\left[W(n|m)P(m,t)-W(m|n)P(n,t)\right],$$

because $T(n|m,t)$ can be considered as the probability density $P(n,t)$ for the initial condition at $t=0$ to be $P(n,t)=\delta_{n,m}$.

8.2.3 Example: Poisson process

When we introduced the one-step process in Chapter 7, we defined it based on the transition rate $W(n|m)$, and derived the solution. We can do it the other way around. For example, we can consider a discrete process defined for the integer space n with the transition probability

$$P(n_1,t_1|n_0,t_0)=T(n_1|n_0,t_1-t_0)=\begin{cases} 0 & \text{for } n_1<n_0 \\ \frac{(r(t_1-t_0))^{n_1-n_0}}{(n_1-n_0)!}e^{-r(t_1-t_0)} & \text{for } n_1\geq n_0. \end{cases} \tag{8.15}$$

This is clearly a homogeneous process, and it is straightforward to confirm that eq. (8.15) satisfies the CK equation (8.7) (left for the exercise), thus this is also a Markov process.

For small Δt, we can perform the expansion as

$$T(n_1|n_0,\Delta t)=\begin{cases} 0 & \text{for } n_1<n_0 \\ 1-r\Delta t+O(\Delta t^2) & \text{for } n_1=n_0, \\ r\Delta t+O(\Delta t^2) & \text{for } n_1=n_0+1, \\ O(\Delta t^2) & \text{for } n_1>n_0+1, \end{cases} \tag{8.16}$$

By comparing this with the expansion (8.9), we see that this process gives $F(\Delta t)=1-r\Delta t+O(\Delta t^2)$ and

$$W(n_1|n_0)=r\delta_{n_1,n_0+1}. \tag{8.17}$$

This leads to the master equation

$$\frac{\partial}{\partial t}P(n,t) = \sum_{m=-\infty}^{\infty}\left[r\delta_{n,m+1}P(m,t)-r\delta_{m,n+1}P(n,t)\right] \tag{8.18}$$

$$= rP(n-1,t)-rP(n,t). \tag{8.19}$$

This is the master equation of the Poisson process that we considered in Chapter 7.

This is a formal way of defining the Poisson process. You can find some mathematically oriented textbooks on a stochastic process that start with the definition of a homogeneous Markov process, and then define the transition property in a short time (like eq. (8.16)) to define the Poisson process, and derive the master equation. This chapter hopefully makes you more prepared to consult textbooks with such a style.

8.2.4 Continuous variable case

It is possible to extend what we did so far to continuum variable case. The logic is completely parallel, just many sums are replaced by integrals. The summary is given in Appendix D.

8.3 Kramers-Moyal Expansion

Now we can relate the Markov process to the master equations, which describe the time evolution of a probability distribution. By the way, in the early part of this course, we learned the Fokker-Planck equations, which also describe the time evolution of a probability distribution. They are not the same in general, but they are certainly related in some situations.

In this section, we learn an expansion method of a master equation, called the **Kramers-Moyal (KM) expansion**, which can be formally applied to any master equation. The expansion enables us to relate a master equation to a Fokker-Plank equation and thus relate to the things we learned in the first half of the course.

Let us perform the Kramers-Moyal expansion to a general master equation for discrete variable n:

$$\frac{\partial}{\partial t}P(n,t) = \sum_{all\ m} \left[W(n|m)P(m,t) - W(m|n)P(n,t)\right]. \tag{8.20}$$

Here, $W(n|m)$ is the transition rate from state m to n.

We re-express $W(n|m)$ with using the *jump size* $\Delta = n - m$, as

$$W(n|m) = W(n|n-\Delta). \tag{8.21}$$

Then we have

$$\begin{aligned}\frac{\partial}{\partial t}P(n,t) &= \sum_{all\ \Delta} \left[W(n|n-\Delta)P(n-\Delta,t) - W(n-\Delta|n)P(n,t)\right] \\ &= \sum_{all\ \Delta} \left[W(n|n-\Delta)P(n-\Delta,t) - W(n+\Delta|n)P(n,t)\right].\end{aligned} \tag{8.22}$$

In the last equality, I used the relation

$$\sum_{all\ \Delta} W(n-\Delta|n)P(n,t) = \sum_{all\ \Delta'} W(n+\Delta'|n)P(n,t).$$

For a cleaner notation, we here introduce a *step operator* E^{Δ}, which acts on any function of n, $f(n)$, as follows:

$$E^{\Delta}f(n) = f(n+\Delta). \tag{8.23}$$

Then we can rewrite eq. (8.22) as

$$\frac{\partial}{\partial t}P(n,t) = \sum_{all\ \Delta} \left[(E^{-\Delta}-1)W(n+\Delta|n)P(n,t)\right]. \tag{8.24}$$

If $f(n)$ is a smooth function of n, we can formally Tayler expand the step operator E^{Δ} as follows:

$$\begin{aligned}E^{\Delta}f(n) &= f(n+\Delta) \\ &= f(n) + \sum_{k=1}^{\infty} \frac{\Delta^k}{k!}\frac{\partial^k}{\partial n^k}f(n) = \left(1 + \sum_{k=1}^{\infty} \frac{\Delta^k}{k!}\frac{\partial^k}{\partial n^k}\right) f(n),\end{aligned}$$

namely,

$$E^{\Delta} = 1 + \sum_{k=1}^{\infty} \frac{\Delta^k}{k!}\frac{\partial^k}{\partial n^k}. \tag{8.25}$$

By using the relation (8.25), eq. (8.24) becomes (pay attention to the sign)

$$\frac{\partial}{\partial t}P(n,t) = \sum_{all\ \Delta}\left(\sum_{k=1}^{\infty}\frac{(-\Delta)^k}{k!}\frac{\partial^k}{\partial n^k}[W(n+\Delta|n)P(n,t)]\right) \tag{8.26}$$

$$= \sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\frac{\partial^k}{\partial n^k}[a_k(n)P(n,t)], \tag{8.27}$$

$$a_k(n) \equiv \sum_{all\Delta}\Delta^k W(n+\Delta|n). \tag{8.28}$$

$a_k(n)$ is called *the jump moment*. Note that $k=0$ of this expression is the same as (8.12).

Equation (8.27) is the Kramers-Moyal expansion of the master equation (8.20). Since it takes all the orders of the expansion, this is formally equivalent to the master equation.

8.3.1 Condition to truncate the KM expansion

When several conditions are met, the higher-order terms in this expansion are expected to be smaller. The conditions are:

- The jump happens only to the close positions. In other words, there is some distance $\delta > 0$, beyond which the transition rate is zero.

  $$W(n+\Delta|n) \approx 0 \quad \text{for} \quad |\Delta| > \delta.$$

  Otherwise, the jump moments may diverge.

- When $|\Delta| < \delta$, the transition rate $W(n+\Delta|n)$ is a slowly varying function of n.

- $P(n,t)$ varies slowly with n.

If these conditions are satisfied, we can expect that the higher order term of the derivative is smaller. If we truncate the expansion up to the second order, we get

$$\frac{\partial}{\partial t}P(n,t) = -\frac{\partial}{\partial n}[a_1(n)P(n,t)] + \frac{1}{2}\frac{\partial^2}{\partial n^2}[a_2(n)P(n,t)], \tag{8.29}$$

which has the form of the Fokker-Planck equation!

8.3.2 Example: a one-step process

We can do this explicitly to the master equation of a one-step process

$$\frac{\partial}{\partial t}P(n,t) = r(n+1)P(n+1,t) - r(n)P(n,t) + g(n-1)P(n-1,t) - g(n)P(n,t). \tag{8.30}$$

(Note that $r(n)$ and $g(n)$ are functions of n.) In this process, transition happens from state n to $n \pm 1$ and *vise versa*, i.e., the jump size Δ takes ± 1.

When we make the Tayler expansion to the first and the third term, we get

$$\frac{\partial}{\partial t}P(n,t) = \sum_{k=1}^{\infty}\frac{1}{k!}\frac{\partial^k}{\partial n^k}[r(n)P(n,t)] + \sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\frac{\partial^k}{\partial n^k}[g(n)P(n,t)]. \tag{8.31}$$

This is the Kramers-Moyal expansion of the one-step process.

If we truncate the KM expansion up to the second order in derivative, we get

$$\frac{\partial}{\partial t}P(n,t) = \frac{\partial}{\partial n}[(r(n)-g(n))P(n,t)] + \frac{1}{2}\frac{\partial^2}{\partial n^2}[(r(n)+g(n))P(n,t)]. \tag{8.32}$$

This has the functional form of the Fokker-Planck equation.

The error by truncation will be smaller if $r(n)P(n,t)$ and $g(n)P(n,t)$ are smooth functions of n and do not change much when n changes by the jump size of the one-step process (one). However, note that there is no solid reason why one truncates at the second-order expansion. Of course, as we know from the previous analysis of Fokker-Planck equations, if we truncate at the first order, it only describes the drift; the second-order term gives the diffusive term, i.e., probability distribution may spread out and we can talk about variance. Thus, we may say that we need at least up to the second order if we want to talk about fluctuations.

Examples of one-step processes

Random walk Random walk is given by $r(n) = g(n) = c$ (c.f. (7.28)). In this case, (8.32) gives

$$\frac{\partial}{\partial t}P(n,t) = c\frac{\partial^2}{\partial n^2}[P(n,t)], \tag{8.33}$$

which is a diffusion equation (with $D = c$).

Simple chemical reaction Consider the simple chemical reaction, where a molecule is produced with a rate of $\alpha\Omega$ and degraded with a rate of Γ per molecule. In this case, the KM expansion up to the second order gives

$$\frac{\partial}{\partial t}P(n,t) = \frac{\partial}{\partial n}\left[(\Gamma n - \alpha\Omega)\,P(n,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial n^2}\left[(\Gamma n + \alpha\Omega)\,P(n,t)\right]. \tag{8.34}$$

Note that the first term could be interpreted as "flux due to the force from a potential (see lecture 7) $U(n) \propto \frac{1}{2}\Gamma\left(n - \frac{\alpha\Omega}{\Gamma}\right)^2$. The second derivative term has n-dependence, i.e. the "diffusion" depends on n.

Decay process Consider the decay process, where the number of particles N decays with a constant rate of γ. This is a special case of the simple chemical reaction, with $\alpha = 0$ and $\Gamma = \gamma$. In this case, the KM expansion up to the second order gives

$$\frac{\partial}{\partial t}P(n,t) = \frac{\partial}{\partial n}\left[\gamma n P(n,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial n^2}\left[\gamma n P(n,t)\right]. \tag{8.35}$$

This corresponds to the FP equation that you would get for the Ito-type SDE

$$dN = -\gamma N dt + \sqrt{\gamma N}\cdot dB. \tag{8.36}$$

This is telling that our argument in chapter 5 for modelling the decay process by an SDE was correct. In general, the KM expansion can be used to find a corresponding approximation SDE of a stochastic process.

Coalescence process This is kept for exercise.

8.3.3 Continuum variable case

It should be noted that the Kramers-Moyal expansion can be done for continuous variable cases, too. The derivation and application to the diffusion process are given in appendix E.

8.4 Summary

- Markov process is a memory-less stochastic process.
- Markov process satisfies the Chapman-Kolmogorov equation.

- KM expansion is possible for any master equation. For a discrete variable case, we get

$$\begin{aligned}\frac{\partial}{\partial t}P(n,t) &= \sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\frac{\partial^k}{\partial n^k}[a_k(n)P(n,t)],\\ a_k(n) &\equiv \sum_{all\,\Delta}\Delta^k W(n+\Delta|n).\end{aligned}$$

- Truncation of the expansion at the second order gives an FP equation as an approximation, which enables us to apply methods we learned for FP equations to the problem that was originally expressed in terms of the master equation.

8.5 Exercise

1. Derive (8.4) from the memoryless property. Hint: Use the definition of the conditional probability (8.1).
2. Confirm that eq. (8.15) satisfies the Chapman-Kolmogorov Equation (8.5). hint: $(a+b)^k = \sum_{m=0}^{k}\frac{k!}{m!(k-m)!}a^m b^{k-m}$.
3. Consider the coalescence reaction $A + A \to A$, which was defined in (7.44).
   (a) Perform the KM expansion of the master equation to show

$$\frac{\partial P(n,t)}{\partial t} = \frac{\lambda}{\Omega}\sum_{k=1}^{\infty}\frac{1}{k!}\frac{\partial^k}{\partial n^k}\left[n(n-1)P(n,t)\right].$$

   (b) Truncate the KM expansion at the 2nd order. Unfortunately, it is not very easy to solve either. Use the obtained FP equation, and assuming $n^k P(n,t) \to 0$ for $k = 0, 1, 2, 3$ and $n(\partial/\partial n)(n(n-1)P(n,t) \to 0)$ for $n \to \infty$, show

$$\frac{d}{dt}\langle N\rangle = -\frac{\lambda}{\Omega}\left(\langle N^2\rangle - \langle N\rangle\right).$$

   So the truncation of KM approximation keeps the correlations in this order.

Advanced exercises

4. Consider the protein production process considered in exercise 7b in chapter 7. Perform KM expansion of the master equation for $P(m,p,t)$. Note that there are two variables, therefore there will be derivatives by m and p in the final result. Truncate the expansion in the second order to derive a Fokker-Planck equation.

8.6 Further reading

This chapter is based on [2, 19, 9].

Chapter 9

System size expansion: The Coalescence process as an example

In the Kramers-Moyal expansion, we expanded the master equation by assuming "jump size" is limited. This method works fairly universally, but it is not a very systematic expansion, in the sense that we do not have a well-defined parameter that denotes how small the higher-order terms are.

In many examples so far, we dealt with the quantity that can be measured as to "concentration" or "density". If we consider a chemical reaction, the reaction rate is better defined by considering the number of molecules divided by the volume of the system. In such a system, the noise comes very often due to the finiteness of the number of molecules, and the noise becomes negligibly small in the infinite "volume" with a constant density limit. Then it is convenient to formulate expansion around this infinite volume, deterministic limit to take into account the noise from finite size.

In this chapter, we learn how to do it by taking the coalescence process that we have been looking at as a specific example. We re-express the process so that the limit of *constant density with large volume* is easy to take, because the limit will give a reference deterministic system described by the rate equations - Stochastic treatment will then address the stochasticity due to the finite size of the system. By doing this, it becomes also possible to expand the master equation around the deterministic limit, which is called **(van Kampen's) system size expansion** or **Ω-expansion**.

The method is applicable for general "chemical reactions" or stochastic processes where the "density" is well defined, so even though we take a specific example here, follow the logic carefully so that you can apply it to another example. The merit of taking an example is that you get a concrete feeling of the order of the expansion. It is of course possible to perform the expansion more formally for general "chemical reactions", and that is given in the appendix F.

9.1 Coalescence process

Let's recap the coalescence process, that we considered in section 7.2.3. There are N particles in a system of volume of Ω, and if two particles (molecules) collide, they merge and become one particle (Coalescence), i.e., the reaction is $A + A \to A$. The rate for N particles to coalesce to become $N-1$ particles is given by

$$\lambda N \frac{N-1}{\Omega},$$

and the master equation for probability $P(n,t)$ to have n particles at time t is given by eq. (7.44), which is

$$\frac{\partial}{\partial t}P(n,t) = \frac{\lambda}{\Omega}(n+1)nP(n+1,t) - \frac{\lambda}{\Omega}n(n-1)P(n,t) = \frac{\lambda}{\Omega}\left[\left(E^1-1\right)n(n-1)P(n,t)\right], \tag{9.1}$$

for $n \geq 0$. Here, we used the step operator defined as eq. (8.23). We have also calculated in the exercise that the average number of particles obeys the equation (7.46)

$$\frac{d}{dt}\langle N(t)\rangle = -\frac{\lambda}{\Omega}(\langle N(t)^2\rangle - \langle N(t)\rangle). \tag{9.2}$$

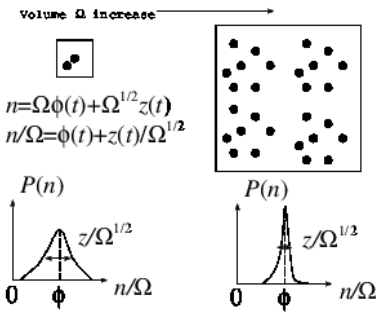


Figure 9.1: Schematic description of the system size expansion.

9.1.1 Describing the system by using the density and large volume limit

The first step to make an approximation is to rewrite the variable from the number of particles to the density of the particles:

$$X \equiv \frac{N}{\Omega}.$$

If we increase the volume Ω *with keeping the average density constant*, we normally reach the thermodynamics limit, where the density

$$\phi \equiv \lim_{\Omega\to\infty} X \tag{9.3}$$

solely determine the dynamics. Now, dividing both sides of (9.2) by Ω, we get

$$\frac{d}{dt}\langle X(t)\rangle = -\lambda\left(\langle X(t)^2\rangle - \frac{\langle X(t)\rangle}{\Omega}\right). \tag{9.4}$$

The naive expectation in the thermodynamic limit is that the density fluctuation becomes negligibly small, hence we expect

$$\phi = \lim_{\Omega\to\infty} X = \lim_{\Omega\to\infty}\langle X\rangle$$

and

$$\lim_{\Omega\to\infty}\left[\langle X(t)^2\rangle - \langle X(t)\rangle^2\right] = 0.$$

Also, in the limit, we have

$$\lim_{\Omega\to\infty}\frac{\langle X(t)\rangle}{\Omega} = \lim_{\Omega\to\infty}\frac{\phi}{\Omega} = 0.$$

Therefore, in the large volume with a constant density limit, we expect to have

$$\frac{d}{dt}\phi = -\lambda\phi^2. \tag{9.5}$$

to hold. We then want to expand N around ϕ for finite Ω, to describe the noise in a large but still finite volume system.

9.2 System size expansion

9.2.1 Fluctuation around the density in the large volume limit

Let us divide the number of molecules at time t into two terms:

$$N(t) = \Omega\phi(t) + \sqrt{\Omega}Z(t).$$

Here, ϕ is the average density. In other words, we have

$$\frac{N(t)}{\Omega} = \phi(t) + \frac{Z(t)}{\sqrt{\Omega}}. \tag{9.6}$$

We consider the situation where we change Ω with keeping the density ϕ constant (Fig.9.1). Intuitively $\Omega \to \infty$ limit will be described fully by $\phi(t)$. The part $\sqrt{\Omega}Z$ describes the fluctuation in the finite volume system. The *assumed* square-root dependence on the volume Ω is a *guess* based on the central limit theorem, i.e., variance is proportional to the system size [1]. Note that Z is a stochastic variable. We will derive the equation for the probability distribution of Z for large but finite Ω.

In the approximation we do in this section, we will be able to set

$$\langle Z(t) \rangle = 0 \tag{9.7}$$

all the time (we will see this later). Then from (9.6), we will have

$$\langle N(t) \rangle = \Omega\phi(t) \tag{9.8}$$

$$\langle (N(t) - \langle N(t) \rangle)^2 \rangle = \Omega \langle Z(t)^2 \rangle. \tag{9.9}$$

Namely, we are assuming that, when we change Ω, keeping the density $\phi(t) = \langle N(t) \rangle / \Omega$ constant, the variance of N will be proportional to the volume Ω.

9.2.2 Derivation

Outline of the expansion

Under these assumptions, we derive the equation for the probability distribution for Z, $Q(z,t)$, and simplify it.

The outline of the derivation is as follows:

1. Convert the master equation for $P(n,t)$ to the equation for $Q(z,t)$.
2. (System size expansion): Expand the equation around the large Ω limit (in power of $1/\sqrt{\Omega}$).
3. (Linear Noise Approximation, LNA): Take the expanded equation up to order $1/\sqrt{\Omega}$ relative to the deterministic order to derive a simplified equation for $Q(z,t)$.

Distribution for Z

Our starting point is the master equation (9.1), which can be slightly reformulated as

$$\frac{\partial}{\partial t} P(n,t) = \Omega\lambda \left[\left(E^1 - 1 \right) \frac{n}{\Omega} \cdot \frac{(n-1)}{\Omega} P(n,t) \right] \tag{9.10}$$

We need to convert $P(n,t)$ to the probability $Q(z,t)dz$, the probability that Z takes the value between z and $z + dz$ (i.e. $Q(z,t)$ is the probability density). By noticing that Z changes by $1/\sqrt{\Omega}$ when N change by 1 by fluctuation, we have the conversion

$$P(n,t) = \Omega^{-1/2} Q(z,t). \tag{9.11}$$

Time derivative

First, we consider the LHS of (9.10), the partial time derivative. We should note that the derivative is for constant n and varying t, which is NOT constant z. Namely

$$\frac{\partial}{\partial t} P(n,t) \equiv \left. \frac{d}{dt} P(n,t) \right|_{n=const.} = \Omega^{-1/2} \left. \frac{d}{dt} Q(z,t) \right|_{n=const.} = \Omega^{-1/2} \left[\frac{\partial}{\partial t} Q(z,t) + \left(\frac{dz}{dt} \right)_{n=const} \frac{\partial}{\partial z} Q(z,t) \right]. \tag{9.12}$$

The time derivative of $n = \Omega\phi(t) + \sqrt{\Omega} z(t)$ with constant n gives

$$0 = \Omega \frac{d\phi(t)}{dt} + \sqrt{\Omega} \frac{dz(t)}{dt}. \tag{9.13}$$

[1] Obviously the method is applicable only for the cases where this assumption is correct. If this is not correct, the expansion will fail. Such an example will be given in the next section.

