Lecture Notes on Nonequilibrium Statistical Physics
(A Work in Progress)

Daniel Arovas
Department of Physics
University of California, San Diego

September 26, 2018

Contents

Chapter 1

Fundamentals of Probability

1.1 References

- C. Gardiner, *Stochastic Methods* (4th edition, Springer-Verlag, 2010)
  Very clear and complete text on stochastic methods with many applications.

- J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (Wiley, 2000)
  A thorough textbook on Bayesian methods.

- D. Williams, *Weighing the Odds: A Course in Probability and Statistics* (Cambridge, 2001)
  A good overall statistics textbook, according to a mathematician colleague.

- E. T. Jaynes, *Probability Theory* (Cambridge, 2007)
  An extensive, descriptive, and highly opinionated presentation, with a strongly Bayesian approach.

- A. N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea, 1956)
  The *Urtext* of mathematical probability theory.

1.2 Statistical Properties of Random Walks

1.2.1 One-dimensional random walk

Consider the mechanical system depicted in Fig. 1.1, a version of which is often sold in novelty shops. A ball is released from the top, which cascades consecutively through N levels. The details of each ball's motion are governed by Newton's laws of motion. However, to predict where any given ball will end up in the bottom row is difficult, because the ball's trajectory depends sensitively on its initial conditions, and may even be influenced by random vibrations of the entire apparatus. We therefore abandon all hope of integrating the equations of motion and treat the system statistically. That is, we assume, at each level, that the ball moves to the right with probability p and to the left with probability $q = 1 - p$. If there is no bias in the system, then $p = q = \frac{1}{2}$. The position X_N after N steps may be written

$$X = \sum_{j=1}^{N} \sigma_j \, , \tag{1.1}$$

where $\sigma_j = +1$ if the ball moves to the right at level j, and $\sigma_j = -1$ if the ball moves to the left at level j. At each level, the probability for these two outcomes is given by

$$P_\sigma = p \, \delta_{\sigma,+1} + q \, \delta_{\sigma,-1} = \begin{cases} p & \text{if } \sigma = +1 \\ q & \text{if } \sigma = -1 \, . \end{cases} \tag{1.2}$$

This is a normalized discrete probability distribution of the type discussed in section 1.5 below. The multivariate distribution for all the steps is then

$$P(\sigma_1 \, , \, \ldots \, , \sigma_N) = \prod_{j=1}^{N} P(\sigma_j) \, . \tag{1.3}$$

Our system is equivalent to a one-dimensional *random walk*. Imagine an inebriated pedestrian on a sidewalk taking steps to the right and left at random. After N steps, the pedestrian's location is X.

Now let's compute the average of X:

$$\langle X \rangle = \Big\langle \sum_{j=1}^{N} \sigma_j \Big\rangle = N \langle \sigma \rangle = N \sum_{\sigma = \pm 1} \sigma \, P(\sigma) = N(p - q) = N(2p - 1) \, . \tag{1.4}$$

This could be identified as an *equation of state* for our system, as it relates a measurable quantity X to the number of steps N and the local bias p. Next, let's compute the average of X^2:

$$\langle X^2 \rangle = \sum_{j=1}^{N} \sum_{j'=1}^{N} \langle \sigma_j \sigma_{j'} \rangle = N^2 (p - q)^2 + 4Npq \, . \tag{1.5}$$

Here we have used

$$\langle \sigma_j \sigma_{j'} \rangle = \delta_{jj'} + \big(1 - \delta_{jj'}\big)(p - q)^2 = \begin{cases} 1 & \text{if } j = j' \\ (p - q)^2 & \text{if } j \neq j' \, . \end{cases} \tag{1.6}$$
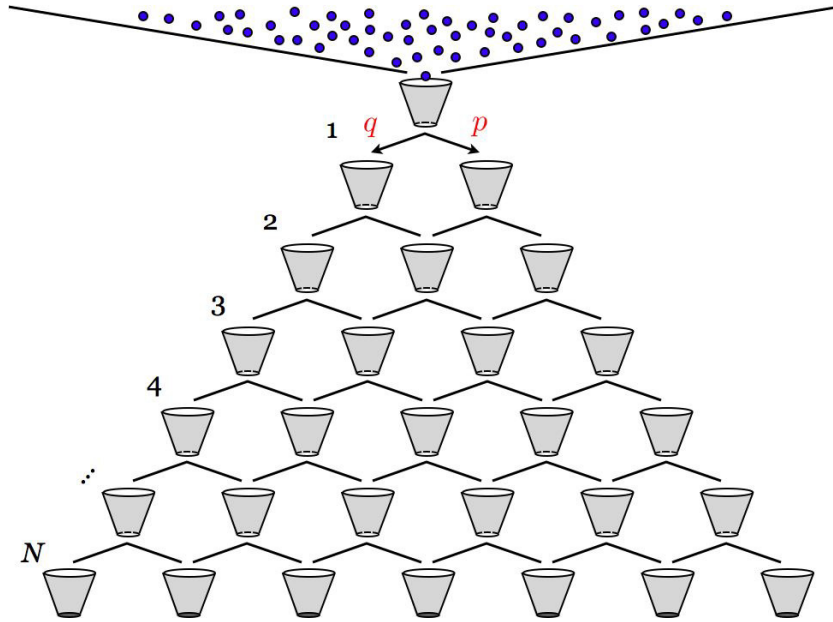
Figure 1.1: The falling ball system, which mimics a one-dimensional random walk.

Note that $\langle X^2 \rangle \geq \langle X \rangle^2$, which must be so because

$$\text{Var}(X) = \langle (\Delta X)^2 \rangle \equiv \left\langle \left(X - \langle X \rangle \right)^2 \right\rangle = \langle X^2 \rangle - \langle X \rangle^2 \ . \tag{1.7}$$

This is called the *variance* of X. We have $\text{Var}(X) = 4Npq$. The *root mean square* deviation, ΔX_{rms}, is the square root of the variance: $\Delta X_{\text{rms}} = \sqrt{\text{Var}(X)}$. Note that the mean value of X is linearly proportional to N^1, but the RMS fluctuations ΔX_{rms} are proportional to $N^{1/2}$. In the limit $N \to \infty$ then, the ratio $\Delta X_{\text{rms}}/\langle X \rangle$ vanishes as $N^{-1/2}$. This is a consequence of the central limit theorem (see §1.5.2 below), and we shall meet up with it again on several occasions.

We can do even better. We can find the complete probability distribution for X. It is given by

$$P_{N,X} = \binom{N}{N_{\text{R}}} p^{N_{\text{R}}} q^{N_{\text{L}}} \ , \tag{1.8}$$

where $N_{\text{R/L}}$ are the numbers of steps taken to the right/left, with $N = N_{\text{R}} + N_{\text{L}}$, and $X = N_{\text{R}} - N_{\text{L}}$. There are many independent ways to take N_{R} steps to the right. For example, our first N_{R} steps could all be to the right, and the remaining $N_{\text{L}} = N - N_{\text{R}}$ steps would then all be to the left. Or our final N_{R} steps could all be to the right. For each of these independent possibilities, the probability is $p^{N_{\text{R}}} q^{N_{\text{L}}}$. How many possibilities are there? Elementary combinatorics tells us this number is

$$\binom{N}{N_{\text{R}}} = \frac{N!}{N_{\text{R}}! \, N_{\text{L}}!} \ . \tag{1.9}$$

Note that $N \pm X = 2N_{\text{R/L}}$, so we can replace $N_{\text{R/L}} = \frac{1}{2}(N \pm X)$. Thus,

$$P_{N,X} = \frac{N!}{\left(\frac{N+X}{2} \right)! \left(\frac{N-X}{2} \right)!} \, p^{(N+X)/2} \, q^{(N-X)/2} \ . \tag{1.10}$$

---

[1]The exception is the unbiased case $p = q = \frac{1}{2}$, where $\langle X \rangle = 0$.
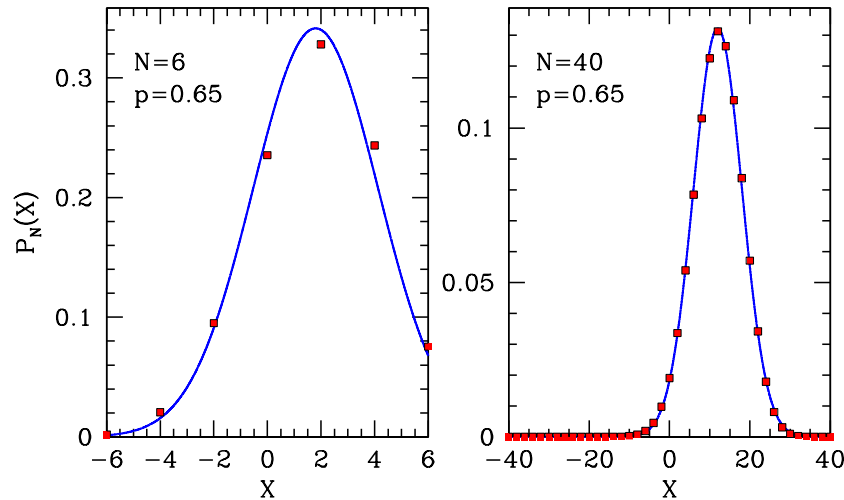
Figure 1.2: Comparison of exact distribution of eqn. 1.10 (red squares) with the Gaussian distribution of eqn. 1.19 (blue line).

1.2.2   Thermodynamic limit

Consider the limit $N \to \infty$ but with $x \equiv X/N$ finite. This is analogous to what is called the *thermodynamic limit* in statistical mechanics. Since N is large, x may be considered a continuous variable. We evaluate $\ln P_{N,X}$ using Stirling's asymptotic expansion

$$\ln N! \simeq N \ln N - N + \mathcal{O}(\ln N) \,. \tag{1.11}$$

We then have

$$\ln P_{N,X} \simeq N \ln N - N - \tfrac{1}{2}N(1+x)\ln\left[\tfrac{1}{2}N(1+x)\right] + \tfrac{1}{2}N(1+x)$$

$$- \tfrac{1}{2}N(1-x)\ln\left[\tfrac{1}{2}N(1-x)\right] + \tfrac{1}{2}N(1-x) + \tfrac{1}{2}N(1+x)\ln p + \tfrac{1}{2}N(1-x)\ln q \tag{1.12}$$

$$= -N\left[\left(\tfrac{1+x}{2}\right)\ln\left(\tfrac{1+x}{2}\right) + \left(\tfrac{1-x}{2}\right)\ln\left(\tfrac{1-x}{2}\right)\right] + N\left[\left(\tfrac{1+x}{2}\right)\ln p + \left(\tfrac{1-x}{2}\right)\ln q\right].$$

Notice that the terms proportional to $N \ln N$ have all cancelled, leaving us with a quantity which is linear in N. We may therefore write $\ln P_{N,X} = -Nf(x) + \mathcal{O}(\ln N)$, where

$$f(x) = \left[\left(\tfrac{1+x}{2}\right)\ln\left(\tfrac{1+x}{2}\right) + \left(\tfrac{1-x}{2}\right)\ln\left(\tfrac{1-x}{2}\right)\right] - \left[\left(\tfrac{1+x}{2}\right)\ln p + \left(\tfrac{1-x}{2}\right)\ln q\right]. \tag{1.13}$$

We have just shown that in the large N limit we may write

$$P_{N,X} = \mathcal{C}\, e^{-Nf(X/N)} \,, \tag{1.14}$$

where $\mathcal{C}$ is a normalization constant[2]. Since N is by assumption large, the function $P_{N,X}$ is dominated by the minimum (or minima) of $f(x)$, where the probability is maximized. To find the minimum of $f(x)$,

---

[2]The origin of $\mathcal{C}$ lies in the $\mathcal{O}(\ln N)$ and $\mathcal{O}(N^0)$ terms in the asymptotic expansion of $\ln N!$. We have ignored these terms here. Accounting for them carefully reproduces the correct value of $\mathcal{C}$ in eqn. 1.20.

we set $f'(x) = 0$, where

$$f'(x) = \tfrac{1}{2} \ln \left(\frac{q}{p} \cdot \frac{1+x}{1-x} \right) . \tag{1.15}$$

Setting $f'(x) = 0$, we obtain

$$\frac{1+x}{1-x} = \frac{p}{q} \quad \Rightarrow \quad \bar{x} = p - q . \tag{1.16}$$

We also have

$$f''(x) = \frac{1}{1 - x^2} , \tag{1.17}$$

so invoking Taylor's theorem,

$$f(x) = f(\bar{x}) + \tfrac{1}{2} f''(\bar{x}) \, (x - \bar{x})^2 + \ldots . \tag{1.18}$$

Putting it all together, we have

$$P_{N,X} \approx \mathcal{C} \, \exp \left[- \frac{N(x - \bar{x})^2}{8pq} \right] = \mathcal{C} \, \exp \left[- \frac{(X - \bar{X})^2}{8Npq} \right] , \tag{1.19}$$

where $\bar{X} = \langle X \rangle = N(p - q) = N\bar{x}$. The constant $\mathcal{C}$ is determined by the normalization condition,

$$\sum_{X=-\infty}^{\infty} P_{N,X} \approx \tfrac{1}{2} \int_{-\infty}^{\infty} dX \, \mathcal{C} \, \exp \left[- \frac{(X - \bar{X})^2}{8Npq} \right] = \sqrt{2\pi Npq} \, \mathcal{C} , \tag{1.20}$$

and thus $\mathcal{C} = 1/\sqrt{2\pi Npq}$. Why don't we go beyond second order in the Taylor expansion of $f(x)$? We will find out in §1.5.2 below.

1.2.3 Entropy and energy

The function $f(x)$ can be written as a sum of two contributions, $f(x) = e(x) - s(x)$, where

$$s(x) = -\left(\tfrac{1+x}{2}\right) \ln \left(\tfrac{1+x}{2}\right) - \left(\tfrac{1-x}{2}\right) \ln \left(\tfrac{1-x}{2}\right)$$
$$e(x) = -\tfrac{1}{2} \ln(pq) - \tfrac{1}{2} x \ln(p/q) . \tag{1.21}$$

The function $S(N, x) \equiv Ns(x)$ is analogous to the *statistical entropy* of our system[3]. We have

$$S(N, x) = Ns(x) = \ln \binom{N}{N_{\mathrm{R}}} = \ln \binom{N}{\tfrac{1}{2} N(1 + x)} . \tag{1.22}$$

Thus, *the statistical entropy is the logarithm of the number of ways the system can be configured so as to yield the same value of X (at fixed N)*. The second contribution to $f(x)$ is the energy term. We write

$$E(N, x) = Ne(x) = -\tfrac{1}{2} N \ln(pq) - \tfrac{1}{2} Nx \ln(p/q) . \tag{1.23}$$

---

[3]The function $s(x)$ is the *specific entropy*.

The energy term biases the probability $P_{N,X} = \exp(S - E)$ so that *low energy configurations are more probable than high energy configurations*. For our system, we see that when $p < q$ (*i.e.* $p < \frac{1}{2}$), the energy is minimized by taking x as small as possible (meaning as negative as possible). The smallest possible allowed value of $x = X/N$ is $x = -1$. Conversely, when $p > q$ (*i.e.* $p > \frac{1}{2}$), the energy is minimized by taking x as large as possible, which means $x = 1$. The average value of x, as we have computed explicitly, is $\bar{x} = p - q = 2p - 1$, which falls somewhere in between these two extremes.

In actual thermodynamic systems, entropy and energy are not dimensionless. What we have called S here is really S/k_{B}, which is the entropy in units of Boltzmann's constant. And what we have called E here is really $E/k_{\mathrm{B}}T$, which is energy in units of Boltzmann's constant times temperature.

1.3   Basic Concepts in Probability Theory

Here we recite the basics of probability theory.

1.3.1   Fundamental definitions

The natural mathematical setting is set theory. *Sets* are generalized collections of *objects*. The basics: $\omega \in A$ is a binary relation which says that the object ω is an *element* of the set A. Another binary relation is *set inclusion*. If all members of A are in B, we write $A \subseteq B$. The *union* of sets A and B is denoted $A \cup B$ and the *intersection* of A and B is denoted $A \cap B$. The *Cartesian product* of A and B, denoted $A \times B$, is the set of all ordered elements (a, b) where $a \in A$ and $b \in B$.

Some details: If ω is not in A, we write $\omega \notin A$. Sets may also be objects, so we may speak of sets of sets, but typically the sets which will concern us are simple discrete collections of numbers, such as the possible rolls of a die {1,2,3,4,5,6}, or the real numbers $\mathbb{R}$, or Cartesian products such as $\mathbb{R}^N$. If $A \subseteq B$ but $A \neq B$, we say that A is a *proper subset* of B and write $A \subset B$. Another binary operation is the *set difference* $A \backslash B$, which contains all ω such that $\omega \in A$ and $\omega \notin B$.

In probability theory, each object ω is identified as an *event*. We denote by Ω the set of all events, and $\emptyset$ denotes the set of no events. There are three basic axioms of probability:

 i) To each set A is associated a non-negative real number $P(A)$, which is called the probability of A.

 ii) $P(\Omega) = 1$.

 iii) If $\{A_i\}$ is a collection of disjoint sets, *i.e.* if $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i) \,. \tag{1.24}$$

From these axioms follow a number of conclusions. Among them, let $\neg A = \Omega \backslash A$ be the *complement* of A, *i.e.* the set of all events *not* in A. Then since $A \cup \neg A = \Omega$, we have $P(\neg A) = 1 - P(A)$. Taking $A = \Omega$, we conclude $P(\emptyset) = 0$.

The meaning of $P(A)$ is that if events ω are chosen from Ω *at random*, then the relative frequency for $\omega \in A$ approaches $P(A)$ as the number of trials tends to infinity. But what do we mean by 'at random'? One meaning we can impart to the notion of randomness is that a process is random if its outcomes can be accurately modeled using the axioms of probability. This entails the identification of a *probability space* Ω as well as a *probability measure* P. For example, in the microcanonical ensemble of classical statistical physics, the space Ω is the collection of phase space points $\varphi = \{q_1, \ldots, q_n, p_1, \ldots, p_n\}$ and the probability measure is $d\mu = \Sigma^{-1}(E) \prod_{i=1}^{n} dq_i \, dp_i \, \delta(E - H(q, p))$, so that for $A \in \Omega$ the probability of A is $P(A) = \int d\mu \, \chi_A(\varphi)$, where $\chi_A(\varphi) = 1$ if $\varphi \in A$ and $\chi_A(\varphi) = 0$ if $\varphi \notin A$ is the *characteristic function* of A. The quantity $\Sigma(E)$ is determined by normalization: $\int d\mu = 1$.

1.3.2 Bayesian statistics

We now introduce two additional probabilities. The *joint probability* for sets A and B together is written $P(A \cap B)$. That is, $P(A \cap B) = \text{Prob}[\omega \in A \text{ and } \omega \in B]$. For example, A might denote the set of all politicians, B the set of all American citizens, and C the set of all living humans with an IQ greater than 60. Then $A \cap B$ would be the set of all politicians who are also American citizens, *etc. Exercise: estimate* $P(A \cap B \cap C)$.

The *conditional probability* of B given A is written $P(B|A)$. We can compute the joint probability $P(A \cap B) = P(B \cap A)$ in two ways:

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) . \tag{1.25}$$

Thus,

$$P(A|B) = \frac{P(B|A) \, P(A)}{P(B)} , \tag{1.26}$$

a result known as *Bayes' theorem*. Now suppose the 'event space' is partitioned as $\{A_i\}$. Then

$$P(B) = \sum_i P(B|A_i) \, P(A_i) . \tag{1.27}$$

We then have

$$P(A_i|B) = \frac{P(B|A_i) \, P(A_i)}{\sum_j P(B|A_j) \, P(A_j)} , \tag{1.28}$$

a result sometimes known as the *extended form of Bayes' theorem*. When the event space is a 'binary partition' $\{A, \neg A\}$, we have

$$P(A|B) = \frac{P(B|A) \, P(A)}{P(B|A) \, P(A) + P(B|\neg A) \, P(\neg A)} . \tag{1.29}$$

Note that $P(A|B) + P(\neg A|B) = 1$ (which follows from $\neg \neg A = A$).

As an example, consider the following problem in epidemiology. Suppose there is a rare but highly contagious disease A which occurs in 0.01% of the general population. Suppose further that there is a simple test for the disease which is accurate 99.99% of the time. That is, out of every 10,000 tests, the correct answer is returned 9,999 times, and the incorrect answer is returned only once. Now let us

administer the test to a large group of people from the general population. Those who test positive are quarantined. Question: what is the probability that someone chosen at random from the quarantine group actually has the disease? We use Bayes' theorem with the binary partition $\{A, \neg A\}$. Let B denote the event that an individual tests positive. Anyone from the quarantine group has tested positive. Given this datum, we want to know the probability that that person has the disease. That is, we want $P(A|B)$. Applying eqn. 1.29 with

$$P(A) = 0.0001 \quad , \quad P(\neg A) = 0.9999 \quad , \quad P(B|A) = 0.9999 \quad , \quad P(B|\neg A) = 0.0001 \; ,$$

we find $P(A|B) = \frac{1}{2}$. That is, there is only a 50% chance that someone who tested positive actually has the disease, despite the test being 99.99% accurate! The reason is that, given the rarity of the disease in the general population, the number of false positives is statistically equal to the number of true positives.

In the above example, we had $P(B|A) + P(B|\neg A) = 1$, but this is not generally the case. What is true instead is $P(B|A) + P(\neg B|A) = 1$. Epidemiologists define the *sensitivity* of a binary classification test as the fraction of actual positives which are correctly identified, and the *specificity* as the fraction of actual negatives that are correctly identified. Thus, $\mathsf{se} = P(B|A)$ is the sensitivity and $\mathsf{sp} = P(\neg B|\neg A)$ is the specificity. We then have $P(B|\neg A) = 1 - P(\neg B|\neg A)$. Therefore,

$$P(B|A) + P(B|\neg A) = 1 + P(B|A) - P(\neg B|\neg A) = 1 + \mathsf{se} - \mathsf{sp} \; . \tag{1.30}$$

In our previous example, $\mathsf{se} = \mathsf{sp} = 0.9999$, in which case the RHS above gives 1. In general, if $P(A) \equiv f$ is the fraction of the population which is afflicted, then

$$P(\text{infected} \,|\, \text{positive}) = \frac{f \cdot \mathsf{se}}{f \cdot \mathsf{se} + (1 - f) \cdot (1 - \mathsf{sp})} \; . \tag{1.31}$$

For continuous distributions, we speak of a probability *density*. We then have

$$P(y) = \int dx \, P(y|x) \, P(x) \tag{1.32}$$

and

$$P(x|y) = \frac{P(y|x) \, P(x)}{\int dx' \, P(y|x') \, P(x')} \; . \tag{1.33}$$

The range of integration may depend on the specific application.

The quantities $P(A_i)$ are called the *prior distribution*. Clearly in order to compute $P(B)$ or $P(A_i|B)$ we must know the priors, and this is usually the weakest link in the Bayesian chain of reasoning. If our prior distribution is not accurate, Bayes' theorem will generate incorrect results. One approach to approximating prior probabilities $P(A_i)$ is to derive them from a *maximum entropy construction*.

1.3.3 Random variables and their averages

Consider an abstract probability space $\mathcal{X}$ whose elements (*i.e.* events) are labeled by x. The average of any function $f(x)$ is denoted as $\mathbb{E}f$ or $\langle f \rangle$, and is defined for discrete sets as

$$\mathbb{E}f = \langle f \rangle = \sum_{x \in \mathcal{X}} f(x) \, P(x) \; , \tag{1.34}$$

where $P(x)$ is the probability of x. For continuous sets, we have

$$\mathbb{E}f = \langle f \rangle = \int_{\mathcal{X}} dx \, f(x) \, P(x) \,. \tag{1.35}$$

Typically for continuous sets we have $\mathcal{X} = \mathbb{R}$ or $\mathcal{X} = \mathbb{R}_{\geq 0}$. Gardiner and other authors introduce an extra symbol, X, to denote a *random variable*, with $X(x) = x$ being its value. This is formally useful but notationally confusing, so we'll avoid it here and speak loosely of x as a random variable.

When there are two random variables $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we have $\Omega = \mathcal{X} \times \mathcal{Y}$ is the product space, and

$$\mathbb{E}f(x, y) = \langle f(x, y) \rangle = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f(x, y) \, P(x, y) \,, \tag{1.36}$$

with the obvious generalization to continuous sets. This generalizes to higher rank products, *i.e.* $x_i \in \mathcal{X}_i$ with $i \in \{1, \dots, N\}$. The *covariance* of x_i and x_j is defined as

$$C_{ij} \equiv \big\langle \big(x_i - \langle x_i \rangle\big)\big(x_j - \langle x_j \rangle\big) \big\rangle = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \,. \tag{1.37}$$

If $f(x)$ is a convex function then one has

$$\mathbb{E}f(x) \geq f(\mathbb{E}x) \,. \tag{1.38}$$

For continuous functions, $f(x)$ is convex if $f''(x) \geq 0$ everywhere[4]. If $f(x)$ is convex on some interval $[a, b]$ then for $x_{1,2} \in [a, b]$ we must have

$$f\big(\lambda x_1 + (1 - \lambda)x_2\big) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \,, \tag{1.39}$$

where $\lambda \in [0, 1]$. This is easily generalized to

$$f\Big(\sum_n p_n x_n\Big) \leq \sum_n p_n f(x_n) \,, \tag{1.40}$$

where $p_n = P(x_n)$, a result known as *Jensen's theorem*.

1.4 Entropy and Probability

1.4.1 Entropy and information theory

It was shown in the classic 1948 work of Claude Shannon that entropy is in fact a measure of *information*[5]. Suppose we observe that a particular event occurs with probability p. We associate with this observation an amount of information $I(p)$. The information $I(p)$ should satisfy certain desiderata:

---

[4] A function $g(x)$ is *concave* if $-g(x)$ is convex.

[5] See 'An Introduction to Information Theory and Entropy' by T. Carter, Santa Fe Complex Systems Summer School, June 2011. Available online at `http://astarte.csustan.edu/$\sim$tom/SFI-CSSS/info-theory/info-lec.pdf`.

1 Information is non-negative, *i.e.* $I(p) \geq 0$.

2 If two events occur independently so their joint probability is $p_1 \, p_2$, then their information is additive, *i.e.* $I(p_1 p_2) = I(p_1) + I(p_2)$.

3 $I(p)$ is a continuous function of p.

4 There is no information content to an event which is always observed, *i.e.* $I(1) = 0$.

From these four properties, it is easy to show that the only possible function $I(p)$ is

$$I(p) = -A \, \ln p \,, \tag{1.41}$$

where A is an arbitrary constant that can be absorbed into the base of the logarithm, since $\log_b x = \ln x / \ln b$. We will take $A = 1$ and use e as the base, so $I(p) = -\ln p$. Another common choice is to take the base of the logarithm to be 2, so $I(p) = -\log_2 p$. In this latter case, the units of information are known as *bits*. Note that $I(0) = \infty$. This means that the observation of an extremely rare event carries a great deal of information[6]

Now suppose we have a set of events labeled by an integer n which occur with probabilities $\{p_n\}$. What is the expected amount of information in N observations? Since event n occurs an average of $N p_n$ times, and the information content in p_n is $-\ln p_n$, we have that the average information per observation is

$$S = \frac{\langle I_N \rangle}{N} = -\sum_n p_n \ln p_n \,, \tag{1.42}$$

which is known as the entropy of the distribution. Thus, maximizing S is equivalent to maximizing the *information* content per observation.

Consider, for example, the information content of course grades. As we shall see, if the only constraint on the probability distribution is that of overall normalization, then S is maximized when all the probabilities p_n are equal. The binary entropy is then $S = \log_2 \Gamma$, since $p_n = 1/\Gamma$. Thus, for pass/fail grading, the maximum average information per grade is $-\log_2(\frac{1}{2}) = \log_2 2 = 1$ bit. If only A, B, C, D, and F grades are assigned, then the maximum average information per grade is $\log_2 5 = 2.32$ bits. If we expand the grade options to include {A+, A, A-, B+, B, B-, C+, C, C-, D, F}, then the maximum average information per grade is $\log_2 11 = 3.46$ bits.

Equivalently, consider, following the discussion in vol. 1 of Kardar, a random sequence $\{n_1, n_2, \ldots, n_N\}$ where each element n_j takes one of K possible values. There are then K^N such possible sequences, and to specify one of them requires $\log_2(K^N) = N \log_2 K$ bits of information. However, if the value n occurs with probability p_n, then on average it will occur $N_n = N p_n$ times in a sequence of length N, and the total number of such sequences will be

$$g(N) = \frac{N!}{\prod_{n=1}^{K} N_n!} \,. \tag{1.43}$$

---

[6]My colleague John McGreevy refers to $I(p)$ as the *surprise* of observing an event which occurs with probability p. I like this very much.

In general, this is far less that the total possible number K^N, and the number of bits necessary to specify one from among these $g(N)$ possibilities is

$$\log_2 g(N) = \log_2(N!) - \sum_{n=1}^{K} \log_2(N_n!) \approx -N \sum_{n=1}^{K} p_n \log_2 p_n \,, \tag{1.44}$$

up to terms of order unity. Here we have invoked Stirling's approximation. If the distribution is uniform, then we have $p_n = \frac{1}{K}$ for all $n \in \{1, \dots, K\}$, and $\log_2 g(N) = N \log_2 K$.

1.4.2 Probability distributions from maximum entropy

We have shown how one can proceed from a probability distribution and compute various averages. We now seek to go in the other direction, and determine the full probability distribution based on a knowledge of certain averages.

At first, this seems impossible. Suppose we want to reproduce the full probability distribution for an N-step random walk from knowledge of the average $\langle X \rangle = (2p-1)N$, where p is the probability of moving to the right at each step (see §1.2 above). The problem seems ridiculously underdetermined, since there are 2^N possible configurations for an N-step random walk: $\sigma_j = \pm 1$ for $j = 1, \dots, N$. Overall normalization requires

$$\sum_{\{\sigma_j\}} P(\sigma_1, \dots, \sigma_N) = 1 \,, \tag{1.45}$$

but this just imposes one constraint on the 2^N probabilities $P(\sigma_1, \dots, \sigma_N)$, leaving $2^N - 1$ overall parameters. What principle allows us to reconstruct the full probability distribution

$$P(\sigma_1, \dots, \sigma_N) = \prod_{j=1}^{N} \left(p \, \delta_{\sigma_j, 1} + q \, \delta_{\sigma_j, -1} \right) = \prod_{j=1}^{N} p^{(1+\sigma_j)/2} \, q^{(1-\sigma_j)/2} \,, \tag{1.46}$$

corresponding to N independent steps?

The principle of maximum entropy

The entropy of a discrete probability distribution $\{p_n\}$ is defined as

$$S = -\sum_{n} p_n \ln p_n \,, \tag{1.47}$$

where here we take e as the base of the logarithm. The entropy may therefore be regarded as a function of the probability distribution: $S = S(\{p_n\})$. One special property of the entropy is the following. Suppose we have two independent normalized distributions $\{p_a^{\mathsf{A}}\}$ and $\{p_b^{\mathsf{B}}\}$. The joint probability for

events a and b is then $P_{a,b} = p_a^{\mathsf{A}} \, p_b^{\mathsf{B}}$. The entropy of the joint distribution is then

$$
\begin{aligned}
S &= -\sum_a \sum_b P_{a,b} \ln P_{a,b} = -\sum_a \sum_b p_a^{\mathsf{A}} \, p_b^{\mathsf{B}} \ln \left(p_a^{\mathsf{A}} \, p_b^{\mathsf{B}} \right) = -\sum_a \sum_b p_a^{\mathsf{A}} \, p_b^{\mathsf{B}} \left(\ln p_a^{\mathsf{A}} + \ln p_b^{\mathsf{B}} \right) \\
&= -\sum_a p_a^{\mathsf{A}} \ln p_a^{\mathsf{A}} \cdot \sum_b p_b^{\mathsf{B}} - \sum_b p_b^{\mathsf{B}} \ln p_b^{\mathsf{B}} \cdot \sum_a p_a^{\mathsf{A}} = -\sum_a p_a^{\mathsf{A}} \ln p_a^{\mathsf{A}} - \sum_b p_b^{\mathsf{B}} \ln p_b^{\mathsf{B}} \\
&= S^{\mathsf{A}} + S^{\mathsf{B}} \; .
\end{aligned}
$$

Thus, the entropy of a joint distribution formed from two independent distributions is additive.

Suppose all we knew about $\{p_n\}$ was that it was normalized. Then $\sum_n p_n = 1$. This is a constraint on the values $\{p_n\}$. Let us now extremize the entropy S with respect to the distribution $\{p_n\}$, but subject to the normalization constraint. We do this using Lagrange's method of undetermined multipliers. We define

$$
S^*\big(\{p_n\}, \lambda\big) = -\sum_n p_n \ln p_n - \lambda \Big(\sum_n p_n - 1 \Big) \tag{1.48}
$$

and we freely extremize S^* over all its arguments. Thus, for all n we have

$$
\begin{aligned}
0 &= \frac{\partial S^*}{\partial p_n} = -\big(\ln p_n + 1 + \lambda \big) \\
0 &= \frac{\partial S^*}{\partial \lambda} = \sum_n p_n - 1 \; .
\end{aligned}
\tag{1.49}
$$

From the first of these equations, we obtain $p_n = e^{-(1+\lambda)}$, and from the second we obtain

$$
\sum_n p_n = e^{-(1+\lambda)} \cdot \sum_n 1 = \varGamma \, e^{-(1+\lambda)} \; , \tag{1.50}
$$

where $\varGamma \equiv \sum_n 1$ is the total number of possible events. Thus, $p_n = 1/\varGamma$, which says that all events are equally probable.

Now suppose we know one other piece of information, which is the average value $X = \sum_n X_n \, p_n$ of some quantity. We now extremize S subject to two constraints, and so we define

$$
S^*\big(\{p_n\}, \lambda_0, \lambda_1\big) = -\sum_n p_n \ln p_n - \lambda_0 \Big(\sum_n p_n - 1 \Big) - \lambda_1 \Big(\sum_n X_n \, p_n - X \Big) \; . \tag{1.51}
$$

We then have

$$
\frac{\partial S^*}{\partial p_n} = -\big(\ln p_n + 1 + \lambda_0 + \lambda_1 \, X_n \big) = 0 \; , \tag{1.52}
$$

which yields the two-parameter distribution

$$
p_n = e^{-(1+\lambda_0)} \, e^{-\lambda_1 X_n} \; . \tag{1.53}
$$

To fully determine the distribution $\{p_n\}$ we need to invoke the two equations $\sum_n p_n = 1$ and $\sum_n X_n \, p_n = X$, which come from extremizing S^* with respect to λ_0 and λ_1, respectively:

$$
\begin{aligned}
1 &= e^{-(1+\lambda_0)} \sum_n e^{-\lambda_1 X_n} \\
X &= e^{-(1+\lambda_0)} \sum_n X_n \, e^{-\lambda_1 X_n} \; .
\end{aligned}
\tag{1.54}
$$

General formulation

The generalization to K extra pieces of information (plus normalization) is immediately apparent. We have

$$X^a = \sum_n X_n^a \, p_n \,, \tag{1.55}$$

and therefore we define

$$S^*\big(\{p_n\}, \{\lambda_a\}\big) = -\sum_n p_n \ln p_n - \sum_{a=0}^{K} \lambda_a \Big(\sum_n X_n^a \, p_n - X^a \Big) \,, \tag{1.56}$$

with $X_n^{(a=0)} \equiv X^{(a=0)} = 1$. Then the optimal distribution which extremizes S subject to the $K+1$ constraints is

$$
\begin{aligned}
p_n &= \exp\left\{ -1 - \sum_{a=0}^{K} \lambda_a \, X_n^a \right\} \\
&= \frac{1}{Z} \exp\left\{ -\sum_{a=1}^{K} \lambda_a \, X_n^a \right\} \,,
\end{aligned}
\tag{1.57}
$$

where $Z = e^{1+\lambda_0}$ is determined by normalization: $\sum_n p_n = 1$. This is a $(K+1)$-parameter distribution, with $\{\lambda_0, \lambda_1, \ldots, \lambda_K\}$ determined by the $K+1$ constraints in eqn. 1.55.

Example

As an example, consider the random walk problem. We have two pieces of information:

$$
\begin{aligned}
&\sum_{\sigma_1} \cdots \sum_{\sigma_N} P(\sigma_1, \ldots, \sigma_N) = 1 \\
&\sum_{\sigma_1} \cdots \sum_{\sigma_N} P(\sigma_1, \ldots, \sigma_N) \sum_{j=1}^{N} \sigma_j = X \,.
\end{aligned}
\tag{1.58}
$$

Here the discrete label n from §1.4.2 ranges over 2^N possible values, and may be written as an N digit binary number $r_N \cdots r_1$, where $r_j = \frac{1}{2}(1 + \sigma_j)$ is 0 or 1. Extremizing S subject to these constraints, we obtain

$$P(\sigma_1, \ldots, \sigma_N) = \mathcal{C} \, \exp\left\{ -\lambda \sum_j \sigma_j \right\} = \mathcal{C} \prod_{j=1}^{N} e^{-\lambda \sigma_j} \,, \tag{1.59}$$

where $\mathcal{C} \equiv e^{-(1+\lambda_0)}$ and $\lambda \equiv \lambda_1$. Normalization then requires

$$\operatorname{Tr} P \equiv \sum_{\{\sigma_j\}} P(\sigma_1, \ldots, \sigma_N) = \mathcal{C} \left(e^{\lambda} + e^{-\lambda} \right)^N \,, \tag{1.60}$$

hence $\mathcal{C} = (\cosh \lambda)^{-N}$. We then have

$$P(\sigma_1, \ldots, \sigma_N) = \prod_{j=1}^{N} \frac{e^{-\lambda \sigma_j}}{e^{\lambda} + e^{-\lambda}} = \prod_{j=1}^{N} \left(p \, \delta_{\sigma_j, 1} + q \, \delta_{\sigma_j, -1} \right) , \tag{1.61}$$

where

$$p = \frac{e^{-\lambda}}{e^{\lambda} + e^{-\lambda}} \quad , \quad q = 1 - p = \frac{e^{\lambda}}{e^{\lambda} + e^{-\lambda}} . \tag{1.62}$$

We then have $X = (2p - 1)N$, which determines $p = \frac{1}{2}(N + X)$, and we have recovered the Bernoulli distribution.

Of course there are no miracles[7], and there are an infinite family of distributions for which $X = (2p - 1)N$ that are not Bernoulli. For example, we could have imposed another constraint, such as $E = \sum_{j=1}^{N-1} \sigma_j \, \sigma_{j+1}$. This would result in the distribution

$$P(\sigma_1, \ldots, \sigma_N) = \frac{1}{Z} \exp \left\{ - \lambda_1 \sum_{j=1}^{N} \sigma_j - \lambda_2 \sum_{j=1}^{N-1} \sigma_j \, \sigma_{j+1} \right\} , \tag{1.63}$$

with $Z(\lambda_1, \lambda_2)$ determined by normalization: $\sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}) = 1$. This is the one-dimensional Ising chain of classical equilibrium statistical physics. Defining the transfer matrix $R_{ss'} = e^{-\lambda_1 (s + s')/2} \, e^{-\lambda_2 s s'}$ with $s, s' = \pm 1$,

$$R = \begin{pmatrix} e^{-\lambda_1 - \lambda_2} & e^{\lambda_2} \\ e^{\lambda_2} & e^{\lambda_1 - \lambda_2} \end{pmatrix}$$
$$= e^{-\lambda_2} \cosh \lambda_1 \, \mathbb{I} + e^{\lambda_2} \, \tau^x - e^{-\lambda_2} \sinh \lambda_1 \, \tau^z , \tag{1.64}$$

where τ^x and τ^z are Pauli matrices, we have that

$$Z_{\text{ring}} = \mathsf{Tr} \left(R^N \right) \quad , \quad Z_{\text{chain}} = \mathsf{Tr} \left(R^{N-1} S \right) , \tag{1.65}$$

where $S_{ss'} = e^{-\lambda_1 (s + s')/2}$, *i.e.*

$$S = \begin{pmatrix} e^{-\lambda_1} & 1 \\ 1 & e^{\lambda_1} \end{pmatrix}$$
$$= \cosh \lambda_1 \, \mathbb{I} + \tau^x - \sinh \lambda_1 \, \tau^z . \tag{1.66}$$

The appropriate case here is that of the chain, but in the thermodynamic limit $N \to \infty$ both chain and ring yield identical results, so we will examine here the results for the ring, which are somewhat easier to obtain. Clearly $Z_{\text{ring}} = \zeta_+^N + \zeta_-^N$, where $\zeta_\pm$ are the eigenvalues of R:

$$\zeta_\pm = e^{-\lambda_2} \cosh \lambda_1 \pm \sqrt{e^{-2\lambda_2} \sinh^2 \lambda_1 + e^{2\lambda_2}} \quad . \tag{1.67}$$

In the thermodynamic limit, the ζ_+ eigenvalue dominates, and $Z_{\text{ring}} \simeq \zeta_+^N$. We now have

$$X = \left\langle \sum_{j=1}^{N} \sigma_j \right\rangle = - \frac{\partial \ln Z}{\partial \lambda_1} = - \frac{N \sinh \lambda_1}{\sqrt{\sinh^2 \lambda_1 + e^{4\lambda_2}}} . \tag{1.68}$$

---

[7]See §10 of *An Enquiry Concerning Human Understanding* by David Hume (1748).

We also have $E = -\partial \ln Z / \partial \lambda_2$. These two equations determine the Lagrange multipliers $\lambda_1(X, E, N)$ and $\lambda_2(X, E, N)$. In the thermodynamic limit, we have $\lambda_i = \lambda_i(X/N, E/N)$. Thus, if we fix $X/N = 2p-1$ alone, there is a continuous one-parameter family of distributions, parametrized $\varepsilon = E/N$, which satisfy the constraint on X.

So what is it about the maximum entropy approach that is so compelling? Maximum entropy gives us a calculable distribution which is consistent with maximum ignorance given our known constraints. In that sense, it is as unbiased as possible, from an information theoretic point of view. As a starting point, a maximum entropy distribution may be improved upon, using Bayesian methods for example (see §1.6.2 below).

1.4.3 Continuous probability distributions

Suppose we have a continuous probability density $P(\varphi)$ defined over some set Ω. We have observables

$$X^a = \int_\Omega d\mu \, X^a(\varphi) \, P(\varphi) \,, \tag{1.69}$$

where $d\mu$ is the appropriate integration measure. We assume $d\mu = \prod_{j=1}^{D} d\varphi_j$, where D is the dimension of Ω. Then we extremize the functional

$$S^*[P(\varphi), \{\lambda_a\}] = -\int_\Omega d\mu \, P(\varphi) \ln P(\varphi) - \sum_{a=0}^{K} \lambda_a \left(\int_\Omega d\mu \, P(\varphi) \, X^a(\varphi) - X^a \right) \tag{1.70}$$

with respect to $P(\varphi)$ and with respect to $\{\lambda_a\}$. Again, $X^0(\varphi) \equiv X^0 \equiv 1$. This yields the following result:

$$\ln P(\varphi) = -1 - \sum_{a=0}^{K} \lambda_a \, X^a(\varphi) \,. \tag{1.71}$$

The $K + 1$ Lagrange multipliers $\{\lambda_a\}$ are then determined from the $K + 1$ constraint equations in eqn. 1.69.

As an example, consider a distribution $P(x)$ over the real numbers $\mathbb{R}$. We constrain

$$\int_{-\infty}^{\infty} dx \, P(x) = 1 \quad , \quad \int_{-\infty}^{\infty} dx \, x \, P(x) = \mu \quad , \quad \int_{-\infty}^{\infty} dx \, x^2 \, P(x) = \mu^2 + \sigma^2 \,. \tag{1.72}$$

Extremizing the entropy, we then obtain

$$P(x) = \mathcal{C} \, e^{-\lambda_1 x - \lambda_2 x^2} \,, \tag{1.73}$$

where $\mathcal{C} = e^{-(1+\lambda_0)}$. We already know the answer:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \,. \tag{1.74}$$

In other words, $\lambda_1 = -\mu/\sigma^2$ and $\lambda_2 = 1/2\sigma^2$, with $\mathcal{C} = (2\pi\sigma^2)^{-1/2} \exp(-\mu^2/2\sigma^2)$.

1.5   General Aspects of Probability Distributions

1.5.1   Discrete and continuous distributions

Consider a system whose possible configurations $|\,n\,\rangle$ can be labeled by a discrete variable $n \in \mathcal{C}$, where $\mathcal{C}$ is the set of possible configurations. The total number of possible configurations, which is to say the *order* of the set $\mathcal{C}$, may be finite or infinite. Next, consider an ensemble of such systems, and let P_n denote the probability that a given random element from that ensemble is in the state (configuration) $|\,n\,\rangle$. The collection $\{P_n\}$ forms a *discrete probability distribution*. We assume that the distribution is *normalized*, meaning

$$\sum_{n \in \mathcal{C}} P_n = 1 \ . \tag{1.75}$$

Now let A_n be a quantity which takes values depending on n. The average of A is given by

$$\langle A \rangle = \sum_{n \in \mathcal{C}} P_n \, A_n \ . \tag{1.76}$$

Typically, $\mathcal{C}$ is the set of integers ($\mathbb{Z}$) or some subset thereof, but it could be any countable set. As an example, consider the throw of a single six-sided die. Then $P_n = \frac{1}{6}$ for each $n \in \{1, \ldots, 6\}$. Let $A_n = 0$ if n is even and 1 if n is odd. Then find $\langle A \rangle = \frac{1}{2}$, *i.e.* on average half the throws of the die will result in an even number.

It may be that the system's configurations are described by several discrete variables $\{n_1, n_2, n_3, \ldots\}$. We can combine these into a vector $\boldsymbol{n}$ and then we write $P_{\boldsymbol{n}}$ for the discrete distribution, with $\sum_{\boldsymbol{n}} P_{\boldsymbol{n}} = 1$.

Another possibility is that the system's configurations are parameterized by a collection of continuous variables, $\boldsymbol{\varphi} = \{\varphi_1, \ldots, \varphi_n\}$. We write $\boldsymbol{\varphi} \in \Omega$, where Ω is the phase space (or configuration space) of the system. Let $d\mu$ be a *measure* on this space. In general, we can write

$$d\mu = W(\varphi_1, \ldots, \varphi_n) \, d\varphi_1 \, d\varphi_2 \cdots d\varphi_n \ . \tag{1.77}$$

The phase space measure used in classical statistical mechanics gives equal weight W to equal phase space volumes:

$$d\mu = \mathcal{C} \prod_{\sigma=1}^{r} dq_\sigma \, dp_\sigma \ , \tag{1.78}$$

where $\mathcal{C}$ is a constant we shall discuss later on below[8].

Any continuous probability distribution $P(\boldsymbol{\varphi})$ is normalized according to

$$\int_\Omega d\mu \, P(\boldsymbol{\varphi}) = 1 \ . \tag{1.79}$$

---

[8]Such a measure is invariant with respect to canonical transformations, which are the broad class of transformations among coordinates and momenta which leave Hamilton's equations of motion invariant, and which preserve phase space volumes under Hamiltonian evolution. For this reason $d\mu$ is called an *invariant phase space measure*.

The average of a function $A(\varphi)$ on configuration space is then

$$\langle A \rangle = \int_\Omega d\mu \, P(\varphi) \, A(\varphi) \, . \tag{1.80}$$

For example, consider the Gaussian distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-(x-\mu)^2/2\sigma^2} \, . \tag{1.81}$$

From the result[9]

$$\int_{-\infty}^{\infty} dx \, e^{-\alpha x^2} \, e^{-\beta x} = \sqrt{\frac{\pi}{\alpha}} \, e^{\beta^2/4\alpha} \, , \tag{1.82}$$

we see that $P(x)$ is normalized. One can then compute

$$\begin{aligned} \langle x \rangle &= \mu \\ \langle x^2 \rangle - \langle x \rangle^2 &= \sigma^2 \, . \end{aligned} \tag{1.83}$$

We call μ the *mean* and σ the *standard deviation* of the distribution, eqn. 1.81.

The quantity $P(\varphi)$ is called the *distribution* or *probability density*. One has

$$P(\varphi) \, d\mu = \text{probability that configuration lies within volume } d\mu \text{ centered at } \varphi$$

For example, consider the probability density $P = 1$ normalized on the interval $x \in [0, 1]$. The probability that some x chosen at random will be *exactly* $\frac{1}{2}$, say, is infinitesimal – one would have to specify each of the infinitely many digits of x. However, we can say that $x \in [0.45 \, , \, 0.55]$ with probability $\frac{1}{10}$.

If x is distributed according to $P_1(x)$, then the probability distribution on the product space $(x_1 \, , \, x_2)$ is simply the product of the distributions: $P_2(x_1, x_2) = P_1(x_1) \, P_1(x_2)$. Suppose we have a function $\phi(x_1, \ldots, x_N)$. How is it distributed? Let $P(\phi)$ be the distribution for ϕ. We then have

$$\begin{aligned} P(\phi) &= \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_N \, P_N(x_1, \ldots, x_N) \, \delta\Big(\phi(x_1, \ldots, x_N) - \phi\Big) \\ &= \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_N \, P_1(x_1) \cdots P_1(x_N) \, \delta\Big(\phi(x_1, \ldots, x_N) - \phi\Big) \, , \end{aligned} \tag{1.84}$$

where the second line is appropriate if the $\{x_j\}$ are themselves distributed independently. Note that

$$\int_{-\infty}^{\infty} d\phi \, P(\phi) = 1 \, , \tag{1.85}$$

so $P(\phi)$ is itself normalized.

---

[9]Memorize this!

1.5.2 Central limit theorem

In particular, consider the distribution function of the sum $X = \sum_{i=1}^{N} x_i$. We will be particularly interested in the case where N is large. For general N, though, we have

$$P_N(X) = \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_N \, P_1(x_1) \cdots P_1(x_N) \, \delta\big(x_1 + x_2 + \ldots + x_N - X\big) \, . \tag{1.86}$$

It is convenient to compute the Fourier transform[10] of $P(X)$:

$$\begin{aligned}
\hat{P}_N(k) &= \int_{-\infty}^{\infty} dX \, P_N(X) \, e^{-ikX} \\
&= \int_{-\infty}^{\infty} dX \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_N \, P_1(x_1) \cdots P_1(x_N) \, \delta\big(x_1 + \ldots + x_N - X\big) \, e^{-ikX} = \big[\hat{P}_1(k)\big]^N \, ,
\end{aligned} \tag{1.87}$$

where

$$\hat{P}_1(k) = \int_{-\infty}^{\infty} dx \, P_1(x) \, e^{-ikx} \tag{1.88}$$

is the Fourier transform of the single variable distribution $P_1(x)$. The distribution $P_N(X)$ is a *convolution* of the individual $P_1(x_i)$ distributions. We have therefore proven that *the Fourier transform of a convolution is the product of the Fourier transforms*.

OK, now we can write for $\hat{P}_1(k)$

$$\begin{aligned}
\hat{P}_1(k) &= \int_{-\infty}^{\infty} dx \, P_1(x) \left(1 - ikx - \tfrac{1}{2} \, k^2 x^2 + \tfrac{1}{6} \, i \, k^3 \, x^3 + \ldots \right) \\
&= 1 - ik\langle x \rangle - \tfrac{1}{2} \, k^2 \langle x^2 \rangle + \tfrac{1}{6} \, i \, k^3 \langle x^3 \rangle + \ldots \, .
\end{aligned} \tag{1.89}$$

---

[10]Jean Baptiste Joseph Fourier (1768-1830) had an illustrious career. The son of a tailor, and orphaned at age eight, Fourier's ignoble status rendered him ineligible to receive a commission in the scientific corps of the French army. A Benedictine minister at the École Royale Militaire of Auxerre remarked, "Fourier, not being noble, could not enter the artillery, although he were a second Newton." Fourier prepared for the priesthood but his affinity for mathematics proved overwhelming, and so he left the abbey and soon thereafter accepted a military lectureship position. Despite his initial support for revolution in France, in 1794 Fourier ran afoul of a rival sect while on a trip to Orleans and was arrested and very nearly guillotined. Fortunately the Reign of Terror ended soon after the death of Robespierre, and Fourier was released. He went on Napoleon Bonaparte's 1798 expedition to Egypt, where he was appointed governor of Lower Egypt. His organizational skills impressed Napoleon, and upon return to France he was appointed to a position of prefect in Grenoble. It was in Grenoble that Fourier performed his landmark studies of heat, and his famous work on partial differential equations and Fourier series. It seems that Fourier's fascination with heat began in Egypt, where he developed an appreciation of desert climate. His fascination developed into an obsession, and he became convinced that heat could promote a healthy body. He would cover himself in blankets, like a mummy, in his heated apartment, even during the middle of summer. On May 4, 1830, Fourier, so arrayed, tripped and fell down a flight of stairs. This aggravated a developing heart condition, which he refused to treat with anything other than more heat. Two weeks later, he died. Fourier's is one of the 72 names of scientists, engineers and other luminaries which are engraved on the Eiffel Tower.

Thus,

$$\ln \hat{P}_1(k) = -i\mu k - \tfrac{1}{2}\sigma^2 k^2 + \tfrac{1}{6} i\, \gamma^3\, k^3 + \dots \; , \tag{1.90}$$

where

$$\begin{aligned}
\mu &= \langle x \rangle \\
\sigma^2 &= \langle x^2 \rangle - \langle x \rangle^2 \\
\gamma^3 &= \langle x^3 \rangle - 3\langle x^2 \rangle \langle x \rangle + 2\langle x \rangle^3
\end{aligned} \tag{1.91}$$

We can now write

$$\big[\hat{P}_1(k)\big]^N = e^{-iN\mu k}\, e^{-N\sigma^2 k^2/2}\, e^{iN\gamma^3 k^3/6} \cdots \tag{1.92}$$

Now for the inverse transform. In computing $P_N(X)$, we will expand the term $e^{iN\gamma^3 k^3/6}$ and all subsequent terms in the above product as a power series in k. We then have

$$\begin{aligned}
P_N(X) &= \int\limits_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ik(X-N\mu)}\, e^{-N\sigma^2 k^2/2} \left\{ 1 + \tfrac{1}{6}\, i\, N\gamma^3 k^3 + \dots \right\} \\
&= \left(1 - \frac{\gamma^3}{6} N \frac{\partial^3}{\partial X^3} + \dots \right) \frac{1}{\sqrt{2\pi N\sigma^2}}\, e^{-(X-N\mu)^2/2N\sigma^2} \\
&= \left(1 - \frac{\gamma^3}{6} N^{-1/2} \frac{\partial^3}{\partial \xi^3} + \dots \right) \frac{1}{\sqrt{2\pi N\sigma^2}}\, e^{-\xi^2/2\sigma^2} \; .
\end{aligned} \tag{1.93}$$

In going from the second line to the third, we have written $X = N\mu + \sqrt{N}\,\xi$, in which case $\partial_X = N^{-1/2}\,\partial_\xi$, and the non-Gaussian terms give a subleading contribution which vanishes in the $N \to \infty$ limit. We have just proven the *central limit theorem*: in the limit $N \to \infty$, the distribution of a sum of N independent random variables x_i is a Gaussian with mean $N\mu$ and standard deviation $\sqrt{N}\,\sigma$. Our only assumptions are that the mean μ and standard deviation σ exist for the distribution $P_1(x)$. Note that $P_1(x)$ itself need not be a Gaussian – it could be a very peculiar distribution indeed, but so long as its first and second moment exist, where the kth moment is simply $\langle x^k \rangle$, the distribution of the sum $X = \sum_{i=1}^{N} x_i$ is a Gaussian.

1.5.3 Moments and cumulants

Consider a general multivariate distribution $P(x_1, \dots, x_N)$ and define the multivariate Fourier transform

$$\hat{P}(k_1, \dots, k_N) = \int\limits_{-\infty}^{\infty} dx_1 \cdots \int\limits_{-\infty}^{\infty} dx_N\, P(x_1, \dots, x_N)\, \exp\left(-i \sum_{j=1}^{N} k_j x_j \right) . \tag{1.94}$$

The inverse relation is

$$P(x_1, \dots, x_N) = \int\limits_{-\infty}^{\infty} \frac{dk_1}{2\pi} \cdots \int\limits_{-\infty}^{\infty} \frac{dk_N}{2\pi}\, \hat{P}(k_1, \dots, k_N)\, \exp\left(+i \sum_{j=1}^{N} k_j x_j \right) . \tag{1.95}$$

Acting on $\hat{P}(\boldsymbol{k})$, the differential operator $i\frac{\partial}{\partial k_i}$ brings down from the exponential a factor of x_i inside the integral. Thus,

$$\left[\left(i\frac{\partial}{\partial k_1}\right)^{m_1}\cdots\left(i\frac{\partial}{\partial k_N}\right)^{m_N}\hat{P}(\boldsymbol{k})\right]_{\boldsymbol{k}=0} = \left\langle x_1^{m_1}\cdots x_N^{m_N}\right\rangle . \tag{1.96}$$

Similarly, we can reconstruct the distribution from its moments, *viz.*

$$\hat{P}(\boldsymbol{k}) = \sum_{m_1=0}^{\infty}\cdots\sum_{m_N=0}^{\infty}\frac{(-ik_1)^{m_1}}{m_1!}\cdots\frac{(-ik_N)^{m_N}}{m_N!}\left\langle x_1^{m_1}\cdots x_N^{m_N}\right\rangle . \tag{1.97}$$

The *cumulants* $\langle\!\langle x_1^{m_1}\cdots x_N^{m_N}\rangle\!\rangle$ are defined by the Taylor expansion of $\ln\hat{P}(\boldsymbol{k})$:

$$\ln\hat{P}(\boldsymbol{k}) = \sum_{m_1=0}^{\infty}\cdots\sum_{m_N=0}^{\infty}\frac{(-ik_1)^{m_1}}{m_1!}\cdots\frac{(-ik_N)^{m_N}}{m_N!}\langle\!\langle x_1^{m_1}\cdots x_N^{m_N}\rangle\!\rangle . \tag{1.98}$$

There is no general form for the cumulants. It is straightforward to derive the following low order results:

$$\begin{aligned}
\langle\!\langle x_i\rangle\!\rangle &= \langle x_i\rangle \\
\langle\!\langle x_i x_j\rangle\!\rangle &= \langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle \\
\langle\!\langle x_i x_j x_k\rangle\!\rangle &= \langle x_i x_j x_k\rangle - \langle x_i x_j\rangle\langle x_k\rangle - \langle x_j x_k\rangle\langle x_i\rangle - \langle x_k x_i\rangle\langle x_j\rangle + 2\langle x_i\rangle\langle x_j\rangle\langle x_k\rangle .
\end{aligned} \tag{1.99}$$

1.5.4 Multidimensional Gaussian integral

Consider the multivariable Gaussian distribution,

$$P(\boldsymbol{x}) \equiv \left(\frac{\det A}{(2\pi)^n}\right)^{1/2}\exp\left(-\tfrac{1}{2}\,x_i\,A_{ij}\,x_j\right) , \tag{1.100}$$

where A is a positive definite matrix of rank n. A mathematical result which is extremely important throughout physics is the following:

$$Z(\boldsymbol{b}) = \left(\frac{\det A}{(2\pi)^n}\right)^{1/2}\int\limits_{-\infty}^{\infty}dx_1\cdots\int\limits_{-\infty}^{\infty}dx_n\,\exp\left(-\tfrac{1}{2}\,x_i\,A_{ij}\,x_j + b_i\,x_i\right) = \exp\left(\tfrac{1}{2}\,b_i\,A_{ij}^{-1}\,b_j\right) . \tag{1.101}$$

Here, the vector $\boldsymbol{b} = (b_1,\ldots,b_n)$ is identified as a *source*. Since $Z(0) = 1$, we have that the distribution $P(\boldsymbol{x})$ is normalized. Now consider averages of the form

$$\begin{aligned}
\left\langle x_{j_1}\cdots x_{j_{2k}}\right\rangle &= \int d^n x\,P(\boldsymbol{x})\,x_{j_1}\cdots x_{j_{2k}} = \frac{\partial^n Z(\boldsymbol{b})}{\partial b_{j_1}\cdots\partial b_{j_{2k}}}\bigg|_{\boldsymbol{b}=0} \\
&= \sum_{\text{contractions}} A_{j_{\sigma(1)}j_{\sigma(2)}}^{-1}\cdots A_{j_{\sigma(2k-1)}j_{\sigma(2k)}}^{-1} .
\end{aligned} \tag{1.102}$$

The sum in the last term is over all *contractions* of the indices $\{j_1, \ldots, j_{2k}\}$. A contraction is an arrangement of the $2k$ indices into k pairs. There are $C_{2k} = (2k)!/2^k k!$ possible such contractions. To obtain this result for C_k, we start with the first index and then find a mate among the remaining $2k-1$ indices. Then we choose the next unpaired index and find a mate among the remaining $2k - 3$ indices. Proceeding in this manner, we have

$$C_{2k} = (2k - 1) \cdot (2k - 3) \cdots 3 \cdot 1 = \frac{(2k)!}{2^k k!} \ . \tag{1.103}$$

Equivalently, we can take all possible permutations of the $2k$ indices, and then divide by $2^k k!$ since permutation within a given pair results in the same contraction and permutation among the k pairs results in the same contraction. For example, for $k = 2$, we have $C_4 = 3$, and

$$\langle\, x_{j_1} x_{j_2} x_{j_3} x_{j_4} \,\rangle = A^{-1}_{j_1 j_2} A^{-1}_{j_3 j_4} + A^{-1}_{j_1 j_3} A^{-1}_{j_2 j_4} + A^{-1}_{j_1 j_4} A^{-1}_{j_2 j_3} \ . \tag{1.104}$$

If we define $b_i = ik_i$, we have

$$\hat{P}(\boldsymbol{k}) = \exp\left(-\tfrac{1}{2}\, k_i\, A^{-1}_{ij}\, k_j\right), \tag{1.105}$$

from which we read off the cumulants $\langle\!\langle x_i x_j \rangle\!\rangle = A^{-1}_{ij}$, with all higher order cumulants vanishing.

1.6 Bayesian Statistical Inference

1.6.1 Frequentists and Bayesians

There field of statistical inference is roughly divided into two schools of practice: frequentism and Bayesianism. You can find several articles on the web discussing the differences in these two approaches. In both cases we would like to model observable data x by a distribution. The distribution in general depends on one or more parameters $\boldsymbol{\theta}$. The basic worldviews of the two approaches are as follows:

Frequentism: Data x are a random sample drawn from an infinite pool at some *frequency*. The underlying parameters $\boldsymbol{\theta}$, which are to be estimated, remain fixed during this process. There is no information prior to the model specification. The experimental conditions under which the data are collected are presumed to be controlled and repeatable. Results are generally expressed in terms of *confidence intervals* and *confidence levels*, obtained via *statistical hypothesis testing*. Probabilities have meaning only for data yet to be collected. Calculations generally are computationally straightforward.

Bayesianism: The only data x which matter are those which have been observed. The parameters $\boldsymbol{\theta}$ are unknown and described probabilistically using a *prior distribution*, which is generally based on some available information but which also may be at least partially subjective. The priors are then to be *updated* based on observed data x. Results are expressed in terms of *posterior distributions* and *credible intervals*. Calculations can be computationally intensive.

In essence, frequentists say *the data are random and the parameters are fixed.* while Bayesians say *the data are fixed and the parameters are random*[11]. Overall, frequentism has dominated over the past several hundred years, but Bayesianism has been coming on strong of late, and many physicists seem naturally drawn to the Bayesian perspective.

1.6.2   Updating Bayesian priors

Given *data D* and a *hypothesis H*, Bayes' theorem tells us

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)} \ .$$

(1.106)

Typically the data is in the form of a set of values $x = \{x_1, \ldots, x_N\}$, and the hypothesis in the form of a set of parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_K\}$. It is notationally helpful to express distributions of x and distributions of x conditioned on $\boldsymbol{\theta}$ using the symbol f, and distributions of $\boldsymbol{\theta}$ and distributions of $\boldsymbol{\theta}$ conditioned on x using the symbol π, rather than using the symbol P everywhere. We then have

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{\int\limits_{\Theta} d\boldsymbol{\theta}'\ f(\boldsymbol{x}|\boldsymbol{\theta}')\,\pi(\boldsymbol{\theta}')} \ ,$$

(1.107)

where $\Theta \ni \boldsymbol{\theta}$ is the space of parameters. Note that $\int_{\Theta} d\boldsymbol{\theta}\ \pi(\boldsymbol{\theta}|\boldsymbol{x}) = 1$. The denominator of the RHS is simply $f(\boldsymbol{x})$, which is independent of $\boldsymbol{\theta}$, hence $\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})$. We call $\pi(\boldsymbol{\theta})$ the *prior* for $\boldsymbol{\theta}$, $f(\boldsymbol{x}|\boldsymbol{\theta})$ the *likelihood* of $\boldsymbol{x}$ given $\boldsymbol{\theta}$, and $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ the *posterior* for $\boldsymbol{\theta}$ given $\boldsymbol{x}$. The idea here is that while our initial guess at the $\boldsymbol{\theta}$ distribution is given by the prior $\pi(\boldsymbol{\theta})$, after taking data, we should *update* this distribution to the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{x})$. The likelihood $f(\boldsymbol{x}|\boldsymbol{\theta})$ is entailed by our model for the phenomenon which produces the data. We can use the posterior to find the distribution of new data points $\boldsymbol{y}$, called the *posterior predictive distribution*,

$$f(\boldsymbol{y}|\boldsymbol{x}) = \int\limits_{\Theta} d\boldsymbol{\theta}\ f(\boldsymbol{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\boldsymbol{x}) \ .$$

(1.108)

This is the update of the *prior predictive distribution*,

$$f(\boldsymbol{x}) = \int\limits_{\Theta} d\boldsymbol{\theta}\ f(\boldsymbol{x}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}) \ .$$

(1.109)

Example: coin flipping

Consider a model of coin flipping based on a standard Bernoulli distribution, where $\theta \in [0,1]$ is the probability for heads ($x = 1$) and $1 - \theta$ the probability for tails ($x = 0$). That is,

$$\begin{aligned} f(x_1, \ldots, x_N|\theta) &= \prod_{j=1}^{N} \left[(1-\theta)\,\delta_{x_j,0} + \theta\,\delta_{x_j,1} \right] \\ &= \theta^X (1-\theta)^{N-X} \ , \end{aligned}$$

(1.110)

---

[11]"A frequentist is a person whose long-run ambition is to be wrong 5% of the time. A Bayesian is one who, vaguely expecting a horse, and catching glimpse of a donkey, strongly believes he has seen a mule." – Charles Annis.

where $X = \sum_{j=1}^{N} x_j$ is the observed total number of heads, and $N - X$ the corresponding number of tails. We now need a prior $\pi(\theta)$. We choose the Beta distribution,

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\mathsf{B}(\alpha, \beta)} \,, \tag{1.111}$$

where $\mathsf{B}(\alpha, \beta) = \Gamma(\alpha)\,\Gamma(\beta)/\Gamma(\alpha+\beta)$ is the Beta function. One can check that $\pi(\theta)$ is normalized on the unit interval: $\int_0^1 d\theta\,\pi(\theta) = 1$ for all positive α, β. Even if we limit ourselves to this form of the prior, different Bayesians might bring different assumptions about the values of α and β. Note that if we choose $\alpha = \beta = 1$, the prior distribution for θ is flat, with $\pi(\theta) = 1$.

We now compute the posterior distribution for θ:

$$\pi(\theta|x_1, \ldots, x_N) = \frac{f(x_1, \ldots, x_N|\theta)\,\pi(\theta)}{\int_0^1 d\theta'\,f(x_1, \ldots, x_N|\theta')\,\pi(\theta')} = \frac{\theta^{X+\alpha-1}(1-\theta)^{N-X+\beta-1}}{\mathsf{B}(X+\alpha, N-X+\beta)} \,. \tag{1.112}$$

Thus, we retain the form of the Beta distribution, but with updated parameters,

$$\begin{aligned} \alpha' &= X + \alpha \\ \beta' &= N - X + \beta \,. \end{aligned} \tag{1.113}$$

The fact that the functional form of the prior is retained by the posterior is generally *not* the case in Bayesian updating. We can also compute the prior predictive,

$$\begin{aligned} f(x_1, \ldots, x_N) &= \int_0^1 d\theta\,f(x_1, \ldots, x_N|\theta)\,\pi(\theta) \\ &= \frac{1}{\mathsf{B}(\alpha, \beta)} \int_0^1 d\theta\,\theta^{X+\alpha-1}(1-\theta)^{N-X+\beta-1} = \frac{\mathsf{B}(X+\alpha, N-X+\beta)}{\mathsf{B}(\alpha, \beta)} \,. \end{aligned} \tag{1.114}$$

The posterior predictive is then

$$\begin{aligned} f(y_1, \ldots, y_M|x_1, \ldots, x_N) &= \int_0^1 d\theta\,f(y_1, \ldots, y_M|\theta)\,\pi(\theta|x_1, \ldots, x_N) \\ &= \frac{1}{\mathsf{B}(X+\alpha, N-X+\beta)} \int_0^1 d\theta\,\theta^{X+Y+\alpha-1}(1-\theta)^{N-X+M-Y+\beta-1} \\ &= \frac{\mathsf{B}(X+Y+\alpha, N-X+M-Y+\beta)}{\mathsf{B}(X+\alpha, N-X+\beta)} \,. \end{aligned} \tag{1.115}$$

1.6.3 Hyperparameters and conjugate priors

In the above example, θ is a *parameter* of the Bernoulli distribution, *i.e.* the likelihood, while quantities α and β are *hyperparameters* which enter the prior $\pi(\theta)$. Accordingly, we could have written $\pi(\theta|\alpha, \beta)$ for

the prior. We then have for the posterior

$$\pi(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{\alpha}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\boldsymbol{\alpha})}{\int\limits_{\Theta} d\boldsymbol{\theta}'\ f(\boldsymbol{x}|\boldsymbol{\theta}')\,\pi(\boldsymbol{\theta}'|\boldsymbol{\alpha})}\ , \tag{1.116}$$

replacing eqn. 1.107, *etc.*, where $\boldsymbol{\alpha} \in A$ is the vector of hyperparameters. The hyperparameters can also be distributed, according to a *hyperprior* $\rho(\boldsymbol{\alpha})$, and the hyperpriors can further be parameterized by *hyperhyperparameters*, which can have their own distributions, *ad nauseum*.

What use is all this? We've already seen a compelling example: when the posterior is of the same form as the prior, the Bayesian update can be viewed as an automorphism of the hyperparameter space A, *i.e.* one set of hyperparameters $\boldsymbol{\alpha}$ is mapped to a new set of hyperparameters $\widetilde{\boldsymbol{\alpha}}$.

> **Definition:** A parametric family of distributions $\mathcal{P} = \{\pi(\boldsymbol{\theta}|\boldsymbol{\alpha})\,|\,\boldsymbol{\theta} \in \Theta,\ \boldsymbol{\alpha} \in A\}$ is called a *conjugate family* for a family of distributions $\{f(\boldsymbol{x}|\boldsymbol{\theta})\,|\,\boldsymbol{x} \in \mathcal{X},\ \boldsymbol{\theta} \in \Theta\}$ if, for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\alpha} \in A$,
>
> $$\pi(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{\alpha}) \equiv \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\boldsymbol{\alpha})}{\int\limits_{\Theta} d\boldsymbol{\theta}'\ f(\boldsymbol{x}|\boldsymbol{\theta}')\,\pi(\boldsymbol{\theta}'|\boldsymbol{\alpha})} \in \mathcal{P}\ . \tag{1.117}$$
>
> That is, $\pi(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{\alpha}) = \pi(\boldsymbol{\theta}|\widetilde{\boldsymbol{\alpha}})$ for some $\widetilde{\boldsymbol{\alpha}} \in A$, with $\widetilde{\boldsymbol{\alpha}} = \widetilde{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{x})$.

As an example, consider the conjugate Bayesian analysis of the Gaussian distribution. We assume a likelihood

$$f(\boldsymbol{x}|u,s) = (2\pi s^2)^{-N/2} \exp\left\{ -\frac{1}{2s^2} \sum_{j=1}^{N} (x_j - u)^2 \right\}. \tag{1.118}$$

The parameters here are $\boldsymbol{\theta} = \{u, s\}$. Now consider the prior distribution

$$\pi(u,s|\mu_0,\sigma_0) = (2\pi\sigma_0^2)^{-1/2}\,\exp\left\{ -\frac{(u - \mu_0)^2}{2\sigma_0^2} \right\}. \tag{1.119}$$

Note that the prior distribution is independent of the parameter s and only depends on u and the hyperparameters $\boldsymbol{\alpha} = (\mu_0, \sigma_0)$. We now compute the posterior:

$$\begin{aligned} \pi(u,s|\boldsymbol{x},\mu_0,\sigma_0) &\propto f(\boldsymbol{x}|u,s)\,\pi(u,s|\mu_0,\sigma_0) \\ &= \exp\left\{ -\left(\frac{1}{2\sigma_0^2} + \frac{N}{2s^2}\right)u^2 + \left(\frac{\mu_0}{\sigma_0^2} + \frac{N\langle x\rangle}{s^2}\right)u - \left(\frac{\mu_0^2}{2\sigma_0^2} + \frac{N\langle x^2\rangle}{2s^2}\right) \right\}, \end{aligned} \tag{1.120}$$

with $\langle x\rangle = \frac{1}{N}\sum_{j=1}^{N} x_j$ and $\langle x^2\rangle = \frac{1}{N}\sum_{j=1}^{N} x_j^2$. This is also a Gaussian distribution for u, and after supplying the appropriate normalization one finds

$$\pi(u,s|\boldsymbol{x},\mu_0,\sigma_0) = (2\pi\sigma_1^2)^{-1/2} \exp\left\{ -\frac{(u - \mu_1)^2}{2\sigma_1^2} \right\}\ , \tag{1.121}$$

with

$$\mu_1 = \mu_0 + \frac{N\left(\langle x \rangle - \mu_0\right)\sigma_0^2}{s^2 + N\sigma_0^2}$$
$$\sigma_1^2 = \frac{s^2\sigma_0^2}{s^2 + N\sigma_0^2} \quad . \tag{1.122}$$

Thus, the posterior is among the same family as the prior, and we have derived the update rule for the hyperparameters $(\mu_0, \sigma_0) \to (\mu_1, \sigma_1)$. Note that $\sigma_1 < \sigma_0$, so the updated Gaussian prior is sharper than the original. The updated mean μ_1 shifts in the direction of $\langle x \rangle$ obtained from the data set.

1.6.4 The problem with priors

We might think that the for the coin flipping problem, the flat prior $\pi(\theta) = 1$ is an appropriate initial one, since it does not privilege any value of θ. This prior therefore seems 'objective' or 'unbiased', also called 'uninformative'. But suppose we make a change of variables, mapping the interval $\theta \in [0, 1]$ to the entire real line according to $\zeta = \ln\left[\theta/(1-\theta)\right]$. In terms of the new parameter ζ, we write the prior as $\tilde{\pi}(\zeta)$. Clearly $\pi(\theta)\, d\theta = \tilde{\pi}(\zeta)\, d\zeta$, so $\tilde{\pi}(\zeta) = \pi(\theta)\, d\theta/d\zeta$. For our example, find $\tilde{\pi}(\zeta) = \frac{1}{4}\mathrm{sech}^2(\zeta/2)$, which is not flat. Thus what was uninformative in terms of θ has become very informative in terms of the new parameter ζ. Is there any truly unbiased way of selecting a Bayesian prior?

One approach, advocated by E. T. Jaynes, is to choose the prior distribution $\pi(\boldsymbol{\theta})$ according to the principle of maximum entropy. For continuous parameter spaces, we must first define a parameter space metric so as to be able to 'count' the number of different parameter states. The entropy of a distribution $\pi(\boldsymbol{\theta})$ is then dependent on this metric: $S = -\int d\mu(\boldsymbol{\theta})\, \pi(\boldsymbol{\theta})\, \ln \pi(\boldsymbol{\theta})$.

Another approach, due to Jeffreys, is to derive a parameterization-independent prior from the likelihood $f(\boldsymbol{x}|\boldsymbol{\theta})$ using the so-called *Fisher information matrix*,

$$\begin{aligned} I_{ij}(\boldsymbol{\theta}) &= -\mathbb{E}_\theta\left(\frac{\partial^2 \ln f(\boldsymbol{x}|\boldsymbol{\theta})}{\partial\theta_i\,\partial\theta_j}\right) \\ &= -\int d\boldsymbol{x}\; f(\boldsymbol{x}|\boldsymbol{\theta})\, \frac{\partial^2 \ln f(\boldsymbol{x}|\boldsymbol{\theta})}{\partial\theta_i\,\partial\theta_j} \quad . \end{aligned} \tag{1.123}$$

The *Jeffreys prior* $\pi_{_{\mathrm{J}}}(\boldsymbol{\theta})$ is defined as

$$\pi_{_{\mathrm{J}}}(\boldsymbol{\theta}) \propto \sqrt{\det I(\boldsymbol{\theta})} \quad . \tag{1.124}$$

One can check that the Jeffries prior is invariant under reparameterization. As an example, consider the Bernoulli process, for which $\ln f(\boldsymbol{x}|\theta) = X \ln\theta + (N - X)\ln(1-\theta)$, where $X = \sum_{j=1}^N x_j$. Then

$$-\frac{d^2 \ln p(\boldsymbol{x}|\theta)}{d\theta^2} = \frac{X}{\theta^2} + \frac{N - X}{(1-\theta)^2} \ , \tag{1.125}$$

and since $\mathbb{E}_\theta\, X = N\theta$, we have

$$I(\theta) = \frac{N}{\theta(1-\theta)} \qquad \Rightarrow \qquad \pi_{_{\mathrm{J}}}(\theta) = \frac{1}{\pi}\frac{1}{\sqrt{\theta(1-\theta)}} \ , \tag{1.126}$$

which felicitously corresponds to a Beta distribution with $\alpha = \beta = \frac{1}{2}$. In this example the Jeffries prior turned out to be a conjugate prior, but in general this is not the case.

We can try to implement the Jeffreys procedure for a two-parameter family where each x_j is normally distributed with mean μ and standard deviation σ. Let the parameters be $(\theta_1, \theta_2) = (\mu, \sigma)$. Then

$$- \ln f(\boldsymbol{x}|\boldsymbol{\theta}) = N \ln \sqrt{2\pi} + N \ln \sigma + \frac{1}{2\sigma^2} \sum_{j=1}^{N} (x_j - \mu)^2 \,, \tag{1.127}$$

and the Fisher information matrix is

$$I(\boldsymbol{\theta}) = -\frac{\partial^2 \ln f(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \theta_i \, \partial \theta_j} = \begin{pmatrix} N\sigma^{-2} & \sigma^{-3} \sum_j (x_j - \mu) \\ \sigma^{-3} \sum_j (x_j - \mu) & -N\sigma^{-2} + 3\sigma^{-4} \sum_j (x_j - \mu)^2 \end{pmatrix} . \tag{1.128}$$

Taking the expectation value, we have $\mathbb{E}\,(x_j - \mu) = 0$ and $\mathbb{E}\,(x_j - \mu)^2 = \sigma^2$, hence

$$\mathbb{E}\,I(\boldsymbol{\theta}) = \begin{pmatrix} N\sigma^{-2} & 0 \\ 0 & 2N\sigma^{-2} \end{pmatrix} \tag{1.129}$$

and the Jeffries prior is $\pi_{\mathrm{J}}(\mu, \sigma) \propto \sigma^{-2}$. This is problematic because if we choose a flat metric on the (μ, σ) upper half plane, the Jeffries prior is not normalizable. Note also that the Jeffreys prior no longer resembles a Gaussian, and hence is not a conjugate prior.

Chapter 2

Stochastic Processes

2.1 References

- C. Gardiner, *Stochastic Methods* (4th edition, Springer-Verlag, 2010)
  Very clear and complete text on stochastic methods, with many applications.

- N. G. Van Kampen *Stochastic Processes in Physics and Chemistry* (3rd edition, North-Holland, 2007)
  Another standard text. Very readable, but less comprehensive than Gardiner.

- Z. Schuss, *Theory and Applications of Stochastic Processes* (Springer-Verlag, 2010)
  In-depth discussion of continuous path stochastic processes and connections to partial differential equations.

- R. Mahnke, J. Kaupužs, and I. Lubashevsky, *Physics of Stochastic Processes* (Wiley, 2009)
  Introductory sections are sometimes overly formal, but a good selection of topics.

- A. N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea, 1956)
  The *Urtext* of mathematical probability theory.

2.2 Introduction to Stochastic Processes

A *stochastic process* is one which is partially random, *i.e.* it is not wholly deterministic. Typically the randomness is due to phenomena at the microscale, such as the effect of fluid molecules on a small particle, such as a piece of dust in the air. The resulting motion (called *Brownian motion* in the case of particles moving in a fluid) can be described only in a statistical sense. That is, the full motion of the system is a *functional* of one or more independent random variables. The motion is then described by its averages with respect to the various random distributions.

2.2.1 Diffusion and Brownian motion

Fick's law (1855) is a phenomenological relationship between number current $\boldsymbol{j}$ and number density gradient $\boldsymbol{\nabla} n$, given by $\boldsymbol{j} = -D\,\boldsymbol{\nabla} n$. Combining this with the continuity equation $\partial_t n + \boldsymbol{\nabla}\cdot\boldsymbol{j}$, one arrives at the diffusion equation[1],

$$\frac{\partial n}{\partial t} = \boldsymbol{\nabla}\cdot(D\,\boldsymbol{\nabla} n)\;. \tag{2.1}$$

Note that the diffusion constant D may be position-dependent. The applicability of Fick's law was experimentally verified in many different contexts and has applicability to a wide range of transport phenomena in physics, chemistry, biology, ecology, geology, *etc.*

The eponymous Robert Brown, a botanist, reported in 1827 on the random motions of pollen grains suspended in water, which he viewed through a microscope. Apparently this phenomenon attracted little attention until the work of Einstein (1905) and Smoluchowski (1906), who showed how it is described by kinetic theory, in which the notion of randomness is essential, and also connecting it to Fick's laws of diffusion. Einstein began with the ideal gas law for osmotic pressure, $p = n k_{\mathrm{B}} T$. In steady state, the osmotic force per unit volume acting on the solute (*e.g.* pollen in water), $-\boldsymbol{\nabla} p$, must be balanced by viscous forces. Assuming the solute consists of spherical particles of radius a, the viscous force per unit volume is given by the hydrodynamic *Stokes drag* per particle $\boldsymbol{F} = -6\pi\eta a\boldsymbol{v}$ times the number density n, where η is the dynamical viscosity of the solvent. Thus, $\boldsymbol{j} = n\boldsymbol{v} = -D\,\boldsymbol{\nabla} n$, where $D = k_{\mathrm{B}} T/6\pi a\eta$.

To connect this to kinetic theory, Einstein reasoned that the solute particles were being buffeted about randomly by the solvent, and he treated this problem statistically. While a given pollen grain is not significantly effected by any single collision with a water molecule, after some characteristic microscopic time τ the grain has effectively forgotten it initial conditions. Assuming there are no global currents, on average each grain's velocity is zero. Einstein posited that over an interval τ, the number of grains which move a distance within $d^3\Delta$ of $\boldsymbol{\Delta}$ is $n\,\phi(\boldsymbol{\Delta})\,d^3\Delta$, where $\phi(\boldsymbol{\Delta}) = \phi(|\boldsymbol{\Delta}|)$ is isotropic and also normalized according to $\int d^3\Delta\,\phi(\Delta) = 1$. Then

$$n(\boldsymbol{x}, t+\tau) = \int d^3\Delta\, n(\boldsymbol{x}-\boldsymbol{\Delta}, t)\,\phi(\boldsymbol{\Delta})\;, \tag{2.2}$$

Taylor expanding in both space and time, to lowest order in τ one recovers the diffusion equation,

---

[1] The equation $\boldsymbol{j} = -D\,\boldsymbol{\nabla} n$ is sometimes called Fick's first law, and the continuity equation $\partial_t n = -\boldsymbol{\nabla}\cdot\boldsymbol{j}$ Fick's second law.

$\partial_t n = D \, \nabla^2 n$, where the diffusion constant is given by

$$D = \frac{1}{6\tau} \int d^3 \Delta \, \phi(\Delta) \, \Delta^2 \ . \tag{2.3}$$

The diffusion equation with constant D is easily solved by taking the spatial Fourier transform. One then has, in d spatial dimensions,

$$\frac{\partial \hat{n}(\mathbf{k}, t)}{\partial t} = -D k^2 \hat{n}(\mathbf{k}, t) \qquad \Rightarrow \qquad n(\mathbf{x}, t) = \int \frac{d^d k}{(2\pi)^d} \, \hat{n}(\mathbf{k}, t_0) \, e^{-D k^2 (t - t_0)} \, e^{i \mathbf{k} \cdot \mathbf{x}} \ . \tag{2.4}$$

If $n(\mathbf{x}, t_0) = \delta(\mathbf{x} - \mathbf{x}_0)$, corresponding to $\hat{n}(\mathbf{k}, t_0) = e^{-i \mathbf{k} \cdot \mathbf{x}_0}$, we have

$$n(\mathbf{x}, t) = \left(4\pi D |t - t_0|\right)^{-d/2} \exp\left\{ -\frac{(\mathbf{x} - \mathbf{x}_0)^2}{4D |t - t_0|} \right\} \ , \tag{2.5}$$

where d is the dimension of space.

WTF just happened?

We're so used to diffusion processes that most of us overlook a rather striking aspect of the above solution to the diffusion equation. At $t = t_0$, the probability density is $P(\mathbf{x}, t = t_0) = \delta(\mathbf{x} - \mathbf{x}_0)$, which means all the particles are sitting at $\mathbf{x} = \mathbf{x}_0$. For any $t > t_0$, the solution is given by Eqn. 2.5, which is nonzero for all $\mathbf{x}$. If we take a value of $\mathbf{x}$ such that $|\mathbf{x} - \mathbf{x}_0| > ct$, where c is the speed of light, we see that there is a finite probability, however small, for particles to diffuse at superluminal speeds. Clearly this is nonsense. The error lies in the diffusion equation itself, which does not recognize any limiting propagation speed. For most processes, this defect is harmless, as we are not interested in the extreme tails of the distribution. Diffusion phenomena and the applicability of the diffusion equation are well-established in virtually every branch of science. To account for a finite propagation speed, one is forced to consider various generalizations of the diffusion equation. Some examples are discussed in the appendix §2.7.

2.2.2 Langevin equation

Consider a particle of mass M subjected to dissipative and random forcing. We'll examine this system in one dimension to gain an understanding of the essential physics. We write

$$\dot{u} + \gamma u = \frac{F}{M} + \eta(t) \ . \tag{2.6}$$

Here, u is the particle's velocity, γ is the damping rate due to friction, F is a constant external force, and $\eta(t)$ is a *stochastic random force*. This equation, known as the Langevin equation, describes a ballistic particle being buffeted by random forcing events[2]. Think of a particle of dust as it moves in the atmosphere. F would then represent the external force due to gravity and $\eta(t)$ the random forcing due to

---

[2]See the appendix in §2.8 for the solution of the Langevin equation for a particle in a harmonic well.

interaction with the air molecules. For a sphere of radius a moving in a fluid of dynamical viscosity η, hydrodynamics gives $\gamma = 6\pi\eta a/M$, where M is the mass of the particle. It is illustrative to compute γ in some setting. Consider a micron sized droplet ($a = 10^{-4}$ cm) of some liquid of density $\rho \sim 1.0\,\mathrm{g/cm}^3$ moving in air at $T = 20^\circ$ C. The viscosity of air is $\eta = 1.8 \times 10^{-4}\,\mathrm{g/cm \cdot s}$ at this temperature[3]. If the droplet density is constant, then $\gamma = 9\eta/2\rho a^2 = 8.1 \times 10^4\,\mathrm{s}^{-1}$, hence the time scale for viscous relaxation of the particle is $\tau = \gamma^{-1} = 12\,\mu$s. We should stress that the viscous damping on the particle is of course due to the fluid molecules, in some average 'coarse-grained' sense. The random component to the force $\eta(t)$ would then represent the fluctuations with respect to this average.

We can easily integrate this equation:

$$\frac{d}{dt}\left(u\,e^{\gamma t}\right) = \frac{F}{M}\,e^{\gamma t} + \eta(t)\,e^{\gamma t}$$

$$u(t) = u(0)\,e^{-\gamma t} + \frac{F}{\gamma M}\left(1 - e^{-\gamma t}\right) + \int\limits_{0}^{t} ds\,\eta(s)\,e^{\gamma(s-t)} \tag{2.7}$$

Note that $u(t)$ is indeed a functional of the random function $\eta(t)$. We can therefore only compute averages in order to describe the motion of the system.

The first average we will compute is that of v itself. In so doing, we assume that $\eta(t)$ has zero mean: $\langle\eta(t)\rangle = 0$. Then

$$\langle u(t)\rangle = u(0)\,e^{-\gamma t} + \frac{F}{\gamma M}\left(1 - e^{-\gamma t}\right). \tag{2.8}$$

On the time scale γ^{-1}, the initial conditions $u(0)$ are effectively forgotten, and asymptotically for $t \gg \gamma^{-1}$ we have $\langle u(t)\rangle \to F/\gamma M$, which is the terminal momentum.

Next, consider

$$\langle u^2(t)\rangle = \langle u(t)\rangle^2 + \int\limits_{0}^{t} ds_1 \int\limits_{0}^{t} ds_2\,e^{\gamma(s_1-t)}\,e^{\gamma(s_2-t)}\,\langle\eta(s_1)\,\eta(s_2)\rangle. \tag{2.9}$$

We now need to know the two-time correlator $\langle\eta(s_1)\,\eta(s_2)\rangle$. We assume that the correlator is a function only of the time difference $\Delta s = s_1 - s_2$, and that the random force $\eta(s)$ has zero average, $\langle\eta(s)\rangle = 0$, and autocorrelation

$$\langle\eta(s_1)\,\eta(s_2)\rangle = \phi(s_1 - s_2). \tag{2.10}$$

The function $\phi(s)$ is the *autocorrelation function* of the random force. A macroscopic object moving in a fluid is constantly buffeted by fluid particles over its entire perimeter. These different fluid particles are almost completely uncorrelated, hence $\phi(s)$ is basically nonzero except on a very small time scale τ_ϕ, which is the time a single fluid particle spends interacting with the object. We can take $\tau_\phi \to 0$ and approximate

$$\phi(s) \approx \Gamma\,\delta(s). \tag{2.11}$$

We shall determine the value of Γ from equilibrium thermodynamic considerations below.

---

[3]The cgs unit of viscosity is the *Poise* (P). $1\,\mathrm{P} = 1\,\mathrm{g/cm \cdot s}$.

With this form for $\phi(s)$, we can easily calculate the equal time momentum autocorrelation:

$$
\begin{aligned}
\langle u^2(t)\rangle &= \langle u(t)\rangle^2 + \Gamma \int\limits_0^t ds\, e^{2\gamma(s-t)} \\
&= \langle u(t)\rangle^2 + \frac{\Gamma}{2\gamma}\left(1 - e^{-2\gamma t}\right) .
\end{aligned}
\tag{2.12}
$$

Consider the case where $F = 0$ and the limit $t \gg \gamma^{-1}$. We demand that the object thermalize at temperature T. Thus, we impose the condition

$$
\left\langle \tfrac{1}{2}Mu^2(t) \right\rangle = \tfrac{1}{2}k_{\mathrm{B}}T \qquad \Longrightarrow \qquad \Gamma = \frac{2\gamma k_{\mathrm{B}}T}{M} \quad .
\tag{2.13}
$$

This fixes the value of Γ.

