Part I. Rudiments

This part provides a rudimentary background of probability and statistical mechanics. It is expected that the reader has already learned elementary probability theory and equilibrium statistical mechanics. Therefore, it should be boring to repeat the usual elementary courses for these topic.

Chapter 1 outlines the measure-theoretical probability theory as intuitively as possible: the probability of an event is a 'volume' of our confidence in the occurrence of the event. Then, the most important law: the law of large numbers and its refinements, large deviation and central limit theorems are discussed. The law of large numbers makes statistical thermodynamics possible, and large deviation theory gives the basis for most elementary nonequilibrium statistical mechanical approaches. Appendix outlines measure theory.

Chapter 2 is a review of the framework of equilibrium statistical mechanics along its history. It was born from kinetic theory. Elementary kinetic theory and the basic part of Boltzmann's kinetic theory are outlined. However, we are not interested in the theoretical framework applicable only to special cases (dilute gases in this case). Therefore, kinetic theories are not discussed at length. Based on the expression of work (or equivalently the microscopic expression of heat) due to Einstein, $A = -k_BT \log Z$ is derived. Then, Jarzynski's equality is discussed. Eventually, we will learn that we have not yet understood the nature of the kind of motion we call heat. This is perhaps a major reason why we cannot understand nonequilibrium phenomena very well.

Chapter 1

Rudiments of Probability and Statistics

Almost surely the reader has already learned elementary probability theory. The purpose of this section is not to introduce elementary probability theory. We introduce the measure-theoretical probability as intuitively as possible. The reader should at least briefly reflect on what probability must be. Then, the three pillars of probability theory, crucial to statistical physics, are outlined in simple terms. They are the law of large numbers and its two refinements, large deviation theory and central limit theorems. (Stochastic processes will be defined in **1.4.2** but will be discussed later).

Their relevance to statistical physics is roughly as follows: the law of large numbers allows us to observe the reproducible macroscopic world (macroscopic phenomenology). The large deviation theory tells us variational principles behind the macroscopic phenomenology and fluctuations around the macroscopic results. Central limit theorems allow us to extract the macroscopic picture of the world (= renormalization group theory).

Recommended reference of probability theory (for a second introduction):
R Durrett, *Probability: Theory and Examples* (Wadsworth & Brooks Cole, Pacific Grove, Ca, 1991)[1]

1.1 Probability, an elementary review

Starting with a reflection on what 'probability' should be, we outline very intuitively the basic idea of 'measure theoretical probability.' Appendix to this chapter is an introduction to measure theory.

The following book contains the most pedestrian introduction to measure-theoretical

---

[1] This is a role model of a textbook of any serious subject. However, it may not be the first book of probability the reader should read; it is a book for those who have firm basic knowledge.

probability theory I have ever encountered:
Z. Brzeźniak and T. Zastawniak, *Basic Stochastic Processes — a course through exercises* (Springer Verlag, 1999)

1.1.1 Probability as a confidence level.

Probability of an event is a measure of our confidence (or conviction) level in 0-1 scale for the event to happen.
Discussion 1. Here, the word 'confidence' is used intuitively, but what is it? Your confidence must be tested. Your confidence must be useful, if you follow it. This means that your confidence must enhance your reproductive success. □

Suppose there is a box containing white balls and red balls. The balls are all identical except for the colors, and we cannot see inside the box. What is our confidence level of picking a red ball from the box? Our confidence level C in 0-1 scale must be a monotone increasing function of the percentage $100p\%$ of red balls in the box. We expect on the average we pick a red ball once in $1/p$ picks (assuming that we return the ball after every pick and shake the box well; this is the conclusion of the law of large numbers →**1.5.3**). Therefore, $C = p$ is a choice. That is, if one can repeat trials, the relative frequency of a success is a good measure of confidence in the success.