So we have $dz/dt = -\sqrt{\Omega}d\phi/dt$. Therefore

$$\frac{\partial}{\partial t}P(n,t) = \Omega^{-1/2}\left[\frac{\partial}{\partial t}Q(z,t) - \sqrt{\Omega}\frac{d\phi(t)}{dt}\frac{\partial}{\partial z}Q(z,t)\right]. \tag{9.14}$$

Expansion of RHS

When we change the variable from n to z, RHS of the mater equation (9.10)becomes

$$RHS = \sqrt{\Omega}\lambda\left[\left(E^1 - 1\right)\left(\phi + \frac{z}{\sqrt{\Omega}}\right)\cdot\left(\phi + \frac{z}{\sqrt{\Omega}} - \frac{1}{\Omega}\right)Q(z,t)\right] \tag{9.15}$$

First, let us expand the step operator. Previously when we did the KM expansion, we expanded the step operator in terms of n, but now our variable is z. Therefore, we do the following expansion: From the definition (8.23)[2]

$$\begin{aligned}
& E^{\Delta}f(n) = f(n+\Delta) \\
& = f(\Omega\phi + \sqrt{\Omega}z + \Delta) = f(\Omega\phi + \sqrt{\Omega}(z + \Delta/\sqrt{\Omega})) \\
& = f(\Omega\phi + \sqrt{\Omega}z) + \frac{\Delta}{\sqrt{\Omega}}\frac{\partial}{\partial z}f(\Omega\phi + \sqrt{\Omega}z) + \frac{1}{2}\left(\frac{\Delta}{\sqrt{\Omega}}\right)^2\frac{\partial^2}{\partial z^2}f(\Omega\phi + \sqrt{\Omega}z) + \cdots \\
= & \left[1 + \frac{\Delta}{\sqrt{\Omega}}\frac{\partial}{\partial z} + \frac{\Delta^2}{2\Omega}\frac{\partial^2}{\partial z^2} + O(\Omega^{-3/2})\right]f(n).
\end{aligned} \tag{9.16}$$

In addition, because the reaction rates depend on Ω, we can also expand them. Specifically, we use

$$\left(\phi + \frac{z}{\sqrt{\Omega}}\right)\cdot\left(\phi + \frac{z}{\sqrt{\Omega}} - \frac{1}{\Omega}\right) = \phi^2 + \frac{2\phi z}{\sqrt{\Omega}} + O(1/\Omega). \tag{9.17}$$

Expansion of the master equation

By substituting (9.14), (9.16) with $\Delta = 1$, and (9.17) into (9.10), we get

$$\begin{aligned}
& \Omega^{-1/2}\left[\frac{\partial}{\partial t}Q(z,t) - \sqrt{\Omega}\frac{d\phi(t)}{dt}\frac{\partial}{\partial z}Q(z,t)\right] \\
= & \sqrt{\Omega}\left[\left(1 + \frac{1}{\sqrt{\Omega}}\frac{\partial}{\partial z} + \frac{1}{2\Omega}\frac{\partial^2}{\partial z^2} + O(\Omega^{-3/2})\right) - 1\right] \\
& \times\lambda\left(\phi^2 + \frac{2\phi z}{\sqrt{\Omega}} + O(1/\Omega)\right)Q(z,t).
\end{aligned} \tag{9.18}$$

In order that this equation makes physical sense in $\Omega \to \infty$, each order of $\sqrt{\Omega}$ should balance.

9.3 Linear noise approximation

9.3.1 Equation of order 1

The largest terms are the order one (Ω^0) terms, which gives

$$\begin{aligned}
& -\frac{d\phi(t)}{dt}\frac{\partial}{\partial z}Q(z,t) = \lambda\frac{\partial}{\partial z}\phi^2 Q(z,t) \\
\Leftrightarrow & \left[\frac{d\phi(t)}{dt} + \lambda\phi^2\right]\frac{\partial}{\partial z}Q(z,t) = 0.
\end{aligned} \tag{9.19}$$

[2]If you have a problem with this expansion, you can do the following intermediate steps: Define $g(z) \equiv= f(\Omega\phi + \sqrt{\Omega}z)$. Then $g(z + k/\sqrt{\Omega}) = f(\Omega\phi + \sqrt{\Omega}z)$. Do the Tayler expansion of $g(z + k/\sqrt{\Omega})$ around z, and then use $g(z) = f(n)$ to reach the final expression.

By demanding that this relation should hold when $Q(z,t)$ is a function of z and not a constant (which we will see just below), we get the deterministic rate equation

$$\frac{d\phi(t)}{dt} = -\lambda\phi^2 \tag{9.20}$$

which is exactly as expected, see (9.5).

9.3.2 Equation of order $1/\sqrt{\Omega}$

This is order $\Omega^{-1/2}$ relative to the deterministic order, so it is the *linear order in noise* (see (9.6)). Approximation up to this order is called the *linear noise approximation*.

$$\begin{aligned}\frac{\partial}{\partial t}Q(z,t) &= \left(\frac{\partial}{\partial z}\right)(2\lambda\phi z)\,Q(z,t) + \left(\frac{1}{2}\frac{\partial^2}{\partial z^2}\right)\lambda\phi^2 Q(z,t) \\ &= 2\lambda\phi\frac{\partial}{\partial z}\left(zQ(z,t)\right) + \frac{1}{2}\lambda\phi^2\frac{\partial^2}{\partial z^2}Q(z,t).\end{aligned} \tag{9.21}$$

9.3.3 "Coefficients" depends on $\phi(t)$

For convenience, we define the following functions:

$$J(\phi) = -2\lambda\phi, \tag{9.22}$$

and

$$D(\phi) = \lambda\phi^2. \tag{9.23}$$

By using these, we can write

$$\frac{\partial}{\partial t}Q(z,t) = -J(\phi)\frac{\partial}{\partial z}\left(zQ(z,t)\right)) + \frac{1}{2}D(\phi)\frac{\partial^2}{\partial z^2}Q(z,t). \tag{9.24}$$

This have the form of the Fokker-Planck equation! But note that J and D depends on $\phi(t)$, which changes over time according to the rate equation.

In general, the LNA of a one-variable master equation for chemical reactions gives the form (9.24); the difference of different systems will be reflected in the functional form of $J(\phi)$ and $D(\phi)$. See appendix F for the general form.

9.3.4 Average, variance, and distribution

Equation for $\langle Z\rangle$

From (9.24), we can derive an equation for $\langle Z(t)\rangle$, by multiplying both sides with z and integrating over z. Using integration by parts with the boundary condition that all $z^2Q \to 0$ for $z \to \pm\infty$, $z(\partial Q/\partial z) \to 0$ for $z \to \pm\infty$, etc., we get

$$\frac{d}{dt}\langle Z(t)\rangle = J(\phi)\langle Z(t)\rangle. \tag{9.25}$$

As the initial condition, we can always have $\langle Z(0)\rangle = 0$ without loss of generality (We just define that $\phi(0) = \langle N(0)\rangle/\Omega$). Once we do this, we have

$$\langle Z(t)\rangle = 0 \tag{9.26}$$

at any time.

LNA gives the Gaussian distribution

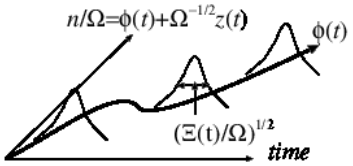
$$Q(z,t) = \frac{1}{\sqrt{2\pi\Xi(t)}} \exp\left[-\frac{z^2}{2\Xi(t)}\right], \tag{9.27}$$

$$\dot{\Xi}(t) = 2J(\phi(t))\Xi(t) + D(\phi(t)), \tag{9.28}$$

$$\Xi(t) = \langle Z(t)^2 \rangle. \tag{9.29}$$

is the solution of (9.24). Namely, the distribution is always Gaussian with an average of zero (cf. eq.(9.26)), but the variance $\Xi(t)$ changes over time.

So we can have the following view: The average density follows the deterministic rate equation (9.20), and fluctuation around it (described by Z) is Gaussian with variance determined by (9.28).



9.3.5 Solving the mean and variance

From the LNA of the coalescence process, we have obtained

$$\dot{\phi} = -\lambda\phi^2, \tag{9.30}$$

$$\frac{\partial}{\partial t}Q(z,t) = 2\lambda\phi\frac{\partial}{\partial z}(zQ(z,t)) + \frac{1}{2}\lambda\phi^2\frac{\partial^2}{\partial z^2}Q(z,t). \tag{9.31}$$

The rate equation (9.30) is solvable (exercise), and we get

$$\phi(t) = \frac{1}{1+\lambda t}. \tag{9.32}$$

when $\phi(0) = 1$. Further, comparing equation for the variance (9.28) and (9.31), we get

$$\dot{\Xi}(t) = -4\lambda\phi(t)\Xi(t) + \lambda\phi(t)^2, \tag{9.33}$$

and this is also a closed equation given (9.32) (exercise) to give

$$\Xi(t) = \frac{1}{3(1+\lambda t)}\left[1 - (1+\lambda t)^{-3}\right] \tag{9.34}$$

under the initial condition $\Xi(0) = 0$. So, the LNA made the problem solvable!

Deriving these solutions and comparing the approximated solution with the numerical simulation are left for the exercise. You will see it is quite a good approximation for this example as long as the number of the particle N is bigger than 10. When N becomes very small, or equivalently when Ω is small, the approximation becomes worse, which is natural because our expansion is around the large Ω limit.

9.4 Summary

- In the thermodynamic (large volume) limit, the density $\phi(t)$ obeys a rate equation.
- Assuming that $N(t) = \phi(t)\Omega + \sqrt{\Omega}Z$, one can perform systematic expansion of the master equation around the large volume limit (System size expansion).
- The linear noise approximation is obtained at $O(1/\sqrt{\Omega})$ relative to the large volume limit. The obtained expression is formally solvable, giving the Gaussian distribution with the variance dependent on the time-dependent density $\phi(t)$.

9.5 Exercise

1. Derive (9.25).
2. Show that (9.27) with (9.28) is a solution of (9.24).

   Hint: Substitute (9.27) to (9.24) and take partial derivatives.
3. Perform the system size expansion and linear noise approximation of the decay process, where each particle has a constant rate γ to decay, i.e., the number of the particles in the system N obeys

$$N \to N - 1 \quad \text{with a rate} \quad \gamma N.$$

4. LNA of the coalescence process
   (a) Show that the solution of (9.30) with $\phi(0) = 1$ is given by (9.32).
   (b) Solve this equation (9.33) using the solution (9.32) to show (9.34).
   (c) * Simulate the coalescence process $A+A \to A$ with $\lambda = 1$, $\Omega = 100$, with initial condition $N = 100$ (Therefore the initial density $\phi_0 = 1$, and the initial variance is zero). Collect 100 samples and plot the average and the variance of the number of the particles per system (N/Ω) as a function of time. Compare the results with the LNA prediction (9.32) and (9.34) by plotting the data and the LNA result together.

   the log-log plot is recommended for the average vs time plot, and semi-log (log for the time axis) is recommended for variance vs time plot. Pay attention to the factor Ω when comparing Ξ and the simulation.

   Tips for computation: You can use either the Gillespie algorithm or the time-driven algorithm. If you use the Gillespie algorithm, you need to take special care when you average over samples as a function of time: We need to take averages at the same time point, while in the Gillespie algorithm, the time intervals of events are different in different samples.

   If you use the time-driven algorithm, it is straightforward to take averages over samples simultaneously. Therefore I recommend the time-driven method as the first try (though slower to compute). To get reasonable results, the time step dt needs to be small enough compared to the total reaction rate for the given moment - i.e., $(\lambda N(N-1)/\Omega)dt$ should be reasonably smaller than 1.

Advanced exercise

5. Perform system size expansion and LNA to the multi-component chemical reaction scheme (F.11). Apply it to the gene expression scheme in exercise 7 in Chapter 7 (including both mRNA and protein), and see how the variance of the protein number depends on the parameters. (Answer is in appendix F but try it yourself first.)

9.6 Further reading

This chapter is based on [2, 9].

Chapter 10

Interacting in space: Traffic flow models

When we were dealing with the Brownian motion, we consider the time evolution of the position of the particle, but we did not consider the interaction between particles. When we were dealing with the chemical reactions, particles interacted but we did not consider the position of the particles explicitly.

In this chapter, we consider interacting particles in space - we go for one of the simplest and yet the non-trivial cases, hard-core repulsion of particles in one dimension with the flow in one direction. This can be considered as a model for “traffic flow".

10.1 Traffic jam / Density wave formation

10.1.1 Real phenomena

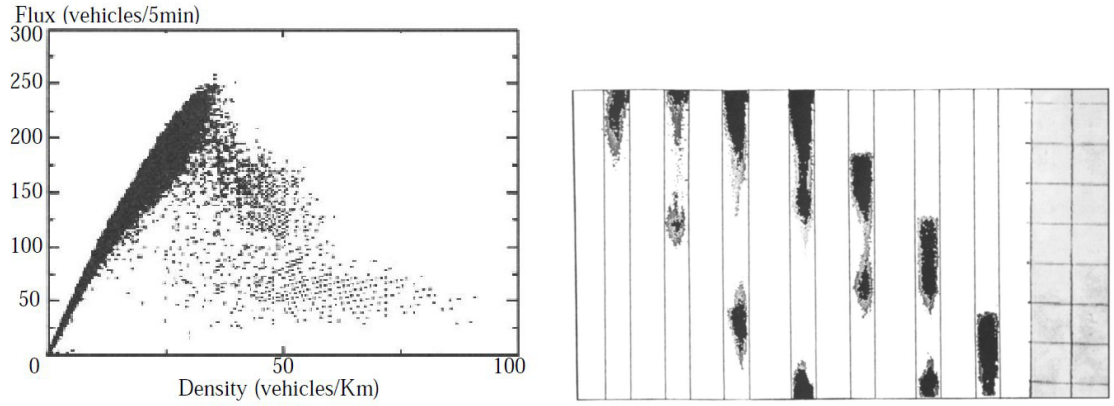


Figure 10.1: Left: Fundamental diagram from Tomei-highway, Japan. Right: Density wave formation in granular pipe flow, from [20].

You may think that traffic jams occur only when there is a clear cause, such as an accident, construction work, etc., and that it is hard to characterize traffic jams. The data, however, indicates something universal and robust about the phenomena. Fig. 10.1 left shows a so-called fundamental diagram, where the flux of cars is plotted as a function of car density for one of the exceptionally busy highways in Japan. We see that flux grows linearly with density when the density is low, as intuitively expected. Above a certain density, here it seems to be about 30 vehicles/Km, the flux drops significantly. This is an indication of a traffic jam. The drop of flux above a critical density is a quite robust feature and has been observed in many places many times.

Cars are not the only things that "jam". Fig. 10.1 right shows a sequence of snapshots of granular flow in a vertical pipe. A density wave is spontaneously formed. This is also a kind of "traffic jam". It is known that many flowing systems in quasi-one-dimensional configuration show traffic jam or density wave formation.

Another example in a very different length scale is the protein synthesis from a messenger RNA (translation process). In the translation process, a protein complex called ribosome binds to mRNA, reads the information on it expressed by a set of 3 nucleotides (codon), and translates it to the sequence of amino acids. At a given moment, many ribosomes can translate the same mRNA (Fig. 10.2). ASEP model, which we study in this chapter, was first introduced as a model of ribosome kinetics.

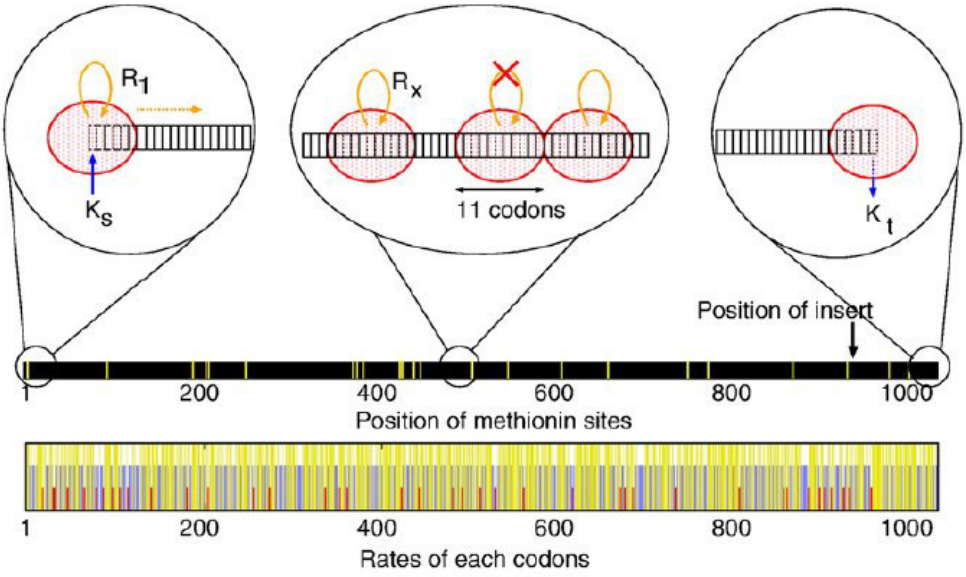


Figure 10.2: Schematic description of ribosome traffic on mRNA, from [21]

10.1.2 Modeling traffic flow

There are various models to describe the traffic jam problem. All of the following can be found in the literature:

- Hydrodynamic model (Partial differential equation)
- Car-following model (Differential equations with time delay)
- Optimal velocity model (Coupled Ordinary differential equations)
- Asymmetric Simple Exclusion Process (ASEP) and extensions (Continuous/Discrete-time, discrete space, discrete states)
  - Cellular Automata (CA) (Discrete space and time, Discrete states)

In this chapter, we focus on ASEP and a simple related CA, which are straightforward stochastic models and yet show various non-trivial features, giving us some insight into the traffic jam formation.

10.2 Totally Asymmetric Simple Exclusion Process (TASEP)

10.2.1 ASEP and TASEP

First let us introduce the ASEP model. The situation is schematically described in Fig. 10.3. There are many particles on a one-dimensional lattice of Length L (L sites). At each site of the lattice, there is either

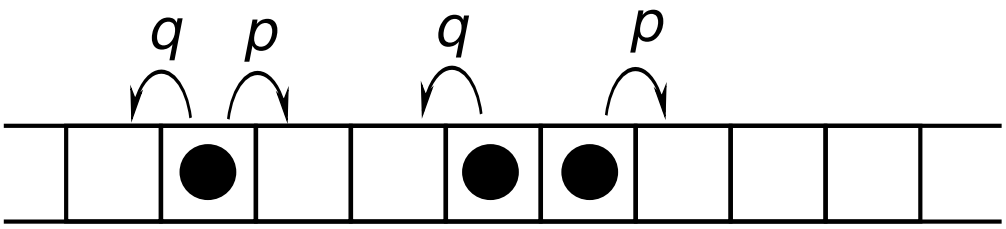


Figure 10.3: Schematic description of ASEP.

one particle or none (empty), but there is no multiple occupancy. This reflects the fact that particles exclude each other and cannot overlap.

In such a system, we consider the following dynamics:

- During each time step Δt, each particle has probability $p\Delta t$ of attempting a jump to its right and probability $q\Delta t$ of attempting a jump to its left. A jump can only succeed if the target site is empty.

This is the most general setup of Asymmetric (i.e. $p \neq q$ possible), Simple Exclusion (i.e. particles cannot overlap) Process. Especially, the following cases are often considered separately:

- $p = q$: Symmetric.
- $p \neq q$: Asymmetric.
- $q = 0$ and $p \neq 0$: Totally Asymmetric.

In this chapter, we mainly consider the Totally Asymmetric Simple Exclusion Process (TASEP), i.e., $q = 0$ and $p \neq 0$. This is a good model for traffic flow, where particles (can be cars) move only in one direction.

10.2.2 Two possible dynamics

The dynamics above are not enough to define the process when there are many particles moving. One can consider the following two cases:

(A) The limit of $\Delta t = dt \to 0$; Continuous-time dynamics with a hopping rate p (Random walk in continuous time if there is only one particle). In the limit, only one particle jumps per dt at most.

(B) Finite Δt (then Δt can be set to one without loss of generality), i.e., time is always discrete. For example, for a car to move its size always take finite time, so if we model traffic flow with discrete space, one may argue that time should also be discrete.

When one says (T)ASEP, it often means the dynamics (A), continuous-time dynamics, since it can be treated analytically and has been provided with a lot of knowledge about flowing systems with particle interaction. Case (B) has been studied well as a car-traffic model.

Case (A) and (B) should be simulated in different ways, and the outcomes are also different. Here, we first consider the continuous-time case (A), then move on to the discrete-time case (B).

10.3 TASEP in continuous time

Let us consider the continuous-time TASEP for a while, which we can analyze with an analytical method better than the discrete-time case. In fact, many quantities in continuous time ASEP has been solved exactly, though the calculations are often rather complicated. Interested readers can refer to [22] and references therein (there are also more recent developments). We here focus on a few interesting behaviours of TASEP and an approximation method to understand them in an easier way.

10.3.1 Numerical simulation of the continuous-time TASEP

One way to simulate it is by applying the Gillespie algorithm, by listing the movements of all particles as possible next events. This can, however, be somewhat heavy to code. The more common way to simulate it is to use a "time-driven" method, where one simulates the movement for a finite time step Δt. The important thing to note is that within Δt, many particles can move, but at a given (infinitely short) instance, at most only one particle can move. This leads to the *random sequential updates*.

(A) Random sequential updates

1. Choose a lattice site in the system randomly.
2. If there is a particle in the chosen site, and there is no particle in front, then make it hop one site with the probability per time step $p\Delta t$. Otherwise, do nothing.
3. Repeat 1 and 2 for L times (All the lattice sites are updated once on average).
4. Proceed time by Δt.

By repeating procedures 1 to 4, we can simulate the long-time behaviour. This method does not give the time resolution smaller than Δt, but it reproduces the behaviour properly for the time scale bigger than Δt, as long as Δt is chosen so that $p\Delta t \leq 1$. We use this method in the exercise to simulate the model behaviour and compare it with the analytical results we derive in the following.

10.3.2 Master Equation

Do we know a good tool to analyze the TASEP? How about writing down the master equation for TASEP?

First, let us consider the case with only one particle in the system. Probability $P_x(t)$ to have the particle at position x at time t obeys:

$$\dot{P}_x = p\left[P_{x-1}(t) - P_x(t)\right].$$

If there are two particles in the system, we should consider probability $P_{x_1,x_2}(t)$ to have the particle 1 at position x_1 and particle 2 at position x_2 (we limit $x_1 < x_2$ since a particle cannot overtake others) at time t. This obeys:

$$\dot{P}_{x_1,x_2} = p\left[P_{x_1-1,x_2}(t) + P_{x_1,x_2-1}(t) - 2P_{x_1,x_2}(t)\right] \quad \text{for} \quad x_2 - x_1 \geq 2.$$
$$\dot{P}_{x_1,x_2} = p\left[P_{x_1-1,x_2}(t) - P_{x_1,x_2}(t)\right] \quad \text{for} \quad x_2 - x_1 = 1.$$

By denoting the position of the i-th particle as x_i, we can extend this to N-particle case:

$$\dot{P}_{x_1,x_2,\cdots,x_N} = \sum_{j=1}^{N} p\left[P_{x_1,x_2,\cdots,x_{j-1},x_j-1,x_{j+1},\cdots,x_N} - P_{x_1,x_2,\cdots,x_{j-1},x_j,x_{j+1},\cdots,x_N}\right] \tag{10.1}$$

for $x_j - x_{j-1} \geq 2$ for any j. Otherwise, the equation needs to be properly modified.

10.3.3 Mean field approximation of TASEP

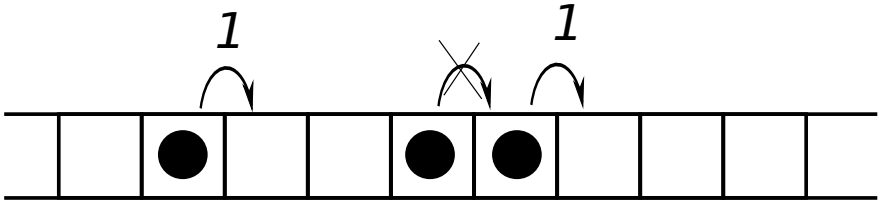


Figure 10.4: Schematic description of TASEP.

Though we can write down the master equation, it is quite complex to solve. Actually, we can construct a better quantity to analyze TASEP. In this section, we set the hopping rate to the right $p = 1$ without loss of generality (Fig. 10.4).