We can now compute the general momentum autocorrelator:

$$
\begin{aligned}
\langle u(t)\, u(t')\rangle - \langle u(t)\rangle\langle u(t')\rangle &= \int\limits_0^t ds \int\limits_0^{t'} ds'\, e^{\gamma(s-t)}\, e^{\gamma(s'-t')}\, \langle \eta(s)\,\eta(s')\rangle \\
&= \frac{\Gamma}{2\gamma}\, e^{-\gamma|t-t'|} \qquad (t, t' \to \infty\,,\ |t-t'|\ \text{finite}) .
\end{aligned}
\tag{2.14}
$$

Let's now compute the position $x(t)$. We find

$$
x(t) = \langle x(t)\rangle + \frac{1}{M} \int\limits_0^t ds \int\limits_0^s ds_1\, \eta(s_1)\, e^{\gamma(s_1 - s)} ,
\tag{2.15}
$$

where

$$
\langle x(t)\rangle = x(0) + \frac{1}{\gamma}\left(u(0) - \frac{F}{\gamma M}\right)\left(1 - e^{-\gamma t}\right) + \frac{Ft}{\gamma M} .
\tag{2.16}
$$

Note that for $\gamma t \ll 1$ we have $\langle x(t)\rangle = x(0) + u(0)\,t + \tfrac{1}{2}M^{-1}Ft^2 + \mathcal{O}(t^3)$, as is appropriate for ballistic particles moving under the influence of a constant force. This long time limit of course agrees with our earlier evaluation for the terminal velocity, $\langle u(\infty)\rangle = F/\gamma M$. We next compute the position autocorrelation:

$$
\begin{aligned}
\langle x(t)\, x(t')\rangle - \langle x(t)\rangle\langle x(t')\rangle &= \frac{1}{M^2} \int\limits_0^t ds \int\limits_0^{t'} ds'\, e^{-\gamma(s+s')} \int\limits_0^s ds_1 \int\limits_0^{s'} ds_1'\, e^{\gamma(s_1+s_2)}\, \langle \eta(s_1)\,\eta(s_2)\rangle \\
&= \frac{2k_{\mathrm{B}}T}{\gamma M}\,\mathrm{min}(t,t') + \mathcal{O}(1) .
\end{aligned}
$$

In particular, the equal time autocorrelator is

$$
\langle x^2(t)\rangle - \langle x(t)\rangle^2 = \frac{2k_{\mathrm{B}}T\,t}{\gamma M} \equiv 2D\,t ,
\tag{2.17}
$$

at long times, up to terms of order unity. Here, $D = \Gamma/2\gamma^2 = k_{\mathrm{B}}T/\gamma M$ is the *diffusion constant*. For a liquid droplet of radius $a = 1\,\mu\mathrm{m}$ moving in air at $T = 293\,\mathrm{K}$, for which $\eta = 1.8 \times 10^{-4}\,\mathrm{P}$, we have

$$D = \frac{k_{\mathrm{B}}T}{6\pi\eta a} = \frac{(1.38 \times 10^{-16}\,\mathrm{erg/K})\,(293\,\mathrm{K})}{6\pi\,(1.8 \times 10^{-4}\,\mathrm{P})\,(10^{-4}\,\mathrm{cm})} = 1.19 \times 10^{-7}\,\mathrm{cm}^2/\mathrm{s}\,. \tag{2.18}$$

This result presumes that the droplet is large enough compared to the intermolecular distance in the fluid that one can adopt a continuum approach and use the Navier-Stokes equations, and then assuming a laminar flow.

If we consider molecular diffusion, the situation is quite a bit different. The diffusion constant is then $D = \ell^2/2\tau$, where ℓ is the mean free path and τ is the collision time. Elementary kinetic theory gives that the mean free path ℓ, collision time τ, number density n, and total scattering cross section σ are related by[4] $\ell = \bar{v}\tau = 1/\sqrt{2}\,n\sigma$, where $\bar{v} = \sqrt{8k_{\mathrm{B}}T/\pi m}$ is the average particle speed. Approximating the particles as hard spheres, we have $\sigma = 4\pi a^2$, where a is the hard sphere radius. At $T = 293\,\mathrm{K}$, and $p = 1\,\mathrm{atm}$, we have $n = p/k_{\mathrm{B}}T = 2.51 \times 10^{19}\,\mathrm{cm}^{-3}$. Since air is predominantly composed of N_2 molecules, we take $a = 1.90 \times 10^{-8}\,\mathrm{cm}$ and $m = 28.0\,\mathrm{amu} = 4.65 \times 10^{-23}\,\mathrm{g}$, which are appropriate for N_2. We find an average speed of $\bar{v} = 471\,\mathrm{m/s}$ and a mean free path of $\ell = 6.21 \times 10^{-6}\,\mathrm{cm}$. Thus, $D = \frac{1}{2}\ell\bar{v} = 0.146\,\mathrm{cm}^2/\mathrm{s}$. Though much larger than the diffusion constant for large droplets, this is still too small to explain common experiences. Suppose we set the characteristic distance scale at $d = 10\,\mathrm{cm}$ and we ask how much time a point source would take to diffuse out to this radius. The answer is $\Delta t = d^2/2D = 343\,\mathrm{s}$, which is between five and six minutes. Yet if someone in the next seat emits a foul odor, you detect the offending emission in on the order of a second. What this tells us is that diffusion isn't the only transport process involved in these and like phenomena. More important are *convection* currents which distribute the scent much more rapidly.

2.3 Distributions and Functionals

2.3.1 Basic definitions

Let $x \in \mathbb{R}$ be a random variable, and $P(x)$ a probability distribution for x. The average of any function $\phi(x)$ is then

$$\langle \phi(x) \rangle = \int\limits_{-\infty}^{\infty} dx\, P(x)\, \phi(x) \left/ \int\limits_{-\infty}^{\infty} dx\, P(x) \right. . \tag{2.19}$$

Let $\eta(t)$ be a random *function* of t, with $\eta(t) \in \mathbb{R}$, and let $P\big[\eta(t)\big]$ be the probability distribution *functional* for $\eta(t)$. Then if $\Phi\big[\eta(t)\big]$ is a functional of $\eta(t)$, the average of Φ is given by

$$\int D\eta\, P\big[\eta(t)\big]\, \Phi\big[\eta(t)\big] \left/ \int D\eta\, P\big[\eta(t)\big] \right. \tag{2.20}$$

---

[4]The scattering time τ is related to the particle density n, total scattering cross section σ, and mean speed $\bar{v}$ through the relation $n\sigma\bar{v}_{\mathrm{rel}}\tau = 1$, which says that on average one scattering event occurs in a cylinder of cross section σ and length $\bar{v}_{\mathrm{rel}}\tau$. Here $\bar{v}_{\mathrm{rel}} = \sqrt{2}\,\bar{v}$ is the mean relative speed of a pair of particles.
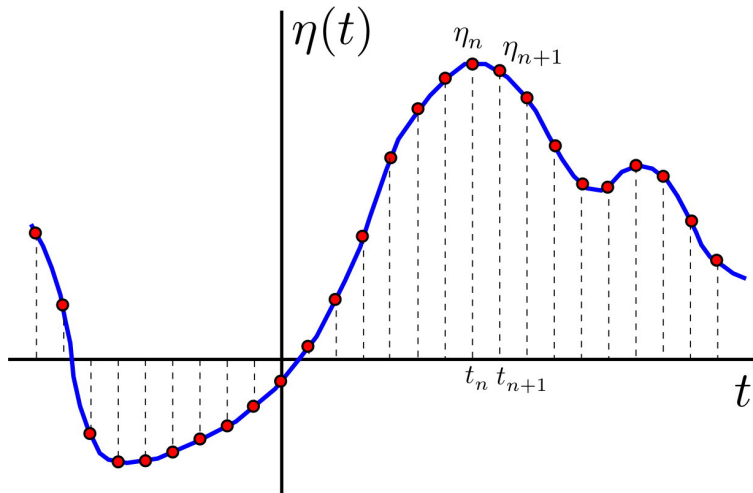
Figure 2.1: Discretization of a continuous function $\eta(t)$. Upon discretization, a functional $\Phi\big[\eta(t)\big]$ becomes an ordinary multivariable function $\Phi(\{\eta_j\})$.

The expression $\int D\eta \, P[\eta] \, \Phi[\eta]$ is a *functional integral*. A functional integral is a continuum limit of a multivariable integral. Suppose $\eta(t)$ were defined on a set of t values $t_n = n\tau$. A functional of $\eta(t)$ becomes a multivariable function of the values $\eta_n \equiv \eta(t_n)$. The metric then becomes $D\eta = \prod_n d\eta_n$.

In fact, for our purposes we will not need to know any details about the functional measure $D\eta$; we will finesse this delicate issue[5]. Consider the *generating functional*,

$$Z\big[J(t)\big] = \int D\eta \, P[\eta] \, \exp\left\{ \int\limits_{-\infty}^{\infty} dt \, J(t)\, \eta(t) \right\} . \tag{2.21}$$

It is clear that

$$\frac{1}{Z[J]} \frac{\delta^n Z[J]}{\delta J(t_1) \cdots \delta J(t_n)}\bigg|_{J(t)=0} = \big\langle \eta(t_1) \cdots \eta(t_n) \big\rangle . \tag{2.22}$$

The function $J(t)$ is an arbitrary *source function*. We functionally differentiate with respect to it in order to find the η-field correlators. The functional derivative $\delta Z\big[J(t)\big]/\delta J(s)$ can be computed by substituting $J(t) \to J(t) + \epsilon\,\delta(t-s)$ inside the functional $Z[J]$, and then taking the ordinary derivative with respect to ε, i.e.

$$\frac{\delta Z\big[J(t)\big]}{\delta J(s)} = \frac{dZ\big[J(t) + \varepsilon\,\delta(t-s)\big]}{d\varepsilon}\bigg|_{\varepsilon=0} . \tag{2.23}$$

Thus the functional derivative $\delta Z\big[J(t)\big]/\delta J(s)$ tells us how the functional $Z[J]$ changes when the function $J(t)$ is replaced by $J(t) + \varepsilon\,\delta(t-s)$. Equivalently, one may eschew this ε prescription and use the familiar chain rule from differential calculus, supplemented by the rule $\delta J(t)\big/\delta J(s) = \delta(t-s)$.

---

[5]A discussion of measure for functional integrals is found in R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*.

Let's compute the generating functional for a class of distributions of the Gaussian form,

$$P[\eta] = \exp\left\{-\frac{1}{2\Gamma}\int_{-\infty}^{\infty} dt\,\left(\tau^2\,\dot{\eta}^2 + \eta^2\right)\right\}$$

$$= \exp\left\{-\frac{1}{2\Gamma}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\left(1 + \omega^2\tau^2\right)|\hat{\eta}(\omega)|^2\right\}\;.$$

(2.24)

Then Fourier transforming the source function $J(t)$, it is easy to see that

$$Z[J] = Z[0]\cdot\exp\left\{\frac{\Gamma}{2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\frac{|\hat{J}(\omega)|^2}{1 + \omega^2\tau^2}\right\}\;.$$

(2.25)

Note that with $\eta(t)\in\mathbb{R}$ and $J(t)\in\mathbb{R}$ we have $\eta^*(\omega) = \eta(-\omega)$ and $\hat{J}^*(\omega) = \hat{J}(-\omega)$. Transforming back to real time, we have

$$Z[J] = Z[0]\cdot\exp\left\{\frac{1}{2}\int_{-\infty}^{\infty}dt\int_{-\infty}^{\infty}dt'\,J(t)\,G(t - t')\,J(t')\right\}\;,$$

(2.26)

where

$$G(s) = \frac{\Gamma}{2\tau}\,e^{-|s|/\tau}\qquad,\qquad \widehat{G}(\omega) = \frac{\Gamma}{1 + \omega^2\tau^2}$$

(2.27)

is the *Green's function*, in real and Fourier space. Note that

$$\int_{-\infty}^{\infty}ds\,G(s) = \widehat{G}(0) = \Gamma\;.$$

(2.28)

We can now compute

$$\langle\eta(t_1)\,\eta(t_2)\rangle = G(t_1 - t_2)$$

(2.29)

$$\langle\eta(t_1)\,\eta(t_2)\,\eta(t_3)\,\eta(t_4)\rangle = G(t_1 - t_2)\,G(t_3 - t_4) + G(t_1 - t_3)\,G(t_2 - t_4)$$
$$+ G(t_1 - t_4)\,G(t_2 - t_3)\;.$$

(2.30)

The generalization is now easy to prove, and is known as *Wick's theorem*:

$$\langle\eta(t_1)\cdots\eta(t_{2n})\rangle = \sum_{\text{contractions}} G(t_{i_1} - t_{i_2})\cdots G(t_{i_{2n-1}} - t_{i_{2n}})\;,$$

(2.31)

where the sum is over all distinct *contractions* of the sequence $1\cdot 2\cdots 2n$ into products of pairs. How many terms are there? Some simple combinatorics answers this question. Choose the index 1. There are $(2n - 1)$ other time indices with which it can be contracted. Now choose another index. There are $(2n - 3)$ indices with which *that* index can be contracted. And so on. We thus obtain

$$C(n)\equiv\left\{\begin{array}{c}\text{\# of contractions}\\\text{of 1-2-3}\cdots 2n\end{array}\right\} = (2n - 1)(2n - 3)\cdots 3\cdot 1 = \frac{(2n)!}{2^n\,n!}\;.$$

(2.32)

2.3.2 Correlations for the Langevin equation

Now suppose we have the Langevin equation

$$\frac{du}{dt} + \gamma u = \eta(t) \tag{2.33}$$

with $u(0) = 0$. We wish to compute the joint probability density

$$P(u_1, t_1; \ldots; u_N, t_N) = \left\langle \delta\big(u_1 - u(t_1)\big) \cdots \delta\big(u_N - u(t_N)\big) \right\rangle, \tag{2.34}$$

where the average is over all realizations of the random variable $\eta(t)$:

$$\left\langle F\big[\eta(t)\big] \right\rangle = \int D\eta \, P\big[\eta(t)\big] \, F\big[\eta(t)\big] . \tag{2.35}$$

Using the integral representation of the Dirac δ-function, we have

$$P(u_1, t_1; \ldots; u_N, t_N) = \int\limits_0^\infty \frac{d\omega_1}{2\pi} \cdots \int\limits_0^\infty \frac{d\omega_N}{2\pi} \, e^{-i(\omega_1 u_1 + \ldots + \omega_N u_N)} \left\langle e^{i\omega_1 u(t_1)} \cdots e^{i\omega_N u(t_N)} \right\rangle. \tag{2.36}$$

Now integrating the Langevin equation with the initial condition $u(0) = 0$ gives

$$u(t_j) = \int\limits_0^{t_j} dt \, e^{\gamma(t-t_j)} \, \eta(t) , \tag{2.37}$$

and therefore we may write

$$\sum_{j=1}^N \omega_j \, u(t_j) = \int\limits_{-\infty}^\infty dt \, f(t) \, \eta(t) \tag{2.38}$$

with

$$f(t) = \sum_{j=1}^N \omega_j \, e^{\gamma(t-t_j)} \, \Theta(t) \, \Theta(t_j - t) . \tag{2.39}$$

We assume that the random variable $\eta(t)$ is distributed as a Gaussian, with $\left\langle \eta(t) \, \eta(t') \right\rangle = G(t - t')$, as described above. Using our previous results, we may perform the functional integral over $\eta(t)$ to obtain

$$\left\langle \exp i \int\limits_{-\infty}^\infty dt \, f(t) \, \eta(t) \right\rangle = \exp\left\{ -\frac{1}{2} \int\limits_{-\infty}^\infty dt \int\limits_{-\infty}^\infty dt' \, G(t - t') \, f(t) \, f(t') \right\}$$

$$= \exp\left\{ -\frac{1}{2} \sum_{j,j'=1}^N M_{jj'} \, \omega_j \, \omega_{j'} \right\}, \tag{2.40}$$

where $M_{jj'} = M(t_j, t_{j'})$ with

$$M(t, t') = \int\limits_0^t ds \int\limits_0^{t'} ds' \, G(s - s') \, e^{\gamma(s-t)} \, e^{\gamma(s'-t')} . \tag{2.41}$$

We now have

$$
\begin{aligned}
P(u_1, t_1; \ldots; u_N, t_N) &= \int\limits_0^\infty \frac{d\omega_1}{2\pi} \cdots \int\limits_0^\infty \frac{d\omega_N}{2\pi} \, e^{-i(\omega_1 u_1 + \ldots + \omega_N u_N)} \, \exp\left\{ -\tfrac{1}{2} \sum_{j,j'=1}^N M_{jj'}\, \omega_j\, \omega_{j'} \right\} \\
&= \det{}^{-1/2}(2\pi M) \, \exp\left\{ -\tfrac{1}{2} \sum_{j,j'=1}^N M_{jj'}^{-1}\, u_j\, u_{j'} \right\} .
\end{aligned}
\tag{2.42}
$$

In the limit $G(s) = \Gamma\, \delta(s)$, we have

$$
\begin{aligned}
M_{jj'} &= \Gamma \int\limits_0^{\min(t_j, t_{j'})} dt\; e^{2\gamma t}\, e^{-\gamma(t_j + t_{j'})} \\
&= \frac{\Gamma}{2\gamma} \left(e^{-\gamma|t_j - t_{j'}|} - e^{-\gamma(t_j + t_{j'})} \right) .
\end{aligned}
\tag{2.43}
$$

From this and the previous expression, we have, assuming $t_{1,2} \gg \gamma^{-1}$ but making no assumptions about the size of $|t_1 - t_2|$,

$$
P(u_1, t_1) = \sqrt{\frac{\gamma}{\pi\Gamma}}\; e^{-\gamma u_1^2/\Gamma} .
\tag{2.44}
$$

The conditional distribution $P(u_1, t_1 \,|\, u_2, t_2) = P(u_1, t_1; u_2, t_2)/P(u_2, t_2)$ is found to be

$$
P(u_1, t_1 \,|\, u_2, t_2) = \sqrt{\frac{\gamma/\pi\Gamma}{1 - e^{-2\gamma(t_1 - t_2)}}}\; \exp\left\{ -\frac{\gamma}{\Gamma} \cdot \frac{\left(u_1 - e^{-\gamma(t_1 - t_2)}\, u_2\right)^2}{1 - e^{-2\gamma(t_1 - t_2)}} \right\} .
\tag{2.45}
$$

Note that $P(u_1, t_1 \,|\, u_2, t_2)$ tends to $P(u_1, t_1)$ independent of the most recent condition, in the limit $t_1 - t_2 \gg \gamma^{-1}$.

As we shall discuss below, a *Markov process* is one where, at any given time, the statistical properties of the subsequent evolution are fully determined by state of the system at that time. Equivalently, every *conditional probability* depends only on the *most recent condition*. Is $u(t)$ a continuous time Markov process? Yes it is! The reason is that $u(t)$ satisfies a first order differential equation, hence only the initial condition on u is necessary in order to derive its probability distribution at any time in the future. Explicitly, we can compute $P(u_1 t_1 \,|\, u_2 t_2, u_3 t_3)$ and show that it is independent of u_3 and t_3 for $t_1 > t_2 > t_3$. This is true regardless of the relative sizes of $t_j - t_{j+1}$ and γ^{-1}.

While $u(t)$ defines a Markov process, its integral $x(t)$ does not. This is because more information than the initial value of x is necessary in order to integrate forward to a solution at future times. Since $x(t)$ satisfies a second order ODE, its conditional probabilities should in principle depend only on the *two most recent conditions*. We could also consider the evolution of the pair $\varphi = (x, u)$ in *phase space*, writing

$$
\frac{d}{dt} \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -\gamma \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} + \begin{pmatrix} 0 \\ \eta(t) \end{pmatrix} ,
\tag{2.46}
$$

or $\dot{\varphi} = A\varphi + \boldsymbol{\eta}(t)$, where A is the above 2×2 matrix, and the stochastic term $\boldsymbol{\eta}(t)$ has only a lower component. The paths $\varphi(t)$ are also Markovian, because they are determined by a first order set of

coupled ODEs. In the limit where $t_j - t_{j+1} \gg \gamma^{-1}$, $x(t)$ effectively becomes Markovian, because we interrogate the paths on time scales where the separations are such that the particle has 'forgotten' its initial velocity.

2.3.3 General ODEs with random forcing

Now let's make a leap to the general n^{th} order linear autonomous inhomogeneous ODE

$$\mathcal{L}_t\, x(t) = \eta(t) \,, \tag{2.47}$$

where $\eta(t)$ is a random function and where

$$\mathcal{L}_t = a_n \frac{d^n}{dt^n} + a_{n-1} \frac{d^{n-1}}{dt^{n-1}} + \cdots + a_1 \frac{d}{dt} + a_0 \tag{2.48}$$

is an nth order differential operator. We are free, without loss of generality, to choose $a_n = 1$. In the appendix in §2.9 we solve this equation using a Fourier transform method. But if we want to impose a boundary condition at $t = 0$, it is more appropriate to consider a *Laplace* transform.

The Laplace transform $\check{x}(z)$ is obtained from a function $x(t)$ via

$$\check{x}(z) = \int\limits_0^\infty dt\, e^{-zt}\, x(t) \,. \tag{2.49}$$

The inverse transform is given by

$$x(t) = \frac{1}{2\pi i} \int\limits_{c-i\infty}^{c+i\infty} dz\, e^{zt}\, \check{x}(z) \,, \tag{2.50}$$

where the integration contour is a straight line which lies to the right of any singularities of $\check{x}(z)$ in the complex z plane. Now let's take the Laplace transform of Eqn. 2.47. Note that integration by parts yields

$$\int\limits_0^\infty dt\, e^{-zt}\, \frac{df}{dt} = z\check{f}(z) - f(0) \tag{2.51}$$

for any function $f(t)$. Applying this result iteratively, we find that the Laplace transform of Eqn. 2.47 is

$$L(z)\, \check{x}(z) = \check{\eta}(z) + R_0(z) \,, \tag{2.52}$$

where

$$L(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0 \tag{2.53}$$

is an nth order polynomial in z with coefficients a_j for $j \in \{0, \ldots, n\}$, and

$$R_0(z) = a_n\, x^{(n-1)}(0) + \left(za_n + a_{n-1}\right) x^{(n-2)}(0) + \cdots + \left(z^{n-1}a_n + \ldots + a_1\right) x(0) \tag{2.54}$$

and $x^{(k)}(t) = d^k x/dt^k$. We now have

$$\check{x}(z) = \frac{1}{L(z)} \left\{ \check{\eta}(z) + R_0(z) \right\} . \tag{2.55}$$

The formal solution to Eqn. 2.47 is then given by the inverse Laplace transform. One finds

$$x(t) = \int_0^t dt' \, K(t - t') \, \eta(t') + x_{\mathrm{h}}(t) , \tag{2.56}$$

where $x_{\mathrm{h}}(t)$ is a solution to the homogeneous equation $\mathcal{L}_t \, x(t) = 0$, and

$$K(s) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} dz \, \frac{e^{zs}}{L(z)} = \sum_{l=1}^{n} \frac{e^{z_l s}}{L'(z_l)} . \tag{2.57}$$

Note that $K(s)$ vanishes for $s < 0$ because then we can close the contour in the far right half plane. The RHS of the above equation follows from the fundamental theorem of algebra, which allows us to factor $L(z)$ as

$$L(z) = a_n(z - z_1) \cdots (z - z_n) , \tag{2.58}$$

with all the roots z_l lying to the left of the contour. In deriving the RHS of Eqn. 2.57, we assume that all roots are distinct[6]. The general solution to the homogeneous equation is

$$x_{\mathrm{h}}(t) = \sum_{l=1}^{n} A_l \, e^{z_l t} , \tag{2.59}$$

again assuming the roots are nondegenerate[7]. In order that the homogeneous solution not grow with time, we must have $\mathrm{Re}\,(z_l) \leq 0$ for all l.

For example, if $\mathcal{L}_t = \frac{d}{dt} + \gamma$, then $L(z) = z + \gamma$ and $K(s) = e^{-\gamma s}$. If $\mathcal{L}_t = \frac{d^2}{dt^2} + \gamma \frac{d}{dt}$, then $L(z) = z^2 + \gamma z$ and $K(s) = (1 - e^{-\gamma s})/\gamma$.

Let us assume that all the initial derivatives $d^k x(t)/dt^k$ vanish at $t = 0$, hence $x_{\mathrm{h}}(t) = 0$. Now let us compute the generalization of Eqn. 2.36,

$$P(x_1, t_1; \ldots; x_N, t_N) = \int_0^\infty \frac{d\omega_1}{2\pi} \cdots \int_0^\infty \frac{d\omega_N}{2\pi} \, e^{-i(\omega_1 x_1 + \ldots + \omega_N x_N)} \left\langle e^{i\omega_1 x(t_1)} \cdots e^{i\omega_N x(t_N)} \right\rangle$$

$$= \det{}^{-1/2}(2\pi M) \, \exp\left\{ -\tfrac{1}{2} \sum_{j,j'=1}^{N} M_{jj'}^{-1} \, x_j \, x_{j'} \right\} , \tag{2.60}$$

where

$$M(t, t') = \int_0^t ds \int_0^{t'} ds' \, G(s - s') \, K(t - s) \, K(t' - s') , \tag{2.61}$$

---

[6]If two or more roots are degenerate, one can still use this result by first inserting a small spacing ε between the degenerate roots and then taking $\varepsilon \to 0$.

[7]If a particular root z_j appears k times, then one has solutions of the form $e^{z_j t}$, $t\,e^{z_j t}$, $\ldots$ $t^{k-1}\,e^{z_j t}$.

with $G(s - s') = \langle \eta(s) \, \eta(s') \rangle$ as before. For $t \gg \gamma^{-1}$, we have $K(s) = \gamma^{-1}$, and if we take $G(s - s') = \Gamma \, \delta(s - s')$ we obtain $M(t, t') = \Gamma \min(t, t')/\gamma^2 = 2D \min(t, t')$. We then have $P(x, t) = \exp\left(- x^2/4Dt\right)/\sqrt{4\pi Dt}$, as expected.

2.4 The Fokker-Planck Equation

2.4.1 Basic derivation

Suppose $x(t)$ is a stochastic variable. We define the quantity

$$\delta x(t) \equiv x(t + \delta t) - x(t) , \tag{2.62}$$

and we assume

$$\langle \delta x(t) \rangle = F_1\big(x(t)\big) \, \delta t \tag{2.63}$$

$$\langle \big[\delta x(t)\big]^2 \rangle = F_2\big(x(t)\big) \, \delta t \tag{2.64}$$

but $\langle \big[\delta x(t)\big]^n \rangle = \mathcal{O}\big((\delta t)^2\big)$ for $n > 2$. The $n = 1$ term is due to *drift* and the $n = 2$ term is due to *diffusion*. Now consider the conditional probability density, $P(x, t \,|\, x_0, t_0)$, defined to be the probability distribution for $x \equiv x(t)$ given that $x(t_0) = x_0$. The conditional probability density satisfies the composition rule,

$$P(x_2, t_2 \,|\, x_0, t_0) = \int\limits_{-\infty}^{\infty} dx_1 \, P(x_2, t_2 \,|\, x_1, t_1) \, P(x_1, t_1 \,|\, x_0, t_0) , \tag{2.65}$$

for any value of t_1. This is also known as the *Chapman-Kolmogorov equation*. In words, what it says is that the probability density for a particle being at x_2 at time t_2, given that it was at x_0 at time t_0, is given by the product of the probability density for being at x_2 at time t_2 given that it was at x_1 at t_1, multiplied by that for being at x_1 at t_1 given it was at x_0 at t_0, integrated over x_1. This should be intuitively obvious, since if we pick any time $t_1 \in [t_0, t_2]$, then the particle had to be *somewhere* at that time. What is perhaps not obvious is why the conditional probability $P(x_2, t_2 \,|\, x_1, t_1)$ does not also depend on (x_0, t_0). This is so if the system is described by a *Markov process*, about we shall have more to say below in §2.6.1. At any rate, a picture is worth a thousand words: see Fig. 2.2.

Proceeding, we may write

$$P(x, t + \delta t \,|\, x_0, t_0) = \int\limits_{-\infty}^{\infty} dx' \, P(x, t + \delta t \,|\, x', t) \, P(x', t \,|\, x_0, t_0) . \tag{2.66}$$

Now

$$
\begin{aligned}
P(x, t + \delta t \,|\, x', t) &= \big\langle \delta\big(x - \delta x(t) - x'\big) \big\rangle \\
&= \left\{ 1 + \langle \delta x(t) \rangle \frac{d}{dx'} + \tfrac{1}{2} \langle \big[\delta x(t)\big]^2 \rangle \frac{d^2}{dx'^2} + \dots \right\} \delta(x - x') \\
&= \delta(x - x') + F_1(x') \frac{d\,\delta(x - x')}{dx'} \, \delta t + \tfrac{1}{2} F_2(x') \frac{d^2\delta(x - x')}{dx'^2} \, \delta t + \mathcal{O}\big((\delta t)^2\big) ,
\end{aligned}
\tag{2.67}
$$
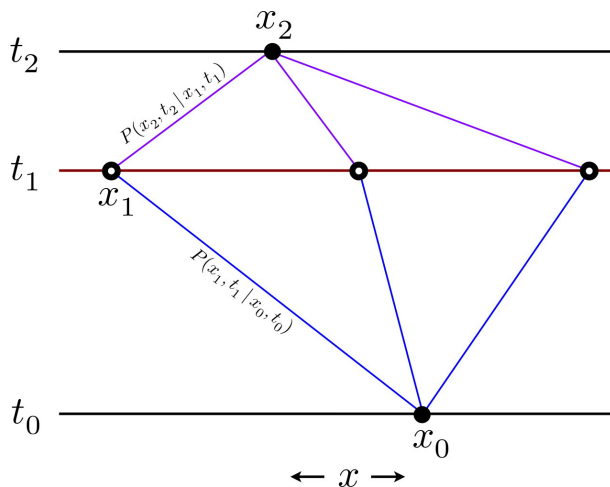
Figure 2.2: Interpretive sketch of the mathematics behind the Chapman-Kolmogorov equation.

where the average is over the random variables. We now insert this result into eqn. 2.66, integrate by parts, divide by δt, and then take the limit $\delta t \to 0$. The result is the Fokker-Planck equation,

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\big[F_1(x)\,P(x,t)\big] + \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\big[F_2(x)\,P(x,t)\big] \; . \tag{2.68}$$

2.4.2 Brownian motion redux

Let's apply our Fokker-Planck equation to a description of Brownian motion. From our earlier results, we have $F_1(x) = F/\gamma M$ and $F_2(x) = 2D$. A formal proof of these results is left as an exercise for the reader. The Fokker-Planck equation is then

$$\frac{\partial P}{\partial t} = -u\,\frac{\partial P}{\partial x} + D\,\frac{\partial^2 P}{\partial x^2} \; , \tag{2.69}$$

where $u = F/\gamma M$ is the average terminal velocity. If we make a Galilean transformation and define $y = x - ut$ and $s = t$, then our Fokker-Planck equation takes the form

$$\frac{\partial P}{\partial s} = D\,\frac{\partial^2 P}{\partial y^2} \; . \tag{2.70}$$

This is known as the *diffusion equation*. Eqn. 2.69 is also a diffusion equation, rendered in a moving frame.

While the Galilean transformation is illuminating, we can easily solve eqn. 2.69 without it. Let's take a

look at this equation after Fourier transforming from x to q:

$$P(x,t) = \int\limits_{-\infty}^{\infty} \frac{dq}{2\pi}\, e^{iqx}\, \hat{P}(q,t) \tag{2.71}$$

$$\hat{P}(q,t) = \int\limits_{-\infty}^{\infty} dx\, e^{-iqx}\, P(x,t)\, . \tag{2.72}$$

Then as should be well known to you by now, we can replace the operator $\frac{\partial}{\partial x}$ with multiplication by iq, resulting in

$$\frac{\partial}{\partial t}\, \hat{P}(q,t) = -(Dq^2 + iqu)\, \hat{P}(q,t)\, , \tag{2.73}$$

with solution

$$\hat{P}(q,t) = e^{-Dq^2 t}\, e^{-iqut}\, \hat{P}(q,0)\, . \tag{2.74}$$

We now apply the inverse transform to get back to x-space:

$$\begin{aligned}
P(x,t) &= \int\limits_{-\infty}^{\infty} \frac{dq}{2\pi}\, e^{iqx}\, e^{-Dq^2 t}\, e^{-iqut} \int\limits_{-\infty}^{\infty} dx'\, e^{-iqx'}\, P(x',0) \\
&= \int\limits_{-\infty}^{\infty} dx'\, P(x',0) \int\limits_{-\infty}^{\infty} \frac{dq}{2\pi}\, e^{-Dq^2 t}\, e^{iq(x-ut-x')} = \int\limits_{-\infty}^{\infty} dx'\, K(x-x',t)\, P(x',0)\, ,
\end{aligned} \tag{2.75}$$

where

$$K(x,t) = \frac{1}{\sqrt{4\pi Dt}}\, e^{-(x-ut)^2/4Dt} \tag{2.76}$$

is the *diffusion kernel*. We now have a recipe for obtaining $P(x,t)$ given the initial conditions $P(x,0)$. If $P(x,0) = \delta(x)$, describing a particle confined to an infinitesimal region about the origin, then $P(x,t) = K(x,t)$ is the probability distribution for finding the particle at x at time t. There are two aspects to $K(x,t)$ which merit comment. The first is that the center of the distribution moves with velocity u. This is due to the presence of the external force. The second is that the standard deviation $\sigma = \sqrt{2Dt}$ is increasing in time, so the distribution is not only shifting its center but it is also getting broader as time evolves. This movement of the center and broadening are what we have called *drift* and *diffusion*, respectively.

2.4.3 Ornstein-Uhlenbeck process

Starting from any initial condition $P(x,0)$, the Fokker-Planck equation for Brownian motion, even with drift, inexorably evolves the distribution $P(x,t)$ toward an infinitesimal probability uniformly spread throughout all space. Consider now the Fokker-Planck equation with $F_2(x) = 2D$ as before, but with $F_1(x) = -\beta x$. Thus we have diffusion but also drift, where the local velocity is $-\beta x$. For $x > 0$, probability which diffuses to the right will also drift to the left, so there is a competition between drift and diffusion. Who wins?

We can solve this model exactly. Starting with the FPE

$$\partial_t P = \partial_x (\beta x P) + D\,\partial_x^2 P \ , \tag{2.77}$$

we first Fourier transform

$$\hat{P}(k,t) = \int\limits_{-\infty}^{\infty} dx\, P(x,t)\, e^{-ikx} \ . \tag{2.78}$$

Expressed in terms of independent variables k and t, one finds that the FPE becomes

$$\partial_t \hat{P} + \beta k\,\partial_k \hat{P} = -Dk^2 \hat{P} \ . \tag{2.79}$$

This is known as a *quasilinear partial differential equation*, and a general method of solution for such equations is the *method of characteristics*, which is briefly reviewed in appendix §2.10. A quasilinear PDE in N independent variables can be transformed into $N+1$ coupled ODEs. Applying the method to Eqn. 2.79, one finds

$$\hat{P}(k,t) = \hat{P}\big(k\,e^{-\beta t}, t=0\big)\ \exp\left\{ -\frac{D}{2\beta}\left(1 - e^{-2\beta t}\right)k^2 \right\} \ . \tag{2.80}$$

Suppose $P(x,0) = \delta(x - x_0)$, in which case $\hat{P}(k,0) = e^{-ikx_0}$. We may now apply the inverse Fourier transform to obtain

$$P(x,t) = \sqrt{\frac{\beta}{2\pi D}\cdot\frac{1}{1 - e^{-2\beta t}}}\ \exp\left\{ -\frac{\beta}{2D}\,\frac{\big(x - x_0\,e^{-\beta t}\big)^2}{1 - e^{-2\beta t}} \right\} \ . \tag{2.81}$$

Taking the limit $t \to \infty$, we obtain the asymptotic distribution

$$P(x, t\to\infty) = \sqrt{\frac{\beta}{2\pi D}}\ e^{-\beta x^2/2D} \ , \tag{2.82}$$

which is a Gaussian centered at $x = 0$, with standard deviation $\sigma = \sqrt{D/\beta}$.

Physically, the drift term $F_1(x) = -\beta x$ arises when the particle is confined to a harmonic well. The equation of motion is then $\ddot{x} + \gamma\dot{x} + \omega_0^2 x = \eta$, which is discussed in the appendix, §2.8. If we average over the random forcing, then setting the acceleration to zero yields the local drift velocity $v_{\text{drift}} = -\omega_0^2\, x/\gamma$, hence $\beta = \omega_0^2/\gamma$. Solving by Laplace transform, one has $L(z) = z^2 + \gamma z + \omega_0^2$, with roots $z_\pm = -\frac{\gamma}{2} \pm \sqrt{\frac{\gamma^2}{4} - \omega_0^2}$, and

$$K(s) = \frac{e^{z_+ s} - e^{z_- s}}{z_+ - z_-}\,\Theta(s) \ . \tag{2.83}$$

Note that $\text{Re}\,(z_\pm) < 0$. Plugging this result into Eqn. 2.61 and integrating, we find

$$\lim_{t\to\infty} M(t,t) = \frac{\gamma\Gamma}{\omega_0^2} \ , \tag{2.84}$$

hence the asymptotic distribution is

$$P(x, t\to\infty) = \sqrt{\frac{\gamma\omega_0^2}{2\pi\Gamma}}\ e^{-\gamma\omega_0^2 x^2/2\Gamma} \ . \tag{2.85}$$

Comparing with Eqn. 2.82, we once again find $D = \Gamma/2\gamma^2$. Does the Langevin particle in a harmonic well describe an Ornstein-Uhlenbeck process for finite t? It does in the limit $\gamma \to \infty$, $\omega_0 \to \infty$, $\Gamma \to \infty$, with $\beta = \omega_0^2/\gamma$ and $D = \Gamma/2\gamma^2$ finite. In this limit, one has $M(t,t) = \beta^{-1}D\left(1 - e^{-\beta t}\right)$. For $\gamma < \infty$, the velocity relaxation time is finite, and on time scales shorter than γ^{-1} the path $x(t)$ is not Markovian.

In the Ornstein-Uhlenbeck model, drift would like to collapse the distribution to a delta-function at $x = 0$, whereas diffusion would like to spread the distribution infinitely thinly over all space. In that sense, both terms represent extremist inclinations. Yet in the limit $t \to \infty$, drift and diffusion gracefully arrive at a grand compromise, with neither achieving its ultimate goal. The asymptotic distribution is centered about $x = 0$, but has a finite width. There is a lesson here for the United States Congress, if only they understood math.

2.5 The Master Equation

Let $P_i(t)$ be the probability that the system is in a quantum or classical state i at time t. Then write

$$\frac{dP_i}{dt} = \sum_j \left(W_{ij}\,P_j - W_{ji}\,P_i\right), \tag{2.86}$$

where W_{ij} is the rate at which j makes a transition to i. This is known as the *Master equation*. Note that we can recast the Master equation in the form

$$\frac{dP_i}{dt} = -\sum_j \Gamma_{ij}\,P_j, \tag{2.87}$$

with

$$\Gamma_{ij} = \begin{cases} -W_{ij} & \text{if } i \neq j \\ \sum_k' W_{kj} & \text{if } i = j, \end{cases} \tag{2.88}$$

where the prime on the sum indicates that $k = j$ is to be excluded. The constraints on the W_{ij} are that $W_{ij} \geq 0$ for all i, j, and we may take $W_{ii} \equiv 0$ (no sum on i). Fermi's Golden Rule of quantum mechanics says that

$$W_{ij} = \frac{2\pi}{\hbar}\left|\langle i|\hat{V}|j\rangle\right|^2 \rho(E_j), \tag{2.89}$$

where $\hat{H}_0|i\rangle = E_i|i\rangle$, $\hat{V}$ is an additional potential which leads to transitions, and $\rho(E_i)$ is the density of final states at energy E_i. The fact that $W_{ij} \geq 0$ means that if each $P_i(t = 0) \geq 0$, then $P_i(t) \geq 0$ for all $t \geq 0$. To see this, suppose that at some time $t > 0$ one of the probabilities P_i is crossing zero and about to become negative. But then eqn. 2.86 says that $\dot{P}_i(t) = \sum_j W_{ij}P_j(t) \geq 0$. So $P_i(t)$ can never become negative.

2.5.1 Equilibrium distribution and detailed balance

If the transition rates W_{ij} are themselves time-independent, then we may formally write

$$P_i(t) = \left(e^{-\Gamma t}\right)_{ij} P_j(0). \tag{2.90}$$

Here we have used the Einstein 'summation convention' in which repeated indices are summed over (in this case, the j index). Note that

$$\sum_i \Gamma_{ij} = 0 \,, \tag{2.91}$$

which says that the total probability $\sum_i P_i$ is conserved:

$$\frac{d}{dt} \sum_i P_i = -\sum_{i,j} \Gamma_{ij} P_j = -\sum_j \left(P_j \sum_i \Gamma_{ij} \right) = 0 \,. \tag{2.92}$$

We conclude that $\vec{\phi} = (1, 1, \ldots, 1)$ is a left eigenvector of Γ with eigenvalue $\lambda = 0$. The corresponding right eigenvector, which we write as P_i^{eq}, satisfies $\Gamma_{ij} P_j^{\mathrm{eq}} = 0$, and is a stationary (*i.e.* time independent) solution to the Master equation. Generally, there is only one right/left eigenvector pair corresponding to $\lambda = 0$, in which case any initial probability distribution $P_i(0)$ converges to P_i^{eq} as $t \to \infty$.

In equilibrium, the net rate of transitions into a state $|\, i \,\rangle$ is equal to the rate of transitions out of $|\, i \,\rangle$. If, for each state $|\, j \,\rangle$ the transition rate from $|\, i \,\rangle$ to $|\, j \,\rangle$ is equal to the transition rate from $|\, j \,\rangle$ to $|\, i \,\rangle$, we say that the rates satisfy the condition of *detailed balance*. In other words,

$$W_{ij}\, P_j^{\mathrm{eq}} = W_{ji}\, P_i^{\mathrm{eq}}. \tag{2.93}$$

Assuming $W_{ij} \neq 0$ and $P_j^{\mathrm{eq}} \neq 0$, we can divide to obtain

$$\frac{W_{ji}}{W_{ij}} = \frac{P_j^{\mathrm{eq}}}{P_i^{\mathrm{eq}}} \,. \tag{2.94}$$

Note that detailed balance is a stronger condition than that required for a stationary solution to the Master equation.

If $\Gamma = \Gamma^{\mathrm{t}}$ is symmetric, then the right eigenvectors and left eigenvectors are transposes of each other, hence $P^{\mathrm{eq}} = 1/N$, where N is the dimension of Γ. The system then satisfies the conditions of detailed balance. See Appendix II (§2.5.3) for an example of this formalism applied to a model of radioactive decay.

2.5.2 Boltzmann's H-theorem

Suppose for the moment that Γ is a symmetric matrix, *i.e.* $\Gamma_{ij} = \Gamma_{ji}$. Then construct the function

$$H(t) = \sum_i P_i(t)\, \ln P_i(t) \,. \tag{2.95}$$

Then

$$\begin{aligned} \frac{dH}{dt} &= \sum_i \frac{dP_i}{dt} \left(1 + \ln P_i \right) = \sum_i \frac{dP_i}{dt} \ln P_i \\ &= -\sum_{i,j} \Gamma_{ij}\, P_j \ln P_i \\ &= \sum_{i,j} \Gamma_{ij}\, P_j \left(\ln P_j - \ln P_i \right) \,, \end{aligned} \tag{2.96}$$

where we have used $\sum_i \Gamma_{ij} = 0$. Now switch $i \leftrightarrow j$ in the above sum and add the terms to get

$$\frac{dH}{dt} = \frac{1}{2} \sum_{i,j} \Gamma_{ij} \left(P_i - P_j \right) \left(\ln P_i - \ln P_j \right) . \tag{2.97}$$

Note that the $i = j$ term does not contribute to the sum. For $i \neq j$ we have $\Gamma_{ij} = -W_{ij} \leq 0$, and using the result

$$(x - y) \left(\ln x - \ln y \right) \geq 0 , \tag{2.98}$$

we conclude

$$\frac{dH}{dt} \leq 0 . \tag{2.99}$$

In equilibrium, P_i^{eq} is a constant, independent of i. We write

$$P_i^{\text{eq}} = \frac{1}{\Omega} \quad , \quad \Omega = \sum_i 1 \quad \Longrightarrow \quad H = -\ln \Omega . \tag{2.100}$$

If $\Gamma_{ij} \neq \Gamma_{ji}$, we can still prove a version of the H-theorem. Define a new symmetric matrix

$$\overline{W}_{ij} \equiv W_{ij} \, P_j^{\text{eq}} = W_{ji} \, P_i^{\text{eq}} = \overline{W}_{ji} , \tag{2.101}$$

and the generalized H-function,

$$H(t) \equiv \sum_i P_i(t) \, \ln \left(\frac{P_i(t)}{P_i^{\text{eq}}} \right) . \tag{2.102}$$

Then

$$\frac{dH}{dt} = -\frac{1}{2} \sum_{i,j} \overline{W}_{ij} \left(\frac{P_i}{P_i^{\text{eq}}} - \frac{P_j}{P_j^{\text{eq}}} \right) \left[\ln \left(\frac{P_i}{P_i^{\text{eq}}} \right) - \ln \left(\frac{P_j}{P_j^{\text{eq}}} \right) \right] \leq 0 . \tag{2.103}$$

2.5.3 Formal solution to the Master equation

Recall the Master equation $\dot{P}_i = -\Gamma_{ij} \, P_j$. The matrix Γ_{ij} is real but not necessarily symmetric. For such a matrix, the left eigenvectors ϕ_i^α and the right eigenvectors ψ_j^β are not the same: general different:

$$\begin{aligned} \phi_i^\alpha \, \Gamma_{ij} &= \lambda_\alpha \, \phi_j^\alpha \\ \Gamma_{ij} \, \psi_j^\beta &= \lambda_\beta \, \psi_i^\beta . \end{aligned} \tag{2.104}$$

Note that the eigenvalue equation for the right eigenvectors is $\Gamma \psi = \lambda \psi$ while that for the left eigenvectors is $\Gamma^{\text{t}} \phi = \lambda \phi$. The characteristic polynomial is the same in both cases:

$$F(\lambda) \equiv \det \left(\lambda - \Gamma \right) = \det \left(\lambda - \Gamma^{\text{t}} \right) , \tag{2.105}$$

which means that the left and right eigenvalues are the same. Note also that $\left[F(\lambda) \right]^* = F(\lambda^*)$, hence the eigenvalues are either real or appear in complex conjugate pairs. Multiplying the eigenvector equation

for ϕ^α on the right by ψ_j^β and summing over j, and multiplying the eigenvector equation for ψ^β on the left by ϕ_i^α and summing over i, and subtracting the two results yields

$$\left(\lambda_\alpha - \lambda_\beta\right)\left\langle\,\phi^\alpha\,\middle|\,\psi^\beta\,\right\rangle = 0\;, \tag{2.106}$$

where the inner product is

$$\left\langle\,\phi\,\middle|\,\psi\,\right\rangle = \sum_i \phi_i\,\psi_i\;. \tag{2.107}$$

We can now demand

$$\left\langle\,\phi^\alpha\,\middle|\,\psi^\beta\,\right\rangle = \delta_{\alpha\beta}\;, \tag{2.108}$$

in which case we can write

$$\Gamma = \sum_\alpha \lambda_\alpha\,\middle|\,\psi^\alpha\,\rangle\langle\,\phi^\alpha\,\middle|\qquad\Longleftrightarrow\qquad \Gamma_{ij} = \sum_\alpha \lambda_\alpha\,\psi_i^\alpha\,\phi_j^\alpha\;. \tag{2.109}$$

We have seen that $\vec{\phi} = (1, 1, \ldots, 1)$ is a left eigenvector with eigenvalue $\lambda = 0$, since $\sum_i \Gamma_{ij} = 0$. We do not know *a priori* the corresponding right eigenvector, which depends on other details of Γ_{ij}. Now let's expand $P_i(t)$ in the right eigenvectors of Γ, writing

$$P_i(t) = \sum_\alpha C_\alpha(t)\,\psi_i^\alpha\;. \tag{2.110}$$

Then

$$\begin{aligned}\frac{dP_i}{dt} &= \sum_\alpha \frac{dC_\alpha}{dt}\,\psi_i^\alpha\\[4pt] &= -\Gamma_{ij}\,P_j = -\sum_\alpha C_\alpha\,\Gamma_{ij}\,\psi_j^\alpha = -\sum_\alpha \lambda_\alpha\,C_\alpha\,\psi_i^\alpha\;,\end{aligned} \tag{2.111}$$

and linear independence of the eigenvectors $|\,\psi^\alpha\,\rangle$ allows us to conclude

$$\frac{dC_\alpha}{dt} = -\lambda_\alpha\,C_\alpha\qquad\Longrightarrow\qquad C_\alpha(t) = C_\alpha(0)\,e^{-\lambda_\alpha t}\;. \tag{2.112}$$

Hence, we can write

$$P_i(t) = \sum_\alpha C_\alpha(0)\,e^{-\lambda_\alpha t}\,\psi_i^\alpha\;. \tag{2.113}$$

It is now easy to see that $\mathrm{Re}\,(\lambda_\alpha) \geq 0$ for all λ, or else the probabilities will become negative. For suppose $\mathrm{Re}\,(\lambda_\alpha) < 0$ for some α. Then as $t \to \infty$, the sum in eqn. 2.113 will be dominated by the term for which λ_α has the largest negative real part; all other contributions will be subleading. But we must have $\sum_i \psi_i^\alpha = 0$ since $|\,\psi^\alpha\,\rangle$ must be orthogonal to the left eigenvector $\vec{\phi}^{\alpha=0} = (1, 1, \ldots, 1)$. Therefore, at least one component of ψ_i^α (*i.e.* for some value of i) must have a negative real part, which means a negative probability![8] As we have already proven that an initial nonnegative distribution $\{P_i(t = 0)\}$ will remain nonnegative under the evolution of the Master equation, we conclude that $P_i(t) \to P_i^{\mathrm{eq}}$ as $t \to \infty$, relaxing to the $\lambda = 0$ right eigenvector, with $\mathrm{Re}\,(\lambda_\alpha) \geq 0\;\forall\,\alpha$.

---

[8]Since the probability $P_i(t)$ is real, if the eigenvalue with the smallest (*i.e.* largest negative) real part is complex, there will be a corresponding complex conjugate eigenvalue, and summing over all eigenvectors will result in a real value for $P_i(t)$.

Poisson process

Consider the *Poisson process*, for which

$$W_{mn} = \begin{cases} \lambda & \text{if } m = n+1 \\ 0 & \text{if } m \neq n+1 \,. \end{cases} \tag{2.114}$$

We then have

$$\frac{dP_n}{dt} = \lambda (P_{n-1} - P_n) \,. \tag{2.115}$$

The *generating function* $P(z,t) = \sum_{n=0}^{\infty} z^n P_n(t)$ then satisfies

$$\frac{\partial P}{\partial t} = \lambda (z-1) P \quad \Rightarrow \quad P(z,t) = e^{(z-1)\lambda t} P(z,0) \,. \tag{2.116}$$

If the initial distribution is $P_n(0) = \delta_{n,0}$, then

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \,, \tag{2.117}$$

which is known as the *Poisson distribution*. If we define $\alpha \equiv \lambda t$, then from $P_n = \alpha^n e^{-\alpha}/n!$ we have

$$\langle n^k \rangle = e^{-\alpha} \left(\alpha \frac{\partial}{\partial \alpha} \right)^k e^{\alpha} \,. \tag{2.118}$$

Thus, $\langle n \rangle = \alpha$, $\langle n^2 \rangle = \alpha^2 + \alpha$, *etc.*

Radioactive decay

Consider a group of atoms, some of which are in an excited state which can undergo nuclear decay. Let $P_n(t)$ be the probability that n atoms are excited at some time t. We then model the decay dynamics by

$$W_{mn} = \begin{cases} 0 & \text{if } m \geq n \\ n\gamma & \text{if } m = n-1 \\ 0 & \text{if } m < n-1 \,. \end{cases} \tag{2.119}$$

Here, γ is the decay rate of an individual atom, which can be determined from quantum mechanics. The Master equation then tells us

$$\frac{dP_n}{dt} = (n+1)\gamma P_{n+1} - n\gamma P_n \,. \tag{2.120}$$

The interpretation here is as follows: let $| n \rangle$ denote a state in which n atoms are excited. Then $P_n(t) = \left| \langle \psi(t) | n \rangle \right|^2$. Then $P_n(t)$ will increase due to spontaneous transitions from $| n+1 \rangle$ to $| n \rangle$, and will decrease due to spontaneous transitions from $| n \rangle$ to $| n-1 \rangle$.

The average number of particles in the system is $N(t) = \sum_{n=0}^{\infty} n P_n(t)$. Note that

$$\frac{dN}{dt} = \sum_{n=0}^{\infty} n \Big[(n+1)\gamma P_{n+1} - n\gamma P_n \Big] = -\gamma \sum_{n=0}^{\infty} n P_n = -\gamma N \,. \tag{2.121}$$

Thus, $N(t) = N(0) \, e^{-\gamma t}$. The relaxation time is $\tau = \gamma^{-1}$, and the equilibrium distribution is $P_n^{\mathrm{eq}} = \delta_{n,0}$, which satisfies detailed balance.

Making use again of the generating function $P(z,t) = \sum_{n=0}^{\infty} z^n \, P_n(t)$, we derive the PDE

$$\frac{\partial P}{\partial t} = \gamma \sum_{n=0}^{\infty} z^n \Big[(n+1) \, P_{n+1} - n \, P_n \Big] = \gamma \, \frac{\partial P}{\partial z} - \gamma z \, \frac{\partial P}{\partial z} \; . \tag{2.122}$$

Thus, we have $\partial_t P = \gamma (1-z) \, \partial_z P$, which is solved by any function $f(\xi)$, where $\xi = \gamma t - \ln(1-z)$. Thus, we can write $P(z,t) = f\big(\gamma t - \ln(1-z)\big)$. Setting $t = 0$ we have $P(z,0) = f\big(-\ln(1-z)\big)$, whence $f(u) = P(1 - e^{-u}, 0)$ is now given in terms of the initial distribution $P(z, t=0)$. Thus, the full solution for $P(z,t)$ is

$$P(z,t) = P\big(1 + (z-1) \, e^{-\gamma t} \, , \, 0\big) \; . \tag{2.123}$$

The total probability is $P(z\!=\!1,t) = \sum_{n=0}^{\infty} P_n$, which clearly is conserved: $P(1,t) = P(1,0)$. The average particle number is then $N(t) = \partial_z \, P(z,t)\big|_{z=1} = e^{-\gamma t} \, P(1,0) = e^{-\gamma t} \, N(0)$.

2.6 Formal Theory of Stochastic Processes

Here we follow the presentation in chapter 3 in the book by C. Gardiner. Given a time-dependent random variable $\boldsymbol{X}(t)$, we define the probability distribution

$$P(\boldsymbol{x},t) = \big\langle \delta\big(\boldsymbol{x} - \boldsymbol{X}(t)\big) \big\rangle \; , \tag{2.124}$$

where the average is over different realizations of the random process. $P(\boldsymbol{x},t)$ is a density with units L^{-d}. This distribution is normalized according to $\int d\boldsymbol{x} \, P(\boldsymbol{x},t) = 1$, where $d\boldsymbol{x} = d^d x$ is the differential for the spatial volume, and does not involve time. If we integrate over some region A, we obtain

$$P_A(t) = \int_A d\boldsymbol{x} \, P(\boldsymbol{x},t) = \text{probability that } \boldsymbol{X}(t) \in A \; . \tag{2.125}$$

We define the *joint probability distributions* as follows:

$$P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2 \, ; \, \ldots \, ; \, \boldsymbol{x}_N, t_N) = \Big\langle \delta\big(\boldsymbol{x}_1 - \boldsymbol{X}(t_1)\big) \cdots \delta\big(\boldsymbol{x}_N - \boldsymbol{X}(t_N)\big) \Big\rangle \; . \tag{2.126}$$

From the joint probabilities we may form *conditional probability distributions*

$$P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2 \, ; \, \ldots \, ; \, \boldsymbol{x}_N, t_N \, | \, \boldsymbol{y}_1, \tau_1 \, ; \, \ldots \, ; \, \boldsymbol{y}_M, \tau_M) = \frac{P(\boldsymbol{x}_1, t_1 \, ; \, \ldots \, ; \, \boldsymbol{x}_N, t_N \, ; \, \boldsymbol{y}_1, \tau_1 \, ; \, \ldots \, ; \, \boldsymbol{y}_M, \tau_M)}{P(\boldsymbol{y}_1, \tau_1 \, ; \, \ldots \, ; \, \boldsymbol{y}_M, \tau_M)} \; . \tag{2.127}$$

Although the times can be in any order, by convention we order them so they decrease from left to right:

$$t_1 > \cdots > t_N > \tau_1 > \cdots \tau_M \; . \tag{2.128}$$

2.6.1 Markov processes

In a *Markov process*, any conditional probability is determined by its *most recent condition*. Thus,

$$P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2 \, ; \, \ldots \, ; \, \boldsymbol{x}_N, t_N \,|\, \boldsymbol{y}_1, \tau_1 \, ; \, \ldots \, ; \, \boldsymbol{y}_M, \tau_M) = P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2 \, ; \, \ldots \, ; \, \boldsymbol{x}_N, t_N \,|\, \boldsymbol{y}_1, \tau_1) \,, \qquad (2.129)$$

where the ordering of the times is as in Eqn. 2.128. This definition entails that *all* probabilities may be constructed from $P(\boldsymbol{x}, t)$ and from the conditional $P(\boldsymbol{x}, t \,|\, \boldsymbol{y}, \tau)$. Clearly $P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2) = P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) \, P(\boldsymbol{x}_2, t_2)$. At the next level, we have

$$\begin{aligned} P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2 \, ; \, \boldsymbol{x}_3, t_3) &= P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2 \, ; \, \boldsymbol{x}_3, t_3) \, P(\boldsymbol{x}_2, t_2 \, ; \, \boldsymbol{x}_3, t_3) \\ &= P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) \, P(\boldsymbol{x}_2, t_2 \,|\, \boldsymbol{x}_3, t_3) \, P(\boldsymbol{x}_3, t_3) \,. \end{aligned}$$

Proceeding thusly, we have

$$P(\boldsymbol{x}_1, t_1 \, ; \, \ldots \, ; \, \boldsymbol{x}_N, t_N) = P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) \, P(\boldsymbol{x}_2, t_2 \,|\, \boldsymbol{x}_3, t_3) \cdots P(\boldsymbol{x}_{N-1}, t_{N-1} \,|\, \boldsymbol{x}_N, t_N) \, P(\boldsymbol{x}_N, t_N) \,, \quad (2.130)$$

so long as $t_1 > t_2 > \ldots > t_N$.

Chapman-Kolmogorov equation

The probability density $P(\boldsymbol{x}_1, t_1)$ can be obtained from the joint probability density $P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2)$ by integrating over $\boldsymbol{x}_2$:

$$P(\boldsymbol{x}_1, t_1) = \int d\boldsymbol{x}_2 \, P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2) = \int d\boldsymbol{x}_2 \, P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) \, P(\boldsymbol{x}_2, t_2) \,. \qquad (2.131)$$

Similarly[9],

$$P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_3, t_3) = \int d\boldsymbol{x}_2 \, P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2 \, ; \, \boldsymbol{x}_3, t_3) \, P(\boldsymbol{x}_2, t_2 \,|\, \boldsymbol{x}_3, t_3) \,. \qquad (2.132)$$

For Markov processes, then,

$$P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_3, t_3) = \int d\boldsymbol{x}_2 \, P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) \, P(\boldsymbol{x}_2, t_2 \,|\, \boldsymbol{x}_3, t_3) \,. \qquad (2.133)$$

For discrete spaces, we have $\int d\boldsymbol{x} \to \sum_{\boldsymbol{x}}$, and $\sum_{\boldsymbol{x}_2} P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) \, P(\boldsymbol{x}_2, t_2 \,|\, \boldsymbol{x}_3, t_3)$ is a matrix multiplication.

Do Markov processes exist in nature and are they continuous?

A random walk in which each step is independently and identically distributed is a Markov process. Consider now the following arrangement. You are given a bag of marbles, an initial fraction p_0 of which are red, q_0 of which are green, and r_0 of which are blue, with $p_0 + q_0 + r_0 = 1$. Let $\sigma_j = +1$, 0, or -1 according to whether the jth marble selected is red, green, or blue, respectively, and define

---

[9]Because $P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_3 \,|\, \boldsymbol{x}_3, t_3) = \big[P(\boldsymbol{x}_1, t_1 \, ; \, \boldsymbol{x}_2, t_2 \, ; \, \boldsymbol{x}_3, t_3) / P(\boldsymbol{x}_2, t_2 \, ; \, \boldsymbol{x}_3, t_3) \big] \cdot \big[P(\boldsymbol{x}_2, t_2 \, ; \, \boldsymbol{x}_3, t_3) / P(\boldsymbol{x}_3, t_3) \big]$.

$X_n = \sum_{j=1}^{n} \sigma_j$, which would correspond to the position of a random walker who steps either to the right ($\sigma_j = +1$), remain stationary ($\sigma_j = 0$), or steps left ($\sigma_j = -1$) during each discrete time interval. If the bag is infinite, then $\{X_1, X_2, \ldots\}$ is a Markov process. The probability for $\sigma_j = +1$ remains at $p = p_0$ and is unaffected by the withdrawal of any finite number of marbles from the bag. But if the contents of the bag are *finite*, then the probability p changes with discrete time, and in such a way that cannot be determined from the instantaneous value of X_n alone. Note that if there were only two colors of marbles, and $\sigma_j \in \{+1, -1\}$, then given $X_0 = 0$ and knowledge of the initial number of marbles in the bag, specifying X_n tells us everything we need to know about the composition of the bag at time n. But with three possibilities $\sigma_j \in \{+1, 0, -1\}$ we need to know the entire history in order to determine the current values of p, q, and r. The reason is that the sequences 0000, $001\bar{1}$, $\bar{1}11\bar{1}$ (with $\bar{1} \equiv -1$) all have the same effect on the displacement X, but result in a different composition of marbles remaining in the bag.

In physical systems, processes we might model as random have a finite correlation time. We saw above that the correlator of the random force $\eta(t)$ in the Langevin equation is written $\langle \eta(t)\,\eta(t+s)\rangle = \phi(s)$, where $\phi(s)$ decays to zero on a time scale τ_ϕ. For time differences $|s| < \tau_\phi$, the system is not Markovian. In addition, the system itself may exhibit some memory. For example, in the Langevin equation $\dot{u} + \gamma u = \eta(t)$, there is a time scale γ^{-1} over which the variable $p(t)$ forgets its previous history. Still, if $\tau_\phi = 0$, $u(t)$ is a Markov process, because the equation is first order and therefore only the most recent condition is necessary in order to integrate forward from some past time $t = t_0$ to construct the statistical ensemble of functions $u(t)$ for $t > t_0$. For second order equations, such as $\ddot{x} + \gamma\dot{x} = \eta(t)$, two initial conditions are required, hence diffusion paths $X(t)$ are only Markovian on time scales beyond γ^{-1}, over which the memory of the initial velocity is lost. More generally, if φ is an N-component vector in phase space, and

$$\frac{d\varphi_i}{dt} = A_i(\boldsymbol{\varphi}, t) + B_{ij}(\boldsymbol{\varphi}, t)\,\eta_j(t)\;, \tag{2.134}$$

where we may choose $\langle \eta_i(t)\,\eta_j(t')\rangle = \delta_{ij}\,\delta(t - t')$, then the path $\varphi(t)$ is a Markov process.

While a random variable $\boldsymbol{X}(t)$ may take values in a continuum, as a function of time it may still exhibit discontinuous jumps. That is to say, even though time t may evolve continuously, the sample paths $\boldsymbol{X}(t)$ may be discontinuous. As an example, consider the Brownian motion of a particle moving in a gas or fluid. On the scale of the autocorrelation time, the velocity changes discontinuously, while the position $\boldsymbol{X}(t)$ evolves continuously (although not smoothly). The condition that sample paths $\boldsymbol{X}(t)$ evolve continuously is known as the *Lindeberg condition*,

$$\lim_{\tau \to 0}\frac{1}{\tau}\int_{|\boldsymbol{x}-\boldsymbol{y}|>\varepsilon} d\boldsymbol{y}\, P(\boldsymbol{y}, t+\tau \,|\, \boldsymbol{x}, t) = 0\;. \tag{2.135}$$

If this condition is satisfied, then the sample paths $\boldsymbol{X}(t)$ are continuous with probability one. Two examples:

(1) **Wiener process**: As we shall discuss below, this is a pure diffusion process with no drift or jumps, with

$$P(x, t \,|\, x', t') = \frac{1}{\sqrt{4\pi D|t - t'|}}\,\exp\!\left(-\frac{(x - x')^2}{4D|t - t'|}\right) \tag{2.136}$$
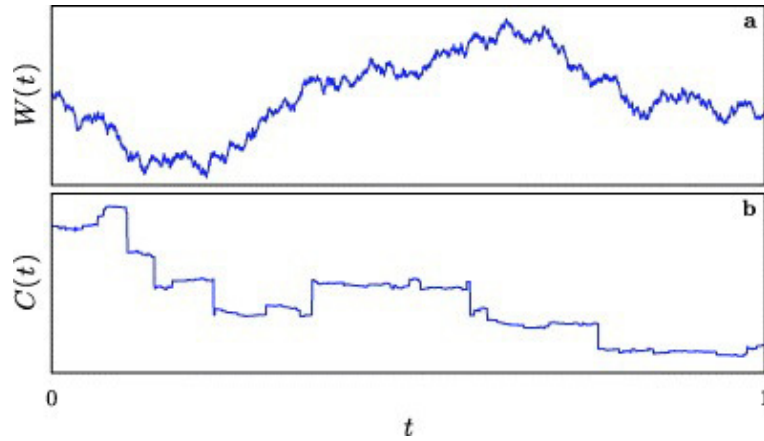
Figure 2.3: (a) Wiener process sample path $W(t)$. (b) Cauchy process sample path $C(t)$. From K. Jacobs and D. A. Steck, *New J. Phys.* **13**, 013016 (2011).

in one space dimension. The Lindeberg condition is satisfied, and the sample paths $X(t)$ are continuous.

(2) **Cauchy process**: This is a process in which sample paths exhibit finite jumps, and hence are not continuous. In one space dimension,

$$P(x, t \,|\, x', t') = \frac{|t - t'|}{\pi\big[(x - x')^2 + (t - t')^2\big]} \ .$$

(2.137)

Note that in both this case and the Wiener process described above, we have $\lim_{t-t' \to 0} P(xt \,|\, x't') = \delta(x - x')$. However in this example the Lindeberg condition is not satisfied.

To simulate, given $x_n = X(t = n\tau)$, choose $y \in \mathcal{D}_b(x_n)$, where $\mathcal{D}_b(x_n)$ is a ball of radius $b > \varepsilon$ centered at x_n. Then evaluate the probability $p \equiv P(y, (n+1)\tau \,|\, x, n\tau)$. If p exceeds a random number drawn from a uniform distribution on $[0, 1]$, accept and set $x_{n+1} = X\big((n+1)\tau\big) = y$. Else reject and choose a new y and proceed as before.

2.6.2 Martingales

A *Martingale* is a stochastic process for which the *conditional average* of the random variable $X(t)$ does not change from its most recent condition. That is,

$$\big\langle x(t) \,\big|\, \{y_1\, \tau_1 \,;\, y_2, \tau_2 \,;\, \ldots \,;\, y_M, \tau_M\} \big\rangle = \int d x\, P(x, t \,|\, y_1, \tau_1 \,;\, \ldots \,;\, y_M, \tau_M)\, x = y_1 \ .$$

(2.138)

In this sense, a Martingale is a stochastic process which represents a 'fair game'. Not every Martingale is a Markov process, and not every Markov process is a Martingale. The Wiener process is a Martingale.

One very important fact about Martingales, which we will here derive in $d = 1$ dimension. For $t_1 > t_2$,

$$
\begin{aligned}
\langle x(t_1)\, x(t_2)\rangle &= \int dx_1 \int dx_2\, P(x_1,t_1\,;\,x_2,t_2)\, x\, x_2 = \int dx_1 \int dx_2\, P(x_1,t_1\,;\,x_2,t_2)\, P(x_2,t_2)\, x_1\, x_2 \\
&= \int dx_2\, P(x_2,t_2)\, x_2 \int dx_1\, P(x_1,t_1\,|\,x_2,t_2)\, x_1 = \int dx_2\, P(x_2,t_2)\, x_2^2 \\
&= \langle x^2(t_2)\rangle\,.
\end{aligned}
\tag{2.139}
$$

One can further show that, for $t_2 > t_2 > t_3$,

$$
\Big\langle \big[x(t_1) - x(t_2)\big]\big[x(t_2) - x(t_3)\big]\Big\rangle = 0\,,
\tag{2.140}
$$

which says that *at the level of pair correlations*, past performance provides no prediction of future results.

2.6.3 Differential Chapman-Kolmogorov equations

Suppose the following conditions apply:

$$
|\boldsymbol{y} - \boldsymbol{x}| > \varepsilon \quad \Longrightarrow \quad \lim_{\tau \to 0} \frac{1}{\tau}\, P(\boldsymbol{y}, t + \tau\,|\,\boldsymbol{x}, t) = W(\boldsymbol{y}\,|\,\boldsymbol{x}, t)
\tag{2.141}
$$

$$
\lim_{\tau \to 0} \frac{1}{\tau} \int_{|\boldsymbol{y}-\boldsymbol{x}|<\varepsilon} d\boldsymbol{y}\, (y_\mu - x_\mu)\, P(\boldsymbol{y}, t + \tau\,|\,\boldsymbol{x}, t) = A_\mu(\boldsymbol{x}, t) + \mathcal{O}(\varepsilon)
\tag{2.142}
$$

$$
\lim_{\tau \to 0} \frac{1}{\tau} \int_{|\boldsymbol{y}-\boldsymbol{x}|<\varepsilon} d\boldsymbol{y}\, (y_\mu - x_\mu)\, (y_\nu - x_\nu)\, P(\boldsymbol{y}, t + \tau\,|\,\boldsymbol{x}, t) = B_{\mu\nu}(\boldsymbol{x}, t) + \mathcal{O}(\varepsilon)\,,
\tag{2.143}
$$

where the last two conditions hold uniformly in $\boldsymbol{x}$, t, and ε. Then following §3.4.1 and §3.6 of Gardiner, one obtains the *forward differential Chapman-Kolmogorov equation* (DCK+),

$$
\begin{aligned}
\frac{\partial P(\boldsymbol{x}, t\,|\,\boldsymbol{x}', t')}{\partial t} &= -\sum_\mu \frac{\partial}{\partial x_\mu}\Big[A_\mu(\boldsymbol{x}, t)\, P(\boldsymbol{x}, t\,|\,\boldsymbol{x}', t')\Big] + \frac{1}{2}\sum_{\mu,\nu} \frac{\partial^2}{\partial x_\mu\, \partial x_\nu}\Big[B_{\mu\nu}(\boldsymbol{x}, t)\, P(\boldsymbol{x}, t\,|\,\boldsymbol{x}', t')\Big] \\
&\quad + \int d\boldsymbol{y}\, \Big[W(\boldsymbol{x}\,|\,\boldsymbol{y}, t)\, P(\boldsymbol{y}, t\,|\,\boldsymbol{x}', t') - W(\boldsymbol{y}\,|\,\boldsymbol{x}, t)\, P(\boldsymbol{x}, t\,|\,\boldsymbol{x}', t')\Big]\,,
\end{aligned}
\tag{2.144}
$$

and the *backward differential Chapman-Kolmogorov equation* (DCK−),

$$
\begin{aligned}
\frac{\partial P(\boldsymbol{x}, t\,|\,\boldsymbol{x}', t')}{\partial t'} &= -\sum_\mu A_\mu(\boldsymbol{x}', t')\, \frac{\partial P(\boldsymbol{x}, t\,|\,\boldsymbol{x}', t')}{\partial x'_\mu} + \frac{1}{2}\sum_{\mu,\nu} B_{\mu\nu}(\boldsymbol{x}', t')\, \frac{\partial^2 P(\boldsymbol{x}, t\,|\,\boldsymbol{x}', t')}{\partial x'_\mu\, \partial x'_\nu} \\
&\quad + \int d\boldsymbol{y}\, W(\boldsymbol{y}\,|\,\boldsymbol{x}', t')\Big[P(\boldsymbol{x}, t\,|\,\boldsymbol{x}', t') - P(\boldsymbol{x}, t\,|\,\boldsymbol{y}, t')\Big]\,.
\end{aligned}
\tag{2.145}
$$

Note that the Lindeberg condition requires that

$$
\lim_{\tau \to 0} \frac{1}{\tau} \int_{|\boldsymbol{x}-\boldsymbol{y}|>\varepsilon} d\boldsymbol{y}\, P(\boldsymbol{y}, t + \tau\,|\,\boldsymbol{x}, t) = \int_{|\boldsymbol{x}-\boldsymbol{y}|>\varepsilon} d\boldsymbol{y}\, W(\boldsymbol{y}\,|\,\boldsymbol{x}, t) = 0\,,
\tag{2.146}
$$

which must hold for any $\varepsilon > 0$. Taking the limit $\varepsilon \to 0$, we conclude[10] $W(\boldsymbol{y} \,|\, \boldsymbol{x}, t) = 0$ *if* the Lindeberg condition is satisfied. If there are any jump processes, *i.e.* if $W(\boldsymbol{y} \,|\, \boldsymbol{x}, t)$ does not identically vanish for all values of its arguments, then Lindeberg is violated, and the paths are discontinuous.

Some applications:

(1) **Master equation**: If $A_\mu(\boldsymbol{x}, t) = 0$ and $B_{\mu\nu}(\boldsymbol{x}, t) = 0$, then we have from DCK+,

$$\frac{\partial P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t')}{\partial t} = \int d\boldsymbol{y} \, \Big[W(\boldsymbol{x} \,|\, \boldsymbol{y}, t) \, P(\boldsymbol{y}, t \,|\, \boldsymbol{x}', t') - W(\boldsymbol{y} \,|\, \boldsymbol{x}, t) \, P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t') \Big] . \tag{2.147}$$

Let's integrate this equation over a time interval Δt. Assuming $P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t) = \delta(\boldsymbol{x} - \boldsymbol{x}')$, we have

$$P(\boldsymbol{x}, t + \Delta t \,|\, \boldsymbol{x}', t) = \Big[1 - \Delta t \int d\boldsymbol{y} \, W(\boldsymbol{y} \,|\, \boldsymbol{x}', t) \Big] \delta(\boldsymbol{x} - \boldsymbol{x}') + W(\boldsymbol{x} \,|\, \boldsymbol{x}', t) \, \Delta t . \tag{2.148}$$

Thus,

$$Q(\boldsymbol{x}', t + \Delta t, t) = 1 - \Delta t \int d\boldsymbol{y} \, W(\boldsymbol{y} \,|\, \boldsymbol{x}', t) \tag{2.149}$$

is the probability for a particle to remain at $\boldsymbol{x}'$ over the interval $[t, t + \Delta t]$ given that it was at $\boldsymbol{x}'$ at time t. Iterating this relation, we find

$$Q(\boldsymbol{x}, t, t_0) = \Big(1 - \Lambda(\boldsymbol{x}, t - \Delta t) \, \Delta t \Big) \Big(1 - \Lambda(\boldsymbol{x}, t - 2\Delta t) \, \Delta t \Big) \cdots \Big(1 - \Lambda(\boldsymbol{x}, t_0) \, \Delta t \Big) \overbrace{Q(\boldsymbol{x}, t_0, t_0)}^{1}$$

$$= \mathrm{P} \exp\Big\{ - \int_{t_0}^{t} dt' \, \Lambda(\boldsymbol{x}, t') \Big\} , \tag{2.150}$$

where $\Lambda(\boldsymbol{x}, t) = \int d\boldsymbol{y} \, W(\boldsymbol{y} \,|\, \boldsymbol{x}, t)$ and P is the path ordering operator which places earlier times to the right.

The interpretation of the function $W(\boldsymbol{y} \,|\, \boldsymbol{x}, t)$ is that it is the probability density *rate* for the random variable $\boldsymbol{X}$ to jump from $\boldsymbol{x}$ to $\boldsymbol{y}$ at time t. Thus, the dimensions of $W(\boldsymbol{y} \,|\, \boldsymbol{x}, t)$ are $L^{-d} T^{-1}$. Such processes are called *jump processes*. For discrete state spaces, the Master equation takes the form

$$\frac{\partial P(\boldsymbol{n}, t \,|\, \boldsymbol{n}', t')}{\partial t} = \sum_{\boldsymbol{m}} \Big[W(\boldsymbol{n} \,|\, \boldsymbol{m}, t) \, P(\boldsymbol{m}, t \,|\, \boldsymbol{n}', t') - W(\boldsymbol{m} \,|\, \boldsymbol{n}, t) \, P(\boldsymbol{n}, t \,|\, \boldsymbol{n}', t') \Big] . \tag{2.151}$$

Here $W(\boldsymbol{n} \,|\, \boldsymbol{m}, t)$ has units T^{-1}, and corresponds to the rate of transitions from state $\boldsymbol{m}$ to state $\boldsymbol{n}$ at time t.

---

[10]What about the case $\boldsymbol{y} = \boldsymbol{x}$, which occurs for $\varepsilon = 0$, which is never actually reached throughout the limiting procedure? The quantity $W(\boldsymbol{x} \,|\, \boldsymbol{x}, t)$ corresponds to the rate at which the system jumps from $\boldsymbol{x}$ to $\boldsymbol{x}$ at time t, which is not a jump process at all. Note that the contribution from $\boldsymbol{y} = \boldsymbol{x}$ cancels from the DCK$\pm$ equations. In other words, we can set $W(\boldsymbol{x} \,|\, \boldsymbol{x}, t) \equiv 0$.

(2) **Fokker-Planck equation**: If $W(\boldsymbol{x} \mid \boldsymbol{y}, t) = 0$, DCK+ gives

$$\frac{\partial P(\boldsymbol{x}, t \mid \boldsymbol{x}', t')}{\partial t} = -\sum_{\mu} \frac{\partial}{\partial x_{\mu}} \Big[A_{\mu}(\boldsymbol{x}, t)\, P(\boldsymbol{x}, t \mid \boldsymbol{x}', t') \Big] + \tfrac{1}{2} \sum_{\mu, \nu} \frac{\partial^2}{\partial x_{\mu}\, \partial x_{\nu}} \Big[B_{\mu\nu}(\boldsymbol{x}, t)\, P(\boldsymbol{x}, t \mid \boldsymbol{x}', t') \Big] \,,$$

(2.152)

which is a more general form of the Fokker-Planck equation we studied in §2.4 above. Defining the average $\langle F(\boldsymbol{x}, t) \rangle = \int d^d x\, F(\boldsymbol{x}, t)\, P(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$, via integration by parts we derive

$$\frac{d}{dt} \langle x_{\mu} \rangle = \langle A_{\mu} \rangle$$

$$\frac{d}{dt} \langle x_{\mu}\, x_{\nu} \rangle = \langle x_{\mu}\, A_{\nu} \rangle + \langle A_{\mu}\, x_{\nu} \rangle + \tfrac{1}{2} \langle D_{\mu\nu} + D_{\nu\mu} \rangle \quad .$$

(2.153)

For the case where $A_{\mu}(\boldsymbol{x}, t)$ and $B_{\mu\nu}(\boldsymbol{x}, t)$ are constants independent of $\boldsymbol{x}$ and t, we have the solution

$$P(\boldsymbol{x}, t \mid \boldsymbol{x}', t') = \det{}^{-1/2} \big[2\pi B\, \Delta t \big] \exp \left\{ -\frac{1}{2\,\Delta t} \left(\Delta x_{\mu} - A_{\mu}\, \Delta t \right) B_{\mu\nu}^{-1} \left(\Delta x_{\nu} - A_{\nu}\, \Delta t \right) \right\} \,, \quad (2.154)$$

where $\Delta \boldsymbol{x} \equiv \boldsymbol{x} - \boldsymbol{x}'$ and $\Delta t \equiv t - t'$. This is normalized so that the integral over $\boldsymbol{x}$ is unity. If we subtract out the *drift* $\boldsymbol{A}\, \Delta t$, then clearly

$$\Big\langle \left(\Delta x_{\nu} - A_{\nu}\, \Delta t \right) \left(\Delta x_{\mu} - A_{\mu}\, \Delta t \right) \Big\rangle = B_{\mu\nu}\, \Delta t \,, \quad (2.155)$$

which is diffusive.

(3) **Liouville equation**: If $W(\boldsymbol{x} \mid \boldsymbol{y}, t) = 0$ and $B_{\mu\nu}(\boldsymbol{x}, t) = 0$, then DCK+ gives

$$\frac{\partial P(\boldsymbol{x}, t \mid \boldsymbol{x}', t')}{\partial t} = -\sum_{\mu} \frac{\partial}{\partial x_{\mu}} \Big[A_{\mu}(\boldsymbol{x}, t)\, P(\boldsymbol{x}, t \mid \boldsymbol{x}', t') \Big] \,. \quad (2.156)$$

This is Liouville's equation from classical mechanics, also known as the continuity equation. Suppressing the $(\boldsymbol{x}', t')$ variables, the above equation is equivalent to

$$\frac{\partial \varrho}{\partial t} + \boldsymbol{\nabla} \cdot (\varrho\, \boldsymbol{v}) = 0 \,, \quad (2.157)$$

where $\varrho(\boldsymbol{x}, t) = P(\boldsymbol{x}, t \mid \boldsymbol{x}', t')$ and $\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{A}(\boldsymbol{x}, t)$. The product of $\boldsymbol{A}$ and P is the current is $\boldsymbol{j} = \varrho \boldsymbol{v}$. To find the general solution, we assume the initial conditions are $P(\boldsymbol{x}, t \mid \boldsymbol{x}', t) = \delta(\boldsymbol{x} - \boldsymbol{x}')$. Then if $\boldsymbol{x}(t; \boldsymbol{x}')$ is the solution to the ODE

$$\frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{A}\big(\boldsymbol{x}(t), t \big) \quad (2.158)$$

with boundary condition $\boldsymbol{x}(t') = \boldsymbol{x}'$, then by applying the chain rule, we see that

$$P(\boldsymbol{x}, t \mid \boldsymbol{x}', t') = \delta\big(\boldsymbol{x} - \boldsymbol{x}(t; \boldsymbol{x}') \big) \quad (2.159)$$

solves the Liouville equation. Thus, the probability density remains a δ-function for all time.

2.6.4 Stationary Markov processes and ergodic properties

Stationary Markov processes satisfy a time translation invariance:

$$P(\boldsymbol{x}_1, t_1\,;\,\ldots\,;\,\boldsymbol{x}_N, t_N) = P(\boldsymbol{x}_1, t_1 + \tau\,;\,\ldots\,;\,\boldsymbol{x}_N, t_N + \tau)\,. \tag{2.160}$$

This means

$$\begin{aligned}
P(\boldsymbol{x}, t) &= P(\boldsymbol{x}) \\
P(\boldsymbol{x}_1, t_1 \,|\, \boldsymbol{x}_2, t_2) &= P(\boldsymbol{x}_1, t_1 - t_2 \,|\, \boldsymbol{x}_2, 0)\,.
\end{aligned} \tag{2.161}$$

Consider the case of one space dimension and define the time average

$$\overline{X}_T \equiv \frac{1}{T} \int\limits_{T/2}^{T/2} dt\, x(t)\,. \tag{2.162}$$

We use a bar to denote time averages and angular brackets $\langle \cdots \rangle$ to denote averages over the randomness. Thus, $\langle \overline{X}_T \rangle = \langle x \rangle$, which is time-independent for a stationary Markov process. The variance of $\overline{X}_T$ is

$$\mathsf{Var}\big(\overline{X}_T\big) = \frac{1}{T^2} \int\limits_{T/2}^{T/2} dt \int\limits_{T/2}^{T/2} dt'\, \big\langle x(t)\, x(t') \big\rangle_{\mathrm{c}}\,, \tag{2.163}$$

where the connected average is $\langle AB \rangle_{\mathrm{c}} = \langle AB \rangle - \langle A \rangle \langle B \rangle$. We define

$$C(t_1 - t_2) \equiv \big\langle x(t_1)\, x(t_2) \big\rangle = \int\limits_{-\infty}^{\infty} dx_1 \int\limits_{-\infty}^{\infty} dx_2\, x_1\, x_2\, P(x_1, t_1\,;\, x_2, t_2)\,. \tag{2.164}$$

If $C(\tau)$ decays to zero sufficiently rapidly with τ, for example as an exponential $e^{-\gamma\tau}$, then $\mathsf{Var}\big(\overline{X}_T\big) \to 0$ as $T \to \infty$, which means that $X_{T \to \infty} = \langle x \rangle$. Thus the time average is the ensemble average, which means the process is *ergodic*.

Wiener-Khinchin theorem

Define the quantity

$$\hat{x}_T(\omega) = \int\limits_{T/2}^{T/2} dt\, x(t)\, e^{i\omega t}\,. \tag{2.165}$$

The *spectral function* $S_T(\omega)$ is given by

$$S_T(\omega) = \left\langle \frac{1}{T} \left| \hat{x}_T(\omega) \right|^2 \right\rangle\,. \tag{2.166}$$

We are interested in the limit $T \to \infty$. Does $S(\omega) \equiv S_{T \to \infty}(\omega)$ exist?

Observe that

$$
\left\langle \left| \hat{x}_T(\omega) \right|^2 \right\rangle = \int\limits_{T/2}^{T/2} dt_1 \int\limits_{T/2}^{T/2} dt_2 \; e^{i\omega(t_2 - t_1)} \; \overbrace{\left\langle x(t_1)\, x(t_2) \right\rangle}^{C(t_1 - t_2)}
$$

$$
= \int\limits_{-T}^{T} d\tau \, e^{-i\omega\tau} \, C(\tau) \left(T - |\tau| \right) .
\tag{2.167}
$$

Thus,

$$
S(\omega) = \lim_{T \to \infty} \int\limits_{-\infty}^{\infty} d\tau \, e^{-i\omega\tau} \, C(\tau) \left(1 - \frac{|\tau|}{T} \right) \Theta(T - |\tau|) = \int\limits_{-\infty}^{\infty} d\tau \, e^{-i\omega\tau} \, C(\tau) .
\tag{2.168}
$$

The second equality above follows from Lebesgue's dominated convergence theorem, which you can look up on Wikipedia[11]. We therefore conclude the limit exists and is given by the Fourier transform of the correlation function $C(\tau) = \left\langle x(t)\, x(t + \tau) \right\rangle$.

2.6.5   Approach to stationary solution

We have seen, for example, how in general an arbitrary initial state of the Master equation will converge exponentially to an equilibrium distribution. For stationary Markov processes, the conditional distribution $P(x, t \,|\, x', t')$ converges to an equilibrium distribution $P_{\mathrm{eq}}(x)$ as $t - t' \to \infty$. How can we understand this convergence in terms of the differential Chapman-Kolmogorov equation? We summarize here the results in §3.7.3 of Gardiner.

Suppose $P_1(x, t)$ and $P_2(x, t)$ are each solutions to the DCK+ equation, and furthermore that $W(x \,|\, x', t)$, $A_\mu(x, t)$, and $B_{\mu\nu}(x, t)$ are all independent of t. Define the *Lyapunov functional*

$$
K[P_1, P_2, t] = \int dx \left(P_1 \ln(P_1/P_2) + P_2 - P_1 \right) .
\tag{2.169}
$$

Since $P_{1,2}(x, t)$ are both normalized, the integrals of the last two terms inside the big round brackets cancel. Nevertheless, it is helpful to express K in this way since, factoring out P_1 from the terms inside the brackets, we may use $f(z) = z - \ln z - 1 \geq 0$ for $z \in \mathbb{R}_+$, where $z = P_2/P_1$. Thus, $K \geq 0$, and the minimum value is obtained for $P_1(x, t) = P_2(x, t)$.

Next, evaluate the time derivative $\dot{K}$:

$$
\frac{dK}{dt} = \int dx \left\{ \frac{\partial P_1}{\partial t} \cdot \left[\ln P_1 - \ln P_2 + 1 \right] - \frac{\partial P_2}{\partial t} \cdot \frac{P_1}{P_2} \right\} .
\tag{2.170}
$$

---

[11]If we define the one parameter family of functions $C_T(\tau) = C(\tau) \left(1 - \frac{|\tau|}{T} \right) \Theta(T - |\tau|)$, then as $T \to \infty$ the function $C_T(\tau)\, e^{-i\omega\tau}$ converges pointwise to $C(\tau)\, e^{-i\omega\tau}$, and if $|C(\tau)|$ is integrable on $\mathbb{R}$, the theorem guarantees the second equality in Eqn. 2.168.

We now use DCK+ to obtain $\partial_t P_{1,2}$ and evaluate the contributions due to drift, diffusion, and jump processes. One finds

$$\left(\frac{dK}{dt}\right)_{\text{drift}} = -\sum_\mu \int d\boldsymbol{x} \, \frac{\partial}{\partial x_\mu} \Big[A_\mu \, P_1 \, \ln\left(P_1/P_2\right) \Big] \tag{2.171}$$

$$\left(\frac{dK}{dt}\right)_{\text{diff}} = -\frac{1}{2}\sum_{\mu,\nu} \int d\boldsymbol{x} \, B_{\mu\nu} \, \frac{\partial \ln(P_1/P_2)}{\partial x_\mu} \, \frac{\partial \ln(P_1/P_2)}{\partial x_\nu} + \frac{1}{2} \int d\boldsymbol{x} \, \frac{\partial^2}{\partial x_\mu \, \partial x_\nu} \Big[B_{\mu\nu} \, P_1 \, \ln(P_1/P_2) \Big] \tag{2.172}$$

$$\left(\frac{dK}{dt}\right)_{\text{jump}} = \int d\boldsymbol{x} \int d\boldsymbol{x}' \, W(\boldsymbol{x}\,|\,\boldsymbol{x}') \, P_2(\boldsymbol{x}',t) \Big[\phi' \ln(\phi/\phi') - \phi + \phi' \Big] \,, \tag{2.173}$$

where $\phi(\boldsymbol{x},t) \equiv P_1(\boldsymbol{x},t)/P_2(\boldsymbol{x},t)$ in the last line. Dropping the total derivative terms, which we may set to zero at spatial infinity, we see that $\dot{K}_{\text{drift}} = 0$, $\dot{K}_{\text{diff}} \le 0$, and $\dot{K}_{\text{jump}} \le 0$. Barring pathological cases[12], one has that $K(t)$ is a nonnegative decreasing function. Since $K = 0$ when $P_1(\boldsymbol{x},t) = P_2(\boldsymbol{x},t) = P_{\text{eq}}(\boldsymbol{x})$, we see that the Lyapunov analysis confirms that K is strictly decreasing. If we set $P_2(\boldsymbol{x},t) = P_{\text{eq}}(\boldsymbol{x})$, we conclude that $P_1(\boldsymbol{x},t)$ converges to $P_{\text{eq}}(\boldsymbol{x})$ as $t \to \infty$.