Let us consider another example. What do we mean, when we say that there will be a 70% chance of rain tomorrow? In this case, in contrast to the preceding example, we cannot repeat tomorrow again and again. However, the message at least means that the weather person has this much of confidence in the rain tomorrow. We could imagine that she looked up all the data of similar situations in the past, or that she did some extensive hydrodynamic calculation and the error bar suggested this confidence level, etc. Thus, even in this case, essentially, an ensemble of days with similar conditions is conceivable, and the relative frequency gives us a measure of our confidence level.[2]

1.1.2 Probability is something like volume.

If two events never occur simultaneously, we say these two events are *mutually exclusive*.[3] For example, suppose a box contains white, red, and yellow balls. We assume that we cannot see the inside of the box and the balls are identical except for their colors. Picking a red ball and picking a yellow ball are two mutually exclusive events.

Suppose our confidence level of picking a red ball be C_R and that for a yellow ball be C_Y. Then, what is our confidence level of picking a red or a yellow ball? It is sensible to require that this confidence level is $C_R + C_Y$. This is just in accordance with the choice '$C = p$' (or the additivity justifies that this choice is the best).
Discussion 1. In the above it is said $C_R + C_Y$ is 'sensible.' Why or in what sense is it sensible? □

This strongly suggests that our confidence level in 0-1 scale = probability is something like a volume or weight of an object: if an object R has volume or weight W_R and another object Y weight W_Y, the weight of the nonoverlapping (= mutually exclusive) compound object made by joining R and Y must be $W_R + W_Y$.

Our confidence level for an event that surely happens is one; for example, we can surely pick a ball irrespective of its color in the above box example, so, if C_W is our confidence level

---

[2] In any case, the confidence level, if meaningful at all to our lives, must be empirically reasonable. Is there any objective way to check the reasonableness of a confidence level without any use of a kind of ensemble?

[3] Do not confuse this with the concept of (statistical) *independence* (→**1.2.2**). If two events are (statistically) independent, the occurrence of one does not tell us anything about the other event.

of picking a white ball, $C_R + C_Y + C_W = 1$. Thus, probability looks like volume or weight whose total volume (weight or mass) is unity.

Suppose a shape is drawn in a square of area 1. If we pepper points on the square uniformly, then the probability = our confidence level of a point to land on the shape should be proportional to its area.[4] Thus, again it is intuitively plausible that probability and area or volume are closely related.

Discussion 2. 'Uniformly' is a tricky word. The reader should have smelled a similar difficulty as encountered in the principle of equal probability in equilibrium statistical mechanics (→**2.6.2**). Uniform with respect to what? Even if we could answer this question, is the thing we base our uniformity really uniform (with respect to what?)? Critically think whether we can really have objective 'uniformity.'[5] (For example, in Euclidean geometry the uniformity of the space must be assumed axiomatically. How is it done? Even if the reader scrutinizes Euclid, she will never find any relevant statement. That is, as is of course well known, Euclid's axiomatic system is (grossly) incomplete.) □

1.1.3 What is probability theory?

A particular measure of the extent of confidence may be realistic or unrealistic (rational or irrational), but this is not the concern of probability theory. Probability theory studies the general properties of the confidence level of an event.

The most crucial observation is that the confidence level behaves (mathematically or logically) just as volume or weight as we have seen in **1.1.2**. Let us write our confidence level = probability of event A as $P(A)$. If two events A and B are mutually exclusive, $P(A$ or B occurs) $= P(A) + P(B)$ (→**1.1.5**(2)). Such basic properties were first clearly recognized as axioms of probability by Kolmogorov in 1933.[6,7]

1.1.4 Elementary and compound events.

An event which cannot (or need not) be analyzed further into a combination of events is called an *elementary event.* For example, to rain tomorrow can be regarded as an elementary event, but to rain or to snow tomorrow is a combination of two elementary events, and is called a *compound event.*

Let us denote by Ω the totality of elementary events allowed under a given situation or to a system.[8] Any event under consideration can be identified with a subset of Ω. Therefore, a subset of Ω is called an *event.* When we say an event A occurs, we mean that one of the elements in A occurs.