ASEP expressed in terms of indicator

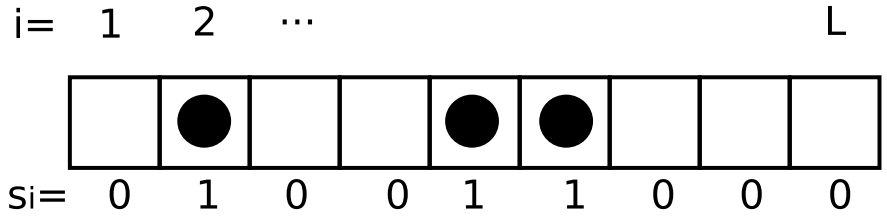


Figure 10.5: Definition of the indicator.

To describe the system in a more understandable way, we define an indicator s_i, which satisfy

$$s_i = \begin{cases} 1 & \text{when the site } i \text{ is occupied by a particle.} \\ 0 & \text{when the site } i \text{ is empty.} \end{cases} \tag{10.2}$$

Then the state of the system is defined by the sequence $\{s_i\}$, as shown in Fig. 10.5.

Equation for the average site occupancy Considering Fig. 10.4, we have for infinitesimally small dt

- If the site i is empty and the site $i-1$ is occupied, the particle at the site $i-1$ moves from the site $i-1$ to site i with the probability dt.

Further, we note that $s_{i-1}(1-s_i)$ is 1 if the site $i-1$ is occupied and the size i is empty, and otherwise, it is zero.

Therefore, for average occupancy $\langle s_i \rangle$ obeys:

$$\frac{d\langle s_i \rangle}{dt} = \langle s_{i-1}(1-s_i) \rangle - \langle s_i(1-s_{i+1}) \rangle. \tag{10.3}$$

Correlation between sites (10.3) can be re-written as

$$\frac{d\langle s_i \rangle}{dt} = \langle s_{i-1} \rangle - \langle s_{i-1}s_i \rangle - \langle s_i \rangle + \langle s_i s_{i+1} \rangle. \tag{10.4}$$

Namely, equation for s_i involves the two-site correlations, $\langle s_{i-1}s_i \rangle$ etc.

This actually continues: two-site correlation depends on the three-site correlation, etc. etc. Therefore, this equation is not closed and cannot be solved without approximations. The problem is essentially an N-body problem.

Mean-field approximation

To make (10.3) solvable, let us make a simple approximation that the deviation of the occupancy from the mean is small, i.e.,

$$s_i = \langle s_i \rangle + \tilde{s}_i$$

with $|\tilde{s}_i| \ll \langle s_i \rangle$. Note that $\langle \tilde{s}_i \rangle = 0$ so that the mean is recovered. Then, we have

$$\langle s_i s_j \rangle = \langle (\langle s_i \rangle + \tilde{s}_i)\,(\langle s_j \rangle + \tilde{s}_j) \rangle = \langle s_i \rangle \langle s_j \rangle + \langle \tilde{s}_i \tilde{s}_j \rangle \text{ etc.}$$

Since $\langle \tilde{s}_i \tilde{s}_j \rangle$ is the second order in the small deviation from the mean, we ignore this term - practically, we ignore the correlation, i.e., we assume

$$\langle s_i s_j \rangle \approx \langle s_i \rangle \langle s_j \rangle.$$

This is a mean-field approximation. With defining the average "density" at the site i as $n_i \equiv \langle s_i \rangle$, mean-field approximation of eq. (10.3) becomes

$$\dot{n}_i = n_{i-1}(1-n_i) - n_i(1-n_{i+1}). \tag{10.5}$$

This equation gives us various insights into the system. For example, by performing a further approximation of continuous space, we can get the *inviscid Burgers equation*, which is a famous partial differential equation known to give a shock wave solution. The relation can be further extended to the KPZ (Karder-Parisi-Zhang) equation. The KPZ equation is a famous equation to describe a wide class of interface growth phenomena. For interested readers, the derivation is given in appendix G.

Behavior under periodic boundary conditions

Let's consider the situation where the system is a circuit of length L, i.e., the site $L+1$ is actually site 1. Such a boundary condition is called the periodic boundary condition. In this case, we have

$$\begin{aligned} \dot{n}_i &= n_{i-1}(1-n_i) - n_i(1-n_{i+1}) \quad \text{for} \quad 2 \le i \le L-1, \\ \dot{n}_1 &= n_L(1-n_1) - n_1(1-n_2), \\ \dot{n}_L &= n_{L-1}(1-n_L) - n_L(1-n_1). \end{aligned}$$

It is easy to see that $n_i = \rho$ with a constant ρ for any i is a steady state solution, i.e. solution of $\dot{n}_i = 0$. ρ can take any value between 0 and 1. The flux (how often particles cross a fixed site per unit time) is in this case given by

$$\begin{aligned} J &= (\text{rate to hop}) \times (\text{prob. being at site } i) \times (\text{prob. for the site } i+1 \text{ empty}) \\ &= \rho(1-\rho). \end{aligned} \tag{10.6}$$

Since the rate to hop, p, is one. If p is not one, we will have

$$J = p\rho(1-\rho). \tag{10.7}$$

This solution will satisfy the master equation (10.1) under the periodic BC because the probability to be a particular configuration is constant. Namely, the constant density solution is actually an exact solution.

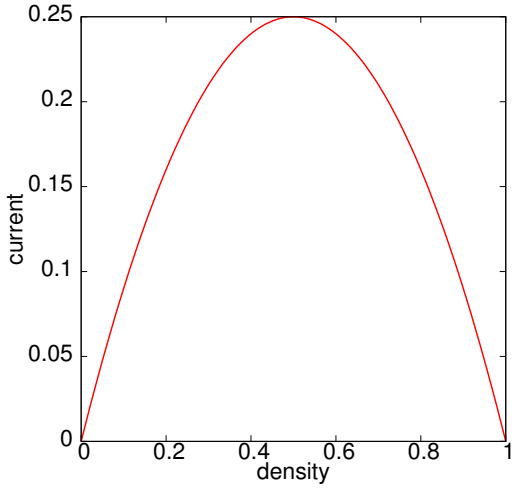


Figure 10.6: $J = p\rho(1-\rho)$ with $p = 1$.

The flux (10.7) is plotted as a function of ρ in Fig. 10.6. We see that the flux increase in $0 < \rho < 1/2$ and decreases in $1/2 < \rho < 1$, but it does not have a sharp transition as observed in the real traffic data.

Behavior in infinite system

If the system is infinite, we will only have

$$\dot{n}_i = n_{i-1}(1-n_i) - n_i(1-n_{i+1}). \tag{10.8}$$

The system again has the constant density $n_i = \rho$ solution as a steady state, as expected from taking $L \to \infty$ limit in the periodic boundary condition.

When the system is infinite, there is another possibility for the steady state, called shock. Consider the situation where $n_i = 0$ for $i < k$ with an integer k, while $n_i = 1$ for $i \geq k$. This situation is clearly a steady-state solution satisfying the mean field equation, noting that around k we have

$$\begin{aligned} \dot{n}_{k-1} &= n_{k-2}(1-n_{k-1}) - n_{k-1}(1-n_k), \\ \dot{n}_k &= n_{k-1}(1-n_k) - n_k(1-n_{k+1}), \\ \dot{n}_{k+1} &= n_k(1-n_{k+1}) - n_{k+1}(1-n_{k+2}). \end{aligned}$$

Physically this means that all the sites to the left of k are empty, and all the sites to the right of k are occupied. In this configuration, no particle can move, thus the current is $J = 0$. Clearly, this is also an exact solution.

This shock situation resembles an extreme traffic jam where no one can move. This is not possible under the periodic boundary condition, namely the boundary condition is strongly affecting the system's behaviour. This motivates us to study TASEP under the open boundary condition.

10.4 TASEP under the open boundary condition

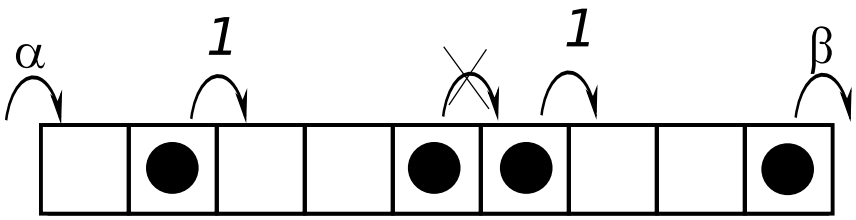


Figure 10.7: TASEP under the open boundary condition.

How about under the open boundary condition, as shown in Fig.10.7? Here, α is the rate for a new particle to come in, and β is the rate for a particle at the site L to go out from the system. What is the steady state for a given α and β?

10.4.1 How to simulate the Open BC with random sequential update

As you can see in Fig. 10.7, new particles need to be added once in a while. To deal with this, we consider additional "site" at $x = 0$ as follows.

1. Out of L sites ($x = 1$ to L) plus the additional site $x = 0$, choose one site randomly (Namely there are $L + 1$ sites to be chosen). Suppose the chosen site is j.

2. If $1 \leq j \leq L - 1$: if there is a particle at j and there is no particle at $j + 1$, then make it hop with the probability Δt. If $j = L$, and there is a particle at j, make it hop according to the probability per time step $\beta\Delta t$. If $j = 0$, and there is no particle at the site 1, add a particle to the site 1 with the probability $\alpha\Delta t$. Otherwise, do nothing.

3. Repeat 1 and 2 for $L + 1$ times.

4. Proceed time by Δt.

10.4.2 Phase diagram

The TASEP under OBC is known to show qualitatively different phases depending on α and β, and the boundary between phases is sharp, as shown in Fig.10.8. The qualitative description of the phases is as follows.

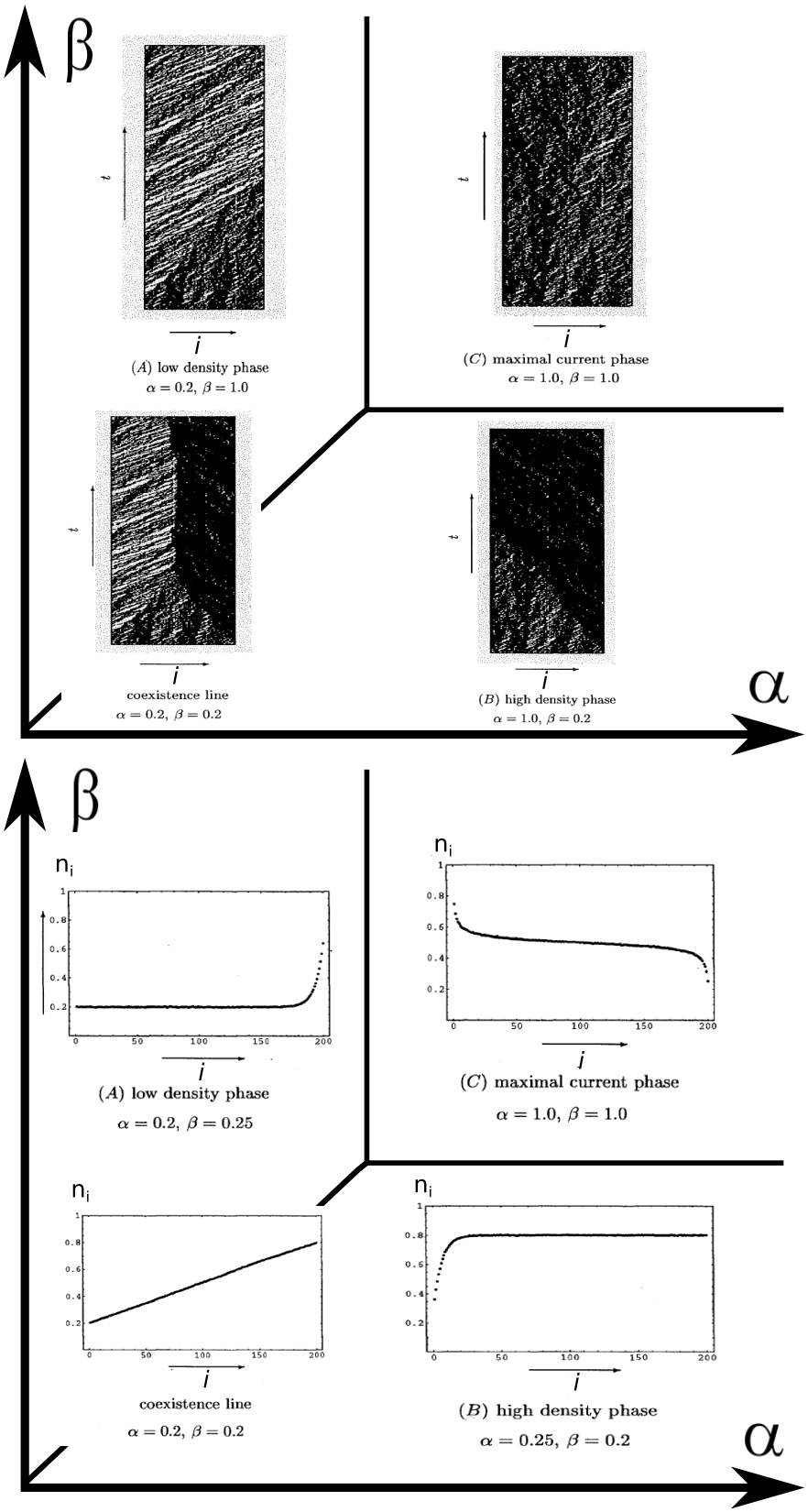


Figure 10.8: Top: Spatio-temporal plot Bottom: Long-time averaged density $n_i = \langle s_i \rangle$ in the steady state as a function of position.

- Low-density phase. The particle input rate α is rather low compared to bulk transport capacity and the exit rate β. Thus input rate α limit the current.
- High-density phase. The particle exit rate β is rather low compared to bulk transport capacity and the entrance rate α. Thus, the output rate β limit the current.
- Maximum current phase. Both α and β are rather high, and the bulk transport capacity (maximum in figure 10.6) limits the current.

We estimate the phase boundary by using mean-field approximation.

10.4.3 Mean field approximation

We again make the mean-field approximation for the average "density" $n_i \equiv \langle s_i \rangle$.

$$\dot{n}_i = n_{i-1}(1-n_i) - n_i(1-n_{i+1}), \tag{10.9}$$
$$\dot{n}_1 = \alpha(1-n_1) - n_1(1-n_2), \tag{10.10}$$
$$\dot{n}_L = n_{L-1}(1-n_L) - \beta n_L. \tag{10.11}$$

The first term of eq. (10.10) denotes that a new particle can enter the system with the rate α if the first site is empty, and the second term of eq. (10.11) denote that if there is a particle at site L it will go out with the rate β. We now try to solve this for a steady-state of very large system (L is large), to see how the flux (constant all over the system in the steady state) and the density in the middle of the system behaves.

low density phase

In this case, we assume that the equation for the entrance eq.(10.10) is dominating (i.e., entrance is the rate-limiting step, therefore it determines the overall flux), thus the density does not change so rapidly around the entrance, i.e. $n_1 = n_2 = n$. Under this condition in the steady state ($\dot{n} = 0$), we get

$$n = \alpha, \tag{10.12}$$
$$J = J_\alpha = \alpha(1-n) = \alpha(1-\alpha). \tag{10.13}$$

high density phase

In this case, we assume that the equation for the exit eq.(10.11) is dominating, thus the density does not change so rapidly around the exit, i.e. $n_{L-1} = n_L = n$. Under this condition in the steady state, we get

$$n = 1-\beta, \tag{10.14}$$
$$J = J_\beta = \beta n = \beta(1-\beta). \tag{10.15}$$

maximum current phase

In this case, we assume that the equation for the bulk eq.(10.9) is dominating, thus the density is almost constant in the bulk, $n_i = n$. Under this condition in the steady state, we get $J = n(1-n)$. The maximum current can be obtained when $n = n_{max} = 1/2$ (See figure 10.6), thus

$$J_{max} = n_{max}(1-n_{max}) = \frac{1}{4}. \tag{10.16}$$

Phase boundary

The boundary between the low density phase and the high-density phase is apparently on the line of $\alpha = \beta$, where two solutions (10.13) and (10.15) become identical.

Since the maximum possible current in the bulk is given by J_{max}, the phase boundary between the maximum current phase and the low (high) density phase should be where $J_\alpha = J_{max}$ ($J_\beta = J_{max}$). This happens when $\alpha = 1/2$ ($\beta = 1/2$).

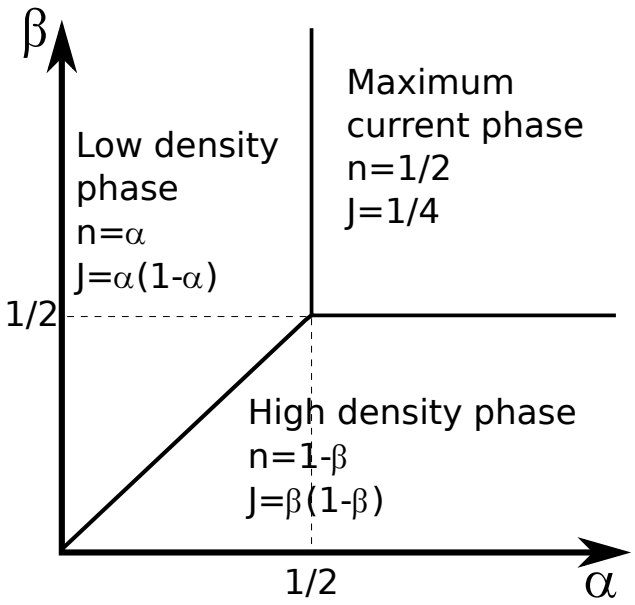


Figure 10.9: Summary of phase diagram obtained by the mean field approximation.

Summary of phase diagram

The obtained phase diagram is summarized in Fig. 10.9. This means that *the boundary condition changes the bulk behaviour completely*. This is called the *boundary induced phase transition*, which is unique to the far-from-equilibrium systems. The order parameter (the parameter that characterizes the phase transition, normalized to take a value between 0 and 1 and show sharp change at the phase boundary) is the density n.

10.5 Discrete time: Parallel update and Nagel-Schreckenberg (NaSch) model

Next, let us think about the situation where the time step is actually discrete (B), representing the time for a car needs to move by its length. Following conventions, we fix the time step $\Delta t = 1$, and instead, reintroduce the probability p to hop to the right per time step (Fig. 10.10). In this case, the update rule should be different from the continuous-time case, because many can move simultaneously; in other words, the update of the position of the particles can happen in parallel. Therefore, here we introduce a *parallel update*.

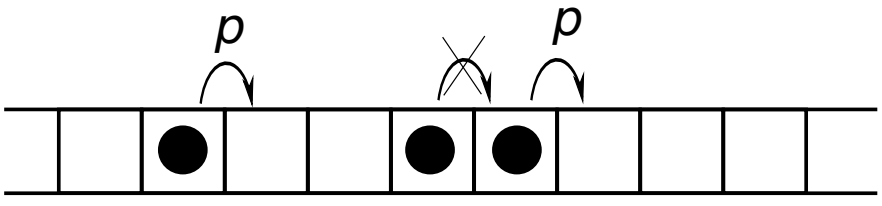


Figure 10.10: TASEP with a parallel update. We fix $\Delta t = 1$, but make p variable.

10.5.1 (B) Parallel update

To make the later extension of the model easier, and also to make it clear why the random sequential update and the parallel update are different, let us distinguish particles that moved and not moved for each time

step. For that purpose, we introduce the concept of "velocity" of each particle and define the dynamics as follows.

Each particle has its id i, velocity v_i and is located at position x_i. The particle id is assigned so that $x_1 < x_2 \cdots < x_N$.

1. *Acceleration.* If $v_i = 0$, the velocity of the i-th particle is increased by one, i.e.
$$v_i \to \min(v_i + 1, 1). \tag{10.17}$$

2. *Deceleration due to other particles.* If $x_{i+1} - x_i \leq v_i$, the velocity of the i-th particle is decreased to $x_{i+1} - x_i - 1$, i.e.,
$$v_i \to \min(v_i, x_{i+1} - x_i - 1). \tag{10.18}$$

3. *Randomization.* If $v_i > 0$, the velocity of the i-th particle is decreased randomly by one unit with probability $(1 - p)$, i.e.,
$$v_i \to \max(v_i - 1, 0) \quad \text{with probability} \quad 1 - p. \tag{10.19}$$

4. *Particle movement.* After doing 1-3 for all the particles, each particle is moved forward according to its new velocity, i.e.,
$$x_i \to x_i + v_i. \tag{10.20}$$

Note that the rule means that we check if a car can move before anyone moves in rules 1 and 2, and then make the move stochastic by stopping the move with probability $1 - p$ in rule 3; this is equivalent to move with a probability in p, but the judgement and the move are done in parallel. Therefore, this is different from the random sequential update.

Configuration of the system with velocity

Figure 10.11: TASEP with velocity.

Notation in Fig. 10.11 shows the configuration of cars just after the movement, showing their velocity v_i at the corner. (This is a snapshot just after step 4 of the update.)

"Garden of Eden" states

See the configuration in Fig. 10.12: Is this configuration possible?

By introducing the notion of velocity, it is easy to see that the state that (0,1) or (1,1) cannot be reached by parallel update unless it is given as the initial configuration. The first means two particles are overlapping in the previous time step, and the second means the two-particle that is next to each other moved simultaneously in the previous time step.

Such a state – a state that cannot appear after one time step no matter what the initial configuration– can be often found in cellar automaton, and it is called Garden of Eden (GoE) states or prohibited states. The existence of GoE is the reason why it is different from the continuous-time system. In the continuous-time case, (1,1) is possible. Since this prohibited state is "congested" in the sense that the second particle is blocked, it is easy to imagine that the absence of these states results in higher flux for the discrete-time case than that in the continuous-time case. Indeed, we observe higher flux in the parallel updates when we simulate it (exercise).

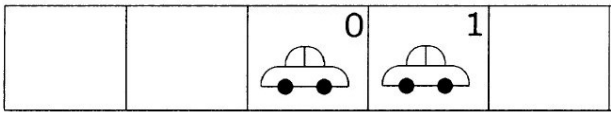


Figure 10.12: Hypothetical configuration for TASEP.

10.5.2 NaSch model

By allowing larger velocity, we can define a more realistic car-traffic model, **Nagel-Schereckenberg (NaSch) model**.

1. *Acceleration.* If $v_i < v_{max}$, the velocity of the i-th particle is increased by one, i.e.

$$v_i \to \min(v_i + 1, v_{max}). \tag{10.21}$$

2. *Deceleration due to other particles.* If $x_{i+1} - x_i \leq v_i$, the velocity of the i-th particle is decreased to $x_{i+1} - x_i - 1$, i.e.,

$$v_i \to \min(v_i, x_{i+1} - x_i - 1). \tag{10.22}$$

3. *Randomization.* If $v_i > 0$, the velocity of the i-th particle is decreased randomly by one unit with probability $(1 - p)$, i.e.,

$$v_i \to \max(v_i - 1, 0) \quad \text{with probability} \quad 1 - p. \tag{10.23}$$

4. *Particle movement.* After doing 1-3 for all the particles, Each particle is moved forward according to its new velocity, i.e.,

$$x_i \to x_i + v_i. \tag{10.24}$$

NaSch model is an extension of the model by adding maximum velocity, v_{max}, as a new parameter. TASEP with the parallel update is the case with $v_{max} = 1$. For traffic jam formation, It is important the randomization comes after the deceleration due to other particles. This introduces an "overreaction" which mimics people breaking more than necessary, and it can be a seed for a traffic jam when the density is high.