2.7 Appendix : Nonlinear diffusion

2.7.1 PDEs with infinite propagation speed

Starting from an initial probability density $P(\boldsymbol{x}, t=0) = \delta(\boldsymbol{x})$, we saw how Fickian diffusion, described by the equation $\partial_t P = \boldsymbol{\nabla} \cdot (D \, \boldsymbol{\nabla} P)$, gives rise to the solution

$$P(\boldsymbol{x},t) = (4\pi Dt)^{-d/2} \, e^{-\boldsymbol{x}^2/4Dt} \,, \tag{2.174}$$

for all $t > 0$, assuming D is a constant. As remarked in §2.2.1, this violates any physical limits on the speed of particle propagation, including that set by special relativity, because $P(\boldsymbol{x},t) > 0$ for all $\boldsymbol{x}$ at any finite value of t.

It's perhaps good to step back at this point and recall the solution to the one-dimensional discrete random walk, where after each time increment the walker moves to the right ($\Delta X = 1$) with probability p and to the left ($\Delta X = -1$) with probability $1 - p$. To make things even simpler we'll consider the case with no drift, *i.e.* $p = \frac{1}{2}$. The distribution for X after N time steps is of the binomial form:

$$P_N(X) = 2^{-N} \binom{N}{\frac{1}{2}(N-X)} \,. \tag{2.175}$$

Invoking Stirling's asymptotic result $\ln K! = K \ln K - K + \mathcal{O}(\ln K)$ for $K \gg 1$, one has[13]

$$P_N(X) \simeq \sqrt{\frac{2}{\pi N}} \, e^{-X^2/2N} \,. \tag{2.176}$$

---

[12]See Gardiner, §3.7.3.

[13]The prefactor in this equation seems to be twice the expected $(2\pi N)^{-1/2}$, but since each step results in $\Delta X = \pm 1$, if we start from $X_0 = 0$ then after N steps X will be even if N is even and odd if N is odd. Therefore the continuum limit for the normalization condition on $P_N(X)$ is $\sum_X P_N(X) \approx \frac{1}{2} \int_{-\infty}^{\infty} dX \, P_N(X) = 1$.

We note that the distribution in Eqn. 2.175 is cut off at $|X| = N$, so that $P_N(X) = 0$ for $|X| > N$. This reflects the fact that the walker travels at a fixed speed of one step per time interval. This feature is lost in Eqn. 2.176, because the approximation which led to this result is not valid in the tails of the distribution. One might wonder about the results of §2.3 in this context, since we ultimately obtained a diffusion form for $P(x,t)$ using an exact functional averaging method. However, since we assumed a Gaussian probability functional for the random forcing $\eta(t)$, there is a finite probability for arbitrarily large values of the forcing. For example, consider the distribution of the integrated force $\phi = \int_{t_1}^{t_2} dt\, \eta(t)$:

$$P(\phi, \Delta t) = \left\langle \delta\left(\phi - \int_{t_1}^{t_2} dt\, \eta(t)\right) \right\rangle = \frac{1}{\sqrt{2\pi\Gamma\,\Delta t}}\, e^{-\phi^2/2\Gamma\Delta t}\ , \tag{2.177}$$

where $\Delta t = t_2 - t_1$. This distribution is nonzero for arbitrarily large values of ϕ.

Mathematically, the diffusion equation is an example of what is known as a *parabolic* partial differential equation. The Navier-Stokes equations of hydrodynamics are also parabolic PDEs. The other two classes are called *elliptical* and *hyperbolic*. Paradigmatic examples of these classes include Laplace's equation (elliptical) and the Helmholtz equation (hyperbolic). Hyperbolic equations propagate information at finite propagation speed. For second order PDEs of the form

$$A_{ij}\frac{\partial^2 \Psi}{\partial x_i\,\partial x_j} + B_i\frac{\partial \Psi}{\partial x_i} + C\Psi = S\ \ , \tag{2.178}$$

the PDE is elliptic if the matrix A is positive definite or negative definite, parabolic if A has one zero eigenvalue, and hyperbolic if A is nondegenerate and indefinite (*i.e.* one positive and one negative eigenvalue). Accordingly, one way to remedy the unphysical propagation speed in the diffusion equation is to deform it to a hyperbolic PDE such as the *telegrapher's equation*,

$$\tau\,\frac{\partial^2 \Psi}{\partial t^2} + \frac{\partial \Psi}{\partial t} + \gamma\Psi = D\,\frac{\partial^2 \Psi}{\partial x^2}\ . \tag{2.179}$$

When $\gamma = 0$, the solution for the initial condition $\Psi(x,0) = \delta(x)$ is

$$\Psi(x,t) = \frac{1}{\sqrt{4Dt}}\, e^{-t/2\tau}\, I_0\!\left(\sqrt{\left(\frac{t}{2\tau}\right)^2 - \frac{x^2}{4D\tau}}\right) \Theta\!\left(\sqrt{D/\tau}\,t - |x|\right) . \tag{2.180}$$

Note that $\Psi(x,t)$ vanishes for $|x| > ct$, where $c = \sqrt{D/\tau}$ is the maximum propagation speed. One can check that in the limit $\tau \to 0$ one recovers the familiar diffusion kernel.

The telegrapher's equation

To derive the telegrapher's equation, consider the section of a transmission line shown in Fig. 2.4. Let $V(x,t)$ be the electrical potential on the top line, with $V = 0$ on the bottom (*i.e.* ground). Per unit length a, the potential drop along the top line is $\Delta V = a\,\partial_x V = -IR - L\,\partial_t I$, and the current drop is $\Delta I = a\,\partial_x I = -GV - C\,\partial_t V$. Differentiating the first equation with respect to x and using the second for $\partial_x I$, one arrives at Eqn. 2.179 with $\tau = LC/(RC+GL)$, $\gamma = RG/(RC+GL)$, and $D = a^2/(RC+GL)$.

2.7.2 The porous medium and p-Laplacian equations

Another way to remedy this problem with the diffusion equation is to consider some nonlinear extensions thereof[14]. Two such examples have been popular in the mathematical literature, the *porous medium equation* (PME),

$$\frac{\partial u}{\partial t} = \nabla^2\left(u^m\right),\tag{2.181}$$

and the *p-Laplacian equation*,

$$\frac{\partial u}{\partial t} = \boldsymbol{\nabla}\cdot\left(|\boldsymbol{\nabla}u|^{p-2}\,\boldsymbol{\nabla}u\right).\tag{2.182}$$

Both these equations introduce a nonlinearity whereby the diffusion constant D depends on the field u. For example, the PME can be rewritten $\partial_t u = \boldsymbol{\nabla}\cdot\left(m\,u^{m-1}\boldsymbol{\nabla}u\right)$, whence $D = m\,u^{m-1}$. For the p-Laplacian equation, $D = |\boldsymbol{\nabla}u|^{p-2}$. These nonlinearities strangle the diffusion when u or $|\boldsymbol{\nabla}u|$ gets small, preventing the solution from advancing infinitely fast.

As its name betokens, the PME describes fluid flow in a porous medium. A fluid moving through a porous medium is described by three fundamental equations:

(i) *Continuity:* In a medium with porosity ε, the continuity equation becomes $\varepsilon\,\partial_t\varrho + \nabla\cdot(\varrho v) = 0$, where ϱ is the fluid density. This is because in a volume Ω where the fluid density is changing at a rate $\partial_t\varrho$, the rate of change of fluid mass is $\varepsilon\Omega\,\partial_t\varrho$.

(ii) *Darcy's law:* First articulated in 1856 by the French hydrologist Henry Darcy, this says that the flow velocity is directly proportional to the pressure gradient according to the relation $v = -(K/\mu)\boldsymbol{\nabla}p$, where the *permeability K* depends on the medium but not on the fluid, and μ is the shear viscosity of the fluid.

(iii) *Fluid equation of state:* This is a relation between the pressure p and the density ϱ of the fluid. For ideal gases, $p = A\,\varrho^\gamma$ where A is a constant and $\gamma = c_p/c_V$ is the specific heat ratio.

Putting these three equations together, we obtain

$$\frac{\partial\varrho}{\partial t} = C\,\nabla^2\left(\varrho^m\right),\tag{2.183}$$

where $C = A\gamma k/(k+1)\varepsilon\mu$ and $m = 1 + \gamma$.

---

[14]See J. L. Vazquez, *The Porous Medium Equation* (Oxford, 2006).
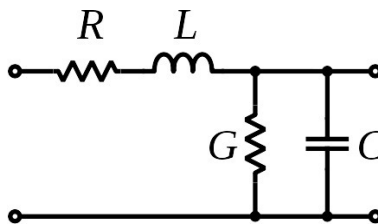
Figure 2.4: Repeating unit of a transmission line. *Credit: Wikipedia*

2.7.3 Illustrative solutions

A class of solution to the PME was discussed in the Russian literature in the early 1950's in a series of papers by Zeldovich, Kompaneets, and Barenblatt. The ZKB solution, which is isotropic in d space dimensions, is of the scaling form,

$$U(r,t) = t^{-\alpha} F\big(r\,t^{-\alpha/d}\big) \qquad ; \qquad F(\xi) = \big(C - k\,\xi^2\big)_{+}^{\frac{1}{m-1}} \quad , \tag{2.184}$$

where $r = |\boldsymbol{x}|$,

$$\alpha = \frac{d}{(m-1)d + 2} \quad , \qquad k = \frac{m-1}{2m}\cdot\frac{1}{(m-1)d+2} \quad , \tag{2.185}$$

and the $+$ subscript in the definition of $F(\xi)$ in Eqn. 2.184 indicates that the function is cut off and vanishes when the quantity inside the round brackets becomes negative. We also take $m > 1$, which means that $\alpha < \frac{1}{2}d$. The quantity C is determined by initial conditions. The scaling form is motivated by the fact that the PME conserves the integral of $u(\boldsymbol{x},t)$ over all space, provided the current $\boldsymbol{j} = -mu^{m-1}\boldsymbol{\nabla} u$ vanishes at spatial infinity. Explicitly, we have

$$\int d^d x\, U(\boldsymbol{x},t) = \Omega_d \int\limits_0^\infty dr\, r^{d-1}\, t^{-\alpha}\, F\big(r\,t^{-\alpha/d}\big) = \Omega_d \int\limits_0^\infty ds\, s^{d-1}\, F(s) \, , \tag{2.186}$$

where Ω_d is the total solid angle in d space dimensions. The above integral is therefore independent of t, which means that the integral of U is conserved. Therefore as $t \to 0$, we must have $U(\boldsymbol{x}, t=0) = A\,\delta(\boldsymbol{x})$, where A is a constant which can be expressed in terms of C, m, and d. We plot the behavior of this solution for the case $m = 2$ and $d = 1$ in Fig. 2.5, and compare and contrast it to the solution of the diffusion equation. Note that the solutions to the PME have compact support, *i.e.* they vanish identically for $r > \sqrt{C/k}\ t^{\alpha/d}$, which is consistent with a finite maximum speed of propagation. A similar point source solution to the p-Laplacian equation in $d = 1$ was obtained by Barenblatt:

$$U(x,t) = t^{-m}\Big(C - k\,|\xi|^{1+m^{-1}}\Big)^{\frac{m}{m-1}} \, , \tag{2.187}$$

for arbitrary $C > 0$, with $\xi = x\,t^{-1/2m}$, and $k = (m-1)(2m)^{-(m+1)/m}$.

To derive the ZKB solution of the porous medium equation, it is useful to write the PME in terms of the 'pressure' variable $v = \frac{m}{m-1}\,u^{m-1}$. The PME then takes the form

$$\frac{\partial v}{\partial t} = (m-1)\,v\,\nabla^2 v + (\boldsymbol{\nabla} v)^2 \, . \tag{2.188}$$

We seek an isotropic solution in d space dimensions, and posit the scaling form

$$V(\boldsymbol{x},t) = t^{-\lambda}\,G\big(r\,t^{-\mu}\big) \, , \tag{2.189}$$

where $r = |\boldsymbol{x}|$. Acting on isotropic functions, the Laplacian is given by $\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{d-1}{r}\frac{\partial}{\partial r}$. Defining $\xi = r\,t^{-\mu}$, we have

$$\frac{\partial V}{\partial t} = -t^{-1}\Big[\lambda\,G + \mu\,\xi\,G'\Big] \quad , \qquad \frac{\partial V}{\partial r} = t^{-(\lambda+\mu)}\,G' \quad , \qquad \frac{\partial^2 V}{\partial r^2} = t^{-(\lambda+2\mu)}\,G'' \quad , \tag{2.190}$$
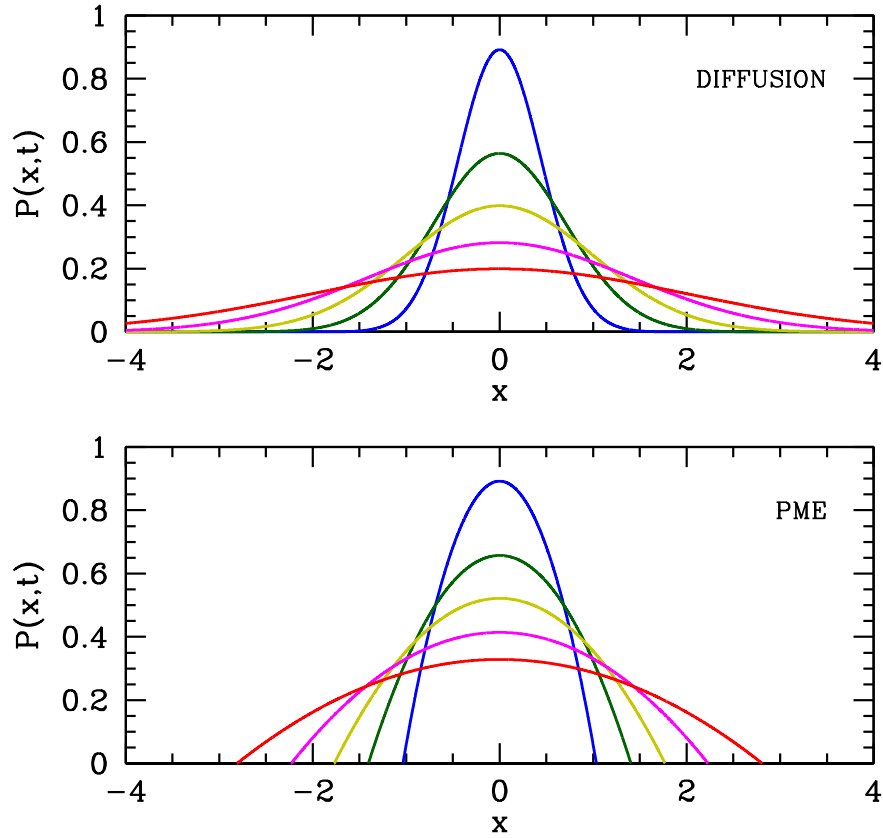
Figure 2.5: Top panel: evolution of the diffusion equation with $D = 1$ and $\sigma = 1$ for times $t = 0.1$, 0.25, 0.5, 1.0, and 2.0. Bottom panel: evolution of the porous medium equation with $m = 2$ and $d = 1$ and C chosen so that $P(x = 0, t = 0.1)$ is equal to the corresponding value in the top panel (*i.e.* the peak of the blue curve).

whence

$$-\left[\lambda G + \mu \xi G'\right] t^{-1} = \left[(m-1) G G'' + (m-1)(d-1) \xi^{-1} G G' + (G')^2\right] t^{-2(\lambda+\mu)} . \qquad (2.191)$$

At this point we can read off the result $\lambda + \mu = \frac{1}{2}$ and eliminate the t variable, which validates our initial scaling form hypothesis. What remains is

$$\lambda G + \mu \xi G' + (m-1) G G'' + (m-1)(d-1) \xi^{-1} G G' + (G')^2 = 0 . \qquad (2.192)$$

Inspection now shows that this equation has a solution of the form $G(\xi) = A - b\,\xi^2$. Plugging this in, we find

$$\lambda = (m-1)\,\alpha \quad , \quad \mu = \frac{\alpha}{d} \quad , \quad b = \frac{\alpha}{2d} \quad , \quad \alpha \equiv \frac{d}{(m-1)\,d + 2} . \qquad (2.193)$$

The quadratic function $G(\xi) = A - b\xi^2$ goes negative for $\xi^2 > A/b$, which is clearly unphysical in the context of diffusion. To remedy this, Zeldovich *et al.* proposed to take the maximum value of $G(\xi)$ and zero. Clearly $G = 0$ is a solution, hence $G(\xi) = \left(A - b\xi^2\right)_+$ is a solution for $|\xi| < \sqrt{A/b}$ and

for $|\xi| > \sqrt{A/b}$, but what about the points $\xi = \pm\sqrt{A/b}$? The concern is that the second derivative $G''(\xi)$ has a delta function singularity at those points, owing to the discontinuity of $G'(\xi)$. However, an examination of Eqn. 2.192 shows that G'' is multiplied by G, and we know that $\lim_{x\to 0} x\,\delta(x) = 0$. The remaining nonzero terms in this equation are then $[\mu\,\xi + G'(\xi)]G'(\xi)$, which agreeably vanishes. So we have a solution of the form[15]

$$V(\boldsymbol{x},t) = \frac{1}{t}\left(A'\,t^{2\alpha/d} - \alpha\,\boldsymbol{x}^2\right)_+ \;,$$

(2.194)

where $A' = 2dA$.

2.8   Appendix : Langevin equation for a particle in a harmonic well

Consider next the equation

$$\ddot{X} + \gamma\dot{X} + \omega_0^2 X = \frac{F}{M} + \eta(t)\;,$$

(2.195)

where F is a constant force. We write $X = x_0 + x$ and measure x relative to the potential minimum $x_0 = F/M\omega_0^2$, yielding

$$\ddot{x} + \gamma\,\dot{x} + \omega_0^2\,x = \eta(t)\;.$$

(2.196)

We solve via Laplace transform. Recall

$$\check{x}(z) = \int\limits_0^\infty dt\,e^{-zt}\,x(t)$$

$$x(t) = \int\limits_{\mathcal{C}}\frac{dz}{2\pi i}\,e^{+zt}\,\check{x}(z)\;,$$

(2.197)

where the contour $\mathcal{C}$ proceeds from $c - i\infty$ to $c + i\infty$ such that all poles of the integrand lie to the left of $\mathcal{C}$. Then

$$\int\limits_0^\infty dt\,e^{-zt}\left(\ddot{x} + \gamma\,\dot{x} + \omega_0^2\,x\right) = -(z+\gamma)\,x(0) - \dot{x}(0) + \left(z^2 + \gamma z + \omega_0^2\right)\check{x}(z)$$

$$= \int\limits_0^\infty dt\,e^{-zt}\,\eta(t) = \check{\eta}(z)\;.$$

(2.198)

Thus, we have

$$\check{x}(z) = \frac{(z+\gamma)\,x(0) + \dot{x}(0)}{z^2 + \gamma z + \omega_0^2} + \frac{1}{z^2 + \gamma z + \omega_0^2}\int\limits_0^\infty dt\,e^{-zt}\,\eta(t)\;.$$

(2.199)

---

[15]Actually the result $\lim_{x\to 0} x\,\delta(x) = 0$ is valid in the distribution sense, *i.e.* underneath an integral, provided $x\,\delta(x)$ is multiplied by a nonsingular function of x. Thus, Eqn. 2.194 constitutes a *weak solution* to the pressure form of the porous medium equation 2.188. Zeldovich *et al.* found numerically that cutting off the negative part of $A - b\,\xi^2$ is appropriate. Mathematically, Vazquez has shown that when the initial data are taken within a suitable class of integrable functions, the weak solution exists and is unique.

Now we may write

$$z^2 + \gamma z + \omega_0^2 = (z - z_+)(z - z_-) \,, \tag{2.200}$$

where $z_\pm = -\frac{1}{2}\gamma \pm \sqrt{\frac{1}{4}\gamma^2 - \omega_0^2}$. Note that $\mathrm{Re}\,(z_\pm) \leq 0$ and that $z_\mp = -\gamma - z_\pm$.

Performing the inverse Laplace transform, we obtain

$$x(t) = \frac{x(0)}{z_+ - z_-}\left(z_+\, e^{z_- t} - z_-\, e^{z_+ t}\right) + \frac{\dot{x}(0)}{z_+ - z_-}\left(e^{z_+ t} - e^{z_- t}\right) + \int\limits_0^\infty \!ds\, K(t-s)\,\eta(s) \quad, \tag{2.201}$$

where

$$K(t - s) = \frac{\Theta(t - s)}{(z_+ - z_-)}\left(e^{z_+ (t-s)} - e^{z_- (t-s)}\right) \tag{2.202}$$

is the *response kernel* and $\Theta(t - s)$ is the step function which is unity for $t > s$ and zero otherwise. The response is *causal*, *i.e.* $x(t)$ depends on $\eta(s)$ for all previous times $s < t$, but not for future times $s > t$. Note that $K(\tau)$ decays exponentially for $\tau \to \infty$, if $\mathrm{Re}(z_\pm) < 0$. The marginal case where $\omega_0 = 0$ and $z_+ = 0$ corresponds to the diffusion calculation we performed in the previous section.

It is now easy to compute

$$\langle x^2(t)\rangle_{\mathrm{c}} = \Gamma \int\limits_0^t ds\, K^2(s) = \frac{\Gamma}{2\omega_0^2\gamma} \quad (t \to \infty) \tag{2.203}$$

$$\langle \dot{x}^2(t)\rangle_{\mathrm{c}} = \Gamma \int\limits_0^t ds\, \dot{K}^2(s) = \frac{\Gamma}{2\gamma} \quad (t \to \infty) \,, \tag{2.204}$$

where the connected average is defined by $\langle AB\rangle_{\mathrm{c}} = \langle AB\rangle - \langle A\rangle\langle B\rangle$. Therefore,

$$\left\langle \tfrac{1}{2}M\dot{x}^2 + \tfrac{1}{2}M\omega_0^2 x^2 \right\rangle_{t\to\infty} = \frac{M\Gamma}{2\gamma} \,. \tag{2.205}$$

Setting this equal to $2 \times \frac{1}{2}k_{\mathrm{B}}T$ by equipartition again yields $\Gamma = 2\gamma k_{\mathrm{B}}T/M$.

2.9 Appendix : General Linear Autonomous Inhomogeneous ODEs

2.9.1 Solution by Fourier transform

We can also solve general autonomous linear inhomogeneous ODEs of the form

$$\frac{d^n x}{dt^n} + a_{n-1}\frac{d^{n-1}x}{dt^{n-1}} + \ldots + a_1\frac{dx}{dt} + a_0\, x = \xi(t) \,. \tag{2.206}$$

We can write this as

$$\mathcal{L}_t\, x(t) = \xi(t) \,, \tag{2.207}$$

where $\mathcal{L}_t$ is the nth order differential operator

$$\mathcal{L}_t = \frac{d^n}{dt^n} + a_{n-1}\frac{d^{n-1}}{dt^{n-1}} + \ldots + a_1\frac{d}{dt} + a_0 \,. \tag{2.208}$$

The general solution to the inhomogeneous equation is given by

$$x(t) = x_{\mathrm{h}}(t) + \int\limits_{-\infty}^{\infty} dt' \, G(t,t') \, \xi(t') \,, \tag{2.209}$$

where $G(t,t')$ is the Green's function. Note that $\mathcal{L}_t \, x_{\mathrm{h}}(t) = 0$. Thus, in order for eqns. 2.207 and 2.209 to be true, we must have

$$\mathcal{L}_t \, x(t) = \overbrace{\mathcal{L}_t \, x_{\mathrm{h}}(t)}^{\text{this vanishes}} + \int\limits_{-\infty}^{\infty} dt' \, \mathcal{L}_t \, G(t,t') \, \xi(t') = \xi(t) \,, \tag{2.210}$$

which means that

$$\mathcal{L}_t \, G(t,t') = \delta(t - t') \,, \tag{2.211}$$

where $\delta(t - t')$ is the Dirac δ-function.

If the differential equation $\mathcal{L}_t \, x(t) = \xi(t)$ is defined over some finite or semi-infinite t interval with prescribed boundary conditions on $x(t)$ at the endpoints, then $G(t,t')$ will depend on t and t' separately. For the case we are now considering, let the interval be the entire real line $t \in (-\infty, \infty)$. Then $G(t,t') = G(t - t')$ is a function of the single variable $t - t'$.

Note that $\mathcal{L}_t = \mathcal{L}\!\left(\frac{d}{dt}\right)$ may be considered a function of the differential operator $\frac{d}{dt}$. If we now Fourier transform the equation $\mathcal{L}_t \, x(t) = \xi(t)$, we obtain

$$\int\limits_{-\infty}^{\infty} dt \, e^{i\omega t} \, \xi(t) = \int\limits_{-\infty}^{\infty} dt \, e^{i\omega t} \left\{ \frac{d^n}{dt^n} + a_{n-1}\frac{d^{n-1}}{dt^{n-1}} + \ldots + a_1\frac{d}{dt} + a_0 \right\} x(t)$$

$$= \int\limits_{-\infty}^{\infty} dt \, e^{i\omega t} \left\{ (-i\omega)^n + a_{n-1}(-i\omega)^{n-1} + \ldots + a_1(-i\omega) + a_0 \right\} x(t) \,. \tag{2.212}$$

Thus, if we define

$$\hat{\mathcal{L}}(\omega) = \sum_{k=0}^{n} a_k \, (-i\omega)^k \,, \tag{2.213}$$

then we have $\hat{\mathcal{L}}(\omega) \, \hat{x}(\omega) = \hat{\xi}(\omega)$, where $a_n \equiv 1$. According to the Fundamental Theorem of Algebra, the nth degree polynomial $\hat{\mathcal{L}}(\omega)$ may be uniquely factored over the complex ω plane into a product over n roots:

$$\hat{\mathcal{L}}(\omega) = (-i)^n \, (\omega - \omega_1)(\omega - \omega_2)\cdots(\omega - \omega_n) \,. \tag{2.214}$$

If the $\{a_k\}$ are all real, then $\left[\hat{\mathcal{L}}(\omega)\right]^* = \hat{\mathcal{L}}(-\omega^*)$, hence if Ω is a root then so is $-\Omega^*$. Thus, the roots appear in pairs which are symmetric about the imaginary axis. *I.e.* if $\Omega = a + ib$ is a root, then so is $-\Omega^* = -a + ib$.

The general solution to the homogeneous equation is

$$x_{\rm h}(t) = \sum_{\sigma=1}^{n} A_\sigma \, e^{-i\omega_\sigma t} \,, \tag{2.215}$$

which involves n arbitrary complex constants A_i. The susceptibility, or Green's function in Fourier space, $\hat{G}(\omega)$ is then

$$\hat{G}(\omega) = \frac{1}{\hat{\mathcal{L}}(\omega)} = \frac{i^n}{(\omega - \omega_1)(\omega - \omega_2)\cdots(\omega - \omega_n)} \,, \tag{2.216}$$

Note that $\left[\hat{G}(\omega)\right]^* = \hat{G}(-\omega)$, which is equivalent to the statement that $G(t-t')$ is a real function of its argument. The general solution to the inhomogeneous equation is then

$$x(t) = x_{\rm h}(t) + \int\limits_{-\infty}^{\infty} dt' \, G(t-t') \, \xi(t') \,, \tag{2.217}$$

where $x_{\rm h}(t)$ is the solution to the homogeneous equation, *i.e.* with zero forcing, and where

$$
\begin{aligned}
G(t-t') &= \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi} \, e^{-i\omega(t-t')} \, \hat{G}(\omega) \\
&= i^n \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi} \, \frac{e^{-i\omega(t-t')}}{(\omega - \omega_1)(\omega - \omega_2)\cdots(\omega - \omega_n)} \\
&= \sum_{\sigma=1}^{n} \frac{e^{-i\omega_\sigma(t-t')}}{i\,\hat{\mathcal{L}}'(\omega_\sigma)} \, \Theta(t-t') \,,
\end{aligned}
\tag{2.218}
$$

where we assume that $\operatorname{Im}\omega_\sigma < 0$ for all σ. This guarantees *causality* – the *response* $x(t)$ to the *influence* $\xi(t')$ is nonzero only for $t > t'$.

As an example, consider the familiar case

$$
\begin{aligned}
\hat{\mathcal{L}}(\omega) &= -\omega^2 - i\gamma\omega + \omega_0^2 \\
&= -(\omega - \omega_+)\,(\omega - \omega_-) \,,
\end{aligned}
\tag{2.219}
$$

with $\omega_\pm = -\frac{i}{2}\gamma \pm \beta$, and $\beta = \sqrt{\omega_0^2 - \frac{1}{4}\gamma^2}$. This yields $\hat{\mathcal{L}}'(\omega_\pm) = \mp(\omega_+ - \omega_-) = \mp 2\beta$, hence according to equation 2.218,

$$
\begin{aligned}
G(s) &= \left\{ \frac{e^{-i\omega_+ s}}{i\mathcal{L}'(\omega_+)} + \frac{e^{-i\omega_- s}}{i\mathcal{L}'(\omega_-)} \right\} \Theta(s) \\
&= \left\{ \frac{e^{-\gamma s/2}\, e^{-i\beta s}}{-2i\beta} + \frac{e^{-\gamma s/2}\, e^{i\beta s}}{2i\beta} \right\} \Theta(s) = \beta^{-1}\, e^{-\gamma s/2}\, \sin(\beta s)\, \Theta(s) \,.
\end{aligned}
\tag{2.220}
$$

Now let us evaluate the two-point correlation function $\langle x(t)\, x(t')\rangle$, assuming the noise is correlated according to $\langle \xi(s)\, \xi(s')\rangle = \phi(s - s')$. We assume $t, t' \to \infty$ so the transient contribution x_{h} is negligible. We then have

$$\langle x(t)\, x(t')\rangle = \int\limits_{-\infty}^{\infty} ds \int\limits_{-\infty}^{\infty} ds'\, G(t - s)\, G(t' - s')\, \langle \xi(s)\, \xi(s')\rangle = \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \hat{\phi}(\omega)\, \big|\hat{G}(\omega)\big|^2\, e^{i\omega(t-t')}\,. \tag{2.221}$$

2.9.2 Higher order ODEs

Note that any n^{th} order ODE, of the general form

$$\frac{d^n x}{dt^n} = F\left(x,\, \frac{dx}{dt},\, \dots,\, \frac{d^{n-1} x}{dt^{n-1}} \right), \tag{2.222}$$

may be represented by the first order system $\dot{\boldsymbol{\varphi}} = \boldsymbol{V}(\boldsymbol{\varphi})$. To see this, define $\varphi_k = d^{k-1} x / dt^{k-1}$, with $k = 1, \dots, n$. Thus, for $k < n$ we have $\dot{\varphi}_k = \varphi_{k+1}$, and $\dot{\varphi}_n = F$. In other words,

$$\frac{d}{dt} \overbrace{\begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_{n-1} \\ \varphi_n \end{pmatrix}}^{\dot{\boldsymbol{\varphi}}} = \overbrace{\begin{pmatrix} \varphi_2 \\ \vdots \\ \varphi_n \\ F(\varphi_1, \dots, \varphi_p) \end{pmatrix}}^{\boldsymbol{V}(\boldsymbol{\varphi})}. \tag{2.223}$$

An inhomogeneous linear n^{th} order ODE,

$$\frac{d^n x}{dt^n} + a_{n-1}\, \frac{d^{n-1} x}{dt^{n-1}} + \dots + a_1\, \frac{dx}{dt} + a_0\, x = \xi(t) \tag{2.224}$$

may be written in matrix form, as

$$\frac{d}{dt} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_n \end{pmatrix} = \overbrace{\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{pmatrix}}^{Q} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_n \end{pmatrix} + \overbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ \xi(t) \end{pmatrix}}^{\boldsymbol{\xi}}. \tag{2.225}$$

Thus,

$$\dot{\boldsymbol{\varphi}} = Q\, \boldsymbol{\varphi} + \boldsymbol{\xi}\,, \tag{2.226}$$

and if the coefficients c_k are time-independent, *i.e.* the ODE is *autonomous*.

For the homogeneous case where $\xi(t) = 0$, the solution is obtained by exponentiating the constant matrix Qt:

$$\boldsymbol{\varphi}(t) = \exp(Qt)\, \boldsymbol{\varphi}(0)\,; \tag{2.227}$$

the exponential of a matrix may be given meaning by its Taylor series expansion. If the ODE is not autonomous, then $Q = Q(t)$ is time-dependent, and the solution is given by the path-ordered exponential,

$$\varphi(t) = \mathsf{P} \exp \left\{ \int_0^t dt' \, Q(t') \right\} \varphi(0) \,, \tag{2.228}$$

where P is the *path ordering operator* which places earlier times to the right. As defined, the equation $\dot{\varphi} = \boldsymbol{V}(\varphi)$ is autonomous, since the t-advance mapping g_t depends only on t and on no other time variable. However, by extending the phase space $\mathbb{M} \ni \varphi$ from $\mathbb{M} \to \mathbb{M} \times \mathbb{R}$, which is of dimension $n+1$, one can describe arbitrary time-dependent ODEs.

In general, path ordered exponentials are difficult to compute analytically. We will henceforth consider the autonomous case where Q is a constant matrix in time. We will assume the matrix Q is real, but other than that it has no helpful symmetries. We can however decompose it into left and right eigenvectors:

$$Q_{ij} = \sum_{\sigma=1}^{n} \nu_\sigma \, R_{\sigma,i} \, L_{\sigma,j} \,. \tag{2.229}$$

Or, in bra-ket notation, $Q = \sum_\sigma \nu_\sigma \,|R_\sigma\rangle\langle L_\sigma|$. We adopt the normalization convention $\langle \, L_\sigma \,|\, R_{\sigma'} \, \rangle = \delta_{\sigma\sigma'}$, where $\{\nu_\sigma\}$ are the eigenvalues of Q. The eigenvalues may be real or imaginary. Since the characteristic polynomial $P(\nu) = \det(\nu \, \mathbb{I} - Q)$ has real coefficients, we know that the eigenvalues of Q are either real or come in complex conjugate pairs.

Consider, for example, the $n = 2$ system we studied earlier. Then

$$Q = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{pmatrix} \,. \tag{2.230}$$

The eigenvalues are as before: $\nu_\pm = -\frac{1}{2}\gamma \pm \sqrt{\frac{1}{4}\gamma^2 - \omega_0^2}$. The left and right eigenvectors are

$$L_\pm = \frac{\pm 1}{\nu_+ - \nu_-} \left(-\nu_\mp \quad 1 \right) \qquad , \qquad R_\pm = \begin{pmatrix} 1 \\ \nu_\pm \end{pmatrix} \,. \tag{2.231}$$

The utility of working in a left-right eigenbasis is apparent once we reflect upon the result

$$f(Q) = \sum_{\sigma=1}^{n} f(\nu_\sigma) \, | \, R_\sigma \, \rangle \langle \, L_\sigma \, | \tag{2.232}$$

for any function f. Thus, the solution to the general autonomous homogeneous case is

$$\left| \, \varphi(t) \, \right\rangle = \sum_{\sigma=1}^{n} e^{\nu_\sigma t} \, | \, R_\sigma \, \rangle \langle \, L_\sigma \, | \, \varphi(0) \, \rangle$$
$$\varphi_i(t) = \sum_{\sigma=1}^{n} e^{\nu_\sigma t} \, R_{\sigma,i} \sum_{j=1}^{n} L_{\sigma,j} \, \varphi_j(0) \,. \tag{2.233}$$

If $\mathrm{Re}\,(\nu_\sigma) \leq 0$ for all σ, then the initial conditions $\varphi(0)$ are forgotten on time scales $\tau_\sigma = \nu_\sigma^{-1}$. Physicality demands that this is the case.

Now let's consider the inhomogeneous case where $\xi(t) \neq 0$. We begin by recasting eqn. 2.226 in the form

$$\frac{d}{dt}\left(e^{-Qt}\,\varphi\right) = e^{-Qt}\,\xi(t) \ . \tag{2.234}$$

We can integrate this directly:

$$\varphi(t) = e^{Qt}\,\varphi(0) + \int\limits_0^t ds\, e^{Q(t-s)}\,\xi(s) \ . \tag{2.235}$$

In component notation,

$$\varphi_i(t) = \sum_{\sigma=1}^n e^{\nu_\sigma t}\, R_{\sigma,i}\left\langle\, L_\sigma \,\middle|\, \varphi(0) \,\right\rangle + \sum_{\sigma=1}^n R_{\sigma,i} \int\limits_0^t ds\, e^{\nu_\sigma(t-s)}\left\langle\, L_\sigma \,\middle|\, \xi(s) \,\right\rangle. \tag{2.236}$$

Note that the first term on the RHS is the solution to the homogeneous equation, as must be the case when $\xi(s) = 0$.

The solution in eqn. 2.236 holds for general Q and $\xi(s)$. For the particular form of Q and $\xi(s)$ in eqn. 2.225, we can proceed further. For starters, $\langle L_\sigma | \xi(s)\rangle = L_{\sigma,n}\,\xi(s)$. We can further exploit a special feature of the Q matrix to analytically determine all its left and right eigenvectors. Applying Q to the right eigenvector $|R_\sigma\rangle$, we find $R_{\sigma,j} = \nu_\sigma\, R_{\sigma,j-1}$ for $j > 1$. We are free to choose $R_{\sigma,1} = 1$ for all σ and defer the issue of normalization to the derivation of the left eigenvectors. Thus, we obtain the pleasingly simple result, $R_{\sigma,k} = \nu_\sigma^{k-1}$. Applying Q to the left eigenvector $\langle L_\sigma|$, we obtain

$$\begin{aligned} -a_0\, L_{\sigma,n} &= \nu_\sigma\, L_{\sigma,1} \\ L_{\sigma,j-1} - a_{j-1}\, L_{\sigma,n} &= \nu_\sigma\, L_{\sigma,j} \qquad (j>1) \ . \end{aligned} \tag{2.237}$$

From these equations we may derive

$$L_{\sigma,k} = -\frac{L_{\sigma,n}}{\nu_\sigma}\sum_{j=0}^{k-1} a_j\, \nu_\sigma^{j-k-1} = \frac{L_{\sigma,n}}{\nu_\sigma}\sum_{j=k}^{n} a_j\, \nu_\sigma^{j-k-1} \ . \tag{2.238}$$

The equality in the above equation is derived using the result $P(\nu_\sigma) = \sum_{j=0}^n a_j\, \nu_\sigma^j = 0$. Recall also that $a_n \equiv 1$. We now impose the normalization condition,

$$\sum_{k=1}^n L_{\sigma,k}\, R_{\sigma,k} = 1 \ . \tag{2.239}$$

This condition determines our last remaining unknown quantity (for a given σ), $L_{\sigma,p}$:

$$\left\langle\, L_\sigma \,\middle|\, R_\sigma \,\right\rangle = L_{\sigma,n}\sum_{k=1}^n k\, a_k\, \nu_\sigma^{k-1} = P'(\nu_\sigma)\, L_{\sigma,n} \ , \tag{2.240}$$

where $P'(\nu)$ is the first derivative of the characteristic polynomial. Thus, we find $L_{\sigma,n} = 1/P'(\nu_\sigma)$.

Now let us evaluate the general two-point correlation function,

$$C_{jj'}(t,t') \equiv \langle \varphi_j(t)\,\varphi_{j'}(t') \rangle - \langle \varphi_j(t) \rangle \langle \varphi_{j'}(t') \rangle \,. \tag{2.241}$$

We write

$$\langle \xi(s)\,\xi(s') \rangle = \phi(s - s') = \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \hat{\phi}(\omega)\, e^{-i\omega(s - s')} \,. \tag{2.242}$$

When $\hat{\phi}(\omega)$ is constant, we have $\langle \xi(s)\,\xi(s') \rangle = \hat{\phi}(t)\,\delta(s - s')$. This is the case of so-called *white noise*, when all frequencies contribute equally. The more general case when $\hat{\phi}(\omega)$ is frequency-dependent is known as *colored noise*. Appealing to eqn. 2.236, we have

$$C_{jj'}(t,t') = \sum_{\sigma,\sigma'} \frac{\nu_\sigma^{j-1}}{P'(\nu_\sigma)} \frac{\nu_{\sigma'}^{j'-1}}{P'(\nu_{\sigma'})} \int\limits_0^t ds\, e^{\nu_\sigma(t-s)} \int\limits_0^{t'} ds'\, e^{\nu_{\sigma'}(t'-s')}\, \phi(s - s') \tag{2.243}$$

$$= \sum_{\sigma,\sigma'} \frac{\nu_\sigma^{j-1}}{P'(\nu_\sigma)} \frac{\nu_{\sigma'}^{j'-1}}{P'(\nu_{\sigma'})} \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \frac{\hat{\phi}(\omega)\,(e^{-i\omega t} - e^{\nu_\sigma t})(e^{i\omega t'} - e^{\nu_{\sigma'} t'})}{(\omega - i\nu_\sigma)(\omega + i\nu_{\sigma'})} \,. \tag{2.244}$$

In the limit $t, t' \to \infty$, assuming $\mathrm{Re}\,(\nu_\sigma) < 0$ for all σ (*i.e.* no diffusion), the exponentials $e^{\nu_\sigma t}$ and $e^{\nu_{\sigma'} t'}$ may be neglected, and we then have

$$C_{jj'}(t,t') = \sum_{\sigma,\sigma'} \frac{\nu_\sigma^{j-1}}{P'(\nu_\sigma)} \frac{\nu_{\sigma'}^{j'-1}}{P'(\nu_{\sigma'})} \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \frac{\hat{\phi}(\omega)\, e^{-i\omega(t-t')}}{(\omega - i\nu_\sigma)(\omega + i\nu_{\sigma'})} \,. \tag{2.245}$$

2.9.3 Kramers-Krönig relations

Suppose $\hat{\chi}(\omega) \equiv \hat{G}(\omega)$ is analytic in the UHP[16]. Then for all ν, we must have

$$\int\limits_{-\infty}^{\infty} \frac{d\nu}{2\pi}\, \frac{\hat{\chi}(\nu)}{\nu - \omega + i\epsilon} = 0 \,, \tag{2.246}$$

where ϵ is a positive infinitesimal. The reason is simple: just close the contour in the UHP, assuming $\hat{\chi}(\omega)$ vanishes sufficiently rapidly that Jordan's lemma can be applied. Clearly this is an extremely weak restriction on $\hat{\chi}(\omega)$, given the fact that the denominator already causes the integrand to vanish as $|\omega|^{-1}$.

Let us examine the function

$$\frac{1}{\nu - \omega + i\epsilon} = \frac{\nu - \omega}{(\nu - \omega)^2 + \epsilon^2} - \frac{i\epsilon}{(\nu - \omega)^2 + \epsilon^2} \,. \tag{2.247}$$

---

[16]In this section, we use the notation $\hat{\chi}(\omega)$ for the susceptibility, rather than $\hat{G}(\omega)$

which we have separated into real and imaginary parts. Under an integral sign, the first term, in the limit $\epsilon \to 0$, is equivalent to taking a *principal part* of the integral. That is, for any function $F(\nu)$ which is regular at $\nu = \omega$,

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \frac{d\nu}{2\pi} \frac{\nu - \omega}{(\nu - \omega)^2 + \epsilon^2} F(\nu) \equiv \wp \int_{-\infty}^{\infty} \frac{d\nu}{2\pi} \frac{F(\nu)}{\nu - \omega} \ . \tag{2.248}$$

The *principal part* symbol $\wp$ means that the singularity at $\nu = \omega$ is elided, either by smoothing out the function $1/(\nu - \epsilon)$ as above, or by simply cutting out a region of integration of width ϵ on either side of $\nu = \omega$.

The imaginary part is more interesting. Let us write

$$h(u) \equiv \frac{\epsilon}{u^2 + \epsilon^2} \ . \tag{2.249}$$

For $|u| \gg \epsilon$, $h(u) \simeq \epsilon/u^2$, which vanishes as $\epsilon \to 0$. For $u = 0$, $h(0) = 1/\epsilon$ which diverges as $\epsilon \to 0$. Thus, $h(u)$ has a huge peak at $u = 0$ and rapidly decays to 0 as one moves off the peak in either direction a distance greater that ϵ. Finally, note that

$$\int_{-\infty}^{\infty} du \, h(u) = \pi \ , \tag{2.250}$$

a result which itself is easy to show using contour integration. Putting it all together, this tells us that

$$\lim_{\epsilon \to 0} \frac{\epsilon}{u^2 + \epsilon^2} = \pi \delta(u) \ . \tag{2.251}$$

Thus, for positive infinitesimal ϵ,

$$\frac{1}{u \pm i\epsilon} = \frac{\wp}{u} \mp i\pi\delta(u) \ , \tag{2.252}$$

a most useful result.

We now return to our initial result 2.246, and we separate $\hat{\chi}(\omega)$ into real and imaginary parts:

$$\hat{\chi}(\omega) = \hat{\chi}'(\omega) + i\hat{\chi}''(\omega) \ . \tag{2.253}$$

(In this equation, the primes do not indicate differentiation with respect to argument.) We therefore have, for every real value of ω,

$$0 = \int_{-\infty}^{\infty} \frac{d\nu}{2\pi} \left[\chi'(\nu) + i\chi''(\nu) \right] \left[\frac{\wp}{\nu - \omega} - i\pi\delta(\nu - \omega) \right] \ . \tag{2.254}$$

Taking the real and imaginary parts of this equation, we derive the *Kramers-Krönig relations*:

$$\chi'(\omega) = +\wp \int_{-\infty}^{\infty} \frac{d\nu}{\pi} \frac{\hat{\chi}''(\nu)}{\nu - \omega} \tag{2.255}$$

$$\chi''(\omega) = -\wp \int_{-\infty}^{\infty} \frac{d\nu}{\pi} \frac{\hat{\chi}'(\nu)}{\nu - \omega} \ . \tag{2.256}$$

2.10 Appendix : Method of Characteristics

2.10.1 Quasilinear partial differential equations

Consider the quasilinear PDE

$$a_1(\boldsymbol{x}, \phi)\, \frac{\partial \phi}{\partial x_1} + a_2(\boldsymbol{x}, \phi)\, \frac{\partial \phi}{\partial x_2} + \ldots + a_N(\boldsymbol{x}, \phi)\, \frac{\partial \phi}{\partial x_N} = b(\boldsymbol{x}, \phi)\, . \tag{2.257}$$

This PDE is called 'quasilinear' because it is linear in the derivatives $\partial \phi / \partial x_j$. The N independent variables are the elements of the vector $\boldsymbol{x} = (x_1, \ldots, x_N)$. A solution is a function $\phi(\boldsymbol{x})$ which satisfies the PDE.

Now consider a curve $\boldsymbol{x}(s)$ parameterized by a single real variable s satisfying

$$\frac{dx_j}{ds} = a_j\big(\boldsymbol{x}, \phi(\boldsymbol{x})\big)\, , \tag{2.258}$$

where $\phi(\boldsymbol{x})$ is a solution of eqn. 2.257. Along such a curve, which is called a *characteristic*, the variation of ϕ is

$$\frac{d\phi}{ds} = \sum_{j=1}^{N} \frac{\partial \phi}{\partial x_j}\, \frac{dx_j}{ds} = b\big(\boldsymbol{x}(s), \phi\big)\, . \tag{2.259}$$

Thus, we have converted our PDE into a set of $N+1$ ODEs. To integrate, we must supply some initial conditions of the form

$$g\big(\boldsymbol{x}, \phi\big)\Big|_{s=0} = 0\, . \tag{2.260}$$

This defines an $(N-1)$-dimensional hypersurface, parameterized by $\{\zeta_1, \ldots, \zeta_{N-1}\}$:

$$\begin{aligned} x_j(s=0) &= h_j(\zeta_1, \ldots, \zeta_{N-1}) \qquad , \qquad j \in \{1, \ldots, N\} \\ \phi(s=0) &= f(\zeta_1, \ldots, \zeta_{N-1})\, . \end{aligned} \tag{2.261}$$

If we can solve for all the characteristic curves, then the solution of the PDE follows. For every $\boldsymbol{x}$, we identify the characteristic curve upon which $\boldsymbol{x}$ lies. The characteristics are identified by their parameters $(\zeta_1, \ldots, \zeta_{N-1})$. The solution is then $\phi(\boldsymbol{x}) = \phi(s; \zeta_1, \ldots, \zeta_{N-1})$. If two or more characteristics cross, the solution is multi-valued, or a shock has occurred.

2.10.2 Example

Consider the PDE

$$\phi_t + t^2\, \phi_x = -x\, \phi\, . \tag{2.262}$$

We identify $a_1(t, x, \phi) = 1$ and $a_2(t, x, \phi) = t^2$, as well as $b(t, x, \phi) = -x\, \phi$. The characteristics are curves $\big(t(s), x(s)\big)$ satisfing

$$\frac{dt}{ds} = 1 \qquad , \qquad \frac{dx}{ds} = t^2\, . \tag{2.263}$$

The variation of ϕ along each of the characteristics is given by

$$\frac{d\phi}{ds} = -x\,\phi\ . \tag{2.264}$$

The initial data are expressed parametrically as

$$t(s=0) = 0 \quad , \quad x(s=0) = \zeta \quad , \quad \phi(s=0) = f(\zeta) \quad . \tag{2.265}$$

We now solve for the characteristics. We have

$$\frac{dt}{ds} = 1 \quad \Rightarrow \quad t(s,\zeta) = s\ . \tag{2.266}$$

It then follows that

$$\frac{dx}{ds} = t^2 = s^2 \quad \Rightarrow \quad x(s,\zeta) = \zeta + \tfrac{1}{3}s^3\ . \tag{2.267}$$

Finally, we have

$$\frac{d\phi}{ds} = -x\,\phi = -\left(\zeta + \tfrac{1}{3}s^3\right)\phi \quad \Rightarrow \quad \phi(s,\zeta) = f(\zeta)\,\exp\left(-\tfrac{1}{12}s^4 - s\zeta\right)\ . \tag{2.268}$$

We may now eliminate (ζ, s) in favor of (x,t), writing $s = t$ and $\zeta = x - \tfrac{1}{3}t^3$, yielding the solution

$$\phi(x,t) = \phi\left(x - \tfrac{1}{3}t^3, t=0\right)\,\exp\left(\tfrac{1}{4}t^4 - xt\right)\ . \tag{2.269}$$

Chapter 3

Stochastic Calculus

3.1 References

- C. Gardiner, *Stochastic Methods* (4th edition, Springer-Verlag, 2010)
  Very clear and complete text on stochastic methods, with many applications.

- Z. Schuss, *Theory and Applications of Stochastic Processes* (Springer-Verlag, 2010)
  In-depth discussion of continuous path stochastic processes and connections to partial differential equations.

- R. Mahnke, J. Kaupužs, and I. Lubashevsky, *Physics of Stochastic Processes* (Wiley, 2009)
  Introductory sections are sometimes overly formal, but a good selection of topics.

- H. Riecke, *Introduction to Stochastic Processes and Stochastic Differential Equations* (unpublished, 2010)
  Good set of lecture notes, often following Gardiner. Available online at:
  http://people.esam.northwestern.edu/~riecke/Vorlesungen/442/Notes/notes_442.pdf

- J. L. McCauley, *Dynamics of Markets* (2nd edition, Cambridge, 2009)
  A physics-friendly discussion of stochastic market dynamics. Crisp and readable. Despite this being the second edition, there are alas a great many typographical errors.

3.2   Gaussian White Noise

Consider a generalized Langevin equation of the form

$$\frac{du}{dt} = f(u,t) + g(u,t)\,\eta(t)\;, \tag{3.1}$$

where $\eta(t)$ is a Gaussian random function with zero mean and

$$\big\langle \eta(t)\,\eta(t')\big\rangle = \phi(t-t')\;. \tag{3.2}$$

The spectral function of the noise is given by the Fourier transform,

$$\hat{\phi}(\omega) = \int\limits_{-\infty}^{\infty} ds\,\phi(s)\,e^{-i\omega s} = \lim_{T\to\infty}\left\langle \frac{1}{T}\,\big|\hat{\eta}_T(\omega)\big|^2\right\rangle\;, \tag{3.3}$$

using the notation of §2.6.3. When $\phi(s) = \Gamma\,\delta(s)$, we have $\hat{\phi}(\omega) = \Gamma$, *i.e.* independent of frequency. This is the case of *Gaussian white noise*. When $\hat{\phi}(\omega)$ has a nontrivial dependence on frequency, the noise is said to be *colored*. Gaussian white noise has an infinite variance $\phi(0)$, which leads to problems. In particular, the derivative $\dot{u}$ strictly speaking does not exist because the function $\eta(t)$ is not continuous.

As an example of the sort of problem this presents, consider the differential equation $\dot{u}(t) = \eta(t)\,u(t)$. Let's integrate this over a time period Δt from t_j to t_{j+1}, where $t_j = j\,\Delta t$. We then have $u(t_{j+1}) = \big(1 + \eta(t_j)\Delta t\big)\,u(t_j)$. Thus, we find

$$u(t_N) = \Big(1 + \eta(t_{N-1})\,\Delta t\Big)\cdots\Big(1 + \eta(t_0)\,\Delta t\Big)\,u(t_0)\;. \tag{3.4}$$

Now let's compute the average $\langle u(t_N)\rangle$. Since $\eta(t_j)$ is uncorrelated with $\eta(t_k)$ for all $k \neq j$, we can take the average of each of the terms individually, and since $\eta(t_j)$ has zero mean, we conclude that $\langle u(t_N)\rangle = u(t_0)$. On average, there is no drift.

Now let's take a continuum limit of the above result, which is to say $\Delta t \to 0$ with $N\Delta t$ finite. Setting $t_0 = 0$ and $t_N = t$, we have

$$u(t) = u(0)\,\exp\left\{\int\limits_0^t ds\,\eta(s)\right\}\;, \tag{3.5}$$

and for Gaussian $\eta(s)$ we have

$$\big\langle u(t)\big\rangle = u(0)\,\exp\left\{\tfrac{1}{2}\int\limits_0^t ds\int\limits_0^t ds'\,\big\langle \eta(s)\,\eta(s')\big\rangle\right\} = u(0)\,e^{\Gamma t/2}\;. \tag{3.6}$$

In the continuum expression, we find there is *noise-induced drift*. The continuum limit of our discrete calculation has failed to match the continuum results. Clearly we have a problem that we must resolve. The origin of the problem is the aforementioned infinite variance of $\eta(t)$. This means that the Langevin equation 3.1 is not well-defined, and in order to get a definite answer we must provide a prescription regarding how it is to be integrated[1].

---

[1]We will see that Eqn. 3.4 corresponds to the *Itô* prescription and Eqn. 3.5 to the *Stratonovich* prescription.

3.3 Stochastic Integration

3.3.1 Langevin equation in differential form

We can make sense of Eqn. 3.1 by writing it in differential form,

$$du = f(u,t)\,dt + g(u,t)\,dW(t)\ ,\tag{3.7}$$

where

$$W(t) = \int_0^t ds\,\eta(s)\ .\tag{3.8}$$

This is because $W(t)$ is described by a Wiener process, for which the sample paths are continuous with probability unity. We shall henceforth take $\Gamma \equiv 1$, in which case $W(t)$ is Gaussianly distributed with $\langle W(t)\rangle = 0$ and

$$\langle W(t)\,W(t')\rangle = \mathsf{min}(t,t')\ .\tag{3.9}$$

The solution to Eqn. 3.7 is formally

$$u(t) = u(0) + \int_0^t ds\,f\big(u(s),s\big) + \int_0^t dW(s)\,g\big(u(s),s\big)\ .\tag{3.10}$$

Note that Eqn. 3.9 implies

$$\frac{d}{dt'}\langle W(t)\,W(t')\rangle = \Theta(t-t')\quad\Rightarrow\quad \langle \frac{dW(t)}{dt}\,\frac{dW(t')}{dt'}\rangle = \langle \eta(t)\,\eta(t')\rangle = \delta(t-t')\quad.\tag{3.11}$$

3.3.2 Defining the stochastic integral

Let $F(t)$ be an arbitrary function of time, and let $\{t_j\}$ be a discretization of the interval $[0,t]$ with $j \in \{0,\dots,N\}$. The simplest example to consider is $t_j = j\,\Delta t$ where $\Delta t = t/N$. Consider the quantity

$$S_N(\alpha) = \sum_{j=0}^{N-1}\Big[(1-\alpha)\,F(t_j) + \alpha\,F(t_{j+1})\Big]\Big[W(t_{j+1}) - W(t_j)\Big]\ ,\tag{3.12}$$

where $\alpha \in [0,1]$. Note that the first term in brackets on the RHS can be approximated as

$$F(\tau_j) = (1-\alpha)\,F(t_j) + \alpha\,F(t_{j+1})\ ,\tag{3.13}$$

where $\tau_j \equiv (1-\alpha)\,t_j + \alpha\,t_{j+1} \in [t_j, t_{j+1}]$. To abbreviate notation, we will write $F(t_j) = F_j$, $W(t_j) = W_j$, etc. We may take $t_0 \equiv 0$ and $W_0 \equiv 0$. The quantities $\Delta W_j \equiv W_{j+1} - W_j$ are *independently and Gaussianly distributed with zero mean for each j*. This means $\langle \Delta W_j\rangle = 0$ and

$$\langle \Delta W_j\,\Delta W_k\rangle = \langle (\Delta W_j)^2\rangle\,\delta_{jk} = \Delta t_j\,\delta_{jk}\quad,\tag{3.14}$$

where $\Delta t_j \equiv t_{j+1} - t_j$. Wick's theorem then tells us

$$\langle \Delta W_j \, \Delta W_k \, \Delta W_l \, \Delta W_m \rangle = \langle \Delta W_j \, \Delta W_k \rangle \langle \Delta W_l \, \Delta W_m \rangle + \langle \Delta W_j \, \Delta W_l \rangle \langle \Delta W_k \, \Delta W_m \rangle + \langle \Delta W_j \, \Delta W_m \rangle \langle \Delta W_k \, \Delta W_l \rangle$$

$$= \Delta t_j \, \Delta t_l \, \delta_{jk} \, \delta_{lm} + \Delta t_j \, \Delta t_k \, \delta_{jl} \, \delta_{km} + \Delta t_j \, \Delta t_k \, \delta_{jm} \, \delta_{kl} \quad . \tag{3.15}$$

EXERCISE: Show that $\langle W_N^2 \rangle = t$ and $\langle W_N^4 \rangle = 3t^2$.

The expression in Eqn. 3.12 would converge to the integral

$$S = \int_0^t dW(s) \, F(s) \tag{3.16}$$

independent of α were it not for the fact that $\Delta W_j / \Delta t_j$ has infinite variance in the limit $N \to \infty$. Instead, we will find that $S_N(\alpha)$ in general depends on the value of α. For example, the *Itô integral* is defined as the $N \to \infty$ limit of $S_N(\alpha)$ with $\alpha = 0$, whereas the *Stratonovich integral* is defined as the $N \to \infty$ limit of $S_N(\alpha)$ with $\alpha = \frac{1}{2}$.

We now define the stochastic integral

$$\int_0^t dW(s) \, \big[F(s)\big]_\alpha \equiv \text{ms} - \lim_{N \to \infty} \sum_{j=0}^{N-1} \Big[(1-\alpha) \, F(t_j) + \alpha \, F(t_{j+1})\Big] \Big[W(t_{j+1}) - W(t_j)\Big] \quad , \tag{3.17}$$

where ms-lim stands for *mean square limit*. We say that a sequence S_N converges to S in the mean square if $\lim_{N \to \infty} \langle (S_N - S)^2 \rangle = 0$. Consider, for example, the sequence $S_N = \sum_{j=0}^{N-1} (\Delta W_j)^2$. We now take averages, using $\langle (\Delta W_j)^2 \rangle = t_{j+1} - t_j \equiv \Delta t_j$. Clearly $S = \langle S_N \rangle = t$. We also have

$$\langle S_N^2 \rangle = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \langle (\Delta W_j)^2 \, (\Delta W_k)^2 \rangle = (N^2 + 2N)(\Delta t)^2 = t^2 + \frac{2t^2}{N} \quad , \tag{3.18}$$

where we have used Eqn. 3.15. Thus, $\langle (S_N - S)^2 \rangle = 2t^2/N \to 0$ in the $N \to \infty$ limit. So S_N converges to t in the mean square.

Next, consider the case where $F(t) = W(t)$. We find

$$S_N(\alpha) = \sum_{j=0}^{N-1} \Big[(1-\alpha) \, W(t_j) + \alpha \, W(t_{j+1})\Big] \Big[W(t_{j+1}) - W(t_j)\Big] = \sum_{j=0}^{N-1} \big(W_j + \alpha \, \Delta W_j\big) \, \Delta W_j$$

$$= \tfrac{1}{2} \sum_{j=0}^{N-1} \Big[\big(W_j + \Delta W_j\big)^2 - W_j^2 + (2\alpha - 1)\big(\Delta W_j\big)^2\Big] = \tfrac{1}{2} W_N^2 + (\alpha - \tfrac{1}{2}) \sum_{j=0}^{N-1} \big(\Delta W_j\big)^2 \quad . \tag{3.19}$$

Taking the average,

$$\langle S_N(\alpha) \rangle = \tfrac{1}{2} t_N + (\alpha - \tfrac{1}{2}) \sum_{j=0}^{N-1} (t_{j+1} - t_j) = \alpha \, t \quad . \tag{3.20}$$

Does S_N converge to $\langle S_N \rangle = \alpha t$ in the mean square? Let's define $Q_N \equiv \sum_{j=0}^{N-1} (\Delta W_j)^2$, which is the sequence we analyzed previously. Then $S_N = \frac{1}{2} W_N^2 + (\alpha - \frac{1}{2}) Q_N$. We then have

$$\langle S_N^2 \rangle = \tfrac{1}{4} \langle W_N^4 \rangle + (\alpha - \tfrac{1}{2}) \langle W_N^2 Q_N \rangle + (\alpha - \tfrac{1}{2})^2 \langle Q_N^2 \rangle \quad , \tag{3.21}$$

with

$$\langle W_N^4 \rangle = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \langle \Delta W_j \, \Delta W_k \, \Delta W_l \, \Delta W_m \rangle = 3N^2 (\Delta t)^2$$

$$\langle W_N^2 \, Q_N \rangle = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \langle \Delta W_j \, \Delta W_k \, (\Delta W_l)^2 \rangle = (N^2 + 2N)(\Delta t)^2 \tag{3.22}$$

$$\langle Q_N^2 \rangle = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \langle (\Delta W_j)^2 \, (\Delta W_k)^2 \rangle = (N^2 + 2N)(\Delta t)^2 \quad .$$

Therefore

$$\langle S_N^2 \rangle = \left(\alpha^2 + \tfrac{1}{2} \right) t^2 + \left(\alpha^2 - \tfrac{1}{4} \right) \cdot \frac{2t^2}{N} \quad . \tag{3.23}$$

Therefore $\langle (S_N - \alpha t)^2 \rangle = \frac{1}{2} t^2 + \mathcal{O}(N^{-1})$ and S_N does *not* converge to αt in the mean square! However, if we take

$$S \equiv \int_0^t dW(s) \, [W(s)]_\alpha = \tfrac{1}{2} W^2(t) + \left(\alpha - \tfrac{1}{2} \right) t \,, \tag{3.24}$$

we have $S_N - S = (\alpha - \frac{1}{2})(Q_N - t)$, S_N converges to S in the mean square. What happened in this example is that $Q_N = \sum_{j=0}^{N-1} (\Delta W_j)^2$ has zero variance in the limit $N \to \infty$, but W_N^2 has finite variance. Therefore S_N has finite variance, and it cannot converge in the mean square to any expression which has zero variance.

3.3.3 Summary of properties of the Itô stochastic integral

For the properties below, it is useful to define the notion of a *nonanticipating function $F(t)$* as one which is independent of the difference $W(s) - W(t)$ for all $s > t$ at any given t. An example of such a function would be any Itô integral of the form $\int_0^t dW(s) \, G(s)$ or $\int_0^t dW(s) \, G[W(s)]$, where we drop the $[\cdots]_\alpha$ notation since the Itô integral is specified. We then have:[2]

(i) The Itô integral $\int_0^t dW(s) \, F(s)$ exists for all smooth nonanticipating functions $F(s)$.

(ii) $[dW(t)]^2 = dt$ but $[dW(t)]^{2+2p} = 0$ for any $p > 0$. This is because

$$\int_0^t [dW(s)]^2 \, F(s) = \mathrm{ms-}\lim_{N \to \infty} \sum_{j=0}^{N-1} F_j \, (\Delta W_j)^2 = \int_0^t ds \, F(s) \,, \tag{3.25}$$

---

[2]See Gardiner §4.2.7.

and because $\left\langle (\Delta W_j)^{2+2p} \right\rangle \propto (\Delta t)^{1+p}$ for $p > 0$. For the same reason, we may neglect products such as $dt\, dW(t)$.

(iii) We see in (ii) that the mth power of the differential $dW(t)$ is negligible for $m > 2$. If, on the other hand, we take the differential of the mth power of $W(t)$, we obtain

$$
\begin{aligned}
d[W^m(t)] &= \big[W(t) + dW(t)\big]^m - \big[W(t)\big]^m \\
&= \sum_{k=1}^{m} \binom{m}{k} W^{m-k}(t) \big[dW(t)\big]^k \\
&= m\, W^{m-1}(t)\, dW(t) + \tfrac{1}{2} m(m-1)\, W^{m-2}(t)\, dt + o(dt^2) \, .
\end{aligned}
\tag{3.26}
$$

Evaluating the above expression for $m = n + 1$ and integrating, we have

$$
\begin{aligned}
\int_0^t d\big[W^{n+1}(s)\big] &= W^{n+1}(t) - W^{n+1}(0) \\
&= (n+1) \int_0^t dW(s)\, W^n(s) + \tfrac{1}{2} n(n+1) \int_0^t ds\, W^{n-1}(s) \, ,
\end{aligned}
\tag{3.27}
$$

and therefore

$$
\int_0^t dW(s)\, W^n(s) = \frac{W^{n+1}(t) - W^{n+1}(0)}{n+1} - \tfrac{1}{2} n \int_0^t ds\, W^{n-1}(s) \, .
\tag{3.28}
$$

(iv) Consider the differential of a function $f\big[W(t), t\big]$:

$$
\begin{aligned}
df\big[W(t), t\big] &= \frac{\partial f}{\partial W}\, dW + \frac{\partial f}{\partial t}\, dt + \frac{1}{2} \frac{\partial^2 f}{\partial W^2}\, (dW)^2 + \frac{\partial^2 f}{\partial W\, \partial t}\, dW\, dt + \frac{1}{2} \frac{\partial^2 f}{\partial t^2}\, (dt)^2 + \dots \\
&= \left(\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial W^2} \right) dt + \frac{\partial f}{\partial W}\, dW + o(dt) \, .
\end{aligned}
\tag{3.29}
$$

For example, for $f = \exp(W)$, we have $d\, e^{W(t)} = e^{W(t)} \big(dW(t) + \tfrac{1}{2} dt\big)$. This is known as *Itô's formula*. As an example of the usefulness of Itô's formula, consider the function $f\big[W(t), t\big] = W^2(t) - t$, for which Itô's formula yields $df = 2W\, dW$. Integrating the differential df, we thereby recover the result,

$$
\int_0^t dW(s)\, W(s) = \tfrac{1}{2} W^2(t) - \tfrac{1}{2} t \, .
\tag{3.30}
$$

(v) If $F(t)$ is nonanticipating, then

$$
\left\langle \int_0^t dW(s)\, F(s) \right\rangle = 0 \, .
\tag{3.31}
$$

Again, this is true for the Itô integral but not the Stratonovich integral.

(vi) The correlator of two Itô integrals of nonanticipating functions $F(s)$ and $G(s')$ is given by

$$\left\langle \int_0^t dW(s)\, F(s) \int_0^{t'} dW(s')\, G(s') \right\rangle = \int_0^{\tilde{t}} ds\, F(s)\, G(s) \,, \tag{3.32}$$

where $\tilde{t} = \min(t, t')$. This result was previously obtained by writing $dW(s) = \eta(s)\, ds$ and then invoking the correlator $\langle \eta(s)\, \eta(s') \rangle = \delta(s - s')$.

(vii) Oftentimes we encounter stochastic integrals in which the integrand contains a factor of $\delta(t - t_1)$ or $\delta(t - t_2)$, where the range of integration is the interval $[t_1, t_2]$. Appealing to the discretization defined in §3.3.2, it is straightforward to show

$$I_1 = \int_{t_1}^{t_2} dt\, f(t)\, \delta(t - t_1) = (1 - \alpha)\, f(t_1)$$

$$I_2 = \int_{t_1}^{t_2} dt\, f(t)\, \delta(t - t_2) = \alpha\, f(t_2) \,. \tag{3.33}$$

Thus, for Itô, $I_1 = f(t_1)$ and $I_2 = 0$, whereas for Stratonovich $I_1 = \frac{1}{2}\, f(t_1)$ and $I_2 = \frac{1}{2}\, f(t_2)$.

3.3.4 Fokker-Planck equation

We saw in §2.4 how the drift and diffusion relations

$$\langle \delta u(t) \rangle = F_1\big(u(t)\big)\, \delta t \qquad , \qquad \langle [\delta u(t)]^2 \rangle = F_2\big(u(t)\big)\, \delta t \,, \tag{3.34}$$

where $\delta u(t) = u(t + \delta t) - u(t)$, results in a Fokker-Planck equation

$$\frac{\partial P(u, t)}{\partial t} = -\frac{\partial}{\partial u}\Big[F_1(u)\, P(u, t) \Big] + \frac{1}{2}\, \frac{\partial^2}{\partial u^2}\Big[F_2(u)\, P(u, t) \Big] \,. \tag{3.35}$$

Consider now the differential Langevin equation

$$du = f(u, t)\, dt + g(u, t)\, dW(t) \,. \tag{3.36}$$

Let's integrate over the interval $[0, t]$, and work only to order t in $u(t) - u_0$, where $u_0 \equiv u(0)$. We then have

$$u(t) - u_0 = \int_0^t ds\, f\big(u(s)\big) + \int_0^t dW(s)\, g\big(u(s)\big)$$

$$= f(u_0)\, t + g(u_0) \int_0^t dW(s) + g'(u_0) \int_0^t dW(s)\, \big[u(s) - u_0\big] + \ldots \tag{3.37}$$

$$= f(u_0)\, t + g(u_0)\, W(t) + f(u_0)\, g'(u_0) \int_0^t dW(s)\, s + g(u_0)\, g'(u_0) \int_0^t dW(s)\, W(s) + \ldots \,,$$

where $W(t) = \int\limits_0^t ds\, \eta(s) = 0$, hence $W(0) = 0$. Averaging, we find

$$\langle u(t) - u_0 \rangle = f(u_0)\, t + \alpha\, g(u_0)\, g'(u_0)\, t + \ldots \tag{3.38}$$

and

$$\left\langle \left[u(t) - u_0 \right]^2 \right\rangle = g^2(u_0)\, t + \ldots \tag{3.39}$$

After a brief calculation, we obtain

$$\begin{aligned}
F_1(u) &= f(u) + \alpha\, g(u)\, g'(u) \\
F_2(u) &= g^2(u)\,.
\end{aligned} \tag{3.40}$$

We see how, for any choice other than the Itô value $\alpha = 0$, there is a noise-induced drift.

3.4 Stochastic Differential Equations

The general form we are considering is

$$du = f(u,t)\, dt + g(u,t)\, dW\,. \tag{3.41}$$

This is a *stochastic differential equation* (SDE). We are here concerned with (i) change of variables, (ii) multivariable formulations, and (iii) differences between Itô and Stratonovich solutions.

3.4.1 Itô change of variables formula

Suppose we change variables from u to $v(u,t)$. Then

$$\begin{aligned}
dv &= \frac{\partial v}{\partial t}\, dt + \frac{\partial v}{\partial u}\, du + \frac{1}{2}\frac{\partial^2 v}{\partial u^2}\,(du)^2 + o(dt) \\[2mm]
&= \left(\frac{\partial v}{\partial t} + f\,\frac{\partial v}{\partial u} + \tfrac{1}{2}g^2\,\frac{\partial^2 v}{\partial u^2} \right) dt + g\,\frac{\partial v}{\partial u}\, dW + o(dt)\,,
\end{aligned} \tag{3.42}$$

where we have used $(dW)^2 = dt$. Note that if $v = v(u)$ we do not have the $\frac{\partial v}{\partial t}\, dt$ term. This change of variables formula is only valid for the Itô case. In §3.4.5 below, we will derive the corresponding result for the Stratonovich case, and show that it satisfies the familiar chain rule.

EXERCISE: Derive the change of variables formula for general α. *Hint:* First integrate the SDE over a small but finite time interval Δt_j to obtain

$$\begin{aligned}
\Delta u_j &= f_j\,\Delta t_j + \left[(1-\alpha)\, g_j + \alpha\, g_{j+1} \right] \Delta W_j \\[1mm]
&= \left[f_j + \alpha\, g_j\, g_j' \right] \Delta t_j + g_j\,\Delta W_j\quad,
\end{aligned} \tag{3.43}$$

up to unimportant terms, where $u_j = u(t_j)$, $f_j = f(u_j, t_j)$, $g_j = g(u_j, t_j)$, and $g_j' = \frac{\partial g}{\partial u}\big|_{u_j, t_j}$.

Example: Kubo oscillator

As an example, consider the Kubo oscillator[3],

$$du = i\omega\, u\, dt + i\lambda\, u\, dW \ . \tag{3.44}$$

This can be interpreted as a linear oscillator with a fluctuating frequency. If $\lambda = 0$, we have $\dot{u} = i\omega u$, with solution $u(t) = u(0)\, e^{i\omega t}$. We now implement two changes of variables:

 (i) First, we define $v = u\, e^{-i\omega t}$. Plugging this into Eqn. 3.42, we obtain

$$dv = i\lambda\, v\, dW \ . \tag{3.45}$$

 (ii) Second, we write $y = \ln v$. Appealing once again to the Itô change of variables formula, we find

$$dy = \tfrac{1}{2}\lambda^2\, dt + i\lambda\, dW \ . \tag{3.46}$$

The solution is therefore

$$y(t) = y(0) + \tfrac{1}{2}\lambda^2 t + i\lambda\, W(t) \quad \Longrightarrow \quad u(t) = u(0)\, e^{i\omega t}\, e^{\lambda^2 t/2}\, e^{i\lambda W(t)} \ . \tag{3.47}$$

Averaging over the Gaussian random variable W, we have

$$\langle u(t) \rangle = u(0)\, e^{i\omega t}\, e^{\lambda^2 t/2}\, e^{-\lambda^2 \langle W^2(t) \rangle /2} = u(0)\, e^{i\omega t} \ . \tag{3.48}$$

Thus, the average of $u(t)$ behaves as if it is unperturbed by the fluctuating piece. *There is no noise-induced drift.* We can also compute the correlator,

$$\langle u(t)\, u^*(t') \rangle = \big|u(0)\big|^2\, e^{i\omega(t-t')}\, e^{\lambda^2 \min(t,t')} \ . \tag{3.49}$$

Thus, $\langle \big|u(t)\big|^2 \rangle = \big|u(0)\big|^2\, e^{\lambda^2 t}$. If $u(0)$ is also a stochastic variable, we must average over it as well.

3.4.2 Solvability by change of variables

Following Riecke[4], we ask under what conditions the SDE $du = f(u,t)\, dt + g(u,t)\, dW$ can be transformed to

$$dv = \alpha(t)\, dt + \beta(t)\, dW \ , \tag{3.50}$$

which can be directly integrated via Itô. From Itô's change of variables formula Eqn. 3.42, we have

$$dv = \left(\frac{\partial v}{\partial t} + f\, \frac{\partial v}{\partial u} + \tfrac{1}{2} g^2\, \frac{\partial^2 v}{\partial u^2} \right) dt + g\, \frac{\partial v}{\partial u}\, dW \ , \tag{3.51}$$

hence

$$\alpha(t) = \frac{\partial v}{\partial t} + f\, \frac{\partial v}{\partial u} + \tfrac{1}{2} g^2\, \frac{\partial^2 v}{\partial u^2} \quad , \quad \beta(t) = g\, \frac{\partial v}{\partial u} \ . \tag{3.52}$$

---

[3]See Riecke, §5.4.1 and Gardiner §4.5.3.
[4]See Riecke, §5.4.2.

We therefore have

$$\frac{\partial v}{\partial u} = \frac{\beta(t)}{g(u,t)} \qquad \Rightarrow \qquad \frac{\partial^2 v}{\partial u^2} = -\frac{\beta}{g^2}\frac{\partial g}{\partial u} \qquad , \qquad \frac{\partial^2 v}{\partial u\,\partial t} = \frac{1}{g}\frac{d\beta}{dt} - \frac{\beta}{g^2}\frac{\partial g}{\partial t} \ . \tag{3.53}$$

Setting $\partial\alpha/\partial u = 0$ then results in

$$\frac{1}{g}\frac{d\beta}{dt} - \frac{\beta}{g^2}\frac{\partial g}{\partial t} + \frac{\partial}{\partial u}\left[\beta\,\frac{f}{g} - \tfrac{1}{2}\beta\,\frac{\partial g}{\partial u}\right] = 0 \ , \tag{3.54}$$

or

$$\frac{d\ln\beta}{dt} = \frac{\partial\ln g}{\partial t} - g\,\frac{\partial}{\partial u}\left(\frac{f}{g}\right) + \tfrac{1}{2}\,g\,\frac{\partial^2 g}{\partial u^2} \ . \tag{3.55}$$

The LHS of the above equation is a function of t alone, hence the solvability condition becomes

$$\frac{\partial}{\partial u}\left[\frac{\partial\ln g}{\partial t} - g\,\frac{\partial}{\partial u}\left(\frac{f}{g}\right) + \tfrac{1}{2}\,g\,\frac{\partial^2 g}{\partial u^2}\right] = 0 \ . \tag{3.56}$$

If the above condition holds, one can find a u-independent $\beta(t)$, and from the second of Eqn. 3.52 one then obtains $\partial v/\partial u$. Plugging this into the first of Eqn. 3.52 then yields $\alpha(t)$, which is itself guaranteed to be u-independent.

3.4.3   Multicomponent SDE

Let $\boldsymbol{u} = \{u_1, \ldots, u_K\}$ and consider the SDE

$$du_a = A_a\,dt + B_{ab}\,dW_b \ , \tag{3.57}$$

where repeated indices are summed over, and where

$$\langle dW_b\,dW_c \rangle = \delta_{bc}\,dt \ . \tag{3.58}$$

Now suppose $f(\boldsymbol{u})$ is a scalar function of the collection $\{u_1, \ldots, u_K\}$. We then have

$$\begin{aligned}
df &= \frac{\partial f}{\partial u_a}\,du_a + \frac{1}{2}\,\frac{\partial^2 f}{\partial u_a\,\partial u_b}\,du_a\,du_b + o(dt) \\
&= \frac{\partial f}{\partial u_a}\left(A_a\,dt + B_{ab}\,dW_b\right) + \frac{1}{2}\,\frac{\partial^2 f}{\partial u_a\,\partial u_b}\left(A_a\,dt + B_{aa'}\,dW_{a'}\right)\left(A_b\,dt + B_{bb'}\,dW_{b'}\right) + o(dt) \\
&= \left[A_a\,\frac{\partial f}{\partial u_a} + \frac{1}{2}\,\frac{\partial^2 f}{\partial u_a\,\partial u_b}\left(BB^{\mathrm{t}}\right)_{ba}\right]dt + \frac{\partial f}{\partial u_a}\,B_{ab}\,dW_b + o(dt) \ .
\end{aligned} \tag{3.59}$$

We also may derive the Fokker-Planck equation,

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial u_a}\left(A_a P\right) + \frac{1}{2}\,\frac{\partial^2}{\partial u_a\,\partial u_b}\left[\left(BB^{\mathrm{t}}\right)_{ab}P\right] \ . \tag{3.60}$$

3.4.4 SDEs with general α expressed as Itô SDEs ($\alpha = 0$)

We return to the single component case and the SDE

$$du = f(u,t)\, dt + g(u,t)\, dW(t) \quad . \tag{3.61}$$

Formally, we can write

$$u(t) - u(0) = \int_0^t ds\, f\big(u(s), s\big) + \int_0^t dW(s)\, g\big(u(s), s\big) \ . \tag{3.62}$$

The second term on the RHS is defined via its discretization, with

$$\int_0^t dW(s)\, \Big[g\big(u(s), s\big)\Big]_\alpha \equiv \text{ms} - \lim_{N \to \infty} \sum_{j=0}^{N-1} g\big((1-\alpha)u_j + \alpha\, u_{j+1}, t_j\big)\, \Delta W_j$$

$$= \text{ms} - \lim_{N \to \infty} \sum_{j=0}^{N-1} \left[g(u_j, t_j)\, \Delta W_j + \alpha\, \frac{\partial g}{\partial u}(u_j, t_j)\, (u_{j+1} - u_j)\, \Delta W_j \right] \ . \tag{3.63}$$

Now if u satisfies the SDE $du = f\, dt + g\, dW$, then

$$u_{j+1} - u_j = f(u_j, t_j)\, \Delta t_j + g(u_j, t_j)\, \Delta W_j \ , \tag{3.64}$$

where $\Delta t_j = t_{j+1} - t_j$, and inserting this into the previous equation gives

$$\int_0^t dW(s)\, \Big[g\big(u(s), s\big)\Big]_\alpha = \text{ms} - \lim_{N \to \infty} \sum_{j=0}^{N-1} \left[g(u_j, t_j)\, \Delta W_j + \alpha\, f(u_j, t_j)\, \frac{\partial g}{\partial u}(u_j, t_j)\, \Delta t_j\, \Delta W_j + \alpha\, g(u_j, t_j)\, \frac{\partial g}{\partial u}(u_j, t_j)\, (\Delta W_j)^2 \right]$$

$$= \int_0^t dW(s)\, \Big[g\big(u(s), s\big)\Big]_0 + \alpha \int_0^t ds\, g\big(u(s), s\big)\, \frac{\partial g}{\partial u}\big(u(s), s\big) \ , \tag{3.65}$$

where the stochastic integral with $\alpha = 0$ found on the last line above is the Itô integral. Thus, the solution of the stochastic differential equation Eqn. 3.61, using the prescription of stochastic integration for general α, is equivalent to the solution using the Itô prescription ($\alpha = 0$) if we substitute

$$f_{\text{I}}(u,t) = f(u,t) + \alpha\, g(u,t)\, \frac{\partial g(u,t)}{\partial u} \quad , \qquad g_{\text{I}}(u,t) = g(u,t) \ , \tag{3.66}$$

where the I subscript denotes the Itô case. In particular, since $\alpha = \frac{1}{2}$ for the Stratonovich case,

$$du = f\, dt + g\, dW \quad \text{[Itô]} \qquad \Longrightarrow \qquad du = \left(f - \tfrac{1}{2}\, g\, \frac{\partial g}{\partial u} \right) dt + g\, dW \quad \text{[Stratonovich]}$$

$$du = f\, dt + g\, dW \quad \text{[Stratonovich]} \qquad \Longrightarrow \qquad du = \left(f + \tfrac{1}{2}\, g\, \frac{\partial g}{\partial u} \right) dt + g\, dW \quad \text{[Itô]} \ .$$

Kubo oscillator as a Stratonovich SDE

Consider the case of the Kubo oscillator, for which $f = i\omega u$ and $g = i\lambda u$. Viewed as a Stratonovich SDE, we transform to Itô form to obtain

$$du = \left(i\omega - \tfrac{1}{2}\lambda^2\right)u\,dt + i\lambda u\,dW \ . \tag{3.67}$$

Solving as in §3.4.1, we find

$$u(t) = u(0)\,e^{i\omega t}\,e^{i\lambda W(t)} \ , \tag{3.68}$$

hence

$$\langle u(t) \rangle = u(0)\,e^{i\omega t}\,e^{-\lambda^2 t/2} \qquad , \qquad \langle u(t)\,u^*(t') \rangle = \left|u(0)\right|^2 e^{i\omega(t-t')}\,e^{-\lambda^2|t-t'|/2} \ . \tag{3.69}$$

We see that there is noise-induced drift and decay in the Stratonovich case.

Multivariable case

Suppose we have

$$\begin{aligned}
du_a &= A_a\,dt + B_{ab}\,dW_b \quad (\alpha-\text{discretization}) \\
&= \widetilde{A}_a\,dt + \widetilde{B}_{ab}\,dW_b \quad (\text{Itô}) \ .
\end{aligned} \tag{3.70}$$

Using $\langle\,dW_a\,dW_b\,\rangle = \delta_{ab}\,dt$, applying the above derivation in §3.4.3, we obtain

$$\widetilde{A}_a = A_a + \alpha\,\frac{\partial B_{ac}}{\partial u_b}\,B^{\mathrm{t}}_{cb} \qquad , \qquad \widetilde{B}_{ab} = B_{ab} \ , \tag{3.71}$$

where repeated indices are summed. The resulting Fokker-Planck equation is then

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial u_a}\left[\left(A_a + \alpha\,\frac{\partial B_{ac}}{\partial u_b}\,B^{\mathrm{t}}_{cb}\right)P\right] + \frac{1}{2}\frac{\partial^2}{\partial u_a\,\partial u_b}\left[(BB^{\mathrm{t}})_{ab}\,P\right] \ . \tag{3.72}$$

When $\alpha = \tfrac{1}{2}$, we obtain the Stratonovich form,

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial u_a}\Big(A_a\,P\Big) + \frac{1}{2}\frac{\partial}{\partial u_a}\left[B_{ac}\,\frac{\partial}{\partial u_b}\Big(B^{\mathrm{t}}_{cb}\,P\Big)\right] \ . \tag{3.73}$$

3.4.5   Change of variables in the Stratonovich case

We saw in Eqn. 3.42 how a change of variables leads to a new SDE in the Itô case. What happens in the Stratonovich case? To see this, we write the Stratonovich SDE,

$$du = f\,dt + g\,dW \ , \tag{3.74}$$

in its Itô form,

$$du = \left(f + \tfrac{1}{2}\,g\,\frac{\partial g}{\partial u}\right)dt + g\,dW \ , \tag{3.75}$$

and now effect the change of variables $v = v(u)$. We leave the general case of $v = v(u, t)$ to the student. Applying Eqn. 3.42, we find

$$
\begin{aligned}
dv &= \left[\left(f + \frac{1}{2} g \frac{\partial g}{\partial u} \right) \frac{dv}{du} + \frac{1}{2} \frac{d^2v}{du^2} g^2 \right] dt + \frac{\partial v}{\partial u} g \, dW \\
&= \left[\frac{f}{u'} + \frac{1}{2} \frac{\partial g}{\partial v} \frac{g}{(u')^2} - \frac{1}{2} \frac{g^2 \, u''}{(u')^3} \right] dt + \frac{g}{u'} \, dW \, ,
\end{aligned}
\tag{3.76}
$$

where $u' = du/dv$ and $u'' = d^2u/dv^2$. Now that everything in the last line above is expressed in terms of v and t, we transform back to the Stratonovich form, resulting in

$$
dv = \tilde{f} \, dt + \tilde{g} \, dW \, ,
\tag{3.77}
$$

with

$$
\tilde{f} = \frac{f}{u'} + \frac{1}{2} \frac{\partial g}{\partial v} \frac{g}{(u')^2} - \frac{1}{2} \frac{g^2 \, u''}{(u')^3} - \frac{1}{2} \left(\frac{g}{u'} \right) \frac{\partial}{\partial v} \left(\frac{g}{u'} \right) = \frac{f}{u'}
\tag{3.78}
$$

and

$$
\tilde{g} = \frac{g}{u'} \, .
\tag{3.79}
$$

Thus,

$$
dv = \frac{1}{u'} \left[f \, dt + g \, dW \right] = \frac{dv}{du} \, du \, ,
\tag{3.80}
$$

which satisfies the familiar chain rule!

3.5 Applications

3.5.1 Ornstein-Uhlenbeck redux

The Ornstein-Uhlenbeck process is described by the SDE

$$
dx = -\beta \, x \, dt + \sqrt{2D} \, dW(t) \, .
\tag{3.81}
$$

Since the coefficient of dW is independent of x, this equation is the same when the Itô prescription is taken. Changing variables to $y = x \, e^{\beta t}$, we have

$$
dy = \sqrt{2D} \, e^{\beta t} \, dW(t) \, ,
\tag{3.82}
$$

with solution

$$
x(t) = x(0) \, e^{-\beta t} + \sqrt{2D} \int_0^t dW(s) \, e^{-\beta(t-s)} \, .
\tag{3.83}
$$

We may now compute

$$
\langle x(t) \rangle = x(0) \, e^{-\beta t} \qquad , \qquad \left\langle \left(x(t) - x(0) \, e^{-\beta t} \right)^2 \right\rangle = \frac{D}{\beta} \left(1 - e^{-2\beta t} \right) \, .
\tag{3.84}
$$

The correlation function is also easily calculable:

$$
\begin{aligned}
\left\langle x(t)\, x(t')\right\rangle_{\mathrm{c}} &= \left\langle x(t)\, x(t')\right\rangle - \left\langle x(t)\right\rangle\left\langle x(t')\right\rangle \\[2mm]
&= 2D \left\langle \int\limits_{0}^{t} dW(s)\, e^{-\beta(t-s)} \int\limits_{0}^{t'} dW(s')\, e^{-\beta(t'-s')} \right\rangle \\[2mm]
&= 2D\, e^{-\beta(t+t')} \int\limits_{0}^{\min(t,t')} ds\, e^{2\beta s} = \frac{D}{\beta}\left(e^{-\beta|t-t'|} - e^{-\beta(t+t')} \right).
\end{aligned}
\tag{3.85}
$$

3.5.2 Time-dependence

Consider the SDE,

$$
du = \alpha(t)\, u\, dt + \beta(t)\, u\, dW(t) .
\tag{3.86}
$$

Writing $v = \ln u$ and appealing the the Itô change of variables formula in Eqn. 3.42, we have

$$
dv = \left(\alpha(t) - \tfrac{1}{2}\beta^2(t) \right) dt + \beta(t)\, dW(t) ,
\tag{3.87}
$$

which may be directly integrated to yield

$$
u(t) = u(0)\, \exp\left\{ \int\limits_{0}^{t} ds\, \left[\alpha(s) - \tfrac{1}{2}\beta^2(s) \right] + \int\limits_{0}^{t} dW(s)\, \beta(s) \right\} .
\tag{3.88}
$$

Using the general result for the average of the exponential of a Gaussian random variable, $\left\langle \exp(\phi) \right\rangle = \exp\left(\tfrac{1}{2}\langle \phi^2 \rangle \right)$, we have

$$
\left\langle u^n(t) \right\rangle = u^n(0)\, \exp\left\{ \int\limits_{0}^{t} ds\, \left[n\,\alpha(s) + \tfrac{1}{2}n(n-1)\,\beta^2(s) \right] \right\} .
\tag{3.89}
$$

3.5.3 Colored noise

We can model colored noise using the following artifice[5]. We saw above how the Ornstein-Uhlenbeck process yields a correlation function

$$
C(s) = \left\langle u(t)\, u(t+s) \right\rangle = \frac{D}{\beta}\, e^{-\beta|s|} ,
\tag{3.90}
$$

in the limit $t \to \infty$. This means that the spectral function is

$$
\hat{C}(\omega) = \int\limits_{-\infty}^{\infty} ds\, C(s)\, e^{-i\omega s} = \frac{2D}{\beta^2 + \omega^2} ,
\tag{3.91}
$$

---

[5]See Riecke §5.6.

which has spectral variation. We henceforth set $2D \equiv \beta^2$ so that $C(s) = \frac{1}{2}\beta\, e^{-\beta|s|}$, and $\hat{C}(\omega) = \beta^2/(\beta^2 + \omega^2)$. Note that $\hat{C}(0) = \int\limits_{-\infty}^{\infty} ds\, C(s) = 1$.