This model is known to show traffic-jam-like structures, as demonstrated in Fig. 10.13.

Deterministic case of NaSch model

When we take $p = 1$, then particles will always get to the maximum velocity v_{max}, unless the particles are blocked by other particles. Otherwise, the maximum velocity is equal to (maximum possible distance between particles -1). For an infinite system, this becomes $1/\rho - 1$, thus current is given by $\rho(1/\rho - 1) = 1 - \rho$. In summary, the current is given by

$$J = \min(\rho v_{max}, 1 - \rho). \tag{10.25}$$

Thus there is a transition to a congested state by changing ρ: the flux suddenly changes at $\rho = \rho_{crit}$ as shown in Fig. 10.14, which is given by

$$\rho_{crit} v_{max} = 1 - \rho_{crit} \quad \Leftrightarrow \quad \rho_{crit} = 1/(v_{max} + 1). \tag{10.26}$$

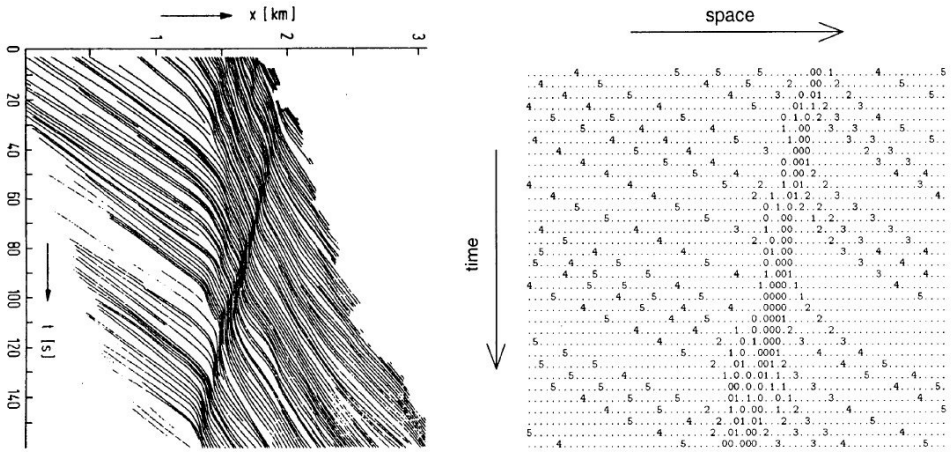


Fig. 3. Trajectories of single cars showing spontaneous jam formation. Left: Empirical data; Right: Computer simulation using the NaSch model. Each number $0, 1, \ldots, v_{\max} = 5$ gives the velocity of the corresponding car.

Figure 10.13: Comparison of real data and NaSch model.

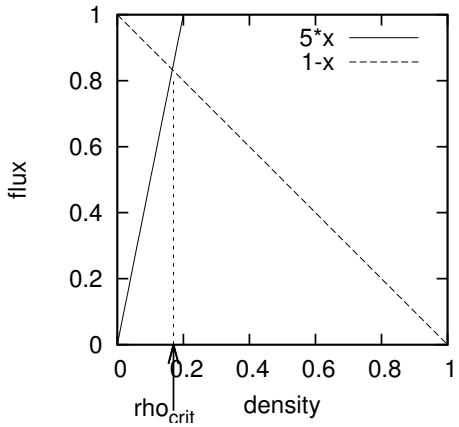


Figure 10.14: Deterministic estimate of the flux.

How a jam is spontaneously formed and propagates in NaSch model

With $p < 1$, the transition in the deterministic model will be smeared out, but this term is important for the spontaneous formation of jams.

- With the probability $1 - p$, the driver brakes more than actually necessary to avoid the collision.
- When the density is rather high, this will cause the drivers behind him also to slow down, eventually leading to traffic jams.

10.6 Summary

- TASEP is a simple stochastic model of traffic flow.
- TASEP with continuous time can be analyzed by mean field approximation;

$$\dot{n}_i = n_{i-1}(1 - n_i) - n_i(1 - n_{i+1}).$$

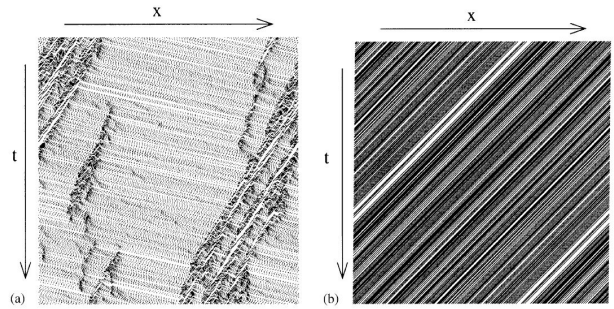


Figure 10.15: Spatiotemporal plot obtained by NaSch model with $v_{max} = 5$. Left: $p = 0.75, \rho = 0.2$. Right: $p = 1, \rho = 0.5$.

- Under the open boundary condition, TASEP shows boundary-induced phase transition, which can be quantitatively analyzed by the mean-field equations.
- NaSch model is an extension of TASEP with a parallel update. Spontaneous traffic jam formation can be seen in NaSch model. Random braking plays an important role in forming a traffic jam.

10.7 Exercise

1. Consider the original ASEP where neither p nor q is zero in continuous time. Write down the differential equation for the average density $n_i = \langle s_i \rangle$ by applying the mean-field approximation.
2. In Fig. 10.8 bottom, long time average density is plotted as a function of the position for various phases in the steady-state of TASEP under the open boundary condition.
   (a) For the low-density phase (A) with $\alpha = 0.2$ and $\beta = 0.25$, we see that the density rises around the exit. Using the mean-field result, evaluate the density at the exit n_L in the steady-state, and confirm that it agrees with the figure. (Hint: The flux should be constant all over the system at the steady state.)
   (b) For the high-density phase (B) with $\alpha = 0.25$ and $\beta = 0.2$, we see that the density rises just after the entrance. Using the mean-field result, evaluate the density at the entrance n_1 in the steady-state, and confirm that it agrees with the figure.
   (c) For the maximum current phase (C) with $\alpha = 1.0$ and $\beta = 1.0$, we see higher density at the entrance and lower density at the exit. Using the mean-field result, evaluate the density at the entrance n_1 and at the exit n_L in the steady state, and confirm that it agrees with the figure.
3. * Simulate continuous time TASEP.
   (a) Simulate the TASEP with (A) random sequential update with $L = 100$ with periodic boundary conditions. Set $\Delta t = 1$ and $p = 0.5$.
      i. First, make the spatiotemporal plot (the horizontal axis is position, the vertical axis is time, and the location of the particles are shown in this "spatiotemporal" space) and see how the system behaves.
      w
      ii. Plot the steady state current J as a function of the density ρ (Try ρ=0.1 to 0.9 with step 0.1). Confirm that it matches with the theoretical value (10.7).
   (b) Perform TASEP simulation in continuous time under the open boundary condition. Set $\Delta t = 1$ and $p = 1$ in the simulation. Try parameter sets that lead to low-density phase, high-density phase, and maximum current phase, and see the Spatio-temporal plot for each phase. Measure the current J for $\alpha = 0.2$ with changing β from 0 to 1. Confirm the agreement with the mean-field estimation. Do the same for $\alpha = 0.7$.

4. * Simulate NaSch model.

   (a) Simulate NaSch model with $v_{max} = 1$ with $L = 100$. Set $\Delta t = 1$ and $p = 0.5$. Plot the steady state current J as a function of the density ρ (Try ρ=0.1 to 0.9 with step 0.1). Confirm that the flux is higher than the case with the continuous-time TASEP (cf. exercise 3-a-ii).

   (b) Simulate NaSch model with $v_{max} = 5$ under the periodic boundary condition with $L = 100$. Plot the spatiotemporal plot for

   - $p = 0.75$; $\rho = 0.2, 0.5, 0.8$.
   - $p = 0.25$; $\rho = 0.2, 0.5, 0.8$.

   How does the behaviour depend on the parameters?

10.8 Further reading

This chapter is based on [22, 23].

Chapter 11

Predator-prey Dynamics

We have been mostly working on one variable systems, except for a few examples in the exercises. Here we take an example topic that (almost) everything we have learned so far can be used to analyze the system, but for a two-variable system. The chosen example is the *predator-prey dynamics*, where we consider a population of agents - for example, hare that grows and is eaten by another agent, for example, lynx, that needs to eat hares to grow, while they can die. Interestingly, it has been found that the population of such a system sometimes shows an oscillation, as observed by MacLulich in 1937 (Fig. 11.1). What is the mechanism of the oscillation of the number of preys $N(t)$ and predators $M(t)$? There are many possibilities - and here, we learn one of the possible mechanisms, where the stochasticity of the system plays an important role. The content of this chapter is based on ref. [24] with small modifications.

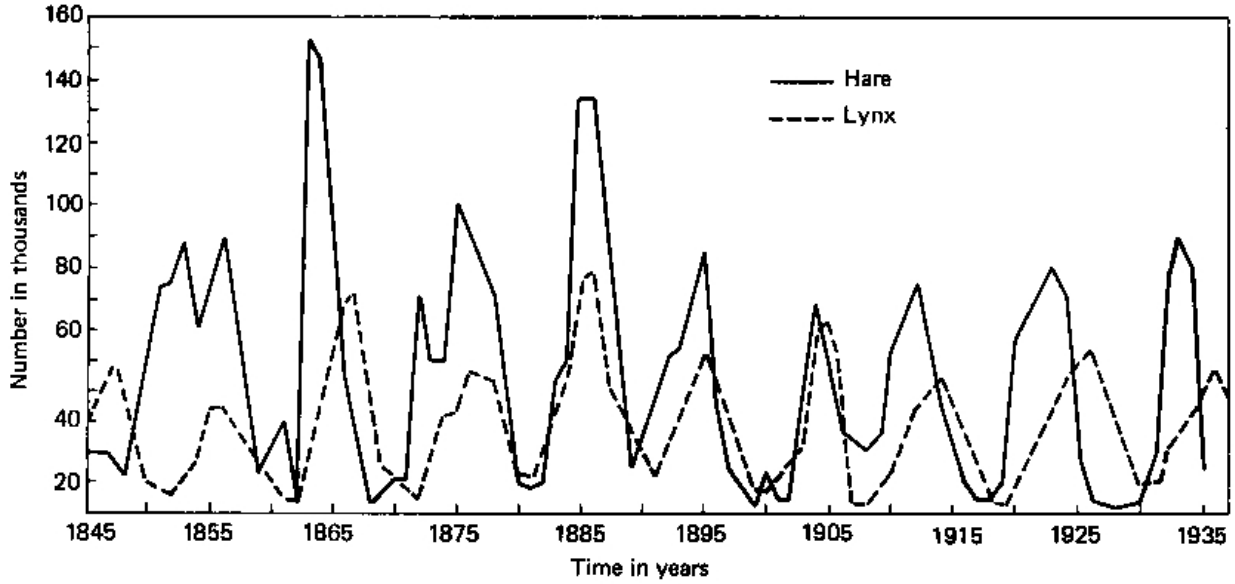


Figure 9-3. Changes in the abundance of the lynx and the snowshoe hare, as indicated by the number of pelts received by the Hudson's Bay Company. This is a classic case of cyclic oscillation in population density. (Redrawn from MacLulich 1937.)

Figure 11.1: Predator-prey oscillation.

11.1 Agent-based rule of predator-prey interaction

Let us consider $N(t)$ preys and $M(t)$ predators in the system. For later convenience, let me express preys as A and Predator as B, and define each "reaction" of growth and death.

Prey growth For simplicity, we assume that one prey can "double" at the rate α per prey. If that is the only rule, it means $A \to 2A$ at a constant rate, but then A will keep increasing exponentially if there is no B to kill it. So we also assume that there is a maximum integer number K of the preys that the world can support - this can be due to the limit of the available food for them. We assume that this limit kicks in as a form of *logistic growth*, i.e., the rate of $A \to 2A$ per prey is given as $\alpha \cdot (1 - N(t)/K)$. In other words,

$$N \to N + 1 \quad \text{at the rate} \quad \alpha \cdot N(1 - N/K).$$

Note that, for this model, we limit the range of N being $0 \le N \le K$, which makes the rate defined above always non-negative. Because N can increase only one at a time and K is an integer, when N hits exactly K the growth rate becomes zero, and hence N cannot exceed K as long as the initial condition satisfies $N \le K$.

Prey death by predation We use the simplest form here, too, where we assume that when the two meet, there is a probability that a prey is lost, i.e., $A + B \to B$. The probability of one prey meeting a predator depends on the predator's density. Here, let us use the carrying capacity K as the "volume" of the system since it is natural that K is proportional to the total system size. Namely, we consider N/K as the prey density and M/K as the predator density. We then set the rule for the prey death by predation using a constant β as

$$N \to N - 1 \quad \text{at the rate} \quad \beta \cdot N \cdot \frac{M}{K}.$$

Predator growth by eating a prey Predator can grow only by eating preys. If it is lynx, it probably needs to eat a certain number of the preys before it "doubles". Keeping track of predation event however complicates the problem, so, for now, we make a very simple assumption: Once in a while, upon a predation event, a prey dies and the predator doubles, i.e., $A + B \to 2B$, with some rate per predation event. Then, by using a new constant γ, we have

$$M \to M + 1 \quad \text{and} \quad N \to N - 1 \quad \text{at the rate} \quad \gamma \cdot M \cdot \frac{N}{K}.$$

Predator death Predator dies, and here again, for simplicity, we assume it happens at a constant rate of δ per predator. This gives

$$M \to M - 1 \quad \text{at the rate} \quad \delta \cdot M.$$

These rules define the dynamics. It is left for exercise to perform a Gillespie simulation of this model. Does the model show interesting behaviour?

11.2 Rate equations: Deterministic Lotoka-Volterra model with logistic growth for prey

To have some insight into the system's dynamics, let us think about the rate equation, i.e., the deterministic equation to describe the behaviour of the density of prey $x = N/K$ and predator $y = M/K$ in the large system size limit ($K \to \infty$) with keeping the density constant, hence the noise is negligible. For example, for the prey, we naturally have (and can be derived by using the master equation (11.7) if you want to make sure)

$$\frac{d\langle N \rangle}{dt} = \left\langle \alpha N \left(1 - \frac{N}{K}\right) \right\rangle - \left\langle (\beta + \gamma) \cdot N \cdot \frac{M}{K} \right\rangle,$$

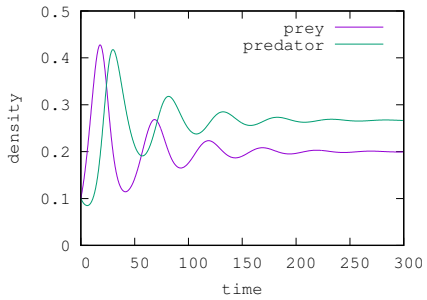


Figure 11.2: An example trajectory of the deterministic rate equation for the prey population density x and predator population density y. The parameters are $\alpha = 0.2$, $\beta = 0.1$, $\gamma = 0.5$, $\delta = 0.1$.

where the average change of N is given by the average rate of increase minus the average rate of decrease. By dividing both sides by K to convert the unit into density, and assuming that in the large system size limit, there is no noise, hence we can approximate $\langle (N/K)(M/K) \rangle$ as $\langle N/K \rangle \langle M/K \rangle = xy$ (mean-field approximation), and doing similar treatment also for $d\langle M \rangle / dt$, we get a well-studied simple deterministic model as follows:

$$\begin{aligned} \frac{dx}{dt} &= \alpha x (1-x) - (\beta + \gamma) xy = x \left[\alpha (1-x) - (\beta + \gamma) y \right], \\ \frac{dy}{dt} &= \gamma xy - \delta y = y \left[\gamma x - \delta \right]. \end{aligned} \tag{11.1}$$

The functional form of predator growth is called logistic growth, and the functional form of the predator-prey interaction is called the Lotoka-Volterra interaction.

This model has multiple fixed points (steady states). The trivial one is

$$x = 0, \quad y = 0, \tag{11.2}$$

i.e., both the predator and the prey went extinct. Let's call this the "full extinction fixed point". Another relatively easy case, which we now call the "prey-only fixed point" is

$$x = 1, \quad y = 0, \tag{11.3}$$

i.e., there is no predator, and the hare population is at the carrying capacity of the system. The last fixed point, which is probably the most interesting and we now call the "coexistence fixed point", is

$$x = \frac{\delta}{\gamma}, \quad y = \frac{\alpha}{\beta + \gamma} \left(1 - \frac{\delta}{\gamma} \right). \tag{11.4}$$

This coexistence solution is feasible (i.e., both densities are positive values) only if $\delta/\gamma < 1$, i.e., the prey population needed to sustain the predator is below the carrying capacity. Let's assume that this is true, so the coexistence fixed point is feasible.

Now, if we start with the system with non-zero predators and preys, where does the system go? If you know *linear stability analysis* of a dynamical system, you can perform it around the fixed points to see if they are stable. Here, we don't have time to cover the linear stability analysis, so we just look at the numerical simulation of the system at a specific parameter set in Fig. 11.2. The parameters are given in the figure caption. We see that the system shows dumped oscillation to converge to the coexistence fixed point. If we perform the linear stability analysis of the feasible coexistence fixed point, it is possible to show that the coexistence fixed point is always linearly stable. This means that the rate equation never shows sustained oscillation.

11.3 Back to a stochastic system: Quasi-oscillation

11.3.1 An absorbing state

Before we go into the oscillation, let us quickly touch that the stochastic model with discrete predator and prey numbers can deal with the extinction of a species. Suppose we start a stochastic Gillespie simulation of the model, with the initial condition very close to the full extinction fixed point. For example, having one predator and one prey in the system. If you simulate the rate equation with the initial condition being very low predator and prey concentration, with the parameter presented in Fig. 11.2 you will actually find that deterministically the system will go to the coexistence fixed point. This is intuitive because, if the prey and the predator density is very low, the predator-prey interaction term (product of the two densities) is a lot smaller than the prey's exponential growth - hence the prey grows first, and eventually predator also starts to grow and they find a balance.

But think about what could actually happen. The events that can happen after the given initial condition are as follows:

- The prey increases from 1 to 2 at the rate $\alpha(1-1/K)$
- The prey decreases from 1 to 0 by predation at the rate β/K
- The prey decreases from 1 to 0 and the predator increases from 1 to 2 at the rate γ/K
- The predator decreases from 1 to 0 at the rate δ

So, there is a finite probability that the prey disappear to 0 in the next event. Then, the only possible next event is for the predator to disappear. In other words, there is a finite probability to hit exactly $(0,0)$, and once the system is there, there is no way to come out of it.

A state that, once entered, cannot be left, is called an absorbing state. The full extinction state in the stochastic version of the model is an absorbing state. Even though deterministically the fixed point is unstable, in a stochastic system there is a finite probability to hit this absorbing state. Understanding how often such an event happens can be important when we are considering a small population.

In general, the rate to reach an absorbing state is often an important quantity to know for stochastic processes. The methods that have been developed to analyse the first passage problem are useful to tackle such problems.

11.3.2 Quasi-oscillation

Now, let us talk about the oscillation we have seen in the experimental data. Interestingly, if we simulate the stochastic version of the model with a finite K, we observe "sustained oscillation" even if we choose the parameter that the rate equation only shows a dumped oscillation. It is fun to see this by actually programming it by yourself (exercise), but you can also see an example trajectory in Fig. 11.3 left. This behaviour could be understood better by applying the tools we have learned in this couse - so let's have a look at it.

Master equation

The first step for the analytic calculation is to write a master equation for probability to have n prey and m predators at time t, $P(n,m,t)$. You can see this is a one-step process, where the value of the variables changes only one by one. For later convenience, let us define two kinds of step operators as

$$E_n^{\Delta} f(n,m) = f(n+\Delta,m), \tag{11.5}$$
$$E_m^{\Delta} f(n,m) = f(n,m+\Delta). \tag{11.6}$$

Then the master equation becomes

$$\begin{aligned}\dot{P}(n,m,t) &= \left(E_n^{-1}-1\right)\alpha n\left(1-\frac{n}{K}\right)P(n,m,t) + \left(E_n^{+1}-1\right)\beta\frac{m}{K}nP(n,m,t) \\ &\quad + \left(E_m^{-1}E_n^{+1}-1\right)\gamma\frac{n}{K}mP(n,m,t) + \left(E_m^{+1}-1\right)\delta mP(n,m,t).\end{aligned} \tag{11.7}$$

Note that "the predator growth by eating a prey" term has the product of two step-operators, reflecting one reaction causes a change of both N and M.

Kramers-Moyal expansion, SDE, and behaviour around the fixed point

The master equation itself is still somewhat difficult to deal with, so let's perform the Kramers-Moyal expansion. Because we have two variables, n and m, we need to expand both of them, but the expansion of the step operator is in parallel as before (see eq. (8.25)). We then get

$$\begin{aligned}\frac{\partial}{\partial t}P(n,m,t) &= \left(\sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\frac{\partial^k}{\partial n^k}\right)\left[\alpha n\left(1-\frac{n}{K}\right)P(n,m,t)\right]+\left(\sum_{k=1}^{\infty}\frac{(+1)^k}{k!}\frac{\partial^k}{\partial n^k}\right)\left[\beta\frac{m}{K}nP(n,m,t)\right]\\ &+\left[\left(1+\sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\frac{\partial^k}{\partial m^k}\right)\left(1+\sum_{j=1}^{\infty}\frac{(+1)^j}{j!}\frac{\partial^j}{\partial n^j}\right)-1\right]\left[\gamma\frac{n}{K}mP(n,m,t)\right]\\ &+\left(\sum_{k=1}^{\infty}\frac{(+1)^k}{k!}\frac{\partial^k}{\partial m^k}\right)\left[\delta mP(n,m,t)\right]. \end{aligned} \tag{11.8}$$

Note the cross-terms in the γ term. By truncating the expansion up to the second order, we obtain

$$\begin{aligned}\frac{\partial}{\partial t}P(n,m,t) &= -\frac{\partial}{\partial n}\left[\left(\alpha n\left(1-\frac{n}{K}\right)-(\beta+\gamma)\frac{m}{K}n\right)P(n,m,t)\right]+\frac{1}{2}\frac{\partial^2}{\partial n^2}\left[\left(\alpha n\left(1-\frac{n}{K}\right)+(\beta+\gamma)\frac{m}{K}n\right)P(n,m,t)\right]\\ &-\frac{\partial}{\partial m}\left[\left(\gamma\frac{n}{K}m-\delta m\right)P(n,m,t)\right]+\frac{1}{2}\frac{\partial^2}{\partial m^2}\left[\left(\gamma\frac{n}{K}m+\delta m\right)P(n,m,t)\right]\\ &-\frac{\partial^2}{\partial m\partial n}\left[\gamma\frac{n}{K}mP(n,m,t)\right]. \end{aligned} \tag{11.9}$$

This is somewhat better - at least the flux term (the first-order derivative term) does give the flux that is consistent with what is expected from the rate equation.