Consider now a quantity $x(t)$ which is *driven* by the OU process, *viz.*

$$
\begin{aligned}
du &= -\beta\, u\, dt + \beta\, dW(t) \\
\frac{dx}{dt} &= a(t)\, x + b(t)\, u(t)\, x \, ,
\end{aligned}
\tag{3.92}
$$

where $a(t)$ and $b(t)$ may be time-dependent. The second of these is an ordinary differential equation and not a SDE since $u(t)$ is a continuous function, even though it is stochastic. As we saw above, the solution for $u(t)$ is

$$
u(t) = u(0)\, e^{-\beta t} + \beta \int\limits_0^t dW(s)\, e^{-\beta(t-s)} \, .
\tag{3.93}
$$

Therefore

$$
x(t) = x(0)\, \exp\left\{ \int\limits_0^t ds\, a(s) + u(0)\int\limits_0^t ds\, b(s)\, e^{-\beta s} + \beta \int\limits_0^t ds\, b(s) \int\limits_0^s dW(s')\, e^{-\beta(s-s')} \right\} \, .
\tag{3.94}
$$

It is convenient to reexpress the last term in brackets such that

$$
x(t) = x(0)\, \exp\left\{ \int\limits_0^t ds\, a(s) + u(0)\int\limits_0^t ds\, b(s)\, e^{-\beta s} + \beta \int\limits_0^t dW(s') \int\limits_{s'}^t ds\, b(s)\, e^{-\beta(s-s')} \right\} \, .
\tag{3.95}
$$

Now let us take the $\beta \to \infty$ limit. We know that for any smooth function $f(s)$ that

$$
\lim_{\beta\to\infty} \beta \int\limits_{s'}^t ds\, b(s)\, e^{-\beta(s-s')} = b(s') \, ,
\tag{3.96}
$$

hence

$$
\lim_{\beta\to\infty} x(t) = x(0)\, \exp\left\{ \int\limits_0^t ds\, a(s) + \int\limits_0^t dW(s)\, b(s) \right\} \, .
\tag{3.97}
$$

Now since $\langle u(t)\, u(t')\rangle = C(t-t') = \delta(t-t')$ in the $\beta \to \infty$ limit, we might as well regard $x(t)$ as being stochastically forced by a Wiener process and describe its evolution using the SDE,

$$
dx = a(t)\, x\, dt + b(t)\, x\, dW(t) \qquad (\alpha = ??) \, .
\tag{3.98}
$$

As we have learned, the integration of SDEs is a negotiable transaction, which requires fixing a value of the interval parameter α. What value of α do we mean for the above equation? We can establish this by transforming it to an Itô SDE with $\alpha = 0$, using the prescription in Eqn. 3.66. Thus, with α as yet undetermined, the Itô form of the above equation is

$$
dx = \left[a(t) + \alpha\, b^2(t) \right] x\, dt + b(t)\, x\, dW(t) \, .
\tag{3.99}
$$

Now we use the Itô change of variables formula 3.42 to write this as a SDE for $y = \ln x$:

$$dy = \Big[a(t) + (\alpha - \tfrac{1}{2})\, b^2(t)\Big]\, dt + b(t)\, dW(t)\,, \tag{3.100}$$

which may be integrated directly, yielding

$$x(t) = x(0)\, \exp\left\{\int_0^t ds\,\Big[a(s) + (\alpha - \tfrac{1}{2})\, b^2(s)\Big] + \int_0^t dW(s)\, b(s)\right\}\,. \tag{3.101}$$

Comparing with Eqn. 3.97, we see that $\alpha = \tfrac{1}{2}$, *i.e.* Stratonovich form.

Finally, what of the correlations? Consider the case where $a(t) \to i\nu$ and $b(t) \to i\lambda$ are complex constants, in which case we have a colored noise version of the Kubo oscillator:

$$\begin{aligned} du &= -\beta\, u\, dt + \beta\, dW(t) \\ \frac{dz}{dt} &= i\nu\, z + i\lambda\, u(t)\, z\,. \end{aligned} \tag{3.102}$$

The solution is

$$z(t) = z(0)\, \exp\left\{i\nu t + \frac{i\lambda}{\beta}\, u(0)\, \big(1 - e^{-\beta t}\big) + i\lambda \int_0^t dW(s)\, \big(1 - e^{-\beta(t-s)}\big)\right\}\,. \tag{3.103}$$

This matches the Stratonovich solution to the Kubo oscillator, $z(t) = z(0)\, e^{i\nu t}\, e^{i\lambda W(t)}$ in the limit $\beta \to \infty$, as we should by now expect. The average oscillator coordinate is

$$\big\langle z(t)\big\rangle = z(0)\, \exp\left\{i\nu t + \frac{i\lambda}{\beta}\, u(0)\, \big(1 - e^{-\beta t}\big) - \tfrac{1}{2}\lambda^2 t + \frac{\lambda^2}{2\beta}\big(1 - e^{-\beta t}\big)^2\right\}\,. \tag{3.104}$$

As $\beta \to \infty$ we recover the result from Eqn. 3.69. For $\beta \to 0$, the stochastic variable $u(t)$ is fixed at $u(0)$, and $z(t) = z(0)\, \exp\big(i[\nu + \lambda\, u(0)]\, t\big)$, which is correct.

Let's now compute the correlation function $\big\langle z(t)\, z^*(t')\big\rangle$ in the limit $t, t' \to \infty$, where it becomes a function of $t - t'$ alone due to decay of the transients arising from the initial conditions. It is left as an exercise to the reader to show that

$$Y(s) = \lim_{t \to \infty} \big\langle z(t+s)\, z^*(t)\big\rangle = |z(0)|^2\, \exp\left\{i\nu s - \tfrac{1}{2}\lambda^2\, |s| + \frac{\lambda^2}{2\beta}\big(1 - e^{-\beta|s|}\big)\right\}\,. \tag{3.105}$$

As $\beta \to \infty$, we again recover the result from Eqn. 3.69, and for $\beta = 0$ (which is taken *after* $t \to \infty$), we also obtain the expected result. We see that the coloration of the noise affects the correlator $Y(s)$, resulting in a different time dependence and hence a different spectral function $\hat{Y}(\omega)$.

3.5.4 Remarks about financial markets

Let p be the price of a financial asset, such as a single share of stock. We model the dynamics of $p(t)$ by a stochastic process described by the SDE

$$dp = r(p,t)\, dt + \sqrt{2D(p,t)}\, dW(t)\,, \qquad (3.106)$$

where $r(p,t)$ and $D(p,t)$ represent drift and diffusion terms. We might set $r(p,t) = \mu(t)\, p$, where $\mu(t)$ is the current interest rate being paid by banks. What about diffusion? In the late 1950's, M. Osborne noted that stock prices are approximately log-normally distributed. To model this, we can take $D = \frac{1}{2}\lambda^2 p^2$. Thus, our SDE is

$$dp = \mu\, p\, dt + \lambda\, p\, dW(t)\,. \qquad (3.107)$$

As we shall now see, this will lead to some problematic consequences.

We've solved this equation many times before. Changing variables to $x = \ln p$, we have $dx = \left(\mu - \frac{1}{2}\lambda^2\right) dt + \lambda\, dW$, and assuming μ and λ are time-independent, we have

$$p(t) = p(0)\, e^{\mu t}\, e^{-\lambda^2 t/2}\, e^{\lambda W(t)}\,. \qquad (3.108)$$

Averaging, we obtain the moments

$$\big\langle p^n(t)\big\rangle = p^n(0)\, e^{n\mu t}\, e^{n(n-1)\lambda^2 t/2}\,. \qquad (3.109)$$

To appreciate the consequences of this result, let's compute the instantaneous variance,

$$\begin{aligned}
\mathsf{Var}\, p(t) &= \big\langle p^2(t)\big\rangle - \big\langle p(t)\big\rangle^2 \\
&= p^2(0)\, e^{2\mu t}\big(e^{\lambda^2 t} - 1\big)\,.
\end{aligned} \qquad (3.110)$$

The ratio of the standard deviation to the mean is therefore growing exponentially, and the distribution keeps getting broader *ad infinitum*.

Another way to see what is happening is to examine the associated Fokker-Planck equation,

$$\frac{\partial P}{\partial t} = -\mu\, \frac{\partial}{\partial p}\big(pP\big) + \tfrac{1}{2}\lambda^2\, \frac{\partial^2}{\partial p^2}\big(p^2 P\big)\,. \qquad (3.111)$$

Let's look for a stationary solution by setting the LHS to zero. We integrate once on p to cancel one power of $\frac{d}{dp}$, and set the associated constant of integration to zero, because $P(p=\infty,t)=0$. This leaves

$$\frac{d}{dp}\big(p^2 P\big) = \frac{2\mu}{\lambda^2}\, pP = \frac{2\mu}{\lambda^2 p}\,\big(p^2 P\big)\,. \qquad (3.112)$$

The solution is a power law,

$$P(p) = C\, p^{2\mu\lambda^{-2} - 2}\,, \qquad (3.113)$$

However, no pure power law distribution is normalizable on the interval $[0,\infty)$, so there is no meaningful steady state for this system. If markets can be modeled by such a stochastic differential equation, then this result is a refutation of Adam Smith's "invisible hand", which is the notion that markets should in time approach some sort of stable equilibrium.

Stochastic variance

A more realistic model is obtained by writing[6]

$$dp = \mu\, p\, dt + \sqrt{v(p,t)}\, p\, dW(t)\,, \tag{3.114}$$

where $v(p,t)$ is strongly nonlinear and nonseparable in p and t. Another approach is to assume the variance $v(t)$ is itself stochastic. We write

$$\begin{aligned} dp &= \mu\, p\, dt + \sqrt{v(t)}\, p\, dW(t) \\ dv &= f(p,v,t)\, dt + g(p,v,t) \left[\cos\theta\, dW(t) + \sin\theta\, dY(t)\right], \end{aligned} \tag{3.115}$$

where $W(t)$ and $Y(t)$ are independent Wiener processes. The variance $v(t)$ of stock prices is observed to relax on long-ish time scales of $\gamma^{-1} \approx 22$ days. This is particularly true for aggregate quantities such as market indices (*e.g.* the Dow-Jones Industrial Average (DJIA) or the Deutscher Aktien-Index (DAX)). One typically assumes

$$f(p,v,t) = \gamma(v_\infty - v)\,, \tag{3.116}$$

describing a drift toward $v = v_\infty$, similar to the drift in the Ornstein-Uhlenbeck model. As for the diffusive term $g(p,v,t)$, two popular models are the Heston and Hull-White models:

$$g(p,v,t) = \begin{cases} \kappa\sqrt{v} & \text{Heston} \\ \beta v & \text{Hull-White}\,. \end{cases} \tag{3.117}$$

Empirically, $\theta \approx \frac{\pi}{2}$, which we shall henceforth assume.

The Fokker-Planck equation for the distribution of the variance, $P(v,t)$, is given by

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial v}\left[\gamma(v - v_\infty)\, P\right] + \frac{1}{2}\frac{\partial^2}{\partial v^2}\left[g^2(v)P\right]\,. \tag{3.118}$$

We seek a steady state solution for which the LHS vanishes. Assuming $vP(v) \to 0$ for $v \to \infty$, we integrate setting the associated constant of integration to zero. This results in the equation

$$\frac{d}{dv}\left[g^2(v)\, P(v)\right] = 2\gamma\left(\frac{v_\infty - v}{g^2(v)}\right) g^2(v)\, P(v)\,, \tag{3.119}$$

with solution

$$P(v) = \frac{1}{g^2(v)}\, \exp\left\{2\gamma \int^{v} dv' \left(\frac{v_\infty - v'}{g^2(v')}\right)\right\}\,. \tag{3.120}$$

For the Heston model, we find

$$P_{\text{H}}(v) = C_{\text{H}}\, v^{(2\gamma v_\infty \kappa^{-2} - 1)}\, e^{-2\gamma v/\gamma^2}\,, \tag{3.121}$$

whereas for the Hull-White model,

$$P_{\text{HW}}(v) = C_{\text{HW}}\, v^{-2(1+\gamma\beta^{-2})}\, e^{-2\gamma v_\infty/\beta^2 v}\,. \tag{3.122}$$

---

[6]See the discussion in McCauley, §4.5 and chapter 6.
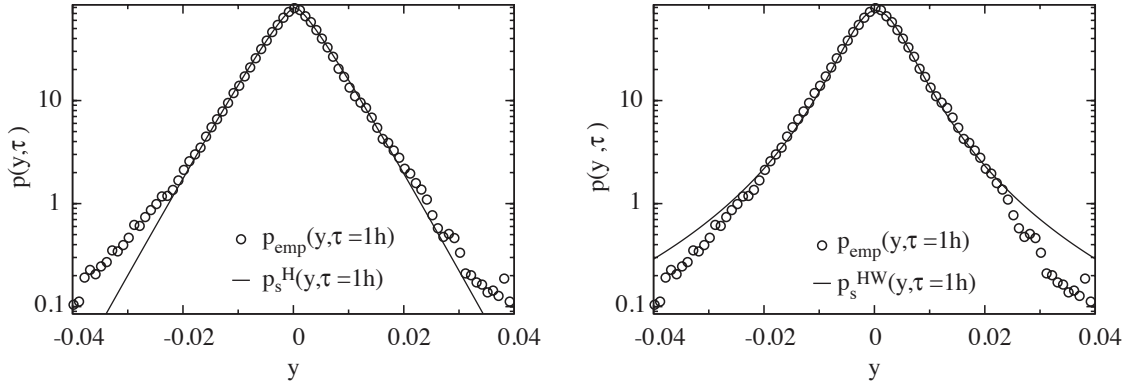
Figure 3.1: Comparison of predictions of the Heston model (left) and the Hull-White model (right) with the empirical probability distribution $P(y, \tau)$ for logarithmic returns of the German DAX index between Feb. 5, 1996 and Dec. 28, 2001 (open circles). Parameters for the Heston model are $r = 1.36$, $v_\infty = 5.15 \times 10^{-5}\,\mathrm{h}^{-1}$, $\mu = 3.03 \times 10^{-4}\,\mathrm{h}^{-1}$. Parameters for the Hull-White model are $s = 0.08$, $v_\infty = 3.21 \times 10^{-4}\,\mathrm{h}^{-1}$, and $\mu = 2.97 \times 10^{-4}\,\mathrm{h}^{-1}$. The time interval was taken to be $\tau = 1\,\mathrm{h}$. From R. Remer and R. Mahnke, *Physica A* **344**, 236 (2004).

Note that both distributions are normalizable. The explicit normalized forms are:

$$
\begin{aligned}
P_{\mathrm{H}}(v) &= \frac{r^r}{\Gamma(r)\, v_\infty} \left(\frac{v}{v_\infty} \right)^{r-1} \exp\left(-rv/v_\infty \right) \\
P_{\mathrm{HW}}(v) &= \frac{s^s}{\Gamma(s)\, v_\infty} \left(\frac{v_\infty}{v} \right)^{s+2} \exp\left(-sv_\infty/v \right) ,
\end{aligned}
\tag{3.123}
$$

with $r = 2\gamma v_\infty/\kappa^2$ and $s = 2\gamma/\beta^2$. Note that the tails of the Heston model variance distribution are exponential with a power law prefactor, while those of the Hull-White model are power law "fat tails".

The SDE for the logarithmic price $x = \ln\big[p(t)/p(0)\big]$, obtained from Itô's change of variables formula, is

$$
dx = \tilde{\mu}\, dt + \sqrt{v}\, dW(t) ,
\tag{3.124}
$$

where $\tilde{\mu} = \mu - \frac{1}{2} v$. Here we assume that v is approximately constant in time as $x(t)$ fluctuates. This is akin to the Born-Oppenheimer approximation in quantum mechanics – we regard $v(t)$ as the "slow variable" and $x(t)$ as the "fast variable". Integrating this over a short time interval τ, we have

$$
y = \tilde{\mu}\tau + \sqrt{v}\, \Delta W ,
\tag{3.125}
$$

with $y = x(t+\tau) - x(t)$ and $\Delta W = W(t+\tau) - W(t)$. This says that $y - \tilde{\mu}\tau$ is distributed normally with variance $\big\langle (\sqrt{v}\, \Delta W)^2 \big\rangle = v\tau$, hence

$$
P(y, \tau \,|\, v) = (2\pi v\tau)^{-1/2} \exp\left\{ -\frac{(y - \tilde{\mu}\tau)^2}{2v\tau} \right\} .
\tag{3.126}
$$

To find the distribution $P(y, \tau)$ of the logarithmic returns y, we must integrate over v with a weight $P(v)$, the steady state distribution of the variance:

$$P(y, \tau) = \int\limits_0^\infty dv \, P(y, \tau \,|\, v) \, P(v) \,. \tag{3.127}$$

The results for the Heston and Hull-White models are shown in Fig. 3.1, where they are compared with empirical data from the DAX.

Chapter 4

The Fokker-Planck and Master Equations

4.1 References

- C. Gardiner, *Stochastic Methods* (4th edition, Springer-Verlag, 2010)
  Very clear and complete text on stochastic methods, with many applications.

- N. G. Van Kampen *Stochastic Processes in Physics and Chemistry* (3rd edition, North-Holland, 2007)
  Another standard text. Very readable, but less comprehensive than Gardiner.

- Z. Schuss, *Theory and Applications of Stochastic Processes* (Springer-Verlag, 2010)
  In-depth discussion of continuous path stochastic processes and connections to partial differential equations.

- R. Mahnke, J. Kaupužs, and I. Lubashevsky, *Physics of Stochastic Processes* (Wiley, 2009)
  Introductory sections are sometimes overly formal, but a good selection of topics.

4.2   Fokker-Planck Equation

Here we mainly follow the discussion in chapter 5 of Gardiner, and chapter 4 of Mahnke *et al.*

4.2.1   Forward and backward time equations

We have already met the Fokker-Planck equation,

$$\frac{\partial P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')}{\partial t} = -\frac{\partial}{\partial x_i}\Big[A_i(\boldsymbol{x},t)\,P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')\Big] + \frac{1}{2}\frac{\partial^2}{\partial x_i\,\partial x_j}\Big[B_{ij}(\boldsymbol{x},t)\,P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')\Big] . \tag{4.1}$$

Defining the probability flux,

$$J_i(\boldsymbol{x},t\,|\,\boldsymbol{x}',t') = A_i(\boldsymbol{x},t)\,P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t') - \frac{1}{2}\frac{\partial}{\partial x_j}\Big[B_{ij}(\boldsymbol{x},t)\,P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')\Big] , \tag{4.2}$$

the Fokker-Planck equation takes the form of the continuity equation,

$$\frac{\partial P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{J}(\boldsymbol{x},t\,|\,\boldsymbol{x}',t') = 0 . \tag{4.3}$$

The corresponding backward Fokker-Planck equation is given by

$$-\frac{\partial P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')}{\partial t'} = +A_i(\boldsymbol{x}',t')\,\frac{\partial P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')}{\partial x_i'} + \tfrac{1}{2}B_{ij}(\boldsymbol{x}',t')\,\frac{\partial^2 P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')}{\partial x_i'\,\partial x_j'} . \tag{4.4}$$

The initial conditions in both cases may be taken to be

$$P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t) = \delta(\boldsymbol{x}-\boldsymbol{x}') . \tag{4.5}$$

4.2.2   Surfaces and boundary conditions

Forward equation

Integrating Eqn. 4.3 over some region Ω, we have

$$\frac{d}{dt}\int\limits_{\Omega} d\boldsymbol{x}\,P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t') = -\int\limits_{\partial\Omega} d\Sigma\,\hat{\boldsymbol{n}}\cdot\boldsymbol{J}(\boldsymbol{x},t\,|\,\boldsymbol{x}',t') , \tag{4.6}$$

where $\hat{\boldsymbol{n}}$ is locally normal to the surface $\partial\Omega$. At surfaces we need to specify boundary conditions. Generally these fall into one of three types:

   (i) *Reflecting* surfaces satisfy $\hat{\boldsymbol{n}}\cdot\boldsymbol{J}(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')\big|_{\Sigma} = 0$ at the surface Σ.

   (ii) *Absorbing* surfaces satisfy $P(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')\big|_{\Sigma} = 0$.

(iii) *Continuity* at a surface entails

$$P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t')\big|_{\Sigma_+} = P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t')\big|_{\Sigma_-} \quad , \quad \hat{\boldsymbol{n}}\cdot\boldsymbol{J}(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t')\big|_{\Sigma_+} = \hat{\boldsymbol{n}}\cdot\boldsymbol{J}(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t')\big|_{\Sigma_-} \ . \tag{4.7}$$

These conditions may be enforced even if the functions $A_i(\boldsymbol{x}, t)$ and $B_{ij}(\boldsymbol{x}, t)$ may be discontinuous across Σ.

Backward equation

For the backward FPE, we have the following[1]:

(i) *Reflecting* surfaces satisfy $n_i(\boldsymbol{x}') \, B_{ij}(\boldsymbol{x}') \, \frac{\partial}{\partial x'_j} P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t')\big|_{\Sigma} = 0$ for $\boldsymbol{x}' \in \Sigma$.

(ii) *Absorbing* surfaces satisfy $P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t')\big|_{\Sigma} = 0$.

4.2.3 One-dimensional Fokker-Planck equation

Consider the Fokker-Planck equation in $d = 1$. On an infinite interval $x \in (-\infty, +\infty)$, normalization requires $P(\pm\infty, t) = 0$, which generally[2] implies $\partial_x P(\pm\infty, t) = 0$. On a finite interval $x \in [a, b]$, we may impose periodic boundary conditions $P(a) = P(b)$ and $J(a) = J(b)$.

Recall that the Fokker-Planck equation follows from the stochastic differential equation

$$dx = f(x, t)\, dt + g(x, t)\, dW(t) \ , \tag{4.8}$$

with $f(x, t) = A(x, t)$ and $g(x, t) = \sqrt{B(x, t)}$, and where $W(t)$ is a Wiener process. In general[3], a solution to the above Itô SDE exists and is unique provided the quantities f and g satisfy a *Lipschitz condition*, which says that there exists a $K > 0$ such that $\big|f(x, t) - f(y, t)\big| + \big|g(x, t) - g(y, t)\big| < K|x - y|$ for $x, y \in [a, b]$[4]. Coupled with this is a *growth condition* which says that there exists an $L > 0$ such that $f^2(x, t) + g^2(x, t) < L(1 + x^2)$ for $x \in [a, b]$. If these two conditions are satisfied for $t \in [0, T]$, then there is a unique solution on this time interval.

Now suppose $B(a, t) = 0$, so there is no diffusion at the left endpoint. The left boundary is then said to be *prescribed*. From the Lipschitz condition on $\sqrt{B}$, this says that $B(x, t)$ vanishes no slower than $(x - a)^2$, which says that $\partial_x B(a, t) = 0$. Consider the above SDE with the condition $B(a, t) = 0$. We see that

(i) If $A(a, t) > 0$, a particle at a will enter the region $[a, b]$ with probability one. This is called an *entrance boundary*.

(ii) If $A(a, t) < 0$, a particle at a will exit the region $[a, b]$ with probability one. This is called an *exit boundary*.

---

[1]See Gardiner, §5.1.2.
[2]*I.e.* for well-behaved functions which you would take home to meet your mother.
[3]See L. Arnold, *Stochastic Differential Equations* (Dover, 2012).
[4]One can choose convenient dimensionless units for all quantities.

(iii) If $A(a, t) = 0$, a particle at a remain fixed with probability one. This is called a *natural boundary*.

Mutatis mutandis, similar considerations hold at $x = b$, where $A(b, t) > 0$ for an exit and $A(b, t) < 0$ for an entrance.

Stationary solutions

We now look for stationary solutions $P(x, t) = P_{\text{eq}}(x)$. We assume $A(x, t) = A(x)$ and $B(x, t) = B(x)$. Then

$$J = A(x)\, P_{\text{eq}}(x) - \frac{1}{2} \frac{d}{dx} \big[B(x)\, P_{\text{eq}}(x) \big] = \text{constant} . \tag{4.9}$$

Define the function

$$\psi(x) = \exp \left\{ 2 \int_a^x dx'\, \frac{A(x')}{B(x')} \right\} , \tag{4.10}$$

so $\psi'(x) = 2\psi(x)A(x)/B(x)$. Then

$$\frac{d}{dx} \left(\frac{B(x)\, P_{\text{eq}}(x)}{\psi(x)} \right) = -\frac{2J}{\psi(x)} , \tag{4.11}$$

with solution

$$P_{\text{eq}}(x) = \frac{B(a)}{B(x)} \cdot \frac{\psi(x)}{\psi(a)} \cdot P_{\text{eq}}(a) - \frac{2J\,\psi(x)}{B(x)} \int_a^x \frac{dx'}{\psi(x')} . \tag{4.12}$$

Note $\psi(a) = 1$. We now consider two different boundary conditions.

Zero current : In this case $J = 0$ and we have

$$P_{\text{eq}}(x) = \frac{B(a)}{B(x)} \cdot \frac{\psi(x)}{\psi(a)} \cdot P_{\text{eq}}(a) . \tag{4.13}$$

The unknown quantity $P(a)$ is then determined by normalization: $\int_a^b dx\, P_{\text{eq}}(x) = 1$.

Periodic boundary conditions : Here we invoke $P(a) = P(b)$, which requires a specific value for J,

$$J = \frac{P_{\text{eq}}(a)}{2} \left[\frac{B(a)}{\psi(a)} - \frac{B(b)}{\psi(b)} \right] \bigg/ \int_a^b \frac{dx'}{\psi(x')} . \tag{4.14}$$

This leaves one remaining unknown, $P_{\text{eq}}(a)$, which again is determined by normalization.

Examples

We conclude this section with two examples. The first is diffusion in a gravitational field, for which the Langevin equation takes the form

$$dx = -v_{\rm D} \, dt + \sqrt{2D} \, dW(t) \,, \tag{4.15}$$

where the drift velocity is $v_{\rm D} = g/\gamma$, with γ the frictional damping constant ($F_{\rm fr} = -\gamma M \dot{x}$) and g the acceleration due to gravity. Thus, the Fokker-Planck equation is $\partial_t P = v_{\rm D} \, \partial_x P + D \, \partial_x^2 P$, whence the solution with a reflecting $(J = 0)$ condition at $x = 0$ is

$$P_{\rm eq}(x) = \frac{D}{v_{\rm D}} \, \exp\!\left(-v_{\rm D} \, x/D\right) \,, \tag{4.16}$$

where we have normalized $P(x)$ on the interval $x \in [0, +\infty)$. This steady state distribution reflects the fact that particles tend to fall to the bottom. If we apply instead periodic boundary conditions at $x = 0$ and $x = L$, the solution is a constant $P(x) = P(0) = P(L)$. In this case the particles fall through the bottom $x = 0$ only to return at the top $x = L$ and keep falling, like in the game *Portal*[5].

Our second example is that of the Ornstein-Uhlenbeck process, described by $\partial_t P = \partial_x(\beta x P) + D \, \partial_x^2 P$. The steady state solution is

$$P_{\rm eq}(x) = P_{\rm eq}(0) \, \exp\!\left(-\beta x^2/2D\right) \,. \tag{4.17}$$

This is normalizable over the real line $x \in (-\infty, \infty)$. On a finite interval, we write

$$P_{\rm eq}(x) = P_{\rm eq}(a) \, e^{\beta(a^2 - x^2)/2D} \,. \tag{4.18}$$

4.2.4 Eigenfunction expansions for Fokker-Planck

We saw in §4.2.1 how the (forward) Fokker-Planck equation could be written as

$$\frac{\partial P(x,t)}{\partial t} = \mathcal{L} \, P(x,t) \quad , \quad \mathcal{L} = -\frac{\partial}{\partial x} \, A(x) + \frac{1}{2} \frac{\partial^2}{\partial x^2} \, B(x) \,, \tag{4.19}$$

and how the stationary state solution $P_{\rm eq}(x)$ satisfies $J = A \, P_{\rm eq} - \frac{1}{2}\partial_x(B \, P_{\rm eq})$. Consider the operator

$$\widetilde{\mathcal{L}} = +A(x) \, \frac{\partial}{\partial x} + \frac{1}{2} \, B(x) \, \frac{\partial^2}{\partial x^2} \,, \tag{4.20}$$

where, relative to $\mathcal{L}$, the sign of the leading term is reversed. It is straightforward to show that, for any functions f and g,

$$\left\langle f \,\middle|\, \widetilde{\mathcal{L}} \,\middle|\, g \right\rangle - \left\langle g \,\middle|\, \mathcal{L} \,\middle|\, f \right\rangle = \left[g \, J_f - f K_g \right]_a^b \,, \tag{4.21}$$

where

$$\left\langle g \,\middle|\, \mathcal{L} \,\middle|\, f \right\rangle = \int\limits_0^a dx \, g(x) \, \mathcal{L} \, f(x) \,, \tag{4.22}$$

---

[5]The cake is a lie.

and $J_f = Af - \frac{1}{2}(Bf)'$ and $K_g = -\frac{1}{2}Bg'$. Thus we conclude that $\widetilde{\mathcal{L}} = \mathcal{L}^\dagger$, the adjoint of $\mathcal{L}$, if either (i) J_f and K_g vanish at the boundaries $x = a$ and $x = b$ (reflecting conditions), or (ii) the functions f and g vanish at the boundaries (absorbing conditions).

We can use the zero current steady state distribution $P_{\text{eq}}(x)$, for which $J = AP_{\text{eq}} - \frac{1}{2}\partial_x(BP_{\text{eq}}) = 0$, to convert between solutions of the forward and backward time Fokker-Planck equations. Suppose $P(x,t)$ satisfies $\partial_t P = \mathcal{L}P$. Then define $Q(x,t) \equiv P(x,t)/P_{\text{eq}}(x)$, in which case

Define $P(x,t) = P_{\text{eq}}(x)\, Q(x,t)$. Then

$$
\begin{aligned}
\partial_t P = P_{\text{eq}}\, \partial_t Q &= -\partial_x(AP_{\text{eq}}Q) + \tfrac{1}{2}\partial_x^2(BP_{\text{eq}}Q) \\
&= \left\{ -\partial_x(AP_{\text{eq}}) + \tfrac{1}{2}\partial_x^2(BP_{\text{eq}}) \right\} Q + \left\{ -A\,\partial_x Q + \tfrac{1}{2}B\,\partial_x^2 Q \right\} P_{\text{eq}} + \partial_x(BP_{\text{eq}})\,\partial_x Q \quad (4.23) \\
&= \left\{ A\,\partial_x Q + \tfrac{1}{2}B\,\partial_x^2 Q \right\} P_{\text{eq}} \,,
\end{aligned}
$$

where we have used $\partial_x(BP_{\text{eq}}) = 2AP_{\text{eq}}$. Thus, we have that $Q(x,t)$ satisfies $\partial_t Q = \widetilde{\mathcal{L}}Q$. We saw in §4.2.1 how the (forward) Fokker-Planck equation could be written as

$$
\frac{\partial Q(x,t)}{\partial t} = \mathcal{L}^\dagger Q(x,t) \qquad , \qquad \mathcal{L}^\dagger = A(x)\,\frac{\partial}{\partial x} + \frac{1}{2}\,B(x)\,\frac{\partial^2}{\partial x^2} \,, \tag{4.24}
$$

which is the backward Fokker-Planck equation when written in terms of the time variable $s = -t$.

Now let us seek eigenfunctions $P_n(x)$ and $Q_n(x)$ which satisfy[6]

$$
\mathcal{L}P_n(x) = -\lambda_n P_n(x) \qquad , \qquad \mathcal{L}^\dagger Q_n(x) = -\lambda_n Q_n(x) \quad . \tag{4.25}
$$

where now $A(x,t) = A(x)$ and $B(x,t) = B(x)$ are assumed to be time-independent. If the functions $P_n(x)$ and $Q_n(x)$ form complete sets, then a solution to the Fokker-Planck equations for $P(x,t)$ and $Q(x,t)$ is of the form[7]

$$
P(x,t) = \sum_n C_n\, P_n(x)\, e^{-\lambda_n t} \qquad , \qquad Q(x,t) = \sum_n C_n\, Q_n(x)\, e^{-\lambda_n t} \quad . \tag{4.26}
$$

To elicit the linear algebraic structure here, we invoke Eqn. 4.25 and write

$$
(\lambda_m - \lambda_n)\, Q_m(x)\, P_n(x) = Q_m(x)\, \mathcal{L}\, P_n(x) - P_n(x)\, \mathcal{L}^\dagger\, Q_m(x) \,. \tag{4.27}
$$

Next we integrate over the interval $[a,b]$, which gives

$$
(\lambda_m - \lambda_n) \int_a^b dx\, Q_m(x)\, P_n(x) = \Big[Q_m(x)\, J_n(x) - K_m(x)\, P_n(x) \Big]_a^b = 0 \,, \tag{4.28}
$$

where $J_n(x) = A(x)\, P_n(x) - \frac{1}{2}\partial_x\big[B(x)\, P_n(x)\big]$ and $K_m(x) = -\frac{1}{2}B(x)\,\partial_x Q_m(x)$. For absorbing boundary conditions, the functions $P_n(x)$ and $Q_n(x)$ vanish at $x = a$ and $x = b$, so the RHS above vanishes. For

---

[6]In the eigensystem, the partial differential operators $\frac{\partial}{\partial x}$ in $\mathcal{L}$ and $\mathcal{L}^\dagger$ may be regarded as ordinary differential operators $\frac{d}{dx}$.

[7]Since $P_n(x) = P_{\text{eq}}(x)\, Q_n(x)$, the same expansion coefficients $\{C_n\}$ appear in both sums.

reflecting boundaries, it is the currents J_n and $K_m(x)$ which vanish at the boundaries. Thus $(\lambda_m - \lambda_n) \langle Q_m | P_n \rangle = 0$, where the inner product is

$$\langle Q | P \rangle \equiv \int\limits_a^b dx\, Q(x)\, P(x)\ . \tag{4.29}$$

Thus we obtain the familiar result from Sturm-Liouville theory that when the eigenvalues differ, the corresponding eigenfunctions are orthogonal. In the case of eigenvalue degeneracy, we can invoke the Gram-Schmidt procedure, in which case we may adopt the general normalization

$$\langle Q_m | P_n \rangle = \int\limits_a^b dx\, Q_m(x)\, P_n(x) = \int\limits_a^b dx\, P_{\mathsf{eq}}(x)\, Q_m(x)\, Q_n(x) = \int\limits_a^b dx\, \frac{P_m(x)\, P_n(x)}{P_{\mathsf{eq}}(x)} = \delta_{mn}\ . \tag{4.30}$$

A general solution to the Fokker-Planck equation with reflecting boundaries may now be written as

$$P(x,t) = \sum_n C_n\, P_n(x)\, e^{-\lambda_n t}\ , \tag{4.31}$$

where the expansion coefficients $\{C_n\}$ are given by

$$C_n = \int\limits_a^b dx\, Q_n(x)\, P(x,0) = \langle Q_n | P(0) \rangle\ . \tag{4.32}$$

Suppose our initial condition is $P(x,0\,|\,x_0,0) = \delta(x - x_0)$. Then $C_n = Q_n(x_0)$, and

$$P(x,t\,|\,x_0,0) = \sum_n Q_n(x_0)\, P_n(x)\, e^{-\lambda_n t}\ . \tag{4.33}$$

We may now take averages, such as

$$\langle F(x(t)) \rangle = \int\limits_a^b dx\, F(x) \sum_n Q_n(x_0)\, P_n(x)\, e^{-\lambda_n t}\ . \tag{4.34}$$

Furthermore, if we also average over $x_0 = x(0)$, assuming is is distributed according to $P_{\mathsf{eq}}(x_0)$, we have the correlator

$$\langle x(t)\, x(0) \rangle = \int\limits_a^b dx_0 \int\limits_a^b dx\, x\, x_0\, P(x,t\,|\,x_0,0)\, P_{\mathsf{eq}}(x_0)$$

$$= \sum_n \left[\int\limits_a^b dx\, x\, P_n(x) \right]^2 e^{-\lambda_n t} = \sum_n \left| \langle x | P_n \rangle \right|^2 e^{-\lambda_n t}\ . \tag{4.35}$$

Absorbing boundaries

At an *absorbing boundary* $x = a$, one has $P(a) = Q(a) = 0$. We may still use the function $P_{\text{eq}}(x)$ obtained from the $J = 0$ reflecting boundary conditions to convert between forward and backward Fokker-Planck equation solutions.

Next we consider some simple examples of the eigenfunction formalism.

Heat equation

We consider the simplest possible Fokker-Planck equation,

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial x^2} , \tag{4.36}$$

which is of course the one-dimensional diffusion equation. We choose our interval to be $x \in [0, L]$.

Reflecting boundaries : The normalized steady state solution is simply $P_{\text{eq}}(x) = 1/L$. The eigenfunctions are $P_0(x) = P_{\text{eq}}(x)$ and

$$P_n(x) = \frac{\sqrt{2}}{L} \cos\left(\frac{n\pi x}{L}\right) \quad , \qquad Q_n(x) = \sqrt{2} \cos\left(\frac{n\pi x}{L}\right) \tag{4.37}$$

for $n > 0$. The eigenvalues are $\lambda_n = D\,(n\pi/L)^2$. We then have

$$P(x,t \,|\, x_0, 0) = \frac{1}{L} + \frac{2}{L} \sum_{n=1}^{\infty} \cos\left(\frac{n\pi x_0}{L}\right) \cos\left(\frac{n\pi x}{L}\right) e^{-\lambda_n t} . \tag{4.38}$$

Note that as $t \to \infty$ one has $P(x, \infty \,|\, x_0, 0) = 1/L$, which says that $P(x,t)$ relaxes to $P_{\text{eq}}(x)$. Both boundaries are natural boundaries, which prevent probability flux from entering or leaking out of the region $[0, L]$.

Absorbing boundaries : Now we have

$$P_n(x) = \frac{\sqrt{2}}{L} \sin\left(\frac{n\pi x}{L}\right) \quad , \qquad Q_n(x) = \sqrt{2} \sin\left(\frac{n\pi x}{L}\right) \tag{4.39}$$

and

$$P(x,t \,|\, x_0, 0) = \frac{2}{L} \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x_0}{L}\right) \sin\left(\frac{n\pi x}{L}\right) e^{-\lambda_n t} , \tag{4.40}$$

again with $\lambda_n = D\,(n\pi/L)^2$. Since $\lambda_n > 0$ for all allowed n, we have $P(x, \infty \,|\, x_0, 0) = 0$, and all the probability leaks out by diffusion. The current is $J(x) = -DP'(x)$, which does not vanish at the boundaries.

Mixed boundaries : Now suppose $x = 0$ is an absorbing boundary and $x = L$ a reflecting boundary. Then

$$P_n(x) = \frac{\sqrt{2}}{L} \sin\left(\frac{(2n+1)\pi x}{2L}\right) \quad , \qquad Q_n(x) = \sqrt{2} \sin\left(\frac{(2n+1)\pi x}{2L}\right) \tag{4.41}$$

with $n \geq 0$. The eigenvalues are $\lambda_n = D\left(\left(n + \tfrac{1}{2}\right)\pi/L\right)^2$.

We can write the eigenfunctions in all three cases in the form $P_n(x) = \frac{\sqrt{2}}{L}\sin(k_n x + \delta)$, where $k_n = n\pi x/L$ or $(n + \tfrac{1}{2})\pi x/L$ and $\delta = 0$ or $\delta = \tfrac{1}{2}\pi$, with $\lambda_n = D\,k_n^2$. One then has

$$
\langle\, x\,|\,P_n\,\rangle = \begin{cases} \tfrac{1}{2}L & \text{reflecting},\, n = 0 \\ -\left(\sqrt{8}/Lk_n^2\right)\delta_{n,\text{odd}} & \text{reflecting},\, n > 0 \\ (-1)^{n+1}\sqrt{2}/k_n & \text{absorbing},\, n > 0 \\ (-1)^{n+1}\sqrt{2}/Lk_n^2 & \text{half reflecting, half absorbing},\, n > 0\,. \end{cases} \tag{4.42}
$$

Note that when a zero mode $\lambda_{\min} = 0$ is part of the spectrum, one has $P_0(x) = P_{\text{eq}}(x)$, to which $P(x,t)$ relaxes in the $t \to \infty$ limit. When one or both of the boundaries is absorbing, the lowest eigenvalue $\lambda_{\min} > 0$ is finite, hence $P(x, t \to \infty) \to 0$, *i.e.* all the probability eventually leaks out of the interval.

Ornstein-Uhlenbeck process

The Fokker-Planck equation for the OU process is $\partial_t P = \partial_x(\beta x P) + D\partial_x^2 P$. Over the real line $x \in \mathbb{R}$, the normalized steady state distribution is $P_{\text{eq}}(x) = (\beta/2\pi D)^{1/2}\exp(-\beta x^2/2D)$. The eigenvalue equation for $Q_n(x)$ is

$$
D\frac{d^2 Q_n}{dx^2} - \beta x\frac{dQ_n}{dx} = -\lambda_n\,Q_n(x)\,. \tag{4.43}
$$

Changing variables to $\xi = x/\ell$, where $\ell = (2D/\beta)^{1/2}$, we obtain $Q_n'' - 2\xi Q_n' + (2\lambda_n/\beta)\,Q_n = 0$, which is Hermite's equation. The eigenvalues are $\lambda_n = n\beta$, and the normalized eigenfunctions are then

$$
Q_n(x) = \frac{1}{\sqrt{2^n\,n!}}\,H_n\left(x/\ell\right)
$$
$$
P_n(x) = \frac{1}{\sqrt{2^n\,n!\,\pi\ell^2}}\,H_n\left(x/\ell\right)e^{-x^2/\ell^2}\,, \tag{4.44}
$$

which satisfy the orthonormality relation $\langle Q_m|P_n\rangle = \delta_{mn}$. Since $H_1(\xi) = 2\xi$, one has $\langle x|P_n\rangle = (\ell/\sqrt{2})\,\delta_{n,1}$, hence the correlator is given by $\langle x(t)\,x(0)\rangle = \tfrac{1}{2}\ell^2\,e^{-\beta t}$.

4.2.5 First passage problems

Suppose we have a particle on an interval $x \in [a, b]$ with absorbing boundary conditions, which means that particles are removed as soon as they get to $x = a$ or $x = b$ and not replaced. Following Gardiner[8], define the quantity

$$
G(x, t) = \int_a^b dx'\,P(x', t\,|\,x, 0)\,. \tag{4.45}
$$

---

[8]See Gardiner §5.5.

Thus, $G(x,t)$ is the probability that $x(t) \in [a,b]$ given that $x(0) = x$. Since the boundary conditions are absorbing, there is no reentrance into the region, which means that $G(x,t)$ is strictly decreasing as a function of time, and that

$$-\frac{\partial G(x,t)}{\partial t}\, dt = \text{probability, starting from } x \text{ at } t = 0, \text{ to exit } [a,b] \text{ during time interval } [t, t+dt]\,. \quad (4.46)$$

If we assume the process is autonomous, then

$$G(x,t) = \int\limits_{a}^{b} dx'\, P(x',0\,|\,x,-t)\,, \quad (4.47)$$

which satisfies the backward Fokker-Planck equation,

$$\frac{\partial G}{\partial t} = A\,\frac{\partial G}{\partial x} + \tfrac{1}{2} B\,\frac{\partial^2 G}{\partial x^2} = \mathcal{L}^{\dagger} G\,. \quad (4.48)$$

We may average functions of the exit time t according to

$$\langle f(t) \rangle_x = \int\limits_{0}^{\infty} dt\, f(t) \left(-\frac{\partial G(x,t)}{\partial t} \right)\,. \quad (4.49)$$

In particular, the *mean exit time $T(x)$* is given by

$$T(x) = \langle t \rangle_x = \int\limits_{0}^{\infty} dt\, t \left(-\frac{\partial G(x,t)}{\partial t} \right) = \int\limits_{0}^{\infty} dt\, G(x,t)\,. \quad (4.50)$$

From the Fokker-Planck equation for $G(x,t)$, the mean exit time $T(x)$ satisfies the ODE

$$\frac{1}{2} B(x)\,\frac{d^2 T}{dx^2} + A(x)\,\frac{dT}{dx} = -1\,. \quad (4.51)$$

This is derived by applying the operator $\mathcal{L}^{\dagger} = \tfrac{1}{2} B(x)\frac{\partial^2}{\partial x^2} + A(x)\frac{\partial}{\partial x}$ to the above expression for $T(x)$. Acting on the integrand $G(x,t)$, this produces $\frac{\partial G}{\partial t}$, according to Eq. 4.48, hence $\int\limits_{0}^{\infty} dt\,\partial_t G(x,t) = G(x,\infty) - G(x,0) = -1$.

To solve Eqn. 4.51, we once again invoke the services of the function

$$\psi_1(x) = \exp\left\{ \int\limits_{a}^{x} dx'\, \frac{2A(x')}{B(x')} \right\}\,, \quad (4.52)$$

which satisfies $\psi_1'(x)/\psi_1(x) = 2A(x)/B(x)$. Thus, we may reexpress eqn. 4.51 as

$$T'' + \frac{\psi_1'}{\psi_1}\, T' = -\frac{2}{B} \quad\Rightarrow\quad \left(\psi_1\, T' \right)' = -\frac{2\,\psi_1}{B}\,. \quad (4.53)$$

We may integrate this to obtain

$$T'(x) = \frac{T'(a)}{\psi_1(x)} - \frac{\psi_2(x)}{\psi_1(x)} \ , \tag{4.54}$$

where we have defined

$$\psi_2(x) = 2\int\limits_a^x dx' \ \frac{\psi_1(x')}{B(x')} \ . \tag{4.55}$$

Note that $\psi_1(a) = 1$ and $\psi_2(a) = 0$. We now integrate one last time to obtain

$$T(x) = T(a) + T'(a)\,\psi_3(x) - \psi_4(x) \ , \tag{4.56}$$

where

$$\psi_3(x) = \int\limits_a^x \frac{dx'}{\psi_1(x')} \quad , \quad \psi_4(x) = \int\limits_a^x dx' \ \frac{\psi_2(x')}{\psi_1(x')} \ . \tag{4.57}$$

Note that $\psi_3(a) = \psi_4(a) = 0$

Eqn. 4.56 involves two constants of integration, $T(a)$ and $T'(a)$, which are to be determined by imposing two boundary conditions. For an absorbing boundary at a, we have $T(a) = 0$. To determine the second unknown $T'(a)$, we impose the condition $T(b) = 0$, which yields $T'(a) = \psi_4(b)/\psi_3(b)$. The final result for the mean exit time is then

$$T(x) = \frac{\psi_3(x)\,\psi_4(b) - \psi_3(b)\,\psi_4(x)}{\psi_3(b)} \ . \tag{4.58}$$

As an example, consider the case of pure diffusion: $A(x) = 0$ and $B(x) = 2D$. Then

$$\psi_1(x) = 1 \quad , \quad \psi_2(x) = (x-a)/D \quad , \quad \psi_3(x) = (x-a) \quad , \quad \psi_4(x) = (x-a)^2/2D \ , \tag{4.59}$$

whence

$$T(x) = \frac{(x-a)(b-x)}{2D} \ . \tag{4.60}$$

A particle starting in the middle $x = \frac{1}{2}(a+b)$ at time $t = 0$ will then exit the region in an average time $(b-a)^2/8D$.

One absorbing, one reflecting boundary

Suppose the boundary at a is now reflecting, while that at b remains absorbing. We then have the boundary conditions $\partial_x G(a,t) = 0$ and $G(b,t) = 0$, which entails $T'(a) = 0$ and $T(b) = 0$. Then the general result of Eqn. 4.56 then gives $T(x) = T(a) - \psi_4(x)$. Requiring $T(b) = 0$ then yields the result

$$T(x) = T(b) - \psi_4(x) = 2\int\limits_x^b \frac{dy}{\psi_1(y)} \int\limits_a^y dz \ \frac{\psi_1(z)}{B(z)} \qquad (x = a \text{ reflecting}, \ x = b \text{ absorbing}) \ . \tag{4.61}$$

Under the opposite condition, where the boundary at a is absorbing while that at b is reflecting, we have $T(a) = 0$ and $T'(b) = 0$. Eqn. 4.56 then gives $T(x) = T'(a)\,\psi_3(x) - \psi_4(x)$, and imposing $T'(b) = 0$ entails $T'(a) = \psi_2(b)$, hence

$$T(x) = \psi_2(b)\,\psi_3(x) - \psi_4(x) = 2\int\limits_a^x \frac{dy}{\psi_1(y)} \int\limits_y^b dz\,\frac{\psi_1(z)}{B(z)} \qquad (x = a \text{ absorbing}, \, x = b \text{ reflecting}) . \qquad (4.62)$$

Escape through either boundary

Define the quantities

$$G_a(x,t) = -\int\limits_t^\infty dt'\, J(a,t'\,|\,x,0) = \int\limits_t^\infty dt'\left\{ -A(a)\,P(a,t'\,|\,x,0) + \tfrac{1}{2}\partial_a\Big[B(a)\,P(a,t'\,|\,x,0)\Big]\right\}$$

$$\qquad (4.63)$$

$$G_b(x,t) = +\int\limits_t^\infty dt'\, J(b,t'\,|\,x,0) = \int\limits_t^\infty dt'\left\{ +A(b)\,P(b,t'\,|\,x,0) - \tfrac{1}{2}\partial_b\Big[B(b)\,P(b,t'\,|\,x,0)\Big]\right\} .$$

Since $-J(a,t\,|\,x,0)$ is the left-moving probability flux at $x = a$, $G_a(x,t)$ represents the probability that a particle starting at $x \in [a,b]$ exits at a sometime after a time t. The second expression for $G_b(x,t)$ yields the probability that a particle starting at x exits at b sometime after t. Note that

$$G_a(x,t) + G_b(x,t) = \int\limits_t^\infty dt' \int\limits_a^b dx'\, \partial_{x'}\left\{ A(x')\,P(x',t'\,|\,x,0) - \tfrac{1}{2}\partial_{x'}\Big[B(x')\,P(x',t\,|\,x,0)\Big]\right\}$$

$$\qquad (4.64)$$

$$= \int\limits_t^\infty dt' \int\limits_a^b dx'\,\Big[-\partial_{t'}\,P(x',t'\,|\,x,0)\Big] = \int\limits_a^b dx'\,P(x',t\,|\,x,0) = G(x,t) ,$$

which is the total probability starting from x to exit the region after t.

Since $P(a,t'\,|\,x,0)$ satisfies the backward Fokker-Planck equation, *i.e.* $\mathcal{L}^\dagger\,P(a,t'\,|\,x,0) = \partial_{t'}P(a,t'\,|\,x,0)$, we have

$$\mathcal{L}^\dagger\,G_a(x,t) = J(a,t\,|\,x,0) = +\partial_t G_a(x,t)$$

$$\mathcal{L}^\dagger\,G_b(x,t) = J(b,t\,|\,x,0) = -\partial_t G_b(x,t) . \qquad (4.65)$$

Now let us evaluate the above equations in the limit $t \to 0$. Since $P(x',0\,|\,x,0) = \delta(x - x')$, there can only be an infinitesimal particle current at any finite distance from the initial point x at an infinitesimal value of the elapsed time t. Therefore we have

$$\mathcal{L}^\dagger\,G_c(x,0) = \left\{ A(x)\,\frac{\partial}{\partial x} + \frac{1}{2}\,B(x)\,\frac{\partial^2}{\partial x^2}\right\} G_c(x,0) = 0 . \qquad (4.66)$$

Thus, $G_c(x,0)$ is the total probability for exit via $c \in \{a,b\}$ over all time, conditioned at starting at x at time 0. The boundary conditions here are

$$G_a(a,0) = 1 \quad , \quad G_a(b,0) = 0 \quad ; \quad G_b(b,0) = 1 \quad , \quad G_b(a,0) = 0 \quad , \tag{4.67}$$

which says that a particle starting at a is immediately removed with probability unity and therefore can never exit through b, and *vice versa*. Solving using the function $\psi_1(x) = \exp \int\limits_a^x dx\, 2A(x')/B(x')$, we have

$$
\begin{aligned}
G_a(x,0) &= \int\limits_x^b dy\, \psi_1(y) \Big/ \int\limits_a^b dz\, \psi_1(x) \\[2mm]
G_b(x,0) &= \int\limits_a^x dy\, \psi_1(y) \Big/ \int\limits_a^b dz\, \psi_1(x) \ .
\end{aligned}
\tag{4.68}
$$

Note $G_a(x,0) + G_b(x,0) = 1$, which says that eventually the particle exits via either a or b. We next define

$$T_c(x) = \int\limits_0^\infty dt\, \frac{G_c(x,t)}{G_c(x,0)} \ , \tag{4.69}$$

which is the mean exit time through c, given that the particle did exit through that boundary. This then satisfies

$$\mathcal{L}^\dagger \Big[G_c(x,0)\, T_c(x) \Big] = -G_c(x,0) \ . \tag{4.70}$$

For pure diffusion, $A(x) = 0$ and $B(x) = 2D$, and we found $\psi_1(x) = 1$. Therefore

$$G_a(x,0) = \frac{b-x}{b-a} \quad , \quad G_b(x,0) = \frac{x-a}{b-a} \quad . \tag{4.71}$$

We may then solve the equations

$$D \frac{d^2}{dx^2} \Big[G_c(x,0)\, T_c(x) \Big] = -G_c(x,0) \tag{4.72}$$

to obtain

$$T_a(x) = \frac{(x-a)(2b-x-a)}{6D} \quad , \quad T_b(x) = \frac{(b-x)(b+x-2a)}{6D} \quad . \tag{4.73}$$

Note that

$$G_a(x,0)\, T_a(x) + G_b(x,0)\, T_b(x) = \frac{(x-a)(b-x)}{2D} = T(x) \ , \tag{4.74}$$

which we found previously in Eqn. 4.60.

4.2.6   Escape from a metastable potential minimum

In the presence of a local potential $U(x)$, the local drift velocity is $-U'(x)/\gamma m$, where m is the particle's mass and γ its frictional damping ($F_{\mathrm{fr}} = -\gamma m \dot{x}$). An example potential $U(x)$ is depicted in Fig. 4.1. Gardiner in §5.5.3 begins with the equation

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left(\frac{U'(x)}{\gamma m}\,P\right) + D\,\frac{\partial^2 P}{\partial x^2} \quad , \tag{4.75}$$

which resembles a Fokker-Planck equation for $P(x,t)$ with drift $v_{\mathrm{D}}(x) = -U'(x)/\gamma m$. However, Eqn. 4.75 is not a Fokker-Planck equation but rather something called the *Smoluchowski equation*. Recall that the position $x(t)$ of a Brownian particle *does not execute a Markov process*. So where does Eqn. 4.75 come from, and under what conditions is it valid?

It is the two-component phase space vector $\varphi = (x, v)$ which executes a Markov process, and for whose conditional probability density we can derive a Fokker-Planck equation, and not the position x alone. The Brownian motion problem may be written as two coupled first order differential equations,

$$\begin{aligned} dx &= v\,dt \\ dv &= -\left[\frac{1}{m}\,U'(x) + \gamma v\right]dt + \sqrt{\Gamma}\,dW(t) \quad , \end{aligned} \tag{4.76}$$

where $\Gamma = 2\gamma k_{\mathrm{B}}T/m = 2\gamma^2 D$, and where $W(t)$ is a Wiener process. The first of these is an ODE and the second an SDE. Viewed as a multicomponent SDE, the Fokker-Planck equation for $P(x,v,t)$ is

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\big(vP\big) + \frac{\partial}{\partial v}\left[\left(\frac{U'(x)}{m} + \gamma v\right)P\right] + \frac{\gamma k_{\mathrm{B}}T}{m}\,\frac{\partial^2 P}{\partial v^2} \ . \tag{4.77}$$

Suppose though that the damping γ is large. Then we can *approximate* the second equation in 4.76 by assuming v rapidly relaxes, which is to say $dv \approx 0$. Then we have

$$v\,dt \approx -\frac{1}{\gamma m}\,U'(x)\,dt + \sqrt{2D}\,dW(t) \tag{4.78}$$

and replacing v in the first equation with this expression we obtain the SDE

$$dx = v_{\mathrm{D}}(x)\,dt + \sqrt{2D}\,dW(t) \quad , \tag{4.79}$$

which immediately yields the Smoluchowski equation 4.75. This procedure is tantamount to an adiabatic elimination of the fast variable. It is valid only in the limit of large damping $\gamma = 6\pi\eta a/m$, which is to say large fluid viscosity η.

Taking the Smoluchowski equation as our point of departure, the steady state distribution is then found to be

$$P_{\mathsf{eq}}(x) = C\,e^{-U(x)/k_{\mathrm{B}}T} \ , \tag{4.80}$$

where we invoke the result $D = k_{\mathrm{B}}T/\gamma m$ from §2.2.2. We now consider the first passage time $T(x\,|\,x_0)$ for a particle starting at $x = x_0$ escaping to a point $x \approx x^*$ in the vicinity of the local potential maximum. We
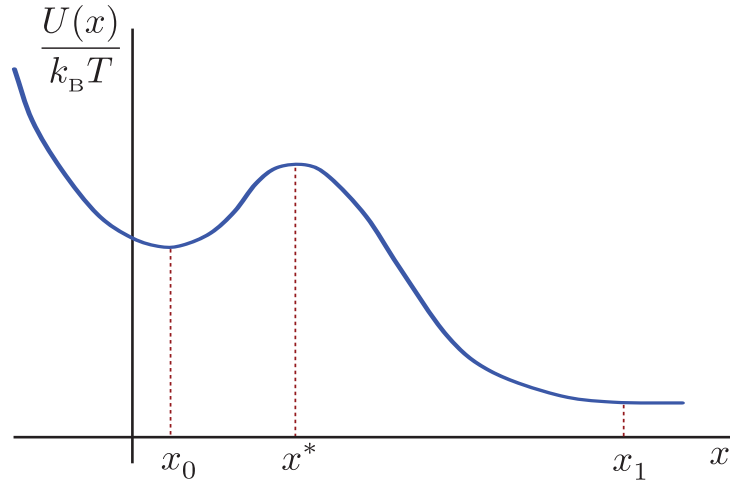
Figure 4.1: Escape from a metastable potential minimum.

apply the result of our previous analysis, with (a, b, x) in Eqn. 4.61 replaced by $(-\infty, x, x_0)$, respectively, and $x \gtrsim x^*$. Note that $A(x) = -U'(x)/\gamma m$, and $B(x) = 2D$, hence

$$\ln \psi_1(x) = \int\limits_a^x dx' \, \frac{2A(x')}{B(x')} = \frac{U(a) - U(x)}{k_{\mathrm{B}}T} \quad . \tag{4.81}$$

Formally we may have $U(a) = \infty$, but it drops out of the expression for the mean exit time,

$$T(x \,|\, x_0) = \frac{1}{D} \int\limits_{x_0}^x \frac{dy}{\psi_1(y)} \int\limits_{-\infty}^y dz \, \psi_1(z) = \frac{1}{D} \int\limits_{x_0}^x dy \, e^{U(y)/k_{\mathrm{B}}T} \int\limits_{-\infty}^y dz \, e^{-U(z)/k_{\mathrm{B}}T} . \tag{4.82}$$

The above integrals can be approximated as follows. Expand $U(x)$ about the local extrema at x_0 and x^* as

$$\begin{aligned}
U(x_0 + \delta x) &= U(x_0) + \tfrac{1}{2}K_0(\delta x)^2 + \dots \\
U(x^* + \delta x) &= U(x^*) - \tfrac{1}{2}K^*(\delta x)^2 + \dots \;,
\end{aligned} \tag{4.83}$$

where $K_0 = U''(x_0)$ and $K^* = -U''(x^*)$. At low temperatures, integrand $e^{-U(z)/k_{\mathrm{B}}T}$ is dominated by the region $z \approx x_0$, hence

$$\int\limits_{-\infty}^y dz \, e^{-U(z)/k_{\mathrm{B}}T} \approx \left(\frac{2\pi k_{\mathrm{B}}T}{K_0} \right)^{1/2} e^{-U(x_0)/k_{\mathrm{B}}T} . \tag{4.84}$$

Similarly, the integrand $e^{U(y)/k_{\mathrm{B}}T}$ is dominated by the region $y \approx x^*$, so for x somewhere between x^* and x_1, we may write[9]

$$\int\limits_{x_0}^x dy \, e^{U(y)/k_{\mathrm{B}}T} \approx \left(\frac{2\pi k_{\mathrm{B}}T}{K^*} \right)^{1/2} e^{U(x^*)/k_{\mathrm{B}}T} . \tag{4.85}$$

---

[9]We take $x > x^*$ to lie somewhere on the downslope of the potential curve, on the other side of the barrier from the metastable minimum.

We then have

$$T(x_1 \,|\, x_0) \approx \frac{2\pi k_{\mathrm{B}} T}{D\sqrt{K_0 K^*}} \, \exp\!\left(\frac{U(x^*) - U(x_0)}{k_{\mathrm{B}} T}\right) . \tag{4.86}$$

Known as the *Arrhenius law*, this is one of the most ubiquitous results in nonequilibrium statistical physics, with abundant consequences for chemistry, biology, and many other fields of science. With $\Delta E = U(x^*) - U(x_0)$, the energy necessary to surmount the barrier, the escape *rate* is seen to be proportional to $\exp(-\Delta E / k_{\mathrm{B}} T)$.

4.2.7 Detailed balance

Let φ denote a coordinate vector in phase space. In classical mechanics, $\varphi = (q, p)$ consists of all the generalized coordinates and generalized momenta. The condition of *detailed balance* says that each individual transition balances precisely with its time reverse, resulting in no net probability currents in equilibrium. Note that this is a much stronger condition than conservation of probability.

In terms of joint probability densities, detailed balance may be stated as follows:

$$P(\varphi, t \,;\, \varphi', t') = P(\varphi'^{\mathcal{T}}, -t' \,;\, \varphi^{\mathcal{T}}, -t) = P(\varphi'^{\mathcal{T}}, t \,;\, \varphi^{\mathcal{T}}, t') , \tag{4.87}$$

where we have assumed time translation invariance. Here, $\varphi^{\mathcal{T}}$ is the time reverse of φ. This is accomplished by multiplying each component φ_i by a quantity $\varepsilon_i = \pm 1$. For positions $\varepsilon = +1$, while for momenta $\varepsilon = -1$. If we define the diagonal matrix $\varepsilon_{ij} = \varepsilon_i \, \delta_{ij}$ (no sum on i), then $\varphi_i^{\mathcal{T}} = \varepsilon_{ij} \varphi_j$ (implied sum on j). Thus we may rewrite the above equation as

$$P(\varphi, t \,;\, \varphi', t') = P(\varepsilon \, \varphi', t \,;\, \varepsilon \, \varphi, t') . \tag{4.88}$$

In terms of the conditional probability distributions, we have

$$P(\varphi, t \,|\, \varphi', 0) \, P_{\mathrm{eq}}(\varphi') = P(\varepsilon \varphi', t \,|\, \varepsilon \varphi, 0) \, P_{\mathrm{eq}}(\varepsilon \varphi) , \tag{4.89}$$

where $P_{\mathrm{eq}}(\varphi)$ is the equilibrium distribution, which we assume holds at time $t' = 0$. Now in the limit $t \to 0$ we have $P(\varphi, t \to 0 \,|\, \varphi', 0) = \delta(\varphi - \varphi')$, and we therefore conclude

$$P_{\mathrm{eq}}(\varepsilon \varphi) = P_{\mathrm{eq}}(\varphi) . \tag{4.90}$$

The equilibrium distribution $P_{\mathrm{eq}}(\varphi)$ is time-reversal invariant. Thus, detailed balance entails

$$P(\varphi, t \,|\, \varphi', 0) \, P_{\mathrm{eq}}(\varphi') = P(\varepsilon \varphi', t \,|\, \varepsilon \varphi, 0) \, P_{\mathrm{eq}}(\varphi) . \tag{4.91}$$

One then has

$$\langle \varphi_i \rangle = \int d\varphi \, P_{\mathrm{eq}}(\varphi) \, \varphi_i = \varepsilon_i \, \langle \varphi_i \rangle$$

$$G_{ij}(t) \equiv \langle \varphi_i(t) \, \varphi_j(0) \rangle = \int d\varphi \int d\varphi' \, \varphi_i \, \varphi_j' \, P(\varphi, t \,|\, \varphi', 0) \, P_{\mathrm{eq}}(\varphi') = \varepsilon_i \, \varepsilon_i \, G_{ji}(t) . \tag{4.92}$$

Thus, as a matrix, $G(t) = \varepsilon \, G^{\mathrm{t}}(t) \, \varepsilon$.

The conditions under which detailed balance holds are[10]

$$W(\varphi \,|\, \varphi') \, P_{\text{eq}}(\varphi') = W(\varepsilon\varphi' \,|\, \varepsilon\varphi) \, P_{\text{eq}}(\varphi)$$

$$\Big[A_i(\varphi) + \varepsilon_i \, A_i(\varepsilon\varphi) \Big] P_{\text{eq}}(\varphi) = \frac{\partial}{\partial \varphi_j} \Big[B_{ij}(\varphi) \, P_{\text{eq}}(\varphi) \Big] \tag{4.93}$$

$$\varepsilon_i \varepsilon_j B_{ij}(\varepsilon\varphi) = B_{ij}(\varphi) \quad \text{(no sum on } i \text{ and } j) \,.$$

Detailed balance for the Fokker-Planck equation

It is useful to define the *reversible* and *irreversible drift* as

$$R_i(\varphi) \equiv \frac{1}{2} \Big[A_i(\varphi) + \varepsilon_i \, A_i(\varepsilon\varphi) \Big]$$

$$I_i(\varphi) \equiv \frac{1}{2} \Big[A_i(\varphi) - \varepsilon_i \, A_i(\varepsilon\varphi) \Big] \,. \tag{4.94}$$

Then we may subtract $\partial_i \big[\varepsilon_i A_i(\varepsilon\varphi) \, P_{\text{eq}}(\varphi) \big] - \frac{1}{2} \partial_i \partial_j \big[\varepsilon_i \varepsilon_j \, B_{ij}(\varepsilon\varphi) \, P_{\text{eq}}(\varphi) \big]$ from $\partial_i \big[A_i(\varphi) \, P_{\text{eq}}(\varphi) \big] - \frac{1}{2} \partial_i \partial_j \big[B_{ij}(\varphi) \, P_{\text{eq}}(\varphi) \big]$ to obtain

$$\sum_i \frac{\partial}{\partial \varphi_i} \Big[I_i(\varphi) \, P_{\text{eq}}(\varphi) \Big] = 0 \quad \Rightarrow \quad \sum_i \left\{ \frac{\partial I_i(\varphi)}{\partial \varphi_i} + I_i(\varphi) \, \frac{\partial \ln P_{\text{eq}}(\varphi)}{\partial \varphi_i} \right\} = 0 \,. \tag{4.95}$$

We may now write the second of Eqn. 4.93 as

$$R_i(\varphi) = \tfrac{1}{2} \partial_j \, B_{ij}(\varphi) + \tfrac{1}{2} B_{ij}(\varphi) \, \partial_j \ln P_{\text{eq}}(\varphi) \,, \tag{4.96}$$

or, assuming the matrix B is invertible,

$$\partial_k \ln P_{\text{eq}}(\varphi) = 2 \, B_{ki}^{-1} \big(R_i - \tfrac{1}{2} \partial_j B_{ij} \big) \equiv Z_k(\varphi) \,. \tag{4.97}$$

Since the LHS above is a gradient, the condition that $P_{\text{eq}}(\varphi)$ exists is tantamount to

$$\frac{\partial Z_i}{\partial \varphi_j} = \frac{\partial Z_j}{\partial \varphi_i} \tag{4.98}$$

for all i and j. If this is the case, then we have

$$P_{\text{eq}}(\varphi) = \exp \int\limits^{\varphi} d\varphi' \cdot \mathbf{Z}(\varphi') \,. \tag{4.99}$$

Because of the condition 4.98, the integral on the RHS may be taken along any path. The constant associated with the undetermined lower limit of integration is set by overall normalization.

---

[10]See Gardiner, §6.3.5.

Brownian motion in a local potential

Recall that the Brownian motion problem may be written as two coupled first order differential equations,

$$
\begin{aligned}
dx &= v\, dt \\
dv &= -\left[\frac{1}{m}\, U'(x) + \gamma v\right] dt + \sqrt{\Gamma}\, dW(t)\ ,
\end{aligned}
\tag{4.100}
$$

where $\Gamma = 2\gamma k_{\mathrm{B}} T/m = 2\gamma^2 D$, and where $W(t)$ is a Wiener process. The first of these is an ODE and the second an SDE. Viewed as a multicomponent SDE with

$$
\boldsymbol{\varphi} = \begin{pmatrix} x \\ v \end{pmatrix} \qquad,\qquad A_i(\boldsymbol{\varphi}) = \begin{pmatrix} v \\ -\frac{U'(x)}{m} - \gamma v \end{pmatrix} \qquad,\qquad B_{ij}(\boldsymbol{\varphi}) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{2\gamma k_{\mathrm{B}} T}{m} \end{pmatrix} \ .
\tag{4.101}
$$

We have already derived in Eqn. 4.77 the associated Fokker-Planck equation for $P(x,v,t)$.

The time reversal eigenvalues are $\varepsilon_1 = +1$ for x and $\varepsilon_2 = -1$ for v. We then have

$$
\boldsymbol{R}(\boldsymbol{\varphi}) = \begin{pmatrix} 0 \\ -\gamma v \end{pmatrix} \qquad,\qquad \boldsymbol{I}(\boldsymbol{\varphi}) = \begin{pmatrix} v \\ -\frac{U'(x)}{m} \end{pmatrix} \ .
\tag{4.102}
$$

As the B matrix is not invertible, we appeal to Eqn. 4.96. The upper component vanishes, and the lower component yields

$$
-\gamma v = \frac{\gamma k_{\mathrm{B}} T}{m}\, \frac{\partial \ln P_{\mathrm{eq}}}{\partial v}\ ,
\tag{4.103}
$$

which says $P_{\mathrm{eq}}(x,v) = F(x)\,\exp(-mv^2/2k_{\mathrm{B}}T)$. To find $F(x)$, we use Eqn. 4.95, which says

$$
\begin{aligned}
0 &= \overbrace{\frac{\partial I_1}{\partial x}}^{0} + \overbrace{\frac{\partial I_2}{\partial v}}^{0} + I_1\, \frac{\partial \ln P_{\mathrm{eq}}}{\partial x} + I_2\, \frac{\partial \ln P_{\mathrm{eq}}}{\partial v} \\[2mm]
&= v\, \frac{\partial \ln F}{\partial x} - \frac{U'(x)}{m}\left(-\frac{mv}{k_{\mathrm{B}} T}\right) \qquad \Rightarrow \qquad F(x) = C\, e^{-U(x)/k_{\mathrm{B}} T}\ .
\end{aligned}
\tag{4.104}
$$

Thus,

$$
P_{\mathrm{eq}}(x,v) = C\, e^{-mv^2/2k_{\mathrm{B}} T}\, e^{-U(x)/k_{\mathrm{B}} T}\ .
\tag{4.105}
$$

4.2.8 Multicomponent Ornstein-Uhlenbeck process

In §3.4.3 we considered the case of coupled SDEs,

$$
d\varphi_i = A_i(\boldsymbol{\varphi})\, dt + \beta_{ij}(\boldsymbol{\varphi})\, dW_j(t)\ ,
\tag{4.106}
$$

where $\langle W_i(t)\, W_j(t')\rangle = \delta_{ij}\,\min(t, t')$. We showed in §3.4.3 that such a multicomponent SDE leads to the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial \varphi_i}\left(A_i\,P\right) + \frac{1}{2}\,\frac{\partial^2}{\partial \varphi_i\,\partial \varphi_j}\left(B_{ij}\,P\right)\,, \tag{4.107}$$

where $B = \beta\beta^{\mathrm{t}}$, i.e. $B_{ij} = \sum_k \beta_{ik}\beta_{jk}$.

Now consider such a process with

$$A_i(\boldsymbol{\varphi}) = A_{ij}\,\varphi_j \qquad , \qquad B_{ij}(\boldsymbol{\varphi}) = B_{ij} \quad , \tag{4.108}$$

where A_{ij} and B_{ij} are independent of $\boldsymbol{\varphi}$. The detailed balance conditions are written as $\varepsilon B\varepsilon = B$, and

$$\left(A + \varepsilon A\,\varepsilon\right)\boldsymbol{\varphi} = B\,\boldsymbol{\nabla}\ln P_{\mathrm{eq}}(\boldsymbol{\varphi})\,. \tag{4.109}$$

This equation says that $P_{\mathrm{eq}}(\boldsymbol{\varphi})$ must be a Gaussian, which we write as

$$P_{\mathrm{eq}}(\boldsymbol{\varphi}) = P_{\mathrm{eq}}(0)\,\exp\left[-\tfrac{1}{2}\,\varphi_i\,M_{ij}^{-1}\,\varphi_j\right]\,, \tag{4.110}$$

Obviously we can take M^{-1} to be symmetric, since any antisymmetric part of M^{-1} is projected out in the expression $\varphi_i\,M_{ij}^{-1}\,\varphi_j$. Thus M is also symmetric. Substituting this solution into the stationary Fokker-Planck equation $\partial_i\big[A_{ij}\varphi_j P_{\mathrm{eq}}\big] = \tfrac{1}{2}\,\partial_i\partial_j\big(B_{ij}P_{\mathrm{eq}}\big)$ yields

$$\operatorname{Tr} A + \tfrac{1}{2}\operatorname{Tr}\left(BM^{-1}\right) = \varphi_i\left[M^{-1}A + \tfrac{1}{2}\,M^{-1}BM^{-1}\right]_{ij}\varphi_j = 0\,. \tag{4.111}$$

This must be satisfied for all $\boldsymbol{\varphi}$, hence both the LHS and RHS of this equation must vanish separately. This entails

$$A + MA^{\mathrm{t}}M^{-1} + BM^{-1} = 0\,. \tag{4.112}$$

We now invoke the detailed balance condition of Eqn. 4.109, which says

$$A + \varepsilon A\,\varepsilon + BM^{-1} = 0\,. \tag{4.113}$$

Combining this with our previous result, we conclude

$$\varepsilon AM = (AM)^{\mathrm{t}}\varepsilon\,, \tag{4.114}$$

which are known as the *Onsager conditions*. If we define the phenomenological force

$$\boldsymbol{F} = \boldsymbol{\nabla}\ln P_{\mathrm{eq}} = -M^{-1}\boldsymbol{\varphi}\,, \tag{4.115}$$

then we have

$$\frac{d\langle\boldsymbol{\varphi}\rangle}{dt} = A\,\langle\boldsymbol{\varphi}\rangle = -AM\boldsymbol{F}\,, \tag{4.116}$$

and defining $L = -AM$ which relates the *fluxes* $\boldsymbol{J} = \langle\dot{\boldsymbol{\varphi}}\rangle$ to the *forces* $\boldsymbol{F}$, viz. $J_i = L_{ik}\,F_k$, we have the celebrated *Onsager relations*, $\varepsilon L\varepsilon = L^{\mathrm{t}}$. A more general formulation, allowing for the presence of a magnetic field, is

$$L_{ik}(\boldsymbol{B}) = \varepsilon_i\,\varepsilon_k\,L_{ki}(-\boldsymbol{B})\,. \tag{4.117}$$

We shall meet up with the Onsager relations again when we study the Boltzmann equation.
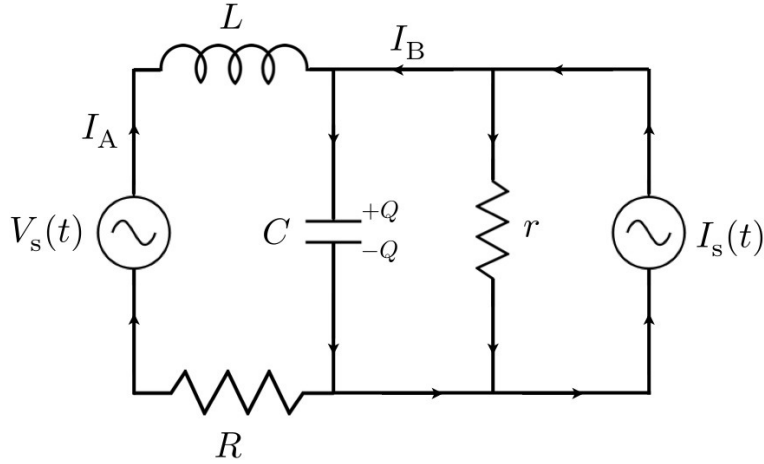
Figure 4.2: Electrical circuit containing a fluctuating voltage source $V_{\mathrm{s}}(t)$ and a fluctuating current source $I_{\mathrm{s}}(t)$.

4.2.9   Nyquist's theorem

Consider the electrical circuit in Fig. 4.2. Kirchoff's laws say that the current flowing through the resistor r is $I_{\mathrm{S}} - I_{\mathrm{B}}$, and that

$$(I_{\mathrm{S}} - I_{\mathrm{B}})\, r \;=\; \frac{Q}{C} = V_{\mathrm{S}} - L\,\frac{dI_{\mathrm{A}}}{dt} - RI_{\mathrm{A}} \tag{4.118}$$

and

$$\frac{dQ}{dt} = I_{\mathrm{A}} + I_{\mathrm{B}} \,. \tag{4.119}$$

Thus, we have the coupled ODEs for Q and I_{A},

$$\frac{dQ}{dt} = I_{\mathrm{A}} - \frac{Q}{rC} + I_{\mathrm{S}}(t)$$
$$\frac{dI_{\mathrm{A}}}{dt} = -\frac{R\,I_{\mathrm{A}}}{L} - \frac{Q}{LC} + \frac{V_{\mathrm{S}}(t)}{L} \,. \tag{4.120}$$

If we assume $V_{\mathrm{S}}(t)$ and $I_{\mathrm{S}}(t)$ are fluctuating sources each described by a Wiener process, we may write

$$V_{\mathrm{S}}(t)\, dt = \sqrt{\Gamma_V}\, dW_V(t) \qquad , \qquad I_{\mathrm{S}}(t)\, dt = \sqrt{\Gamma_I}\, dW_I(t) \,. \tag{4.121}$$

Then

$$dQ = \left(-\frac{Q}{rC} + I_{\mathrm{A}}\right) dt + \sqrt{\Gamma_I}\, dW_I(t)$$
$$dI_{\mathrm{A}} = -\left(\frac{Q}{LC} + \frac{R\,I_{\mathrm{A}}}{L}\right) dt + \frac{1}{L}\,\sqrt{\Gamma_V}\, dW_V(t) \,. \tag{4.122}$$

We now see that Eqn. 4.122 describes a two component Ornstein-Uhlenbeck process, with $\varphi^{\mathrm{t}} = (Q, I_{\mathrm{A}})$, and

$$A_{ij} = -\begin{pmatrix} 1/rC & -1 \\ 1/LC & R/L \end{pmatrix} \quad , \qquad B_{ij} = \begin{pmatrix} \Gamma_I & 0 \\ 0 & \Gamma_V/L^2 \end{pmatrix} \quad . \tag{4.123}$$

The ε matrix for this problem is $\varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ since charge is even and current odd under time reversal. Thus,

$$A + \varepsilon A \varepsilon = -\begin{pmatrix} 2/rC & 0 \\ 0 & 2R/L \end{pmatrix} = -BM^{-1} \,, \tag{4.124}$$

from which we may obtain M^{-1} and then

$$M = \begin{pmatrix} \Gamma_I \, rC/2 & 0 \\ 0 & \Gamma_V/2LR \end{pmatrix} \quad . \tag{4.125}$$

The equilibrium distribution is then

$$P_{\mathrm{eq}}(Q, I_{\mathrm{A}}) = \mathcal{N} \, \exp\left\{ -\frac{Q^2}{rC\Gamma_I} - \frac{RL\, I_{\mathrm{A}}^2}{\Gamma_V} \right\} \,. \tag{4.126}$$

We now demand that equipartition hold, *i.e.*

$$\left\langle \frac{Q^2}{2C} \right\rangle = \left\langle \frac{LI_{\mathrm{A}}^2}{2} \right\rangle = \tfrac{1}{2}k_{\mathrm{B}}T \,, \tag{4.127}$$

which fixes

$$\Gamma_V = 2Rk_{\mathrm{B}}T \quad , \qquad \Gamma_I = 2k_{\mathrm{B}}T/r \quad . \tag{4.128}$$

Therefore, the current and voltage fluctuations are given by

$$\left\langle V_{\mathrm{S}}(0)\, V_{\mathrm{S}}(t) \right\rangle = 2k_{\mathrm{B}}TR\,\delta(t) \quad , \qquad \left\langle I_{\mathrm{S}}(0)\, I_{\mathrm{S}}(t) \right\rangle = \frac{2k_{\mathrm{B}}T}{r}\,\delta(t) \quad , \qquad \left\langle V_{\mathrm{S}}(0)\, I_{\mathrm{S}}(t) \right\rangle = 0 \quad . \tag{4.129}$$

4.3 Master Equation

In §2.6.3 we showed that the differential Chapman-Kolmogorov equation with only jump processes yielded the Master equation,

$$\frac{\partial P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t')}{\partial t} = \int d\boldsymbol{y} \left[W(\boldsymbol{x} \,|\, \boldsymbol{y}, t)\, P(\boldsymbol{y}, t \,|\, \boldsymbol{x}', t') - W(\boldsymbol{y} \,|\, \boldsymbol{x}, t)\, P(\boldsymbol{x}, t \,|\, \boldsymbol{x}', t') \right] . \tag{4.130}$$

Here $W(\boldsymbol{x} \,|\, \boldsymbol{y}, t)$ is the rate density of transitions from $\boldsymbol{y}$ to $\boldsymbol{x}$ at time t, and has dimensions $T^{-1}L^{-d}$. On a discrete state space, we have

$$\frac{\partial P(\boldsymbol{n}, t \,|\, \boldsymbol{n}', t')}{\partial t} = \sum_{\boldsymbol{m}} \left[W(\boldsymbol{n} \,|\, \boldsymbol{m}, t)\, P(\boldsymbol{m}, t \,|\, \boldsymbol{n}', t') - W(\boldsymbol{m} \,|\, \boldsymbol{n}, t)\, P(\boldsymbol{n}, t \,|\, \boldsymbol{n}', t') \right] , \tag{4.131}$$

where $W(\boldsymbol{n} \,|\, \boldsymbol{m}, t)$ is the rate of transitions from $\boldsymbol{m}$ to $\boldsymbol{n}$ at time t, with dimensions T^{-1}.

4.3.1 Birth-death processes

The simplest case is that of one variable n, which represents the number of individuals in a population. Thus $n \geq 0$ and $P(n, t \,|\, n', t') = 0$ if $n < 0$ or $n' < 0$. If we assume that births and deaths happen individually and at with a time-independent rate, then we may write

$$W(n \,|\, m, t) = t^+(m)\, \delta_{n, m+1} + t^-(m)\, \delta_{n, m-1} \,. \tag{4.132}$$

Here $t^+(m)$ is the rate for $m \to m + 1$, and $t^-(m)$ is the rate for $m \to m - 1$. We require $t^-(0) = 0$, since the dying rate for an entirely dead population must be zero[11]. We then have the Master equation

$$\frac{\partial P(n, t \,|\, n_0, t_0)}{\partial t} = t^+(n-1)\, P(n-1, t \,|\, n_0, t_0) + t^-(n+1)\, P(n+1, t \,|\, n_0, t_0) - \left[t^+(n) + t^-(n) \right] P(n, t \,|\, n_0, t_0) \,. \tag{4.133}$$

This may be written in the form

$$\frac{\partial P(n, t \,|\, n_0, t_0)}{\partial t} + \Delta J(n, t \,|\, n_0, t_0) = 0 \,, \tag{4.134}$$

where the lattice current operator on the link $(n, n+1)$ is

$$J(n, t \,|\, n_0, t_0) = t^+(n)\, P(n, t \,|\, n_0, t_0) - t^-(n+1)\, P(n+1, t \,|\, n_0, t_0) \,. \tag{4.135}$$

The lattice derivative Δ is defined by

$$\Delta f(n) = f(n) - f(n-1) \,, \tag{4.136}$$

for any lattice function $f(n)$. One then has

$$\frac{d\langle n \rangle_t}{dt} = \sum_{n=0}^{\infty} \left[t^+(n) - t^-(n) \right] P(n, t \,|\, n_0, t_0) = \left\langle t^+(n) \right\rangle_t - \left\langle t^-(n) \right\rangle \,. \tag{4.137}$$

Steady state solution

We now seek a steady state solution $P_{\text{eq}}(n)$, as we did in the case of the Fokker-Planck equation. This entails $\Delta_n J(n) = 0$, where we suppress the initial conditions (n_0, t_0). Now $J(-1) = 0$ because $t^-(0) = 0$ and $P(-1) = 0$, hence $0 = J(0) - J(-1)$ entails $J(0) = 0$, and since $0 = \Delta_n J(n)$ we have $J(n) = 0$ for all $n \geq 0$. Therefore

$$P_{\text{eq}}(j + 1) = \frac{t^+(j)}{t^-(j+1)}\, P_{\text{eq}}(j) \quad, \tag{4.138}$$

which means

$$P_{\text{eq}}(n) = P_{\text{eq}}(0) \prod_{j=1}^{n} \frac{t^+(j-1)}{t^-(j)} \,. \tag{4.139}$$

---

[11]We neglect here the important possibility of zombies.

4.3.2 Examples: reaction kinetics

First example

Consider the example in Gardiner §11.1.2, which is the reaction

$$X \underset{k_1}{\overset{k_2}{\rightleftharpoons}} A \, . \tag{4.140}$$

We assume the concentration $[A] = a$ is fixed, and denote the number of X reactants to be n. The rates are $t^-(n) = k_2 n$ and $t^+(n) = k_1 a$, hence we have the Master equation

$$\partial_t P(n,t) = k_2(n+1) \, P(n+1,t) + k_1 a \, P(n-1,t) - \big(k_2 n + k_1 a\big) \, P(n,t) \, , \tag{4.141}$$

with $P(-1,t) \equiv 0$. We solve this using the generating function formalism, defining

$$\tilde{P}(z,t) = \sum_{n=0}^{\infty} z^n \, P(n,t) \, . \tag{4.142}$$

Note that $\tilde{P}(1,t) = \sum_{n=0}^{\infty} P(n,t) = 1$ by normalization. Multiplying both sides of Eqn. 4.141 by z^n and then summing from $n = 0$ to $n = \infty$, we obtain

$$\partial_t \tilde{P}(z,t) = k_1 a \overbrace{\sum_{n=0}^{\infty} P(n-1,t) \, z^n}^{zP(z,t)} - k_1 a \overbrace{\sum_{n=0}^{\infty} P(n,t) \, z^n}^{P(z,t)} + k_2 \overbrace{\sum_{n=0}^{\infty} (n+1) \, P(n+1,t) \, z^n}^{\partial_z P(z,t)} - k_2 \overbrace{\sum_{n=0}^{\infty} n \, P(n,t) \, z^n}^{z\partial_z P(z,t)}$$

$$= (z-1) \left\{ k_1 a \, \tilde{P}(z,t) - k_2 \, \partial_z \tilde{P}(z,t) \right\} \, . \tag{4.143}$$

We now define the function $\tilde{Q}(z,t)$ via

$$\tilde{P}(z,t) = e^{k_1 a z / k_2} \, \tilde{Q}(z,t) \, , \tag{4.144}$$

so that

$$\partial_t \tilde{Q} + k_2 (z-1) \, \partial_z \tilde{Q} = 0 \, , \tag{4.145}$$

and defining $w = -\ln(1-z)$, this is recast as $\partial_t \tilde{Q} - k_2 \partial_w \tilde{Q} = 0$, whose solution is

$$\tilde{Q}(z,t) = F(w + k_2 t) \, , \tag{4.146}$$

where F is an arbitrary function of its argument. To determine the function $F(w)$, we invoke our initial conditions,

$$\tilde{Q}(z,0) = e^{-k_1 a z / k_2} \, \tilde{P}(z,0) = F(w) \, . \tag{4.147}$$

We then have

$$F(w) = \exp\left\{ -\frac{k_1 a}{k_2}(1 - e^{-w}) \right\} \tilde{P}\big(1 - e^{-w}, 0\big) \, , \tag{4.148}$$

and hence

$$\tilde{P}(z,t) = \exp\left\{-\frac{k_1 a}{k_2}(1-z)(1-e^{-k_2 t})\right\} \tilde{P}\big(1 - (1-z)\, e^{-k_2 t}, 0\big) \,. \tag{4.149}$$

We may then obtain $P(n,t)$ via contour integration, *i.e.* by extracting the coefficient of z^n in the above expression:

$$P(n,t) = \frac{1}{2\pi i} \oint_{|z|=1} \frac{dz}{z^{n+1}}\, \tilde{P}(z,t) \,. \tag{4.150}$$

Note that setting $t = 0$ in Eqn. 4.149 yields the identity $\tilde{P}(z,0) = \tilde{P}(z,0)$. As $t \to \infty$, we have the steady state result

$$\tilde{P}(z,\infty) = e^{k_1 a(z-1)/k_2} \qquad \Rightarrow \qquad P(n,\infty) = \frac{\lambda^n}{n!}\, e^{-\lambda} \,, \tag{4.151}$$

where $\lambda = k_1 a/k_2$, which is a Poisson distribution. Indeed, suppose we start at $t = 0$ with the Poisson distribution $P(n,0) = e^{-\alpha_0}\alpha_0^n/n!$. Then $\tilde{P}(z,0) = \exp\big[\alpha_0(z-1)\big]$, and Eqn. 4.149 gives

$$\tilde{P}(z,t) = \exp\left\{-\frac{k_1 a}{k_2}(1-z)(1-e^{-k_2 t})\right\} \exp\left\{-\alpha_0(1-z)\, e^{-k_2 t}\right\} = e^{\alpha(t)\,(z-1)} \,, \tag{4.152}$$

where

$$\alpha(t) = \alpha_0\, e^{-k_2 t} + \frac{k_1}{k_2}\, a\left(1 - e^{-k_2 t}\right) \,. \tag{4.153}$$

Thus, $\alpha(0) = \alpha_0$ and $\alpha(\infty) = k_1 a/k_2 = \lambda$. The distribution is Poisson all along, with a time evolving Poisson parameter $\alpha(t)$. The situation is somewhat reminiscent of the case of updating conjugate Bayesian priors, where the prior distribution was matched with the likelihood function so that the updated prior retains the same functional form.