It may be even easier to understand the system behaviour by rewriting this into the stochastic differential equations (SDEs). It is straightforward to show that the following Ito-type SDEs give the Fokker-Planck equation given above (exercise):

$$dN = \left[\alpha N\left(1-\frac{N}{K}\right)-\frac{(\beta+\gamma)}{K}MN\right]dt+\sum_{k=1}^{4}b_{1k}(N,M)\cdot dB_k, \tag{11.10}$$

$$dM = \left[\frac{\gamma}{K}NM-\delta M\right]dt+\sum_{k=1}^{4}b_{2k}(N,M)\cdot dB_k, \tag{11.11}$$

with B_k's are independent Wiener processes, and

$$b(N,M)=\begin{pmatrix}\sqrt{\alpha N\left(1-\frac{N}{K}\right)} & -\sqrt{\beta N\frac{M}{K}} & -\sqrt{\gamma M\frac{N}{K}} & 0\\ 0 & 0 & \sqrt{\gamma M\frac{N}{K}} & \sqrt{\delta M}.\end{pmatrix} \tag{11.12}$$

You may wonder where the 4 noise terms and the matrix b come from. The 4 noise terms represent 4 reactions, and the matrix b_{ik} is the square root of each reaction rate times how much the component changes: $i=1$ ($i=2$) corresponds to the prey (predator) component, and if the k-th reaction changes N (M) by $+1$ or -1 is determining the sign.

Note that the deterministic part of eqs. (11.10) and (11.11) is equivalent to the rate equations (11.1) is we replace N/K with x and M/K with y. This set of SDEs (11.10) and (11.11) makes us view the dynamics as a deterministic equation plus noise.

We are interested in the quasi-oscillation around the coexistence fixed point. So now, let us consider N and M are actually close to the coexistence fixed point but fluctuating around it, and let us take the spirit

of the system size expansion by assuming

$$N(t) = K\frac{\delta}{\gamma} + \sqrt{K}X(t), \tag{11.13}$$

$$M(t) = K\frac{\alpha}{\beta+\gamma}\left(1-\frac{\delta}{\gamma}\right) + \sqrt{K}Y(t). \tag{11.14}$$

Putting this back to the SDEs and taking the most relevant order in K (which is the term proportional to $\sqrt{K}$, left for advanced exercise), we get

$$dX = \left[-\frac{\alpha\delta}{\gamma}X - \frac{(\beta+\gamma)\delta}{\gamma}Y\right]dt + \sum_{k=1}^{4}\tilde{b}_{1k}\cdot dB_k, \tag{11.15}$$

$$dY = \left[\frac{\alpha\gamma}{\beta+\gamma}\left(1-\frac{\delta}{\gamma}\right)X\right]dt + \sum_{k=1}^{4}\tilde{b}_{2k}\cdot dB_k, \tag{11.16}$$

with

$$\tilde{b} = \begin{pmatrix} \sqrt{\alpha\frac{\delta}{\gamma}\left(1-\frac{\delta}{\gamma}\right)} & -\sqrt{\beta\frac{\delta}{\gamma}\frac{\alpha}{\beta+\gamma}\left(1-\frac{\delta}{\gamma}\right)} & -\sqrt{\gamma\frac{\delta}{\gamma}\frac{\alpha}{\beta+\gamma}\left(1-\frac{\delta}{\gamma}\right)} & 0 \\ 0 & 0 & \sqrt{\gamma\frac{\delta}{\gamma}\frac{\alpha}{\beta+\gamma}\left(1-\frac{\delta}{\gamma}\right)} & \sqrt{\delta\frac{\alpha}{\beta+\gamma}\left(1-\frac{\delta}{\gamma}\right)}. \end{pmatrix} \tag{11.17}$$

Note that, if we write down the corresponding Fokker-Plank equation for the distribution of X and Y, it will be the same form as what you will get if you did the system size expansion and the linear noise approximation, and then put the coexistence fixed point for the deterministic density.

Note that the coefficients in the deterministic part of eqs. (11.15) and (11.16) are given by the Jacobian matrix J of the rate equation (11.1) at the coexistence fixed point

$$J = \begin{pmatrix} -\alpha\frac{\delta}{\gamma} & -(\beta+\gamma)\frac{\delta}{\gamma} \\ \frac{\gamma\alpha}{\beta+\gamma}\left(1-\frac{\delta}{\gamma}\right) & 0 \end{pmatrix}. \tag{11.18}$$

So, if there is no noise term, the equation becomes a linear equation

$$\frac{d\vec{X}}{dt} = J\vec{X},$$

where

$$\vec{X} \equiv \begin{pmatrix} X \\ Y \end{pmatrix}.$$

It is straightforward to see that the solution of such a linear equation can be expressed by using the eigenvalues λ_1, λ_2 and the corresponding right eigenvectors $\vec{u}_1$, $\vec{u}_2$ as

$$\vec{X} = \sum_{j=1}^{2} c_j e^{\lambda_j t}\vec{u}_j,$$

hence the eigenvalue of the Jacobian tells how the system behaves without the noise.

The eigenvalues of the Jacobian (11.18) is given by

$$\lambda_\pm = \frac{1}{2}\left(-\frac{\alpha\delta}{\gamma} \pm \sqrt{\left(\frac{\alpha\delta}{\gamma}\right)^2 - 4\alpha\delta\left(1-\frac{\delta}{\gamma}\right)}\right) = \frac{\alpha\delta}{2\gamma}\left(-1 \pm \sqrt{1-4\frac{\gamma^2}{\alpha\delta}\left(1-\frac{\delta}{\gamma}\right)}\right). \tag{11.19}$$

Because we consider the case $\left(1-\frac{\delta}{\gamma}\right) > 0$ (feasible coexistence point), the real parts of both of the eigenvalues are always negative. If

$$1-4\frac{\gamma^2}{\alpha\delta}\left(1-\frac{\delta}{\gamma}\right) < 0,$$

the eigenvalues are complex. If there is no noise, the trajectory will show a dumped oscillation and the frequency is given by the imaginary part of the eigenvalue, i.e., $\frac{\alpha\delta}{2\gamma}\sqrt{-1+4\frac{\gamma^2}{\alpha\delta}\left(1-\frac{\delta}{\gamma}\right)}/2\pi$.

This gives an intuitive explanation for quasi-cycle: Deterministically the system is decaying to the fixed point with dumped oscillation as the real part of the eigenvalue of the Jacobian is negative and the imaginary part is non-zero, but the noise keeps kicking the system out of the fixed point. Hence the system appears to follow a noisy oscillation.

We can say more quantitatively about the frequency distribution of this quasi-cycle. The SDEs (11.15) and (11.16) can be expressed as Langevin equations

$$\frac{dX}{dt} = J_{11}X + J_{12}Y + \sum_{k=1}^{4}\tilde{b}_{1k}R_k(t), \tag{11.20}$$

$$\frac{dY}{dt} = J_{21}X + J_{22}Y + \sum_{k=1}^{4}\tilde{b}_{2k}R_k(t), \tag{11.21}$$

with J defined as Jacobian (11.18) and

$$\langle R_i(t)\rangle = 0, \langle R_i(t)R_j(t')\rangle = 0 \quad \text{for} \quad i \neq j, \tag{11.22}$$

$$\langle R_i(t)R_i(t')\rangle = \delta(t-t'). \tag{11.23}$$

It is easy to calculate the power spectrum density $P_X(\omega)$ and $P_Y(\omega)$ from (11.20) and (11.21), which corresponds to the Fourier transform of the auto-correlation in the steady state fluctuation (Wiener-Khintin theorem). Hence, the peak of the power spectrum density will tell us the frequency distribution of the quasi-oscillation. Here, the power spectrum density is defined as before, by using

$$\tilde{X}(\omega) = \int_{-\infty}^{\infty} X(t)e^{i\omega t}dt, \quad \tilde{Y}(\omega) = \int_{-\infty}^{\infty} Y(t)e^{i\omega t}dt \tag{11.24}$$

as

$$2\pi\delta(\omega-\omega')P_X(\omega) = \langle\tilde{X}(\omega)\tilde{X}^*(\omega')\rangle, \quad 2\pi\delta(\omega-\omega')P_Y(\omega) = \langle\tilde{Y}(\omega)\tilde{Y}^*(\omega')\rangle. \tag{11.25}$$

Using the Fourier-transformed Langevin equations

$$-i\omega\tilde{X}(\omega) = J_{11}\tilde{X}(\omega) + J_{12}\tilde{Y}(\omega) + \sum_{k=1}^{4}\tilde{b}_{1k}\tilde{R}_k(\omega), \tag{11.26}$$

$$-i\omega\tilde{Y}(\omega) = J_{21}\tilde{X}(\omega) + J_{22}\tilde{Y}(\omega) + \sum_{k=1}^{4}\tilde{b}_{2k}\tilde{R}_k(\omega), \tag{11.27}$$

(here $\tilde{R}_i(\omega) = \int_{-\infty}^{\infty} R_i(t)e^{i\omega t}dt$), it is straightforward to show (advanced exercise)

$$P_X(\omega) = \frac{J_{12}^2\sum_{k=1}^4\tilde{b}_{1k}^2 - 2J_{12}J_{22}\sum_{k=1}^4\tilde{b}_{1k}\tilde{b}_{2k} + (J_{22}^2+\omega^2)\sum_{k=1}^4\tilde{b}_{2k}^2}{(J_{21}J_{12}-J_{11}J_{22}+\omega^2)^2 + (J_{11}+J_{22})^2\omega^2}, \tag{11.28}$$

$$P_Y(\omega) = \frac{J_{21}^2\sum_{k=1}^4\tilde{b}_{2k}^2 - 2J_{21}J_{11}\sum_{k=1}^4\tilde{b}_{2k}b_{1k} + (J_{11}^2+\omega^2)\sum_{k=1}^4\tilde{b}_{1k}^2}{(J_{21}J_{12}-J_{11}J_{22}+\omega^2)^2 + (J_{11}+J_{22})^2\omega^2}. \tag{11.29}$$

If you plot these power spectrum densities, it has a peak at a finite frequency, indicating that the fluctuation contains periodic oscillatory components.

In Fig. 11.3, the simulated trajectories and the power spectrum densities are shown. The theoretical prediction (11.28)-(11.29) agrees well with the data. The peak matches the frequency expected from the eigenvalues.

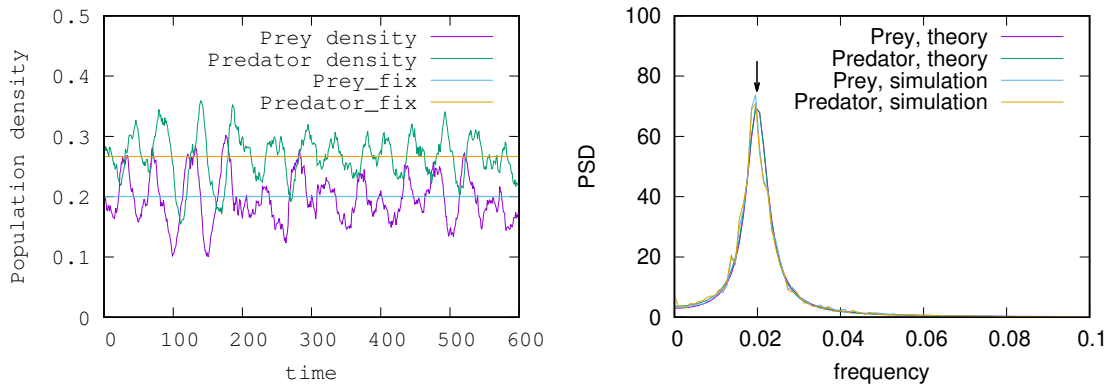


Figure 11.3: Left: An example trajectory obtained by Gillespie simulation. Plotted quantities are Density N/K and M/K. The horizontal lines show the coexistence fixed-point. Right: Comparison between the theoretically obtained power spectrum densities (11.28)-(11.29) and numerically calculated power spectrum density from Gillespie simulation (power spectrum densities were calculated from independent 64 simulations for time duration 2^{10} with initial condition close to the coexistence fixed point and averaged over the samples.). The arrow indicates the frequency expected from the complex part of the eigenvalue of Jacobian, $\frac{\alpha\delta}{2\gamma}\sqrt{-1+4\frac{\gamma^2}{\alpha\delta}\left(1-\frac{\delta}{\gamma}\right)}/2\pi$. Parameters used are $K = 1000$, $\alpha = 0.2$, $\beta = 0.1$, $\gamma = 0.5$, $\delta = 0.1$.

11.4 Summary

- Predator-Prey system was analyzed as an example of a 2-variable stochastic system.
- Even if the deterministic equation gives a stable fixed point with dumped oscillation, the noise can induce "quasi-oscillation".
- Using the expansion of the master equations and SDEs, it is possible to calculate power spectrum density. The calculated power spectrum density gives a peak at a specific frequency, consistent with the quasi-oscillation seen in the simulation.

11.5 Exercise

1. Show that the SDEs (11.10) and (11.11) gives the Fokker-Planck equation (11.9).
2. * Simulating predator-prey model.
   (a) Perform Gillespie simulation of the agent-based predator-prey model defined in section 11.1. Use the following parameters: $K = 1000$, $\alpha = 0.2$, $\beta = 0.1$, $\gamma = 0.5$, and $\delta = 0.1$. Plot the time course of the population densities N/K and M/K. Compare it with what is expected from the mean-field model's fixed points.
   (b) If the initial condition is not too close to zero for both of the populations, the behaviour should look like noisy oscillation around the coexistence fixed point. Is the period of this quasi-oscillation consistent with what you expect from the eigenvalues 11.19? Calculate the expected typical period from eq. (11.19).
   (c) Plot the theoretically evaluated power spectrum (11.28) and (11.29) as a function of the angular frequency ω. Does the peak correspond to the typical oscillation period you see?

Advanced exercise

3. Derive SDEs (11.15) and (11.16) by substituting eqs. (11.13) and (11.14) into eqs. (11.10) and (11.11) and collecting the highest order of K (which will be $\sqrt{K}$).

4. Derive the power spectrum densities (11.28) and (11.29) from (11.26) and (11.27).

5. * Numerically integrate the SDEs for the noise term (11.15) and (11.16) and plot the time course. Does it look similar to what you got in the Gillespie simulation? Tips: When you compare, pay attention to the correspondence between (N, M) and (X, Y).

6. * Numerically calculate the power spectrum density of the Gillespie simulation and confirm that it agrees with the theory. Note: It may take some time to collect enough data. Also, there are some tools to do power spectrum density calculation but they may have different normalizations.

11.6 Further reading

This chapter is based on [24, 9]

Appendix A

Proof of the central limit theorem

In this chapter, we summarize one of the simplest proof of the central limit theorem. For that, we fist introduce an convenient tool, called characteristic function.

A.1 Characteristic function

In general, to calculate the *n-th order moments* $\langle X^n \rangle$ of a stochastic variable X with a distribution $P(x)$, it is convenient to consider **the characteristic function**. The characteristic function of a probability distribution $P(x)$, $G(k)$, is just the Fourier transform of $P(x)$:

$$G(k) = \int_{-\infty}^{\infty} P(x) e^{ikx} dx = \langle e^{ikX} \rangle. \tag{A.1}$$

The inverse transform is of course

$$P(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(k) e^{-ikx} dk. \tag{A.2}$$

The characteristic function exists for any probability distribution.

Moments

IF one can expand the characteristic function with k, we have

$$G(k) = \int_{-\infty}^{\infty} P(x) \sum_{p=0}^{\infty} \frac{(ikx)^p}{p!} dx = \sum_{p=0}^{\infty} \frac{(ik)^p}{p!} \int_{-\infty}^{\infty} x^p P(x) dx = \sum_{p=0}^{\infty} \frac{(ik)^p}{p!} \langle X^p \rangle. \tag{A.3}$$

From this, it is clear that the finiteness of the *p-th order moments* $\langle X^p \rangle$ is required for the expansion of the characteristic function to the p-th order. We can in tern calculate the p-th moment if the p-th order derivative of $G(k)$ at $k = 0$ is finite, i.e.,

$$\langle X^p \rangle = \frac{1}{i^p} \cdot \frac{\partial^p G(k)}{\partial k^p}|_{k=0}. \tag{A.4}$$

Note that

$$\langle X^0 \rangle = \langle 1 \rangle = G(0) = \int_{-\infty}^{\infty} P(x) dx = 1, \tag{A.5}$$

due to the normalization condition for $P(x)$.

Cases with diverging moments

It is useful to know how characteristic function appears if (higher) moments are diverging. In the case of the *Cauchy distribution*

$$\frac{1}{\pi}\left[\frac{1}{1+x^2}\right],$$

the characteristic function is[1]

$$\exp(-|k|).$$

Since the characteristic function is singular even at order k (due to the absolute value), we can see from the characteristic function that even the first order moment (mean) is not well defined for the Cauchy distribution.

If a probability distribution $Q(x)$ has a characteristic function is

$$G_Q(k) = \exp\left[-\frac{a^2}{2}k^2 + Ak^\gamma + \cdots\right] \tag{A.6}$$

with $2 < \gamma < 3$. We will get

$$\langle X\rangle = 0, \langle X^2\rangle = a^2, \tag{A.7}$$

but the third moment $\langle X^3\rangle$ will diverge because the third derivative of $G_Q(k)$ diverges at $k = 0$.

Characteristic function for Gaussian distribution

The characteristic function of the Gaussian distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \tag{A.8}$$

is given by

$$G(k) = \exp\left(imk - \frac{\sigma^2}{2}k^2\right). \tag{A.9}$$

We can of course derive this directly from definition, though we have already seen this in another exercise when we solved the diffusion equation by using Fourier transformation. It is easy to obtain the moments for Gaussian distribution such as (1.24), (1.25), (1.28) from the characteristic function (A.9).

A.2 Proof of the central limit theorem using the characteristic functions

We now prove the central limit theorem. Let us review the set up. We draw n real stochastic variables Z_j independently from the identical distribution $P(z)$, which has zero mean ($\langle Z\rangle = 0$) and finite variance σ_z^2; higher order moments can diverge. This means that the characteristic function of $P(z)$ has the form

$$G(k) = \exp\left[-\frac{\sigma_z^2}{2}k^2 + A|k|^\alpha + \cdots\right], \tag{A.10}$$

with $\alpha > 2$, and "$\cdots$" denotes higher order in k. If for example α is between 2 and 3, then the third moment will diverge, but we still have zero mean and variance σ_z^2. We then construct

$$Y_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_j, \tag{A.11}$$

and we would like to know the probability distribution of Y_n, $P_n(y)$, in $n \to \infty$ limit.

[1] This is easily calculated by using residue theorem of complex integral.

Construction of $P_n(y)$ from $P(z)$ For example if $n = 2$, then $\sqrt{2}Y_2 = Z_1 + Z_2$. Now, the probability that Z_1 takes a value between z_1 and $z_1 + dz_1$ AND Z_2 takes a value between z_2 and $z_2 + dz_2$ is

$$P(z_1)dz_1 \cdot P(z_2)dz_2,$$

since the two numbers are independently drawn. The probability that Y_2 takes a value between y and $y+dy$, $P_2(y)dy$, is then given by

$$P_2(y)dy = \left[\int_{z_1+z_2=\sqrt{2}y} [P(z_1)P(z_2)]dz_2\right] dz_1.$$

Taking the condition for the integral explicitly, we can rewrite this to

$$P_2(y) = \sqrt{2}\int_{-\infty}^{\infty} P(\sqrt{2}y - z_2)P(z_2)dz_2. \tag{A.12}$$

Here, the factor $\sqrt{2}$ came from Jacobian $\sqrt{2}dy = dz_1$. Performing the similar procedure for larger n, it is clear that

$$P_n(y) = \sqrt{n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} P(\sqrt{n}y - (z_2 + \cdots + z_n))P(z_2)\cdots P(z_n)dz_2\cdots dz_n. \tag{A.13}$$

Expressing in terms of characteristic function Now let us define the characteristic functions of $P(z)$ and $P_n(y)$ as

$$G(k) = \int_{-\infty}^{\infty} e^{ikz}P(z)dz \tag{A.14}$$

and

$$G_n(k) = \int_{-\infty}^{\infty} e^{iky}P_n(y)dy, \tag{A.15}$$

respectively. Performing the Fourier transform (FT) to the both side of (A.13), we notice that this is just FT of convolution, which results in a very simple form as follows:

$$\begin{aligned}
G_n(k) &= \int_{-\infty}^{\infty} e^{iky}P_n(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{iky}\sqrt{n}P(\sqrt{n}y - (z_2 + \cdots + z_n))P(z_2)\cdots P(z_n)dz_2\cdots dz_n dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i(k/\sqrt{n})u}\sqrt{n}P(u - (z_2 + \cdots + z_n))P(z_2)\cdots P(z_n)dz_2\cdots dz_n\frac{du}{\sqrt{n}} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i(k/\sqrt{n})(u'+z_2+\cdots z_n)}P(u')P(z_2)\cdots P(z_n)dz_2\cdots dz_n du' \\
&= \int_{-\infty}^{\infty} e^{i(k/\sqrt{n})u'}P(u')du'\int_{-\infty}^{\infty} e^{i(k/\sqrt{n})z_2}P(z_2)dz_2\cdots\int_{-\infty}^{\infty} e^{i(k/\sqrt{n})z_n}P(z_n)dz_n \\
&= \left[G\left(\frac{k}{\sqrt{n}}\right)\right]^n,
\end{aligned} \tag{A.16}$$

where from the second to the third line y is converted to a new variable $u = \sqrt{n}y$, and from the third line to the forth line u is converted to $u' = u - (z_2 + \cdots z_n)$.

Since we know that $G(k)$ is expanded as (A.10), by substituting this to (A.16) we have

$$G_n(k) = \exp\left[-n\frac{k^2}{2n}\sigma_z^2 + nA\left|\frac{k}{\sqrt{n}}\right|^{\alpha} + \cdots\right] = \exp\left[-\frac{k^2}{2}\sigma_z^2 + A|k|^{\alpha}n^{1-\alpha/2} + \cdots\right]. \tag{A.17}$$

Since $1 - \alpha/2 < 0$ (remember $\alpha > 2$), we have

$$\lim_{n\to\infty} G_n(k) = \exp\left[-\frac{k^2}{2}\sigma_z^2\right]. \tag{A.18}$$

Compared with (A.9), this is the characteristic function of a Gaussian with zero mean and variance σ_z^2! The inverse FT will finally give the central limit theorem

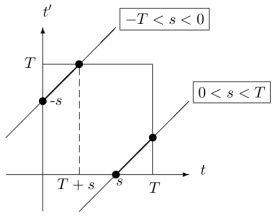
$$\lim_{n\to\infty} P_n(y) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left(-\frac{y^2}{2\sigma_z^2}\right). \tag{A.19}$$

Appendix B

Proof of Wiener-Khinchin theorem for finite time series

Substitute (2.29) to (2.30):

$$\begin{aligned} I(\omega) &= \lim_{T\to\infty}\frac{1}{T}\int_0^T X(t)\exp(i\omega t)\,dt\int_0^T X(t')\exp(-i\omega t')\,dt' \\ &= \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\int_0^T dt'\,[X(t)X(t')\exp(i\omega(t-t'))] \end{aligned} \tag{B.1}$$



Now change variables from (t,t') to (t,s) with $s=t-t'$. Noting the figure to the left (the integral is within the square, and for a given s t moves along the line given by $t=t'+s$), we get

$$I(\omega) = \lim_{T\to\infty}\frac{1}{T}\left[\int_0^T ds\int_s^T dtX(t)X(t-s)e^{i\omega s}+\int_{-T}^0 ds\int_0^{T+s} dtX(t)X(t-s)e^{i\omega s}\right]. \tag{B.2}$$

By changing variable for the first term from t to $\tau=t-s$, we get

$$I(\omega) = \lim_{T\to\infty}\frac{1}{T}\left[\int_0^T dse^{i\omega s}\int_0^{T-s} d\tau X(\tau+s)X(\tau)+\int_{-T}^0 dse^{i\omega s}\int_0^{T+s} dtX(t)X(t-s)\right]. \tag{B.3}$$

By taking $T\to\infty$ limit and using the definition for the time average, this becomes

$$I(\omega)=\left[\int_0^\infty dse^{i\omega s}C(s)+\int_{-\infty}^0 dse^{i\omega s}C(-s)\right]=\int_{-\infty}^\infty e^{i\omega s}C(s), \tag{B.4}$$

where $C(s)=C(-s)$ is used. This is the Wiener-Khinchin theorem (2.24). Note that the inverse Fourier transformation gives

$$C(t)=\frac{1}{2\pi}\int_{-\infty}^\infty I(\omega)e^{-i\omega t}dt. \tag{B.5}$$

What does the power spectrum mean? For finite interval of T, the Fourier expansion has the form

$$A(2\pi f_n)=\int_0^T X(t)e^{i2\pi f_n t}dt,\quad f_n\equiv\frac{n}{T},$$

with

$$X(t) = \frac{1}{T}\sum_{n=-\infty}^{\infty} A(2\pi f_n)e^{-i2\pi f_n t}.$$

If we were considering an oscillator with frequency f, its amplitude squared would be proportional to the energy of that oscillator. From this analogy, we can consider $|A(2\pi f_n)|^2/T^2$ as a measure of the strength or the power of the mode with frequency f_n. Especially, we define the power spectrum or the power spectral density $I_n(2\pi f)$ using the sum of the amplitude for the mode in the frequency between f and $f + \Delta f$ as

$$I_n(2\pi f)\Delta f = \sum_{f \le f_n \le f+\Delta f} \frac{1}{T^2}\langle |A(2\pi f_n)|^2\rangle.$$

Now, when Δf is small, we may assume $A(2\pi f_n) \approx A(2\pi f)$. At the same time, the number of the modes in the interval Δf is $\Delta f/(1/T) = \Delta f T$. Therefore we could write

$$I_n(2\pi f)\Delta f \approx \Delta f T \frac{1}{T^2}\langle |A(2\pi f)|^2\rangle.$$

The limit of $T \to \infty$ gives (2.30).

Appendix C

Derivation of First passage time distribution using reflection principle

This appendix is based on ref. [4]. Let us think about slightly different set up than the text, by starting the standard Brownian motion / Wiener process, i.e.,

$$\frac{dB(t)}{dt} = \hat{R}(t), \quad \langle \hat{R}(t) \rangle = 0, \quad \langle \hat{R}(t)\hat{R}(t') \rangle = \delta(t - t'), \tag{C.1}$$

with $B(0) = 0$.

We consider the time that it takes the Brownian particle to reach a given position, $a > 0$. (So now the particle is starting at $B = 0$, $B = a$ is the absorbing boundary, and the space is open semi-infinite for $B < a$. Since the Brownian motion is symmetric in the motion to the right and the left, this will have the same statistics as the problem we considered in the text.) We derive the probability distribution of T_a, $P(T)$.

Derivation of the first passage time distribution

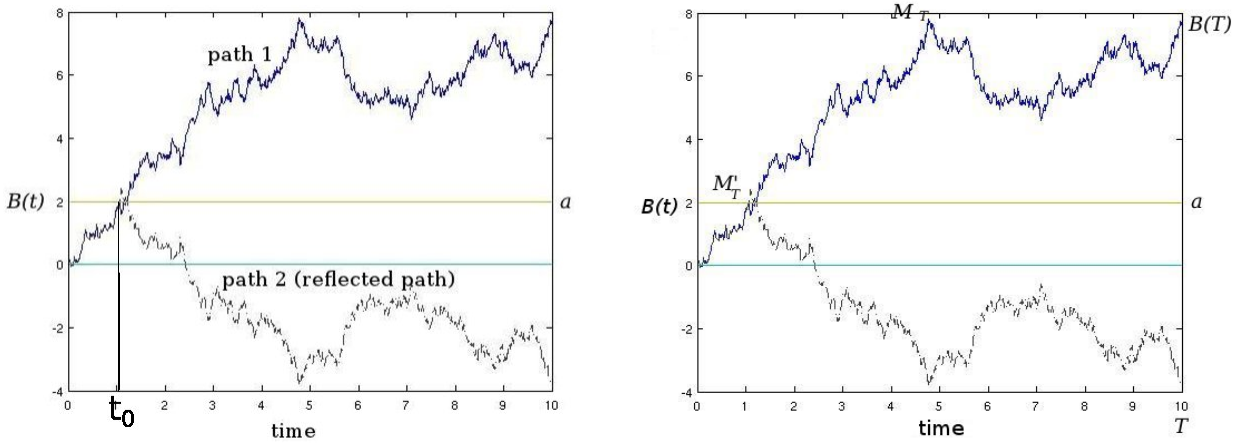


Figure C.1: Schematic description of reflection of a path.

Preparation: reflection of a path We need to use a "trick" to solve this problem.

Consider a sample path 1 which cross the position a at time t_0 (Fig. C.1, solid line). We can also consider another path 2, which is the same as the path 1 until time t_0, but then *reflected by the horizontal line $B = a$ after that* (dashed line).

Q. Which path is more likely to realize?

The reflection principle Since the free Brownian motion is *symmetric*, the probability to move positive direction just after t_0 and the probability to move negative direction just after t_0 are exactly the same. Therefore, we can have the *reflection principle*, i.e., if a sample path is reflected as we did, the new path has the same probability to occur as the first path.

The position at time T and the maximum value taken Now, take many sample path up to time T. Some of samples end up at the the final position $B(T)$ is bigger than or equal to a, $B(T) \geq a$. Look at one of a such path. The maximum value the path took until time T, M_T, surely satisfies $M_T \geq a$ (See Fig. C.1).

Now reflect the path at the time t_0 at which the path crossed a for the first time. For the reflected path, the value at time T will be less than a. The reflected path's maximum value M'_T also satisfies $M'_T \geq a$.

This gives us relationship between the following two quantities:

- Prob($M_T \geq a$): the probability that the maximum value M_T of the path until time T to be bigger or equal to a.
- Prob($B(T) \geq a$): the probability that the value at time T to be bigger or equal to a.

From the argument above, we have

$$\text{Prob}(M_T \geq a) = 2\text{Prob}(B(T) \geq a). \tag{C.2}$$

The first passage time distribution and Prob($M_T \geq a$) How do these concepts have anything to do with the first passage time distribution $P(T)$?

Remember that the first passage time T_a is the time when $B(t)$ cross a for the first time. When a path $B(t)$ satisfy $M_T \geq a$, it had crossed a before T. Namely, Prob($M_T \geq a$) is equal to the probability for a path to have the first passage time smaller than T, i.e., $0 \leq T_a \leq T$. This leads

$$\text{Prob}(M_T \geq a) = \int_0^T P(t)dt. \tag{C.3}$$

By taking derivative by T, we get

$$P(T) = \frac{d}{dT}\text{Prob}(M_T \geq a). \tag{C.4}$$

Therefore, if we know Prob($M_T \geq a$), we can calculate the first passage time distribution using (C.4).

Prob($B(T) \geq a$) We can calculate Prob($B(T) \geq a$) based on what we know about the free diffusion. When the initial condition is $B(0) = 0$, the probability for $B(T)$ to take value x is given by

$$P(x,t) = \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{x^2}{2T}\right) \tag{C.5}$$

(eq. (1.12) with $x_0 = 0, t_0 = 0, t = T, D = 1/2$.) Therefore, the probability for $B(T)$ to be bigger than a is

$$\text{Prob}(B(T) \geq a) = \int_a^\infty P(x,t)dx = \frac{1}{\sqrt{2\pi T}} \int_a^\infty \exp\left(-\frac{x^2}{2T}\right) dx. \tag{C.6}$$

The first passage time distribution Substituting (C.2) to (C.4) and then using (C.6), we get

$$P(T) = \frac{d}{dT}\left[\frac{2}{\sqrt{2\pi T}} \int_a^\infty \exp\left(-\frac{x^2}{2T}\right) dx\right]. \tag{C.7}$$

By converting the integration variable from x to $v = (Ta^2)/x^2$, we get (Note $v > 0$ and $x > 0$ for all the integral that the conversion $x = a\sqrt{T/v}$ can be used)

$$\begin{aligned} P(T) &= \frac{d}{dT}\left[\frac{2}{\sqrt{2\pi T}}\int_0^T \exp\left(-\frac{a^2}{2v}\right)\frac{a\sqrt{T}}{2v^{3/2}}dv\right] \\ &= \frac{d}{dT}\left[\frac{a}{\sqrt{2\pi}}\int_0^T \frac{1}{v^{3/2}}\exp\left(-\frac{a^2}{2v}\right)dv\right] \\ &= \frac{a}{\sqrt{2\pi}}\frac{1}{T^{3/2}}\exp\left(-\frac{a^2}{2T}\right). \end{aligned} \tag{C.8}$$

In summary, the first passage time distribution is given by

$$P(T) = \frac{a}{\sqrt{2\pi}}\frac{1}{T^{3/2}}\exp\left(-\frac{a^2}{2T}\right). \tag{C.9}$$

Appendix D

Derivation of Master equation from Chapmann-Kolmogorov equation for a continuous variable case

D.0.1 Markov process

We can also define the Markov process for a continuous variable case. The transition probability $P(x_1, t_1|x_0, t_0)$ should be defined as the probability that $x_1 < X(t_1) \leq x_1 + dx_1$ under the condition that $X(t_0) = x_0 \quad (t_0 < t_1)$. Formally, it is given by:

$$P(x_1, t_1|x_0, t_0)dx_1 = \frac{P_2(x_0, t_0; x_1, t_1)dx_1}{P(x_0, t_0)}. \tag{D.1}$$

The definition of P_2 is given in eq. (1.6) for general case of n-point probability density, but here I repeat it for P_2:

$$P_2(x_0, t_0; x_1, t_1)dx_0dx_1 = Pr(x_0 < X(t_0) \leq x_0 + dx_0 \text{ and } x_1 < X(t_1) \leq x_1 + dx_1). \tag{D.2}$$

Using the conditional probability, the memory-less property of a Markov process can be formally expressed for $t_0 < t_1 < \cdots < t_{k-1} < t_k$ as

$$P(x_k, t_k|x_0, t_0; x_1, t_1; \cdots; x_{k-1}, t_{k-1}) = P(x_k, t_k|x_{k-1}, t_{k-1}), \tag{D.3}$$

which means that the probability density to be at x_k at time t_k given the condition that it had a specific history of being x_0 at time t_0, x_1 at time t_1, $\cdots$, x_{k-1} at time t_{k-1} depends only on the last time point of the history, i.e., being x_{k-1} at time t_{k-1}.

The memory less property can be also expressed by using the probability for the observed value of $X(t)$ for 2 time points under the given initial condition $X(t_0) = x_0$, $P(x_1, t_1; x_2, t_2|x_0, t_0)$, as

$$P(x_1, t_1; x_2, t_2|x_0, t_0) = P(x_2, t_2|x_1, t_1)P(x_1, t_1|x_0, t_0) \quad (t_0 < t_1 < t_2). \tag{D.4}$$

This also means that we can construct the transition probability between (x_0, t_0) to (x_2, t_2) as

$$P(x_2, t_2|x_0, t_0) = \int_{-\infty}^{\infty} P(x_2, t_2|x_1, t_1)P(x_1, t_1|x_0, t_0)dx_1. \tag{D.5}$$

Eq.(D.5) is the **Chapman-Kolmogorov equation**.

D.0.2 Derivation of master equation

A homogeneous process satisfy

$$P(x_1, t_1|x_0, t_0) = T(x_1|x_0, t_1 - t_0). \tag{D.6}$$

Therefore, for a homogeneous Markov process, we have

$$T(x_2|x_0,\tau+\tau') = \int_{-\infty}^{\infty} T(x_2|x_1,\tau')T(x_1|x_0,\tau)dx_1. \tag{D.7}$$

The derivation of the master equation for the continuous variable case is completely in parallel with the discrete variable case, just keeping in mind that we deal with the probability density: Practically, you just need to replace the discrete values n_1 and n_2 with the continuous values x_1 and x_2, replace the sums over all possible values of n_1 or n_2 with the integrals of x_1 or x_2, and replace the Kronecker's delta δ_{n_1,n_2} with the Dirac delta function for the state space $\delta(x_2-x_1)$.

Normalization condition is

$$\int_{-\infty}^{\infty} T(x_2|x_1,t)dx_2 = 1. \tag{D.8}$$

This means that the probability that the particle located at x_1 will appear somewhere in the system after time t is one. For very short time interval Δt, the transition probability $T(x_2|x_1,\Delta t)$ can be expanded as follows:

$$T(x_2|x_1,\Delta t) = F(\Delta t)\delta(x_2-x_1) + \Delta t\cdot W(x_2|x_1) + O(\Delta t^2), \tag{D.9}$$

where the first term is the probability to stay the position, and the second term is the transition probability for the very short time. Here, $W(x_2|x_1)$ is the transition rate. Therefore, the probability that a transition $x_1 \to x_2$ occurs during Δt is $\Delta t\cdot W(x_2|x_1)$. Substituting (D.9) to (D.8), we get

$$1 = F(\Delta t) + \Delta t\int_{-\infty}^{\infty} W(x_2|x_1)dx_2 + O(\Delta t^2), \tag{D.10}$$

which gives (up to the first order of Δt)

$$F(\Delta t) = 1 - a_0(x_1)\Delta t, \tag{D.11}$$

$$a_0(x_1) = \int_{-\infty}^{\infty} W(x_2|x_1)dx_2. \tag{D.12}$$

By substituting eq.(D.9) to (D.7) with $\tau' = \Delta t$, we get

$$\begin{aligned} T(x_2|x_0,\tau+\Delta t) &= \int \left[(1-a_0(x_1)\Delta t)\delta(x_2-x_1) + \Delta t\cdot W(x_2|x_1) + O(\Delta t^2)\right]\times T(x_1|x_0,\tau)dx_1 \\ &= (1-a_0(x_2)\Delta t)T(x_2|x_0,\tau) + \Delta t\int_{-\infty}^{\infty} W(x_2|x_1)T(x_1|x_0,\tau)dx_1 + O(\Delta t^2), \end{aligned}$$

which leads

$$\begin{aligned} T(x_2|x_0,\tau+\Delta t) - T(x_2|x_0,\tau) &= \Delta t\left[\int_{-\infty}^{\infty} W(x_2|x_1)T(x_1|x_0,\tau)dx_1 - a_0(x_2)T(x_2|x_0,\tau)\right] \\ &\quad + O(\Delta t^2). \end{aligned} \tag{D.13}$$

By dividing both side of (D.13) by Δt and taking the limit of $\Delta t \to 0$, we get

$$\frac{\partial}{\partial\tau}T(x_2|x_0,\tau) = \int_{-\infty}^{\infty}\left[W(x_2|x_1)T(x_1|x_0,\tau) - W(x_1|x_2)T(x_2|x_0,\tau)\right]dx_1. \tag{D.14}$$

Here, the definition of $a_0(x)$ is also used.

More familiar form is

$$\frac{\partial}{\partial t}P(x,t) = \int_{-\infty}^{\infty}\left[W(x|y)P(y,t) - W(y|x)P(x,t)\right]dy,$$

because $T(x|x_0,t)$ can be considered as the probability density $P(x,t)$ for the initial condition at $t=0$ to be $P(x,t) = \delta(x-x_0)$. This is exactly the master equation (7.50).

The transition rate $W(x|y)$ is related to the transition property as

$$T(x_2|x_1,\Delta t) = (1-a_0(x_1)\Delta t)\,\delta(x_2-x_1) + \Delta t\cdot W(x_2|x_1) + O(\Delta t^2), \tag{D.15}$$

with

$$a_0(x_1) = \int_{-\infty}^{\infty} W(x_2|x_1)dx_2. \tag{D.16}$$

D.0.3 Example: Diffusion

Let us consider how diffusion (or Brownian motion in over-damped limit) appears in this frame work. We know that the solution of the diffusion equation with the initial condition being $P(x, t_0) = \delta(x - x_0)$ is given by eq. (1.12). We can do the other-way around: We can consider a process that is defined by the transition probability of the form

$$P(x_1, t_1|x_0, t_0) = T(x_1|x_0, t_1 - t_0) = \frac{1}{\sqrt{4\pi D(t_1 - t_0)}} \exp\left(-\frac{(x_1 - x_0)^2}{4D(t_1 - t_0)}\right). \tag{D.17}$$

Clearly this depends only on the time difference, thus this is a homogeneous process. It is straightforward (but tedious) to confirm that eq. (D.17) satisfies the CK equation (D.7), thus this process is a homogeneous Markov process.

How does the corresponding master equation look like? Well, it is not straightforward to expand it in the form given in eq. (D.15). Instead, in the next chapter, we will learn how to make a formal expansion of mater equation in the state space. By applying it to the transition probability given (D.17), we see that it will in the end give us the diffusion equation.

Appendix E

Kramars-Moyal expansion for a continuous variable case

E.0.1 Continuous variable case

Now let us apply the Kramers-Moyal expansion to a general master equation (7.50):

$$\frac{\partial}{\partial t}P(x,t) = \int_{-\infty}^{\infty} [W(x|y)P(y,t) - W(y|x)P(x,t)]\, dy. \tag{E.1}$$

Here, $W(x|y)$ is the transition rate from position y to x.

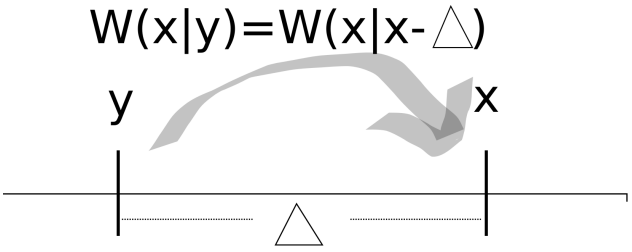


Figure E.1: Schematic figure to describe the jump size.

We re-express $W(x|y)$ with using the *jump size* $\Delta = x - y$ (See Fig. E.1), as

$$W(x|y) = W(x|x-\Delta). \tag{E.2}$$

Then we have

$$\frac{\partial}{\partial t}P(x,t) = \int_{-\infty}^{\infty} [W(x|x-\Delta)P(x-\Delta,t) - W(x-\Delta|x)P(x,t)]\, d\Delta. \tag{E.3}$$

The KM expansion is the expansion of the master equation in terms of the jump size, as we see.

If $W(x-\Delta|x)$ and $P(x,t)$ are smooth function of x, we can formally expand the first term in the integral in (E.3) as

$$\begin{aligned} & W(x|x-\Delta)P(x-\Delta,t) = W([x-\Delta]+\Delta|x-\Delta)P(x-\Delta,t) \\ = \;& W(x+\Delta|x)P(x,t) + \sum_{k=1}^{\infty} \frac{(-\Delta)^k}{k!}\frac{\partial^k}{\partial x^k}[W(x+\Delta|x)P(x,t)]. \end{aligned} \tag{E.4}$$

Substituting (E.4) and

$$\int_{-\infty}^{\infty} W(x-\Delta|x)P(x,t)d\Delta = \int_{\infty}^{-\infty} W(x+\Delta'|x)P(x,t)(-d\Delta') = \int_{-\infty}^{\infty} W(x+\Delta'|x)P(x,t)d\Delta' \tag{E.5}$$

to (E.3), we get

$$\begin{aligned}
\frac{\partial}{\partial t}P(x,t) &= \int_{-\infty}^{\infty}\left[W(x+\Delta|x)P(x,t)+\sum_{k=1}^{\infty}\frac{(-\Delta)^k}{k!}\frac{\partial^k}{\partial x^k}[W(x+\Delta|x)P(x,t)]\right]d\Delta \\
&\quad -\int_{-\infty}^{\infty}W(x+\Delta'|x)P(x,t)d\Delta' \\
&= \int_{-\infty}^{\infty}\left(\sum_{k=1}^{\infty}\frac{(-\Delta)^k}{k!}\frac{\partial^k}{\partial x^k}[W(x+\Delta|x)P(x,t)]\right)d\Delta \qquad \text{(E.6)} \\
&= \sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\frac{\partial^k}{\partial x^k}[a_k(x)P(x,t)], \qquad \text{(E.7)} \\
a_k(x) &\equiv \int_{-\infty}^{\infty}\Delta^k W(x+\Delta|x)d\Delta. \qquad \text{(E.8)}
\end{aligned}$$

$a_k(x)$ is called *the jump moment*. Note that $k=0$ of this expression is the same as (D.16).

Equation (E.7) is the Kramers-Moyal expansion of the master equation.

E.1 Jump moments

The obtained expansion

$$\begin{aligned}
\frac{\partial}{\partial t}P(x,t) &= \sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\frac{\partial^k}{\partial x^k}[a_k(x)P(x,t)], \qquad \text{(E.9)} \\
a_k(x) &\equiv \int_{-\infty}^{\infty}\Delta^k W(x+\Delta|x)d\Delta. \qquad \text{(E.10)}
\end{aligned}$$

is useful to see what kind of "master equation" we get when we have a transition probability expressed in a continuous function. To make full use of it, we consider a bit deeper about the meaning of the jump moments, (E.10).