If we start instead with $P(n,0) = \delta_{n,n_0}$, then we have $\tilde{P}(z,0) = z^{n_0}$, and

$$\tilde{P}(z,t) = \exp\left\{-\frac{k_1 a}{k_2}(1-z)(1-e^{-k_2 t})\right\} \left(1 - (1-z)\, e^{-k_2 t}\right)^{n_0} \,. \tag{4.154}$$

We then have

$$\langle n(t)\rangle = \left.\frac{\partial \tilde{P}(z,t)}{\partial z}\right|_{z=1} = \frac{k_1 a}{k_2}\left(1 - e^{-k_2 t}\right) + n_0\, e^{-k_2 t}$$

$$\langle n^2(t)\rangle = \left(\frac{\partial^2 \tilde{P}(z,t)}{\partial z^2} + \frac{\partial \tilde{P}(z,t)}{\partial z}\right)_{z=1} = \langle n(t)\rangle^2 + \langle n(t)\rangle - n_0\, e^{-2k_2 t} \tag{4.155}$$

$$\mathrm{Var}\big[n(t)\big] = \left(\frac{k_1 a}{k_2} + n_0\, e^{-k_2 t}\right)\left(1 - e^{-k_2 t}\right) \,.$$

Second example

Gardiner next considers the reactions

$$X \underset{k_1}{\overset{k_2}{\rightleftharpoons}} A \qquad , \qquad B + 2X \underset{k_4}{\overset{k_3}{\rightleftharpoons}} 3X \quad , \tag{4.156}$$
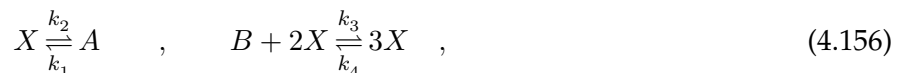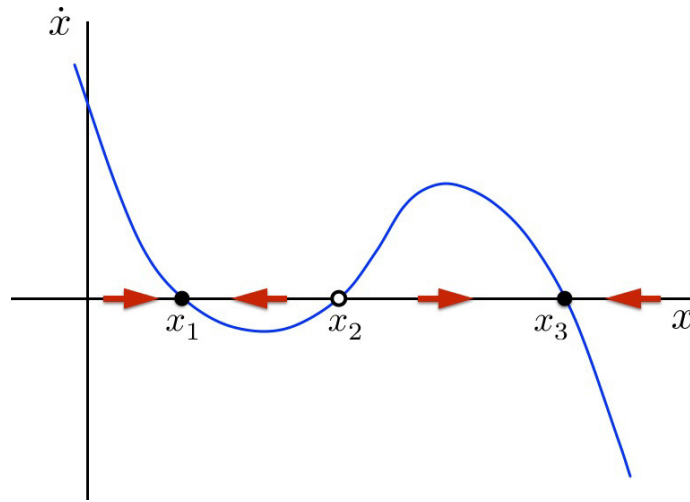
Figure 4.3: Geometric interpretation of the ODE in Eqn. 4.160.

for which we have

$$
\begin{aligned}
t^+(n) &= k_1 a + k_3 b \, n(n-1) \\
t^-(n) &= k_2 n + k_4 n(n-1)(n-2) \ .
\end{aligned}
\tag{4.157}
$$

The reason here is that for the second equation to proceed to the left, we need to select three X molecules to take part in the reaction, and there are $n(n-1)(n-2)$ ordered triples (i,j,k). Now Eqn. 4.137 gives

$$
\frac{d\langle n \rangle}{dt} = k_1 a + k_3 \langle n(n-1) \rangle - k_2 \langle n \rangle - k_4 \langle n(n-1)(n-2) \rangle \ .
\tag{4.158}
$$

For a Poisson distribution $P_n = e^{-\lambda} \lambda^n / n!$, it is easy to see that

$$
\langle n(n-1) \cdots (n-k+1) \rangle = \langle n \rangle^k \quad \text{(Poisson)} \ .
\tag{4.159}
$$

Suppose the distribution $P(n,t)$ is Poissonian for all t. This is not necessarily the case, but we assume it to be so for the purposes of approximation. Then the above equation closes, and with $x = \langle n \rangle$, we have

$$
\begin{aligned}
\frac{dx}{dt} &= -k_4 \, x^3 + k_3 \, x^2 - k_2 \, x + k_1 \, a \\
&= -k_4 (x - x_1)(x - x_2)(x - x_3) \ ,
\end{aligned}
\tag{4.160}
$$

where $x_{1,2,3}$ are the three roots of the cubic on the RHS of the top equation. Since the coefficients of this equation are real numbers, the roots are either real or come in complex conjugate pairs. We know that the product of the roots is $x_1 x_2 x_3 = k_1 a / k_4$ and that the sum is $x_1 + x_2 + x_3 = k_3 / k_4$, both of which are positive. Clearly when x is real and negative, all terms in the cubic are of the same sign, hence there can be no real roots with $x < 0$. We assume three real positive roots with $x_1 < x_2 < x_3$.

Further examining Eqn. 4.160, we see that x_1 and x_3 are stable fixed points and that x_2 is an unstable fixed point of this one-dimensional dynamical system. Thus, there are two possible stable equilibria. If

$x(0) < x_2$ the flow will be toward x_1, while if $x(0) > x_2$ the flow will be toward x_3. We can integrate Eqn. 4.160 using the method of partial fractions. First, we write

$$\frac{1}{(x - x_1)(x - x_2)(x - x_3)} = \frac{A_1}{x - x_1} + \frac{A_2}{x - x_2} + \frac{A_3}{x - x_3} \, , \tag{4.161}$$

with $(x - x_2)(x - x_3)A_1 + (x - x_1)(x - x_3)A_2 + (x - x_1)(x - x_2)A_3 = 1$. This requires

$$
\begin{aligned}
0 &= A_1 + A_2 + A_3 \\
0 &= (x_2 + x_3)A_1 + (x_1 + x_3)A_2 + (x_1 + x_2)A_3 \\
1 &= x_2 x_3 A_1 + x_1 x_3 A_2 + x_1 x_2 A_3 \quad ,
\end{aligned}
\tag{4.162}
$$

with solution

$$A_1 = \frac{1}{(x_2 - x_1)(x_3 - x_1)} \quad , \quad A_2 = -\frac{1}{(x_2 - x_1)(x_3 - x_2)} \quad , \quad A_3 = \frac{1}{(x_3 - x_1)(x_3 - x_2)} \quad . \tag{4.163}$$

Thus, Eqn. 4.160 may be recast as

$$(x_3 - x_2) \, d \ln(x - x_1) - (x_3 - x_1) \, d \ln(x - x_2) + (x_2 - x_1) \, d \ln(x - x_3) = -k_4 (x_2 - x_1)(x_3 - x_1)(x_3 - x_2) \, dt \, . \tag{4.164}$$

The solution is given in terms of $t(x)$:

$$t(x) = \frac{1}{k_4(x_2 - x_1)(x_3 - x_1)} \ln\left(\frac{x_0 - x_1}{x - x_1}\right) \tag{4.165}$$

$$- \frac{1}{k_4(x_2 - x_1)(x_3 - x_2)} \ln\left(\frac{x_0 - x_2}{x - x_2}\right) + \frac{1}{k_4(x_3 - x_1)(x_3 - x_2)} \ln\left(\frac{x_0 - x_2}{x - x_3}\right) ,$$

where $x_0 = x(0)$.

Going back to Eqn. 4.139, we have that the steady state distribution is

$$P_{\text{eq}}(n) = P_{\text{eq}}(0) \prod_{j=1}^{n} \frac{t^+(j - 1)}{t^-(j)} = P_{\text{eq}}(0) \prod_{j=1}^{n} \frac{k_1 \, a + k_3 \, b \, (j - 1)(j - 2)}{k_2 \, j + k_4 \, j \, (j - 1)(j - 2)} \, . \tag{4.166}$$

The product is maximized for when the last term with $j = n$ is unity. If we call this value n^*, then n^* is a root of the equation

$$k_1 \, a + k_3 \, b \, (n - 1)(n - 2) = k_2 \, n + k_4 \, n \, (n - 1)(n - 2) \, . \tag{4.167}$$

If $n \gg 1$ and all the terms are roughly the same size, this equation becomes $k_1 \, a + k_3 \, b \, n^2 = k_2 \, n + k_4 \, n^3$, which is the same as setting the RHS of Eqn. 4.160 to zero in order to find a stationary solution.

4.3.3 Forward and reverse equations and boundary conditions

In §2.6.3 we discussed the forward and backward differential Chapman-Kolmogorov equations, from which, with $A_\mu = 0$ and $B_{\mu\nu} = 0$, we obtain the forward and reverse Master equations,

$$
\begin{aligned}
\frac{\partial P(\boldsymbol{n}, t \,|\, \cdot)}{\partial t} &= \sum_{\boldsymbol{m}} \Big\{ W(\boldsymbol{n} \,|\, \boldsymbol{m}, t) \, P(\boldsymbol{m}, t \,|\, \cdot) - W(\boldsymbol{m} \,|\, \boldsymbol{n}, t) \, P(\boldsymbol{n}, t \,|\, \cdot) \Big\} \\
-\frac{\partial P(\cdot \,|\, \boldsymbol{n}, t)}{\partial t} &= \sum_{\boldsymbol{m}} W(\boldsymbol{m} \,|\, \boldsymbol{n}, t) \Big\{ P(\cdot \,|\, \boldsymbol{m}, t) - P(\cdot \,|\, \boldsymbol{n}, t) \Big\} \quad ,
\end{aligned}
\tag{4.168}
$$

where we have suppressed the initial conditions in the forward equation and the final conditions in the backward equation. Consider the one-dimensional version, and take the transition rates to be

$$W(j' \,|\, j, t) = t^+(j)\, \delta_{j',j+1} + t^-(j)\, \delta_{j',j-1} \quad . \tag{4.169}$$

We may then write

$$\frac{\partial P(n,t\,|\,\cdot)}{\partial t} = \mathcal{L}P(n,t\,|\,\cdot) = \overbrace{\left\{ t^+(n-1)\, P(n-1,t\,|\,\cdot) - t^-(n)\, P(n,t\,|\,\cdot) \right\}}^{J(n-1,t\,|\,\cdot)} - \overbrace{\left\{ t^+(n)\, P(n,t\,|\,\cdot) - t^-(n+1)\, P(n+1,t\,|\,} ^{J(n,t\,|\,\cdot)}$$

$$-\frac{\partial P(\cdot\,|\,n,t)}{\partial t} = \widetilde{\mathcal{L}}P(\cdot\,|\,n,t) = t^+(n)\, \overbrace{\left\{ P(\cdot\,|\,n+1,t) - P(\cdot\,|\,n,t) \right\}}^{K(\cdot\,|\,n+1,t)} - t^-(n)\, \overbrace{\left\{ P(\cdot\,|\,n,t) - P(\cdot\,|\,n-1,t) \right\}}^{K(\cdot\,|\,n,t)} \quad , \tag{4.170}$$

where we have defined the quantities $J(n,t\,|\,\cdot)$ and $K(\cdot\,|\,n,t)$. Here $(\mathcal{L}f)_n = \mathcal{L}_{nn'}\, f_{n'}$ and $(\widetilde{\mathcal{L}}f)_n = \widetilde{\mathcal{L}}_{nn'}\, f_{n'}$, where $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ are matrices, *viz.*

$$\mathcal{L}_{nn'} = t^+(n')\, \delta_{n',n-1} + t^-(n')\, \delta_{n',n+1} - t^+(n')\, \delta_{n',n} - t^-(n')\, \delta_{n',n}$$

$$\widetilde{\mathcal{L}}_{nn'} = t^+(n)\, \delta_{n',n+1} + t^-(n)\, \delta_{n',n-1} - t^+(n)\, \delta_{n',n} - t^-(n)\, \delta_{n',n} \quad . \tag{4.171}$$

Clearly $\widetilde{\mathcal{L}}_{nn'} = \mathcal{L}_{n'n}$, hence $\widetilde{\mathcal{L}} = \mathcal{L}^{\mathrm{t}}$, the matrix transpose, if we can neglect boundary terms. For $n, n' \in \mathbb{Z}$, we could specify $P(\pm\infty, t\,|\,\cdot) = P(\cdot\,|\,\pm\infty, t) = 0$.

Consider now a birth-death process where we focus on a finite interval $n \in \{a, \ldots, b\}$. Define the inner product

$$\langle\, g\,|\,\mathcal{O}\,|\,f\,\rangle = \sum_{n=a}^{b} g(n)\, (\mathcal{O}f)(n) \;. \tag{4.172}$$

One then has

$$\begin{aligned} \langle\, g\,|\,\mathcal{L}\,|\,f\,\rangle - \langle\, f\,|\,\widetilde{\mathcal{L}}\,|\,g\,\rangle = {}& t^-(b+1)\, f(b+1)\, g(b) - t^+(b)\, f(b)\, g(b+1) \\ & + t^+(a-1)\, f(a-1)\, g(a) - t^-(a)\, f(a)\, g(a-1) \;. \end{aligned} \tag{4.173}$$

Thus, if $f(a-1) = g(a-1) = f(b+1) = g(b+1) = 0$, we have $\widetilde{\mathcal{L}} = \mathcal{L}^{\mathrm{t}} = \mathcal{L}^\dagger$, the adjoint. In the suppressed initial and final conditions, we always assume the particle coordinate n lies within the interval.

We now must specify appropriate boundary conditions on our interval. These conditions depend on whether we are invoking the forward or backward Master equation:

Forward equation : For reflecting boundaries, we set $t^-(a) = 0$ and $t^+(b) = 0$, assuring that a particle starting from inside the region can never exit. We also specify $P(a - 1, t\,|\,\cdot) = 0$ and $P(b + 1, t\,|\,\cdot) = 0$ so that no particles can enter from the outside. This is equivalent to specifying that the boundary currents vanish, *i.e.* $J(a - 1, t\,|\,\cdot) = 0$ and $J(b, t\,|\,\cdot) = 0$, respectively. For absorbing boundaries, we choose $t^+(a - 1) = 0$ and $t^-(b + 1) = 0$, which assures

that a particle which exits the region can never reenter. This is equivalent to demanding $P(a-1, t \,|\, \cdot) = 0$ and $P(b+1, t \,|\, \cdot) = 0$, respectively.

Backward equation : From Eqn. 4.170, it is clear that the reflecting conditions $t^-(a) = 0$ and $t^+(b) = 0$ are equivalent to $K(\cdot \,|\, a, t) = 0$ and $K(\cdot \,|\, b+1, t) = 0$, where these functions. Neither of the quantities in the absorbing conditions $t^+(a-1) = 0$ and $t^-(b+1) = 0$ enter in the backward Master equation. The effect of these conditions on the data outside the interval is to preserve $P(\cdot \,|\, a-1, t) = 0$ and $P(\cdot \,|\, b+1, t) = 0$, respectively.

The situation is summarized in Tab. 4.3.3 below.

		conditions		equivalent conditions	
equation	boundary	reflecting	absorbing	reflecting	absorbing
FORWARD	left	$t^-(a) = 0$	$t^+(a-1) = 0$	$J(a-1, t \,|\, \cdot) = 0$	$P(a-1, t \,|\, \cdot)$
	right	$t^+(b) = 0$	$t^-(b+1) = 0$	$J(b, t \,|\, \cdot) = 0$	$P(b+1, t \,|\, \cdot)$
BACKWARD	left	$t^-(a) = 0$	$t^+(a-1) = 0$	$K(\cdot \,|\, a, t) = 0$	$P(\cdot \,|\, a-1, t)$
	right	$t^+(b) = 0$	$t^-(b+1) = 0$	$K(\cdot \,|\, b+1, t) = 0$	$P(\cdot \,|\, b+1, t)$

Table 4.1: Absorbing and reflecting boundary conditions for the Master equation on the interval $\{a, \dots, b\}$.

4.3.4 First passage times

The treatment of first passage times within the Master equation follows that for the Fokker-Planck equation in §4.2.5. If our discrete particle starts at n at time $t_0 = 0$, the probability that it lies within the interval $\{a, \dots, b\}$ at some later time t is

$$G(n, t) = \sum_{n'=a}^{b} P(n', t \,|\, n, 0) = \sum_{n'=a}^{b} P(n', 0 \,|\, n, -t) \quad , \tag{4.174}$$

and therefore $-\partial_t G(n, t) \, dt$ is the probability that the particle exits the interval within the time interval $[t, t+dt]$. Therefore the average first passage time out of the interval, starting at n at time $t_0 = 0$, is

$$T(n) = \int_0^\infty dt \, t \left(-\frac{\partial G(n, t)}{\partial t} \right) = \int_0^\infty dt \, G(n, t) \ . \tag{4.175}$$

Applying $\widetilde{\mathcal{L}}$, we obtain

$$\widetilde{\mathcal{L}}\, T(n) = t^+(n) \left\{ T(n+1) - T(n) \right\} - t^-(n) \left\{ T(n) - T(n-1) \right\} = -1 \quad . \tag{4.176}$$

Let a be a reflecting barrier and b be absorbing. Since $t^-(a) = 0$ we are free to set $T(a-1) = T(a)$. At the right boundary we have $T(b+1) = 0$, because a particle starting at $b+1$ is already outside the interval. Eqn. 4.176 may be written

$$t^+(n)\, \Delta T(n) - t^-(n)\, \Delta T(n-1) = -1 \,, \tag{4.177}$$

with $\Delta T(n) \equiv T(n+1) - T(n)$. Now define the function

$$\phi(n) = \prod_{j=a+1}^{n} \frac{t^-(j)}{t^+(j)} \quad , \tag{4.178}$$

with $\phi(a) \equiv 1$. This satisfies $\phi(n)/\phi(n-1) = t^-(n)/t^+(n)$, and therefore Eqn. 4.177 may be recast as

$$\frac{\Delta T(n)}{\phi(n)} = \frac{\Delta T(n-1)}{\phi(n-1)} - \frac{1}{t^+(n)\, \phi(n)} \quad . \tag{4.179}$$

Since $\Delta T(a) = -1/t^+(a)$ from Eqn. 4.176, the first term on the RHS above vanishes for $n = a$. We then have

$$\Delta T(n) = -\phi(n) \sum_{j=a}^{n} \frac{1}{t^+(j)\, \phi(j)} \quad , \tag{4.180}$$

and therefore, working backward from $T(b+1) = 0$, we have

$$T(n) = \sum_{k=n}^{b} \phi(k) \sum_{j=a}^{k} \frac{1}{t^+(j)\, \phi(j)} \qquad (a \text{ reflecting}, b \text{ absorbing}). \tag{4.181}$$

One may also derive

$$T(n) = \sum_{k=a}^{n} \phi(k) \sum_{j=k}^{b} \frac{1}{t^+(j)\, \phi(j)} \qquad (a \text{ absorbing}, b \text{ reflecting}). \tag{4.182}$$

Example

Suppose $a = 0$ is reflecting and $b = N - 1$ is absorbing, and furthermore suppose that $t^\pm(n) = t^\pm$ are site-independent. Then $\phi(n) = r^{-n}$, where $r \equiv t^+/t^-$. The mean escape time starting from site n is

$$T(n) = \frac{1}{t^+} \sum_{k=n}^{N-1} r^{-k} \sum_{j=0}^{k} r^j$$
$$= \frac{1}{(r-1)^2\, t^+} \left\{ (N-n)(r-1) + r^{-N} - r^{-n} \right\} \quad . \tag{4.183}$$

If $t^+ = t^-$, so the walk is unbiased, then $r = 1$. We can then evaluate by taking $r = 1 + \varepsilon$ with $\varepsilon \to 0$, or, more easily, by evaluating the sum in the first line when $r = 1$. The result is

$$T(n) = \frac{1}{t^+} \left\{ \tfrac{1}{2} N(N-1) - \tfrac{1}{2} n(n+1) + N - n \right\} \qquad (r = 1) . \qquad (4.184)$$

By taking an appropriate limit, we can compare with the Fokker-Planck result of Eqn. 4.61, which for an interval $[a, b]$ with $a = 0$ reflecting and b absorbing yields $T(x) = (b^2 - x^2)/2D$. Consider the Master equation,

$$\frac{\partial P(n,t)}{\partial t} = \beta \left[P(n+1,t) + P(n-1,t) - 2P(n,t) \right] = \beta \frac{\partial^2 P}{\partial n^2} + \tfrac{1}{12} \beta \frac{\partial^4 P}{\partial n^4} + \dots , \qquad (4.185)$$

where $\beta = t^+ = t^-$. Now define $n \equiv Nx/b$, and rescale both time $t \equiv N\tau$ and hopping $\beta \equiv N\gamma$, resulting in

$$\frac{\partial P}{\partial \tau} = D \frac{\partial^2 P}{\partial x^2} + \frac{Db^2}{12N^2} \frac{\partial^4 P}{\partial x^4} + \dots , \qquad (4.186)$$

where $D = b^2 \gamma$ is the diffusion constant. In the continuum limit, $N \to \infty$ and we may drop all terms beyond the first on the RHS, yielding the familiar diffusion equation. Taking this limit, Eqn. 4.184 may be rewritten as $T(x)/N = (N/2t^+ b^2)(b^2 - x^2) = (b^2 - x^2)/2D$, which agrees with the result of Eqn. 4.61.

4.3.5 From Master equation to Fokker-Planck

Let us start with the Master equation,

$$\frac{\partial P(\boldsymbol{x},t)}{\partial t} = \int d\boldsymbol{x}' \left[W(\boldsymbol{x} \,|\, \boldsymbol{x}') \, P(\boldsymbol{x}',t) - W(\boldsymbol{x}' \,|\, \boldsymbol{x}) \, P(\boldsymbol{x},t) \right] , \qquad (4.187)$$

and define $W(\boldsymbol{z} \,|\, \boldsymbol{z}_0) \equiv t(\boldsymbol{z} - \boldsymbol{z}_0 \,|\, \boldsymbol{z}_0)$, which rewrites the rate $W(\boldsymbol{z} \,|\, \boldsymbol{z}_0)$ from $\boldsymbol{z}_0$ to $\boldsymbol{z}$ as a function of $\boldsymbol{z}_0$ and the distance $\boldsymbol{z} - \boldsymbol{z}_0$ to $\boldsymbol{z}$. Then the Master equation may be rewritten as

$$\frac{\partial P(\boldsymbol{x},t)}{\partial t} = \int d\boldsymbol{y} \left[t(\boldsymbol{y} \,|\, \boldsymbol{x} - \boldsymbol{y}) \, P(\boldsymbol{x} - \boldsymbol{y}, t) - t(\boldsymbol{y} \,|\, \boldsymbol{x}) \, P(\boldsymbol{x}, t) \right] . \qquad (4.188)$$

Now expand $t(\boldsymbol{y} \,|\, \boldsymbol{x} - \boldsymbol{y}) \, P(\boldsymbol{x} - \boldsymbol{y})$ as a power series in the jump distance $\boldsymbol{y}$ to obtain[12]

$$\begin{aligned}
\frac{\partial P(\boldsymbol{x},t)}{\partial t} &= \int d\boldsymbol{y} \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \, y_{\alpha_1} \cdots y_{\alpha_n} \frac{\partial^n}{\partial x_{\alpha_1} \cdots \partial x_{\alpha_n}} \left[t(\boldsymbol{y} \,|\, \boldsymbol{x}) \, P(\boldsymbol{x}, t) \right] \\
&= \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x_{\alpha_1} \cdots \partial x_{\alpha_n}} \left[R^{\alpha_1 \cdots \alpha_n}(\boldsymbol{x}) \, P(\boldsymbol{x}, t) \right] ,
\end{aligned} \qquad (4.189)$$

where

$$R^{\alpha_1 \cdots \alpha_n}(\boldsymbol{x}) = \int d\boldsymbol{y} \; y_{\alpha_1} \cdots y_{\alpha_n} \, t(\boldsymbol{y} \,|\, \boldsymbol{x}) . \qquad (4.190)$$

---

[12]We only expand the second argument of $t(\boldsymbol{y} \,|\, \boldsymbol{x} - \boldsymbol{y})$ in $\boldsymbol{y}$. We retain the full $\boldsymbol{y}$-dependence of the first argument.

For $d = 1$ dimension, we may write

$$\frac{\partial P(x,t)}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \Big[R_n(x) \, P(x,t) \Big] \quad , \qquad R_n(x) \equiv \int dy \, y^n \, t(y \,|\, x) \quad . \tag{4.191}$$

This is known as the *Kramers-Moyal* expansion. If we truncate at order $n = 2$, we obtain the Fokker-Planck equation,

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x} \Big[R_1(x) \, P(x,t) \Big] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \Big[R_2(x) \, P(x,t) \Big] \quad . \tag{4.192}$$

The problem is that the FPE here is akin to a Procrustean bed. We have amputated the $n > 2$ terms from the expansion without any justification at all, and we have no reason to expect this will end well. A more systematic approach was devised by N. G. van Kampen, and goes by the name of the *size expansion*. One assumes that there is a large quantity lurking about, which we call Ω. Typically this can be the total system volume, or the total population in the case of an ecological or epidemiological model. One assumes that $t(y \,|\, x)$ obeys a scaling form,

$$t(\Delta z \,|\, z_0) = \Omega \, \tau \Big(\Delta z \, \Big| \, \frac{z_0}{\Omega} \Big) \quad . \tag{4.193}$$

From the second of Eqn. 4.191, we then have

$$R_n(x) = \Omega \int dy \, y^n \, \tau \Big(y \, \Big| \, \frac{x}{\Omega} \Big) \equiv \Omega \, \widetilde{R}_n(x/\Omega) \quad . \tag{4.194}$$

We now proceed by defining

$$x = \Omega \, \phi(t) + \sqrt{\Omega} \, \xi \quad , \tag{4.195}$$

where $\phi(t)$ is an as-yet undetermined function of time, and ξ is to replace x, so that our independent variables are now (ξ, t). We therefore have

$$R_n(x) = \Omega \, \widetilde{R}_n\big(\phi(t) + \Omega^{-1/2}\xi \big) \quad . \tag{4.196}$$

Now we are set to derive a systematic expansion in inverse powers of Ω. We define $P(x,t) = \Pi(\xi, t)$, and we note that $dx = \Omega \, \dot{\phi} \, dt + \sqrt{\Omega} \, d\xi$, hence $d\xi\big|_x = -\sqrt{\Omega} \, \dot{\phi} \, dt$, which means

$$\frac{\partial P(x,t)}{\partial t} = \frac{\partial \Pi(\xi, t)}{\partial t} - \sqrt{\Omega} \, \dot{\phi} \, \frac{\partial \Pi(\xi, t)}{\partial \xi} \quad . \tag{4.197}$$

We therefore have, from Eqn. 4.191,

$$\frac{\partial \Pi(\xi, t)}{\partial t} - \sqrt{\Omega} \, \dot{\phi} \, \frac{\partial \Pi}{\partial \xi} = \sum_{n=1}^{\infty} \frac{(-1)^n \, \Omega^{(2-n)/2}}{n!} \frac{\partial^n}{\partial \xi^n} \Big[\widetilde{R}_n\big(\phi(t) + \Omega^{-1/2}\xi \big) \, \Pi(\xi, t) \Big] \quad . \tag{4.198}$$

Further expanding $\widetilde{R}_n(\phi + \Omega^{-1/2}\xi)$ in powers of $\Omega^{-1/2}$, we obtain

$$\frac{\partial \Pi(\xi, t)}{\partial t} - \sqrt{\Omega} \, \dot{\phi} \, \frac{\partial \Pi}{\partial \xi} = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} \frac{(-1)^n \, \Omega^{(2-n-k)/2}}{n! \, k!} \frac{d^k \widetilde{R}_n(\phi)}{d\phi^k} \bigg|_{\phi(t)} \frac{\partial^n}{\partial \xi^n} \Big[\xi^k \, \Pi(\xi, t) \Big] \quad . \tag{4.199}$$

Let's define an index $l \equiv n + k$, which runs from 1 to ∞. Clearly $n = l - k$, which for fixed l runs from 1 to l. In this way, we can reorder the terms in the sum, according to

$$\sum_{k=0}^{\infty} \sum_{n=1}^{\infty} A(k,n) = \sum_{l=1}^{\infty} \sum_{n=1}^{l} A(l-n, n) \quad . \tag{4.200}$$

The lowest order term on the RHS of Eqn. 4.199 is the term with $n = 1$ and $k = 0$, corresponding to $l = n = 1$ if we eliminate the k index in favor of l. It is equal to $-\sqrt{\Omega}\, \widetilde{R}_1\big(\phi(t)\big)\, \partial_\xi \Pi$, hence if we demand that $\phi(t)$ satisfy

$$\frac{d\phi}{dt} = \widetilde{R}_1(\phi) \quad , \tag{4.201}$$

these terms cancel from either side of the equation. We then have

$$\frac{\partial \Pi(\xi, t)}{\partial t} = \sum_{l=2}^{\infty} \Omega^{(2-l)/2} \sum_{n=1}^{l} \frac{(-1)^n}{n!\,(l-n)!}\, \widetilde{R}_n^{(l-n)}\big(\phi(t)\big) \frac{\partial^n}{\partial \xi^n} \Big[\xi^{l-n}\, \Pi(\xi, t) \Big] \quad , \tag{4.202}$$

where $\widetilde{R}_n^{(k)}(\phi) = d^k \widetilde{R}_n / d\phi^k$. We are now in a position to send $\Omega \to \infty$, in which case only the $l = 2$ term survives, and we are left with

$$\frac{\partial \Pi}{\partial t} = -\widetilde{R}_1'\big(\phi(t)\big) \frac{\partial\,(\xi \Pi)}{\partial \xi} + \tfrac{1}{2} \widetilde{R}_2\big(\phi(t)\big) \frac{\partial^2 \Pi}{\partial \xi^2} \quad , \tag{4.203}$$

which is a Fokker-Planck equation.

Birth-death processes

Consider a birth-death process in which the states $|\,n\,\rangle$ are labeled by nonnegative integers. Let α_n denote the rate of transitions from $|\,n\,\rangle \to |\,n+1\,\rangle$ and let β_n denote the rate of transitions from $|\,n\,\rangle \to |\,n-1\,\rangle$. The Master equation then takes the form[13]

$$\frac{dP_n}{dt} = \alpha_{n-1} P_{n-1} + \beta_{n+1} P_{n+1} - \big(\alpha_n + \beta_n\big) P_n \,, \tag{4.204}$$

where we abbreviate $P_n(t)$ for $P(n, t \,|\, n_0, t_0)$ and suppress the initial conditions (n_0, t_0).

Let us assume we can write $\alpha_n = K\bar{\alpha}(n/K)$ and $\beta_n = K\bar{\beta}(n/K)$, where $K \gg 1$. Define $x \equiv n/K$, so the Master equation becomes

$$\frac{\partial P}{\partial t} = K\bar{\alpha}(x - \tfrac{1}{K})P(x - \tfrac{1}{K}) + K\bar{\beta}(x + \tfrac{1}{K})P(x + \tfrac{1}{K}) - K\big(\bar{\alpha}(x) + \bar{\beta}(x)\big)P(x)$$

$$= -\frac{\partial}{\partial x}\Big[\big(\bar{\alpha}(x) - \bar{\beta}(x)\big)P(x,t)\Big] + \frac{1}{2K}\frac{\partial^2}{\partial x^2}\Big[\big(\bar{\alpha}(x) + \bar{\beta}(x)\big)P(x,t)\Big] + \mathcal{O}(K^{-2}) \quad . \tag{4.205}$$

If we truncate the expansion after the $\mathcal{O}(K^{-1})$ term, we obtain

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\Big[f(x)\,P(x,t)\Big] + \frac{1}{2K}\frac{\partial^2}{\partial x^2}\Big[g(x)\,P(x,t)\Big] \quad , \tag{4.206}$$

---

[13] We further demand $\beta_{n=0} = 0$ and $P_{-1}(t) = 0$ at all times.

where we have defined

$$f(x) \equiv \bar{\alpha}(x) - \bar{\beta}(x) \qquad , \qquad g(x) \equiv \bar{\alpha}(x) + \bar{\beta}(x) \quad . \tag{4.207}$$

This FPE has an equilibrium solution

$$P_{\text{eq}}(x) = \frac{A}{g(x)}\, e^{-K\Phi(x)} \qquad , \qquad \Phi(x) = -2\int\limits_{0}^{x} dx'\, \frac{f(x')}{g(x')} \quad , \tag{4.208}$$

where the constant A is determined by normalization. If K is large, we may expand about the minimum of $\Phi(x)$

$$\begin{aligned}
\Phi(x) &= \Phi(x^*) - \frac{2f(x^*)}{g(x^*)}\,(x - x^*) + \frac{2f(x^*)\,g'(x^*) - 2g(x^*)\,f'(x^*)}{g^2(x^*)}\,(x - x^*)^2 + \dots \\
&= \Phi(x^*) - \frac{2f'(x^*)}{g(x^*)}\,(x - x^*)^2 + \dots \quad .
\end{aligned} \tag{4.209}$$

Thus, we obtain a Gaussian distribution

$$P_{\text{eq}}(x) \simeq \sqrt{\frac{K}{2\pi\sigma^2}}\, e^{-K(x-x^*)^2/2\sigma^2} \quad \text{with} \quad \sigma^2 = -\frac{g(x^*)}{2f'(x^*)} \quad . \tag{4.210}$$

In order that the distribution be normalizable, we must have $f'(x^*) < 0$.

In §4.3.6, we will see how the Fokker-Planck expansion fails to account for the large $\mathcal{O}(K)$ fluctuations about a metastable equilibrium which lead to rare extinction events in this sort of birth-death process.

van Kampen treatment

We now discuss the same birth-death process using van Kampen's size expansion. Assume the distribution $P_n(t)$ has a time-dependent maximum at $n = K\phi(t)$ and a width proportional to $\sqrt{K}$. We expand relative to this maximum, writing $n \equiv K\phi(t) + \sqrt{K}\,\xi$ and we define $P_n(t) \equiv \Pi(\xi, t)$. We now rewrite the Master equation in eqn. 4.204 in terms of $\Pi(\xi, t)$. Since n is an independent variable, we set

$$dn = K\dot{\phi}\, dt + \sqrt{K}\, d\xi \quad \Rightarrow \quad d\xi\big|_{n} = -\sqrt{K}\,\dot{\phi}\, dt \;. \tag{4.211}$$

Therefore

$$\frac{dP_n}{dt} = -\sqrt{K}\,\dot{\phi}\,\frac{\partial\Pi}{\partial\xi} + \frac{\partial\Pi}{\partial t} \;. \tag{4.212}$$

We now write

$$\begin{aligned}
\alpha_{n-1}\, P_{n-1} &= K\,\bar{\alpha}\big(\phi + K^{-1/2}\xi - K^{-1}\big)\, \Pi\big(\xi - K^{-1/2}\big) \\
\beta_{n+1}\, P_{n+1} &= K\,\bar{\beta}\big(\phi + K^{-1/2}\xi + K^{-1}\big)\, \Pi\big(\xi + K^{-1/2}\big) \\
\big(\alpha_n + \beta_n\big)\, P_n &= K\,\bar{\alpha}\big(\phi + K^{-1/2}\xi\big)\, \Pi(\xi) + K\,\bar{\beta}\big(\phi + K^{-1/2}\xi\big)\, \Pi(\xi) \quad ,
\end{aligned} \tag{4.213}$$

and therefore Eqn. 4.204 becomes

$$-\sqrt{K}\frac{\partial \Pi}{\partial \xi}\,\dot{\phi} + \frac{\partial \Pi}{\partial t} = \sqrt{K}\,(\bar{\beta}-\bar{\alpha})\,\frac{\partial \Pi}{\partial \xi} + (\bar{\beta}'-\bar{\alpha}')\,\xi\,\frac{\partial \Pi}{\partial \xi} + (\bar{\beta}'-\bar{\alpha}')\Pi + \tfrac{1}{2}(\bar{\alpha}+\bar{\beta})\,\frac{\partial^2 \Pi}{\partial \xi^2} + \mathcal{O}\big(K^{-1/2}\big) \quad , \text{ (4.214)}$$

where $\bar{\alpha}=\bar{\alpha}(\phi)$ and $\bar{\beta}=\bar{\beta}(\phi)$. Equating terms of order $\sqrt{K}$ yields the equation

$$\dot{\phi} = f(\phi) \equiv \bar{\alpha}(\phi) - \bar{\beta}(\phi) \quad , \tag{4.215}$$

which is a first order ODE for the quantity $\phi(t)$. Equating terms of order K^0 yields the Fokker-Planck equation,

$$\frac{\partial \Pi}{\partial t} = -f'\big(\phi(t)\big)\,\frac{\partial}{\partial \xi}\,\big(\xi\Pi\big) + \tfrac{1}{2}\,g\big(\phi(t)\big)\,\frac{\partial^2 \Pi}{\partial \xi^2} \quad , \tag{4.216}$$

where $g(\phi) \equiv \bar{\alpha}(\phi) + \bar{\beta}(\phi)$. If in the limit $t \to \infty$, eqn. 4.215 evolves to a stable fixed point ϕ^*, then the stationary solution of the Fokker-Planck eqn. 4.216, $\Pi_{\text{eq}}(\xi) = \Pi(\xi, t=\infty)$ must satisfy

$$-f'(\phi^*)\,\frac{\partial}{\partial \xi}\,\big(\xi\,\Pi_{\text{eq}}\big) + \tfrac{1}{2}\,g(\phi^*)\,\frac{\partial^2 \Pi_{\text{eq}}}{\partial \xi^2} = 0 \quad \Rightarrow \quad \Pi_{\text{eq}}(\xi) = \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\xi^2/2\sigma^2} \quad , \tag{4.217}$$

where

$$\sigma^2 = -\frac{g(\phi^*)}{2f'(\phi^*)} \quad . \tag{4.218}$$

Now both α and β are rates, hence both are positive and thus $g(\phi) > 0$. We see that the condition $\sigma^2 > 0$, which is necessary for a normalizable equilibrium distribution, requires $f'(\phi^*) < 0$, which is saying that the fixed point in Eqn. 4.215 is stable.

We thus arrive at the same distribution as in Eqn. 4.210. The virtue of this latter approach is that we have a better picture of how the distribution evolves toward its equilibrium value. The condition of normalizability $f'(x^*) < 0$ is now seen to be connected with the dynamics of location of the instantaneous maximum of $P(x,t)$, namely $x = \phi(t)$. If the dynamics of the FPE in Eqn. 4.216 are fast compared with those of the simple dynamical system in Eqn. 4.215, we may regard the evolution of $\phi(t)$ as adiabatic so far as $\Pi(\xi, t)$ is concerned.

4.3.6 Extinction times in birth-death processes

In §4.3.1 we discussed the Master equation for birth-death processes,

$$\frac{dP_n}{dt} = t^+(n-1)\,P_{n-1} + t^-(n+1)\,P_{n+1} - \big[t^+(n) + t^-(n)\big]\,P_n \quad . \tag{4.219}$$

At the mean field level, we have for the average population $\bar{n} = \sum_n n\,P_n$,

$$\frac{d\bar{n}}{dt} = t^+(\bar{n}) - t^-(\bar{n}) \quad . \tag{4.220}$$

Two models from population biology that merit our attention here:

Susceptible-infected-susceptible (SIS) model : Consider a population of fixed total size N, among which n individuals are infected and the remaining $N - n$ are susceptible. The number of possible contacts between infected and susceptible individuals is then $n(N - n)$, and if the infection rate per contact is Λ/N and the recovery rate of infected individuals is set to unity[14], then we have

$$t^+(n) = \Lambda n \left(1 - \frac{n}{N}\right) \qquad , \qquad t^-(n) = n \quad . \tag{4.221}$$

Verhulst model : Here the birth rate is B and the death rate is unity plus a stabilizing term $(B/N)\,n$ which increases linearly with population size. Thus,

$$t^+(n) = Bn \qquad , \qquad t^-(n) = n + \frac{Bn^2}{N} \quad . \tag{4.222}$$

The mean field dynamics of both models is the same, with

$$\frac{d\bar{n}}{dt} = (\Lambda - 1)\bar{n} - \frac{\Lambda \bar{n}^2}{N} \tag{4.223}$$

for the SIS model; take $\Lambda \to B$ for the Verhulst model. This is known as the *logistic equation*: $\dot{\bar{n}} = r\bar{n}(K - \bar{n})$, with $r = \Lambda/N$ the growth rate and $K = (\Lambda - 1)/\Lambda$ the equilibrium population. If $\Lambda > 1$ then $K > 0$, in which case the fixed point at $\bar{n} = 0$ is unstable and the fixed point at $\bar{n} = K$ is stable. The asymptotic state is one of an equilibrium number K of infected individuals. At $\Lambda = 1$ there is a transcritical bifurcation, and for $0 < \Lambda < 1$ we have $K < 0$, and the unphysical fixed point at $\bar{n} = K$ is unstable, while the fixed point at $\bar{n} = 0$ is stable. The infection inexorably dies out. So the mean field dynamics for $\Lambda > 1$ are a simple flow to the stable fixed point (SFP) at $\bar{n} = K$, and those for $\Lambda < 1$ are a flow to the SFP at $\bar{n} = 0$. In both cases, the approach to the SFP takes a logarithmically infinite amount of time.

Although the mean field solution for $\Lambda > 1$ asymptotically approaches an equilibrium number of infected individuals K, the stochasticity in this problem means that *there is a finite extinction time for the infection.* The extinction time is the first passage time to the state $n = 0$. Once the population of infected individuals goes to zero, there is no way for new infections to spontaneously develop. The mean first passage time was studied in §4.3.4. We have an absorbing boundary at $n = 1$, since $t^+(0) = 0$, and a reflecting boundary at $n = N$, since $t^+(N) = 0$, and Eqn. 4.182 gives the mean first passage time for absorption as

$$T(n) = \sum_{k=1}^{n} \phi(k) \sum_{j=k}^{N} \frac{1}{t^+(j)\,\phi(j)} \quad , \tag{4.224}$$

where[15]

$$\phi(k) = \prod_{l=1}^{k} \frac{t^-(l)}{t^+(l)} \quad . \tag{4.225}$$

---

[14]That is, we measure time in units of the recovery time.

[15]In §4.3.4, we defined $\phi(a) = 1$ where $a = 1$ is the absorbing boundary here, whereas in Eqn. 4.225 we have $\phi(1) = t^+(1)/t^-(1)$. Since the mean first passage time $T(n)$ does not change when all $\phi(n)$ are multiplied by the same constant, we are free to define $\phi(a)$ any way we please. In this chapter it pleases me to define it as described.

The detailed analysis of $T(n)$ is rather tedious, and is described in the appendices to C. Doering *et al.*, *Multiscale Model Simul.* **3**, 283 (2005). For our purposes, it suffices to consider the behavior of the function $\phi(n)$. Let $x \equiv n/N \in [0, 1]$. Then with $y \equiv j/N$ define

$$\rho(y) \equiv \frac{t^+(j)}{t^-(j)} = \Lambda(1 - y) \quad , \tag{4.226}$$

in which case, using the trapezoidal rule, and setting $x \equiv n/N$,

$$-\ln \phi(n) = \sum_{l=1}^{n} \ln \rho(l/N)$$

$$\approx -\tfrac{1}{2} \ln \rho(0) - \tfrac{1}{2} \ln \rho(x) + N \int_0^x du \, \ln \rho(u) \tag{4.227}$$

$$= N \Big\{ \ln \Lambda - (1 - x) \ln \Lambda - (1 - x) \ln(1 - x) - x \Big\} - \ln \Lambda - \tfrac{1}{2} \ln(1 - x) \quad .$$

In the $N \to \infty$ limit, the maximum occurs at $x^* = (\Lambda - 1)/\Lambda$, which for $\Lambda > 1$ is the scaled mean field equilibrium population of infected individuals. For $x \approx x^*$, the mean extinction time for the infection is therefore

$$T(x^*) \sim e^{N\Phi(\Lambda)} \quad , \quad \Phi(\Lambda) = \ln \Lambda - 1 + \Lambda^{-1} \quad . \tag{4.228}$$

The full result, from Doering *et al.*, is

$$T(x^*) = \frac{\Lambda}{(\Lambda - 1)^2} \sqrt{\frac{2\pi}{N}} \, e^{N\left(\ln \Lambda - 1 + \Lambda^{-1}\right)} \times \Big(1 + \mathcal{O}(N^{-1}) \Big) \tag{4.229}$$

The extinction time is exponentially large in the population size.

Below threshold, when $\Lambda < 1$, Doering *et al.* find

$$T(x) = \frac{\ln(Nx)}{1 - \Lambda} + \mathcal{O}(1) \quad , \tag{4.230}$$

which is logarithmic in N. From the mean field dynamics $\dot{\bar{n}} = (\Lambda - 1)\bar{n} - \Lambda \bar{n}^2$, if we are sufficiently close to the SFP at $\bar{n} = 0$, we can neglect the nonlinear term, in which case the solution becomes $\bar{n}(t) = \bar{n}(0) \, e^{(\Lambda - 1)t}$. If we set $\bar{n}(T) \equiv 1$ and $\bar{n}(0) = Nx$, we obtain $T(x) = \ln(Nx)/(1 - \Lambda)$, in agreement with the above expression.

Fokker-Planck solution

Another approach to this problem is to map the Master equation onto a Fokker-Planck equation, as we did in §4.3.5. The corresponding FPE is

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\big(fP\big) + \frac{1}{2N}\frac{\partial^2}{\partial x^2}\big(gP\big) \quad , \tag{4.231}$$

where

$$f(x) = (\Lambda - 1)\,x - \Lambda x^2 = \Lambda\,x\,(x^* - x)$$
$$g(x) = (\Lambda + 1)\,x - \Lambda x^2 = \Lambda\,x\,(x^* + 2\Lambda^{-1} - x) \quad . \tag{4.232}$$

The mean extinction time, from Eqn. 4.63, is

$$T(x) = 2N \int\limits_0^x \frac{dy}{\psi(y)} \int\limits_y^1 dz\, \frac{\psi(z)}{g(z)} \quad , \tag{4.233}$$

where

$$\psi(x) = \exp\left\{ 2N \int\limits_0^x dy\, \frac{f(y)}{g(y)} \right\} \equiv e^{2N\sigma(x)} \tag{4.234}$$

and

$$\sigma(x) = x + 2\Lambda^{-1}\, \ln\!\left(\frac{x^* + 2\Lambda^{-1} - x}{x^* + 2\Lambda^{-1}} \right) \quad . \tag{4.235}$$

Thus,

$$T(x) = \frac{2N}{\Lambda} \int\limits_0^x dy \int\limits_y^1 dz\, \frac{e^{2N\sigma(z)}\, e^{-2N\sigma(y)}}{z(x^* + 2\Lambda^{-1} - z)} \quad . \tag{4.236}$$

The z integral is dominated by $z \approx x^*$, and the y integral by $y \approx 0$. Computing the derivatives for the Taylor series,

$$\sigma(x^*) = \frac{\Lambda - 1}{\Lambda} - \frac{2}{\Lambda}\ln\!\left(\frac{\Lambda + 1}{2} \right) \quad , \quad \sigma'(x^*) = 0 \quad , \quad \sigma''(x^*) = -\tfrac{1}{2}\Lambda \tag{4.237}$$

and also $\sigma(0) = 0$ and $\sigma'(0) = (\Lambda - 1)/(\Lambda + 1)$. One then finds

$$T(x^*) \approx \frac{\Lambda}{(\Lambda - 1)^2} \sqrt{\frac{2\pi}{N\Lambda}}\, e^{2N\sigma(x^*)} \quad . \tag{4.238}$$

Comparison of Master and Fokker-Planck equation predictions for extinction times

How does the FPE result compare with the earlier analysis of the extinction time from the Master equation? If we expand about the threshold value $\Lambda = 1$, writing $\Lambda = 1 + \varepsilon$, we find

$$\Phi(\Lambda) = \ln\Lambda - 1 + \Lambda^{-1} = \tfrac{1}{2}\,\varepsilon^2 - \tfrac{2}{3}\,\varepsilon^3 + \tfrac{3}{4}\,\varepsilon^4 - \tfrac{4}{5}\,\varepsilon^5 + \dots$$

$$2\sigma(x^*) = \frac{2(\Lambda - 1)}{\Lambda} - \frac{4}{\Lambda}\ln\!\left(\frac{\Lambda + 1}{2} \right) = \tfrac{1}{2}\,\varepsilon^2 - \tfrac{2}{3}\,\varepsilon^3 + \tfrac{35}{48}\,\varepsilon^4 - \tfrac{181}{240}\,\varepsilon^5 + \dots \tag{4.239}$$

The difference only begins at fourth order in ε *viz.*

$$\ln T^{\mathrm{ME}}(x^*) - \ln T^{\mathrm{FPE}}(x^*) = N\!\left(\frac{\varepsilon^4}{48} - \frac{11\,\varepsilon^5}{240} + \frac{11\,\varepsilon^6}{160} + \dots \right) + \mathcal{O}(1) \quad , \tag{4.240}$$

where the superscripts indicate Master equation (ME) and Fokker-Planck equation (FPE), respectively. While the term inside the parentheses impressively small when $\varepsilon \ll 1$, it is nevertheless finite, and, critically, *it is multiplied by* N. Thus, the actual mean extinction time, as computed from the original Master equation, is exponentially larger than the Fokker-Planck result.

What are we to learn from this? The origin of the difference lies in the truncations we had to do in order to derive the Fokker-Planck equation itself. The FPE fails to accurately capture the statistics of large deviations from the metastable state. D. Kessler and N. Shnerb, in *J. Stat. Phys.* **127**, 861 (2007), show that the FPE is only valid for fluctuations about the metastable state whose size is $\mathcal{O}(N^{2/3})$, whereas to reach the absorbing state requires a fluctuation of $\mathcal{O}(N)$. As these authors put it, *"In order to get the correct statistics for rare and extreme events one should base the estimate on the exact Master equation that describes the stochastic process. . ."*. They also derive a real space WKB method to extract the correct statistics from the Master equation. Another WKB-like treatment, and one which utilizes the powerful Doi-Peliti field theory formalism, is found in the paper by V. Elgart and A. Kamenev, *Phys. Rev. E* **70**, 041106 (2004).

Chapter 5

The Boltzmann Equation

5.1 References

– H. Smith and H. H. Jensen, *Transport Phenomena* (Oxford, 1989)
An outstanding, thorough, and pellucid presentation of the theory of Boltzmann transport in classical and quantum systems.

– P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge, 2010)
Superb, modern discussion of a broad variety of issues and models in nonequilibrium statistical physics.

– E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics* (Pergamon, 1981)
Volume 10 in the famous Landau and Lifshitz *Course of Theoretical Physics*. Surprisingly readable, and with many applications (some advanced).

– M. Kardar, *Statistical Physics of Particles* (Cambridge, 2007)
A superb modern text, with many insightful presentations of key concepts. Includes a very instructive derivation of the Boltzmann equation starting from the BBGKY hierarchy.

– J. A. McLennan, *Introduction to Non-equilibrium Statistical Mechanics* (Prentice-Hall, 1989)
Though narrow in scope, this book is a good resource on the Boltzmann equation.

– F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, 1987)
This has been perhaps the most popular undergraduate text since it first appeared in 1967, and with good reason. The later chapters discuss transport phenomena at an undergraduate level.

– N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (3$^{\rm rd}$ edition, North-Holland, 2007)
This is a very readable and useful text. A relaxed but meaty presentation.

5.2 Equilibrium, Nonequilibrium and Local Equilibrium

Classical equilibrium statistical mechanics is described by the full N-body distribution,

$$
f^0(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N) =
\begin{cases}
Z_N^{-1} \cdot \frac{1}{N!}\, e^{-\beta \hat{H}_N(\boldsymbol{p},\boldsymbol{x})} & \text{OCE} \\[2ex]
\Xi^{-1} \cdot \frac{1}{N!}\, e^{\beta \mu N} e^{-\beta \hat{H}_N(\boldsymbol{p},\boldsymbol{x})} & \text{GCE} .
\end{cases}
\tag{5.1}
$$

We assume a Hamiltonian of the form

$$
\hat{H}_N = \sum_{i=1}^N \frac{\boldsymbol{p}_i^2}{2m} + \sum_{i=1}^N v(\boldsymbol{x}_i) + \sum_{i<j}^N u(\boldsymbol{x}_i - \boldsymbol{x}_j),
\tag{5.2}
$$

typically with $v = 0$, *i.e.* only two-body interactions. The quantity

$$
f^0(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N;\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N)\, \frac{d^d x_1\, d^d p_1}{h^d} \cdots \frac{d^d x_N\, d^d p_N}{h^d}
\tag{5.3}
$$

is the probability, under equilibrium conditions, of finding N particles in the system, with particle #1 lying within d^3x_1 of $\boldsymbol{x}_1$ and having momentum within $d^d p_1$ of $\boldsymbol{p}_1$, *etc.* The temperature T and chemical potential μ are constants, independent of position. Note that $f(\{\boldsymbol{x}_i\},\{\boldsymbol{p}_i\})$ is dimensionless.

Nonequilibrium statistical mechanics seeks to describe thermodynamic systems which are out of equilibrium, meaning that the distribution function is not given by the Boltzmann distribution above. For a general nonequilibrium setting, it is hopeless to make progress – we'd have to integrate the equations of motion for all the constituent particles. However, typically we are concerned with situations where external forces or constraints are imposed over some macroscopic scale. Examples would include the imposition of a voltage drop across a metal, or a temperature differential across any thermodynamic sample. In such cases, scattering at microscopic length and time scales described by the *mean free path* ℓ and the *collision time* τ work to establish *local equilibrium* throughout the system. A local equilibrium is a state described by a space and time varying temperature $T(\boldsymbol{r},t)$ and chemical potential $\mu(\boldsymbol{r},t)$. As we will see, the Boltzmann distribution with $T = T(\boldsymbol{r},t)$ and $\mu = \mu(\boldsymbol{r},t)$ will not be a solution to the evolution equation governing the distribution function. Rather, the distribution for systems slightly out of equilibrium will be of the form $f = f^0 + \delta f$, where f^0 describes a state of local equilibrium.

We will mainly be interested in the one-body distribution

$$
\begin{aligned}
f(\boldsymbol{r},\boldsymbol{p};t) &= \sum_{i=1}^N \big\langle\, \delta\big(\boldsymbol{x}_i(t) - \boldsymbol{r}\big)\, \delta(\boldsymbol{p}_i(t) - \boldsymbol{p})\, \big\rangle \\
&= N \int \prod_{i=2}^N d^d x_i\, d^d p_i\, f(\boldsymbol{r},\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N;\boldsymbol{p},\boldsymbol{p}_2,\ldots,\boldsymbol{p}_N;t)\,.
\end{aligned}
\tag{5.4}
$$

In this chapter, we will drop the $1/\hbar$ normalization for phase space integration. Thus, $f(\boldsymbol{r},\boldsymbol{p},t)$ has dimensions of h^{-d}, and $f(\boldsymbol{r},\boldsymbol{p},t)\, d^3r\, d^3p$ is the average number of particles found within d^3r of $\boldsymbol{r}$ and d^3p of $\boldsymbol{p}$ at time t.

In the GCE, we sum the RHS above over N. Assuming $v = 0$ so that there is no one-body potential to break translational symmetry, the equilibrium distribution is time-independent and space-independent:

$$f^0(\boldsymbol{r}, \boldsymbol{p}) = n \, (2\pi m k_{\mathrm{B}} T)^{-3/2} \, e^{-\boldsymbol{p}^2/2m k_{\mathrm{B}} T} \, , \tag{5.5}$$

where $n = N/V$ or $n = n(T, \mu)$ is the particle density in the OCE or GCE. From the one-body distribution we can compute things like the particle current, $\boldsymbol{j}$, and the energy current, $\boldsymbol{j}_\varepsilon$:

$$\boldsymbol{j}(\boldsymbol{r}, t) = \int d^d\!p \; f(\boldsymbol{r}, \boldsymbol{p}; t) \, \frac{\boldsymbol{p}}{m} \tag{5.6}$$

$$\boldsymbol{j}_\varepsilon(\boldsymbol{r}, t) = \int d^d\!p \; f(\boldsymbol{r}, \boldsymbol{p}; t) \, \varepsilon(\boldsymbol{p}) \, \frac{\boldsymbol{p}}{m} \, , \tag{5.7}$$

where $\varepsilon(\boldsymbol{p}) = \boldsymbol{p}^2/2m$. Clearly these currents both vanish in equilibrium, when $f = f^0$, since $f^0(\boldsymbol{r}, \boldsymbol{p})$ depends only on $\boldsymbol{p}^2$ and not on the direction of $\boldsymbol{p}$. In a *steady state* nonequilibrium situation, the above quantities are time-independent.

Thermodynamics says that

$$dq = T \, ds = d\varepsilon - \mu \, dn \, , \tag{5.8}$$

where s, ε, and n are entropy density, energy density, and particle density, respectively, and dq is the differential heat density. This relation may be case as one among the corresponding current densities:

$$\boldsymbol{j}_q = T\boldsymbol{j}_s = \boldsymbol{j}_\varepsilon - \mu \, \boldsymbol{j} \, . \tag{5.9}$$

Thus, in a system with no particle flow, $\boldsymbol{j} = 0$ and the *heat current $\boldsymbol{j}_q$* is the same as the energy current $\boldsymbol{j}_\varepsilon$.

When the individual particles are not point particles, they possess angular momentum as well as linear momentum. Following Lifshitz and Pitaevskii, we abbreviate $\Gamma = (\boldsymbol{p}, \boldsymbol{L})$ for these two variables for the case of diatomic molecules, and $\Gamma = (\boldsymbol{p}, \boldsymbol{L}, \hat{\boldsymbol{n}} \cdot \boldsymbol{L})$ in the case of spherical top molecules, where $\hat{\boldsymbol{n}}$ is the symmetry axis of the top. We then have, in $d = 3$ dimensions,

$$d\Gamma = \begin{cases} d^3p & \text{point particles} \\ d^3p \, L \, dL \, d\Omega_L & \text{diatomic molecules} \\ d^3p \, L^2 \, dL \, d\Omega_L \, d\cos\vartheta & \text{symmetric tops} \, , \end{cases} \tag{5.10}$$

where $\vartheta = \cos^{-1}(\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{L}})$. We will call the set Γ the 'kinematic variables'. The instantaneous number density at $\boldsymbol{r}$ is then

$$n(\boldsymbol{r}, t) = \int d\Gamma \; f(\boldsymbol{r}, \Gamma; t) \, . \tag{5.11}$$

One might ask why we do not also keep track of the angular orientation of the individual molecules. There are two reasons. First, the rotations of the molecules are generally extremely rapid, so we are justified in averaging over these motions. Second, the orientation of, say, a rotor does not enter into its energy. While the same can be said of the spatial position in the absence of external fields, (i) in the presence of external fields one must keep track of the position coordinate $\boldsymbol{r}$ since there is physical transport of particles from one region of space to another, and (iii) the collision process, which as we shall see enters the dynamics of the distribution function, takes place in real space.

5.3   Boltzmann Transport Theory

5.3.1   Derivation of the Boltzmann equation

For simplicity of presentation, we assume point particles. Recall that

$$f(\boldsymbol{r}, \boldsymbol{p}, t) \, d^3r \, d^3p \equiv \begin{cases} \text{\# of particles with positions within } d^3r \text{ of} \\ \boldsymbol{r} \text{ and momenta within } d^3p \text{ of } \boldsymbol{p} \text{ at time } t. \end{cases} \tag{5.12}$$

We now ask how the distribution functions $f(\boldsymbol{r}, \boldsymbol{p}, t)$ evolves in time. It is clear that in the absence of collisions, the distribution function must satisfy the continuity equation,

$$\frac{\partial f}{\partial t} + \boldsymbol{\nabla} \cdot (\boldsymbol{u} f) = 0 \,. \tag{5.13}$$

This is just the condition of number conservation for particles. Take care to note that $\boldsymbol{\nabla}$ and $\boldsymbol{u}$ are *six-dimensional phase space* vectors:

$$\boldsymbol{u} = (\, \dot{x} \,,\, \dot{y} \,,\, \dot{z} \,,\, \dot{p}_x \,,\, \dot{p}_y \,,\, \dot{p}_z \,) \tag{5.14}$$

$$\boldsymbol{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}, \frac{\partial}{\partial p_x}, \frac{\partial}{\partial p_y}, \frac{\partial}{\partial p_z} \right) \,. \tag{5.15}$$

The continuity equation describes a distribution in which each constituent particle evolves according to a prescribed dynamics, which for a mechanical system is specified by

$$\frac{d\boldsymbol{r}}{dt} = \frac{\partial H}{\partial \boldsymbol{p}} = \boldsymbol{v}(\boldsymbol{p}) \qquad , \qquad \frac{d\boldsymbol{p}}{dt} = -\frac{\partial H}{\partial \boldsymbol{r}} = \boldsymbol{F}_{\text{ext}} \,, \tag{5.16}$$

where $\boldsymbol{F}$ is an external applied force. Here,

$$H(\boldsymbol{p}, \boldsymbol{r}) = \varepsilon(\boldsymbol{p}) + U_{\text{ext}}(\boldsymbol{r}) \,. \tag{5.17}$$

For example, if the particles are under the influence of gravity, then $U_{\text{ext}}(\boldsymbol{r}) = m\boldsymbol{g} \cdot \boldsymbol{r}$ and $\boldsymbol{F} = -\boldsymbol{\nabla} U_{\text{ext}} = -m\boldsymbol{g}$.

Note that as a consequence of the dynamics, we have $\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0$, *i.e.* phase space flow is *incompressible*, provided that $\varepsilon(\boldsymbol{p})$ is a function of $\boldsymbol{p}$ alone, and not of $\boldsymbol{r}$. Thus, in the absence of collisions, we have

$$\frac{\partial f}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} f = 0 \,. \tag{5.18}$$

The differential operator $D_t \equiv \partial_t + \boldsymbol{u} \cdot \boldsymbol{\nabla}$ is sometimes called the 'convective derivative', because $D_t f$ is the time derivative of f in a comoving frame of reference.

Next we must consider the effect of collisions, which are not accounted for by the semiclassical dynamics. In a collision process, a particle with momentum $\boldsymbol{p}$ and one with momentum $\tilde{\boldsymbol{p}}$ can instantaneously convert into a pair with momenta $\boldsymbol{p}'$ and $\tilde{\boldsymbol{p}}'$, provided total momentum is conserved: $\boldsymbol{p} + \tilde{\boldsymbol{p}} = \boldsymbol{p}' + \tilde{\boldsymbol{p}}'$. This means that $D_t f \neq 0$. Rather, we should write

$$\frac{\partial f}{\partial t} + \dot{\boldsymbol{r}} \cdot \frac{\partial f}{\partial \boldsymbol{r}} + \dot{\boldsymbol{p}} \cdot \frac{\partial f}{\partial \boldsymbol{p}} = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} \tag{5.19}$$

where the right side is known as the *collision integral*. The collision integral is in general a *function* of $\boldsymbol{r}$, $\boldsymbol{p}$, and t and a *functional* of the distribution f.

After a trivial rearrangement of terms, we can write the Boltzmann equation as

$$\frac{\partial f}{\partial t} = \left(\frac{\partial f}{\partial t}\right)_{\text{str}} + \left(\frac{\partial f}{\partial t}\right)_{\text{coll}} , \tag{5.20}$$

where

$$\left(\frac{\partial f}{\partial t}\right)_{\text{str}} \equiv -\dot{\boldsymbol{r}} \cdot \frac{\partial f}{\partial \boldsymbol{r}} - \dot{\boldsymbol{p}} \cdot \frac{\partial f}{\partial \boldsymbol{p}} \tag{5.21}$$

is known as the *streaming term*. Thus, there are two contributions to $\partial f/\partial t$: streaming and collisions.

5.3.2 Collisionless Boltzmann equation

In the absence of collisions, the Boltzmann equation is given by

$$\frac{\partial f}{\partial t} + \frac{\partial \varepsilon}{\partial \boldsymbol{p}} \cdot \frac{\partial f}{\partial \boldsymbol{r}} - \boldsymbol{\nabla} U_{\text{ext}} \cdot \frac{\partial f}{\partial \boldsymbol{p}} = 0 . \tag{5.22}$$

In order to gain some intuition about how the streaming term affects the evolution of the distribution $f(\boldsymbol{r}, \boldsymbol{p}, t)$, consider a case where $\boldsymbol{F}_{\text{ext}} = 0$. We then have

$$\frac{\partial f}{\partial t} + \frac{\boldsymbol{p}}{m} \cdot \frac{\partial f}{\partial \boldsymbol{r}} = 0 . \tag{5.23}$$

Clearly, then, *any* function of the form

$$f(\boldsymbol{r}, \boldsymbol{p}, t) = \varphi\big(\boldsymbol{r} - \boldsymbol{v}(\boldsymbol{p})\, t\, ,\, \boldsymbol{p}\big) \tag{5.24}$$

will be a solution to the collisionless Boltzmann equation, where $\boldsymbol{v}(\boldsymbol{p}) = \frac{\partial \varepsilon}{\partial \boldsymbol{p}}$. One possible solution would be the Boltzmann distribution,

$$f(\boldsymbol{r}, \boldsymbol{p}, t) = e^{\mu/k_{\text{B}}T} e^{-\boldsymbol{p}^2/2mk_{\text{B}}T} , \tag{5.25}$$

which is time-independent[1]. Here we have assumed a ballistic dispersion, $\varepsilon(\boldsymbol{p}) = \boldsymbol{p}^2/2m$.

For a slightly less trivial example, let the initial distribution be $\varphi(\boldsymbol{r}, \boldsymbol{p}) = A\, e^{-\boldsymbol{r}^2/2\sigma^2} e^{-\boldsymbol{p}^2/2\kappa^2}$, so that

$$f(\boldsymbol{r}, \boldsymbol{p}, t) = A\, e^{-\left(\boldsymbol{r} - \frac{\boldsymbol{p}t}{m}\right)^2/2\sigma^2}\, e^{-\boldsymbol{p}^2/2\kappa^2} . \tag{5.26}$$

Consider the one-dimensional version, and rescale position, momentum, and time so that

$$f(x, p, t) = A\, e^{-\frac{1}{2}(\bar{x} - \bar{p}\,\bar{t})^2}\, e^{-\frac{1}{2}\bar{p}^2} . \tag{5.27}$$

Consider the level sets of f, where $f(x, p, t) = A\, e^{-\frac{1}{2}\alpha^2}$. The equation for these sets is

$$\bar{x} = \bar{p}\,\bar{t} \pm \sqrt{\alpha^2 - \bar{p}^2} . \tag{5.28}$$

---

[1]Indeed, any arbitrary function of $\boldsymbol{p}$ alone would be a solution. Ultimately, we require some energy exchanging processes, such as collisions, in order for any initial nonequilibrium distribution to converge to the Boltzmann distribution.
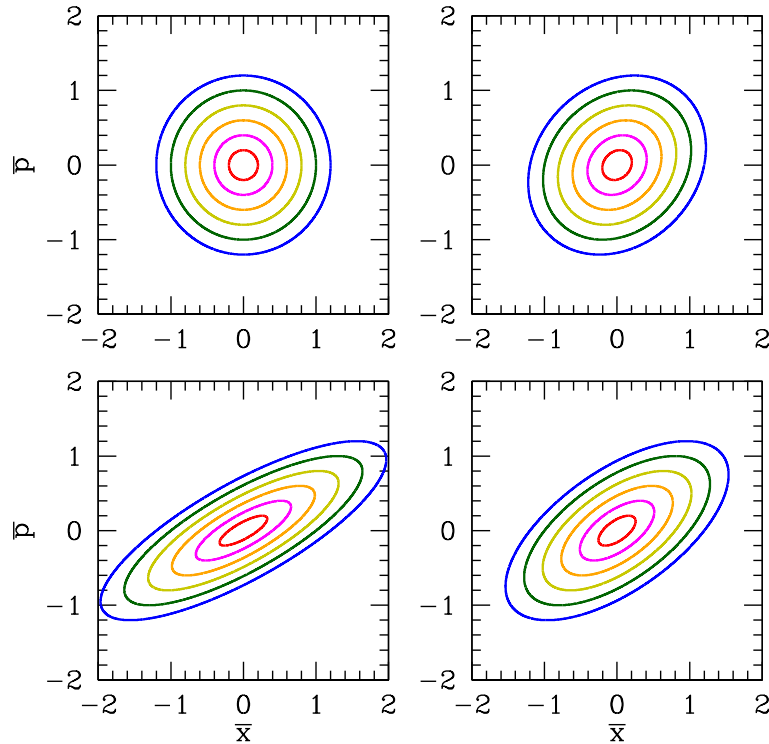
Figure 5.1: Level sets for a sample $f(\bar{x}, \bar{p}, \bar{t}) = A\,e^{-\frac{1}{2}(\bar{x}-\bar{p}\bar{t})^2}e^{-\frac{1}{2}\bar{p}^2}$, for values $f = A\,e^{-\frac{1}{2}\alpha^2}$ with α in equally spaced intervals from $\alpha = 0.2$ (red) to $\alpha = 1.2$ (blue). The time variable $\bar{t}$ is taken to be $\bar{t} = 0.0$ (upper left), 0.2 (upper right), 0.8 (lower right), and 1.3 (lower left).

For fixed $\bar{t}$, these level sets describe the loci in phase space of equal probability densities, with the probability density decreasing exponentially in the parameter α^2. For $\bar{t} = 0$, the initial distribution describes a Gaussian cloud of particles with a Gaussian momentum distribution. As $\bar{t}$ increases, the distribution widens in $\bar{x}$ but not in $\bar{p}$ – each particle moves with a constant momentum, so the set of momentum values never changes. However, the level sets in the $(\bar{x}, \bar{p})$ plane become elliptical, with a semimajor axis oriented at an angle $\theta = \mathrm{ctn}^{-1}(t)$ with respect to the $\bar{x}$ axis. For $\bar{t} > 0$, he particles at the outer edges of the cloud are more likely to be moving away from the center. See the sketches in fig. 5.1

Suppose we add in a constant external force $\boldsymbol{F}_{\mathrm{ext}}$. Then it is easy to show (and left as an exercise to the reader to prove) that any function of the form

$$f(\boldsymbol{r}, \boldsymbol{p}, t) = A\,\varphi\left(\boldsymbol{r} - \frac{\boldsymbol{p}\,t}{m} + \frac{\boldsymbol{F}_{\mathrm{ext}}t^2}{2m}\,,\; \boldsymbol{p} - \frac{\boldsymbol{F}_{\mathrm{ext}}t}{m}\right) \tag{5.29}$$

satisfies the collisionless Boltzmann equation (ballistic dispersion assumed).

5.3.3 Collisional invariants

Consider a function $A(\boldsymbol{r}, \boldsymbol{p})$ of position and momentum. Its average value at time t is

$$A(t) = \int d^3r \, d^3p \, A(\boldsymbol{r}, \boldsymbol{p}) \, f(\boldsymbol{r}, \boldsymbol{p}, t) \,. \tag{5.30}$$

Taking the time derivative,

$$
\begin{aligned}
\frac{dA}{dt} &= \int d^3r \, d^3p \, A(\boldsymbol{r}, \boldsymbol{p}) \, \frac{\partial f}{\partial t} \\
&= \int d^3r \, d^3p \, A(\boldsymbol{r}, \boldsymbol{p}) \left\{ -\frac{\partial}{\partial \boldsymbol{r}} \cdot (\dot{\boldsymbol{r}} f) - \frac{\partial}{\partial \boldsymbol{p}} \cdot (\dot{\boldsymbol{p}} f) + \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} \right\} \\
&= \int d^3r \, d^3p \left\{ \left(\frac{\partial A}{\partial \boldsymbol{r}} \cdot \frac{d\boldsymbol{r}}{dt} + \frac{\partial A}{\partial \boldsymbol{p}} \cdot \frac{d\boldsymbol{p}}{dt} \right) f + A(\boldsymbol{r}, \boldsymbol{p}) \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} \right\} \,.
\end{aligned}
\tag{5.31}
$$

Hence, if A is preserved by the dynamics between collisions, then[2]

$$\frac{dA}{dt} = \frac{\partial A}{\partial \boldsymbol{r}} \cdot \frac{d\boldsymbol{r}}{dt} + \frac{\partial A}{\partial \boldsymbol{p}} \cdot \frac{d\boldsymbol{p}}{dt} = 0 \,. \tag{5.32}$$

We therefore have that the rate of change of A is determined wholly by the collision integral

$$\frac{dA}{dt} = \int d^3r \, d^3p \, A(\boldsymbol{r}, \boldsymbol{p}) \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} \,. \tag{5.33}$$

Quantities which are then conserved in the collisions satisfy $\dot{A} = 0$. Such quantities are called *collisional invariants*. Examples of collisional invariants include the particle number ($A = 1$), the components of the total momentum ($A = p_\mu$) (in the absence of broken translational invariance, due *e.g.* to the presence of walls), and the total energy ($A = \varepsilon(\boldsymbol{p})$).

5.3.4 Scattering processes

What sort of processes contribute to the collision integral? There are two broad classes to consider. The first involves potential scattering, where a particle in state $|\Gamma\rangle$ scatters, in the presence of an external potential, to a state $|\Gamma'\rangle$. Recall that Γ is an abbreviation for the set of kinematic variables, *e.g.* $\Gamma = (\boldsymbol{p}, \boldsymbol{L})$ in the case of a diatomic molecule. For point particles, $\Gamma = (p_x, p_y, p_z)$ and $d\Gamma = d^3p$.

We now define the function $w(\Gamma'|\Gamma)$ such that

$$w(\Gamma'|\Gamma) \, f(\boldsymbol{r}, \Gamma; t) \, d\Gamma \, d\Gamma' = \begin{cases} \text{rate at which a particle within } d\Gamma \text{ of } (\boldsymbol{r}, \Gamma) \\ \text{scatters to within } d\Gamma' \text{ of } (\boldsymbol{r}, \Gamma') \text{ at time } t. \end{cases} \tag{5.34}$$

---

[2]Recall from classical mechanics the definition of the *Poisson bracket*, $\{A, B\} = \frac{\partial A}{\partial \boldsymbol{r}} \cdot \frac{\partial B}{\partial \boldsymbol{p}} - \frac{\partial B}{\partial \boldsymbol{r}} \cdot \frac{\partial A}{\partial \boldsymbol{p}}$. Then from Hamilton's equations $\dot{\boldsymbol{r}} = \frac{\partial H}{\partial \boldsymbol{p}}$ and $\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{r}}$, where $H(\boldsymbol{p}, \boldsymbol{r}, t)$ is the Hamiltonian, we have $\frac{dA}{dt} = \{A, H\}$. Invariants have zero Poisson bracket with the Hamiltonian.
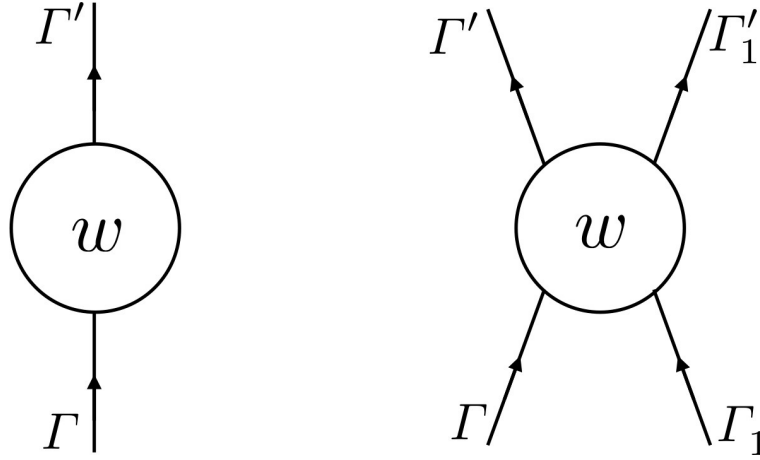
Figure 5.2: Left: single particle scattering process $|\Gamma\rangle \to |\Gamma'\rangle$. Right: two-particle scattering process $|\Gamma\Gamma_1\rangle \to |\Gamma'\Gamma_1'\rangle$.

The units of $w\,d\Gamma$ are therefore $1/T$. The differential scattering cross section for particle scattering is then

$$d\sigma = \frac{w(\Gamma'|\Gamma)}{n\,|\boldsymbol{v}|}\,d\Gamma' \, , \tag{5.35}$$

where $\boldsymbol{v} = \boldsymbol{p}/m$ is the particle's velocity and n the density.

The second class is that of two-particle scattering processes, *i.e.* $|\Gamma\Gamma_1\rangle \to |\Gamma'\Gamma_1'\rangle$. We define the scattering function $w(\Gamma'\Gamma_1' \,|\, \Gamma\Gamma_1)$ by

$$w(\Gamma'\Gamma_1' \,|\, \Gamma\Gamma_1)\, f_2(\boldsymbol{r},\Gamma; \boldsymbol{r}, \Gamma_1\,;t)\,d\Gamma\,d\Gamma_1\,d\Gamma'\,d\Gamma_1' = \begin{cases} \text{rate at which two particles within } d\Gamma \text{ of } (\boldsymbol{r},\Gamma) \\ \text{and within } d\Gamma_1 \text{ of } (\boldsymbol{r},\Gamma_1) \text{ scatter into states within} \\ d\Gamma' \text{ of } (\boldsymbol{r},\Gamma') \text{ and } d\Gamma_1' \text{ of } (\boldsymbol{r},\Gamma_1') \text{ at time } t\,, \end{cases} \tag{5.36}$$

where

$$f_2(\boldsymbol{r},\boldsymbol{p}\,;\,\boldsymbol{r}',\boldsymbol{p}'\,;\,t) = \Big\langle \sum_{i,j} \delta\big(\boldsymbol{x}_i(t) - \boldsymbol{r}\big)\,\delta\big(\boldsymbol{p}_i(t) - \boldsymbol{p}\big)\,\delta\big(\boldsymbol{x}_j(t) - \boldsymbol{r}'\big)\,\delta\big(\boldsymbol{p}_j(t) - \boldsymbol{p}'\big) \Big\rangle \tag{5.37}$$

is the nonequilibrium two-particle distribution for point particles. The differential scattering cross section is

$$d\sigma = \frac{w(\Gamma'\Gamma_1' \,|\, \Gamma\Gamma_1)}{|\boldsymbol{v} - \boldsymbol{v}_1|}\,d\Gamma'\,d\Gamma_1' \, . \tag{5.38}$$

We assume, in both cases, that any scattering occurs *locally*, *i.e.* the particles attain their asymptotic kinematic states on distance scales small compared to the mean interparticle separation. In this case we can treat each scattering process independently. This assumption is particular to rarefied systems, *i.e.* gases, and is not appropriate for dense liquids. The two types of scattering processes are depicted in fig. 5.2.

In computing the collision integral for the state $|\boldsymbol{r},\Gamma\rangle$, we must take care to sum over contributions from transitions *out of* this state, *i.e.* $|\Gamma\rangle \to |\Gamma'\rangle$, which reduce $f(\boldsymbol{r},\Gamma)$, and transitions *into* this state, *i.e.*

$|\Gamma'\rangle \to |\Gamma\rangle$, which increase $f(\boldsymbol{r}, \Gamma)$. Thus, for one-body scattering, we have

$$\frac{D}{Dt} f(\boldsymbol{r}, \Gamma; t) = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \int d\Gamma' \left\{ w(\Gamma \,|\, \Gamma') \, f(\boldsymbol{r}, \Gamma'; t) - w(\Gamma' \,|\, \Gamma) \, f(\boldsymbol{r}, \Gamma; t) \right\} . \tag{5.39}$$

For two-body scattering, we have

$$\begin{aligned}
\frac{D}{Dt} f(\boldsymbol{r}, \Gamma; t) &= \left(\frac{\partial f}{\partial t}\right)_{\text{coll}} \\
&= \int d\Gamma_1 \!\int d\Gamma' \!\int d\Gamma_1' \left\{ w(\Gamma\Gamma_1 \,|\, \Gamma'\Gamma_1') \, f_2(\boldsymbol{r}, \Gamma'; \boldsymbol{r}, \Gamma_1'; t) \right. \\
&\qquad\qquad\qquad\qquad\qquad \left. - w(\Gamma'\Gamma_1' \,|\, \Gamma\Gamma_1) \, f_2(\boldsymbol{r}, \Gamma; \boldsymbol{r}, \Gamma_1; t) \right\} .
\end{aligned} \tag{5.40}$$

Unlike the one-body scattering case, the kinetic equation for two-body scattering does not close, since the LHS involves the one-body distribution $f \equiv f_1$ and the RHS involves the two-body distribution f_2. To close the equations, we make the *approximation*

$$f_2(\boldsymbol{r}, \Gamma'; \tilde{\boldsymbol{r}}, \tilde{\Gamma}; t) \approx f(\boldsymbol{r}, \Gamma; t) \, f(\tilde{\boldsymbol{r}}, \tilde{\Gamma}; t) . \tag{5.41}$$

We then have

$$\begin{aligned}
\frac{D}{Dt} f(\boldsymbol{r}, \Gamma; t) &= \int d\Gamma_1 \!\int d\Gamma' \!\int d\Gamma_1' \left\{ w(\Gamma\Gamma_1 \,|\, \Gamma'\Gamma_1') \, f(\boldsymbol{r}, \Gamma'; t) \, f(\boldsymbol{r}, \Gamma_1'; t) \right. \\
&\qquad\qquad\qquad\qquad\qquad \left. - w(\Gamma'\Gamma_1' \,|\, \Gamma\Gamma_1) \, f(\boldsymbol{r}, \Gamma; t) \, f(\boldsymbol{r}, \Gamma_1; t) \right\} .
\end{aligned} \tag{5.42}$$

5.3.5 Detailed balance

Classical mechanics places some restrictions on the form of the kernel $w(\Gamma\Gamma_1 \,|\, \Gamma'\Gamma_1')$. In particular, if $\Gamma^T = (-\boldsymbol{p}, -\boldsymbol{L})$ denotes the kinematic variables under time reversal, then

$$w(\Gamma'\Gamma_1' \,|\, \Gamma\Gamma_1) = w(\Gamma^T \Gamma_1^T \,|\, \Gamma'^T \Gamma_1'^T) . \tag{5.43}$$

This is because the time reverse of the process $|\Gamma\Gamma_1\rangle \to |\Gamma'\Gamma_1'\rangle$ is $|\Gamma'^{\mathrm{T}} \Gamma_1'^{\mathrm{T}}\rangle \to |\Gamma^{\mathrm{T}} \Gamma_1^{\mathrm{T}}\rangle$.

In equilibrium, we must have

$$w(\Gamma'\Gamma_1' \,|\, \Gamma\Gamma_1) \, f^0(\Gamma) \, f^0(\Gamma_1) \, d^4\Gamma = w(\Gamma^T \Gamma_1^T \,|\, \Gamma'^T \Gamma_1'^T) \, f^0(\Gamma'^T) \, f^0(\Gamma_1'^T) \, d^4\Gamma^T \tag{5.44}$$

where

$$d^4\Gamma \equiv d\Gamma \, d\Gamma_1 \, d\Gamma' d\Gamma_1' \quad , \quad d^4\Gamma^T \equiv d\Gamma^T \, d\Gamma_1^T \, d\Gamma'^T d\Gamma_1'^T . \tag{5.45}$$

Since $d\Gamma = d\Gamma^T$ *etc.*, we may cancel the differentials above, and after invoking eqn. 5.43 and suppressing the common $\boldsymbol{r}$ label, we find

$$f^0(\Gamma) \, f^0(\Gamma_1) = f^0(\Gamma'^T) \, f^0(\Gamma_1'^T) . \tag{5.46}$$

This is the condition of *detailed balance*. For the Boltzmann distribution, we have

$$f^0(\Gamma) = A \, e^{-\varepsilon/k_{\mathrm{B}} T} , \tag{5.47}$$

where A is a constant and where $\varepsilon = \varepsilon(\Gamma)$ is the kinetic energy, *e.g.* $\varepsilon(\Gamma) = \boldsymbol{p}^2/2m$ in the case of point particles. Note that $\varepsilon(\Gamma^T) = \varepsilon(\Gamma)$. Detailed balance is satisfied because the kinematics of the collision requires energy conservation:

$$\varepsilon + \varepsilon_1 = \varepsilon' + \varepsilon_1' \ . \tag{5.48}$$

Since momentum is also kinematically conserved, *i.e.*

$$\boldsymbol{p} + \boldsymbol{p}_1 = \boldsymbol{p}' + \boldsymbol{p}_1' \ , \tag{5.49}$$

any distribution of the form

$$f^0(\Gamma) = A\,e^{-(\varepsilon - \boldsymbol{p} \cdot \boldsymbol{V})/k_{\mathrm{B}}T} \tag{5.50}$$

also satisfies detailed balance, for any velocity parameter $\boldsymbol{V}$. This distribution is appropriate for gases which are flowing with average particle $\boldsymbol{V}$.

In addition to time-reversal, parity is also a symmetry of the microscopic mechanical laws. Under the parity operation P, we have $\boldsymbol{r} \to -\boldsymbol{r}$ and $\boldsymbol{p} \to -\boldsymbol{p}$. Note that a pseudovector such as $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ is unchanged under P. Thus, $\Gamma^P = (-\boldsymbol{p}, \boldsymbol{L})$. Under the combined operation of $C = PT$, we have $\Gamma^C = (\boldsymbol{p}, -\boldsymbol{L})$. If the microscopic Hamiltonian is invariant under C, then we must have

$$w\big(\Gamma' \Gamma_1' \,|\, \Gamma \Gamma_1\big) = w\big(\Gamma^C \Gamma_1^C \,|\, \Gamma'^C \Gamma_1'^C\big) \ . \tag{5.51}$$

For point particles, invariance under T and P then means

$$w(\boldsymbol{p}', \boldsymbol{p}_1' \,|\, \boldsymbol{p}, \boldsymbol{p}_1) = w(\boldsymbol{p}, \boldsymbol{p}_1 \,|\, \boldsymbol{p}', \boldsymbol{p}_1') \ , \tag{5.52}$$

and therefore the collision integral takes the simplified form,

$$\begin{aligned}
\frac{Df(\boldsymbol{p})}{Dt} &= \left(\frac{\partial f}{\partial t}\right)_{\mathrm{coll}} \\
&= \int d^3p_1 \int d^3p' \int d^3p_1'\, w(\boldsymbol{p}', \boldsymbol{p}_1' \,|\, \boldsymbol{p}, \boldsymbol{p}_1) \left\{ f(\boldsymbol{p}')\, f(\boldsymbol{p}_1') - f(\boldsymbol{p})\, f(\boldsymbol{p}_1) \right\} \ ,
\end{aligned} \tag{5.53}$$

where we have suppressed both $\boldsymbol{r}$ and t variables.

The most general statement of detailed balance is

$$\frac{f^0(\Gamma')\, f^0(\Gamma_1')}{f^0(\Gamma)\, f^0(\Gamma_1)} = \frac{w\big(\Gamma' \Gamma_1' \,|\, \Gamma \Gamma_1\big)}{w\big(\Gamma \Gamma_1 \,|\, \Gamma' \Gamma_1'\big)} \ . \tag{5.54}$$

Under this condition, the collision term vanishes for $f = f^0$, which is the equilibrium distribution.

5.3.6 Kinematics and cross section

We can rewrite eqn. 5.53 in the form

$$\frac{Df(\boldsymbol{p})}{Dt} = \int d^3p_1 \int d\Omega \,|\boldsymbol{v} - \boldsymbol{v}_1|\, \frac{\partial \sigma}{\partial \Omega} \left\{ f(\boldsymbol{p}')\, f(\boldsymbol{p}_1') - f(\boldsymbol{p})\, f(\boldsymbol{p}_1) \right\} \ , \tag{5.55}$$

where $\frac{\partial \sigma}{\partial \Omega}$ is the *differential scattering cross section*. If we recast the scattering problem in terms of center-of-mass and relative coordinates, we conclude that the total momentum is conserved by the collision, and furthermore that the energy in the CM frame is conserved, which means that the magnitude of the *relative* momentum is conserved. Thus, we may write $\boldsymbol{p}' - \boldsymbol{p}'_1 = |\boldsymbol{p} - \boldsymbol{p}_1|\,\hat{\boldsymbol{\Omega}}$, where $\hat{\boldsymbol{\Omega}}$ is a unit vector. Then $\boldsymbol{p}'$ and $\boldsymbol{p}'_1$ are determined to be

$$
\begin{aligned}
\boldsymbol{p}' &= \tfrac{1}{2}\left(\boldsymbol{p} + \boldsymbol{p}_1 + |\boldsymbol{p} - \boldsymbol{p}_1|\,\hat{\boldsymbol{\Omega}}\right) \\
\boldsymbol{p}'_1 &= \tfrac{1}{2}\left(\boldsymbol{p} + \boldsymbol{p}_1 - |\boldsymbol{p} - \boldsymbol{p}_1|\,\hat{\boldsymbol{\Omega}}\right).
\end{aligned}
\tag{5.56}
$$

5.3.7 H-theorem

Let's consider the Boltzmann equation with two particle collisions. We define the local (*i.e. $\boldsymbol{r}$-dependent*) quantity

$$
\rho_\varphi(\boldsymbol{r}, t) \equiv \int d\Gamma\,\varphi(\Gamma, f)\,f(\Gamma, \boldsymbol{r}, t).
\tag{5.57}
$$

At this point, $\varphi(\Gamma, f)$ is arbitrary. Note that the $\varphi(\Gamma, f)$ factor has $\boldsymbol{r}$ and t dependence through its dependence on f, which itself is a function of $\boldsymbol{r}$, Γ, and t. We now compute

$$
\begin{aligned}
\frac{\partial \rho_\varphi}{\partial t} &= \int d\Gamma\,\frac{\partial(\varphi f)}{\partial t} = \int d\Gamma\,\frac{\partial(\varphi f)}{\partial f}\,\frac{\partial f}{\partial t} \\
&= -\int d\Gamma\,\boldsymbol{u}\cdot\boldsymbol{\nabla}(\varphi f) - \int d\Gamma\,\frac{\partial(\varphi f)}{\partial f}\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} \\
&= -\oint d\Sigma\,\hat{\boldsymbol{n}}\cdot(\boldsymbol{u}\,\varphi f) - \int d\Gamma\,\frac{\partial(\varphi f)}{\partial f}\left(\frac{\partial f}{\partial t}\right)_{\text{coll}}.
\end{aligned}
\tag{5.58}
$$

The first term on the last line follows from the divergence theorem, and vanishes if we assume $f = 0$ for infinite values of the kinematic variables, which is the only physical possibility. Thus, the rate of change of ρ_φ is entirely due to the collision term. Thus,

$$
\begin{aligned}
\frac{\partial \rho_\varphi}{\partial t} &= \int d\Gamma\!\int d\Gamma_1\!\int d\Gamma'\!\int d\Gamma'_1\,\left\{w(\Gamma'\Gamma'_1\,|\,\Gamma\Gamma_1)\,ff_1\,\chi - w(\Gamma\Gamma_1\,|\,\Gamma'\Gamma'_1)\,f'f'_1\,\chi\right\} \\
&= \int d\Gamma\!\int d\Gamma_1\!\int d\Gamma'\!\int d\Gamma'_1\,w(\Gamma'\Gamma'_1\,|\,\Gamma\Gamma_1)\,ff_1\,(\chi - \chi'),
\end{aligned}
\tag{5.59}
$$

where $f \equiv f(\Gamma)$, $f' \equiv f(\Gamma')$, $f_1 \equiv f(\Gamma_1)$, $f'_1 \equiv f(\Gamma'_1)$, $\chi = \chi(\Gamma)$, with

$$
\chi = \frac{\partial(\varphi f)}{\partial f} = \varphi + f\,\frac{\partial \varphi}{\partial f}.
\tag{5.60}
$$

We now invoke the symmetry

$$
w(\Gamma'\Gamma'_1\,|\,\Gamma\Gamma_1) = w(\Gamma'_1\Gamma'\,|\,\Gamma_1\Gamma),
\tag{5.61}
$$

which allows us to write

$$
\frac{\partial \rho_\varphi}{\partial t} = \tfrac{1}{2}\int d\Gamma\!\int d\Gamma_1\!\int d\Gamma'\!\int d\Gamma'_1\,w(\Gamma'\Gamma'_1\,|\,\Gamma\Gamma_1)\,ff_1\,(\chi + \chi_1 - \chi' - \chi'_1).
\tag{5.62}
$$

This shows that ρ_φ is preserved by the collision term if $\chi(\Gamma)$ is a collisional invariant.