We can rewrite the definition of the jump moments (E.10) as

$$a_k(x)=\int_{-\infty}^{\infty}\Delta^k W(x+\Delta|x)d\Delta=\int_{-\infty}^{\infty}(y-x)^k W(y|x)dy \qquad \text{(E.11)}$$

Now, for later convenience we introduce

$$\begin{aligned}
A_k(x;\tau,t) &\equiv \left\langle[X(t+\tau)-X(t))]^k\right\rangle|_{X(t)=x} \\
&= \int_{-\infty}^{\infty}\left[(y-x)^k P(y,t+\tau|x,t)\right]dy=\int_{-\infty}^{\infty}\left[(y-x)^k T(y|x,\tau)\right]dy, \qquad \text{(E.12)}
\end{aligned}$$

which is the average of $(X(t+\tau)-X(t))^k$ under the initial condition that $X(t)=x$, and the homogeneous property (D.6) was used in the second equality. Using the short time expansion of the transition probability (D.15), we have

$$\begin{aligned}
A_k(x;\Delta t,t) &= \int_{-\infty}^{\infty}(y-x)^k\left[(1-a_0(x_1)\Delta t)\delta(y-x)+\Delta t\cdot W(y|x)+O(\Delta t^2)\right]dy \\
&= \Delta t\int_{-\infty}^{\infty}(y-x)^k\ W(y|x)dy+O(\Delta t^2) \\
&= a_k(x)\Delta t+O(\Delta t^2). \qquad \text{(E.13)}
\end{aligned}$$

Therefore, the jump moment can be expressed as

$$a_k(x)=\lim_{\Delta t\to 0}\frac{A_k(x;\Delta t,t)}{\Delta t}. \qquad \text{(E.14)}$$

E.1.1 KM expansion and diffusion

Now let us see how these relations can be used in a particular example.

As noted in subsection D.0.3, diffusion can be considered as a homogeneous Markov process that has the transition probability (D.17), i.e.,

$$T(x_1|x_0,\tau) = \frac{1}{\sqrt{4\pi D\tau}} \exp\left(-\frac{(x_1-x_0)^2}{4D\tau}\right). \tag{E.15}$$

By knowing this, and by using (E.12) and (E.14), we can calculate the jump moments. We have

$$\begin{aligned} A_k(x;\tau,t) &= \int_{-\infty}^{\infty} (y-x)^k T(y|x,\tau)dy \\ &= \int_{-\infty}^{\infty} (y-x)^k \frac{1}{\sqrt{4\pi D\tau}} \exp\left(-\frac{(y-x)^2}{4D\tau}\right) dy. \\ &= \int_{-\infty}^{\infty} z^k \frac{1}{\sqrt{4\pi D\tau}} \exp\left(-\frac{z^2}{4D\tau}\right) dz, \end{aligned}$$

where in the last equality $z = y - x$ was introduced. This is simply moments for Gaussian distribution. Using (1.24) to (1.28), we have

$$\begin{aligned} A_1(x;\tau,t) &= 0, \\ A_2(x;\tau,t) &= 2D\tau, \\ A_3(x;\tau,t) &= 0, \\ A_4(x;\tau,t) &= 12D^2\tau^2. \end{aligned}$$

Therefore, using (E.14), the jump moments are

$$\begin{aligned} a_1(x) &= 0, \\ a_2(x) &= 2D, \\ a_3(x) &= 0, \\ a_4(x) &= \lim_{\tau\to 0}(12D^2\tau) = 0. \end{aligned}$$

It is easy to confirm that higher jump moments are all zero. Namely, only the second jump moment is non-zero, and the corresponding KM expansion is

$$\frac{\partial}{\partial t}P(x,t) = \frac{1}{2}\cdot\frac{\partial^2}{\partial x^2}\left(2DP(x,t)\right) = D\frac{\partial^2}{\partial x^2}P(x,t), \tag{E.16}$$

which is the diffusion equation. Note that we did not need to perform the truncation of the expansion - higher order terms are exactly zero. In other word, the diffusion equation is equivalent to the "master equation" of the process defined by the transition probability (E.15).

Appendix F

System size expansion and Linear Noise Approximation for multi-component system

For chemical reaction system which is known to obey the rate equation for the large volume limit, the systematic expansion is possible around $\Omega \to \infty$.

The method the Ω-*expansion*, proposed by van Kampen, and especially we make the *linear noise approximation*.

We first introduce the notations for multi-component systems, and then perform the expansion.

F.1 General formulation for Chemical reactions

Now let us generalize the formulation. For convenience, we introduce a few new notations.

Density vector

Consider a system with volume Ω, with M different kinds of chemical components. For $i = 1, 2, \cdots M$, we define the following:

N_i : Number of the molecule of the i-th kind

$x_i \equiv N_i/\Omega$: Density of the molecule of the i-th kind

For convenience, we use the vector expression

$$\vec{x}(t) \equiv \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{pmatrix}, \tag{F.1}$$

$$\vec{N}(t) = \Omega \vec{x}(t). \tag{F.2}$$

The state of the system is defined by $\vec{N}(t)$ (or equivalently $\vec{x}(t)$).

F.1.1 Reaction rates and Stoichiometric matrix

Suppose that there are R kinds of chemical reactions. We assume that for a small time interval Δt, the probability that the reaction j is given by

$$\Omega \tilde{\nu}_j(\vec{x}, \Omega)\Delta t. \tag{F.3}$$

Note that it was written as if the rate is proportional to Ω, because later we would like to consider the situation where we change volume Ω with keeping concentrations constant. Of course $\tilde{\nu}_j$ depends on Ω, too, so it can be anything at this point.

With this reaction, the number of molecules of kind i, N_i, increase by S_{ij} (decrease when S_{ij} is negative). Namely

$$N_i \to N_i + S_{ij}. \tag{F.4}$$

Apparently the *stoichiometric matrix* $\mathbf{S}$ is $M \times R$ matrix.

F.1.2 Large volume limit

Conceptually, the reaction means

$$N_i(t+\Delta t) - N_i(t) \approx \sum_{j=1}^{R} S_{ij}\Omega\tilde{\nu}_j(\vec{x},\Omega)\Delta t + (\text{some noise}). \tag{F.5}$$

though the noise term is not very well defined in this expression.

If we increase the volume Ω *with keeping the average density constant*, we normally reach the deterministic rate equation for the density using the deterministic limit of the reaction rate

$$\vec{\phi} = \lim_{\Omega\to\infty} \vec{x}, \quad \nu_j(\vec{\phi}) = \lim_{\Omega\to\infty} \tilde{\nu}_j(\vec{x},\Omega). \tag{F.6}$$

$$\frac{d\phi_i}{dt} = \sum_{j=1}^{R} S_{ij}\nu_j(\vec{\phi}) \tag{F.7}$$

This can be understood well if we divide (F.5) with $\Omega\Delta t$ and assume that the noise go away in the $\Omega \to \infty$ limit. We sometimes call ν_j as a macroscopic reaction rate.

F.1.3 Master equation for M-component system

We now consider the master equation for having $\vec{n}(t)$ particles at time t, $P(\vec{n},t)$. For this, it is convenient to introduce the *step operator* E_i^k, which acts on a function of $\vec{n}$,

$$f(\vec{n}) = f(n_1, n_2, \cdots, n_M) \tag{F.8}$$

as

$$\begin{aligned} E_i^k f(\vec{n}) &= E_i^k f(n_1, n_2, \cdots, n_M) && \text{(F.9)} \\ &= f(n_1, n_2, \cdots, n_i + k, \cdots n_M). && \text{(F.10)} \end{aligned}$$

Considering the reaction j moves the state $(N_1, N_2, \cdot, N_i, \cdot, N_M)$ to $(N_1, N_2, \cdot, N_i + S_{ij}, \cdot, N_M)$ with the rate $\Omega\tilde{\nu}_j(\vec{x},\Omega) = \Omega\tilde{\nu}_j(\vec{N}/\Omega,\Omega)$, we have (Note that $\prod_{i=1}^{M} E_i^{-S_{ij}} = E_1^{-S_{1j}} E_2^{-S_{2j}} \cdots E_M^{-S_{Mj}}$)

$$\frac{\partial P(\vec{n},t)}{\partial t} = \Omega \sum_{j=1}^{R} \left[\left(\prod_{i=1}^{M} E_i^{-S_{ij}} \right) - 1 \right] \tilde{\nu}_j(\vec{n}/\Omega,\Omega) P(\vec{n},t). \tag{F.11}$$

Let us see a few examples.

Simple chemical reaction with 1-component

Suppose there is only one kind of the molecule ($M = 1$), whose number is N (and hence the density $x = N/\Omega$). If one molecule is produced at the rate $\Omega\alpha$ and degraded at the rate Γ per molecule, we have $R = 2$ reactions and

$$\Omega\tilde{\nu}_1(x,\Omega) = \Omega\alpha, \qquad S_{1,1} = 1, \tag{F.12}$$

$$\Omega\tilde{\nu}_2(x,\Omega) = \Gamma N = \Omega\Gamma x, \qquad S_{1,2} = -1. \tag{F.13}$$

The master equation is given by

$$\begin{aligned}\frac{\partial P(n,t)}{\partial t} &= \Omega\left(E_1^{-1}-1\right)\alpha P(n,t)+\Omega\left(E_1^{1}-1\right)\Gamma\frac{n}{\Omega}P(n,t)\\ &= \Omega\left(\alpha P(n-1,t)-\alpha P(n,t)\right)+\Omega\left(\Gamma\frac{(n+1)}{\Omega}P(n+1,t)-\Gamma\frac{n}{\Omega}P(n,t)\right)\\ &= \Omega\alpha(P(n-1,t)-P(n,t))+\Gamma((n+1)P(n+1,t)-nP(n,t)). \end{aligned} \tag{F.14}$$

This is expected master equation for this one-step process.

The deterministic limit of the density is given by

$$\phi = \lim_{\Omega\to\infty} x \tag{F.15}$$

and the reaction rate is

$$\begin{aligned}\nu_1(\phi) &= \lim_{\Omega\to\infty}\alpha=\alpha,\\ \nu_2(\phi) &= \lim_{\Omega\to\infty}\Gamma x=\Gamma\phi.\end{aligned}$$

Therefore, the rate equation is given by

$$\begin{aligned}\frac{d\phi}{dt} &= S_{1,1}\nu_1(\phi)+S_{1,2}\nu_2(\phi)\\ &= \alpha-\Gamma\phi. \end{aligned} \tag{F.16}$$

This matches with what we expected, (7.43).

F.1.4 Gene expression

The mRNA and protein production in a cell, that we considered in exercise 6 of chapter 7, is a good example of $M=2$ components. There are 4 reactions. Identifying $\tilde{\nu}_j$ and $S_{i,j}$, and taking $\Omega\to\infty$ limit for determining the rate equation.

F.2 System size expansion

Fluctuation around the density in the large volume limit

Suppose the number of the molecules in the system is $\vec{N}$ (there are M kinds), and the volume is Ω. We assume that the molecule numbers is written as

$$N_i(t) = \Omega\phi_i(t)+\sqrt{\Omega}Z_i(t). \tag{F.17}$$

Here, ϕ_i is the density. The part $\sqrt{\Omega}Z_i$ describes the fluctuation, and the square-root dependence is the *guess* based on the Poisson statistics etc. (See e.g. (58)). Note that Z_i is a stochastic variable.

We will derive the equation for the probability distribution of $\vec{Z}$, to see how the fluctuating part Z_i behaves.

Distribution for $\vec{Z}$

Our starting point is the master equation for chemical reactions

$$\frac{\partial P(\vec{n},t)}{\partial t} = \Omega\sum_{j=1}^{R}\left[\left(\prod_{i=1}^{M}E_i^{-S_{ij}}\right)-1\right]\tilde{\nu}_j(\vec{n}/\Omega,\Omega)P(\vec{n},t). \tag{F.18}$$

We first need to convert $P(\vec{n},t)$ to the probability $Q(\vec{z},t)d\vec{z}=Q(\vec{z},t)dz_1dz_2\cdots dz_M$, the probability that $\vec{Z}$ takes the value between $\vec{z}$ and $\vec{z}+d\vec{z}$. By noticing that Z_i changes by $1/\sqrt{\Omega}$ when N_i change by 1 by fluctuation, we have the conversion

$$P(\vec{n},t) = \Omega^{-M/2}Q(\vec{z},t). \tag{F.19}$$

Time derivative

First we consider the LHS of (F.18), the partial time derivative. We should note that the derivative is for constant n and varying t, which is NOT constant z. Namely

$$\frac{\partial}{\partial t}P(\vec{n},t) = \Omega^{-M/2}\left[\frac{\partial}{\partial t}Q(\vec{z},t) + \sum_{i=1}^{M}\left(\frac{dz_i}{dt}\right)_{\vec{n}=const}\frac{\partial}{\partial z_i}Q(\vec{z},t)\right]. \tag{F.20}$$

Taking derivative of $n_i = \Omega\phi_i(t) + \sqrt{\Omega}z_i(t)$ with constant n_i,

$$0 = \Omega\frac{d\phi_i(t)}{dt} + \sqrt{\Omega}\frac{dz_i(t)}{dt}, \tag{F.21}$$

which gives

$$\frac{\partial}{\partial t}P(\vec{n},t) = \Omega^{-M/2}\left[\frac{\partial}{\partial t}Q(\vec{z},t) - \sum_{i=1}^{M}\sqrt{\Omega}\frac{d\phi_i(t)}{dt}\frac{\partial}{\partial z_i}Q(\vec{z},t)\right]. \tag{F.22}$$

Expansion of step operators

In the (F.18), we have step operators E_i^k. In the large Ω limit, we can expand E_i^k as follows. From the definition (F.10)

$$\begin{aligned}
&E_i^k f(\cdots, n_i, \cdots) = f(\cdots, n_i + k, \cdots)\\
&= f(\cdots, \Omega\phi_i + \sqrt{\Omega}z_i + k, \cdots) = f(\cdots, \Omega\phi_i + \sqrt{\Omega}(z_i + k/\sqrt{\Omega}), \cdots)\\
&= f(\cdots, \Omega\phi_i + \sqrt{\Omega}z_i, \cdots) + \frac{k}{\sqrt{\Omega}}\frac{\partial}{\partial z_i}f(\cdots, \Omega\phi_i + \sqrt{\Omega}z_i, \cdots)\\
&+\frac{1}{2}\left(\frac{k}{\sqrt{\Omega}}\right)^2\frac{\partial^2}{\partial z_i^2}f(\cdots, \Omega\phi_i + \sqrt{\Omega}z_i, \cdots) + \cdots\\
= &\left[1 + \frac{k}{\sqrt{\Omega}}\frac{\partial}{\partial z_i} + \frac{k^2}{2\Omega}\frac{\partial^2}{\partial z_i^2} + \cdots\right]f(\cdots, n_i, \cdots).
\end{aligned} \tag{F.23}$$

Therefore, formally we can write

$$E_i^k = \left[1 + \frac{k}{\sqrt{\Omega}}\frac{\partial}{\partial z_i} + \frac{k^2}{2\Omega}\frac{\partial^2}{\partial z_i^2} + O(\Omega^{-3/2})\right] \tag{F.24}$$

Expansion of the reaction rate

Finally we need to make some assumption of the reaction rate.

$$\begin{aligned}
\tilde{\nu}_j\left(\frac{\vec{n}}{\Omega},\Omega\right) &= \tilde{\nu}_j\left(\vec{\phi} + \frac{\vec{z}}{\sqrt{\Omega}},\Omega\right)\\
&= \tilde{\nu}_j\left(\vec{\phi},\Omega\right) + \sum_{i=1}^{M}\frac{\partial}{\partial\phi_i}\tilde{\nu}_j(\vec{\phi},\Omega)\frac{z_i}{\sqrt{\Omega}} + \cdots
\end{aligned} \tag{F.25}$$

Here we further assume that

$$\tilde{\nu}_j\left(\vec{\phi},\Omega\right) = \nu_j(\vec{\phi}) + O(1/\Omega), \tag{F.26}$$

which is consistent with (F.6) $\nu_j(\vec{\phi}) = \lim_{\Omega\to\infty}\tilde{\nu}_j\left(\frac{\vec{n}}{\Omega},\Omega\right)$. and can be confirmed for some examples, .e.g. in the eq.(**??**). In the end we have

$$\tilde{\nu}_j\left(\frac{\vec{n}}{\Omega},\Omega\right) = \nu_j(\vec{\phi}) + \frac{1}{\sqrt{\Omega}}\sum_{i=1}^{M}\frac{\partial}{\partial\phi_i}\nu_j(\vec{\phi})z_i + O(1/\Omega). \tag{F.27}$$

Expansion of the master equation

By substituting (F.22), (F.24), and (F.27) into (F.18), we get ($\Omega^{M/2}$ multiplied for both side)

$$\begin{aligned}
&\left[\frac{\partial}{\partial t}Q(\vec{z},t) - \sum_{i=1}^{M}\sqrt{\Omega}\frac{d\phi_i(t)}{dt}\frac{\partial}{\partial z_i}Q(\vec{z},t)\right] \\
= \;& \Omega\sum_{j=1}^{R}\left[\left(\prod_{i=1}^{M}\left[1+\frac{-S_{ij}}{\sqrt{\Omega}}\frac{\partial}{\partial z_i}+\frac{S_{ij}^2}{2\Omega}\frac{\partial^2}{\partial z_i^2}+O(\Omega^{-3/2})\right]\right)-1\right] \\
& \left(\nu_j(\vec{\phi})+\frac{1}{\sqrt{\Omega}}\sum_{i=1}^{M}\frac{\partial}{\partial\phi_i}\nu_j(\vec{\phi})z_i+O(1/\Omega)\right)Q(\vec{z},t). \qquad \text{(F.28)}
\end{aligned}$$

In order that this equation makes physical sense in $\Omega\to\infty$, each order of $\sqrt{\Omega}$ should balance.

F.3 Linear noise approximation

Preparation

It is useful to note

$$\begin{aligned}
&\left(\prod_{i=1}^{M}\left[1+\frac{-S_{ij}}{\sqrt{\Omega}}\frac{\partial}{\partial z_i}+\frac{S_{ij}^2}{2\Omega}\frac{\partial^2}{\partial z_i^2}+O(\Omega^{-3/2})\right]\right) \\
= \;& \left[1+\frac{-S_{1j}}{\sqrt{\Omega}}\frac{\partial}{\partial z_1}+\frac{S_{1j}^2}{2\Omega}\frac{\partial^2}{\partial z_1^2}+O(\Omega^{-3/2})\right]\times\cdots \\
& \times\left[1+\frac{-S_{Mj}}{\sqrt{\Omega}}\frac{\partial}{\partial z_M}+\frac{S_{Mj}^2}{2\Omega}\frac{\partial^2}{\partial z_M^2}+O(\Omega^{-3/2})\right] \\
= \;& 1-\frac{1}{\sqrt{\Omega}}\sum_{i=1}^{M}S_{ij}\frac{\partial}{\partial z_i}+\frac{1}{2\Omega}\sum_{i=1}^{M}\sum_{i'=1}^{M}S_{ij}S_{i'j}\frac{\partial^2}{\partial z_i\partial z_{i'}}+O(\Omega^{-3/2}) \qquad \text{(F.29)}
\end{aligned}$$

Substituting this to eq. (F.28), we get

$$\begin{aligned}
&\left[\frac{\partial}{\partial t}Q(\vec{z},t) - \sum_{i=1}^{M}\sqrt{\Omega}\frac{d\phi_i(t)}{dt}\frac{\partial}{\partial z_i}Q(\vec{z},t)\right] \\
= \;& \Omega\sum_{j=1}^{R}\left[-\frac{1}{\sqrt{\Omega}}\sum_{i=1}^{M}S_{ij}\frac{\partial}{\partial z_i}+\frac{1}{2\Omega}\sum_{i=1}^{M}\sum_{i'=1}^{M}S_{ij}S_{i'j}\frac{\partial^2}{\partial z_i\partial z_{i'}}+O(\Omega^{-3/2})\right] \\
& \left(\nu_j(\vec{\phi})+\frac{1}{\sqrt{\Omega}}\sum_{i=1}^{M}\frac{\partial}{\partial\phi_i}\nu_j(\vec{\phi})z_i+O(1/\Omega)\right)Q(\vec{z},t) \\
= \;& \sum_{j=1}^{R}\left[-\sqrt{\Omega}\sum_{i=1}^{M}S_{ij}\frac{\partial}{\partial z_i}+\frac{1}{2}\sum_{i=1}^{M}\sum_{i'=1}^{M}S_{ij}S_{i'j}\frac{\partial^2}{\partial z_i\partial z_{i'}}+O(\Omega^{-1/2})\right] \\
& \left(\nu_j(\vec{\phi})+\frac{1}{\sqrt{\Omega}}\sum_{i=1}^{M}\frac{\partial}{\partial\phi_i}\nu_j(\vec{\phi})z_i+O(1/\Omega)\right)Q(\vec{z},t) \qquad \text{(F.30)}
\end{aligned}$$

Equation of the largest order

The largest term is proportional to $\sqrt{\Omega}$, which gives

$$\begin{aligned} -\sum_{i=1}^{M} \frac{d\phi_i(t)}{dt} \frac{\partial}{\partial z_i} Q(\vec{z},t) &= -\sum_{j=1}^{R}\sum_{i=1}^{M} S_{ij} \frac{\partial}{\partial z_i} \nu_j(\vec{\phi}) Q(\vec{z},t) \\ \Leftrightarrow \quad \sum_{i=1}^{M} \left[\frac{d\phi_i(t)}{dt} - \sum_{j=1}^{R} S_{ij}\nu_j(\vec{\phi}) \right] \frac{\partial}{\partial z_i} Q(\vec{z},t) &= 0. \end{aligned} \tag{F.31}$$

This gives the deterministic rate equation

$$\frac{d\phi_i(t)}{dt} = \sum_{j=1}^{R} S_{ij}\nu_j(\vec{\phi}), \tag{F.32}$$

which is exactly as expected, see (F.7).

Equation of the next largest order

This is order $\Omega^{-1/2}$ relative to the previous deterministic order, so it is the *linear order in noise* (see (F.17)). Approximation up to this order is called the *linear noise approximation*.

$$\begin{aligned} \frac{\partial}{\partial t} Q(z,t) &= \sum_{j=1}^{R} \left(-\sum_{i=1}^{M} S_{ij} \frac{\partial}{\partial z_i} \right) \left(\sum_{i'=1}^{M} \frac{\partial}{\partial \phi_{i'}} \nu_j(\vec{\phi}) z_{i'} \right) Q(\vec{z},t) \\ &\quad + \sum_{j=1}^{R} \left(\frac{1}{2} \sum_{i=1}^{M}\sum_{i'=1}^{M} S_{ij}S_{i'j} \frac{\partial^2}{\partial z_i \partial z_{i'}} \right) \nu_j(\vec{\phi}) Q(\vec{z},t) \\ &= -\sum_{i=1}^{M}\sum_{i'=1}^{M} \left[\frac{\partial}{\partial \phi_{i'}} \left(\sum_{j=1}^{R} S_{ij}\nu_j(\vec{\phi}) \right) \right] \frac{\partial}{\partial z_i} \left(z_{i'} Q(\vec{z},t) \right) \\ &\quad + \frac{1}{2} \sum_{i=1}^{M}\sum_{i'=1}^{M} \left[\sum_{j=1}^{R} S_{ij}\nu_j(\vec{\phi}) S_{i'j} \right] \frac{\partial^2}{\partial z_i \partial z_{i'}} Q(\vec{z},t). \end{aligned} \tag{F.33}$$

Jacobian and diffusion matrix

For the convenience, we define the following matrices

$$J_{ij}(\vec{\phi}) \equiv \frac{\partial}{\partial \phi_j} \left(\sum_{k=1}^{R} S_{ik}\nu_k(\vec{\phi}) \right) = \frac{\partial}{\partial \phi_j} \left(\frac{d\phi_i}{dt} \right) \tag{F.34}$$

and

$$D_{ij}(\vec{\phi}) \equiv \sum_{k=1}^{R} S_{ik}\nu_k(\vec{\phi}) S_{jk} = \left(S \cdot \mathrm{diag}[\vec{\nu}] \cdot S^T \right)_{ij} \tag{F.35}$$

Here, $\mathrm{diag}[\vec{\nu}]$ is $R \times R$ diagonal matrix with ii component to be ν_i and the rest of components are zero. Note that J is the Jacobian of the rate equation. D is called the diffusion coefficient matrix. Then

$$\frac{\partial}{\partial t} Q(\vec{z},t) = -\sum_{i,j=1}^{M} J_{ij}(\vec{\phi}) \frac{\partial}{\partial z_i} \left(z_j Q(\vec{z},t) \right)) + \frac{1}{2} \sum_{i,j=1}^{M} D_{ij}(\vec{\phi}) \frac{\partial^2}{\partial z_i \partial z_j} Q(\vec{z},t)). \tag{F.36}$$

This have the form of M-dimensional Fokker-Planck equation.

F.3.1 Moment, co-variance, and distribution

The first moment

It is possible to obtain time evolution equation for

$$\langle \vec{Z}(t) \rangle = \int_{-\infty}^{\infty} \vec{z} Q(\vec{z},t) d\vec{z} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \vec{z} Q(\vec{z},t) dz_1 dz_2 \cdots dz_M \tag{F.37}$$

by multiplying (F.36) with z_k and integrating over $\vec{z}$. Note (integration by part)

$$\begin{aligned}
&\frac{\partial}{\partial t} \int_{-\infty}^{\infty} z_k Q(\vec{z},t) d\vec{z} = \frac{d}{dt} \langle Z_k(t) \rangle. \\
&\int_{-\infty}^{\infty} z_k \frac{\partial}{\partial z_i} \left(z_j Q(\vec{z},t) \right) \right) dz_i = \left[z_k z_j Q(\vec{z},t) \right]_{-\infty}^{\infty} - \delta_{i,k} \int_{-\infty}^{\infty} z_j Q(\vec{z},t)) dz_i \\
&= -\delta_{i,k} \int_{-\infty}^{\infty} z_j Q(\vec{z},t)) dz_i, \\
&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_k \frac{\partial^2}{\partial z_i \partial z_j} Q(\vec{z},t) dz_i dz_j \\
&= \left[z_k \frac{\partial}{\partial z_j} Q(\vec{z},t) \right]_{-\infty}^{\infty} - \delta_{i,k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial}{\partial z_j} Q(\vec{z},t)) dz_i dz_j = 0
\end{aligned} \tag{F.38}$$

Thus we get

$$\begin{aligned}
\frac{d}{dt} \langle Z_k(t) \rangle &= -\sum_{i,j=1}^{M} J_{ij}(\vec{\phi}) \left(-\delta_{i,k} \right) \int_{-\infty}^{\infty} z_j Q(\vec{z},t) d\vec{z} \\
&= \sum_{j=1}^{M} J_{kj}(\vec{\phi}) \langle Z_j(t) \rangle.
\end{aligned} \tag{F.39}$$

Equivalently, the matrix representation is

$$\frac{d}{dt} \langle \vec{Z}(t) \rangle = J(\vec{\phi}) \cdot \langle \vec{Z}(t) \rangle. \tag{F.40}$$

Therefore, if we set $\langle \vec{Z}(0) \rangle = 0$ (i.e., the initial number of the particles coincides with the density, $\langle N(0) \rangle = \Omega \phi(0)$), we always have $\langle \vec{Z}(t) \rangle = 0$. We can do this without loss of generality.

Co-variance

Now let us consider *co-variance*

$$\Xi_{kl} \equiv \langle Z_k Z_l \rangle - \langle Z_k \rangle \langle Z_l \rangle = \langle Z_k Z_l \rangle. \tag{F.41}$$

Similar to the first moment, we can derive equation for the co-variance by multiplying (F.36) with $z_k z_l$ and integrate over $\vec{z}$. Note

$$\begin{aligned}
&\int_{-\infty}^{\infty} z_k z_l \frac{\partial}{\partial z_i} \left(z_j Q(\vec{z},t) \right) \right) dz_i = -\int_{-\infty}^{\infty} \left(\delta_{k,i} z_l + z_k \delta_{l,i} \right) \left(z_j Q(\vec{z},t) \right) \right) dz_i \\
&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_k z_l \frac{\partial^2}{\partial z_i \partial z_j} Q(\vec{z},t) dz_i dz_j \\
&= -\int_{-\infty}^{\infty} (\delta_{i,k} z_l + \delta_{i,l} z_k) \frac{\partial}{\partial z_j} Q(\vec{z},t) dz_i dz_j \\
&= (\delta_{i,k} \delta_{j,l} + \delta_{i,l} \delta_{k,j}) \int_{-\infty}^{\infty} Q(\vec{z},t) dz_i dz_j,
\end{aligned} \tag{F.42}$$

We get

$$\begin{aligned}\frac{d}{dt}\Xi_{kl} &= \sum_{i,j=1}^{M} J_{ij}(\vec{\phi})(\delta_{k,i}\Xi_{jl} + \delta_{l,i}\Xi_{jk}) + \frac{1}{2}\sum_{i,j=1}^{M} D_{ij}(\vec{\phi})(\delta_{i,k}\delta_{j,l} + \delta_{i,l}\delta_{k,j}) \\ &= \sum_{j=1}^{M}(\Xi_{jl}J_{kj} + \Xi_{jk}J_{lj}) + \frac{1}{2}(D_{kl} + D_{lk}). \end{aligned} \tag{F.43}$$

Noting $\Xi_{ij} = \Xi_{ji}$ and $D_{ij} = D_{ji}$, the matrix representation becomes

$$\frac{d}{dt}\Xi = J \cdot \Xi + (J \cdot \Xi)^T + D. \tag{F.44}$$

Here T represent transposed matrix. Noting $(J \cdot \Xi)^T_{kl} = \sum_j J_{lj}\Xi_{jk} = \sum_j J_{lj}\Xi_{kj} = \sum_j \Xi_{kj}J^T_{jl} = \Xi \cdot J^T_{kl}$, we can also write

$$\frac{d}{dt}\Xi = J \cdot \Xi + \Xi \cdot J^T + D. \tag{F.45}$$

Distribution

The solution of (F.36) is Gaussian, characterized by Ξ as

$$Q(\vec{z}, t) = \frac{1}{\sqrt{(2\pi)^N \det \Xi(t)}} \exp\left[-\frac{1}{2}\vec{z}^T \cdot \Xi(t)^{-1} \cdot \vec{z}\right]. \tag{F.46}$$

Steady state

Especially, in the steady state, the density will take a constant value $\vec{\phi}_s$, and so the $J_s = J(\vec{\phi}_s)$ and $D_s = D(\vec{\phi}_s)$. The steady state value of the co-variance matrix Ξ_s is given by

$$J_s \cdot \Xi_s + \Xi_s \cdot J_s^T + D_s = 0. \tag{F.47}$$

F.4 Example: Protein production

We again consider the biochemical reactions in a cell, where the information on DNA is copied to mRNA, and it is translated to proteins.

- The cell volume is Ω.
- mRNA is made at the rate $\alpha\Omega$. (reaction 1)
- mRNA is degraded at the rate Γ_1 per mRNA molecule. (reaction 2)
- Protein is made at the rate β per mRNA molecule. (reaction 3)
- Protein is degraded at the rate Γ_2 per protein molecule. (reaction 4)

Define the number of mRNA molecule as N_1 and the number of protein molecule as N_2 ($M = 2$). $R = 4$ reactions can happen as above.

Stoichiometry matrix and rate vector

Since the reaction rates are constant or linear in number of particles, this problem is actually exactly solvable - for example the second moment of protein number can be calculated by using generating function method. However the calculation is a bit long - so here we apply LNA and estimate how protein number fluctuate depending on the parameters.

We identified that

$$S = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \tag{F.48}$$

and

$$\vec{\nu}(\vec{\phi}) = \begin{pmatrix} \alpha \\ \Gamma_1\phi_1 \\ \beta\phi_1 \\ \Gamma_2\phi_2 \end{pmatrix}. \tag{F.49}$$

Rate equation and Jacobian

The deterministic rate equation is given by

$$\frac{d}{dt}\vec{\phi} = S\vec{\nu} \Leftrightarrow \begin{pmatrix} \dot{\phi}_1 \\ \dot{\phi}_2 \end{pmatrix} = \begin{pmatrix} \alpha - \Gamma_1\phi_1 \\ \beta\phi_1 - \Gamma_2\phi_2 \end{pmatrix} \tag{F.50}$$

From the definition,

$$J = \begin{pmatrix} \frac{\partial}{\partial\phi_1}(\alpha - \Gamma_1\phi_1) & \frac{\partial}{\partial\phi_2}(\alpha - \Gamma_1\phi_1) \\ \frac{\partial}{\partial\phi_1}(\beta\phi_1 - \Gamma_2\phi_2) & \frac{\partial}{\partial\phi_2}(\beta\phi_1 - \Gamma_2\phi_2) \end{pmatrix} = \begin{pmatrix} -\Gamma_1 & 0 \\ \beta & -\Gamma_2 \end{pmatrix}$$

Diffusion matrix

$$\begin{aligned} D &= \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha & 0 & 0 & 0 \\ 0 & \Gamma_1\phi_1 & 0 & 0 \\ 0 & 0 & \beta\phi_1 & 0 \\ 0 & 0 & 0 & \Gamma_2\phi_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ -\Gamma_1\phi_1 & 0 \\ 0 & \beta\phi_1 \\ 0 & -\Gamma_2\phi_2 \end{pmatrix} \qquad \text{(F.51)} \\ &= \begin{pmatrix} \alpha + \Gamma_1\phi_1 & 0 \\ 0 & \beta\phi_1 + \Gamma_2\phi_2 \end{pmatrix}. \qquad \text{(F.52)} \end{aligned}$$

Steady state fluctuation

Now let us calculate the co-variance matrix in the steady state. Using $\phi_{1,s} = \alpha/\Gamma_1$,$\phi_{2,s} = \beta\phi_1, s/\Gamma_2 = (\alpha\beta)/(\Gamma_1\Gamma_2)$, we have

$$J_s = \begin{pmatrix} -\Gamma_1 & 0 \\ \beta & -\Gamma_2 \end{pmatrix}, \quad D_s = \begin{pmatrix} 2\alpha & 0 \\ 0 & 2\alpha\beta/\Gamma_1 \end{pmatrix}. \tag{F.53}$$

Therefore, the steady state gives

$$\begin{aligned} 0 &= \begin{pmatrix} -\Gamma_1\Xi_{11} & -\Gamma_1\Xi_{12} \\ \beta\Xi_{11} - \Gamma_2\Xi_{21} & \beta\Xi_{12} - \Gamma_2\Xi_{22} \end{pmatrix} + \begin{pmatrix} -\Gamma_1\Xi_{11} & \beta\Xi_{11} - \Gamma_2\Xi_{12} \\ -\Gamma_1\Xi_{21} & \beta\Xi_{21} - \Gamma_2\Xi_{22} \end{pmatrix} \\ &\quad + \begin{pmatrix} 2\alpha & 0 \\ 0 & 2\alpha\beta/\Gamma_1 \end{pmatrix} \qquad \text{(F.54)} \\ &= \begin{pmatrix} -2\alpha - 2\Gamma_1\Xi_{11} & \beta\Xi_{11} - (\Gamma_1 + \Gamma_2)\Xi_{12} \\ \beta\Xi_{11} - (\Gamma_1 + \Gamma_2)\Xi_{12} & 2\beta\Xi_{12} - 2\Gamma_2\Xi_{22} + 2\alpha\beta/\Gamma_1 \end{pmatrix} \qquad \text{(F.55)} \end{aligned}$$

where $\Xi_{12} = \Xi_{21}$ was used.

By solving these equations, we get

$$\Xi_{11} = \langle Z_1^2 \rangle = \frac{\alpha}{\Gamma_1}, \tag{F.56}$$

$$\Xi_{12} = \langle Z_1 Z_2 \rangle = \frac{\beta\alpha}{\Gamma_1(\Gamma_1+\Gamma_2)}, \tag{F.57}$$

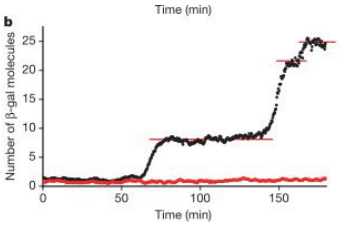
$$\begin{aligned}\Xi_{22} &= \langle Z_2 \rangle = \frac{1}{\Gamma_2}\left(\frac{\beta^2\alpha}{\Gamma_1(\Gamma_1+\Gamma_2)} + \frac{\alpha\beta}{\Gamma_1}\right) \\ &= \frac{\alpha\beta}{\Gamma_1\Gamma_2}\left(1+\frac{\beta}{\Gamma_1+\Gamma_2}\right) \\ &= \phi_{2,s}\left(1+\left(\frac{\beta}{\Gamma_2}\right)\frac{1}{1+\Gamma_1/\Gamma_2}\right).\end{aligned} \tag{F.58}$$

Protein burst

The fractional deviation of the protein number fluctuation (squared):

$$\left(\frac{\sqrt{\langle \Omega Z_2 \rangle}}{\Omega\phi_{2,s}}\right)^2 = \frac{\langle (N_2 - \langle N_2 \rangle)^2 \rangle}{\langle N_2 \rangle^2} = \frac{1}{\langle N_2 \rangle}\left(1+\left(\frac{\beta}{\Gamma_1}\right)\frac{1}{1+\Gamma_2/\Gamma_1}\right). \tag{F.59}$$

Note that β/Γ_1 tells how much protein is made per mRNA ($1/\Gamma_1$ is mRNA life time). When the mRNA is short-lived compared to the protein ($\Gamma_1 \gg \Gamma_2$), this characterize how "busty" protein production is - once a mRNA is made, the number of protein jumps by β/Γ_1, and it determines the fluctuation in the number of proteins. Figure shows the experimentally observed step-wise increase of proteins in E. coli due to the bursty production.



(protein vs. time plotted. from Cai et al. Nature 2006.)

It is actually possible to apply the generating function method to the original master equation exactly and calculate the variance. LNA agrees with the exact solution for this problem.

Appendix G

Burgers equation from the mean field equation of TASEP

G.1 Derivation of Burgers equation

When we make further approximation of *continuous space* for the mean field equation of the density n_i, eq. (10.5), we can derive the *inviscid Burgers equation* as follows:

We take the lattice spacing to be Δx, and express position x for the site i as $x = i\Delta x$. We write the occupancy in the continuous space as $\rho(x,t) = n_i(t)$. We then expand

$$n_{i\pm1}(t) = \rho(x \pm \Delta x, t) = \rho(x,t) + \frac{\partial}{\partial x}\rho(x,t)(\pm\Delta x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\rho(x,t)\Delta x^2 + \cdots$$

Inserting this to (10.5), we get

$$\frac{\partial}{\partial t}\rho(x,t) = -\left(\frac{\partial \rho(x,t)}{\partial x}\right)(1 - 2\rho(x,t))\,\Delta x + O(\Delta x^2). \tag{G.1}$$

We then resale the time t to s with $s = \Delta x t$, so that the particle's velocity (proportional to Δx) remains non-zero in the continuous limit $\Delta x \to 0$.

$$\frac{\partial}{\partial t}\rho = \frac{\partial}{\partial t}s\frac{\partial}{\partial s}\rho = \Delta x\frac{\partial}{\partial s}\rho. \tag{G.2}$$

In the limit of $\Delta x \to 0$, we get

$$\frac{\partial}{\partial s}\rho(x,s) = -\left(\frac{\partial \rho(x,s)}{\partial x}\right)(1 - 2\rho(x,s)) = -\frac{\partial}{\partial x}\left(\rho(x,s)(1-\rho(x,s))\right). \tag{G.3}$$

This is *the inviscid Burgers equation*.

The inviscid Bugers equation is known to have the following *shock solution*, where the density discontinuously jumps.

$$\rho(x,t) = \begin{cases} \rho_- & \text{for } x < x_0(s), \\ \rho_+ & \text{for } x > x_0(s), \end{cases} \tag{G.4}$$

where $0 < \rho_- < \rho_+ < 1$ and the boundary $x_0(s)$ moves to the right with a constant velocity

$$v_s = 1 - \rho_+ - \rho_-. \tag{G.5}$$

G.2 Shock solution

It is easy to show that the velocity of the shock front is given by (G.5) by considering the conservation of mass.

Hint: Consider a small time step Δt. The net transport of the particle at the position of the shock is by $(\rho_+ - \rho_-)v_s\Delta t$. This should match with the amount moved to the right, i.e. (out-flux to the right−influx from the left).

This shock can be related to a "traffic jam", which also shows a sudden change of the density in space. It is also known that the solution of the inviscid Burgers equation often develops a shock and diverges within finite time. For description of the physical system, one should take into account a small diffusion term

$$\frac{\partial}{\partial s}\rho(x,s) + \frac{\partial}{\partial x}\left(\rho(x,s)(1-\rho(x,s))\right) = D_B\frac{\partial^2}{\partial x^2}\rho(x,s) \tag{G.6}$$

This is the simplest hydrodynamics description of traffic jam.

G.3 TASEP to the KPZ equation

TASEP can be mapped to different problems, too, and therefore it has been studied due to various physical motivations. For example, we can map the configuration of the particle to a shape of a surface, as Fig. G.1:

- A site with a particle → slope goes down.
- An empty site → slope goes up.

Therefore, if the slope is always constant (+1 or -1), a surface growth problem can be mapped to TASEP.

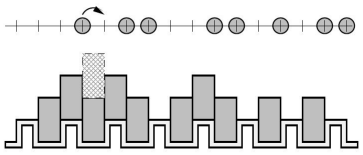


Fig. 2. The single-step growth model and its mapping onto the asymmetric exclusion process with $q = 0$. Note that the addition of a new particle to the surface (shown light-shaded) corresponds to the rightward hop of a particle in the ASEP.

Figure G.1: Mapping TASEP to a surface growth model.

In continuous space, this corresponds to relate the integral of density ρ to the height of a surface h, i.e., $h(x,s) = \int^x (1-\rho(x',s))dx'$. By performing this to the inviscid Burger's equation, we get

$$\frac{\partial}{\partial s}h(x,s) + \left(\frac{\partial h(x,s)}{\partial x}\right)^2 - \frac{\partial}{\partial x}h(x,s) = 0. \tag{G.7}$$

The last term can be removed by using a moving frame. When a diffusion term added to prevent this equation to diverge, and also a noise term is added to this equation to describe randomness in the growth process, the equation is called the KPZ (Karder-Parisi-Zhang) equation. The KPZ equation is a famous equation to describe a wide class of interface growth.

Bibliography

[1] N. Hashitsume, M. Toda, R. Kubo, and N. Saitō, *Statistical physics II: nonequilibrium statistical mechanics*, vol. 30. Springer, 1992.

[2] N. G. Van Kampen, *Stochastic processes in physics and chemistry, 2nd edition*. North Holland, 1992.

[3] P. Langevin, "Sur la théorie du mouvement brownien [on the theory of brownian motion]," *CR Acad. Sci. (Paris)*, vol. 146, pp. 530–533, 1908.

[4] K. Jacobs, *Stochastic processes for physicists: understanding noisy systems*. Cambridge University Press Cambridge, 2010.

[5] H. Risken, *Fokker-Planck Equation*. Springer, 1984.

[6] K. Sneppen and G. Zocchi, *Physics in molecular biology*. Cambridge University Press, 2005.

[7] K. Sneppen, *Models of Life*. Cambridge University Press, 2014.

[8] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: fifty years after kramers," *Reviews of Modern Physics*, vol. 62, no. 2, p. 251, 1990.

[9] M. Scott, "Stochastic processes in physics and engineering.." Lecture notes, avilable from http://www.math.uwaterloo.ca/ mscott/msteaching.html.

[10] C. W. Gardiner *et al.*, *Handbook of stochastic methods*, vol. 3. Springer Berlin, 1985.

[11] P. Reimann, "Brownian motors: noisy transport far from equilibrium," *Physics reports*, vol. 361, no. 2-4, pp. 57–265, 2002.

[12] V. Blickle, T. Speck, L. Helden, U. Seifert, and C. Bechinger, "Thermodynamics of a colloidal particle in a time-dependent non-harmonic potential," *Physical Review Letters*, vol. 96, p. 070603, 2006.

[13] K. Sekimoto, *Stochastic energetics*, vol. 799. Springer, 2010.

[14] D. Williams, *Probability with martingales*. Cambridge university press, 1991.

[15] J. Kelly Jr, "A new interpretation of the information rate, bell systems tech," *J*, vol. 35, pp. 917–926, 1956.

[16] O. Peters, "Optimal leverage from non-ergodicity," *Quantitative Finance*, vol. 11, no. 11, pp. 1593–1602, 2011.

[17] O. Peters and W. Klein, "Ergodicity breaking in geometric brownian motion," *Physical review letters*, vol. 110, no. 10, p. 100603, 2013.

[18] S. Maslov and Y.-C. Zhang, "Optimal investment strategy for risky assets," *International Journal of Theoretical and Applied Finance*, vol. 1, no. 03, pp. 377–387, 1998.

[19] J. Garcia-Palacios, "Introduction to the theory of stochastic processes and brownian motion problems," *arXiv preprint cond-mat/0701242*, 2007.

[20] S. Horikawa, T. Isoda, T. Nakayama, A. Nakahara, and M. Matsushita, "Self-organized critical density waves of granular particles flowing through a pipe," *Physica A: Statistical Mechanics and its Applications*, vol. 233, no. 3, pp. 699–708, 1996.

[21] N. Mitarai, K. Sneppen, and S. Pedersen, "Ribosome collisions and translation efficiency: optimization by codon usage and mrna destabilization," *Journal of molecular biology*, vol. 382, no. 1, pp. 236–245, 2008.

[22] M. Evans and R. Blythe, "Nonequilibrium dynamics in low-dimensional systems," *Physica A: Statistical Mechanics and its Applications*, vol. 313, no. 1, pp. 110–152, 2002.

[23] A. Schadschneider, "Traffic flow: a statistical physics point of view," *Physica A: Statistical Mechanics and its Applications*, vol. 313, no. 1, pp. 153–187, 2002.

[24] A. J. McKane and T. J. Newman, "Predator-prey cycles from resonant amplification of demographic stochasticity," *Physical review letters*, vol. 94, no. 21, p. 218102, 2005.