Now let us consider $\varphi(f) = \ln f$. We define $\mathsf{h} \equiv \rho\big|_{\varphi = \ln f}$. We then have

$$\frac{\partial \mathsf{h}}{\partial t} = -\tfrac{1}{2} \int d\Gamma \int d\Gamma_1 \int d\Gamma' \int d\Gamma'_1 \, w \, f' f'_1 \cdot x \ln x \,, \tag{5.63}$$

where $w \equiv w\big(\Gamma'\Gamma'_1 \,|\, \Gamma\Gamma_1\big)$ and $x \equiv f f_1 / f' f'_1$. We next invoke the result

$$\int d\Gamma' \int d\Gamma'_1 \, w\big(\Gamma'\Gamma'_1 \,|\, \Gamma\Gamma_1\big) = \int d\Gamma' \int d\Gamma'_1 \, w\big(\Gamma\Gamma_1 \,|\, \Gamma'\Gamma'_1\big) \tag{5.64}$$

which is a statement of unitarity of the scattering matrix[3]. Multiplying both sides by $f(\Gamma) \, f(\Gamma_1)$, then integrating over Γ and Γ_1, and finally changing variables $(\Gamma, \Gamma_1) \leftrightarrow (\Gamma', \Gamma'_1)$, we find

$$0 = \int d\Gamma \int d\Gamma_1 \int d\Gamma' \int d\Gamma'_1 \, w \, (f f_1 - f' f'_1) = \int d\Gamma \int d\Gamma_1 \int d\Gamma' \int d\Gamma'_1 \, w \, f' f'_1 \, (x - 1) \,. \tag{5.65}$$

Multiplying this result by $\tfrac{1}{2}$ and adding it to the previous equation for $\dot{\mathsf{h}}$, we arrive at our final result,

$$\frac{\partial \mathsf{h}}{\partial t} = -\tfrac{1}{2} \int d\Gamma \int d\Gamma_1 \int d\Gamma' \int d\Gamma'_1 \, w \, f' f'_1 \, (x \ln x - x + 1) \,. \tag{5.66}$$

Note that w, f', and f'_1 are all nonnegative. It is then easy to prove that the function $g(x) = x \ln x - x + 1$ is nonnegative for all positive x values[4], which therefore entails the important result

$$\frac{\partial \mathsf{h}(\boldsymbol{r}, t)}{\partial t} \leq 0 \,. \tag{5.67}$$

Boltzmann's H function is the space integral of the h density: $\mathsf{H} = \int d^3 r \, \mathsf{h}$.

Thus, everywhere in space, the function $\mathsf{h}(\boldsymbol{r}, t)$ is monotonically decreasing or constant, due to collisions. In equilibrium, $\dot{\mathsf{h}} = 0$ everywhere, which requires $x = 1$, *i.e.*

$$f^0(\Gamma) \, f^0(\Gamma_1) = f^0(\Gamma') \, f^0(\Gamma'_1) \,, \tag{5.68}$$

or, taking the logarithm,

$$\ln f^0(\Gamma) + \ln f^0(\Gamma_1) = \ln f^0(\Gamma') + \ln f^0(\Gamma'_1) \,. \tag{5.69}$$

But this means that $\ln f^0$ is itself a collisional invariant, and if 1, $\boldsymbol{p}$, and ε are the only collisional invariants, then $\ln f^0$ must be expressible in terms of them. Thus,

$$\ln f^0 = \frac{\mu}{k_{\mathrm{B}} T} + \frac{\boldsymbol{V} \cdot \boldsymbol{p}}{k_{\mathrm{B}} T} - \frac{\varepsilon}{k_{\mathrm{B}} T} \,, \tag{5.70}$$

where μ, $\boldsymbol{V}$, and T are constants which parameterize the equilibrium distribution $f^0(\boldsymbol{p})$, corresponding to the chemical potential, flow velocity, and temperature, respectively.

---

[3]See Lifshitz and Pitaevskii, *Physical Kinetics*, §2.

[4]The function $g(x) = x \ln x - x + 1$ satisfies $g'(x) = \ln x$, hence $g'(x) < 0$ on the interval $x \in [0, 1)$ and $g'(x) > 0$ on $x \in (1, \infty]$. Thus, $g(x)$ monotonically decreases from $g(0) = 1$ to $g(1) = 0$, and then monotonically increases to $g(\infty) = \infty$, never becoming negative.

5.4 Weakly Inhomogeneous Gas

Consider a gas which is only weakly out of equilibrium. We follow the treatment in Lifshitz and Pitaevskii, §6. As the gas is only slightly out of equilibrium, we seek a solution to the Boltzmann equation of the form $f = f^0 + \delta f$, where f^0 is describes a local equilibrium. Recall that such a distribution function is annihilated by the collision term in the Boltzmann equation but not by the streaming term, hence a correction δf must be added in order to obtain a solution.

The most general form of local equilibrium is described by the distribution

$$f^0(\boldsymbol{r}, \Gamma) = C \exp\left(\frac{\mu - \varepsilon(\Gamma) + \boldsymbol{V} \cdot \boldsymbol{p}}{k_{\mathrm{B}} T}\right), \tag{5.71}$$

where $\mu = \mu(\boldsymbol{r}, t)$, $T = T(\boldsymbol{r}, t)$, and $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{r}, t)$ vary in both space and time. Note that

$$
\begin{aligned}
df^0 &= \left(d\mu + \boldsymbol{p} \cdot d\boldsymbol{V} + (\varepsilon - \mu - \boldsymbol{V} \cdot \boldsymbol{p})\frac{dT}{T} - d\varepsilon\right)\left(-\frac{\partial f^0}{\partial \varepsilon}\right) \\
&= \left(\frac{1}{n}\,dp + \boldsymbol{p} \cdot d\boldsymbol{V} + (\varepsilon - h)\frac{dT}{T} - d\varepsilon\right)\left(-\frac{\partial f^0}{\partial \varepsilon}\right)
\end{aligned}
\tag{5.72}
$$

where we have assumed $\boldsymbol{V} = 0$ on average, and used

$$
\begin{aligned}
d\mu &= \left(\frac{\partial \mu}{\partial T}\right)_p dT + \left(\frac{\partial \mu}{\partial p}\right)_T dp \\
&= -s\,dT + \frac{1}{n}\,dp,
\end{aligned}
\tag{5.73}
$$

where s is the entropy per particle and n is the number density. We have further written $h = \mu + Ts$, which is the enthalpy per particle. Here, c_p is the heat capacity per particle at constant pressure[5]. Finally, note that when f^0 is the Maxwell-Boltzmann distribution, we have

$$-\frac{\partial f^0}{\partial \varepsilon} = \frac{f^0}{k_{\mathrm{B}} T}. \tag{5.74}$$

The Boltzmann equation is written

$$\left(\frac{\partial}{\partial t} + \frac{\boldsymbol{p}}{m} \cdot \frac{\partial}{\partial \boldsymbol{r}} + \boldsymbol{F} \cdot \frac{\partial}{\partial \boldsymbol{p}}\right)(f^0 + \delta f) = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{coll}}. \tag{5.75}$$

The RHS of this equation must be of order δf because the local equilibrium distribution f^0 is annihilated by the collision integral. We therefore wish to evaluate one of the contributions to the LHS of this equation,

$$
\begin{aligned}
\frac{\partial f^0}{\partial t} + \frac{\boldsymbol{p}}{m} \cdot \frac{\partial f^0}{\partial \boldsymbol{r}} + \boldsymbol{F} \cdot \frac{\partial f^0}{\partial \boldsymbol{p}} = {}& \left(-\frac{\partial f^0}{\partial \varepsilon}\right)\left\{\frac{1}{n}\frac{\partial p}{\partial t} + \frac{\varepsilon - h}{T}\frac{\partial T}{\partial t} + m\boldsymbol{v} \cdot \left[(\boldsymbol{v} \cdot \boldsymbol{\nabla})\,\boldsymbol{V}\right]\right. \\
&\left. + \boldsymbol{v} \cdot \left(m\frac{\partial \boldsymbol{V}}{\partial t} + \frac{1}{n}\boldsymbol{\nabla}p\right) + \frac{\varepsilon - h}{T}\,\boldsymbol{v} \cdot \boldsymbol{\nabla}T - \boldsymbol{F} \cdot \boldsymbol{v}\right\}.
\end{aligned}
\tag{5.76}
$$

---

[5]In the chapter on thermodynamics, we adopted a slightly different definition of c_p as the heat capacity per mole. In this chapter c_p is the heat capacity per particle.

To simplify this, first note that Newton's laws applied to an ideal fluid give $\rho \dot{V} = -\nabla p$, where $\rho = mn$ is the mass density. Corrections to this result, *e.g.* viscosity and nonlinearity in V, are of higher order.

Next, continuity for particle number means $\dot{n} + \nabla \cdot (nV) = 0$. We assume V is zero on average and that all derivatives are small, hence $\nabla \cdot (nV) = V \cdot \nabla n + n \nabla \cdot V \approx n \nabla \cdot V$. Thus,

$$\frac{\partial \ln n}{\partial t} = \frac{\partial \ln p}{\partial t} - \frac{\partial \ln T}{\partial t} = -\nabla \cdot V \,, \tag{5.77}$$

where we have invoked the ideal gas law $n = p/k_\mathrm{B} T$ above.

Next, we invoke conservation of entropy. If s is the entropy per particle, then ns is the entropy per unit volume, in which case we have the continuity equation

$$\frac{\partial (ns)}{\partial t} + \nabla \cdot (nsV) = n\left(\frac{\partial s}{\partial t} + V \cdot \nabla s\right) + s\left(\frac{\partial n}{\partial t} + \nabla \cdot (nV)\right) = 0 \,. \tag{5.78}$$

The second bracketed term on the RHS vanishes because of particle continuity, leaving us with $\dot{s} + V \cdot \nabla s \approx \dot{s} = 0$ (since $V = 0$ on average, and any gradient is first order in smallness). Now thermodynamics says

$$\begin{aligned}
ds &= \left(\frac{\partial s}{\partial T}\right)_p dT + \left(\frac{\partial s}{\partial p}\right)_T dp \\
&= \frac{c_p}{T} \, dT - \frac{k_\mathrm{B}}{p} \, dp \,,
\end{aligned} \tag{5.79}$$

since $T\left(\frac{\partial s}{\partial T}\right)_p = c_p$ and $\left(\frac{\partial s}{\partial p}\right)_T = \left(\frac{\partial v}{\partial T}\right)_p$, where $v = V/N$. Thus,

$$\frac{c_p}{k_\mathrm{B}} \frac{\partial \ln T}{\partial t} - \frac{\partial \ln p}{\partial t} = 0 \,. \tag{5.80}$$

We now have in eqns. 5.77 and 5.80 two equations in the two unknowns $\frac{\partial \ln T}{\partial t}$ and $\frac{\partial \ln p}{\partial t}$, yielding

$$\frac{\partial \ln T}{\partial t} = -\frac{k_\mathrm{B}}{c_V} \nabla \cdot V \tag{5.81}$$

$$\frac{\partial \ln p}{\partial t} = -\frac{c_p}{c_V} \nabla \cdot V \,. \tag{5.82}$$

Thus eqn. 5.76 becomes

$$\begin{aligned}
\frac{\partial f^0}{\partial t} + \frac{p}{m} \cdot \frac{\partial f^0}{\partial r} + F \cdot \frac{\partial f^0}{\partial p} = \left(-\frac{\partial f^0}{\partial \varepsilon}\right) &\left\{ \frac{\varepsilon(\Gamma) - h}{T} \, v \cdot \nabla T + m \, v_\alpha v_\beta \, \mathcal{Q}_{\alpha\beta} \right. \\
&\left. + \frac{h - Tc_p - \varepsilon(\Gamma)}{c_V/k_\mathrm{B}} \, \nabla \cdot V - F \cdot v \right\} \,,
\end{aligned} \tag{5.83}$$

where

$$\mathcal{Q}_{\alpha\beta} = \frac{1}{2}\left(\frac{\partial V_\alpha}{\partial x_\beta} + \frac{\partial V_\beta}{\partial x_\alpha}\right) \,. \tag{5.84}$$

Therefore, the Boltzmann equation takes the form

$$\left\{ \frac{\varepsilon(\Gamma) - h}{T} \, \boldsymbol{v} \cdot \boldsymbol{\nabla} T + m \, v_\alpha v_\beta \, \mathcal{Q}_{\alpha\beta} - \frac{\varepsilon(\Gamma) - h + T c_p}{c_V / k_{\mathrm{B}}} \, \boldsymbol{\nabla} \cdot \boldsymbol{V} - \boldsymbol{F} \cdot \boldsymbol{v} \right\} \frac{f^0}{k_{\mathrm{B}} T} + \frac{\partial \, \delta f}{\partial t} = \left(\frac{\partial f}{\partial t} \right)_{\mathrm{coll}} . \tag{5.85}$$

Notice we have dropped the terms $\boldsymbol{v} \cdot \frac{\partial \, \delta f}{\partial \boldsymbol{r}}$ and $\boldsymbol{F} \cdot \frac{\partial \, \delta f}{\partial \boldsymbol{p}}$, since δf must already be first order in smallness, and both the $\frac{\partial}{\partial \boldsymbol{r}}$ operator as well as $\boldsymbol{F}$ add a second order of smallness, which is negligible. Typically $\frac{\partial \, \delta f}{\partial t}$ is nonzero if the applied force $\boldsymbol{F}(t)$ is time-dependent. We use the convention of summing over repeated indices. Note that $\delta_{\alpha\beta} \, \mathcal{Q}_{\alpha\beta} = \mathcal{Q}_{\alpha\alpha} = \boldsymbol{\nabla} \cdot \boldsymbol{V}$. For ideal gases in which only translational and rotational degrees of freedom are excited, $h = c_{\mathrm{p}} T$.

5.5 Relaxation Time Approximation

5.5.1 Approximation of collision integral

We now consider a very simple model of the collision integral,

$$\left(\frac{\partial f}{\partial t} \right)_{\mathrm{coll}} = - \frac{f - f^0}{\tau} = - \frac{\delta f}{\tau} . \tag{5.86}$$

This model is known as the *relaxation time approximation*. Here, $f^0 = f^0(\boldsymbol{r}, \boldsymbol{p}, t)$ is a distribution function which describes a *local equilibrium* at each position $\boldsymbol{r}$ and time t. The quantity τ is the *relaxation time*, which can in principle be momentum-dependent, but which we shall first consider to be constant. In the absence of streaming terms, we have

$$\frac{\partial \, \delta f}{\partial t} = - \frac{\delta f}{\tau} \qquad \Longrightarrow \qquad \delta f(\boldsymbol{r}, \boldsymbol{p}, t) = \delta f(\boldsymbol{r}, \boldsymbol{p}, 0) \, e^{-t/\tau} . \tag{5.87}$$

The distribution f then relaxes to the equilibrium distribution f^0 on a time scale τ. We note that this approximation is obviously flawed in that all quantities – even the collisional invariants – relax to their equilibrium values on the scale τ. In the Appendix, we consider a model for the collision integral in which the collisional invariants are all preserved, but everything else relaxes to local equilibrium at a single rate.

5.5.2 Computation of the scattering time

Consider two particles with velocities $\boldsymbol{v}$ and $\boldsymbol{v}'$. The average of their relative *speed* is

$$\langle \, |\boldsymbol{v} - \boldsymbol{v}'| \, \rangle = \int \! d^3 v \int \! d^3 v' \; P(\boldsymbol{v}) \, P(\boldsymbol{v}') \, |\boldsymbol{v} - \boldsymbol{v}'| \, , \tag{5.88}$$

where $P(\boldsymbol{v})$ is the Maxwell velocity distribution,

$$P(\boldsymbol{v}) = \left(\frac{m}{2 \pi k_{\mathrm{B}} T} \right)^{3/2} \exp \left(- \frac{m \boldsymbol{v}^2}{2 k_{\mathrm{B}} T} \right) , \tag{5.89}$$
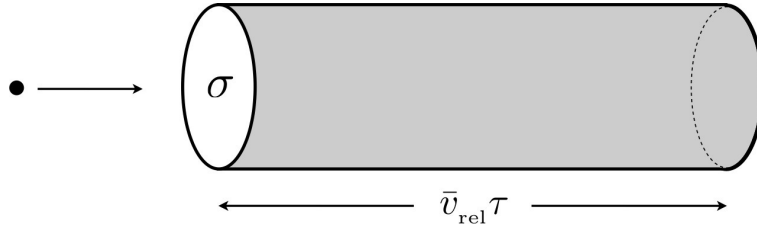
Figure 5.3: Graphic representation of the equation $n\,\sigma\,\bar{v}_{\text{rel}}\,\tau = 1$, which yields the scattering time τ in terms of the number density n, average particle pair relative velocity $\bar{v}_{\text{rel}}$, and two-particle total scattering cross section σ. The equation says that on average there must be one particle within the tube.

which follows from the Boltzmann form of the equilibrium distribution $f^0(\boldsymbol{p})$. It is left as an exercise for the student to verify that

$$\bar{v}_{\text{rel}} \equiv \langle\,|\boldsymbol{v} - \boldsymbol{v}'|\,\rangle = \frac{4}{\sqrt{\pi}}\left(\frac{k_{\text{B}}T}{m}\right)^{1/2}. \tag{5.90}$$

Note that $\bar{v}_{\text{rel}} = \sqrt{2}\,\bar{v}$, where $\bar{v}$ is the average particle speed. Let σ be the total scattering cross section, which for hard spheres is $\sigma = \pi d^2$, where d is the hard sphere diameter. Then the rate at which particles scatter is

$$\frac{1}{\tau} = n\,\bar{v}_{\text{rel}}\,\sigma\,. \tag{5.91}$$

The particle *mean free path* is simply

$$\ell = \bar{v}\,\tau = \frac{1}{\sqrt{2}\,n\,\sigma}\,. \tag{5.92}$$

While the scattering length is not temperature-dependent within this formalism, the scattering time *is* T-dependent, with

$$\tau(T) = \frac{1}{n\,\bar{v}_{\text{rel}}\,\sigma} = \frac{\sqrt{\pi}}{4n\sigma}\left(\frac{m}{k_{\text{B}}T}\right)^{1/2}. \tag{5.93}$$

As $T \to 0$, the collision time diverges as $\tau \propto T^{-1/2}$, because the particles on average move more slowly at lower temperatures. The mean free path, however, is independent of T, and is given by $\ell = 1/\sqrt{2}n\sigma$.

5.5.3 Thermal conductivity

We consider a system with a temperature gradient $\boldsymbol{\nabla}T$ and seek a steady state (*i.e.* time-independent) solution to the Boltzmann equation. We assume $F_\alpha = \mathcal{Q}_{\alpha\beta} = 0$. Appealing to eqn. 5.85, and using the relaxation time approximation for the collision integral, we have

$$\delta f = -\frac{\tau(\varepsilon - c_p\,T)}{k_{\text{B}}T^2}\,(\boldsymbol{v}\cdot\boldsymbol{\nabla}T)\,f^0\,. \tag{5.94}$$

We are now ready to compute the energy and particle currents. In order to compute the *local density* of any quantity $A(\boldsymbol{r},\boldsymbol{p})$, we multiply by the distribution $f(\boldsymbol{r},\boldsymbol{p})$ and integrate over momentum:

$$\rho_A(\boldsymbol{r},t) = \int\! d^3p\,A(\boldsymbol{r},\boldsymbol{p})\,f(\boldsymbol{r},\boldsymbol{p},t)\,, \tag{5.95}$$

For the energy (thermal) current, we let $A = \varepsilon\, v_\alpha = \varepsilon\, p_\alpha/m$, in which case $\rho_A = j_\alpha$. Note that $\int\! d^3p\, \boldsymbol{p}\, f^0 = 0$ since f^0 is isotropic in $\boldsymbol{p}$ even when μ and T depend on $\boldsymbol{r}$. Thus, only δf enters into the calculation of the various currents. Thus, the energy (thermal) current is

$$
\begin{aligned}
j_\varepsilon^\alpha(\boldsymbol{r}) &= \int\! d^3p\, \varepsilon\, v^\alpha\, \delta f \\
&= -\frac{n\tau}{k_\mathrm{B} T^2} \big\langle\, v^\alpha v^\beta\, \varepsilon\, (\varepsilon - c_p\, T)\, \big\rangle \frac{\partial T}{\partial x^\beta}\ ,
\end{aligned}
\tag{5.96}
$$

where the repeated index β is summed over, and where momentum averages are defined relative to the equilibrium distribution, *i.e.*

$$
\langle\, \phi(\boldsymbol{p})\, \rangle = \int\! d^3p\, \phi(\boldsymbol{p})\, f^0(\boldsymbol{p}) \Big/ \int\! d^3p\, f^0(\boldsymbol{p}) = \int\! d^3v\, P(\boldsymbol{v})\, \phi(m\boldsymbol{v})\ .
\tag{5.97}
$$

In this context, it is useful to point out the identity

$$
d^3p\, f^0(\boldsymbol{p}) = n\, d^3v\, P(\boldsymbol{v})\ ,
\tag{5.98}
$$

where

$$
P(\boldsymbol{v}) = \left(\frac{m}{2\pi k_\mathrm{B} T}\right)^{3/2} e^{-m(\boldsymbol{v}-\boldsymbol{V})^2/2k_\mathrm{B} T}
\tag{5.99}
$$

is the Maxwell velocity distribution.

Note that if $\phi = \phi(\varepsilon)$ is a function of the energy, and if $\boldsymbol{V} = 0$, then

$$
d^3p\, f^0(\boldsymbol{p}) = n\, d^3v\, P(\boldsymbol{v}) = n\, \widetilde{P}(\varepsilon)\, d\varepsilon\ ,
\tag{5.100}
$$

where

$$
\widetilde{P}(\varepsilon) = \tfrac{2}{\sqrt{\pi}}(k_\mathrm{B} T)^{-3/2}\, \varepsilon^{1/2}\, e^{-\varepsilon/k_\mathrm{B} T}\ ,
\tag{5.101}
$$

is the Maxwellian distribution of single particle energies. This distribution is normalized with $\int_0^\infty\! d\varepsilon\, \widetilde{P}(\varepsilon) = 1$. Averages with respect to this distribution are given by

$$
\langle\, \phi(\varepsilon)\, \rangle = \int_0^\infty\! d\varepsilon\, \phi(\varepsilon)\, \widetilde{P}(\varepsilon) = \tfrac{2}{\sqrt{\pi}}(k_\mathrm{B} T)^{-3/2}\!\int_0^\infty\! d\varepsilon\, \varepsilon^{1/2}\, \phi(\varepsilon)\, e^{-\varepsilon/k_\mathrm{B} T}\ .
\tag{5.102}
$$

If $\phi(\varepsilon)$ is homogeneous, then for any α we have

$$
\langle\, \varepsilon^\alpha\, \rangle = \tfrac{2}{\sqrt{\pi}}\Gamma\big(\alpha + \tfrac{3}{2}\big)(k_\mathrm{B} T)^\alpha\ .
\tag{5.103}
$$

Due to spatial isotropy, it is clear that we can replace

$$
v^\alpha\, v^\beta \to \tfrac{1}{3}\boldsymbol{v}^2\, \delta_{\alpha\beta} = \frac{2\varepsilon}{3m}\, \delta_{\alpha\beta}
\tag{5.104}
$$

in eqn. 5.96. We then have $\boldsymbol{j}_\varepsilon = -\kappa\, \boldsymbol{\nabla} T$, with

$$
\kappa = \frac{2n\tau}{3mk_\mathrm{B} T^2} \big\langle\, \varepsilon^2(\varepsilon - c_p\, T)\, \big\rangle = \frac{5n\tau k_\mathrm{B}^2 T}{2m} = \tfrac{\pi}{8} n\ell\bar{v}\, c_p\ ,
\tag{5.105}
$$

where we have used $c_p = \tfrac{5}{2}k_\mathrm{B}$ and $\bar{v}^2 = \frac{8k_\mathrm{B} T}{\pi m}$. The quantity κ is called the *thermal conductivity*. Note that $\kappa \propto T^{1/2}$.
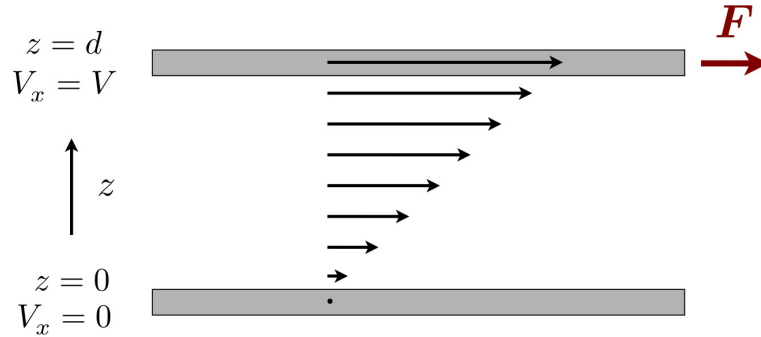
Figure 5.4: *Gedankenexperiment* to measure shear viscosity η in a fluid. The lower plate is fixed. The viscous drag force per unit area on the upper plate is $F_{\mathrm{drag}}/A = -\eta V/d$. This must be balanced by an applied force F.

5.5.4 Viscosity

Consider the situation depicted in fig. 5.4. A fluid filling the space between two large flat plates at $z = 0$ and $z = d$ is set in motion by a force $\boldsymbol{F} = F\hat{\boldsymbol{x}}$ applied to the upper plate; the lower plate is fixed. It is assumed that the fluid's velocity locally matches that of the plates. Fluid particles at the top have an average x-component of their momentum $\langle p_x \rangle = mV$. As these particles move downward toward lower z values, they bring their x-momenta with them. Therefore there is a downward ($-\hat{\boldsymbol{z}}$-directed) flow of $\langle p_x \rangle$. Since x-momentum is constantly being drawn away from $z = d$ plane, this means that there is a $-x$-directed *viscous drag* on the upper plate. The viscous drag force per unit area is given by $F_{\mathrm{drag}}/A = -\eta V/d$, where $V/d = \partial V_x/\partial z$ is the velocity gradient and η is the *shear viscosity*. In steady state, the applied force balances the drag force, *i.e.* $F + F_{\mathrm{drag}} = 0$. Clearly in the steady state the net momentum density of the fluid does not change, and is given by $\frac{1}{2}\rho V \hat{\boldsymbol{x}}$, where ρ is the fluid mass density. The momentum per unit time injected into the fluid by the upper plate at $z = d$ is then extracted by the lower plate at $z = 0$. The *momentum flux density* $\Pi_{xz} = n \langle p_x v_z \rangle$ is the drag force on the upper surface per unit area: $\Pi_{xz} = -\eta \frac{\partial V_x}{\partial z}$. The units of viscosity are $[\eta] = M/LT$.

We now provide some formal definitions of viscosity. As we shall see presently, there is in fact a second type of viscosity, called second viscosity or *bulk viscosity*, which is measurable although not by the type of experiment depicted in fig. 5.4.

The momentum flux tensor $\Pi_{\alpha\beta} = n \langle p_\alpha v_\beta \rangle$ is defined to be the current of momentum component p_α in the direction of increasing x_β. For a gas in motion with average velocity $\boldsymbol{V}$, we have

$$
\begin{aligned}
\Pi_{\alpha\beta} &= nm \langle (V_\alpha + v'_\alpha)(V_\beta + v'_\beta) \rangle \\
&= nm\, V_\alpha V_\beta + nm \langle v'_\alpha v'_\beta \rangle \\
&= nm\, V_\alpha V_\beta + \tfrac{1}{3}nm \langle \boldsymbol{v}'^2 \rangle \delta_{\alpha\beta} \\
&= \rho\, V_\alpha V_\beta + p\, \delta_{\alpha\beta} ,
\end{aligned}
\tag{5.106}
$$

where $\boldsymbol{v}'$ is the particle velocity in a frame moving with velocity $\boldsymbol{V}$, and where we have invoked the ideal gas law $p = nk_{\mathrm{B}}T$. The mass density is $\rho = nm$.

When V is spatially varying,

$$\Pi_{\alpha\beta} = p\,\delta_{\alpha\beta} + \rho\,V_\alpha V_\beta - \tilde{\sigma}_{\alpha\beta}\,, \tag{5.107}$$

where $\tilde{\sigma}_{\alpha\beta}$ is the *viscosity stress tensor*. Any symmetric tensor, such as $\tilde{\sigma}_{\alpha\beta}$, can be decomposed into a sum of (i) a traceless component, and (ii) a component proportional to the identity matrix. Since $\tilde{\sigma}_{\alpha\beta}$ should be, to first order, linear in the spatial derivatives of the components of the velocity field V, there is a unique two-parameter decomposition:

$$\begin{aligned}
\tilde{\sigma}_{\alpha\beta} &= \eta\left(\frac{\partial V_\alpha}{\partial x_\beta} + \frac{\partial V_\beta}{\partial x_\alpha} - \tfrac{2}{3}\,\boldsymbol{\nabla}\!\cdot\!\boldsymbol{V}\,\delta_{\alpha\beta}\right) + \zeta\,\boldsymbol{\nabla}\!\cdot\!\boldsymbol{V}\,\delta_{\alpha\beta} \\
&= 2\eta\left(\mathcal{Q}_{\alpha\beta} - \tfrac{1}{3}\,\mathsf{Tr}\,(\mathcal{Q})\,\delta_{\alpha\beta}\right) + \zeta\,\mathsf{Tr}\,(\mathcal{Q})\,\delta_{\alpha\beta}\,.
\end{aligned} \tag{5.108}$$

The coefficient of the traceless component is η, known as the *shear viscosity*. The coefficient of the component proportional to the identity is ζ, known as the *bulk viscosity*. The full stress tensor $\sigma_{\alpha\beta}$ contains a contribution from the pressure:

$$\sigma_{\alpha\beta} = -p\,\delta_{\alpha\beta} + \tilde{\sigma}_{\alpha\beta}\,. \tag{5.109}$$

The differential force dF_α that a fluid exerts on on a surface element $\hat{n}\,dA$ is

$$dF_\alpha = -\sigma_{\alpha\beta}\,n_\beta\,dA\,, \tag{5.110}$$

where we are using the Einstein summation convention and summing over the repeated index β. We will now compute the shear viscosity η using the Boltzmann equation in the relaxation time approximation.

Appealing again to eqn. 5.85, with $\boldsymbol{F} = 0$ and $h = c_p T$, we find

$$\delta f = -\frac{\tau}{k_{\mathrm{B}}T}\left\{m\,v_\alpha v_\beta\,\mathcal{Q}_{\alpha\beta} + \frac{\varepsilon - c_p T}{T}\,\boldsymbol{v}\cdot\boldsymbol{\nabla}T - \frac{\varepsilon}{c_V/k_{\mathrm{B}}}\,\boldsymbol{\nabla}\!\cdot\!\boldsymbol{V}\right\}f^0\,. \tag{5.111}$$

We assume $\boldsymbol{\nabla}T = \boldsymbol{\nabla}\!\cdot\!\boldsymbol{V} = 0$, and we compute the momentum flux:

$$\begin{aligned}
\Pi_{xz} &= n\int\! d^3p\,p_x v_z\,\delta f \\
&= -\frac{nm^2\tau}{k_{\mathrm{B}}T}\,\mathcal{Q}_{\alpha\beta}\,\langle\,v_x\,v_z\,v_\alpha\,v_\beta\,\rangle \\
&= -\frac{n\tau}{k_{\mathrm{B}}T}\left(\frac{\partial V_x}{\partial z} + \frac{\partial V_z}{\partial x}\right)\langle\,mv_x^2\cdot mv_z^2\,\rangle \\
&= -n\tau k_{\mathrm{B}}T\left(\frac{\partial V_z}{\partial x} + \frac{\partial V_x}{\partial z}\right).
\end{aligned} \tag{5.112}$$

Thus, if $V_x = V_x(z)$, we have

$$\Pi_{xz} = -n\tau k_{\mathrm{B}}T\,\frac{\partial V_x}{\partial z} \tag{5.113}$$

from which we read off the viscosity,

$$\eta = nk_{\mathrm{B}}T\tau = \tfrac{\pi}{8}nm\ell\bar{v}\,. \tag{5.114}$$
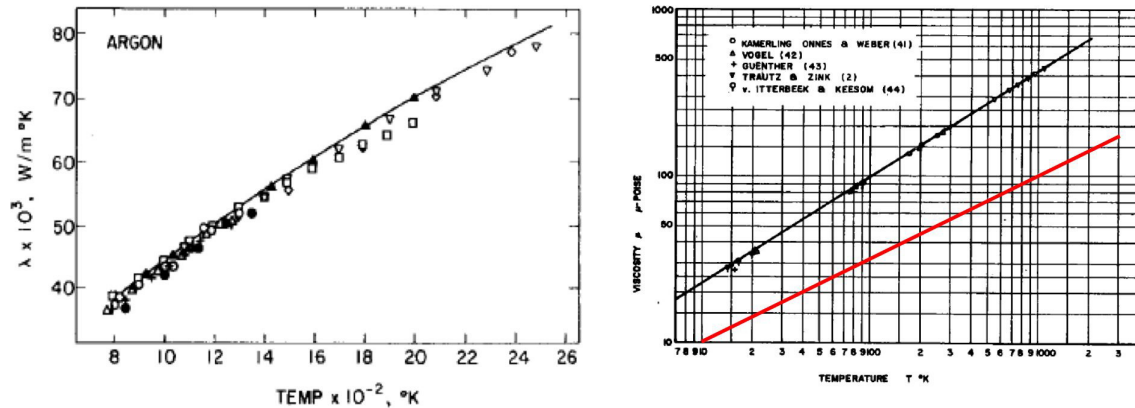
Figure 5.5: Left: thermal conductivity (λ in figure) of Ar between $T = 800\,\mathrm{K}$ and $T = 2600\,\mathrm{K}$. The best fit to a single power law $\lambda = aT^b$ results in $b = 0.651$. Source: G. S. Springer and E. W. Wingeier, *J. Chem Phys.* **59**, 1747 (1972). Right: log-log plot of shear viscosity (μ in figure) of He between $T \approx 15\,\mathrm{K}$ and $T \approx 1000\,\mathrm{K}$. The red line has slope $\frac{1}{2}$. The slope of the data is approximately 0.633. Source: J. Kestin and W. Leidenfrost, *Physica* **25**, 537 (1959).

Note that $\eta(T) \propto T^{1/2}$.

How well do these predictions hold up? In fig. 5.5, we plot data for the thermal conductivity of argon and the shear viscosity of helium. Both show a clear sublinear behavior as a function of temperature, but the slope $d\ln\kappa/dT$ is approximately 0.65 and $d\ln\eta/dT$ is approximately 0.63. Clearly the simple model is not even getting the functional dependence on T right, let alone its coefficient. Still, our crude theory is at least qualitatively correct.

Why do both $\kappa(T)$ as well as $\eta(T)$ decrease at low temperatures? The reason is that the heat current which flows in response to $\boldsymbol{\nabla}T$ as well as the momentum current which flows in response to $\partial V_x/\partial z$ are due to the presence of collisions, which result in momentum and energy transfer between particles. This is true even when total energy and momentum are conserved, which they are not in the relaxation time approximation. Intuitively, we might think that the viscosity should *increase* as the temperature is lowered, since common experience tells us that fluids 'gum up' as they get colder – think of honey as an extreme example. But of course honey is nothing like an ideal gas, and the physics behind the crystallization or glass transition which occurs in real fluids when they get sufficiently cold is completely absent from our approach. In our calculation, viscosity results from collisions, and with no collisions there is no momentum transfer and hence no viscosity. If, for example, the gas particles were to simply pass through each other, as though they were ghosts, then there would be no opposition to maintaining an arbitrary velocity gradient.

5.5.5 Oscillating external force

Suppose a uniform oscillating external force $\boldsymbol{F}_{\mathrm{ext}}(t) = \boldsymbol{F}\,e^{-i\omega t}$ is applied. For a system of charged particles, this force would arise from an external electric field $\boldsymbol{F}_{\mathrm{ext}} = q\boldsymbol{E}\,e^{-i\omega t}$, where q is the charge of each

particle. We'll assume $\nabla T = 0$. The Boltzmann equation is then written

$$\frac{\partial f}{\partial t} + \frac{\boldsymbol{p}}{m} \cdot \frac{\partial f}{\partial \boldsymbol{r}} + \boldsymbol{F} e^{-i\omega t} \cdot \frac{\partial f}{\partial \boldsymbol{p}} = -\frac{f - f^0}{\tau} \ . \tag{5.115}$$

We again write $f = f^0 + \delta f$, and we assume δf is spatially constant. Thus,

$$\frac{\partial \, \delta f}{\partial t} + \boldsymbol{F} e^{-i\omega t} \cdot \boldsymbol{v} \, \frac{\partial f^0}{\partial \varepsilon} = -\frac{\delta f}{\tau} \ . \tag{5.116}$$

If we assume $\delta f(t) = \delta f(\omega) \, e^{-i\omega t}$ then the above differential equation is converted to an algebraic equation, with solution

$$\delta f(t) = -\frac{\tau \, e^{-i\omega t}}{1 - i\omega \tau} \frac{\partial f^0}{\partial \varepsilon} \, \boldsymbol{F} \cdot \boldsymbol{v} \ . \tag{5.117}$$

We now compute the particle current:

$$
\begin{aligned}
j_\alpha(\boldsymbol{r}, t) &= \int d^3 p \, \boldsymbol{v} \, \delta f \\[2mm]
&= \frac{\tau \, e^{-i\omega t}}{1 - i\omega \tau} \cdot \frac{F_\beta}{k_{\rm B} T} \int d^3 p \, f^0(\boldsymbol{p}) \, v_\alpha \, v_\beta \\[2mm]
&= \frac{\tau \, e^{-i\omega t}}{1 - i\omega \tau} \cdot \frac{n F_\alpha}{3 k_{\rm B} T} \int d^3 v \, P(\boldsymbol{v}) \, \boldsymbol{v}^2 \\[2mm]
&= \frac{n\tau}{m} \cdot \frac{F_\alpha \, e^{-i\omega t}}{1 - i\omega \tau} \ .
\end{aligned}
\tag{5.118}
$$

If the particles are electrons, with charge $q = -e$, then the electrical current is $(-e)$ times the particle current. We then obtain

$$j_\alpha^{\rm (elec)}(t) = \frac{ne^2\tau}{m} \cdot \frac{E_\alpha \, e^{-i\omega t}}{1 - i\omega \tau} \equiv \sigma_{\alpha\beta}(\omega) \, E_\beta \, e^{-i\omega t} \ , \tag{5.119}$$

where

$$\sigma_{\alpha\beta}(\omega) = \frac{ne^2\tau}{m} \cdot \frac{1}{1 - i\omega \tau} \, \delta_{\alpha\beta} \tag{5.120}$$

is the frequency-dependent electrical conductivity tensor. Of course for fermions such as electrons, we should be using the Fermi distribution in place of the Maxwell-Boltzmann distribution for $f^0(\boldsymbol{p})$. This affects the relation between n and μ only, and the final result for the conductivity tensor $\sigma_{\alpha\beta}(\omega)$ is unchanged.

5.5.6 Quick and Dirty Treatment of Transport

Suppose we have some averaged intensive quantity ϕ which is spatially dependent through $T(\boldsymbol{r})$ or $\mu(\boldsymbol{r})$ or $\boldsymbol{V}(\boldsymbol{r})$. For simplicity we will write $\phi = \phi(z)$. We wish to compute the current of ϕ across some surface whose equation is $dz = 0$. If the mean free path is ℓ, then the value of ϕ for particles crossing this surface in the $+\hat{z}$ direction is $\phi(z - \ell \cos\theta)$, where θ is the angle the particle's velocity makes with respect to

$\hat{\boldsymbol{z}}$, *i.e.* $\cos\theta = v_z/v$. We perform the same analysis for particles moving in the $-\hat{\boldsymbol{z}}$ direction, for which $\phi = \phi(z + \ell\cos\theta)$. The current of ϕ through this surface is then

$$
\begin{aligned}
\boldsymbol{j}_\phi &= n\hat{\boldsymbol{z}}\int\limits_{v_z>0} d^3v\, P(\boldsymbol{v})\, v_z\, \phi(z - \ell\cos\theta) + n\hat{\boldsymbol{z}}\int\limits_{v_z<0} d^3v\, P(\boldsymbol{v})\, v_z\, \phi(z + \ell\cos\theta) \\
&= -n\ell\,\frac{\partial\phi}{\partial z}\hat{\boldsymbol{z}}\int d^3v\, P(\boldsymbol{v})\,\frac{v_z^2}{v} = -\tfrac{1}{3}n\bar{v}\ell\,\frac{\partial\phi}{\partial z}\,\hat{\boldsymbol{z}}\ ,
\end{aligned}
\tag{5.121}
$$

where $\bar{v} = \sqrt{\frac{8k_B T}{\pi m}}$ is the average particle speed. If the z-dependence of ϕ comes through the dependence of ϕ on the local temperature T, then we have

$$
\boldsymbol{j}_\phi = -\tfrac{1}{3}\,n\ell\bar{v}\,\frac{\partial\phi}{\partial T}\,\boldsymbol{\nabla} T \equiv -K\,\boldsymbol{\nabla} T\ ,
\tag{5.122}
$$

where

$$
K = \tfrac{1}{3}n\ell\bar{v}\,\frac{\partial\phi}{\partial T}
\tag{5.123}
$$

is the transport coefficient. If $\phi = \langle\varepsilon\rangle$, then $\frac{\partial\phi}{\partial T} = c_p$, where c_p is the heat capacity per particle at constant pressure. We then find $\boldsymbol{j}_\varepsilon = -\kappa\,\boldsymbol{\nabla} T$ with thermal conductivity

$$
\kappa = \tfrac{1}{3}n\ell\bar{v}\,c_p\ .
\tag{5.124}
$$

Our Boltzmann equation calculation yielded the same result, but with a prefactor of $\frac{\pi}{8}$ instead of $\frac{1}{3}$.

We can make a similar argument for the viscosity. In this case $\phi = \langle p_x\rangle$ is spatially varying through its dependence on the flow velocity $\boldsymbol{V}(\boldsymbol{r})$. Clearly $\partial\phi/\partial V_x = m$, hence

$$
j_{p_x}^z = \Pi_{xz} = -\tfrac{1}{3}nm\ell\bar{v}\,\frac{\partial V_x}{\partial z}\ ,
\tag{5.125}
$$

from which we identify the viscosity, $\eta = \tfrac{1}{3}nm\ell\bar{v}$. Once again, this agrees in its functional dependences with the Boltzmann equation calculation in the relaxation time approximation. Only the coefficients differ. The ratio of the coefficients is $K_{\text{QDC}}/K_{\text{BRT}} = \frac{8}{3\pi} = 0.849$ in both cases[6].

5.5.7   Thermal diffusivity, kinematic viscosity, and Prandtl number

Suppose, under conditions of constant pressure, we add heat q per unit volume to an ideal gas. We know from thermodynamics that its temperature will then increase by an amount $\Delta T = q/nc_p$. If a heat current $\boldsymbol{j}_q$ flows, then the continuity equation for energy flow requires

$$
nc_p\,\frac{\partial T}{\partial t} + \boldsymbol{\nabla}\cdot\boldsymbol{j}_q = 0\ .
\tag{5.126}
$$

In a system where there is no net particle current, the heat current $\boldsymbol{j}_q$ is the same as the energy current $\boldsymbol{j}_\varepsilon$, and since $\boldsymbol{j}_\varepsilon = -\kappa\,\boldsymbol{\nabla} T$, we obtain a diffusion equation for temperature,

$$
\frac{\partial T}{\partial t} = \frac{\kappa}{nc_p}\,\nabla^2 T\ .
\tag{5.127}
$$

---

[6]Here we abbreviate QDC for '*quick and dirty calculation*' and BRT for '*Boltzmann equation in the relaxation time approximation*'.

Gas	$\eta\,(\mu\mathrm{Pa}\cdot\mathrm{s})$	$\kappa\,(\mathrm{mW/m}\cdot\mathrm{K})$	c_p/k_{B}	Pr
He	19.5	149	2.50	0.682
Ar	22.3	17.4	2.50	0.666
Xe	22.7	5.46	2.50	0.659
H_2	8.67	179	3.47	0.693
N_2	17.6	25.5	3.53	0.721
O_2	20.3	26.0	3.50	0.711
CH_4	11.2	33.5	4.29	0.74
CO_2	14.8	18.1	4.47	0.71
NH_3	10.1	24.6	4.50	0.90

Table 5.1: Viscosities, thermal conductivities, and Prandtl numbers for some common gases at $T = 293\,\mathrm{K}$ and $p = 1\,\mathrm{atm}$. (Source: Table 1.1 of Smith and Jensen, with data for triatomic gases added.)

The combination

$$a \equiv \frac{\kappa}{nc_p} \tag{5.128}$$

is known as the *thermal diffusivity*. Our Boltzmann equation calculation in the relaxation time approximation yielded the result $\kappa = nk_{\mathrm{B}}T\tau c_p/m$. Thus, we find $a = k_{\mathrm{B}}T\tau/m$ via this method. Note that the dimensions of a are the same as for any diffusion constant D, namely $[a] = L^2/T$.

Another quantity with dimensions of L^2/T is the *kinematic viscosity*, $\nu = \eta/\rho$, where $\rho = nm$ is the mass density. We found $\eta = nk_{\mathrm{B}}T\tau$ from the relaxation time approximation calculation, hence $\nu = k_{\mathrm{B}}T\tau/m$. The ratio ν/a, called the *Prandtl number*, $\mathrm{Pr} = \eta c_p/m\kappa$, is dimensionless. According to our calculations, $\mathrm{Pr} = 1$. According to table 5.1, most monatomic gases have $\mathrm{Pr} \approx \frac{2}{3}$.

5.6 Diffusion and the Lorentz model

5.6.1 Failure of the relaxation time approximation

As we remarked above, the relaxation time approximation fails to conserve any of the collisional invariants. It is therefore unsuitable for describing hydrodynamic phenomena such as diffusion. To see this, let $f(\boldsymbol{r}, \boldsymbol{v}, t)$ be the distribution function, here written in terms of position, velocity, and time rather than position, momentum, and time as befor[7]. In the absence of external forces, the Boltzmann equation in the relaxation time approximation is

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \frac{\partial f}{\partial \boldsymbol{r}} = -\frac{f - f^0}{\tau} . \tag{5.129}$$

The density of particles in velocity space is given by

$$\tilde{n}(\boldsymbol{v}, t) = \int d^3r \, f(\boldsymbol{r}, \boldsymbol{v}, t) . \tag{5.130}$$

---

[7]The difference is trivial, since $\boldsymbol{p} = m\boldsymbol{v}$.

In equilibrium, this is the Maxwell distribution times the total number of particles: $\tilde{n}_0(\boldsymbol{v}) = NP_{\mathrm{M}}(\boldsymbol{v})$. The number of particles as a function of time, $N(t) = \int d^3v\, \tilde{n}(\boldsymbol{v}, t)$, should be a constant.

Integrating the Boltzmann equation one has

$$\frac{\partial \tilde{n}}{\partial t} = -\frac{\tilde{n} - \tilde{n}_0}{\tau} \ . \tag{5.131}$$

Thus, with $\delta \tilde{n}(\boldsymbol{v}, t) = \tilde{n}(\boldsymbol{v}, t) - \tilde{n}_0(\boldsymbol{v})$, we have

$$\delta \tilde{n}(\boldsymbol{v}, t) = \delta \tilde{n}(\boldsymbol{v}, 0)\, e^{-t/\tau} \ . \tag{5.132}$$

Thus, $\tilde{n}(\boldsymbol{v}, t)$ decays exponentially to zero with time constant τ, from which it follows that the total particle number exponentially relaxes to N_0. This is physically incorrect; local density perturbations can't just *vanish*. Rather, they *diffuse*.

5.6.2   Modified Boltzmann equation and its solution

To remedy this unphysical aspect, consider the modified Boltzmann equation,

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \frac{\partial f}{\partial \boldsymbol{r}} = \frac{1}{\tau} \left[-f + \int \frac{d\hat{\boldsymbol{v}}}{4\pi}\, f \right] \equiv \frac{1}{\tau} (\mathbb{P} - 1) f \ , \tag{5.133}$$

where $\mathbb{P}$ is a projector onto a space of isotropic functions of $\boldsymbol{v}$: $\mathbb{P}\, F = \int \frac{d\hat{\boldsymbol{v}}}{4\pi}\, F(\boldsymbol{v})$ for any function $F(\boldsymbol{v})$. Note that $\mathbb{P}\, F$ is a function of the speed $v = |\boldsymbol{v}|$. For this modified equation, known as the *Lorentz model*, one finds $\partial_t \tilde{n} = 0$.

The model in eqn. 5.133 is known as the *Lorentz model*[8]. To solve it, we consider the Laplace transform,

$$\hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) = \int\limits_0^\infty dt\, e^{-st} \int d^3r\, e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\, f(\boldsymbol{r}, \boldsymbol{v}, t) \ . \tag{5.134}$$

Taking the Laplace transform of eqn. 5.133, we find

$$\left(s + i\boldsymbol{v} \cdot \boldsymbol{k} + \tau^{-1} \right) \hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) = \tau^{-1}\, \mathbb{P}\, \hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) + f(\boldsymbol{k}, \boldsymbol{v}, t = 0) \ . \tag{5.135}$$

We now solve for $\mathbb{P}\, \hat{f}(\boldsymbol{k}, \boldsymbol{v}, s)$:

$$\hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) = \frac{\tau^{-1}}{s + i\boldsymbol{v} \cdot \boldsymbol{k} + \tau^{-1}}\, \mathbb{P}\, \hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) + \frac{f(\boldsymbol{k}, \boldsymbol{v}, t = 0)}{s + i\boldsymbol{v} \cdot \boldsymbol{k} + \tau^{-1}} \ , \tag{5.136}$$

which entails

$$\mathbb{P}\, \hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) = \left[\int \frac{d\hat{\boldsymbol{v}}}{4\pi}\, \frac{\tau^{-1}}{s + i\boldsymbol{v} \cdot \boldsymbol{k} + \tau^{-1}} \right] \mathbb{P}\, \hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) + \int \frac{d\hat{\boldsymbol{v}}}{4\pi}\, \frac{f(\boldsymbol{k}, \boldsymbol{v}, t = 0)}{s + i\boldsymbol{v} \cdot \boldsymbol{k} + \tau^{-1}} \ . \tag{5.137}$$

---

[8]See the excellent discussion in the book by Krapivsky, Redner, and Ben-Naim, cited in §8.1.

Now we have

$$\int \frac{d\hat{\boldsymbol{v}}}{4\pi} \frac{\tau^{-1}}{s + i\boldsymbol{v}\cdot\boldsymbol{k} + \tau^{-1}} = \int_{-1}^{1} dx \frac{\tau^{-1}}{s + ivkx + \tau^{-1}} \tag{5.138}$$

$$= \frac{1}{vk} \tan^{-1}\left(\frac{vk\tau}{1 + \tau s}\right) .$$

Thus,

$$\mathbb{P} f(\boldsymbol{k}, \boldsymbol{v}, s) = \left[1 - \frac{1}{vk\tau} \tan^{-1}\left(\frac{vk\tau}{1 + \tau s}\right)\right]^{-1} \int \frac{d\hat{\boldsymbol{v}}}{4\pi} \frac{f(\boldsymbol{k}, \boldsymbol{v}, t = 0)}{s + i\boldsymbol{v}\cdot\boldsymbol{k} + \tau^{-1}} . \tag{5.139}$$

We now have the solution to Lorentz's modified Boltzmann equation:

$$\hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) = \frac{\tau^{-1}}{s + i\boldsymbol{v}\cdot\boldsymbol{k} + \tau^{-1}} \left[1 - \frac{1}{vk\tau} \tan^{-1}\left(\frac{vk\tau}{1 + \tau s}\right)\right]^{-1} \int \frac{d\hat{\boldsymbol{v}}}{4\pi} \frac{f(\boldsymbol{k}, \boldsymbol{v}, t = 0)}{s + i\boldsymbol{v}\cdot\boldsymbol{k} + \tau^{-1}}$$

$$+ \frac{f(\boldsymbol{k}, \boldsymbol{v}, t = 0)}{s + i\boldsymbol{v}\cdot\boldsymbol{k} + \tau^{-1}} . \tag{5.140}$$

Let us assume an initial distribution which is perfectly localized in both $\boldsymbol{r}$ and $\boldsymbol{v}$:

$$f(\boldsymbol{r}, \boldsymbol{v}, t = 0) = \delta(\boldsymbol{v} - \boldsymbol{v}_0) . \tag{5.141}$$

For these initial conditions, we find

$$\int \frac{d\hat{\boldsymbol{v}}}{4\pi} \frac{f(\boldsymbol{k}, \boldsymbol{v}, t = 0)}{s + i\boldsymbol{v}\cdot\boldsymbol{k} + \tau^{-1}} = \frac{1}{s + i\boldsymbol{v}_0\cdot\boldsymbol{k} + \tau^{-1}} \cdot \frac{\delta(v - v_0)}{4\pi v_0^2} . \tag{5.142}$$

We further have that

$$1 - \frac{1}{vk\tau} \tan^{-1}\left(\frac{vk\tau}{1 + \tau s}\right) = s\tau + \tfrac{1}{3}k^2 v^2 \tau^2 + \dots , \tag{5.143}$$

and therefore

$$\hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) = \frac{\tau^{-1}}{s + i\boldsymbol{v}\cdot\boldsymbol{k} + \tau^{-1}} \cdot \frac{\tau^{-1}}{s + i\boldsymbol{v}_0\cdot\boldsymbol{k} + \tau^{-1}} \cdot \frac{1}{s + \tfrac{1}{3}v_0^2\, k^2\, \tau + \dots} \cdot \frac{\delta(v - v_0)}{4\pi v_0^2}$$

$$+ \frac{\delta(v - v_0)}{s + i\boldsymbol{v}_0\cdot\boldsymbol{k} + \tau^{-1}} . \tag{5.144}$$

We are interested in the long time limit $t \gg \tau$ for $f(\boldsymbol{r}, \boldsymbol{v}, t)$. This is dominated by $s \sim t^{-1}$, and we assume that τ^{-1} is dominant over s and $i\boldsymbol{v}\cdot\boldsymbol{k}$. We then have

$$\hat{f}(\boldsymbol{k}, \boldsymbol{v}, s) \approx \frac{1}{s + \tfrac{1}{3}v_0^2\, k^2\, \tau} \cdot \frac{\delta(v - v_0)}{4\pi v_0^2} . \tag{5.145}$$

Performing the inverse Laplace and Fourier transforms, we obtain

$$f(\boldsymbol{r}, \boldsymbol{v}, t) = (4\pi Dt)^{-3/2}\, e^{-r^2/4Dt} \cdot \frac{\delta(v - v_0)}{4\pi v_0^2} , \tag{5.146}$$

where the *diffusion constant* is

$$D = \tfrac{1}{3} v_0^2 \, \tau \; . \tag{5.147}$$

The units are $[D] = L^2/T$. Integrating over velocities, we have the density

$$n(\boldsymbol{r}, t) = \int\! d^3v \; f(\boldsymbol{r}, \boldsymbol{v}, t) = (4\pi Dt)^{-3/2} \, e^{-r^2/4Dt} \; . \tag{5.148}$$

Note that

$$\int\! d^3r \; n(\boldsymbol{r}, t) = 1 \tag{5.149}$$

for all time. Total particle number is conserved!

5.7 Linearized Boltzmann Equation

5.7.1 Linearizing the collision integral

We now return to the classical Boltzmann equation and consider a more formal treatment of the collision term in the linear approximation. We will assume time-reversal symmetry, in which case

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \int\! d^3p_1 \!\int\! d^3p' \!\int\! d^3p_1' \; w(\boldsymbol{p}', \boldsymbol{p}_1' \,|\, \boldsymbol{p}, \boldsymbol{p}_1) \left\{ f(\boldsymbol{p}') \, f(\boldsymbol{p}_1') - f(\boldsymbol{p}) \, f(\boldsymbol{p}_1) \right\} . \tag{5.150}$$

The collision integral is nonlinear in the distribution f. We *linearize* by writing

$$f(\boldsymbol{p}) = f^0(\boldsymbol{p}) + f^0(\boldsymbol{p}) \, \psi(\boldsymbol{p}) \; , \tag{5.151}$$

where we assume $\psi(\boldsymbol{p})$ is small. We then have, to first order in ψ,

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = f^0(\boldsymbol{p}) \, \hat{L}\psi + \mathcal{O}(\psi^2) \; , \tag{5.152}$$

where the action of the *linearized collision operator* is given by

$$
\begin{aligned}
\hat{L}\psi &= \int\! d^3p_1 \!\int\! d^3p' \!\int\! d^3p_1' \; w(\boldsymbol{p}', \boldsymbol{p}_1' \,|\, \boldsymbol{p}, \boldsymbol{p}_1) \, f^0(\boldsymbol{p}_1) \left\{ \psi(\boldsymbol{p}') + \psi(\boldsymbol{p}_1') - \psi(\boldsymbol{p}) - \psi(\boldsymbol{p}_1) \right\} \\
&= \int\! d^3p_1 \!\int\! d\Omega \, |\boldsymbol{v} - \boldsymbol{v}_1| \, \frac{\partial \sigma}{\partial \Omega} \, f^0(\boldsymbol{p}_1) \left\{ \psi(\boldsymbol{p}') + \psi(\boldsymbol{p}_1') - \psi(\boldsymbol{p}) - \psi(\boldsymbol{p}_1) \right\} ,
\end{aligned}
\tag{5.153}
$$

where we have invoked eqn. 5.55 to write the RHS in terms of the differential scattering cross section. In deriving the above result, we have made use of the detailed balance relation,

$$f^0(\boldsymbol{p}) \, f^0(\boldsymbol{p}_1) = f^0(\boldsymbol{p}') \, f^0(\boldsymbol{p}_1') \; . \tag{5.154}$$

We have also suppressed the $\boldsymbol{r}$ dependence in writing $f(\boldsymbol{p})$, $f^0(\boldsymbol{p})$, and $\psi(\boldsymbol{p})$.

From eqn. 5.85, we then have the linearized equation

$$\left(\hat{L} - \frac{\partial}{\partial t}\right)\psi = Y, \tag{5.155}$$

where, for point particles,

$$Y = \frac{1}{k_{\mathrm{B}}T}\left\{\frac{\varepsilon(\boldsymbol{p}) - c_p T}{T}\,\boldsymbol{v}\cdot\boldsymbol{\nabla}T + m\,v_\alpha v_\beta\,\mathcal{Q}_{\alpha\beta} - \frac{k_{\mathrm{B}}\,\varepsilon(\boldsymbol{p})}{c_V}\,\boldsymbol{\nabla}\cdot\boldsymbol{V} - \boldsymbol{F}\cdot\boldsymbol{v}\right\}. \tag{5.156}$$

Eqn. 5.155 is an inhomogeneous linear equation, which can be solved by inverting the operator $\hat{L} - \frac{\partial}{\partial t}$.

5.7.2 Linear algebraic properties of $\hat{L}$

Although $\hat{L}$ is an integral operator, it shares many properties with other linear operators with which you are familiar, such as matrices and differential operators. We can define an *inner product*[9],

$$\langle\,\psi_1\,|\,\psi_2\,\rangle \equiv \int d^3p\,f^0(\boldsymbol{p})\,\psi_1(\boldsymbol{p})\,\psi_2(\boldsymbol{p})\,. \tag{5.157}$$

Note that this is not the usual Hilbert space inner product from quantum mechanics, since the factor $f^0(\boldsymbol{p})$ is included in the metric. This is necessary in order that $\hat{L}$ be *self-adjoint*:

$$\langle\,\psi_1\,|\,\hat{L}\psi_2\,\rangle = \langle\,\hat{L}\psi_1\,|\,\psi_2\,\rangle\,. \tag{5.158}$$

We can now define the spectrum of *normalized eigenfunctions* of $\hat{L}$, which we write as $\phi_n(\boldsymbol{p})$. The eigenfunctions satisfy the eigenvalue equation,

$$\hat{L}\phi_n = -\lambda_n\,\phi_n\,, \tag{5.159}$$

and may be chosen to be orthonormal,

$$\langle\,\phi_m\,|\,\phi_n\,\rangle = \delta_{mn}\,. \tag{5.160}$$

Of course, in order to obtain the eigenfunctions ϕ_n we must have detailed knowledge of the function $w(\boldsymbol{p}',\boldsymbol{p}_1'\,|\,\boldsymbol{p},\boldsymbol{p}_1)$.

Recall that there are five collisional invariants, which are the particle number, the three components of the total particle momentum, and the particle energy. To each collisional invariant, there is an associated eigenfunction ϕ_n with eigenvalue $\lambda_n = 0$. One can check that these normalized eigenfunctions are

$$\phi_n(\boldsymbol{p}) = \frac{1}{\sqrt{n}} \tag{5.161}$$

$$\phi_{p_\alpha}(\boldsymbol{p}) = \frac{p_\alpha}{\sqrt{nmk_{\mathrm{B}}T}} \tag{5.162}$$

$$\phi_\varepsilon(\boldsymbol{p}) = \sqrt{\frac{2}{3n}}\left(\frac{\varepsilon(\boldsymbol{p})}{k_{\mathrm{B}}T} - \frac{3}{2}\right). \tag{5.163}$$

---

[9]The requirements of an inner product $\langle f|g\rangle$ are symmetry, linearity, and non-negative definiteness.

If there are no temperature, chemical potential, or bulk velocity gradients, and there are no external forces, then $Y = 0$ and the only changes to the distribution are from collisions. The linearized Boltzmann equation becomes

$$\frac{\partial \psi}{\partial t} = \hat{L} \psi \; . \tag{5.164}$$

We can therefore write the most general solution in the form

$$\psi(\boldsymbol{p}, t) = {\sum_n}' C_n \, \phi_n(\boldsymbol{p}) \, e^{-\lambda_n t} \; , \tag{5.165}$$

where the prime on the sum reminds us that collisional invariants are to be excluded. All the eigenvalues λ_n, aside from the five zero eigenvalues for the collisional invariants, must be positive. Any negative eigenvalue would cause $\psi(\boldsymbol{p}, t)$ to increase without bound, and an initial nonequilibrium distribution would not relax to the equilibrium $f^0(\boldsymbol{p})$, which we regard as unphysical. Henceforth we will drop the prime on the sum but remember that $C_n = 0$ for the five collisional invariants.

Recall also the particle, energy, and thermal (heat) currents,

$$
\begin{aligned}
\boldsymbol{j} &= \int d^3p \, \boldsymbol{v} \, f(\boldsymbol{p}) = \int d^3p \, f^0(\boldsymbol{p}) \, \boldsymbol{v} \, \psi(\boldsymbol{p}) = \langle \, \boldsymbol{v} \, | \, \psi \, \rangle \\
\boldsymbol{j}_\varepsilon &= \int d^3p \, \boldsymbol{v} \, \varepsilon \, f(\boldsymbol{p}) = \int d^3p \, f^0(\boldsymbol{p}) \, \boldsymbol{v} \, \varepsilon \, \psi(\boldsymbol{p}) = \langle \, \boldsymbol{v} \, \varepsilon \, | \, \psi \, \rangle \\
\boldsymbol{j}_q &= \int d^3p \, \boldsymbol{v} \, (\varepsilon - \mu) \, f(\boldsymbol{p}) = \int d^3p \, f^0(\boldsymbol{p}) \, \boldsymbol{v} \, (\varepsilon - \mu) \, \psi(\boldsymbol{p}) = \langle \, \boldsymbol{v} \, (\varepsilon - \mu) \, | \, \psi \, \rangle \; .
\end{aligned}
\tag{5.166}
$$

Note $\boldsymbol{j}_q = \boldsymbol{j}_\varepsilon - \mu \boldsymbol{j}$.

5.7.3   Steady state solution to the linearized Boltzmann equation

Under steady state conditions, there is no time dependence, and the linearized Boltzmann equation takes the form

$$\hat{L} \psi = Y \; . \tag{5.167}$$

We may expand ψ in the eigenfunctions ϕ_n and write $\psi = \sum_n C_n \, \phi_n$. Applying $\hat{L}$ and taking the inner product with ϕ_j, we have

$$C_j = -\frac{1}{\lambda_j} \, \langle \, \phi_j \, | \, Y \, \rangle \; . \tag{5.168}$$

Thus, the formal solution to the linearized Boltzmann equation is

$$\psi(\boldsymbol{p}) = -\sum_n \frac{1}{\lambda_n} \, \langle \, \phi_n \, | \, Y \, \rangle \, \phi_n(\boldsymbol{p}) \; . \tag{5.169}$$

This solution is applicable provided $| \, Y \, \rangle$ is orthogonal to the five collisional invariants.

Thermal conductivity

For the thermal conductivity, we take $\boldsymbol{\nabla} T = \partial_z T\, \hat{\boldsymbol{x}}$, and

$$Y = \frac{1}{k_{\mathrm{B}} T^2}\, \frac{\partial T}{\partial x} \cdot X_\kappa\,, \tag{5.170}$$

where $X_\kappa \equiv (\varepsilon - c_p T)\, v_x$. Under the conditions of no particle flow ($\boldsymbol{j} = 0$), we have $\boldsymbol{j}_q = -\kappa\, \partial_x T\, \hat{\boldsymbol{x}}$. Then we have

$$\langle\, X_\kappa \,|\, \psi \,\rangle = -\kappa\, \frac{\partial T}{\partial x}\,. \tag{5.171}$$

Viscosity

For the viscosity, we take

$$Y = \frac{m}{k_{\mathrm{B}} T}\, \frac{\partial V_x}{\partial y} \cdot X_\eta\,, \tag{5.172}$$

with $X_\eta = v_x\, v_y$. We then

$$\Pi_{xy} = \langle\, m\, v_x\, v_y \,|\, \psi \,\rangle = -\eta\, \frac{\partial V_x}{\partial y}\,. \tag{5.173}$$

Thus,

$$\langle\, X_\eta \,|\, \psi \,\rangle = -\frac{\eta}{m}\, \frac{\partial V_x}{\partial y}\,. \tag{5.174}$$

5.7.4 Variational approach

Following the treatment in chapter 1 of Smith and Jensen, define $\hat{H} \equiv -\hat{L}$. We have that $\hat{H}$ is a positive semidefinite operator, whose only zero eigenvalues correspond to the collisional invariants. We then have the Schwarz inequality,

$$\langle\, \psi \,|\, \hat{H} \,|\, \psi \,\rangle \cdot \langle\, \phi \,|\, \hat{H} \,|\, \phi \,\rangle \geq \langle\, \phi \,|\, \hat{H} \,|\, \psi \,\rangle^2\,, \tag{5.175}$$

for any two Hilbert space vectors $|\,\psi\,\rangle$ and $|\,\phi\,\rangle$. Consider now the above calculation of the thermal conductivity. We have

$$\hat{H}\psi = -\frac{1}{k_{\mathrm{B}} T^2}\, \frac{\partial T}{\partial x}\, X_\kappa \tag{5.176}$$

and therefore

$$\kappa = \frac{k_{\mathrm{B}} T^2}{(\partial T/\partial x)^2}\, \langle\, \psi \,|\, \hat{H} \,|\, \psi \,\rangle \geq \frac{1}{k_{\mathrm{B}} T^2}\, \frac{\langle\, \phi \,|\, X_\kappa \,\rangle^2}{\langle\, \phi \,|\, \hat{H} \,|\, \phi \,\rangle}\,. \tag{5.177}$$

Similarly, for the viscosity, we have

$$\hat{H}\psi = -\frac{m}{k_{\mathrm{B}} T}\, \frac{\partial V_x}{\partial y}\, X_\eta\,, \tag{5.178}$$

from which we derive

$$\eta = \frac{k_{\mathrm{B}}T}{(\partial V_x/\partial y)^2} \, \langle \, \psi \, | \, \hat{H} \, | \, \psi \, \rangle \geq \frac{m^2}{k_{\mathrm{B}}T} \, \frac{\langle \, \phi \, | \, X_\eta \, \rangle^2}{\langle \, \phi \, | \, \hat{H} \, | \, \phi \, \rangle} \; . \tag{5.179}$$

In order to get a good lower bound, we want ϕ in each case to have a good overlap with $X_{\kappa,\eta}$. One approach then is to take $\phi = X_{\kappa,\eta}$, which guarantees that the overlap will be finite (and not zero due to symmetry, for example). We illustrate this method with the viscosity calculation. We have

$$\eta \geq \frac{m^2}{k_{\mathrm{B}}T} \, \frac{\langle \, v_x v_y \, | \, v_x v_y \, \rangle^2}{\langle \, v_x v_y \, | \, \hat{H} \, | \, v_x v_y \, \rangle} \; . \tag{5.180}$$

Now the linearized collision operator $\hat{L}$ acts as

$$\langle \, \phi \, | \, \hat{L} \, | \, \psi \, \rangle = \int \! d^3 p \, g^0(\boldsymbol{p}) \, \phi(\boldsymbol{p}) \! \int \! d^3 p_1 \! \int \! d\Omega \, \frac{\partial \sigma}{\partial \Omega} \, |\boldsymbol{v} - \boldsymbol{v}_1| \, f^0(\boldsymbol{p}_1) \big\{ \psi(\boldsymbol{p}) + \psi(\boldsymbol{p}_1) - \psi(\boldsymbol{p}') - \psi(\boldsymbol{p}'_1) \big\} \; . \tag{5.181}$$

Here the kinematics of the collision guarantee total energy and momentum conservation, so $\boldsymbol{p}'$ and $\boldsymbol{p}'_1$ are determined as in eqn. 5.56.

Now we have

$$d\Omega = \sin \chi \, d\chi \, d\varphi \; , \tag{5.182}$$

where χ is the scattering angle depicted in Fig. 5.6 and φ is the azimuthal angle of the scattering. The differential scattering cross section is obtained by elementary mechanics and is known to be

$$\frac{\partial \sigma}{\partial \Omega} = \left| \frac{d(b^2/2)}{d \sin \chi} \right| \; , \tag{5.183}$$

where b is the *impact parameter*. The scattering angle is

$$\chi(b, u) = \pi - 2 \int\limits_{r_{\mathrm{p}}}^{\infty} dr \, \frac{b}{\sqrt{r^4 - b^2 r^2 - \frac{2U(r)r^4}{\tilde{m}u^2}}} \; , \tag{5.184}$$

where $\tilde{m} = \frac{1}{2}m$ is the reduced mass, and r_{p} is the relative coordinate separation at periapsis, *i.e.* the distance of closest approach, which occurs when $\dot{r} = 0$, *i.e.*

$$\tfrac{1}{2}\tilde{m}u^2 = \frac{\ell^2}{2\tilde{m}r_{\mathrm{p}}^2} + U(r_{\mathrm{p}}) \; , \tag{5.185}$$

where $\ell = \tilde{m}ub$ is the relative coordinate angular momentum.

We work in center-of-mass coordinates, so the velocities are

$$\boldsymbol{v} = \boldsymbol{V} + \tfrac{1}{2}\boldsymbol{u} \qquad\qquad \boldsymbol{v}' = \boldsymbol{V} + \tfrac{1}{2}\boldsymbol{u}' \tag{5.186}$$

$$\boldsymbol{v}_1 = \boldsymbol{V} - \tfrac{1}{2}\boldsymbol{u} \qquad\qquad \boldsymbol{v}'_1 = \boldsymbol{V} - \tfrac{1}{2}\boldsymbol{u}' \qquad\qquad , \tag{5.187}$$

with $|\boldsymbol{u}| = |\boldsymbol{u}'|$ and $\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{u}}' = \cos \chi$. Then if $\psi(\boldsymbol{p}) = v_x v_y$, we have

$$\Delta(\psi) \equiv \psi(\boldsymbol{p}) + \psi(\boldsymbol{p}_1) - \psi(\boldsymbol{p}') - \psi(\boldsymbol{p}'_1) = \tfrac{1}{2}\big(u_x u_y - u'_x u'_y \big) \; . \tag{5.188}$$
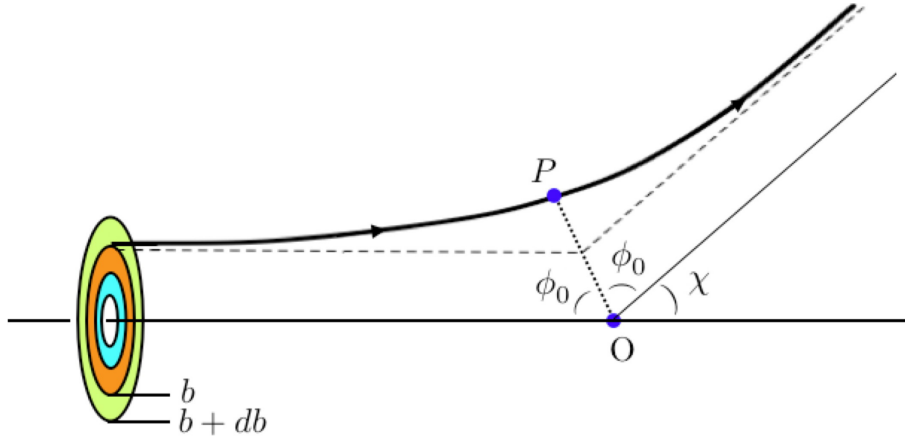
Figure 5.6: Scattering in the CM frame. O is the force center and P is the point of periapsis. The impact parameter is b, and χ is the scattering angle. ϕ_0 is the angle through which the relative coordinate moves between periapsis and infinity.

We may write

$$\boldsymbol{u}' = u \left(\sin\chi \cos\varphi \, \hat{\boldsymbol{e}}_1 + \sin\chi \sin\varphi \, \hat{\boldsymbol{e}}_2 + \cos\chi \, \hat{\boldsymbol{e}}_3 \right) , \tag{5.189}$$

where $\hat{\boldsymbol{e}}_3 = \hat{\boldsymbol{u}}$. With this parameterization, we have

$$\int_0^{2\pi} d\varphi \, \tfrac{1}{2} \left(u_\alpha u_\beta - u'_\alpha u'_\beta \right) = -\pi \sin^2\chi \left(u^2 \, \delta_{\alpha\beta} - 3 u_\alpha u_\beta \right) . \tag{5.190}$$

Note that we have used here the relation

$$e_{1\alpha} \, e_{1\beta} + e_{2\alpha} \, e_{2\beta} + e_{3\alpha} \, e_{3\beta} = \delta_{\alpha\beta} , \tag{5.191}$$

which holds since the LHS is a projector $\sum_{i=1}^3 |\hat{\boldsymbol{e}}_i\rangle\langle\hat{\boldsymbol{e}}_i|$.

It is convenient to define the following integral:

$$R(u) \equiv \int_0^\infty db \, b \, \sin^2\chi(b, u) . \tag{5.192}$$

Since the Jacobian

$$\left| \det \frac{(\partial \boldsymbol{v}, \partial \boldsymbol{v}_1)}{(\partial \boldsymbol{V}, \partial \boldsymbol{u})} \right| = 1 , \tag{5.193}$$

we have

$$\langle\, v_x v_y \,|\, \hat{L} \,|\, v_x v_y \,\rangle = n^2 \left(\frac{m}{2\pi k_{\mathrm{B}} T} \right)^3 \int d^3V \int d^3u \, e^{-m\boldsymbol{V}^2/k_{\mathrm{B}}T} \, e^{-m\boldsymbol{u}^2/4k_{\mathrm{B}}T} \cdot u \cdot \tfrac{3\pi}{2} u_x u_y \cdot R(u) \cdot v_x v_y . \tag{5.194}$$

This yields

$$\langle\, v_x v_y \,|\, \hat{L} \,|\, v_x v_y \,\rangle = \tfrac{\pi}{40} \, n^2 \, \big\langle u^5 \, R(u) \big\rangle , \tag{5.195}$$

where

$$\langle F(u) \rangle \equiv \int\limits_0^\infty du\, u^2\, e^{-mu^2/4k_{\mathrm B}T}\, F(u) \Big/ \int\limits_0^\infty du\, u^2\, e^{-mu^2/4k_{\mathrm B}T} \ . \tag{5.196}$$

It is easy to compute the term in the numerator of eqn. 5.180:

$$\langle\, v_x v_y \,|\, v_x v_y \,\rangle = n \left(\frac{m}{2\pi k_{\mathrm B}T} \right)^{3/2} \int d^3v\, e^{-m\boldsymbol{v}^2/2k_{\mathrm B}T}\, v_x^2\, v_y^2 = n \left(\frac{k_{\mathrm B}T}{m} \right)^2 \ . \tag{5.197}$$

Putting it all together, we find

$$\eta \geq \frac{40\,(k_{\mathrm B}T)^3}{\pi\, m^2} \Big/ \langle u^5\, R(u) \rangle \ . \tag{5.198}$$

The computation for κ is a bit more tedious. One has $\psi(\boldsymbol{p}) = (\varepsilon - c_p T)\, v_x$, in which case

$$\Delta(\psi) = \tfrac{1}{2}m \Big[(\boldsymbol{V} \cdot \boldsymbol{u})\, u_x - (\boldsymbol{V} \cdot \boldsymbol{u}')\, u_x' \Big] \ . \tag{5.199}$$

Ultimately, one obtains the lower bound

$$\kappa \geq \frac{150\, k_{\mathrm B}\, (k_{\mathrm B}T)^3}{\pi\, m^3} \Big/ \langle u^5\, R(u) \rangle \ . \tag{5.200}$$

Thus, independent of the potential, this variational calculation yields a Prandtl number of

$$\mathrm{Pr} = \frac{\nu}{a} = \frac{\eta\, c_p}{m\, \kappa} = \tfrac{2}{3} \ , \tag{5.201}$$

which is very close to what is observed in dilute monatomic gases (see Tab. 5.1).

While the variational expressions for η and κ are complicated functions of the potential, for hard sphere scattering the calculation is simple, because $b = d\sin\phi_0 = d\cos(\tfrac{1}{2}\chi)$, where d is the hard sphere diameter. Thus, the impact parameter b is independent of the relative speed u, and one finds $R(u) = \tfrac{1}{3}d^3$. Then

$$\langle u^5\, R(u) \rangle = \tfrac{1}{3}d^3 \langle u^5 \rangle = \frac{128}{\sqrt{\pi}} \left(\frac{k_{\mathrm B}T}{m} \right)^{5/2} d^2 \tag{5.202}$$

and one finds

$$\eta \geq \frac{5\,(mk_{\mathrm B}T)^{1/2}}{16\sqrt{\pi}\,d^2} \qquad , \qquad \kappa \geq \frac{75\, k_{\mathrm B}}{64\sqrt{\pi}\,d^2} \left(\frac{k_{\mathrm B}T}{m} \right)^{1/2} \ . \tag{5.203}$$

5.8   The Equations of Hydrodynamics

We now derive the equations governing fluid flow. The equations of mass and momentum balance are

$$\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\,\boldsymbol{V}) = 0 \tag{5.204}$$

$$\frac{\partial(\rho\,V_\alpha)}{\partial t} + \frac{\partial\Pi_{\alpha\beta}}{\partial x^\beta} = 0 \ , \tag{5.205}$$

where

$$\Pi_{\alpha\beta} = \rho \, V_\alpha V_\beta + p \, \delta_{\alpha\beta} - \overbrace{\left\{ \eta \left(\frac{\partial V_\alpha}{\partial x_\beta} + \frac{\partial V_\beta}{\partial x_\alpha} - \tfrac{2}{3} \, \boldsymbol{\nabla}\!\cdot\!\boldsymbol{V} \, \delta_{\alpha\beta} \right) + \zeta \, \boldsymbol{\nabla}\!\cdot\!\boldsymbol{V} \, \delta_{\alpha\beta} \right\}}^{\tilde{\sigma}_{\alpha\beta}} \; . \tag{5.206}$$

Substituting the continuity equation into the momentum balance equation, one arrives at

$$\rho \, \frac{\partial \boldsymbol{V}}{\partial t} + \rho \, (\boldsymbol{V}\!\cdot\!\boldsymbol{\nabla})\boldsymbol{V} = -\boldsymbol{\nabla}p + \eta \, \nabla^2 \boldsymbol{V} + (\zeta + \tfrac{1}{3}\eta)\boldsymbol{\nabla}(\boldsymbol{\nabla}\!\cdot\!\boldsymbol{V}) \,, \tag{5.207}$$

which, together with continuity, are known as the *Navier-Stokes equations*. These equations are supplemented by an equation describing the conservation of energy,

$$T \frac{\partial s}{\partial T} + T \, \boldsymbol{\nabla}\!\cdot\!(s\boldsymbol{V}) = \tilde{\sigma}_{\alpha\beta} \, \frac{\partial V_\alpha}{\partial x^\beta} + \boldsymbol{\nabla}\!\cdot\!(\kappa \boldsymbol{\nabla}T) \,. \tag{5.208}$$

Note that the LHS of eqn. 5.207 is $\rho \, D\boldsymbol{V}/Dt$, where D/Dt is the convective derivative. Multiplying by a differential volume, this gives the mass times the acceleration of a differential local fluid element. The RHS, multiplied by the same differential volume, gives the differential force on this fluid element in a frame instantaneously moving with constant velocity $\boldsymbol{V}$. Thus, this is Newton's Second Law for the fluid.

5.9 Nonequilibrium Quantum Transport

5.9.1 Boltzmann equation for quantum systems

Almost everything we have derived thus far can be applied, *mutatis mutandis*, to quantum systems. The main difference is that the distribution f^0 corresponding to local equilibrium is no longer of the Maxwell-Boltzmann form, but rather of the Bose-Einstein or Fermi-Dirac form,

$$f^0(\boldsymbol{r}, \boldsymbol{k}, t) = \left\{ \exp\left(\frac{\varepsilon(\boldsymbol{k}) - \mu(\boldsymbol{r}, t)}{k_{\mathrm{B}} T(\boldsymbol{r}, t)} \right) \mp 1 \right\}^{-1} \,, \tag{5.209}$$

where the top sign applies to bosons and the bottom sign to fermions. Here we shift to the more common notation for quantum systems in which we write the distribution in terms of the wavevector $\boldsymbol{k} = \boldsymbol{p}/\hbar$ rather than the momentum $\boldsymbol{p}$. The quantum distributions satisfy detailed balance with respect to the quantum collision integral

$$\left(\frac{\partial f}{\partial t} \right)_{\mathrm{coll}} = \int\!\frac{d^3k_1}{(2\pi)^3}\!\int\!\frac{d^3k'}{(2\pi)^3}\!\int\!\frac{d^3k_1'}{(2\pi)^3} \; w \left\{ f'f_1' \, (1 \pm f) \, (1 \pm f_1) - f f_1 \, (1 \pm f') \, (1 \pm f_1') \right\} \tag{5.210}$$

where $w = w(\boldsymbol{k}, \boldsymbol{k}_1 \,|\, \boldsymbol{k}', \boldsymbol{k}_1')$, $f = f(\boldsymbol{k})$, $f_1 = f(\boldsymbol{k}_1)$, $f' = f(\boldsymbol{k}')$, and $f_1' = f(\boldsymbol{k}_1')$, and where we have assumed time-reversal and parity symmetry. Detailed balance requires

$$\frac{f}{1 \pm f} \cdot \frac{f_1}{1 \pm f_1} = \frac{f'}{1 \pm f'} \cdot \frac{f_1'}{1 \pm f_1'} \,, \tag{5.211}$$

where $f = f^0$ is the equilibrium distribution. One can check that

$$f = \frac{1}{e^{\beta(\varepsilon - \mu)} \mp 1} \qquad \Longrightarrow \qquad \frac{f}{1 \pm f} = e^{\beta(\mu - \varepsilon)} \,, \tag{5.212}$$

which is the Boltzmann distribution, which we have already shown to satisfy detailed balance. For the streaming term, we have

$$
\begin{aligned}
df^0 &= k_{\mathrm{B}} T \, \frac{\partial f^0}{\partial \varepsilon} \, d\!\left(\frac{\varepsilon - \mu}{k_{\mathrm{B}} T} \right) \\
&= k_{\mathrm{B}} T \, \frac{\partial f^0}{\partial \varepsilon} \left\{ -\frac{d\mu}{k_{\mathrm{B}} T} - \frac{(\varepsilon - \mu)\, dT}{k_{\mathrm{B}} T^2} + \frac{d\varepsilon}{k_{\mathrm{B}} T} \right\} \\
&= -\frac{\partial f^0}{\partial \varepsilon} \left\{ \frac{\partial \mu}{\partial \boldsymbol{r}} \cdot d\boldsymbol{r} + \frac{\varepsilon - \mu}{T} \frac{\partial T}{\partial \boldsymbol{r}} \cdot d\boldsymbol{r} - \frac{\partial \varepsilon}{\partial \boldsymbol{k}} \cdot d\boldsymbol{k} \right\} \,,
\end{aligned}
\tag{5.213}
$$

from which we read off

$$
\begin{aligned}
\frac{\partial f^0}{\partial \boldsymbol{r}} &= -\frac{\partial f^0}{\partial \varepsilon} \left\{ \frac{\partial \mu}{\partial \boldsymbol{r}} + \frac{\varepsilon - \mu}{T} \frac{\partial T}{\partial \boldsymbol{r}} \right\} \\
\frac{\partial f^0}{\partial \boldsymbol{k}} &= \hbar \boldsymbol{v} \, \frac{\partial f^0}{\partial \varepsilon} \,.
\end{aligned}
\tag{5.214}
$$

The most important application is to the theory of electron transport in metals and semiconductors, in which case f^0 is the Fermi distribution. In this case, the quantum collision integral also receives a contribution from one-body scattering in the presence of an external potential $U(\boldsymbol{r})$, which is given by Fermi's Golden Rule:

$$
\begin{aligned}
\left(\frac{\partial f(\boldsymbol{k})}{\partial t} \right)'_{\mathrm{coll}} &= \frac{2\pi}{\hbar} \sum_{\boldsymbol{k}' \in \hat{\Omega}} |\langle\, \boldsymbol{k}' \,|\, U \,|\, \boldsymbol{k} \,\rangle|^2 \left(f(\boldsymbol{k}') - f(\boldsymbol{k}) \right) \delta\!\left(\varepsilon(\boldsymbol{k}) - \varepsilon(\boldsymbol{k}') \right) \\
&= \frac{2\pi}{\hbar V} \int_{\hat{\Omega}} \frac{d^3 k}{(2\pi)^3} \, |\hat{U}(\boldsymbol{k} - \boldsymbol{k}')|^2 \left(f(\boldsymbol{k}') - f(\boldsymbol{k}) \right) \delta\!\left(\varepsilon(\boldsymbol{k}) - \varepsilon(\boldsymbol{k}') \right) \,.
\end{aligned}
\tag{5.215}
$$

The wavevectors are now restricted to the first Brillouin zone, and the dispersion $\varepsilon(\boldsymbol{k})$ is no longer the ballistic form $\varepsilon = \hbar^2 \boldsymbol{k}^2 / 2m$ but rather the dispersion for electrons in a particular energy band (typically the valence band) of a solid[10]. Note that $f = f^0$ satisfies detailed balance with respect to one-body collisions as well[11].

In the presence of a weak electric field $\boldsymbol{E}$ and a (not necessarily weak) magnetic field $\boldsymbol{B}$, we have, within the relaxation time approximation, $f = f^0 + \delta f$ with

$$\frac{\partial \, \delta f}{\partial t} - \frac{e}{\hbar c} \, \boldsymbol{v} \times \boldsymbol{B} \cdot \frac{\partial \, \delta f}{\partial \boldsymbol{k}} - \boldsymbol{v} \cdot \left[e\, \boldsymbol{\mathcal{E}} + \frac{\varepsilon - \mu}{T} \, \boldsymbol{\nabla} T \right] \frac{\partial f^0}{\partial \varepsilon} = -\frac{\delta f}{\tau} \,, \tag{5.216}$$

---

[10]We neglect interband scattering here, which can be important in practical applications, but which is beyond the scope of these notes.

[11]The transition rate from $|\boldsymbol{k}'\rangle$ to $|\boldsymbol{k}\rangle$ is proportional to the matrix element and to the product $f'(1 - f)$. The reverse process is proportional to $f(1 - f')$. Subtracting these factors, one obtains $f' - f$, and therefore the nonlinear terms felicitously cancel in eqn. 5.215.

where $\mathcal{E} = -\boldsymbol{\nabla}(\phi - \mu/e) = \boldsymbol{E} - e^{-1}\boldsymbol{\nabla}\mu$ is the gradient of the 'electrochemical potential' $\phi - e^{-1}\mu$. In deriving the above equation, we have worked to lowest order in small quantities. This entails dropping terms like $\boldsymbol{v} \cdot \frac{\partial\,\delta f}{\partial r}$ (higher order in spatial derivatives) and $\boldsymbol{E} \cdot \frac{\partial\,\delta f}{\partial k}$ (both $\boldsymbol{E}$ and δf are assumed small). Typically τ is energy-dependent, *i.e.* $\tau = \tau\big(\varepsilon(\boldsymbol{k})\big)$.

We can use eqn. 5.216 to compute the electrical current $\boldsymbol{j}$ and the thermal current $\boldsymbol{j}_q$,

$$\boldsymbol{j} = -2e\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\,\boldsymbol{v}\,\delta f \tag{5.217}$$

$$\boldsymbol{j}_q = 2\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\,(\varepsilon - \mu)\,\boldsymbol{v}\,\delta f\ . \tag{5.218}$$

Here the factor of 2 is from spin degeneracy of the electrons (we neglect Zeeman splitting).

In the presence of a time-independent temperature gradient and electric field, linearized Boltzmann equation in the relaxation time approximation has the solution

$$\delta f = -\tau(\varepsilon)\,\boldsymbol{v}\cdot\left(e\mathcal{E} + \frac{\varepsilon-\mu}{T}\,\boldsymbol{\nabla}T\right)\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\ . \tag{5.219}$$

We now consider both the electrical current[12] $\boldsymbol{j}$ as well as the thermal current density $\boldsymbol{j}_q$. One readily obtains

$$\boldsymbol{j} = -2e\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\,\boldsymbol{v}\,\delta f \equiv L_{11}\,\mathcal{E} - L_{12}\,\boldsymbol{\nabla}T \tag{5.220}$$

$$\boldsymbol{j}_q = 2\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\,(\varepsilon - \mu)\,\boldsymbol{v}\,\delta f \equiv L_{21}\,\mathcal{E} - L_{22}\,\boldsymbol{\nabla}T \tag{5.221}$$

where the *transport coefficients* L^{11} *etc.* are matrices:

$$L_{11}^{\alpha\beta} = \frac{e^2}{4\pi^3\hbar}\int d\varepsilon\,\tau(\varepsilon)\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\int dS_\varepsilon\,\frac{v^\alpha v^\beta}{|\boldsymbol{v}|} \tag{5.222}$$

$$L_{21}^{\alpha\beta} = TL_{12}^{\alpha\beta} = -\frac{e}{4\pi^3\hbar}\int d\varepsilon\,\tau(\varepsilon)\,(\varepsilon-\mu)\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\int dS_\varepsilon\,\frac{v^\alpha v^\beta}{|\boldsymbol{v}|} \tag{5.223}$$

$$L_{22}^{\alpha\beta} = \frac{1}{4\pi^3\hbar T}\int d\varepsilon\,\tau(\varepsilon)\,(\varepsilon-\mu)^2\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\int dS_\varepsilon\,\frac{v^\alpha v^\beta}{|\boldsymbol{v}|}\ . \tag{5.224}$$
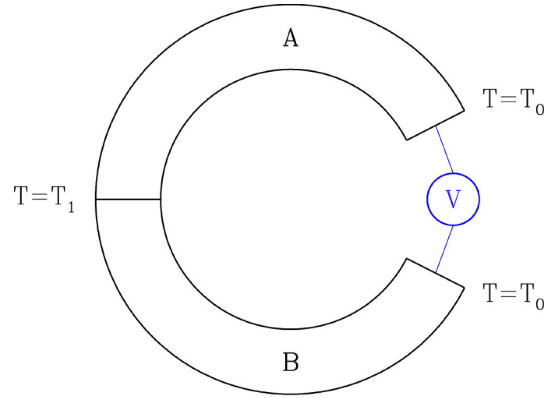
If we define the hierarchy of integral expressions

$$\mathcal{J}_n^{\alpha\beta} \equiv \frac{1}{4\pi^3\hbar}\int d\varepsilon\,\tau(\varepsilon)\,(\varepsilon-\mu)^n\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\int dS_\varepsilon\,\frac{v^\alpha v^\beta}{|\boldsymbol{v}|} \tag{5.225}$$

then we may write

$$L_{11}^{\alpha\beta} = e^2\,\mathcal{J}_0^{\alpha\beta}\quad,\quad L_{21}^{\alpha\beta} = TL_{12}^{\alpha\beta} = -e\,\mathcal{J}_1^{\alpha\beta}\quad,\quad L_{22}^{\alpha\beta} = \frac{1}{T}\,\mathcal{J}_2^{\alpha\beta}\ . \tag{5.226}$$

---

[12]In this section we use $\boldsymbol{j}$ to denote electrical current, rather than particle number current as before.

$$V \;=\; (Q_{\mathrm{B}}-Q_{\mathrm{A}})(T_1-T_0)$$

Figure 5.7: A thermocouple is a junction formed of two dissimilar metals. With no electrical current passing, an electric field is generated in the presence of a temperature gradient, resulting in a voltage $V = V_{\mathrm{A}} - V_{\mathrm{B}}$.

The linear relations in eqn. (5.221) may be recast in the following form:

$$
\begin{aligned}
\boldsymbol{\mathcal{E}} &= \rho \, \boldsymbol{j} + Q \, \boldsymbol{\nabla} T \\
\boldsymbol{j}_q &= \sqcap \, \boldsymbol{j} - \kappa \, \boldsymbol{\nabla} T \; ,
\end{aligned}
\tag{5.227}
$$

where the matrices ρ, Q, $\sqcap$, and κ are given by

$$\rho = L_{11}^{-1} \qquad\qquad\qquad Q = L_{11}^{-1} \, L_{12} \tag{5.228}$$

$$\sqcap = L_{21} \, L_{11}^{-1} \qquad\qquad \kappa = L_{22} - L_{21} \, L_{11}^{-1} \, L_{12} \; , \tag{5.229}$$

or, in terms of the $\mathcal{J}_n$,

$$\rho = \frac{1}{e^2} \, \mathcal{J}_0^{-1} \qquad\qquad\qquad Q = -\frac{1}{e \, T} \, \mathcal{J}_0^{-1} \, \mathcal{J}_1 \tag{5.230}$$

$$\sqcap = -\frac{1}{e} \, \mathcal{J}_1 \, \mathcal{J}_0^{-1} \qquad\qquad \kappa = \frac{1}{T} \left(\mathcal{J}_2 - \mathcal{J}_1 \, \mathcal{J}_0^{-1} \, \mathcal{J}_1 \right) \; , \tag{5.231}$$

These equations describe a wealth of transport phenomena:

- Electrical resistance ($\boldsymbol{\nabla} T = \boldsymbol{B} = 0$)
  An electrical current $\boldsymbol{j}$ will generate an electric field $\boldsymbol{\mathcal{E}} = \rho \boldsymbol{j}$, where ρ is the electrical resistivity.

- Peltier effect ($\boldsymbol{\nabla} T = \boldsymbol{B} = 0$)
  An electrical current $\boldsymbol{j}$ will generate an heat current $\boldsymbol{j}_q = \sqcap \boldsymbol{j}$, where $\sqcap$ is the Peltier coefficient.

- Thermal conduction ($\boldsymbol{j} = \boldsymbol{B} = 0$)
  A temperature gradient $\boldsymbol{\nabla} T$ gives rise to a heat current $\boldsymbol{j}_q = -\kappa \boldsymbol{\nabla} T$, where κ is the thermal conductivity.

- Seebeck effect ($j = B = 0$)
  A temperature gradient ∇T gives rise to an electric field $\mathcal{E} = Q\nabla T$, where Q is the Seebeck coefficient.

One practical way to measure the thermopower is to form a junction between two dissimilar metals, A and B. The junction is held at temperature T_1 and the other ends of the metals are held at temperature T_0. One then measures a voltage difference between the free ends of the metals – this is known as the Seebeck effect. Integrating the electric field from the free end of A to the free end of B gives

$$V_A - V_B = -\int_A^B \mathcal{E} \cdot d\boldsymbol{l} = (Q_B - Q_A)(T_1 - T_0) \,. \tag{5.232}$$

What one measures here is really the difference in thermopowers of the two metals. For an absolute measurement of Q_A, replace B by a superconductor ($Q = 0$ for a superconductor). A device which converts a temperature gradient into an emf is known as a *thermocouple*.

The Peltier effect has practical applications in refrigeration technology. Suppose an electrical current I is passed through a junction between two dissimilar metals, A and B. Due to the difference in Peltier coefficients, there will be a net heat current into the junction of $W = (\Pi_A - \Pi_B)\,I$. Note that this is proportional to I, rather than the familiar I^2 result from Joule heating. The sign of W depends on the direction of the current. If a second junction is added, to make an ABA configuration, then heat absorbed at the first junction will be liberated at the second. [13]

5.9.2 The Heat Equation

We begin with the continuity equations for charge density ρ and energy density ε:

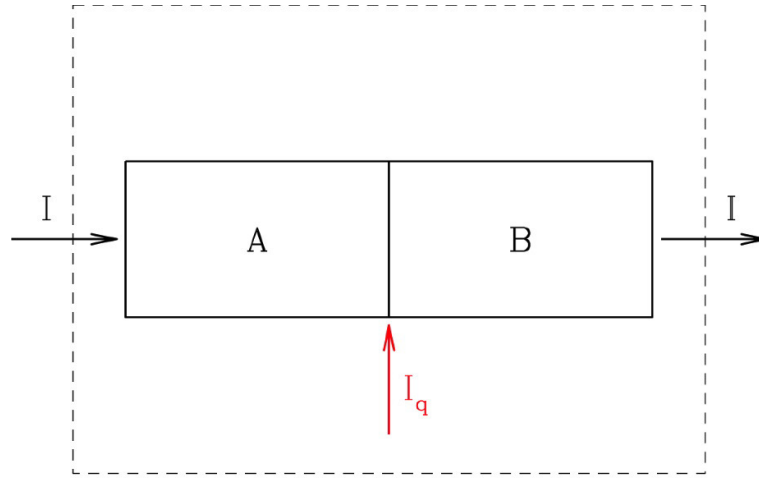$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{j} = 0 \tag{5.233}$$

$$\frac{\partial \varepsilon}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{j}_\varepsilon = \boldsymbol{j} \cdot \boldsymbol{E} \,, \tag{5.234}$$

where $\boldsymbol{E}$ is the electric field[14]. Now we invoke local thermodynamic equilibrium and write

$$\begin{aligned}
\frac{\partial \varepsilon}{\partial t} &= \frac{\partial \varepsilon}{\partial n}\frac{\partial n}{\partial t} + \frac{\partial \varepsilon}{\partial T}\frac{\partial T}{\partial t} \\
&= -\frac{\mu}{e}\frac{\partial \rho}{\partial t} + c_V \frac{\partial T}{\partial t} \,,
\end{aligned} \tag{5.235}$$

---

[13]To create a refrigerator, stick the cold junction inside a thermally insulated box and the hot junction outside the box.
[14]Note that it is $\boldsymbol{E} \cdot \boldsymbol{j}$ and *not* $\mathcal{E} \cdot \boldsymbol{j}$ which is the source term in the energy continuity equation.

$$I_q = (\Pi_B - \Pi_A) \cdot I$$

Figure 5.8: A sketch of a Peltier effect refrigerator. An electrical current I is passed through a junction between two dissimilar metals. If the dotted line represents the boundary of a thermally well-insulated body, then the body cools when $\Pi_B > \Pi_A$, in order to maintain a heat current balance at the junction.

where n is the electron *number density* ($n = -\rho/e$) and c_V is the specific heat. We may now write

$$
\begin{aligned}
c_V \frac{\partial T}{\partial t} &= \frac{\partial \varepsilon}{\partial t} + \frac{\mu}{e} \frac{\partial \rho}{\partial t} \\
&= \boldsymbol{j} \cdot \boldsymbol{E} - \boldsymbol{\nabla} \cdot \boldsymbol{j}_\varepsilon - \frac{\mu}{e} \boldsymbol{\nabla} \cdot \boldsymbol{j} \\
&= \boldsymbol{j} \cdot \boldsymbol{\mathcal{E}} - \boldsymbol{\nabla} \cdot \boldsymbol{j}_q \, .
\end{aligned}
\tag{5.236}
$$

Invoking $\boldsymbol{j}_q = \Pi \boldsymbol{j} - \kappa \boldsymbol{\nabla} T$, we see that if there is no electrical current ($\boldsymbol{j} = 0$), we obtain the *heat equation*

$$
c_V \frac{\partial T}{\partial t} = \kappa_{\alpha\beta} \frac{\partial^2 T}{\partial x^\alpha \, \partial x^\beta} \, .
\tag{5.237}
$$

This results in a time scale τ_T for temperature diffusion $\tau_T = \mathcal{C} L^2 c_V / \kappa$, where L is a typical length scale and $\mathcal{C}$ is a numerical constant. For a cube of size L subjected to a sudden external temperature change, L is the side length and $\mathcal{C} = 1/3\pi^2$ (solve by separation of variables).

5.9.3 Calculation of Transport Coefficients

We will henceforth assume that sufficient crystalline symmetry exists (*e.g.* cubic symmetry) to render all the transport coefficients multiples of the identity matrix. Under such conditions, we may write $\mathcal{J}_n^{\alpha\beta} = \mathcal{J}_n \, \delta_{\alpha\beta}$ with

$$
\mathcal{J}_n = \frac{1}{12\pi^3 \hbar} \int d\varepsilon \, \tau(\varepsilon) \, (\varepsilon - \mu)^n \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \int dS_\varepsilon \, |\boldsymbol{v}| \, .
\tag{5.238}
$$

The low-temperature behavior is extracted using the Sommerfeld expansion,

$$
\mathcal{I} \equiv \int\limits_{-\infty}^{\infty} d\varepsilon \, H(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon} \right) = \pi \mathcal{D} \csc(\pi \mathcal{D}) \, H(\varepsilon) \Big|_{\varepsilon=\mu}
\tag{5.239}
$$

$$
= H(\mu) + \frac{\pi^2}{6} (k_{\mathrm{B}} T)^2 \, H''(\mu) + \dots
\tag{5.240}
$$

where $\mathcal{D} \equiv k_{\mathrm{B}} T \frac{\partial}{\partial \varepsilon}$ is a dimensionless differential operator.[15]

Let us now perform some explicit calculations in the case of a parabolic band with an energy-independent scattering time τ. In this case, one readily finds

$$
\mathcal{J}_n = \frac{\sigma_0}{e^2} \, \mu^{-3/2} \, \pi \mathcal{D} \csc \pi \mathcal{D} \, \varepsilon^{3/2} \, (\varepsilon - \mu)^n \Big|_{\varepsilon=\mu} \, ,
\tag{5.241}
$$

where $\sigma_0 = ne^2\tau/m^*$. Thus,

$$
\begin{aligned}
\mathcal{J}_0 &= \frac{\sigma_0}{e^2} \left[1 + \frac{\pi^2}{8} \frac{(k_{\mathrm{B}} T)^2}{\mu^2} + \dots \right] \\
\mathcal{J}_1 &= \frac{\sigma_0}{e^2} \frac{\pi^2}{2} \frac{(k_{\mathrm{B}} T)^2}{\mu} + \dots \\
\mathcal{J}_2 &= \frac{\sigma_0}{e^2} \frac{\pi^2}{3} (k_{\mathrm{B}} T)^2 + \dots \, ,
\end{aligned}
\tag{5.242}
$$

from which we obtain the low-T results $\rho = \sigma_0^{-1}$,

$$
Q = -\frac{\pi^2}{2} \frac{k_{\mathrm{B}}^2 T}{e \, \varepsilon_{\mathrm{F}}} \qquad \kappa = \frac{\pi^2}{3} \frac{n\tau}{m^*} k_{\mathrm{B}}^2 T \, ,
\tag{5.243}
$$

and of course $\sqcap = TQ$. The predicted universal ratio

$$
\frac{\kappa}{\sigma T} = \frac{\pi^2}{3} (k_{\mathrm{B}}/e)^2 = 2.45 \times 10^{-8} \, \mathrm{V^2 \, K^{-2}} \, ,
\tag{5.244}
$$

is known as the *Wiedemann-Franz law*. Note also that our result for the thermopower is unambiguously negative. In actuality, several nearly free electron metals have positive low-temperature thermopowers (Cs and Li, for example). What went wrong? We have neglected electron-phonon scattering!

5.9.4 Onsager Relations

Transport phenomena are described in general by a set of linear relations,

$$
J_i = L_{ik} \, F_k \, ,
\tag{5.245}
$$

---

[15]Remember that physically the fixed quantities are temperature and total carrier number density (or charge density, in the case of electron and hole bands), and *not* temperature and chemical potential. An equation of state relating n, μ, and T is then inverted to obtain $\mu(n, T)$, so that all results ultimately may be expressed in terms of n and T.

where the $\{F_k\}$ are *generalized forces* and the $\{J_i\}$ are *generalized currents*. Moreover, to each force F_i corresponds a unique conjugate current J_i, such that the rate of internal entropy production is

$$\dot{S} = \sum_i F_i \, J_i \quad \Longrightarrow \quad F_i = \frac{\partial \dot{S}}{\partial J_i} \; . \tag{5.246}$$

The Onsager relations (also known as *Onsager reciprocity*) state that

$$L_{ik}(\boldsymbol{B}) = \eta_i \, \eta_k \, L_{ki}(-\boldsymbol{B}) \; , \tag{5.247}$$

where η_i describes the parity of J_i under time reversal:

$$J_i^T = \eta_i \, J_i \; , \tag{5.248}$$

where J_i^T is the time reverse of J_i. To justify the Onsager relations requires a microscopic description of our nonequilibrium system.

The Onsager relations have some remarkable consequences. For example, they require, for $\boldsymbol{B} = 0$, that the thermal conductivity tensor κ_{ij} of any crystal must be symmetric, independent of the crystal structure. In general,this result does not follow from considerations of crystalline symmetry. It also requires that for every 'off-diagonal' transport phenomenon, *e.g.* the Seebeck effect, there exists a distinct corresponding phenomenon, *e.g.* the Peltier effect.

For the transport coefficients studied, Onsager reciprocity means that in the presence of an external magnetic field,

$$\rho_{\alpha\beta}(\boldsymbol{B}) = \rho_{\beta\alpha}(-\boldsymbol{B}) \tag{5.249}$$

$$\kappa_{\alpha\beta}(\boldsymbol{B}) = \kappa_{\beta\alpha}(-\boldsymbol{B}) \tag{5.250}$$

$$\sqcap_{\alpha\beta}(\boldsymbol{B}) = T \, Q_{\beta\alpha}(-\boldsymbol{B}) \; . \tag{5.251}$$

Let's consider an isotropic system in a weak magnetic field, and expand the transport coefficients to first order in $\boldsymbol{B}$:

$$\rho_{\alpha\beta}(\boldsymbol{B}) = \rho \, \delta_{\alpha\beta} + \nu \, \epsilon_{\alpha\beta\gamma} \, B^\gamma \tag{5.252}$$

$$\kappa_{\alpha\beta}(\boldsymbol{B}) = \kappa \, \delta_{\alpha\beta} + \varpi \, \epsilon_{\alpha\beta\gamma} \, B^\gamma \tag{5.253}$$

$$Q_{\alpha\beta}(\boldsymbol{B}) = Q \, \delta_{\alpha\beta} + \zeta \, \epsilon_{\alpha\beta\gamma} \, B^\gamma \tag{5.254}$$

$$\sqcap_{\alpha\beta}(\boldsymbol{B}) = \sqcap \, \delta_{\alpha\beta} + \theta \, \epsilon_{\alpha\beta\gamma} B^\gamma \; . \tag{5.255}$$

Onsager reciprocity requires $\sqcap = T \, Q$ and $\theta = T \, \zeta$. We can now write

$$\boldsymbol{\mathcal{E}} = \rho \, \boldsymbol{j} + \nu \, \boldsymbol{j} \times \boldsymbol{B} + Q \, \boldsymbol{\nabla} T + \zeta \, \boldsymbol{\nabla} T \times \boldsymbol{B} \tag{5.256}$$

$$\boldsymbol{j}_q = \sqcap \, \boldsymbol{j} + \theta \, \boldsymbol{j} \times \boldsymbol{B} - \kappa \, \boldsymbol{\nabla} T - \varpi \, \boldsymbol{\nabla} T \times \boldsymbol{B} \; . \tag{5.257}$$

There are several new phenomena lurking:

- **Hall effect** ($\frac{\partial T}{\partial x} = \frac{\partial T}{\partial y} = j_y = 0$)

  An electrical current $\boldsymbol{j} = j_x \,\hat{\boldsymbol{x}}$ and a field $\boldsymbol{B} = B_z \,\hat{\boldsymbol{z}}$ yield an electric field $\boldsymbol{\mathcal{E}}$. The Hall coefficient is $R_\mathrm{H} = \mathcal{E}_y / j_x \, B_z = -\nu$.

- **Ettingshausen effect** ($\frac{\partial T}{\partial x} = j_y = j_{q,y} = 0$)

  An electrical current $\boldsymbol{j} = j_x \,\hat{\boldsymbol{x}}$ and a field $\boldsymbol{B} = B_z \,\hat{\boldsymbol{z}}$ yield a temperature gradient $\frac{\partial T}{\partial y}$. The Ettingshausen coefficient is $P = \frac{\partial T}{\partial y} / j_x \, B_z = -\theta / \kappa$.

- **Nernst effect** ($j_x = j_y = \frac{\partial T}{\partial y} = 0$)

  A temperature gradient $\boldsymbol{\nabla} T = \frac{\partial T}{\partial x} \,\hat{\boldsymbol{x}}$ and a field $\boldsymbol{B} = B_z \,\hat{\boldsymbol{z}}$ yield an electric field $\boldsymbol{\mathcal{E}}$. The Nernst coefficient is $\Lambda = \mathcal{E}_y / \frac{\partial T}{\partial x} \, B_z = -\zeta$.

- **Righi-Leduc effect** ($j_x = j_y = \mathcal{E}_y = 0$)

  A temperature gradient $\boldsymbol{\nabla} T = \frac{\partial T}{\partial x} \,\hat{\boldsymbol{x}}$ and a field $\boldsymbol{B} = B_z \,\hat{\boldsymbol{z}}$ yield an orthogonal temperature gradient $\frac{\partial T}{\partial y}$. The Righi-Leduc coefficient is $\mathcal{L} = \frac{\partial T}{\partial y} / \frac{\partial T}{\partial x} \, B_z = \zeta / Q$.

5.10 Appendix : Boltzmann Equation and Collisional Invariants

Problem : The linearized Boltzmann operator $L\psi$ is a complicated functional. Suppose we replace L by $\mathcal{L}$, where

$$
\begin{aligned}
\mathcal{L}\psi = -\gamma\,\psi(\boldsymbol{v},t) + \gamma \left(\frac{m}{2\pi k_\mathrm{B} T} \right)^{3/2} \int d^3\!u \, \exp\left(-\frac{m\boldsymbol{u}^2}{2k_\mathrm{B} T} \right) \\
\times \left\{ 1 + \frac{m}{k_\mathrm{B} T}\,\boldsymbol{u}\cdot\boldsymbol{v} + \frac{2}{3}\left(\frac{m\boldsymbol{u}^2}{2k_\mathrm{B} T} - \frac{3}{2} \right)\left(\frac{m\boldsymbol{v}^2}{2k_\mathrm{B} T} - \frac{3}{2} \right) \right\} \psi(\boldsymbol{u},t) \, .
\end{aligned}
\tag{5.258}
$$

Show that $\mathcal{L}$ shares all the important properties of L. What is the meaning of γ? Expand $\psi(\boldsymbol{v},t)$ in spherical harmonics and Sonine polynomials,

$$
\psi(\boldsymbol{v},t) = \sum_{r\ell m} a_{r\ell m}(t)\, S^r_{\ell+\frac{1}{2}}(x)\, x^{\ell/2}\, Y^\ell_m(\hat{\boldsymbol{n}}),
\tag{5.259}
$$

with $x = mv^2 / 2k_\mathrm{B} T$, and thus express the action of the linearized Boltzmann operator algebraically on the expansion coefficients $a_{r\ell m}(t)$.

The Sonine polynomials $S^n_\alpha(x)$ are a complete, orthogonal set which are convenient to use in the calculation of transport coefficients. They are defined as

$$
S^n_\alpha(x) = \sum_{m=0}^{n} \frac{\Gamma(\alpha + n + 1)\,(-x)^m}{\Gamma(\alpha + m + 1)\,(n - m)!\,m!} \, ,
\tag{5.260}
$$

and satisfy the generalized orthogonality relation

$$
\int\limits_0^\infty dx\, e^{-x}\, x^\alpha\, S^n_\alpha(x)\, S^{n'}_\alpha(x) = \frac{\Gamma(\alpha + n + 1)}{n!}\, \delta_{nn'} \, .
\tag{5.261}
$$

Solution : The 'important properties' of L are that it annihilate the five collisional invariants, *i.e.* 1, $\boldsymbol{v}$, and v^2, and that all other eigenvalues are *negative*. That this is true for $\mathcal{L}$ can be verified by an explicit calculation.

Plugging the conveniently parameterized form of $\psi(\boldsymbol{v}, t)$ into $\mathcal{L}$, we have

$$
\begin{aligned}
\mathcal{L}\psi = -\gamma \sum_{r\ell m} a_{r\ell m}(t)\, S^r_{\ell+\frac{1}{2}}(x)\, x^{\ell/2}\, Y^\ell_m(\hat{\boldsymbol{n}}) \;+\; \frac{\gamma}{2\pi^{3/2}} \sum_{r\ell m} a_{r\ell m}(t) \int_0^\infty dx_1\, x_1^{1/2}\, e^{-x_1} \\
\times \int d\hat{\boldsymbol{n}}_1 \Big[1 + 2\, x^{1/2} x_1^{1/2}\, \hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{n}}_1 + \tfrac{2}{3}\big(x - \tfrac{3}{2}\big)\big(x_1 - \tfrac{3}{2}\big) \Big] S^r_{\ell+\frac{1}{2}}(x_1)\, x_1^{\ell/2}\, Y^\ell_m(\hat{\boldsymbol{n}}_1)\,,
\end{aligned}
\tag{5.262}
$$

where we've used

$$
u = \sqrt{\frac{2k_{\mathrm{B}}T}{m}}\, x_1^{1/2} \qquad,\qquad du = \sqrt{\frac{k_{\mathrm{B}}T}{2m}}\, x_1^{-1/2}\, dx_1\,.
\tag{5.263}
$$

Now recall $Y^0_0(\hat{\boldsymbol{n}}) = \frac{1}{\sqrt{4\pi}}$ and

$$
Y^1_1(\hat{\boldsymbol{n}}) = -\sqrt{\frac{3}{8\pi}}\, \sin\theta\, e^{i\varphi} \qquad Y^1_0(\hat{\boldsymbol{n}}) = \sqrt{\frac{3}{4\pi}}\, \cos\theta \qquad Y^1_{-1}(\hat{\boldsymbol{n}}) = +\sqrt{\frac{3}{8\pi}}\, \sin\theta\, e^{-i\varphi}
$$

$$
S^0_{1/2}(x) = 1 \qquad\qquad S^0_{3/2}(x) = 1 \qquad\qquad S^1_{1/2}(x) = \tfrac{3}{2} - x\,,
$$

which allows us to write

$$
1 = 4\pi\, Y^0_0(\hat{\boldsymbol{n}})\, Y^{0*}_0(\hat{\boldsymbol{n}}_1)
\tag{5.264}
$$

$$
\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{n}}_1 = \frac{4\pi}{3}\Big[Y^1_0(\hat{\boldsymbol{n}})\, Y^{1*}_0(\hat{\boldsymbol{n}}_1) + Y^1_1(\hat{\boldsymbol{n}})\, Y^{1*}_1(\hat{\boldsymbol{n}}_1) + Y^1_{-1}(\hat{\boldsymbol{n}})\, Y^{1*}_{-1}(\hat{\boldsymbol{n}}_1) \Big]\,.
\tag{5.265}
$$

We can do the integrals by appealing to the orthogonality relations for the spherical harmonics and Sonine polynomials:

$$
\int d\hat{\boldsymbol{n}}\, Y^\ell_m(\hat{\boldsymbol{n}})\, Y^{l'\,*}_{m'}(\hat{\boldsymbol{n}}) = \delta_{ll'}\, \delta_{mm'}
\tag{5.266}
$$

$$
\int_0^\infty dx\, e^{-x}\, x^\alpha\, S^n_\alpha(x)\, S^{n'}_\alpha(x) = \frac{\Gamma(n+\alpha+1)}{\Gamma(n+1)}\, \delta_{nn'}\,.
\tag{5.267}
$$

Integrating first over the direction vector $\hat{\boldsymbol{n}}_1$,

$$
\begin{aligned}
\mathcal{L}\psi = &-\gamma \sum_{r\ell m} a_{r\ell m}(t)\, S^r_{\ell+\frac{1}{2}}(x)\, x^{\ell/2}\, Y^\ell_m(\hat{\boldsymbol{n}}) \\
&+ \frac{2\gamma}{\sqrt{\pi}} \sum_{r\ell m} a_{r\ell m}(t) \int_0^\infty dx_1\, x_1^{1/2}\, e^{-x_1} \int d\hat{\boldsymbol{n}}_1 \left[Y^0_0(\hat{\boldsymbol{n}})\, Y^{0*}_0(\hat{\boldsymbol{n}}_1)\, S^0_{1/2}(x)\, S^0_{1/2}(x_1) \right. \\
&\qquad + \tfrac{2}{3}\, x^{1/2} x_1^{1/2} \sum_{m'=-1}^{1} Y^1_{m'}(\hat{\boldsymbol{n}})\, Y^{1\,*}_{m'}(\hat{\boldsymbol{n}}_1)\, S^0_{3/2}(x)\, S^0_{3/2}(x_1) \\
&\qquad \left. + \tfrac{2}{3}\, Y^0_0(\hat{\boldsymbol{n}})\, Y^{0*}_0(\hat{\boldsymbol{n}}_1)\, S^1_{1/2}(x)\, S^1_{1/2}(x_1) \right] S^r_{\ell+\frac{1}{2}}(x_1)\, x_1^{\ell/2}\, Y^\ell_m(\hat{\boldsymbol{n}}_1)\,,
\end{aligned}
\tag{5.268}
$$

we obtain the intermediate result

$$
\begin{aligned}
\mathcal{L}\psi = &-\gamma \sum_{r\ell m} a_{r\ell m}(t)\, S^r_{\ell+\frac{1}{2}}(x)\, x^{\ell/2}\, Y^\ell_m(\hat{\boldsymbol{n}}) \\
&+ \frac{2\gamma}{\sqrt{\pi}} \sum_{r\ell m} a_{r\ell m}(t) \int_0^\infty dx_1\, x_1^{1/2}\, e^{-x_1} \left[Y^0_0(\hat{\boldsymbol{n}})\, \delta_{l0}\, \delta_{m0}\, S^0_{1/2}(x)\, S^0_{1/2}(x_1) \right. \\
&\qquad + \tfrac{2}{3}\, x^{1/2} x_1^{1/2} \sum_{m'=-1}^{1} Y^1_{m'}(\hat{\boldsymbol{n}})\, \delta_{l1}\, \delta_{mm'}\, S^0_{3/2}(x)\, S^0_{3/2}(x_1) \\
&\qquad \left. + \tfrac{2}{3}\, Y^0_0(\hat{\boldsymbol{n}})\, \delta_{l0}\, \delta_{m0}\, S^1_{1/2}(x)\, S^1_{1/2}(x_1) \right] S^r_{\ell+\frac{1}{2}}(x_1)\, x_1^{1/2}\,.
\end{aligned}
\tag{5.269}
$$

Appealing now to the orthogonality of the Sonine polynomials, and recalling that

$$
\Gamma(\tfrac{1}{2}) = \sqrt{\pi} \quad , \quad \Gamma(1) = 1 \quad , \quad \Gamma(z+1) = z\,\Gamma(z)\,,
\tag{5.270}
$$

we integrate over x_1. For the first term in brackets, we invoke the orthogonality relation with $n = 0$ and $\alpha = \frac{1}{2}$, giving $\Gamma(\frac{3}{2}) = \frac{1}{2}\sqrt{\pi}$. For the second bracketed term, we have $n = 0$ but $\alpha = \frac{3}{2}$, and we obtain $\Gamma(\frac{5}{2}) = \frac{3}{2}\Gamma(\frac{3}{2})$, while the third bracketed term involves leads to $n = 1$ and $\alpha = \frac{1}{2}$, also yielding $\Gamma(\frac{5}{2}) = \frac{3}{2}\Gamma(\frac{3}{2})$. Thus, we obtain the simple and pleasing result

$$
\mathcal{L}\psi = -\gamma \sideset{}{'}\sum_{r\ell m} a_{r\ell m}(t)\, S^r_{\ell+\frac{1}{2}}(x)\, x^{\ell/2}\, Y^\ell_m(\hat{\boldsymbol{n}})
\tag{5.271}
$$

where the prime on the sum indicates that the set

$$
\mathrm{CI} = \Big\{ (0,0,0)\,, \quad (1,0,0)\,, \quad (0,1,1)\,, \quad (0,1,0)\,, \quad (0,1,-1) \Big\}
\tag{5.272}
$$

are to be excluded from the sum. But these are just the functions which correspond to the five collisional invariants! Thus, we learn that

$$
\psi_{r\ell m}(\boldsymbol{v}) = \mathcal{N}_{r\ell m}\, S^r_{\ell+\frac{1}{2}}(x)\, x^{\ell/2}\, Y^\ell_m(\hat{\boldsymbol{n}})\,,
\tag{5.273}
$$

is an eigenfunction of $\mathcal{L}$ with eigenvalue $-\gamma$ if (r, ℓ, m) does not correspond to one of the five collisional invariants. In the latter case, the eigenvalue is zero. Thus, the algebraic action of $\mathcal{L}$ on the coefficients $a_{r\ell m}$ is

$$(\mathcal{L}a)_{r\ell m} = \begin{cases} -\gamma\, a_{r\ell m} & \text{if } (r, \ell, m) \notin \text{CI} \\ = 0 & \text{if } (r, \ell, m) \in \text{CI} \end{cases} \tag{5.274}$$

The quantity $\tau = \gamma^{-1}$ is the relaxation time.

It is pretty obvious that $\mathcal{L}$ is self-adjoint, since

$$
\begin{aligned}
\langle\, \phi \,|\, \mathcal{L}\psi \,\rangle &\equiv \int d^3v\, f^0(\boldsymbol{v})\, \phi(\boldsymbol{v})\, \mathcal{L}[\psi(\boldsymbol{v})] \\
&= -\gamma\, n \left(\frac{m}{2\pi k_{\mathrm{B}}T} \right)^{3/2} \int d^3v\, \exp\left(-\frac{m\boldsymbol{v}^2}{2k_{\mathrm{B}}T} \right) \phi(\boldsymbol{v})\, \psi(\boldsymbol{v}) \\
&\quad + \gamma\, n \left(\frac{m}{2\pi k_{\mathrm{B}}T} \right)^{3} \int d^3v \int d^3u\, \exp\left(-\frac{m\boldsymbol{u}^2}{2k_{\mathrm{B}}T} \right) \exp\left(-\frac{m\boldsymbol{v}^2}{2k_{\mathrm{B}}T} \right) \\
&\quad \times \phi(\boldsymbol{v}) \left[1 + \frac{m}{k_{\mathrm{B}}T}\, \boldsymbol{u} \cdot \boldsymbol{v} + \frac{2}{3} \left(\frac{m\boldsymbol{u}^2}{2k_{\mathrm{B}}T} - \frac{3}{2} \right) \left(\frac{m\boldsymbol{v}^2}{2k_{\mathrm{B}}T} - \frac{3}{2} \right) \right] \psi(\boldsymbol{u}) \\
&= \langle\, \mathcal{L}\phi \,|\, \psi \,\rangle\,,
\end{aligned}
\tag{5.275}
$$

where n is the bulk number density and $f^0(\boldsymbol{v})$ is the Maxwellian velocity distribution.

Chapter 6

Applications

6.1 References

– P. L. Krapivsky, S. Redner, E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge, 2010)
  An excellent selection of modern topics.

– A.-L. Barbarási and H. E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge, 1995)
  A very physical approach to the many interesting aspects of surface growth phenomena.

– V. Méndez, S. Fedotov, and W. Horsthemke, *Reaction-Transport Systems* (Springer-Verlag, 2010)
  Covers a broad range of topics in the area of reaction-diffusion systems.

– C. Gardiner, *Stochastic Methods* (4th edition, Springer-Verlag, 2010)
  Very clear and complete text on stochastic methods, with many applications.

– R. Mahnke, J. Kaupužs, and I. Lubashevsky, *Physics of Stochastic Processes* (Wiley, 2009)
  Introductory sections are sometimes overly formal, but a good selection of topics.

6.2 Diffusion

Diffusion is a ubiquitous phenomenon in the physical sciences. Here we briefly discuss some interesting features. Several examples are adapted from the book by Krapivsky, Redner, and Ben-Naim, which we abbreviate as KRB.

6.2.1 Return statistics

We have already studied the statistics of random walks in one dimension and also solutions of the diffusion equation, $\partial_t P = D \, \nabla^2 P$, in arbitrary dimensions,

$$P(\boldsymbol{x}, t) = (4\pi Dt)^{-d/2} \, e^{-\boldsymbol{x}^2/4Dt} \, , \tag{6.1}$$

with $P(\boldsymbol{x}, 0) = \delta(\boldsymbol{x})$. The variance of $\boldsymbol{x}$ at time t is

$$\mathsf{Var}\big[\boldsymbol{x}(t)\big] = \int d^d\!x \; \boldsymbol{x}^2 \, P(\boldsymbol{x}, t) = -\nabla_{\boldsymbol{k}}^2 \, \hat{P}(\boldsymbol{k}, t)\big|_{\boldsymbol{k}=0} = 2dDt \quad , \tag{6.2}$$

since $\hat{P}(\boldsymbol{k}, t) = \hat{P}(\boldsymbol{k}, 0) \, \exp(-Dk^2 t)$, and $\hat{P}(\boldsymbol{k}, 0) = 1$. Thus, the RMS distance of the particle from its initial position, after a time t, is $L(t) = \sqrt{2dDt}$. The diffusion equation is a continuum limit of a Master equation. The instantaneous position of the walker may be written as a sum over d unit vectors $\hat{e}_\mu$ with coefficients that are integer multiples of the lattice spacing a, i.e. $\boldsymbol{R} = a \sum_{\mu=1}^{d} n_\mu \hat{e}_\mu$. The Master equation is

$$\frac{\partial P(\boldsymbol{R}, t)}{\partial t} = \gamma \sum_{\mu=1}^{d} \Big[P(\boldsymbol{R} + a \, \hat{e}_\mu, t) + P(\boldsymbol{R} - a \, \hat{e}_\mu, t) - 2P(\boldsymbol{R}, t) \Big] \, , \tag{6.3}$$

where γ is the hopping rate. If we Taylor expand $P(\boldsymbol{R} + a \, \hat{e}_\mu, t)$ to second order in a, we recover the diffusion equation with $D = \gamma a^2$.

The number of sites visited over a time interval t is simply t, although a given site may be visited more than once. The density of visited sites is then $t/L^d(t) \propto t^{1-d/2}$. Thus, for $d > 2$ the density decreases with t, but for $d < 2$ the density increases, which means that we return to any given site with probability unity. The case $d = 2$ is marginal, and as we shall now see, also yields an infinite number of returns.

We studied first passage problems in §4.2.5 and §4.3.5. For the discrete time random walk on a d-dimensional cubic lattice, let $P(\boldsymbol{R}, t)$ be the probability that the walker is at position $\boldsymbol{R}$ at time $t \in \mathbb{Z}$, having started at $\boldsymbol{R} = 0$ at time $t = 0$. We write $\boldsymbol{R}(t) = \sum_{s=1}^{t} \hat{n}(s)$, where $\hat{n}(s) \in \{\pm \hat{e}_1, \dots, \pm \hat{e}_d\}$. Define $F(\boldsymbol{R}, t)$ to be the probability that the walker's first move onto site $\boldsymbol{R}$ occurs at time step t. Then we must have

$$P(\boldsymbol{R}, t) = \delta_{\boldsymbol{R}, \boldsymbol{0}} \, \delta_{t, 0} + \sum_{s=1}^{t} P(\boldsymbol{0}, t - s) \, F(\boldsymbol{R}, s) \, , \tag{6.4}$$

with $F(\boldsymbol{R}, t = 0) \equiv 0$. Now define

$$\breve{P}(\boldsymbol{R}, z) = \sum_{t=0}^{\infty} P(\boldsymbol{R}, t) \, z^t \, . \tag{6.5}$$

We then have

$$\check{P}(\boldsymbol{R}, z) = \delta_{\boldsymbol{R},\boldsymbol{0}} + \check{P}(\boldsymbol{0}, z)\, \check{F}(\boldsymbol{R}, z) \qquad \Rightarrow \qquad \check{F}(\boldsymbol{R}, z) = \frac{\check{P}(\boldsymbol{R}, z) - \delta_{\boldsymbol{R},\boldsymbol{0}}}{\check{P}(\boldsymbol{0}, z)} \quad . \tag{6.6}$$

Now

$$P(\boldsymbol{R}, t) = \left\langle \delta_{\boldsymbol{R},\boldsymbol{R}(t)} \right\rangle = \int\limits_{\hat{\Omega}} \frac{d^d\!k}{(2\pi)^d}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \left\langle e^{-i\boldsymbol{k}\cdot\boldsymbol{R}(t)} \right\rangle = \int\limits_{\hat{\Omega}} \frac{d^d\!k}{(2\pi)^d}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, \psi^t(\boldsymbol{k}) \,, \tag{6.7}$$

where

$$\psi(\boldsymbol{k}) = \frac{1}{d} \sum_{\mu=1}^{d} \cos k_\mu \quad , \tag{6.8}$$

and $\hat{\Omega}$ is the *first Brillouin zone* of the d-dimensional cubic lattice, which is the $d-$cube defined by $k_\mu \in [-\pi, \pi]$ for all $\mu \in \{1, \ldots, d\}$. We then have

$$\check{P}(\boldsymbol{R}, z) = \int\limits_{\hat{\Omega}} \frac{d^d\!k}{(2\pi)^d}\, \frac{e^{i\boldsymbol{k}\cdot\boldsymbol{R}}}{1 - z\,\psi(\boldsymbol{k})} \quad . \tag{6.9}$$

The expected total number of visits the walker makes to site $\boldsymbol{R}$ is $\nu_d(\boldsymbol{R}) = \sum_t P(\boldsymbol{R}, t) = \check{P}(\boldsymbol{R}, 1)$, hence

$$\nu_d(\boldsymbol{0}) = \check{P}(\boldsymbol{0}, 1) = \int\limits_0^\infty ds\, e^{-s}\, \big[I_0(s/d)\big]^d \quad , \tag{6.10}$$

where $I_0(z)$ is the modified Bessel function. Note that $I_0(z) \sim e^z/\sqrt{2\pi z}$ for large z, so the integral diverges for $d \leq 2$. Numerically, one finds $\nu_{d=3}(\boldsymbol{0}) = 1.517$.

The probability that the walker eventually returns to $\boldsymbol{R} = \boldsymbol{0}$ is

$$\mathcal{R} = \sum_{t=1}^\infty F(\boldsymbol{0}, t) = \check{F}(\boldsymbol{0}, 1) = 1 - \frac{1}{\check{P}(\boldsymbol{0}, 1)} \quad . \tag{6.11}$$

If $\check{P}(\boldsymbol{0}, 1)$ is finite, then $0 < \mathcal{R} < 1$. If on the other hand $\check{P}(\boldsymbol{0}, 1)$ diverges, then $\mathcal{R} = 1$ and the eventual return is certain. As the first Brillouin zone itself is finite, the only possibility for divergence is associated with the point $\boldsymbol{k} = 0$. Taylor expanding the function $\psi(\boldsymbol{k})$ about that point, we find

$$\psi(\boldsymbol{k}) = 1 - \frac{\boldsymbol{k}^2}{2d} + \sum_{\mu=1}^{d} \frac{k_\mu^4}{24d} + \mathcal{O}(k^6) \quad . \tag{6.12}$$

Thus, $1 - \psi(\boldsymbol{k}) \sim k^2/2d$ as $k \to 0$, and $\check{P}(\boldsymbol{0}, 1)$ diverges for $d \leq 2$. For $z \approx 1$, we may approximate

$$\check{P}(\boldsymbol{0}, z) = \int\limits_0^\infty du\, e^{-u} \int\limits_{\hat{\Omega}} \frac{d^d\!k}{(2\pi)^3}\, e^{uz\psi(\boldsymbol{k})} \approx \int\limits_0^\infty du\, e^{-u(1-z)} \left(\int\limits_{-\infty}^\infty \frac{dk}{2\pi}\, e^{-uzk^2/2d}\, e^{-k^2/2\Lambda^2} \right)^{\!d}$$

$$= \left(\frac{d}{2\pi}\right)^{\!d/2} \int\limits_0^\infty du\, e^{-u(1-z)} \left(zu + d\Lambda^{-2}\right)^{-d/2} \tag{6.13}$$

$$\approx \left(\frac{d}{2\pi}\right)^{\!d/2} \frac{\varepsilon^{(d-2)/2}}{1 - \frac{d}{2}} + \text{finite} \,,$$

where $z \equiv 1 - \varepsilon$ and $\Lambda \sim \pi$ is an ultraviolet cutoff, corresponding to the finite size of the Brillouin zone. When $d = 2$, the expression $\varepsilon^{(d-2)/2}/(1 - \frac{d}{2})$ is replaced by $\ln(1/\varepsilon)$, which follows from L'Hospital's rule. As advertised, we have a divergence in the limit $\varepsilon \to 0$ for $d \leq 2$, hence the return probability is $\mathcal{R} = 1$.

We now know that the number of visits to each site diverges as the number of steps t tends to infinity with $d \leq 2$. This prompts the question: for $d \leq 2$, what is the frequency of these visits? Let's compute the number of visits to the origin within T time steps. We have

$$\nu_d(\mathbf{0}, T) = \sum_{t=0}^{T} \left\langle \delta_{\mathbf{R}(t),\mathbf{0}} \right\rangle = \int_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d} \frac{1 - \psi^{T+1}(\mathbf{k})}{1 - \psi(\mathbf{k})} \ . \tag{6.14}$$

The numerator now vanishes for $\mathbf{k} \to \mathbf{0}$ and so the integral is finite. To estimate its value, note that the numerator behaves as

$$1 - \left(1 - \frac{k^2}{2d}\right)^{T+1} \sim 1 - e^{-Tk^2/2d} \tag{6.15}$$

where the RHS is valid for $k^2 = \mathcal{O}(d/T)$. This means that there is an effective infrared cutoff $k_{\min} \sim T^{-1/2}$. The infrared divergence is thus cured, and

$$\nu_d(\mathbf{0}, T) \sim \int_{k_{\min}} dk \ k^{d-3} \sim k_{\min}^{d-2} = T^{1-\frac{d}{2}} \ . \tag{6.16}$$

Therefore the average time between visits to the origin is $\tau_d(T) = T/\nu_d(\mathbf{0}, T) \sim T^{d/2}$. As $T \to \infty$, this, too, diverges. Note that for $d = 2$ we have $\nu_{d=2}(\mathbf{0}, T) \sim \ln T$ and $\tau_{d=2}(T) \sim T/\ln T$.

So there is good news and bad news if you lose your keys in $d \leq 2$ dimensions. The good news is that by executing a random walk, asymptotically you will visit every possible place your keys could be hiding, and each one of them a divergent number of times at that. The bad news is that your lifetime is finite.

6.2.2 Exit problems

Let Σ be a boundary surface (or point in $d = 1$ dimension), and consider the generalization of Eqn. 4.64, *viz.*

$$G_\Sigma(\mathbf{x}, t) = - \int_t^\infty dt' \int_\Sigma dS' \ \hat{\mathbf{n}}' \cdot \mathbf{J}(\mathbf{x}', t' \,|\, \mathbf{x}, 0) \quad , \tag{6.17}$$

which is the probability that a particle starting at $\mathbf{x}$ at time $t = 0$ exits via the surface Σ sometime after t. Applying the operator

$$\widetilde{\mathcal{L}} = +A_i(\mathbf{x}) \frac{\partial}{\partial x_i} + \frac{1}{2} B_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i \, \partial x_j} \tag{6.18}$$

to the previous equation, we have $\widetilde{\mathcal{L}} \mathbf{J}(\mathbf{x}', t \,|\, \mathbf{x}, 0) = \partial_t \mathbf{J}(\mathbf{x}', t \,|\, \mathbf{x}, 0)$, and therefore

$$\frac{\partial G_\Sigma(\mathbf{x}, t)}{\partial t} = \widetilde{\mathcal{L}} \, G_\Sigma(\mathbf{x}, t) = \int_\Sigma dS' \ \hat{\mathbf{n}}' \cdot \mathbf{J}(\mathbf{x}', t \,|\, \mathbf{x}, 0) \quad , \tag{6.19}$$

which says that the rate at which the probability $G_\Sigma(x, t)$ for exit via Σ changes is given by the instantaneous integral of the probability current normal to the surface. If we set $t = 0$, we must have $J(x', 0 \,|\, x, 0) = 0$ if $x \notin \Sigma$, which gives us an equation for the total exit probability via the surface Σ over all time, $\widetilde{\mathcal{L}} G_\Sigma(x, 0) = 0$. This equation is subject to the boundary condition that $G_\Sigma(x, 0) = 1$ if $x \in \Sigma$ and $G_\Sigma(x, 0) = 0$ if $x \in \Sigma'$ where Σ' is an absorbing boundary. To simplify notation, we will define $G_\Sigma(x) \equiv G_\Sigma(x, 0)$. Thus,

$$\left(v_{\mathrm{D}} \cdot \boldsymbol{\nabla} + D_{ij} \, \nabla_i \nabla_j \right) G_\Sigma(x) = 0 \quad , \tag{6.20}$$

where $v_{\mathrm{D}}(x) = A(x)$ is the local drift velocity and $D_{ij}(x) = \frac{1}{2} B_{ij}(x)$ is the local diffusion tensor. When v_{D} is constant and $D_{ij}(x) = D \, \delta_{ij}$ is constant and isotropic, we can define a length scale $\lambda = D/v_{\mathrm{D}}$.

In $d = 1$ dimension, assuming the homogeneity of space is broken only at the boundaries, Eqn. 6.20 takes the form $\partial_x \left(v_{\mathrm{D}} \, G + D \, \partial_x G \right) = 0$. The solution is easily found to be

$$G_\Sigma(x) = C_1 \, e^{-x/\lambda} + C_2 \quad , \tag{6.21}$$

where $C_{1,2}$ are constants of integration. Suppose we have an absorbing boundary at $x = 0$ and Σ denotes the point $x = L$ is the escape boundary. Then

$$G_L(x) = \frac{1 - \exp(-x/\lambda)}{1 - \exp(-L/\lambda)} \quad . \tag{6.22}$$

In the limit $\lambda \to \infty$, *i.e.* $v_{\mathrm{D}} \to 0$, we have $G_L(x, 0) = x/L$. This solution assumes $x \in [0, L]$, and if $x > L$ we have $G_L(x) = e^{(L-x)/\lambda}$. If $\lambda = \infty$, this means $G_L(x > L) = 1$, which means that starting anywhere to the right of $x = L$, there is a 100% chance that the particle will eventually arrive at $x = L$. If $x \in [0, L]$ the probability is less than 100% because the particle may instead be absorbed at $x = 0$.

In $d = 2$ dimensions, if we assume isotropy and a radial drift $v_{\mathrm{D}} = v_{\mathrm{D}} \hat{r}$, then from $\nabla^2 = \partial_r^2 + \frac{1}{r} \partial_r$ we have

$$\left(\frac{1}{\lambda} + \frac{1}{r} \right) \frac{\partial G_\Sigma(r)}{\partial r} + \frac{\partial^2 G_\Sigma(r)}{\partial r^2} = 0 \quad , \tag{6.23}$$

with $\lambda = D/v_{\mathrm{D}}$. We then define the function $W(r)$ such that

$$\frac{\partial \ln W}{\partial r} = \frac{1}{\lambda} + \frac{1}{r} \quad \Rightarrow \quad W(r) = r \, e^{r/\lambda} \quad , \tag{6.24}$$

so that

$$\frac{\partial}{\partial r} \left[W(r) \, \frac{\partial G_\Sigma(r)}{\partial r} \right] = 0 \quad , \tag{6.25}$$

the solution of which is

$$G_\Sigma(r) = C_1 \, E_1(r/\lambda) + C_2 \quad , \tag{6.26}$$

where $E_1(z)$ is the exponential integral,

$$E_1(z) = \int_z^\infty dt \, \frac{e^{-t}}{t} \quad . \tag{6.27}$$

In the limit $\lambda \to \infty$ the solution takes the form $G_\Sigma(r) = C_1' \ln r + C_2'$. If the circle $r = a$ is absorbing and the exit surface is the circle $r = b$, then for $r \in [a, b]$ we have

$$G_b(r) = \frac{E_1(a/\lambda) - E_1(r/\lambda)}{E_1(a/\lambda) - E_1(b/\lambda)} \xrightarrow[\lambda \to \infty]{} \frac{\ln(r/a)}{\ln(b/a)} \quad . \tag{6.28}$$

If $r > b$, then for $\lambda \to \infty$ we have $G_b(r) = 1$ as in the $d = 1$ case, but for finite λ the solution is given by $G_b(r) = E_1(r/\lambda)/E_1(b/\lambda)$.

Finally, consider the case $d > 2$, again assuming spatial isotropy away from the boundaries. We again assume spherical symmetry and purely radial drift. The radial Laplacian is $\nabla^2 = \partial_r^2 + \frac{d-1}{r}\partial_r$, hence we again obtain Eqn. 6.25, but with $W(r) = r^{d-1}e^{r/\lambda}$. Define the generalized exponential integral,

$$E_k(z) = \int_z^\infty dt \; \frac{e^{-t}}{t^k} = \Gamma(1 - k, z) \quad , \tag{6.29}$$

where $\Gamma(a, z)$ is the incomplete gamma function. The general solution may now be written as

$$G_\Sigma(r) = C_1 \, E_{d-1}(r/\lambda) + C_2 \; . \tag{6.30}$$

With an absorbing boundary at $r = a$ and the exit boundary at $r = b > a$, we obtain

$$G_b(r) = \frac{E_{d-1}(a/\lambda) - E_{d-1}(r/\lambda)}{E_{d-1}(a/\lambda) - E_{d-1}(b/\lambda)} \xrightarrow[\lambda \to \infty]{} \frac{(b/a)^{d-2} - (b/r)^{d-2}}{(b/a)^{d-2} - 1} \quad . \tag{6.31}$$

Starting at a point with $r > b$, the solution with $\lambda \to \infty$ is $G_b(r) = (b/r)^{d-2}$, which is less than one. Thus, there is a finite probability $1 - G_b(r)$ that a diffusing particle with no drift will escape to $r = \infty$ without ever hitting the surface at $r = b$.

Mean exit times

The mean exit time from a region Ω via a boundary surface Σ, starting from some point $x \in \Omega$, is

$$T_\Sigma(x) = \int_0^\infty dt \; t \left(-\frac{\partial G_\Sigma(x, t)}{\partial t} \right) \quad . \tag{6.32}$$

This function satisfies the equation $\widetilde{\mathcal{L}} \, T_\Sigma(x) = -1$, subject to boundary conditions $T_\Sigma(x) = 0$ if $x \in \Sigma$. In fact, the moments $T_\Sigma^{(n)}(x) \equiv \langle t^n \rangle = \int_0^\infty dt \; t^{n-1} \, G_\Sigma(x, t)$ satisfy the hierarchical set of equations,

$$\widetilde{\mathcal{L}} \, T_\Sigma^{(n)}(x) = -n \, T_\Sigma^{(n-1)}(x) \quad . \tag{6.33}$$

As is clear, the $n = 1$ level is already closed, since $T_\Sigma^{(0)}(x) = \langle 1 \rangle = 1$.

As an example, consider the case of pure diffusion in d dimensions. We ask what is the mean exit time, starting at a radius r, to pass through a sphere of radius $b > r$. The conditions being rotationally invariant, we solve the radial equation

$$\frac{\partial^2 T_b(r)}{\partial r^2} + \frac{d-1}{r}\frac{\partial T_b(r)}{\partial r} = -\frac{1}{D} \quad , \tag{6.34}$$

subject to $T_b(b) = 0$. We then have

$$T_b(r) = \frac{b^2 - r^2}{2dD} \quad . \tag{6.35}$$

6.2.3 Vicious random walks

Consider two random walkers on the same line, under the condition that the walkers annihilate if they should meet. How long before this tragic event happens? Following KRB, we can think of the *pair* of diffusing one-dimensional walkers as a *single* walker in *two* space dimensions. Annihilation occurs if the two-dimensional walker hits the line $x_1 = x_2$.

Since only the distance to the line matters, it is convenient to recast the diffusion equation in terms of relative and center-of-mass variables $x = x_2 - x_1$ and $X = \frac{1}{2}(x_1 + x_2)$, respectively. From classical mechanics, it should be no surprise that the diffusion equation in these variables becomes

$$\frac{\partial P}{\partial t} = 2D\frac{\partial^2 P}{\partial x^2} + \tfrac{1}{2}D\frac{\partial^2 P}{\partial X^2} \quad . \tag{6.36}$$

Since the value of X is irrelevant to the annihilation problem, we integrate over this variable, which kills off the second term on the RHS above because it is a total derivative, leaving the diffusion equation $\partial_t P = 2D\,\partial_x^2 P$ with a new diffusion constant $D' = 2D$, and an absorbing boundary condition $P(x = 0, t) = 0$. With initial conditions $x(0) = x_0$, we solve using the method of images, *viz.*

$$P(x,t) = \frac{1}{\sqrt{8\pi Dt}}\left\{e^{-(x-x_0)^2/8Dt} - e^{-(x+x_0)^2/8Dt}\right\} . \tag{6.37}$$

Now as we have discussed in §4.2.5, the first passage probability density for a particle starting from $x_0 > 0$ to hit $x = 0$ is

$$F(0,t) = -J(0,t\,|\,x_0,0) = 2D\,\partial_x P(x,t\,|\,x_0,0)\big|_{x=0}$$
$$= \frac{x_0}{\sqrt{8\pi Dt^3}}\,e^{-x_0^2/8Dt} \quad . \tag{6.38}$$

As $t \to \infty$, this decreases as $t^{-3/2}$. We also define the *survival probability* $S(t)$ as

$$S(t\,|\,x_0,0) = 1 - \int\limits_0^t dt'\,F(0,t'\,|\,x_0,0) \quad . \tag{6.39}$$

For our problem, $S(t\,|\,x_0,0) = \mathrm{erf}\left(x_0/\sqrt{8Dt}\right)$, which decays as $t^{-1/2}$ as $t \to \infty$.
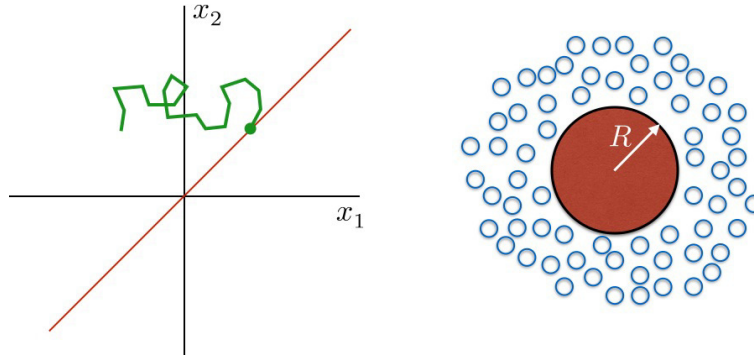
Figure 6.1: Two examples of diffusion problems. Left: vicious random walk. Right: diffusing particles and an absorbing sphere.

6.2.4   Reaction rate problems

Consider an object Ω whose surface is absorbing for some diffusing particles. How does the concentration $c(\boldsymbol{x},t)$ of diffusing particles evolve in the presence of the absorber? To answer this, we solve the diffusion equation $\partial_t c = D \, \nabla^2 c$ subject to the initial conditions $c(\boldsymbol{x} \notin \partial\Omega, t = 0) = c_0$ and the boundary condition $c(\boldsymbol{x} \in \partial\Omega, t) = 0$. It's convenient to define the complementary function $\bar{c}(\boldsymbol{x},t) = c_0 - c(\boldsymbol{x},t)$, which satisfies

$$\frac{\partial \bar{c}}{\partial t} = D \, \nabla^2 \bar{c} \quad , \qquad \bar{c}(\boldsymbol{x} \in \partial\Omega, t) = c_0 \quad , \quad \bar{c}(\boldsymbol{x} \notin \partial\Omega, t = 0) = 0 \; . \tag{6.40}$$

Initially there is a discontinuity in $\bar{c}(\boldsymbol{x}, t = 0)$ at the surface, resulting in a divergent second derivative at that location for $\bar{c}$. This causes $\bar{c}$ to grow there, as the diffusion equation requires, and smooths out the function. Eventually $\bar{c}(\boldsymbol{x},t)$ tends to a limiting function, and we define $\phi(\boldsymbol{x}) = \bar{c}(\boldsymbol{x},\infty)/c_0$. The function $\phi(\boldsymbol{x})$ then satisties

$$\nabla^2 \phi(\boldsymbol{x}) = 0 \quad , \qquad \phi(\boldsymbol{x} \in \partial\Omega) = 1 \quad , \quad \phi(\boldsymbol{x} \to \infty) = 0 \; . \tag{6.41}$$

These are the same equations as for the electrostatic potential $\phi(\boldsymbol{x})$ of a conducting surface of unit electrical potential. In electrostatics, the total surface charge is

$$Q = -\frac{1}{4\pi} \int_{\partial\Omega} dS \; \hat{\boldsymbol{n}} \cdot \boldsymbol{\nabla}\phi \quad . \tag{6.42}$$

The corresponding quantity for the reaction rate problem is the total incident flux of diffusing particles on the surface,

$$K = -\int_{\partial\Omega} dS \; \hat{\boldsymbol{n}} \cdot \boldsymbol{J} = -D \int_{\partial\Omega} dS \; \hat{\boldsymbol{n}} \cdot \boldsymbol{\nabla}\phi \quad . \tag{6.43}$$

In electrostatics, the ratio of the surface charge to the surface potential is the capacitance, which is a purely geometric quantity. Therefore, we have $K = 4\pi D C$, where C is the capacitance. For a sphere of radius R, we have $C = R$. For a disc of radius R, we have $C = 2R/\pi$. KRB provide a couple of other examples, for prolate and oblate ellipsoids of revolution[1]. Note that K as defined above has units

---

[1]For a sphere in d dimensions, the isotropic solution to Laplace's equation with $\phi(R) = 1$ is $\phi(r) = (R/r)^{d-2}$. We then obtain the capacitance $C = (d-2) \, R^{d-2}$.

$[K] = L^d\, T^{-1}$. Multiplying by the concentration c_0 gives the number of diffusing particles per unit time which hit the surface.

What happens in $d \le 2$ dimensions, where we know that random walks are recurrent? Consider, for example, the one-dimensional problem,

$$\frac{\partial c}{\partial t} = D\, \frac{\partial^2 c}{\partial x^2} \quad , \qquad c(x > 0, 0) = c_0 \quad , \quad c(0, t) = 0 \;. \tag{6.44}$$

The solution is $c(x, t) = c_0\, \mathsf{erf}\big(x/\sqrt{4Dt}\big)$, hence $c(x, t \to \infty) = 0$. A similar problem arises in $d = 2$ dimensions. KRB remark how the $d \le 2$ case can be understood in terms of effective *time-dependent* boundary conditions. For a problem with spherical symmetry, we solve the Laplace equation $\nabla^2 c = 0$ subject to the boundary conditions $c(a) = 0$ and $c(b) = 1$, with $b = \sqrt{Dt} > a$ a moving boundary. This yields

$$c(r, t) \simeq \frac{c_0\, r^{2-d} - c_0\, a^{2-d}}{(\sqrt{Dt})^{2-d} - a^{2-d}} \quad (d < 2) \quad , \qquad c(r, t) \simeq \frac{c_0\, \ln(r/a)}{\ln\big(\sqrt{Dt}/a\big)} \quad (d = 2) \;. \tag{6.45}$$

As $t \to \infty$, the reaction slows down, and one finds

$$K_{d<2}(t \to \infty) \simeq (2-d)\,\Omega_d\, D c_0\, (Dt)^{(d-2)/2} \quad , \quad K_{d=2}(t \to \infty) \simeq \frac{4\pi D c_0}{\ln\big(Dt/a^2\big)} \quad , \quad K_{d>2}(t \to \infty) \simeq D c_0\, a^{d-2} \quad , \tag{6.46}$$

where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the total solid angle in d dimensions. How can we understand these results? Recall that a diffusing particle starting a distance outside a spherical surface has a 100% probability of reaching the sphere. Thus, in the limit $t \to \infty$, *all* the diffusing material eventually gets absorbed by the sphere, leaving nothing! For $d > 2$, there is a finite probability *not* to hit the sphere, hence the asymptotic solution $c(\boldsymbol{x}, t = \infty)$ is not identically zero.

6.2.5 Polymers

Linear chain polymers are repeating structures with the chemical formula $(A)_x$, where A is the *formula unit* and x is the *degree of polymerization*. In many cases (*e.g.* polystyrene), $x \gtrsim 10^5$ is not uncommon. For a very readable introduction to the subject, see P. G. de Gennes, *Scaling Concepts in Polymer Physics*.

Quite often a given polymer solution will contain a distribution of x values; this is known as *polydispersity*. Various preparation techniques, such as chromatography, can mitigate the degree of polydispersity. Another morphological feature of polymers is *branching*, in which the polymers do not form linear chains.

Polymers exhibit a *static flexibility* which can be understood as follows. Consider a long chain hydrocarbon with a $-\mathrm{C} - \mathrm{C} - \mathrm{C}-$ backbone. The angle between successive $\mathrm{C} - \mathrm{C}$ bonds is fixed at $\theta \approx 68°$, but the azimuthal angle φ can take one of three possible low-energy values, as shown in the right panel of fig. 6.3. Thus, the relative probabilities of *gauche* and *trans* orientations are

$$\frac{\text{Prob (gauche)}}{\text{Prob (trans)}} = 2\, e^{-\Delta\varepsilon/k_{\mathrm{B}} T} \tag{6.47}$$
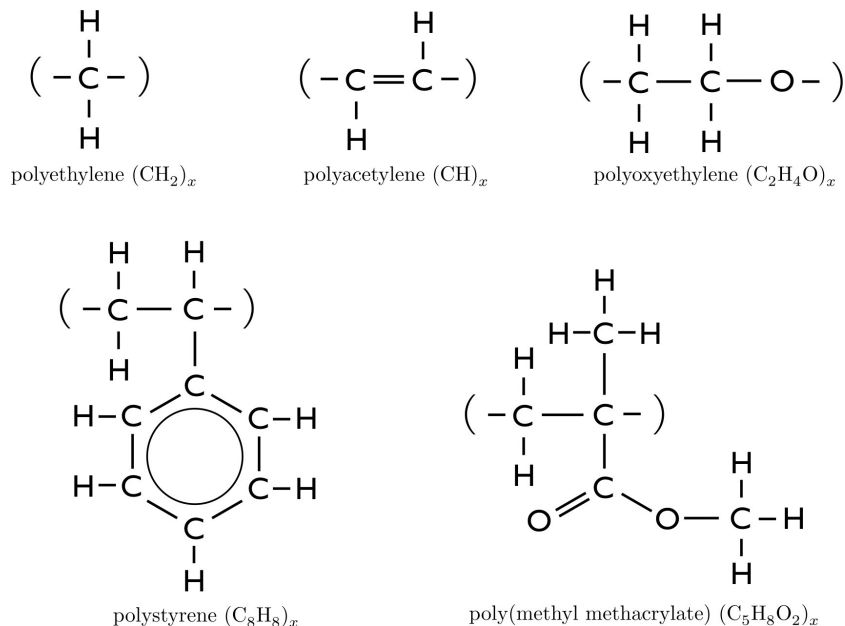
Figure 6.2: Some examples of linear chain polymers.

where $\Delta\varepsilon$ is the energy difference between *trans* and *gauche* configurations. This means that the polymer chain is in fact a *random coil* with a *persistence length*

$$\ell_{\mathrm{p}} = \ell_0 \, e^{\Delta\varepsilon/k_{\mathrm{B}}T} \tag{6.48}$$

where ℓ_0 is a microscopic length scale, roughly given by the length of a formula unit, which is approximately a few Ångstroms (see fig. 6.4). Let L be the total length of the polymer when it is stretched into a straight line. If $\ell_{\mathrm{p}} > L$, the polymer is *rigid*. If $\ell_{\mathrm{p}} \ll L$, the polymer is rigid on the length scale ℓ_{p} but flexible on longer scales. We have

$$\frac{\ell_{\mathrm{p}}}{L} = \frac{1}{N} \, e^{\Delta\varepsilon/k_{\mathrm{B}}T} \quad , \tag{6.49}$$

where we now use N (rather than x) for the degree of polymerization.

In the time domain, the polymer exhibits a *dynamical flexibility* on scales longer than a *persistence time*. The persistence time τ_{p} is the time required for a *trans-gauche* transition. The rate for such transitions is set by the energy barrier B separating *trans* from *gauche* configurations:

$$\tau_{\mathrm{p}} = \tau_0 \, e^{B/k_{\mathrm{B}}T} \tag{6.50}$$

where $\tau_0 \sim 10^{-11}$ s. On frequency scales $\omega \ll \tau_{\mathrm{p}}^{-1}$ the polymer is dynamically flexible. If $\Delta\varepsilon \sim k_{\mathrm{B}}T \ll B$ the polymer is flexible from a static point of view, but dynamically rigid. That is, there are many *gauche* orientations of successive carbon bonds which reflect a quenched disorder. The polymer then forms a frozen random coil, like a twisted coat hanger.
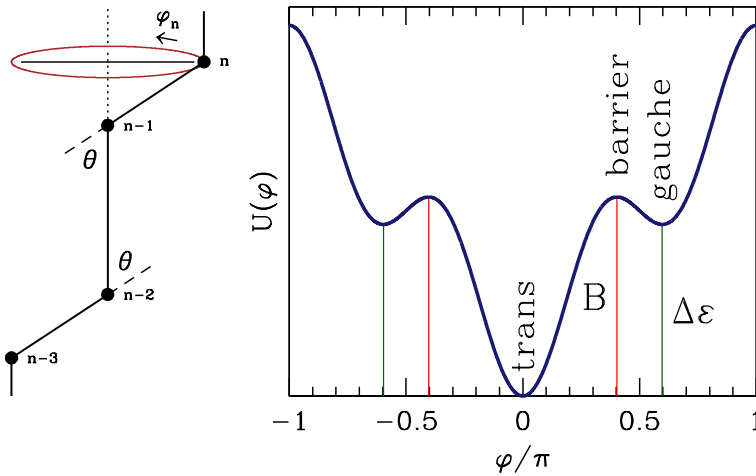
Figure 6.3: Left: *trans* and *gauche* orientations in carbon chains. Right: energy as a function of azimuthal angle φ. There are three low energy states: *trans* ($\varphi = 0$) and *gauche* ($\varphi = \pm\varphi_0$).

Polymers as random walks

A polymer can be modeled by a *self-avoiding random walk* (SAW). That is, on scales longer than $\ell_{\rm p}$, it twists about randomly in space subject to the constraint that it doesn't overlap itself. Before we consider the mathematics of SAWs, let's first recall some aspects of ordinary random walks which are not self-avoiding, which we discussed in §6.2.1 above.

We'll simplify matters further by considering random walks on a hypercubic lattice of dimension d. Such a lattice has coordination number $2d$, *i.e.* there are $2d$ nearest neighbor separations, $\boldsymbol{\delta} = \pm a\,\hat{\boldsymbol{e}}_1 , \pm a\,\hat{\boldsymbol{e}}_2 , \dots , \pm a\,\hat{\boldsymbol{e}}_d$, where a is the lattice spacing. Consider now a random walk of N steps starting at the origin. After N steps the position of the walker is $\boldsymbol{R}_N = \sum_{j=1}^N \boldsymbol{\delta}_j$, where $\boldsymbol{\delta}_j$ takes on one of $2d$ possible values. The quantity N is no longer the degree of polymerization, but something approximating $L/\ell_{\rm p}$, which is the number of persistence lengths in the chain. We assume each step is independent, hence $\langle \delta_j^\alpha\, \delta_{j'}^\beta \rangle = (a^2/d)\, \delta_{jj'} \delta^{\alpha\beta}$ and $\langle \boldsymbol{R}_N^2 \rangle = Na^2$. The full distribution $P_N(\boldsymbol{R})$ is given by

$$P_N(\boldsymbol{R}) = (2d)^{-N} \sum_{\boldsymbol{\delta}_1} \cdots \sum_{\boldsymbol{\delta}_N} \delta_{\boldsymbol{R},\sum_j \boldsymbol{\delta}_j}$$

$$= a^d \int_{-\pi/a}^{\pi/a} \frac{dk_1}{2\pi} \cdots \int_{-\pi/a}^{\pi/a} \frac{dk_d}{2\pi}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \left[\frac{1}{d} \sum_{\mu=1}^d \cos(k_\mu a) \right]^N$$

$$= a^d \int_{\hat{\Omega}} \frac{d^dk}{(2\pi)^d}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, \exp\left[N \ln\left(1 - \frac{1}{2d}\,\boldsymbol{k}^2 a^2 + \dots \right) \right] \tag{6.51}$$

$$\approx \left(\frac{a}{2d} \right)^d \int d^dk\, e^{-N\boldsymbol{k}^2 a^2/2d}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} = \left(\frac{d}{2\pi N} \right)^{d/2} e^{-d\boldsymbol{R}^2/2Na^2} \quad .$$

This is a simple Gaussian, with width $\langle \boldsymbol{R}^2 \rangle = d\cdot(Na^2/d) = Na^2$, as we have already computed. The

quantity $\boldsymbol{R}$ defined here is the *end-to-end vector* of the chain. The RMS end-to-end distance is then $\langle \boldsymbol{R}^2 \rangle^{1/2} = \sqrt{N}a \equiv R_0$.

A related figure of merit is the *radius of gyration*, R_{g}, defined by

$$R_{\mathrm{g}}^2 = \frac{1}{N} \left\langle \sum_{n=1}^{N} \left(\boldsymbol{R}_n - \boldsymbol{R}_{\mathrm{CM}} \right)^2 \right\rangle \quad , \tag{6.52}$$

where $R_{\mathrm{CM}} = \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{R}_j$ is the center of mass position. A brief calculation yields

$$R_g^2 = \frac{1}{6} \left(N + 3 - 4N^{-1} \right) a^2 \sim \frac{Na^2}{6} \quad , \tag{6.53}$$

in all dimensions.

The total number of random walk configurations with end-to-end vector $\boldsymbol{R}$ is then $(2d)^N P_N(\boldsymbol{R})$, so the entropy of a chain at fixed elongation is

$$S(\boldsymbol{R}, N) = k_{\mathrm{B}} \ln \left[(2d)^N P_N(\boldsymbol{R}) \right] = S(0, N) - \frac{dk_{\mathrm{B}} \boldsymbol{R}^2}{2Na^2} \quad . \tag{6.54}$$

If we assume that the energy of the chain is conformation independent, then $E = E_0(N)$ and

$$F(\boldsymbol{R}, N) = F(0, N) + \frac{dk_{\mathrm{B}} T \boldsymbol{R}^2}{2Na^2} \quad . \tag{6.55}$$

In the presence of an external force F_{ext}, the Gibbs free energy is the Legendre transform

$$G(\boldsymbol{F}_{\mathrm{ext}}, N) = F(\boldsymbol{R}, N) - \boldsymbol{F}_{\mathrm{ext}} \cdot \boldsymbol{R} \quad , \tag{6.56}$$

and $\partial G / \partial \boldsymbol{R} = 0$ then gives the relation

$$\langle \boldsymbol{R}(\boldsymbol{F}_{\mathrm{ext}}, N) \rangle = \frac{Na^2}{dk_{\mathrm{B}} T} \boldsymbol{F}_{\mathrm{ext}} \quad . \tag{6.57}$$

This may be considered an equation of state for the polymer.

Following de Gennes, consider a chain with charges $\pm e$ at each end, placed in an external electric field of magnitude $E = 30,000 \, \mathrm{V/cm}$. Let $N = 10^4$, $a = 2 \, \text{Å}$, and $d = 3$. What is the elongation? From the above formula, we have

$$\frac{R}{R_0} = \frac{eER_0}{3k_{\mathrm{B}} T} = 0.8 \ , \tag{6.58}$$

with $R_0 = \sqrt{N}a$ as before.

Structure factor

We can also compute the structure factor,

$$S(\boldsymbol{k}) = \frac{1}{N} \left\langle \sum_{m=1}^{N} \sum_{n=1}^{N} e^{i\boldsymbol{k} \cdot (\boldsymbol{R}_m - \boldsymbol{R}_n)} \right\rangle = 1 + \frac{2}{N} \sum_{m=1}^{N} \sum_{n=1}^{m-1} \left\langle e^{i\boldsymbol{k} \cdot (\boldsymbol{R}_m - \boldsymbol{R}_n)} \right\rangle \quad . \tag{6.59}$$
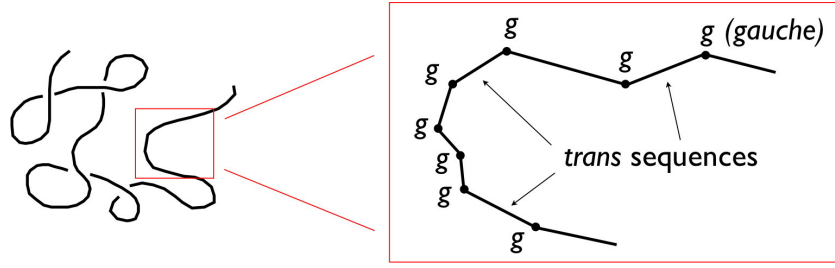
Figure 6.4: The polymer chain as a random coil.

For averages with respect to a Gaussian distribution,

$$\left\langle e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_m-\boldsymbol{R}_n)}\right\rangle = \exp\left\{-\frac{1}{2}\left\langle\left(\boldsymbol{k}\cdot(\boldsymbol{R}_m-\boldsymbol{R}_n)\right)^2\right\rangle\right\} \quad . \tag{6.60}$$

Now for $m > n$ we have $\boldsymbol{R}_m - \boldsymbol{R}_n = \sum_{j=n+1}^{m}\boldsymbol{\delta}_j$, and therefore

$$\left\langle\left(\boldsymbol{k}\cdot(\boldsymbol{R}_m-\boldsymbol{R}_n)\right)^2\right\rangle = \sum_{j=n+1}^{m}\left\langle(\boldsymbol{k}\cdot\boldsymbol{\delta}_j)^2\right\rangle = \frac{1}{d}\,(m-n)\,k^2 a^2 \quad , \tag{6.61}$$

since $\langle\delta_j^\alpha\,\delta_{j'}^\beta\rangle = (a^2/d)\,\delta_{jj'}\delta^{\alpha\beta}$. We then have

$$S(\boldsymbol{k}) = 1 + \frac{2}{N}\sum_{m=1}^{N}\sum_{n=1}^{m-1} e^{-(m-n)\,k^2 a^2/2d} = \frac{N\,(e^{2\mu_{\boldsymbol{k}}}-1)-2\,e^{\mu_{\boldsymbol{k}}}\,(1-e^{-N\mu_{\boldsymbol{k}}})}{N\left(e^{\mu_{\boldsymbol{k}}}-1\right)^2} \quad , \tag{6.62}$$

where $\mu_{\boldsymbol{k}} = k^2 a^2/2d$. In the limit where $N\to\infty$ and $a\to 0$ with $Na^2 = R_0^2$ constant, the structure factor has a *scaling form*, $S(\boldsymbol{k}) = Nf(N\mu_{\boldsymbol{k}}) = (R_0/a)^2\,f(k^2 R_0^2/2d)$, where

$$f(x) = \frac{2}{x^2}\left(e^{-x}-1+x\right) = 1 - \frac{x}{3} + \frac{x^2}{12} + \dots \quad . \tag{6.63}$$

Rouse model

Consider next a polymer chain subjected to stochastic forcing. We model the chain as a collection of mass points connected by springs, with a potential energy $U = \frac{1}{2}k\sum_n\left(\boldsymbol{x}_{n+1}-\boldsymbol{x}_n\right)^2$. This reproduces the distribution of Eqn. 6.51 if we take the spring constant to be $k = 3k_{\mathrm{B}}T/a^2$ and set the equilibrium length of each spring to zero. The equations of motion are then

$$M\ddot{\boldsymbol{x}}_n + \gamma\,\dot{\boldsymbol{x}}_n = -k\left(2\boldsymbol{x}_n - \boldsymbol{x}_{n-1} - \boldsymbol{x}_{n+1}\right) + \boldsymbol{f}_n(t) \quad , \tag{6.64}$$

where $n\in\{1,\dots,N\}$ and $\{f_n^\mu(t)\}$ a set of Gaussian white noise forcings, each with zero mean, and

$$\left\langle f_n^\mu(t)\,f_{n'}^\nu(t')\right\rangle = 2\gamma k_{\mathrm{B}}T\,\delta_{nn'}\,\delta^{\mu\nu}\,\delta(t-t') \quad . \tag{6.65}$$

We define $x_0 \equiv x_1$ and $x_{N+1} \equiv x_N$ so that the end mass points $n = 1$ and $n = N$ experience a restoring force from only one neighbor. We assume the chain is overdamped and set $M \to 0$. We then have

$$\gamma \, \dot{x}_n = -k \sum_{n'=1}^{N} A_{nn'} \, x_{n'} + f_n(t) \quad , \tag{6.66}$$

where

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ 0 & 0 & -1 & \ddots & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 1 \end{pmatrix} \quad . \tag{6.67}$$

The matrix A is real and symmetric. Its eigenfunctions are labeled $\psi_j(n)$, with $j \in \{0, \ldots, N-1\}$:

$$\psi_0(n) = \frac{1}{\sqrt{N}}$$
$$\psi_j(n) = \sqrt{\frac{2}{N}} \, \cos\left(\frac{(2n-1)j\pi}{2N} \right) \quad , \quad j \in \{1, \ldots, N-1\} \tag{6.68}$$

The completeness and orthonormality relations are

$$\sum_{j=0}^{N-1} \psi_j(n) \, \psi_j(n') = \delta_{nn'} \quad , \qquad \sum_{n=1}^{N} \psi_j(n) \, \psi_{j'}(n) = \delta_{jj'} \quad , \tag{6.69}$$

with eigenvalues $\lambda_j = 4 \sin^2\left(\pi j / 2N \right)$. Note that $\lambda_0 = 0$.

We now work in the basis of normal modes $\{\eta_j^\mu\}$, where

$$\eta_j^\mu(t) = \sum_{n=1}^{N} \psi_j(n) \, x_n^\mu(t) \qquad , \qquad x_n^\mu(t) = \sum_{j=0}^{N-1} \psi_j(n) \, \eta_j^\mu(t) \quad . \tag{6.70}$$

We then have

$$\frac{d\eta_j}{dt} = -\frac{1}{\tau_j} \, \eta_j + g_j(t) \quad , \tag{6.71}$$

where the jth relaxation time is

$$\tau_j = \frac{\gamma}{4k \sin^2\left(\pi j / 2N \right)} \tag{6.72}$$

and

$$g_j^\mu(t) = \gamma^{-1} \sum_{n=1}^{N} \psi_j(n) \, f_n^\mu(t) \quad . \tag{6.73}$$

Note that

$$\langle g_j^\mu(t)\, g_{j'}^\nu(t')\rangle = 2\gamma^{-1}k_{\rm B}T\,\delta_{jj'}\,\delta^{\mu\nu}\,\delta(t-t')\quad. \tag{6.74}$$

Integrating Eqn. 6.71, we have for, $j=0$,

$$\boldsymbol{\eta}_0(t) = \boldsymbol{\eta}_0(0) + \int\limits_0^t dt'\,\boldsymbol{g}_0(t')\quad. \tag{6.75}$$

For the $j>0$ modes,

$$\boldsymbol{\eta}_j(t) = \boldsymbol{\eta}_j(0)\,e^{-t/\tau_j} + \int\limits_0^t dt'\,\boldsymbol{g}_j(t')\,e^{(t'-t)/\tau_j}\quad. \tag{6.76}$$

Thus,

$$\begin{aligned}
\langle \eta_0^\mu(t)\,\eta_0^\nu(t')\rangle_{\rm c} &= 2\gamma^{-1}k_{\rm B}T\,\delta^{\mu\nu}\,{\rm min}(t,t')\\[4pt]
\langle \eta_j^\mu(t)\,\eta_j^\nu(t')\rangle_{\rm c} &= \gamma^{-1}k_{\rm B}T\,\delta^{\mu\nu}\,\tau_j\left(e^{-|t-t'|/\tau_j} - e^{-(t+t')/\tau_j}\right)\quad,
\end{aligned} \tag{6.77}$$

where the 'connected average' is defined to be $\langle A(t)\,B(t')\rangle_{\rm c} \equiv \langle A(t)\,B(t')\rangle - \langle A(t)\rangle\langle B(t')\rangle$. Transforming back to the original real space basis, we then have

$$\langle x_n^\mu(t)\,x_{n'}^\nu(t')\rangle_{\rm c} = \frac{2k_{\rm B}T}{N\gamma}\,\delta^{\mu\nu}\,{\rm min}(t,t') + \frac{k_{\rm B}T}{\gamma}\,\delta^{\mu\nu}\sum_{j=1}^{N-1}\tau_j\,\psi_j(n)\,\psi_j(n')\left(e^{-|t-t'|/\tau_j} - e^{-(t+t')/\tau_j}\right)\quad. \tag{6.78}$$

In particular, the 'connected variance' of $\boldsymbol{x}_n(t)$ is

$$\mathsf{CVar}\big[\boldsymbol{x}_n(t)\big] \equiv \big\langle[\boldsymbol{x}_n(t)]^2\big\rangle_{\rm c} = \frac{6k_{\rm B}T}{N\gamma}\,t + \frac{3k_{\rm B}T}{\gamma}\sum_{j=1}^{N-1}\tau_j\,[\psi_j(n)]^2\left(1-e^{-2t/\tau_j}\right)\quad. \tag{6.79}$$

From this we see that at long times, *i.e.* when $t\gg\tau_1$, the motion of $\boldsymbol{x}_n(t)$ is diffusive, with diffusion constant $D = k_{\rm B}T/N\gamma \propto B^{-1}$, which is inversely proportional to the chain length. Recall the Stokes result $\gamma = 6\pi\eta R/M$ for a sphere of radius R and mass M moving in a fluid of dynamical viscosity η. From $D = k_{\rm B}T/\gamma M$, shouldn't we expect the diffusion constant to be $D = k_{\rm B}T/6\pi\eta R \propto N^{-1/2}$, since the radius of gyration of the polymer is $R_{\rm g} \propto N^{1/2}$? This argument smuggles in the assumption that the only dissipation is taking place at the *outer surface* of the polymer, modeled as a ball of radius $R_{\rm g}$. In fact, for a Gaussian random walk in three space dimensions, the density for $r < R_{\rm g}$ is $\rho \propto N^{-1/2}$ since there are N monomers inside a region of volume $(\sqrt{N})^3$. Accounting for Flory swelling due to steric interactions (see below), the density is $\rho \sim N^{-4/5}$, which is even smaller. So as $N \to \infty$, the density within the $r = R_{\rm g}$ effective sphere gets small, which means water molecules can easily penetrate, in which case the entire polymer chain should be considered to be in a dissipative environment, which is what the Rouse model says – each monomer executed overdamped motion.

A careful analysis of Eqn. 6.79 reveals that there is a subdiffusive regime[2] where $\mathsf{CVar}\big[\boldsymbol{x}_n(t)\big]\propto t^{1/2}$. To see this, first take the $N \gg 1$ limit, in which case we may write $\tau_j = N^2\tau_0/j^2$, where $\tau_0 \equiv \gamma/\pi^2 k$ and

---

[2]I am grateful to Jonathan Lam and Olga Dudko for explaining this to me.

$j \in \{1, \ldots, N-1\}$. Let $s \equiv (n - \tfrac{1}{2})/N \in [0,1]$ be the scaled coordinate along the chain. The second term in Eqn. 6.79 is then

$$S(s,t) \equiv \frac{6k_{\mathrm{B}}T}{\gamma} \cdot \frac{\tau_1}{N} \sum_{j=1}^{N-1} \frac{\cos^2(\pi j s)}{j^2} \left(1 - e^{-2j^2 t/\tau_1}\right) \quad . \tag{6.80}$$

Let $\sigma \equiv (t/\tau_1)^{1/2}$. When $t \ll \tau_1$, *i.e.* $\sigma \ll 1$, we have

$$S(s,t) \simeq \frac{6k_{\mathrm{B}}T}{\gamma} \cdot \frac{\tau_1}{N} \, \sigma \int_0^{N\sigma} du \, \frac{\cos^2(\pi u s/\sigma)}{u^2} \left(1 - e^{-2u^2}\right) \quad . \tag{6.81}$$

Since $s/\sigma \gg 1$, we may replace the cosine squared term by its average $\tfrac{1}{2}$. If we further assume $N\sigma \gg 1$, which means we are in the regime $1 \ll t/\tau_0 \ll N^2$, after performing the integral we obtain the result

$$S(s,t) = \frac{3k_{\mathrm{B}}T}{\gamma} \sqrt{2\pi\tau_0 t} \quad , \tag{6.82}$$

provided $s = \mathcal{O}(1)$, *i.e.* the site n is not on either end of the chain. The result in Eqn. 6.82 dominates the first term on the RHS of Eqn. 6.79 since $\tau_0 \ll t \ll \tau_1$. This is the subdiffusive regime.

When $t \gg \tau_1 = N^2 \tau_0$, the exponential on the RHS of Eqn. 6.80 is negligible, and if we again approximate $\cos^2(\pi j s) \simeq \tfrac{1}{2}$, and we extend the upper limit on the sum to infinity, we find $S(t) = (3k_{\mathrm{B}}T/\gamma)(\tau_1/N)(\pi^2/6) \propto t^0$, which is dominated by the leading term on the RHS of Eqn. 6.79. This is the diffusive regime, with $D = k_{\mathrm{B}}T/N\gamma$.

Finally, when $t \ll \tau_0$, the factor $1 - \exp(-2t/\tau_j)$ may be expanded to first order in t. One then obtains $\mathsf{CVar}\big[\boldsymbol{x}_n(t)\big] = (6k_{\mathrm{B}}T/\gamma)\, t$, which is independent of the force constant k. In this regime, the monomers don't have time to respond to the force from their neighbors, hence they each diffuse independently. On such short time scales, however, one should check to make sure that inertial effects can be ignored, *i.e.* that $t \gg M/\gamma$.

One serious defect of the Rouse model is its prediction of the relaxation time of the $j = 1$ mode, $\tau_1 \propto N^2$. The experimentally observed result is $\tau_1 \propto N^{3/2}$. We should stress here that the Rouse model applies to ideal chains. In the theory of polymer solutions, a *theta solvent* is one in which polymer coils act as ideal chains. An extension of the Rouse model, due to my former UCSD colleague Bruno Zimm, accounts for hydrodynamically-mediated interactions between any pair of 'beads' along the chain. Specifically, the Zimm model is given by

$$\frac{dx_n^\mu}{dt} = \sum_{n'} H^{\mu\nu}(\boldsymbol{x}_n - \boldsymbol{x}_{n'}) \Big[k\big(x_{n'+1}^\nu + x_{n'-1}^\nu - 2x_{n'}^\nu\big) + f_{n'}^\nu(t) \Big] \quad , \tag{6.83}$$

where

$$H^{\mu\nu}(\boldsymbol{R}) = \frac{1}{6\pi\eta R}\big(\delta^{\mu\nu} + \hat{R}^\mu \hat{R}^\nu\big) \tag{6.84}$$

is known as the Oseen hydrodynamic tensor (1927) and arises when computing the velocity in a fluid at position $\boldsymbol{R}$ when a point force $\boldsymbol{F} = \boldsymbol{f}\,\delta(\boldsymbol{r})$ is applied at the origin. Typically one replaces $H(\boldsymbol{R})$ by its average over the equilibrium distribution of polymer configurations. Zimm's model more correctly reproduces the behavior of polymers in θ-solvents.

Flory theory of self-avoiding walks

What is missing from the random walk free energy is the effect of *steric interactions*. An argument due to Flory takes these interactions into account in a mean field treatment. Suppose we have a chain of radius R. Then the average monomer density within the chain is $c = N/R^d$. Assuming short-ranged interactions, we should then add a term to the free energy which effectively counts the number of near self-intersections of the chain. This number should be roughly Nc. Thus, we write

$$F(\boldsymbol{R}, N) = F_0 + u(T) \frac{N^2}{R^d} + \tfrac{1}{2} dk_{\text{B}} T \frac{R^2}{Na^2} \quad . \tag{6.85}$$

The effective interaction $u(T)$ is positive in the case of a so-called 'good solvent'.

The free energy is minimized when

$$0 = \frac{\partial F}{\partial R} = -\frac{duN^2}{R^{d+1}} + \frac{dk_{\text{B}}TR}{Na^2} \quad , \tag{6.86}$$

which yields the result

$$R_{\text{F}}(N) = \left(\frac{ua^2}{k_{\text{B}}T} \right)^{1/(d+2)} N^{3/(d+2)} \propto N^{\nu} \quad . \tag{6.87}$$

Thus, we obtain $\nu = 3/(d+2)$. In $d = 1$ this says $\nu = 1$, which is exactly correct because a SAW in $d = 1$ has no option but to keep going in the same direction. In $d = 2$, Flory theory predicts $\nu = \frac{3}{4}$, which is also exact. In $d = 3$, we have $\nu_{d=3} = \frac{3}{5}$, which is extremely close to the numerical value $\nu = 0.5880$. Flory theory is again exact at the SAW upper critical dimension, which is $d = 4$, where $\nu = \frac{1}{2}$, corresponding to a Gaussian random walk[3]. Best. Mean. Field. Theory. Ever.

How well are polymers described as SAWs? Fig. 6.5 shows the radius of gyration R_{g} *versus* molecular weight M for polystyrene chains in a toluene and benzene solvent. The slope is $\nu = d\ln R_{\text{g}}/d\ln M = 0.5936$. Experimental results can vary with concentration and temperature, but generally confirm the validity of the SAW model.

For a SAW under an external force, we compute the Gibbs partition function,

$$Y(\boldsymbol{F}_{\text{ext}}, N) = \int d^dR \, P_N(\boldsymbol{R}) \, e^{\boldsymbol{F}_{\text{ext}} \cdot \boldsymbol{R}/k_{\text{B}}T} = \int d^dx \, f(x) \, e^{s\hat{\boldsymbol{n}} \cdot \boldsymbol{x}} \, , \tag{6.88}$$

where $x = R/R_{\text{F}}$ and $s = k_{\text{B}}T/R_{\text{F}}F_{\text{ext}}$ and $\hat{\boldsymbol{n}} = \hat{\boldsymbol{F}}_{\text{ext}}$. One than has $R(F_{\text{ext}}) = R_{\text{F}} \, \Phi(R_{\text{F}}/\xi)$, where $\xi = k_{\text{B}}T/F_{\text{ext}}$ and $R(F_{\text{ext}}) = F_{\text{ext}}R_{\text{F}}^2/k_{\text{B}}T$. For small values of its argument one has $\Phi(u) \propto u$. For large u it can be shown that $R(F_{\text{ext}}) \propto (F_{\text{ext}}R_{\text{F}}/k_{\text{B}}T)^{2/3}$.

On a lattice of coordination number z, the number of N-step random walks starting from the origin is $\Omega_N = z^N$. If we constrain our random walks to be self-avoiding, the number is reduced to

$$\Omega_N^{\text{SAW}} = \mathcal{C} \, N^{\gamma-1} y^N \quad , \tag{6.89}$$

---

[3]There are logarithmic corrections to the SAW result exactly at $d = 4$, but for all $d > 4$ one has $\nu = \frac{1}{2}$.
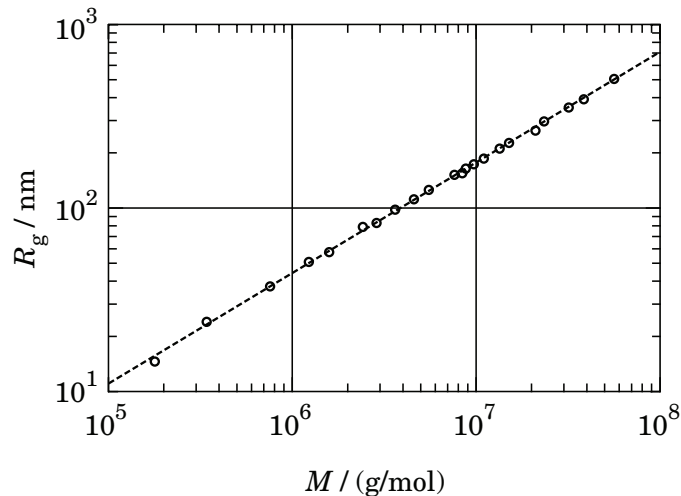
Figure 6.5: Radius of gyration R_g of polystyrene in a toluene and benzene solvent, plotted as a function of molecular weight of the polystyrene. The best fit corresponds to a power law $R_g \propto M^\nu$ with $\nu = 0.5936$. From J. Des Cloizeaux and G. Jannink, *Polymers in Solution: Their Modeling and Structure* (Oxford, 1990).

where $\mathcal{C}$ and γ are dimension-dependent constants, and we expect $y \lesssim z - 1$, since at the very least a SAW cannot immediately double back on itself. In fact, on the cubic lattice one has $z = 6$ but $y = 4.68$, slightly less than $z - 1$. One finds $\gamma_{d=2} \simeq \frac{4}{3}$ and $\gamma_{d=3} \simeq \frac{7}{6}$. The RMS end-to-end distance of the SAW is

$$R_{\mathrm{F}} = a\, N^\nu \quad , \tag{6.90}$$

where a and ν are d-dependent constants, with $\nu_{d=1} = 1$, $\nu_{d=2} \simeq \frac{3}{4}$, and $\nu_{d=3} \simeq \frac{3}{5}$. The distribution $P_N(\boldsymbol{R})$ has a scaling form,

$$P_N(R) = \frac{1}{R_{\mathrm{F}}^d}\, f(R/R_{\mathrm{F}}) \qquad (a \ll R \ll Na) \quad . \tag{6.91}$$

One finds

$$f(x) \sim \begin{cases} x^g & x \ll 1 \\ \exp(-x^\delta) & x \gg 1 \,, \end{cases} \tag{6.92}$$

with $g = (\gamma - 1)/\nu$ and $\delta = 1/(1 - \nu)$.

Polymers and solvents

Consider a solution of monodisperse polymers of length N in a solvent. Let ϕ be the dimensionless monomer concentration, so ϕ/N is the dimensionless polymer concentration and $\phi_{\mathrm{s}} = 1 - \phi$ is the dimensionless solvent concentration. (Dimensionless concentrations are obtained by dividing the corresponding dimensionful concentration by the overall density.) The entropy of mixing for such a system is given by

$$S_{\mathrm{mix}} = -\frac{V k_{\mathrm{B}}}{v_0} \cdot \left\{ \frac{1}{N} \phi \ln \phi + (1 - \phi) \ln(1 - \phi) \right\} , \tag{6.93}$$

where $v_0 \propto a^3$ is the volume per monomer. Accounting for an interaction between the monomer and the solvent, we have that the free energy of mixing is

$$\frac{v_0 F_{\text{mix}}}{V k_{\text{B}} T} = \frac{1}{N} \phi \ln \phi + (1 - \phi) \ln(1 - \phi) + \chi \, \phi(1 - \phi) \,. \tag{6.94}$$

where χ is the dimensionless polymer-solvent interaction, called the *Flory parameter*. This provides a mean field theory of the polymer-solvent system.

The osmotic pressure Π is defined by

$$\Pi = -\frac{\partial F_{\text{mix}}}{\partial V}\bigg|_{N_{\text{p}}} \,, \tag{6.95}$$

which is the variation of the free energy of mixing with respect to volume *holding the number of polymers constant*. The monomer concentration is $\phi = N N_{\text{p}} v_0 / V$, so

$$\frac{\partial}{\partial V}\bigg|_{N_{\text{p}}} = -\frac{\phi^2}{N N_{\text{p}} v_0} \frac{\partial}{\partial \phi}\bigg|_{N_{\text{p}}} \,. \tag{6.96}$$

Now we have

$$F_{\text{mix}} = N N_{\text{p}} k_{\text{B}} T \left\{ \frac{1}{N} \ln \phi + (\phi^{-1} - 1) \ln(1 - \phi) + \chi \, (1 - \phi) \right\} \,, \tag{6.97}$$

and therefore

$$\Pi = \frac{k_{\text{B}} T}{v_0} \left[(N^{-1} - 1) \, \phi - \ln(1 - \phi) - \chi \, \phi^2 \right] \,. \tag{6.98}$$

In the limit of vanishing monomer concentration $\phi \to 0$, we recover

$$\Pi = \frac{\phi \, k_{\text{B}} T}{N v_0} \,, \tag{6.99}$$

which is the ideal gas law for polymers.

For $N^{-1} \ll \phi \ll 1$, we expand the logarithm and obtain

$$\begin{aligned} \frac{v_0 \, \Pi}{k_{\text{B}} T} &= \frac{1}{N} \phi + \tfrac{1}{2}(1 - 2\chi) \, \phi^2 + \mathcal{O}(\phi^3) \\ &\approx \tfrac{1}{2}(1 - 2\chi) \, \phi^2 \,. \end{aligned} \tag{6.100}$$

Note that $\Pi > 0$ only if $\chi < \tfrac{1}{2}$, which is the condition for a 'good solvent'. The case $\chi = \tfrac{1}{2}$ is that of the θ-solvent.

In fact, Eqn. 6.100 is only qualitatively correct. In the limit where $\chi \ll \tfrac{1}{2}$, Flory showed that the individual polymer coils behave much as hard spheres of radius R_{F}. The osmotic pressure then satisfies something analogous to a virial equation of state:

$$\begin{aligned} \frac{\Pi}{k_{\text{B}} T} &= \frac{\phi}{N v_0} + A \left(\frac{\phi}{N v_0} \right)^2 R_{\text{F}}^3 + \ldots \\ &= \frac{\phi}{N v_0} \, h(\phi/\phi^*) \,. \end{aligned} \tag{6.101}$$

This is generalized to a scaling form in the second line, where $h(x)$ is a scaling function, and $\phi^* = Nv_0/R_{\mathrm{F}}^3 \propto N^{-4/5}$, assuming $d = 3$ and $\nu = \frac{3}{5}$ from Flory theory. As $x = \phi/\phi^* \to 0$, we must recover the ideal gas law, so $h(x) = 1 + \mathcal{O}(x)$ in this limit. For $x \to \infty$, we require that the result be independent of the degree of polymerization N. This means $h(x) \propto x^p$ with $\frac{4}{5}p = 1$, *i.e.* $p = \frac{5}{4}$. The result is known as the des Cloiseaux law:

$$\frac{v_0 \Pi}{k_{\mathrm{B}} T} = C \, \phi^{9/4} \; , \tag{6.102}$$

where C is a constant. This is valid for what is known as semi-dilute solutions, where $\phi^* \ll \phi \ll 1$. In the dense limit $\phi \sim 1$, the results do not exhibit this universality, and we must appeal to liquid state theory, which is no fun at all.

6.2.6 Surface growth

We've explored the subject of stochastic differential equations in chapter 3 of these notes. Those examples all involved ordinary SDEs of the form

$$dx = f(x,t)\, dt + g(x,t)\, dW(t) \quad , \tag{6.103}$$

where $W(t)$ is a Wiener process. Many (most?) physical systems of interest are extended objects described by space and time dependent *fields*. In such cases, we might consider an extension of stochastic ordinary differential equations (SODEs) to stochastic partial differential equations (SPDEs), which can be thought of as a continuum limit of coupled SODEs. For example, consider the system of coupled SODEs described by the equations

$$dh_{\boldsymbol{R}}(t) = K \sum_{\mu=1}^{d} \left[h_{\boldsymbol{R}+\boldsymbol{a}_\mu}(t) + h_{\boldsymbol{R}-\boldsymbol{a}_\mu}(t) - 2h_{\boldsymbol{R}}(t) \right] dt + \sqrt{\Gamma}\, a^{-d/2}\, dW_{\boldsymbol{R}}(t) \quad , \tag{6.104}$$

where each $h_{\boldsymbol{R}}(t)$ lives on a site $\boldsymbol{R}$ of a d-dimensional cubic lattice, with $\boldsymbol{a}_\mu = a\,\hat{\boldsymbol{e}}_\mu$ and a being the lattice constant. The Wiener processes $\{W_{\boldsymbol{R}}(t)\}$ are independent at different sites, so

$$\left\langle W_{\boldsymbol{R}}(t)\, W_{\boldsymbol{R}'}(t') \right\rangle = \delta_{\boldsymbol{R},\boldsymbol{R}'} \min(t,t') \, . \tag{6.105}$$

The $a^{-d/2}$ factor in Eqn. 6.104 is in anticipation of the continuum limit where $\boldsymbol{R} \to \boldsymbol{x}$ and $\delta_{\boldsymbol{R},\boldsymbol{R}'} \to a^d\, \delta(\boldsymbol{x} - \boldsymbol{x}')$. Expanding $h(\boldsymbol{R} + \boldsymbol{a}_\mu)$ in a Taylor series, one finds that the first nonvanishing term in the sum on the RHS is at second order, hence the continuum limit is

$$dh = D\, \nabla^2 h\, dt + \sqrt{\Gamma}\, dW(\boldsymbol{x},t) \quad , \tag{6.106}$$

where $D = Ka^2$ and $\left\langle W(\boldsymbol{x},t)\, W(\boldsymbol{r}',t') \right\rangle = \delta(\boldsymbol{x} - \boldsymbol{x}') \min(t,t')$. We can write this as a conventional Langevin equation as well, *viz.*

$$\frac{\partial h}{\partial t} = D\, \nabla^2 h + \eta(\boldsymbol{x},t) \quad , \qquad \left\langle \eta(\boldsymbol{x},t)\, \eta(\boldsymbol{x}',t') \right\rangle = \Gamma\, \delta(\boldsymbol{x} - \boldsymbol{x}')\, \delta(t - t') \quad . \tag{6.107}$$

Note that this SPDE is linear in the field $h(\boldsymbol{x},t)$. It is called the Edwards-Wilkinson equation, and has been applied to the phenomenon of surface growth. In this application, the field $h(\boldsymbol{x},t) = H(\boldsymbol{x},t) -$

$\langle\!\langle H(\boldsymbol{x},t)\rangle\!\rangle$ denotes the fluctuation in the surface height from its space and time average. We now consider the evolution of this SPDE in different space dimensions.

Let the instantaneous variance of the height be the disorder average

$$w^2(t) = \langle h^2(\boldsymbol{x},t)\rangle \quad . \tag{6.108}$$

Assuming spatial homogeneity, this average is independent of the location $\boldsymbol{x}$. Without diffusion, the height $h(\boldsymbol{x},t)$ at each point in space executes its own independent Wiener process, and the local variance is proportional to the elapsed time t. The coefficient is divergent, however, and from the discrete model is known to be Γa^{-d}, which diverges in the continuum limit $a \to 0$. For the continuum equation, dimensional analysis says that $[D] = L^2 T^{-1}$ and $[\Gamma] = L^{d+2}\, T^{-1}$, hence there is a dimensionless parameter $r \equiv D^{(d+2)/2}\, t^{d/2}/\Gamma$, and we expect on dimensional grounds $w^2(t) = Dt\, f(r)$. Since we also expect $w^2 \propto \Gamma$, we have $f(r) = C/r$ with C a constant, which says

$$w^2(t) \stackrel{?}{=} C\Gamma D^{-d/2}\, t^{(2-d)/2} \quad . \tag{6.109}$$

In $d = 1$ this is correct. In $d = 2$, as we shall see, a logarithm appears. For $d > 2$ this makes no sense at all, since it says the height fluctuations decay with time. The problem, as we shall see, is that there is another scale in the problem, arising from a short distance cutoff which we may take to be the lattice constant a itself. This introduces a new dimensionless parameter which is Dt/a^2.

The solution to Eqn. 6.107, with $h(\boldsymbol{x},0) = 0$, is

$$h(\boldsymbol{x},t) = \int d^d x_1 \int_0^t dt_1 \left[4\pi D(t-t_1)\right]^{-d/2} \exp\left\{-\frac{(\boldsymbol{x}-\boldsymbol{x}_1)^2}{4D(t-t_1)}\right\} \eta(\boldsymbol{x}_1,t_1) \, . \tag{6.110}$$

From this we may derive a formal expression for the correlation function,

$$C_d(\boldsymbol{x},s\,;\,t) \equiv \langle h(\boldsymbol{0},t)\, h(\boldsymbol{x},t+s)\rangle \quad . \tag{6.111}$$

Note that the correlator does not depend on $\boldsymbol{x}$, due to spatial isotropy, but does depend on both t and s time variables[4]. We will consider the equal time ($s = 0$) correlator,

$$C_d(\boldsymbol{x},0\,;\,t) = \langle h(\boldsymbol{0},t)\, h(\boldsymbol{x},t)\rangle = \Gamma\, e^{-\boldsymbol{x}^2/4Dt} \int d^d u \int_0^t d\tau\, \frac{e^{-\boldsymbol{u}^2/2D\tau}\, e^{-\boldsymbol{x}\cdot\boldsymbol{u}/2D\tau}}{(4\pi D\tau)^d}$$

$$= \frac{\Gamma\, |\boldsymbol{x}|^{2-d}}{2\pi^{d/2} D} \int_{\boldsymbol{x}^2/8Dt}^{\infty} ds\, \frac{e^{-s}}{s^{(4-d)/2}} = \frac{\Gamma\, |\boldsymbol{x}|^{2-d}}{2\pi^{d/2} D}\, E_{2-\frac{d}{2}}\left(\boldsymbol{x}^2/8Dt\right) \quad , \tag{6.112}$$

where $E_k(z)$ is familiar from Eqn. 6.29. It is also interesting to consider the correlation function for height *differences*,

$$R_d(\boldsymbol{x},t) \equiv \left\langle \left[h(\boldsymbol{x},t) - h(\boldsymbol{0},t)\right]^2\right\rangle = 2\Big[C_d(\boldsymbol{0},0\,;\,t) - C_d(\boldsymbol{x},0\,;\,t)\Big] \quad . \tag{6.113}$$

---

[4]We may assume, without loss of generality, that $s \geq 0$.

For $d = 1$, we integrate by parts once and obtain

$$C_1(x, 0\,;\, t) = \left(\frac{\Gamma^2 t}{2\pi D}\right)^{1/2} e^{-x^2/8Dt} - \frac{\Gamma|x|}{4\sqrt{\pi}\,D}\, E_{1/2}\big(x^2/8Dt\big) \quad . \tag{6.114}$$

In the limit $x \to 0$, the second term on the RHS vanishes, and we obtain $C_1(0\,,0\,;\, t) = \sqrt{\frac{2}{\pi}}\,\Gamma(t/D)^{1/2}$, which agrees with the dimensional analysis. The height difference correlator $R_1(x, t)$ is then

$$R_1(x, t) = \left(\frac{\Gamma^2 t}{2\pi D}\right)^{1/2}\left(1 - e^{-x^2/8Dt}\right) + \frac{\Gamma|x|}{2\sqrt{\pi}\,D}\, E_{1/2}\big(x^2/8Dt\big) \quad . \tag{6.115}$$

As $t \to \infty$, we have $E_{1/2}(0) = \sqrt{\pi}$ and thus $R_1(x, t \to \infty) = \Gamma|x|/2D$, which says that the height function $h(x, t \to \infty)$ is a random walk in the spatial coordinate x.

In $d = 2$, we have

$$C_2(\boldsymbol{x}, 0\,;\, t) = \frac{\Gamma}{2\pi D}\, E_1(\boldsymbol{x}^2/8Dt) = \frac{\Gamma}{2\pi D}\left\{\ln\left(\frac{8Dt}{\boldsymbol{x}^2}\right) - \gamma_{\mathrm{E}} + \mathcal{O}\big(\boldsymbol{x}^2/t\big)\right\} \quad , \tag{6.116}$$

where the expansion is for the long time limit, and where $\gamma_{\mathrm{E}} \simeq 0.577215$ is the Euler-Mascheroni constant. This diverges logarithmically as $t \to \infty$ or $\boldsymbol{x} \to 0$. For $d > 2$, the $t \to \infty$ limit yields

$$C_d(\boldsymbol{x}, 0\,;\, t \to \infty) = \frac{\Gamma(\frac{1}{2}d - 1)}{2\pi^{d/2} D}\, \Gamma\, |\boldsymbol{x}|^{2-d} \quad , \tag{6.117}$$

where one should take care to distinguish the Gamma function $\Gamma(\frac{1}{2}d - 1)$ from the parameter Γ. This is independent of time but diverges as $\boldsymbol{x} \to 0$. The short distance divergence is a pathology which is cured by the introduction of a new length scale a corresponding to an ultraviolet cutoff in the theory. One then replaces $\boldsymbol{x}^2$ in these formulae for $d \geq 2$ with $\mathsf{max}(\boldsymbol{x}^2, a^2)$. We conclude then that for $d > 2$ the random term does not roughen the interface, *i.e.* the height fluctuations do not diverge as $t \to \infty$.

We can derive a scaling form for the space and time dependent correlation function $\langle h(\boldsymbol{x}, t)\, h(\boldsymbol{x}, t')\rangle$ in the limit where t and t' are both large. The Fourier transform of the EW equation is

$$-i\omega \hat{h}(\boldsymbol{k}, \omega) = -D\boldsymbol{k}^2\, \hat{h}(\boldsymbol{k}, \omega) + \hat{\eta}(\boldsymbol{k}, \omega) \quad . \tag{6.118}$$

In Fourier space, the correlations of the stochastic term are given by $\langle \hat{h}(\boldsymbol{k}, \omega)\rangle = 0$ and

$$\langle \hat{\eta}(\boldsymbol{k}, \omega)\, \hat{\eta}(\boldsymbol{k}', \omega')\rangle = (2\pi)^{d+1}\, \Gamma\, \delta(\boldsymbol{k} + \boldsymbol{k}')\, \delta(\omega + \omega') \quad , \tag{6.119}$$

from which we obtain

$$\langle \hat{h}(\boldsymbol{k}, \omega)\, \hat{h}(\boldsymbol{k}', \omega')\rangle = \frac{(2\pi)^{d+1}\, \Gamma\, \delta(\boldsymbol{k} + \boldsymbol{k}')\, \delta(\omega + \omega')}{(D\boldsymbol{k}^2)^2 + \omega^2} \quad . \tag{6.120}$$

Here we have neglected any transients, which is consistent with our assumption that we are in the late time phase. Fourier transforming back to the space-time domain, we obtain the scaling form

$$\langle h(\boldsymbol{x}, t)\, h(\boldsymbol{x}', t')\rangle = A_d\, \frac{\Gamma}{D}\, |\boldsymbol{x} - \boldsymbol{x}'|^{2-d}\, f\left(\frac{D|t - t'|}{|\boldsymbol{x} - \boldsymbol{x}'|^2}\right) \quad , \tag{6.121}$$

where A_d is a d-dependent constant and $f(\zeta)$ is given by

$$f(\zeta) = \int\limits_0^\infty du\, u^{(d-4)/2}\, J_{\frac{d}{2}-1}(u)\, e^{-\zeta u^2} \quad. \tag{6.122}$$

The integral is convergent for $d > 2$, with $f(\zeta \to \infty) \sim \zeta^{(2-d)/2}$.

Generalized EW model

Consider now the more general case

$$-i\omega \hat{h}(\boldsymbol{k}, \omega) = -B|\boldsymbol{k}|^p\, \hat{h}(\boldsymbol{k}, \omega) + \hat{\eta}(\boldsymbol{k}, \omega) \quad. \tag{6.123}$$

Proceeding as before, we obtain

$$\langle h(\boldsymbol{x}, t)\, h(\boldsymbol{x}', t') \rangle = \frac{(2\pi)^{d/2}\Gamma}{4B}\, |\boldsymbol{x} - \boldsymbol{x}'|^{p-d}\, f_{d,p}(\zeta) \quad, \tag{6.124}$$

where $\zeta = B|t - t'|/|\boldsymbol{x} - \boldsymbol{x}'|^p$ is the scaling variable and[5]

$$f_{d,p}(\zeta) = \int\limits_0^\infty du\, u^{\frac{d}{2}-p}\, J_{\frac{d}{2}-1}(u)\, e^{-\zeta u^p} \quad, \tag{6.125}$$

which is convergent for $d > p$, with $f(\zeta \to \infty) \sim \zeta^{(p-d)/p}$.

For $d \le p$ the integral is divergent. If we start with initial conditions $h(\boldsymbol{x}, 0) = 0$, then we find

$$\langle h(\boldsymbol{x}, t)\, h(\boldsymbol{x}', t') \rangle = \frac{(2\pi)^{d/2}\Gamma}{4B}\, |\boldsymbol{x} - \boldsymbol{x}'|^{p-d} \left[f_{d,p}(\zeta) - f_{d,p}(Z) \right] \quad, \tag{6.126}$$

where $Z = B(t + t')/|\boldsymbol{x} - \boldsymbol{x}'|^p$. For $d > p$, when $f_{d,p}(w)$ converges, the second term is negligible as t and t' tend to infinity, with $|t - t'|$ finite. For $d \le p$, we have that $f_{d,p}(w)$ is divergent, however, the difference,

$$f_{d,p}(\zeta) - f_{d,p}(Z) = \int\limits_0^\infty du\, u^{\frac{d}{2}-p}\, J_{\frac{d}{2}-1}(u) \left[e^{-\zeta u^p} - e^{-Z u^p} \right] \tag{6.127}$$

converges. This amounts to imposing a lower limit cutoff on u in Eqn. 6.125 of $u_{\mathrm{min}} \sim Z^{-1/p}$ when $Z \gg 1$. The height-height correlator then behaves as $(t + t')^{(p-d)/p}$, which diverges in the late time limit. For $p = d$ the correlator behaves as $\ln Z$. *Thus, for $d \le p$ the surface roughens.*

---

[5]To derive this result, we invoke

$$\int \frac{d\hat{k}}{\Omega_d}\, e^{iz\hat{k}\cdot\hat{n}} = \Gamma(d/2) \left(\frac{2}{z} \right)^{\frac{d}{2}-1} J_{\frac{d}{2}-1}(z) \quad,$$

where the integral is over the surface of a unit sphere in d space dimensions, and where $\hat{\boldsymbol{n}}$ is any unit vector. The RHS approaches 1 in the limit $z \to 0$.

Kardar-Parisi-Zhang equation

The Edwards-Wilkinson equation is a linear stochastic partial differential equation. A nonlinear extension of the EW equation for surface growth was proposed by Kardar, Parisi, and Zhang, and accordingly is known as the KPZ equation,

$$\frac{\partial h}{\partial t} = D\,\nabla^2 h + \tfrac{1}{2}\lambda\,(\nabla h)^2 + \eta \quad , \tag{6.128}$$

where $\eta(x,t)$ is the same stochastic noise term. On physical grounds, the nonlinearity in this equation is rather generic. It may be transformed to the Burgers equation with noise for a vorticity-free field, via $v \equiv -\lambda\,\nabla h$, whence

$$\frac{\partial v}{\partial t} + (v \cdot \nabla)v = D\,\nabla^2 v - \lambda\,\nabla\eta(x,t) \quad . \tag{6.129}$$

Dimensionally, we still have $[\Gamma] = L^{d+2}\,T^{-1}$ and $[D] = L^2\,T^{-1}$, but now we add $[\lambda] = L\,T^{-1}$ to the mix. There are now two dimensionless parameters $\Gamma^2/D^{d+2}\,t^2$ and $\Gamma\lambda^d/D^{d+1}$. However, because the transverse coordinates x and the height h enter the equation in different ways, we should really distinguish between these coordinates and define a transverse length scale L as well as a height length scale H. In this case, we have

$$[\Gamma] = L^d\,H^2\,T^{-1} \quad , \quad [\lambda] = L^2\,H^{-1}\,T^{-1} \quad , \quad [D] = L^2\,T^{-1} \quad , \tag{6.130}$$

and the only properly dimensionless combination is

$$\kappa = \frac{\Gamma^2\lambda^4}{D^{d+4}} \times t^{2-d} \quad . \tag{6.131}$$

The instantaneous height variance $w^2(t)$ and the spatial correlation length $\xi(t)$ should then scale with units of H^2 and L, respectively, hence we expect

$$w(t) = \frac{D}{\lambda}\,f(\kappa) \quad , \qquad \xi(t) = (Dt)^{1/2}g(\kappa) \quad . \tag{6.132}$$

Note in $d = 1$ we have $\kappa = \Gamma^2\,\lambda^4\,t/D^5$. Applied to the EW equation, where $\lambda = 0$, this analysis recovers $w(t) \sim \Gamma^{1/2}D^{-d/4}\,t^{(2-d)/4}$ and $\xi \sim (Dt)^{1/2}$, but note that our earlier argument was rooted in the linearity of the EW equation, which requires $w \propto \Gamma^{1/2}$. The dimensional argument does not specifically invoke linearity in this way.

There is not much more that can be said about the KPZ equation in dimensions $d > 1$ without resorting to more sophisticated analysis, but in $d = 1$, much is known. For example, a nonlinear transformation known as the *Cole-Hopf transformation*,

$$\psi(x,t) = \exp\left(\frac{\lambda}{2D}\,h(x,t)\right) \quad , \tag{6.133}$$

transforms KPZ to a linear SPDE,

$$\frac{\partial\psi}{\partial t} = D\,\frac{\partial^2\psi}{\partial x^2} + \frac{\lambda}{2D}\,\psi\,\eta \quad . \tag{6.134}$$

This describes diffusion in the presence of a random potential.

The probability distribution $\Pi\big[h(\boldsymbol{x}),t\big]$ for the *field* $h(\boldsymbol{x})$ at time t obeys a *functional Fokker-Planck equation*,

$$\frac{\partial \Pi\big[h(\boldsymbol{x}),t\big]}{\partial t} = \int \! d^dx' \left(\tfrac{1}{2}\Gamma \, \frac{\delta^2}{\delta h(\boldsymbol{x}')^2} - \frac{\delta}{\delta h(\boldsymbol{x}')} \, J(\boldsymbol{x}') \right) \Pi\big[h(\boldsymbol{x}),t\big] \quad , \tag{6.135}$$

where

$$J = D \, \nabla^2 h + \tfrac{1}{2}\lambda \, (\nabla h)^2 \quad . \tag{6.136}$$

To make sense of this and avoid ill-defined expressions like $\delta''(0)$, we may write the functional Fokker-Planck equation as

$$\frac{\partial \Pi\big[h(\boldsymbol{x}),t\big]}{\partial t} = \lim_{\varepsilon \to 0} \int \! d^dx' \left(\tfrac{1}{2}\Gamma \, \frac{\delta^2}{\delta h(\boldsymbol{x}') \, \delta h(\boldsymbol{x}' + \varepsilon)} - \frac{\delta}{\delta h(\boldsymbol{x}')} \, J(\boldsymbol{x}' + \varepsilon) \right) \Pi\big[h(\boldsymbol{x}),t\big] \quad , \tag{6.137}$$

In one dimension, we have the stationary solution

$$\Pi\big[h(x)\big] = \exp\left\{ -\frac{D}{\Gamma} \int\limits_{-\infty}^{\infty} \! dx \left(\frac{\partial h}{\partial x} \right)^2 \right\} \quad . \tag{6.138}$$

When $\lambda = 0$, this solution generalizes to arbitrary d, but for nonzero λ it is valid only for $d = 1$. Because the asymptotic distribution there depends only on the ratio D/Γ, we conclude that the asymptotic behaviors of $w(t)$ and $\xi(t)$ must do the same, in which case we must have $f(\kappa) \propto \kappa^{1/3}$ and $g(\kappa) \propto \kappa^{1/6}$, resulting in

$$w(t) \sim (\Gamma/D)^{2/3} \, (\lambda t)^{1/3} \qquad , \qquad \xi(t) \sim (\Gamma/D)^{1/3} \, (\lambda t)^{2/3} \tag{6.139}$$

for the one-dimensional KPZ equation. The characteristic $w \sim t^{1/3}$ growth is called KPZ growth.

Scaling and exponents

The mean height of a surface is

$$\bar{h}(t) = L^{-d} \int \! d^dx \, h(\boldsymbol{x},t) \quad , \tag{6.140}$$

where the integration is over a region of characteristic linear dimension L. The *interface width* $w(L,t)$ is given by

$$w(L,t) = \left[L^{-d} \int \! d^dx \, \big(h(\boldsymbol{x},t) - \bar{h}(t) \big)^2 \right]^{1/2} \quad . \tag{6.141}$$

Given these intuitive and precise definitions, we introduce the following concepts. The *growth exponent* β is defined such that for $t \ll \tau(L)$ the interface width grows as $w(L, t \ll \tau) \sim t^\beta$. The time $\tau(L) \sim L^z$ is a characteristic scale which increases as a power law with *dynamical critical exponent* z. In the long time limit $t \gg \tau(L)$, the interface width goes as $w(L, t \gg \tau) \sim L^\alpha$, where α is the *roughness exponent*. For $L \to \infty$, the interface width obeys a *scaling relation*

$$w(L,t) \sim L^\alpha \, f\big(t/L^z\big) \quad . \tag{6.142}$$

In order that $w(L, t \ll \tau) \sim t^\beta$, we must have $f(u) \sim u^{\alpha/z}$, in which case we read off $z = \alpha/\beta$, which is a *scaling relation*.

For the EW equation, we may derive the exponents α, β, and z from our calculations of the correlation functions. However there is a slicker way to do this, which is by scaling space x, time t, and height h and demanding the EW equation retain its form. Let us write $x \to x' = bx$, $h \to h' = b^\alpha h$, and $t \to t' = b^z t$. Space derivatives scale as $\nabla \to \nabla' = b^{-1}\nabla$, time derivatives as $\partial_t \to \partial_{t'} = b^{-z}\partial_t$, and the noise as $\eta \to \eta' = b^{-(d+z)/2}\eta$, because

$$\langle \eta(bx, b^z t)\, \eta(bx', b^z t')\rangle = \Gamma\, \delta(bx - bx')\, \delta(b^z t - b^z t') = \Gamma b^{-(d+z)}\, \delta(x - x')\, \delta(t - t') \quad . \tag{6.143}$$

Under this rescaling, then, we have

$$b^{\alpha-z}\frac{\partial h}{\partial t} = b^{\alpha-2} D\, \nabla^2 h + b^{-(d+z)/2}\eta \quad , \tag{6.144}$$

and demanding that the EW equation retain its form means

$$\alpha - z = \alpha - 2 = -\tfrac{1}{2}(d+z) \qquad \Rightarrow \qquad \alpha = \frac{2-d}{2} \quad , \quad \beta = \frac{2-d}{4} \quad , \quad z = 2 \quad , \tag{6.145}$$

where we have used $\beta = \alpha/z$. One can verify that these exponents describe our earlier exact solution.

What happens when we try to apply these scaling arguments to KPZ? Evidently we wind up with a rescaled equation

$$b^{\alpha-z}\frac{\partial h}{\partial t} = b^{\alpha-2} D\, \nabla^2 h + \tfrac{1}{2}b^{2\alpha-2}\lambda\,(\nabla h)^2 + b^{-(d+z)/2}\eta \quad , \tag{6.146}$$

which yields three equations for the two unknowns α and z, *viz.*

$$\alpha - z = \alpha - 2 = 2\alpha - 2 = -\tfrac{1}{2}(d+z) \quad . \tag{6.147}$$

This is overdetermined – clearly something has gone wrong with our scaling arguments. The resolution is that the coefficients D, λ, and Γ themselves are scale-dependent. A proper treatment requires the invocation of renormalization group technology. Still we may argue on general grounds, from the Burgers equation form of KPZ, that the convective derivative,

$$\frac{Dv}{Dt} = \frac{\partial v}{\partial t} + (v \cdot \nabla)\, v \quad , \tag{6.148}$$

must retain its form under rescaling. If we write[6] $v = -\nabla h$ instead of $v = -\lambda\,\nabla h$, then λ multiplies the $(v \cdot \nabla)\, v$ term, and if we set $\lambda = 1$ we conclude that λ should not change under rescaling. Thus leads to the relation $\alpha + z = 2$ in all dimensions. We still have $\beta = \alpha/z$, so we need just one more equation to determine all three exponents. In $d = 1$, Eqn. 6.138 implies a roughening exponent of $\alpha = \tfrac{1}{2}$, hence we conclude for the KPZ equation in $d = 1$ that

$$\alpha = \tfrac{1}{2} \quad , \quad \beta = \tfrac{1}{3} \quad , \quad z = \tfrac{3}{2} \quad . \tag{6.149}$$

These values have been confirmed numerically.

---

[6]Warning! Slick argument imminent!

6.2.7 Lévy flights

We follow the discussion in KRB §2.3. We saw earlier in §1.4.2 how the sum of N independent random variables $X = \sum_{j=1}^{N} x_j$ is distributed as a Gaussian in the $N \to \infty$ limit, a consequence of the central limit theorem. If $p(x)$ is the single step distribution, then $P_N(X) = (2\pi N \sigma^2)^{-1/2} \exp\left[-(X - N\mu)^2 / 2N\sigma^2\right]$, where μ and σ are the mean and standard deviation of $p(x)$, respectively. This presumes that μ and σ exist. Suppose that

$$p(x) = \begin{cases} r \, x^{-(1+r)} & x \geq 1 \\ 0 & x < 1 \, . \end{cases} \tag{6.150}$$

Here we consider a process where each step is to the right ($x > 0$), but we could easily allow for leftward steps as well. The distribution is normalized, and we exclude steps of length less than one so we can retain a simple power law that is still normalizable. Clearly $\mu = \langle x \rangle$ is finite only if $r > 1$ and $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$ is finite only if $r > 2$. What happens if $r < 2$?

For a walk of N steps, the mean and standard deviation of X will necessarily be finite, because each step is itself finite. Let's now ask: what is the typical value of the *largest* among the individual steps $\{x_j\}$? Suppose we demand that the largest of these values be x. Then the probability distribution for x is

$$M_N(x) = N \left[1 - P(x)\right]^{N-1} p(x) \quad , \tag{6.151}$$

where $P(x) = \int_x^\infty dx' \, p(x')$ is the probability that a given step lies in the range $[x, \infty)$. The factor of N above arises because any among the N steps could be the largest. Note that $dP(x) = -p(x) \, dx$, hence

$$\int_0^\infty dx \, M_N(x) = N \int_0^1 dP \, (1 - P)^{N-1} = 1 \, , \tag{6.152}$$

so $M_N(x)$ is normalized. If $P(x) = \mathcal{O}(N^{-1})$, we may write Eqn. 6.151 as $M_N(x) \approx p(x) \, e^{-NP(x)}$ and then extract a typical value for the maximum step $x_{\max}(N)$ by setting $NP(x) \approx 1$, *i.e.* by setting $\int_x^\infty dx' \, p(x') \sim N^{-1}$. For the power law distribution in Eqn. 6.150, this yields $x_{\max}(N) \sim N^{1/r}$. KRB compute the average

$$\langle x_{\max}(N) \rangle = \int_0^\infty dx \, x \, M_N(x) = N \int_0^1 ds \, (1 - s)^{N-1} s^{-1/r} \stackrel{?}{=} \frac{\Gamma(1 - r^{-1}) \, \Gamma(N + 1)}{\Gamma(N + 1 - r^{-1})} \, . \tag{6.153}$$

For $N \to \infty$ this yields $\langle x_{\max}(N) \rangle = \Gamma(1 - r^{-1}) \, N^{1/r}$, which has the same dependence on N, but includes a prefactor. Unfortunately, this prefactor arises from a divergent integral if $r < 1$, as the above equation shows, but which KRB let pass without comment. Indeed, if the average *single* step length diverges, then the average *greatest* step length among N steps *surely* diverges! A more sensible definition of $x_{\max}(N)$ is obtained by setting the integral of $M_N(x)$ up to $x_{\max}(N)$ to some value α on the order of unity, such as $\alpha = \frac{1}{2}$:

$$N \int_0^{x_{\max}} dx \, M_N(x) = \alpha \quad \Rightarrow \quad x_{\max}(N) = \left(\frac{N}{\ln(1/\alpha)}\right)^{1/r} \, . \tag{6.154}$$

This again is proportional to $N^{1/r}$, but with a finite coefficient for all r. We may then write $x_{\max}(N) = C_r \, N^{1/r}$, where C_r is an r-dependent $\mathcal{O}(1)$ constant.

We may now approximate the single-step distribution for an N-step walk as

$$\tilde{p}(x) \equiv p(x) \, \Theta(x_{\max} - x) \bigg/ \int\limits_0^{x_{\max}} dx' \, p(x') \tag{6.155}$$

$$= \frac{r \, x^{-(1+r)}}{1 - x_{\max}^{-r}} \, \Theta(x_{\max} - x) \simeq r \, x^{-(1+r)} \, \Theta(x_{\max} - x) \quad .$$

Then for large N one has

$$\langle x \rangle = \begin{cases} A_r \, N^{(1-r)/r} & \text{if} \quad r < 1 \\ \ln N + A_1 & \text{if} \quad r = 1 \\ r/(r-1) & \text{if} \quad r > 1 \end{cases} \quad \Rightarrow \quad \langle X \rangle = \begin{cases} A_r \, N^{1/r} & \text{if} \quad r < 1 \\ N \ln N + A_1 N & \text{if} \quad r = 1 \\ rN/(r-1) & \text{if} \quad r > 1 \,. \end{cases} \tag{6.156}$$

Similarly,

$$\langle x^2 \rangle = \begin{cases} A'_r \, N^{(2-r)/r} & \text{if} \quad r < 2 \\ \ln N + A'_1 & \text{if} \quad r = 2 \\ r/(r-2) & \text{if} \quad r > 2 \end{cases} \quad \Rightarrow \quad \langle X^2 \rangle - \langle X \rangle^2 = \begin{cases} A'_r \, N^{2/r} & \text{if} \quad r < 2 \\ N \ln N + A'_1 N & \text{if} \quad r = 2 \\ rN/(r-2) & \text{if} \quad r > 2 \,. \end{cases} \tag{6.157}$$

These are examples of *Lévy flights*. The Lévy distribution $L_{\alpha,\beta}(x)$ is defined in terms of its Fourier transform, $\hat{L}_{\alpha,\beta}(k)$,

$$\hat{L}_{\alpha,\beta}(k) = \exp\left\{ i\mu k - \left(1 - i\beta \, \mathsf{sgn}(k) \, \phi(k,\alpha)\right) \sigma^\alpha |k|^\alpha \right\} , \tag{6.158}$$

where

$$\phi(k,\alpha) = \begin{cases} \tan\left(\tfrac{1}{2}\pi\alpha\right) & \text{if} \quad \alpha \neq 1 \\ -\tfrac{2}{\pi} \ln|k| & \text{if} \quad \alpha = 1 \,. \end{cases} \tag{6.159}$$

This is a four parameter distribution, specified by the *index* $\alpha \in [0,2]$, which corresponds to r in Eqn. 6.150, the *skewness* β, the *shift* μ, and the *scale* σ. Of these, the shift and the scale are uninteresting, because

$$L_{\alpha,\beta}(x \, ; \, \mu \, , \sigma) = \int\limits_{-\infty}^{\infty} \frac{dk}{2\pi} \, \hat{L}_{\alpha,\beta}(k) \, e^{ikx} = L_{\alpha,\beta}\left(\frac{x - \mu}{\sigma} \, ; \, \mu = 0 \, , \sigma = 1\right) . \tag{6.160}$$

Without loss of generality, then, we may set $\mu = 0$ and $\sigma = 1$, in which case we are left with the two-parameter family,

$$\hat{L}_{\alpha,\beta}(k) = \exp\left\{ -\left(1 - i\beta \, \mathsf{sgn}(k) \, \tan(\tfrac{1}{2}\pi\alpha)\right)|k|^\alpha \right\} . \tag{6.161}$$

When the skewness vanishes ($\beta = 0$), we obtain the symmetric Lévy distribution, $\hat{L}_{\alpha,0}(k) = \exp(-|k|^\alpha)$. We can compute the inverse Fourier transform analytically in two cases:

$$L_{1,0}(x) = \frac{1}{\pi} \frac{1}{x^2 + 1} \quad , \quad L_{2,0}(x) = \frac{1}{\sqrt{4\pi}} \, e^{-x^2/4} \quad , \tag{6.162}$$
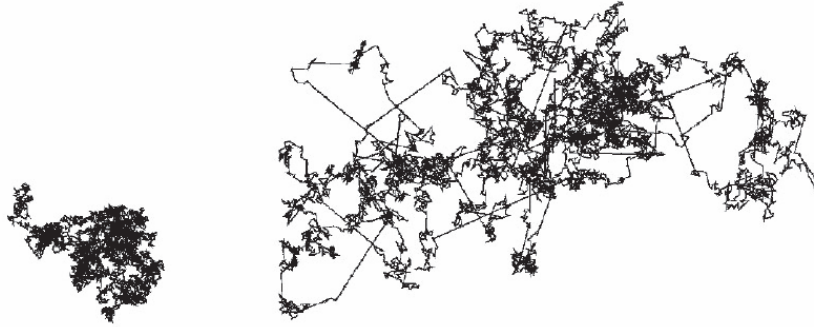
Figure 6.6: Diffusion process (left) and a Lévy flight aith $\alpha = \frac{3}{2}$ (right). Both walks contain approximately $N = 7000$ steps. The Lévy process is characterized by blobs connected by long steps and is superdiffusive. From A. V. Chechkin *et al.* in *Anomalous Transport: Foundations and Applications*, R. Klages *et al.*, eds. (Wiley-VCH, 2008).

which are the Cauchy (Lorentzian) and the Gaussian distributions, respectively. Asymptotically, we have

$$L_{\alpha,0}(x) \sim \frac{\Gamma(1+\alpha)\,\sin(\frac{1}{2}\alpha\pi)}{\pi\,|x|^{1+\alpha}} \quad \big(\,|x| \to \infty\,\big)\,. \tag{6.163}$$

An example of an asymmetric Lévy distribution is the Lévy-Smirnoff form,

$$L_{\frac{1}{2},1}(x) = \frac{1}{\sqrt{2\pi}}\,x^{-3/2}\,\exp\!\left(-\frac{1}{2x}\right)\Theta(x)\,. \tag{6.164}$$

A special property of the Lévy distributions is their *stability*, which means that the distribution of a sum of N independent and identically distributed random Lévy variables itself is a Lévy distribution. If $\hat{P}(k) = \hat{L}_{\alpha,0}(k)$, for example, then for the sum $X = \sum_{j=1}^{N} x_j$ we have $\hat{P}_N(k) = \exp\big(-N|k|^{\alpha}\big)$, and

$$P_N(X) = \frac{1}{N^{\alpha}}\,L_{\alpha,0}\!\left(\frac{X}{N^{1/\alpha}}\right)\,. \tag{6.165}$$

Note that the width of the distribution is $N^{1/\alpha}$, so for $\alpha < 2$ we have $N^{1/\alpha} \gg \sqrt{N}$ as $N \to \infty$, hence the Lévy distribution is much broader than the usual Gaussian.

The Lévy flight arising from a power law distribution of step lengths is *superdiffusive*, with $\langle x^2 \rangle \propto t^{2/r} > t$ and $r < 2$. What happens if the step length size is normally distributed, but the *waiting time* between consecutive steps is power law distributed as $\psi(\tau) = r\,\tau^{-(1+r)}\,\Theta(\tau)$? Following KRB, the maximum waiting time for an N-step process is then obtained from the extremal condition

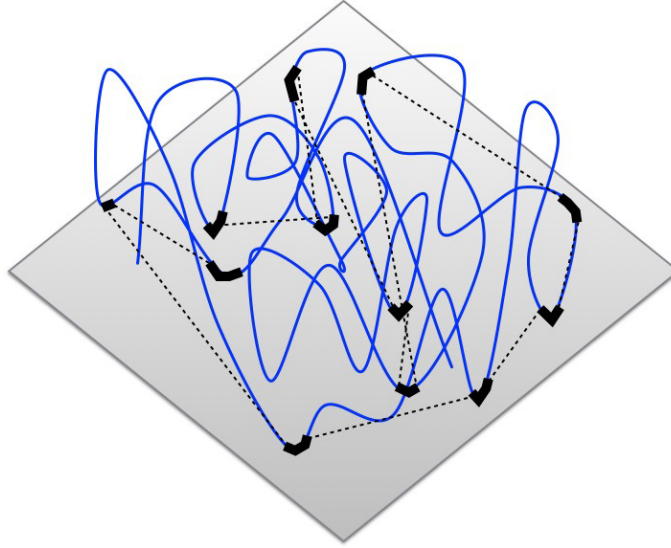$$r\int\limits_{\tau_{\text{max}}}^{\infty} d\tau\,\tau^{-(1+r)} \sim \frac{1}{N}\quad, \tag{6.166}$$

Figure 6.7: A physical example of a Lévy flight, a polymer in contact with a surface. The polymer often leaves the surface to explore three-dimensional space, and touches down again a long distance away from its previous point of contact.

whence $\tau_{\max}(N) \sim N^{1/r}$. The average time to take a step and total time T_N for N steps are then

$$
\langle t \rangle \sim \int_0^{\tau_{\max}} d\tau\, \mu\, \tau^{-r} =
\begin{cases}
B_r\, N^{(1-r)/r} & \text{if} \quad r < 1 \\
\ln N + B_1 & \text{if} \quad r = 1 \\
r/(r-1) & \text{if} \quad r > 1
\end{cases}
\quad \Rightarrow \quad
T_N = N\langle t \rangle =
\begin{cases}
B_r\, N^{1/r} & \text{if} \quad r < 1 \\
N \ln N + B_1 N & \text{if} \quad r = 1 \\
r N/(r-1) & \text{if} \quad r > 1
\end{cases}
$$
(6.167)

and therefore

$$
\langle X^2 \rangle \sim N =
\begin{cases}
B'_r\, T^r & \text{if} \quad r < 1 \\
T/\ln T & \text{if} \quad r = 1 \\
(r-1)\, T/r & \text{if} \quad r > 1 \,.
\end{cases}
$$
(6.168)

For $r < 1$, this process is *subdiffusive*, spreading more slowly than ordinary diffusion.

6.2.8   Holtsmark distribution

Consider a distribution of equal mass objects, which we can imagine to be stars, which are equally dense throughout the universe. We seek the distribution $P(\boldsymbol{F})$ of the force acting on any given star. We will compute this by placing, without loss of generality, our 'test star' at the origin $\boldsymbol{r} = 0$ and then computing the force on it from all stars within a radius R, then take $R \to \infty$ at the end of the calculation. We have that

$$
\boldsymbol{F}(R) = \sum_{j=1}^{N} \boldsymbol{f}_j = -\sum_{j=1}^{N} \frac{GM^2\, \hat{\boldsymbol{r}}_j}{r_j^2} \quad ,
$$
(6.169)

where N is the number of other stars within a sphere of radius R. Assuming the stars are independently and identically distributed with number density n, we have

$$P(\boldsymbol{F}) = V_R^{-N} \int d^3x_1 \cdots \int d^3x_N \, \delta\!\left(\boldsymbol{F} - \sum_{j=1}^{N} \boldsymbol{f}_j\right) \quad,$$

(6.170)

with $V_R = \frac{4}{3}\pi R^3$, the Fourier transform of which is

$$\hat{P}(\boldsymbol{k}) = \int d^3F \, P(\boldsymbol{F}) \, e^{-i\boldsymbol{k}\cdot\boldsymbol{F}} = \left(V_R^{-1} \int\limits_{r<R} d^3r \, e^{-iGM^2\boldsymbol{k}\cdot\hat{\boldsymbol{r}}/r^2}\right)^N$$

$$= \left(1 - \frac{n}{N} \int\limits_{r<R} d^3r \left(1 - e^{-iGM^2\boldsymbol{k}\cdot\hat{\boldsymbol{r}}/r^2}\right)\right)^N = \exp\left(-n\,\hat{\Phi}(\boldsymbol{k})\right) \quad,$$

(6.171)

where we have taken the $N \to \infty$ limit with $n = N/V_R$ fixed, and where we have defined

$$\hat{\Phi}(\boldsymbol{k}) = \int d^3r \left(1 - e^{iGM^2\boldsymbol{k}\cdot\hat{\boldsymbol{r}}/r^2}\right).$$

(6.172)

This integral may be taken over all space, as we shall see. Note that $\boldsymbol{k}$ has dimensions of inverse force. Integrating over the solid angle $\hat{\boldsymbol{r}}$, we have $\hat{\Phi}(\boldsymbol{k})$ is isotropic, with

$$\hat{\Phi}(\boldsymbol{k}) = 4\pi \int\limits_0^\infty dr \, r^2 \left(1 - \frac{\sin\left(GM^2k/r^2\right)}{GM^2k/r^2}\right)$$

$$= 2\pi(GM^2k)^{3/2} \int\limits_0^\infty du \, \frac{u - \sin u}{u^{7/2}} = \tfrac{4}{15}\,(2\pi)^{3/2}\,(GM^2)^{3/2}\,k^{3/2} \quad.$$

(6.173)

We define the dimensional force unit $F_0 \equiv GM^2n^{2/3}$ and the dimensionless wavevector $\boldsymbol{\kappa} \equiv F_0\,\boldsymbol{k}$. Then

$$P(\boldsymbol{F}) = F_0^{-3} \int \frac{d^3\kappa}{(2\pi)^3} \, e^{i\boldsymbol{\kappa}\cdot\boldsymbol{\xi}} \, e^{-C\kappa^{3/2}} \quad,$$

(6.174)

where $\boldsymbol{F} \equiv F_0\,\boldsymbol{\xi}$, with $C = \frac{4}{15}(2\pi)^{3/2} = 4.12$. Thus, the dimensionless force distribution $P(w) = F_0^3\,P(\boldsymbol{F})$ is

$$P(\xi) = \frac{1}{2\pi^2\xi} \int\limits_0^\infty d\kappa \, \kappa \, \sin(\kappa\,\xi) \, \exp\!\left(-C\kappa^{3/2}\right).$$

(6.175)

This expression has two limiting forms. In the weak force limit $\xi \to 0$, we may write $\sin(\kappa\,\xi) \approx \kappa\,\xi$ in which case

$$P(\xi \ll \xi_0) = \frac{1}{2\pi^2} \int\limits_0^\infty d\kappa \, \kappa^2 \, \exp\!\left(-C\kappa^{3/2}\right) = \frac{1}{3\pi^2 C^2} = \frac{75}{128\,\pi^5} = 1.9 \times 10^{-3}\,.$$

(6.176)

Thus, the distribution is flat for $\xi \ll \xi_0 \equiv C^{-2/3} = 0.384$. In the opposite limit $\xi \gg C^{-2/3}$, we expand the exponential in Eqn. 6.175, write $\sin(\kappa\,\xi) = \text{Im}\, e^{i\kappa\xi}$, and introduce a convergence factor $e^{-\epsilon k}$ with $\varepsilon \to 0$ at the end of the calculation. The final result is

$$P(\xi \gg \xi_0) = \frac{1}{2\pi^2\xi}\, \text{Im} \lim_{\epsilon \to 0} \int\limits_0^\infty d\kappa\,\kappa\, e^{i\kappa\,\xi} \left(1 - C\kappa^{3/2} + \dots\right) e^{-\epsilon\kappa} = \tfrac{1}{2}\,\xi^{-9/2}\ . \tag{6.177}$$

For a central force $\boldsymbol{f}(\boldsymbol{r}) = A\,\hat{\boldsymbol{r}}/r^\beta$, one has $n\,\hat{\varPhi}(\boldsymbol{k}) = C_\beta\,(F_0\,k)^{3/\beta}$, with $F_0 = A\,n^{\beta/3}$ and

$$C_\beta = \frac{4\pi}{\beta} \int\limits_0^\infty du\, \frac{u - \sin u}{u^{2+3/\beta}}\quad . \tag{6.178}$$

We are now in position to compute moments of the force distribution. We have

$$\langle F^v \rangle = 4\pi F_0^v \int\limits_0^\infty d\xi\, \xi^{2+v}\, P(\xi) = A_v\, F_0^v\quad , \tag{6.179}$$

with

$$A_v = \frac{\sin(\pi v/2)\,\Gamma(2+v)}{\sin(2\pi v/3)\,\Gamma(1 + \tfrac{2}{3}v)} \cdot \tfrac{4}{3}\, C^v\ . \tag{6.180}$$

The moments are finite provided $v \in \left[-3\,,\,\tfrac{3}{2}\right]$. In the strong force limit, the average force is dominated by the statistically closest other star.

6.3 Aggregation

In the process of *aggregation*, two clusters of different size join irreversibly. Starting from an initial distribution of cluster sizes, the distribution coarsens under the sole constraint of total mass conservation. Aggregation describes physical processes from the accretion of stellar matter to the coagulation of proteins in the production of cheese. Here we follow the pellucid presentation in chapter five of KRB.

6.3.1 Master equation dynamics

The basic aggregation process is schematically described by the reaction

$$A_i + A_j \xrightarrow{K_{ij}} A_{i+j}\quad , \tag{6.181}$$

where A_i denotes a cluster of size/mass i. We do not distinguish between different shapes of clusters; the only relevant variable in describing the cluster is its total mass. The rate constants K_{ij} have dimensions $L^d\, T^{-1}$ and, when multiplied by a concentration c whose dimensions are $[c] = L^{-d}$, yield a reaction rate. The matrix of rate constants is symmetric: $K_{ij} = K_{ji}$.
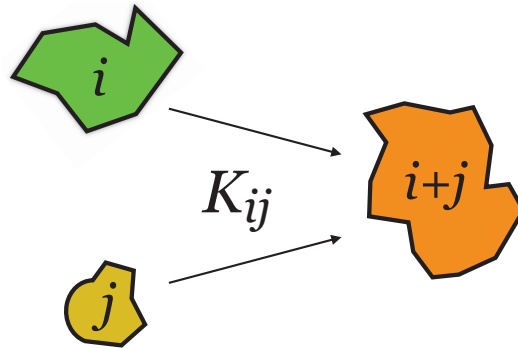
Figure 6.8: Aggregation process in which two clusters of mass i and j combine to form a cluster of mass $i + j$.

Let $c_n(t)$ be the concentration of clusters of mass n at time t. The dynamics of the cluster size concentrations is given, at the mean field level, by a set of nonlinear coupled ODEs,

$$
\begin{aligned}
\frac{dc_n}{dt} &= \frac{1}{2} \sum_{i,j=1}^{\infty} K_{ij}\, c_i\, c_j \left\{ \delta_{n,i+j} - \delta_{n,i} - \delta_{n,j} \right\} \\
&= \frac{1}{2} \sum_{i+j=n} K_{ij}\, c_i\, c_j - c_n \sum_{j=1}^{\infty} K_{nj}\, c_j \quad .
\end{aligned}
\tag{6.182}
$$

Several comments are in order here:

(i) The dynamics here are assumed to be spatially independent. A more realistic model invoking diffusion would entail a set of coupled PDEs of the form

$$
\frac{\partial c_n}{\partial t} = D_n \nabla^2 c_n + \frac{1}{2} \sum_{i+j=n} K_{ij}\, c_i\, c_j - c_n \sum_{j \geq 1} K_{nj}\, c_j \quad ,
\tag{6.183}
$$

where D_n is the diffusion constant for clusters of mass n. If diffusion is fast, the different clusters undergo rapid spatial homogenization, and we can approximate their dynamics by Eqn. 6.182.

(ii) Unlike the Master equation (see §2.5 and §2.6.3), the aggregation dynamics of Eqn. 6.182 are nonlinear in the concentrations. This represents an approximation to a much more complicated hierarchy akin to the BBGKY hierarchy in equilibrium statistical physics. The probability of a reaction $A_i + A_j \to A_{i+j}$ is proportional to the *joint* probability of finding a cluster A_i and a cluster A_j at the same position in space at the same time. If $c_n(\boldsymbol{r}, t) = P(n\,;\, \boldsymbol{r}, t)$ is the probability density to find a cluster of mass n at position $\boldsymbol{r}$ at time t, and $P(n_1, n_2\,;\, \boldsymbol{r}, ;\, t)$ is the probability density for finding two clusters of masses n_1 and n_2 at position $\boldsymbol{r}$ at time t, then we should write

$$
\frac{\partial P(n\,;\, \boldsymbol{r}, t)}{\partial t} = D_n \nabla^2 P(n\,;\, \boldsymbol{r}, t) + \frac{1}{2} \sum_{i+j=n} K_{ij}\, P(i, j\,;\, \boldsymbol{r}, t) - \sum_{j \geq 1} K_{nj}\, P(n, j\,;\, \boldsymbol{r}, t)
\tag{6.184}
$$

This is not a closed set of equations inasmuch as the dynamics of the single cluster distribution is dependent on the two cluster distribution. At the next level of the hierarchy, the rate of change of

the two cluster distribution will be given in terms of the three cluster distribution. To recover Eqn. 6.182, we approximate

$$P(i, j\,;\,\boldsymbol{r}, t) \approx P(i\,;\,\boldsymbol{r}, t)\, P(j\,;\,\boldsymbol{r}, t) = c_i(\boldsymbol{r}, t)\, c_j(\boldsymbol{r}, t) \quad . \tag{6.185}$$

Assuming diffusion rapidly induces spatial uniformity of the cluster densities, we have $c_j(\boldsymbol{r}, t) \approx c_j(t)$.

(iii) The factor of one half on the RHS of Eqn. 6.182 is explained as follows. The number of *pairs* of clusters of masses i and j, with $i \neq j$, is $N_i\, N_j$, where $N_i = V c_i$ where V is the volume. The number of pairs where both clusters have mass k is $\frac{1}{2} N_k(N_k - 1) \approx \frac{1}{2} N_k^2$, where the approximation is valid in the thermodynamic limit. Note that there is no factor of one half for the $j = n$ term in the second sum on the RHS of Eqn. 6.182 because the reaction $A_n + A_n \to A_{2n}$ results in the loss of two A_n clusters, and this factor of two cancels with the above factor of one half.

(iv) Three body aggregation $A_i + A_j + A_k \to A_{i+j+k}$ is ignored on the presumption that the reactants are sufficiently dilute. Note that the aggregation process itself leads to increasing dilution in terms of the number of clusters per unit volume.

6.3.2   Moments of the mass distribution

Define the kth moment of the mass distribution,

$$\nu_k(t) = \sum_{n=1}^{\infty} n^k\, c_n(t) \quad . \tag{6.186}$$

Then from Eqn. 6.182 we have

$$\frac{d\nu_k}{dt} = \frac{1}{2} \sum_{i,j=1}^{\infty} K_{ij}\, c_i\, c_j \left\{ (i+j)^k - i^k - j^k \right\} \quad . \tag{6.187}$$

For $k = 1$ the RHS vanishes, hence $\dot{\nu}_1 = 0$ and the total mass density ν_1 is conserved by the dynamics. This is of course expected, since mass is conserved in each reaction $A_i + A_j \to A_{i+j}$.

6.3.3   Constant kernel model

The general equation 6.182 cannot be solved analytically. A great simplification arises if we assume a *constant kernel* K_{ij} is a constant, independent of i and j, as proposed by Smoluchowski (1917). What justifies such a seemingly radical assumption? As KRB discuss, if we assume the aggregating clusters are executing Brownian motion, then we can use the results of §6.2.4, which says that the rate constant for a diffusing particle to hit a sphere of radius R is $4\pi D R$, where D is the particle's diffusion constant. For two spherical particles of sizes i and j to meet, we have $K_{ij} \approx 4\pi(D_i + D_j)(R_i + R_j)$. Now the diffusion constant for species i is $D_i = k_{\mathrm{B}} T / 6\pi\eta R_i$, where η is the kinematic viscosity of the solvent in which the clusters move. Thus,

$$K_{ij} \approx \frac{k_{\mathrm{B}} T}{6\pi\eta} \left\{ 2 + \left(\frac{i}{j} \right)^{1/3} + \left(\frac{j}{i} \right)^{1/3} \right\} \quad , \tag{6.188}$$

where we have used $R_i \propto i^{1/3}$, for a particle of mass i. This kernel is not constant, but it does share a scale invariance $K_{i,j} = K_{ri,rj}$, for all $r \in \mathbb{Z}_+$, with any constant kernel model. This feature is supposed to give us a warm fuzzy feeling about the constant kernel model. Let's assume, then, that $K_{ij} = 2\alpha$, so

$$\frac{1}{\alpha}\frac{dc_n}{dt} = \sum_{i+j=n} c_i\, c_j - 2\nu_0 c_n = \sum_{j=1}^{n-1} c_j\, c_{n-j} - 2\nu_0 c_n \quad, \tag{6.189}$$

where $\nu_0(t) = \sum_{j=1}^{\infty} c_j(t)$ is the total cluster concentration, accounting for all possible masses, at time t. The resulting hierarchy is

$$\alpha^{-1}\dot{c}_1 = -2\nu_0\, c_1 \qquad\qquad \alpha^{-1}\dot{c}_4 = 2c_1 c_3 + c_2^2 - 2\nu_0\, c_4 \tag{6.190}$$
$$\alpha^{-1}\dot{c}_2 = c_1^2 - 2\nu_0\, c_2 \qquad\qquad \alpha^{-1}\dot{c}_5 = 2c_1 c_4 + 2c_2 c_3 - 2\nu_0\, c_5 \tag{6.191}$$
$$\alpha^{-1}\dot{c}_3 = 2c_1 c_2 - 2\nu_0\, c_3 \qquad\qquad \alpha^{-1}\dot{c}_6 = 2c_1 c_5 + 2c_2 c_4 + c_3^2 - 2\nu_0\, c_6 \,. \tag{6.192}$$

From Eqn. 6.187, $\nu_0(t)$ obeys

$$\dot{\nu}_0(t) = -\alpha\, \nu_0^2 \quad \Rightarrow \quad \nu_0(t) = \frac{\nu_0(0)}{1 + \nu_0(0)\,\alpha\, t} \quad. \tag{6.193}$$

The $k = 1$ moment $\nu_1(t)$ is conserved by the evolution. The equations for the higher moments $\nu_k(t)$ with $k > 1$ are

$$\dot{\nu}_k = \alpha \sum_{l=1}^{k-1} \binom{k}{l} \nu_l\, \nu_{k-l} \quad. \tag{6.194}$$

Generating function solution

Remarkably, the nonlinear hierarchy of the constant kernel model may be solved analytically via the generating function formalism[7]. We define

$$c(z,t) = \sum_{n=1}^{\infty} z^n\, c_n(t) \quad. \tag{6.195}$$

Multiplying both sides of Eqn. 6.189 by z^n and summing on n, we obtain

$$\frac{\partial c(z,t)}{\partial t} = \alpha\, c^2(z,t) - 2\alpha\nu_0(t)\, c(z,t) \quad. \tag{6.196}$$

Subtract from this the equation from $\dot{\nu}_0 = -\alpha\, \nu_0^2$, to obtain

$$\frac{\partial h(z,t)}{\partial t} = -\alpha\, h^2(z,t) \quad \Rightarrow \quad h(z,t) = \frac{h(z,0)}{1 + h(z,0)\,\alpha\, t} \quad, \tag{6.197}$$

where $h(z,t) = \nu_0(t) - c(z,t)$. Therefore

$$c(z,t) = \frac{\nu_0(0)}{1 + \nu_0(0)\,\alpha\, t} - \frac{\nu_0(0) - c(z,0)}{1 + \big[\nu_0(0) - c(z,0)\big]\alpha\, t} \quad. \tag{6.198}$$

---

[7]See §2.5.3 and §4.3.2.

The cluster distribution $c_n(t)$ is the coefficient of z^n in the above expression. Note that $c(z,0) = \sum_j z^j c_j(0)$ is given in terms of the initial cluster distribution, and that $\nu_0(0) = c(z=1, t=0)$.

As an example, consider the initial condition $c_n(0) = \kappa \delta_{n,m}$. We then have $c(z,0) = \kappa z^m$ and thus $\nu_0(0) = \kappa$, and

$$c(z,t) = \frac{\kappa}{1 + \kappa\alpha t} - \frac{\kappa(1 - z^m)}{1 + \kappa\alpha t(1 - z^m)} = \frac{u(1 - u\alpha t)z^m}{1 - u\alpha t z^m} \quad , \tag{6.199}$$

where $u = \kappa/(1 + \kappa\alpha t)$. We can extract the distribution $c_n(t)$ by inspection. Note that $c(z,t)$ contains only integer powers of z^m, because clusters whose mass is an integer multiple of m can only aggregate to produce clusters whose mass is a larger integer multiple of m. One finds

$$c_{lm}(t) = \frac{\kappa(\kappa\alpha t)^{l-1}}{(1 + \kappa\alpha t)^{l+1}} = \frac{1}{\kappa\alpha^2 t^2}\left(1 + \frac{1}{\kappa\alpha t}\right)^{-(l+1)} \quad . \tag{6.200}$$

Note that the RHS does not depend on m, which is a manifestation of the aforementioned scale invariance of the constant kernel (and the diffusion model kernel). The total cluster density agrees with Eqn. 6.193:

$$\nu_0(t) = \sum_{n=1}^{\infty} c_n(t) = \frac{\kappa}{1 + \kappa\alpha t} \quad . \tag{6.201}$$

One can further check that the total mass density is conserved:

$$\nu_1(t) = \sum_{n=1}^{\infty} n\, c_n(t) = \left.\frac{\partial c(z,t)}{\partial z}\right|_{z=1} = m\kappa \quad . \tag{6.202}$$

Asymptotically as $t \to \infty$ with l fixed, we have $c_{lm}(t) \simeq 1/\kappa(\alpha t)^2$, with a universal t^{-2} falloff. For $l \to \infty$ with t fixed, we have that $c_{lm}(t) \sim e^{-l\lambda}$, where $\lambda = \ln(1 + \kappa\alpha t) - \ln(\kappa\alpha t)$. For $t \to \infty$ and $l \to \infty$ with $l \propto t$, we have

$$c_{lm}(t) \simeq \frac{1}{\kappa\alpha^2 t^2}\exp\left(-\frac{l}{\kappa\alpha t}\right) \quad . \tag{6.203}$$

KRB also discuss the case where the initial conditions are given by

$$c_n(0) = \kappa(1 - \lambda)\lambda^n \quad \Rightarrow \quad \nu_0(0) = \kappa \quad , \quad c(z,0) = \frac{\kappa(1 - \lambda)}{1 - \lambda z} \quad . \tag{6.204}$$

Solving for $c(z,t)$, one finds

$$c(z,t) = \frac{\kappa}{1 + \kappa\alpha t} \cdot \frac{1 - \lambda}{1 + \lambda\kappa\alpha t - \lambda(1 + \kappa\alpha t)z} \quad , \tag{6.205}$$

from which we derive

$$c_n(t) = \frac{\kappa(1 - \lambda)}{(1 + \kappa\alpha t)^2}\left(\frac{1 + \kappa\alpha t}{\lambda^{-1} + \kappa\alpha t}\right)^n \quad . \tag{6.206}$$

The asymptotic behavior is the same as for the previous case, where $c_n(0) = \kappa\delta_{n,m}$. The cluster densities $c_n(t)$ fall off as t^{-2} as $t \to \infty$.
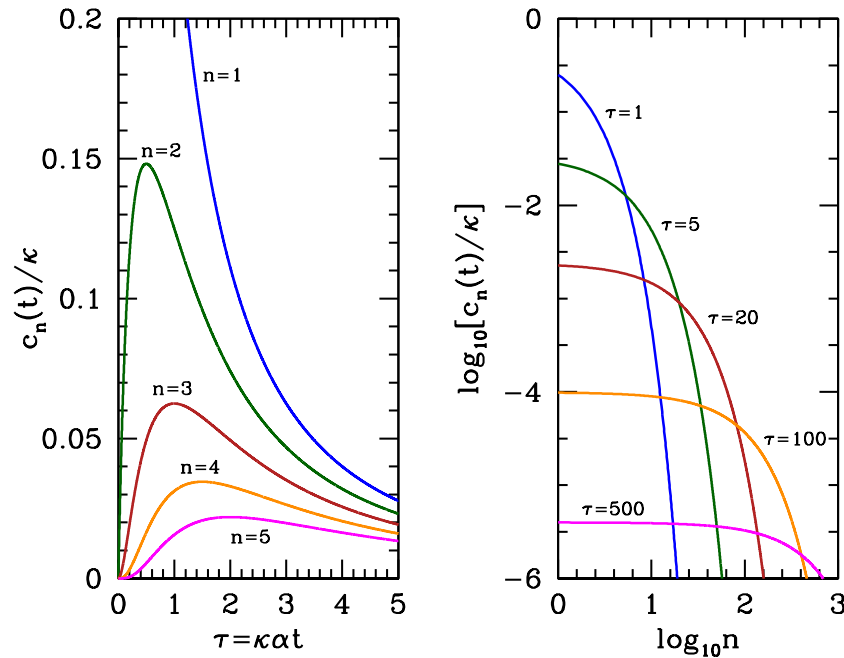
Figure 6.9: Results for the constant kernel model of aggregation with initial conditions $c_n(0) = \kappa\,\delta_{n,1}$. Left panel: cluster densities $c_n(t)$ *versus* dimensionless time $\tau = \kappa\alpha t$. Note that $\kappa^{-1}c_{n=1}(0) = 1$ is off-scale. Right panel: cluster densities $c_n(t)$ *versus* cluster mass n for different times. (Adapted from KRB Fig. 5.2.)

Power law distribution

Consider now the power law distribution,

$$c_n(0) = \frac{\kappa}{\zeta(s)}\,n^{-s} \quad \Rightarrow \quad \nu_0(0) = \kappa \quad , \quad c(z,0) = \frac{\kappa\,\mathrm{Li}_s(z)}{\zeta(s)} \quad , \tag{6.207}$$

where

$$\mathrm{Li}_s(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^s} \tag{6.208}$$

is the *polylogarithm* function, and $\zeta(s) = \mathrm{Li}_s(1)$ is the Riemann zeta function. One has[8]

$$\mathrm{Li}_s(z) = \Gamma(1-s)\left(-\ln z\right)^{s-1} + \sum_{k=0}^{\infty} \frac{\zeta(s-k)}{k!}\left(\ln z\right)^k$$
$$= \zeta(s) + \Gamma(1-s)\left(-\ln z\right)^{s-1} + \mathcal{O}(\ln z) \quad , \tag{6.209}$$

for $s \notin \mathbb{Z}^+$. Note also that $z\frac{d}{dz}\,\mathrm{Li}_s(z) = \mathrm{Li}_{s-1}(z)$. If the zeroth moment $\nu_0(0)$ is to converge, we must have $s > 1$.

---

[8]See §25.12 of the *NIST Handbook of Mathematical Functions.*

If the first moment $\nu_1(t)$, which is constant, converges, then the asymptotics of the cluster densities $c_n(t)$ are of the familiar t^{-2} form. This is the case for $s > 2$. It is therefore interesting to consider the case $s \in [1, 2]$.

From the generating function solution Eqn. 6.198, we have

$$c(z,t) = \frac{\kappa}{1 + \kappa\,\alpha\,t} + \frac{\kappa}{\kappa\,\alpha\,t}\left\{\frac{1}{1 + \kappa\,\alpha\,t - \kappa\,\alpha\,t\,\mathsf{Li}_s(z)/\zeta(s)} - 1\right\} \quad . \tag{6.210}$$

Now

$$1 - \frac{\mathsf{Li}_s(z)}{\zeta(s)} = A_s\,(-\ln z)^{s-1} + \mathcal{O}(\ln z) \quad , \tag{6.211}$$

with $A_s = -\Gamma(1-s)/\zeta(s) = -\pi/\Gamma(s)\,\zeta(s)\sin(\pi s) > 0$. For the asymptotic behavior as $t \to \infty$, we focus on the first term on the RHS above. Then we must compute

$$c_{n>1}(t) \approx \frac{1}{\alpha\,t}\oint\frac{dz}{2\pi i z}\frac{1}{z^n}\frac{1}{1 + A_s\,\kappa\,\alpha\,t\,(-\ln z)^{s-1}} = \frac{f\big(n/\zeta(t)\big)}{\alpha\,t\,\zeta(t)} \quad , \tag{6.212}$$

where

$$\zeta(t) = \big(A_s\,\kappa\,\alpha\,t\big)^{1/(s-1)} \tag{6.213}$$

and

$$f(w) = \mathsf{Re}\int_{-i\pi\zeta}^{i\pi\zeta}\frac{du}{2\pi i}\frac{e^{wu}}{1 + u^{s-1}} \quad . \tag{6.214}$$

In the long time limit, the range of integration may be extended to the entire imaginary axis. Asymptotically,

$$f(w) = \begin{cases} w^{s-2}/\Gamma(s-1) & w \to 0 \\ -w^{-s}/\Gamma(1-s) & w \to \infty \end{cases} \tag{6.215}$$

6.3.4   Aggregation with source terms

Let's now add a source to the RHS of Eqn. 6.189, *viz.*

$$\frac{dc_n}{dt} = \alpha\sum_{i+j=n}c_i\,c_j - 2\alpha\nu_0 c_n + \gamma\delta_{n,m} \quad . \tag{6.216}$$

This says that m-mers are fed into the system at a constant rate γ. The generating function is again $c(z,t) = \sum_{n=1}^{\infty} z^n c_n(t)$ and satisfies

$$\frac{\partial c}{\partial t} = \alpha\,c^2 - 2\alpha\nu_0 c + \gamma z^m \quad . \tag{6.217}$$

We still have $\nu_0 = \sum_n c_n = c(z = 1, t)$, hence

$$\frac{\partial\nu_0}{\partial t} = -\alpha\nu_0^2 + \gamma \quad . \tag{6.218}$$

This may be integrated with the substitution $\nu_0 = (\gamma/\alpha)^{1/2} \tanh \theta$, yielding the equation $d\theta = \sqrt{\alpha\gamma}\, dt$. Assuming $\nu_0(0) = 0$, we have $\theta(0) = 0$ and

$$\nu_0(t) = \sqrt{\frac{\gamma}{\alpha}}\, \tanh\left(\sqrt{\alpha\gamma}\, t\right) \quad . \tag{6.219}$$

As $t \to \infty$, the cluster density tends to a constant $\nu_0(\infty) = \sqrt{\gamma/\alpha}$. Note the difference between the cluster dynamics with the source term and the results in Eqn. 6.193, where there is no source and $\nu_0(t) \sim t^{-1}$ at late times. The limiting constant value in the present calculation reflects a dynamic equilibrium between the source, which constantly introduces new m-mers into the system, and the aggregation process, where $A_m + A_{jm} \to A_{(j+1)m}$.

Subtracting $c(z,t)$ from $\nu_0(t)$ as before, we obtain

$$\frac{\partial}{\partial t}\left(\nu_0 - c\right) = -\left(\nu_0 - c\right)^2 + \gamma(1 - z^m) \quad , \tag{6.220}$$

which can be integrated using the same substitution, resulting in

$$c(z,t) = \sqrt{\frac{\gamma}{\alpha}}\left\{ \tanh\left(\sqrt{\alpha\gamma}\, t\right) - \sqrt{1 - z^m}\, \tanh\left(\sqrt{\alpha\gamma(1 - z^m)}\, t\right) \right\}. \tag{6.221}$$

For late times, we have

$$c(z, t \to \infty) = \sqrt{\frac{\gamma}{\alpha}}\left[1 - \sqrt{1 - z^m}\right] \quad , \tag{6.222}$$

and from the Taylor expansion

$$1 - \sqrt{1 - \varepsilon} = \frac{1}{\sqrt{4\pi}} \sum_{k=1}^{\infty} \frac{\Gamma(k - \frac{1}{2})}{\Gamma(k + 1)}\, \varepsilon^k \quad , \tag{6.223}$$

we have

$$c_{jm}(t \to \infty) = \left(\frac{\gamma}{4\pi\alpha}\right)^{1/2} \frac{\Gamma(j - \frac{1}{2})}{\Gamma(j + 1)} \simeq \left(\frac{\gamma}{4\pi\alpha}\right)^{1/2} j^{-3/2} \quad , \tag{6.224}$$

where the last expression is for $j \gg 1$. Note that, as before, the RHS is independent of m due to the scale invariance of the constant kernel model.

While the zeroth moment of the asymptotic distribution $c_n(t \to \infty)$, *i.e.* ν_0, is finite, the quantities ν_k for all integer $k > 0$ diverge. This is because clusters are being fed into the system at a constant rate. Indeed, while the total mass density $\nu_1(t)$ is conserved with no input, when $\gamma \neq 0$ we have $\dot{\nu}_1 = \gamma m$, hence $\nu_1(t) = \gamma m t$, which diverges linearly with time, as it must.

Following KRB, we may utilize the identity

$$\tanh x = \frac{1}{\pi} \sum_{j=-\infty}^{\infty} \frac{x/\pi}{(x/\pi)^2 + \left(j + \frac{1}{2}\right)^2} \tag{6.225}$$

to write

$$
\begin{aligned}
c(z,t) &= \frac{1}{\pi}\left(\frac{\gamma}{\alpha}\right)^{1/2}\sum_{j=-\infty}^{\infty}\left\{\frac{\tau}{\left(j+\frac{1}{2}\right)^2+\tau^2}-\frac{(1-z^m)\tau}{\left(j+\frac{1}{2}\right)^2+\tau^2-\tau^2 z^m}\right\} \\
&= \frac{1}{\pi}\left(\frac{\gamma}{\alpha}\right)^{1/2}\sum_{j=-\infty}^{\infty}\left(j+\tfrac{1}{2}\right)^2\sum_{k=1}^{\infty}\frac{\tau^{2k-1}}{D_j^{k+1}(\tau)}\,z^{km}\quad,
\end{aligned}
$$

(6.226)

where $\tau\equiv(\alpha\gamma)^{1/2}\,t/\pi$ and $D_j(\tau)=\left(j+\frac{1}{2}\right)^2+\tau^2$. Thus,

$$
c_{km}(t)=\frac{1}{\pi}\left(\frac{\gamma}{\alpha}\right)^{1/2}\tau^{2k-1}\sum_{j=-\infty}^{\infty}\frac{\left(j+\frac{1}{2}\right)^2}{D_j^{k+1}(\tau)}\quad.
$$

(6.227)

When $\tau\to\infty$, we can replace

$$
\sum_{j=-\infty}^{\infty}\frac{\left(j+\frac{1}{2}\right)^2}{D_j^{k+1}(\tau)}\approx\int_{-\infty}^{\infty}du\,\frac{u^2}{(u^2+\tau^2)^{k+1}}=\frac{\sqrt{\pi}}{2}\frac{\Gamma(k-\frac{1}{2})}{\Gamma(k+1)}\,\tau^{1-2k}\quad,
$$

(6.228)

which, combined with the previous equation, recovers Eqn. 6.224.

When $t\to\infty$ and $k\to\infty$ such that k/t^2 is constant, we write

$$
D_j^{-(k+1)}(\tau)=\tau^{-2(k+1)}\left[1+\frac{\left(j+\frac{1}{2}\right)^2}{\tau^2}\right]^{-(k+1)}\approx\tau^{-2(k+1)}\exp\left(-\frac{\left(j+\frac{1}{2}\right)^2 k}{\tau^2}\right)
$$

(6.229)

and thus

$$
c_{km}(t)\simeq\frac{\pi^2}{\alpha^2\gamma\,t^3}\sum_{j=-\infty}^{\infty}\left(j+\tfrac{1}{2}\right)^2\exp\left(-\frac{\left(j+\frac{1}{2}\right)^2 k}{\tau^2}\right)\quad.
$$

(6.230)

For $k\gg\tau^2$ we can retain only the $j=0$ term, in which case

$$
c_{km}(t)\simeq\frac{\pi^2}{4\alpha^2\gamma\,t^3}\exp\left(-\frac{\pi^2 k}{4\alpha\gamma\,t^2}\right)\quad.
$$

(6.231)

6.3.5 Gelation

Consider a group of monomers, each of which has f functional end groups. If two monomers aggregate into a dimer, one end group from each monomer participates in the fusion process, and the resulting dimer has $2f-2$ functional end groups. Generalizing to the case of k monomers, the aggregated k-mer has $(f-2)k+2$ functional end groups (see Fig. 6.10). We then expect the kernel K_{ij} to be of the form

$$
K_{ij}\propto\left[(f-2)\,i+2\right]\left[(f-2)\,j+2\right]\quad.
$$

(6.232)

When $f\to\infty$, we have $K_{ij}\propto ij$, and here we consider the case $K_{ij}=\alpha\,i\,j$. The nonlinear growth of K_{ij} as a function of i and j leads to a phenomenon known as *gelation*, in which a cluster of infinite size develops.
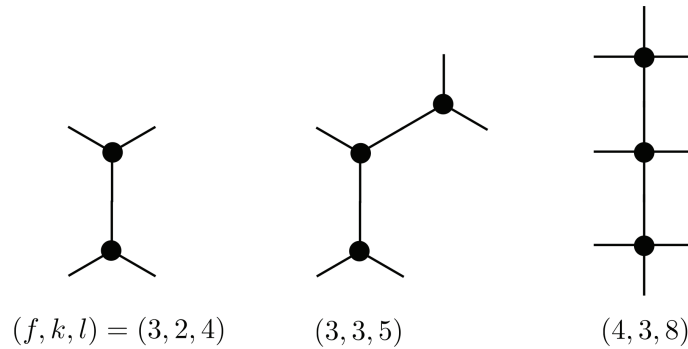
$$(f, k, l) = (3, 2, 4) \qquad (3, 3, 5) \qquad (4, 3, 8)$$

Figure 6.10: Examples of k-mers, each with f functional end groups. The resulting aggregates have $l = (f - 2)k + 2$ functional end groups. (Adapted from KRB Fig. 5.3.)

From the dynamical equations in 6.182, we have

$$\frac{1}{\alpha} \frac{dc_n}{dt} = \frac{1}{2} \sum_{i+j=n} (i\, c_i)(j\, c_j) - n\, c_n \overbrace{\sum_{j=1}^{\infty} j\, c_j}^{\nu_1 = \text{fixed}} \quad . \tag{6.233}$$

We can solve this using a modified generating function, defined as

$$c(u, t) = \sum_{n=1}^{\infty} n\, c_n(t)\, e^{-nu} \quad , \tag{6.234}$$

which satisfies

$$\frac{\partial c}{\partial t} = \frac{1}{2}\alpha \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} (i+j)(i\, c_i)(j\, c_j)\, e^{-(i+j)u} - \alpha\nu_1 \sum_{n=1}^{\infty} n^2 c_n\, e^{-nu}$$
$$= \alpha(\nu_1 - c)\frac{\partial c}{\partial u} \quad . \tag{6.235}$$

Writing $q \equiv c - \nu_1$, we have $\partial_t q + \alpha\, q\, \partial_u q = 0$, which is the *inviscid Burgers equation*. This may be solved using the method of characteristics outlined in §2.10. We introduce a variable s and solve

$$\frac{dt}{ds} = \frac{1}{\alpha} \quad , \qquad \frac{du}{ds} = c - \nu_1 \quad , \qquad \frac{dc}{ds} = 0 \quad . \tag{6.236}$$

The solution is $t = s/\alpha$ and $u = (c - \nu_1)\,\alpha t + \zeta$, where ζ encodes the initial conditions, which are

$$c(u, t = 0) = \sum_{n=1}^{\infty} n\, c_n(0)\, e^{-nu} \quad . \tag{6.237}$$

We assume $c_n(t = 0) = \kappa\, \delta_{n,1}$, in which case $c(u, 0) = \kappa\, e^{-u}$, and therefore $\zeta = -\ln(c/\kappa)$. We then have the implicit solution

$$c(u, t)\, e^{-\alpha\, t\, c(u,t)} = e^{-\kappa\alpha t}\, e^{-u} \quad . \tag{6.238}$$

It is convenient to measure $c_n(t)$ and $c(u,t)$ in units of κ, so we define $\bar{c}_n(t) = c_n(t)/\kappa$ and $\bar{c}(u,t) = c(u,t)/\kappa$. We further define the dimensionless time variable $\tau \equiv \kappa \, \alpha \, t$, so that

$$\bar{c}\, e^{-\tau \bar{c}} = e^{-(u+\tau)} \quad . \tag{6.239}$$

To obtain the $\bar{c}_n(\tau)$, we must invert this to find $\bar{c}(u,\tau)$, extract the coefficient of e^{-nu}, and then divide by n.

To invert the above equation, we invoke a method due to Lagrange. Suppose we have a function $y(x) = \sum_{n=1}^{\infty} A_n \, x^n$ and we wish to invert this to obtain $x(y) = \sum_{n=1}^{\infty} B_n \, y^n$. We have

$$B_n = \oint \frac{dy}{2\pi i} \frac{x(y)}{y^{n+1}} = \oint \frac{dx}{2\pi i} \frac{dy}{dx} \frac{x(y)}{y^{n+1}} = \oint \frac{dx}{2\pi i} \frac{x\, y'(x)}{\left[y(x)\right]^{n+1}} \quad . \tag{6.240}$$

Using our equation as an example, we have $x \equiv \tau \, \bar{c}$, $y(x) = x\, e^{-x}$, and $y = \tau \, e^{-(u+\tau)}$. Then $f'(x) = (1-x)\, e^{-x}$ and the expansion coefficients B_n are

$$\begin{aligned} B_n &= \oint \frac{dx}{2\pi i} \frac{x\,(1-x)\, e^{-x}}{x^{n+1}\, e^{-(n+1)\,x}} = \oint \frac{dx}{2\pi i} \frac{1-x}{x^n}\, e^{nx} \\ &= \frac{n^{n-1}}{(n-1)!} - \frac{n^{n-2}}{(n-2)!} = \frac{n^{n-1}}{n!} \quad . \end{aligned} \tag{6.241}$$

Thus,

$$\bar{c}(u,\tau) = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!}\, \tau^{n-1}\, e^{-n\tau}\, e^{-nu} \quad , \tag{6.242}$$

from which we extract

$$\bar{c}_n(\tau) = \frac{n^{n-2}}{n!}\, \tau^{n-1}\, e^{-n\tau} \quad . \tag{6.243}$$

For $n \gg 1$ we may use Stirling's expansion,

$$\ln n! = n \ln n - n + \tfrac{1}{2} \ln(2\pi n) + \mathcal{O}(n^{-1}) \tag{6.244}$$

to obtain

$$\bar{c}_n(\tau) \simeq \frac{n^{-5/2}}{\sqrt{2\pi}}\, \tau^{-1}\, e^{n(1-\tau+\ln \tau)} \quad . \tag{6.245}$$

The function $f(\tau) \equiv 1 - \tau + \ln \tau$ is concave and nonpositive over $\tau \in (0,\infty)$, with a local maximum at $\tau = 1$ where $f(\tau) = -\tfrac{1}{2}(1-\tau)^2 + \dots$. At the *gelation time* $\tau = 1$, the cluster density distribution becomes a power law $c_n(\tau = 1) \propto n^{-5/2}$, which means that the second and all higher moments are divergent at this point. For both $\tau < 1$ and $\tau > 1$ there is an exponential decrease with n, but for $\tau > 1$ an infinite cluster is present. This is the gel.

We define the gel fraction by

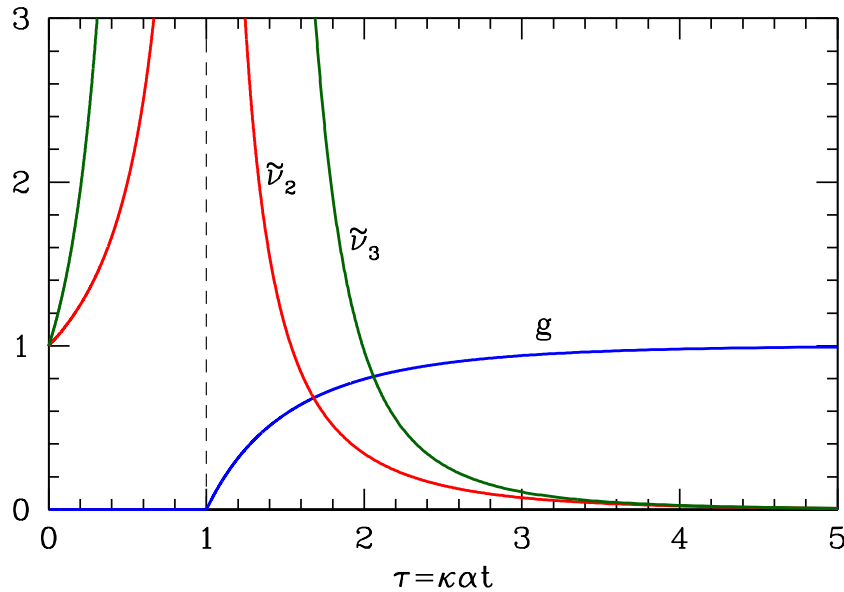$$g \equiv 1 - \bar{c}(0,t) = 1 - \sum_{n=1}^{\infty} n\, \bar{c}_n(t) \quad . \tag{6.246}$$

Figure 6.11: Gelation model time evolution, showing gel fraction $g(\tau)$ and dimensionless moments $\tilde{\nu}_2(\tau)$ and $\tilde{\nu}_3(\tau)$ in terms of dimensionless time $\tau = \kappa\alpha t$, with initial conditions $c_n(t = 0) = \kappa\,\delta_{n,1}$. (Adapted from KRB Fig. 5.4.)

If we plug this into Eqn. 6.239, we obtain

$$\bar{c}(0,\tau) = 1 - g = e^{-g\tau} \quad , \tag{6.247}$$

which is an implicit equation for the time-dependent gelation fraction $g(\tau)$. This equation always has the solution $g = 0$, but for $\tau > 1$ there is a second solution with $g \in (0,1)$. The solution $g(\tau)$ for all $\tau \in [0,\infty)$ is shown as the blue curve in Fig. 6.11. We also show the moments $\tilde{\nu}_2(\tau)$ and $\tilde{\nu}_3(\tau)$, where

$$\tilde{\nu}_k(\tau) = \sum_{n=1}^{\infty} n^k\,\bar{c}_n(\tau) = \left(-\frac{\partial}{\partial u}\right)^{k-1} \bar{c}(u,\tau)\bigg|_{u=0} \quad . \tag{6.248}$$

From Eqn. 6.239 we have

$$\tau\bar{c} - \ln\bar{c} = u + \tau \tag{6.249}$$

and therefore

$$\tilde{\nu}_2(\tau) = -\frac{\partial\bar{c}}{\partial u}\bigg|_{u=0} = \frac{\bar{c}(0,\tau)}{1 - \bar{c}(0,\tau)\,\tau} = \begin{cases} (1-\tau)^{-1} & \text{if } \tau < 1 \\ (e^{g\tau} - \tau)^{-1} & \text{if } \tau > 1 \end{cases} \quad . \tag{6.250}$$

Similarly,

$$\tilde{\nu}_3(t) = \frac{\partial^2\bar{c}(u,t)}{\partial u^2}\bigg|_{u=0} = \frac{\tilde{\nu}_2^3(\tau)}{\bar{c}^2(0,\tau)} \quad . \tag{6.251}$$

The functions $g(\tau)$, $\tilde{\nu}_2(\tau)$, and $\tilde{\nu}_3(\tau)$ are plotted in Fig. 6.11.