1.1.5 Definition of probability.

Let us denote a probability of $A \subset \Omega$ by $P(A)$. It is defined as a function whose domain is

---

[4] This has a practical consequence. See **1.5.6**.

[5] If the reader happens to know the so-called algorithmic probability theory, she may claim that there is an objective definition of uniform distribution, but there all the symbols are *assumed* to be equally probable.

[6] This year, Hitler was elected German Chancellor; Stalin began a massive purge; Roosvelt introduced New Deal.

[7] Kolmogorov's work is a direct response to Hilbert's address to the second International Congress of Mathematicians in Paris held in 1900 (see "Mathematical problems," Bull. AMS **8** 437 (1992)). He named probability as one of the subdisciplines of physics whose axiomatic foundations needed investigation. Chapter 2 of G. Shafer and V. Vovk, *Probability and Finance, it's only a game!* (Wiley-Interscience 2001) is a good summary of the history of probability.

[8] That Ω is treated as a set is implicit here.

the totality of the sets in Ω,[9] and whose range is $[0,1]$, satisfying the following conditions:
(1) Since probability should measure the degree of our confidence on the 0-1 scale, we demand that

$$P(\Omega) = 1; \tag{1.1.1}$$

in other words, something must happen. Then, it is also sensible to assume

$$P(\emptyset) = 0. \tag{1.1.2}$$

(2) Now, consider two mutually exclusive events, A and B. This means that whenever an elementary event in A occurs, no elementary event in B occurs and vice versa. That is, $A \cap B = \emptyset$. As we have already discussed, it is sensible to demand *additivity*

$$P(A \cup B) = P(A) + P(B), \text{ if } A \cap B = \emptyset. \tag{1.1.3}$$

It is convenient to require the additivity to infinitely many sets:
(2') [*σ-additivity*]: Suppose A_i $(i = 1, 2, \cdots)$ are mutually exclusive events ($A_i \cap A_j = \emptyset$ for any i and j $(\neq i)$). Then,

$$P(\cup_i A_i) = \sum_i P(A_i). \tag{1.1.4}$$

In mathematical terms, *probability* P is a σ-additive measure with total weight unity. In Appendix to this Chapter, *measure* is explained informally; any statistical physicist should know its rudiments.

(Ω, P) is called a *probability space*.[10]

Note that for a given totality of elementary events Ω, there are many different probabilities just as there are many different unfair coins. For a coin $\Omega = \{H, T\}$ (head and tail), and $P(H) = p_h$ and $P(T) = 1 - p_h$ define a probability P.

Exercise 1. There are 50 people in a room. What is the probability of finding at least two persons sharing the same birthday? □

Exercise 2. There is a unit circle on which we wish to draw a chord whose length is longer than $\sqrt{3}$. What is the probability of drawing such a chord by randomly drawing it?[11] [Notice that the word 'randomly' is as tricky as 'uniformly' (→**1.1.2**)]. □

1.1.6 On the interpretation of probability

Mathematicians and philosophers have been thinking about the interpretation of probability; as a branch of mathematics the measure theoretical framework may be enough, but to apply probability to the real world some interpretation is needed.

There are two major camps in this respect: the frequentists vs. the 'subjectivists'. The former asserts that probability is interpreted as relative frequency (out of an ensemble of events). The latter points out the difficulty in the interpretation of a unique event (an event that occurs only once, e.g., our lives), and asserts that some probabilities must be interpreted deductively rather than inductively as the frequentists advocate.[12]

---

[9] Such a function is called a set-theoretical function; volume (→**1.A.2**) is an example. Not all the sets have volume; not all the events have probability. Therefore, precisely speaking, we must deal only the events whose probability is meaningful. See **1.A.20**.

[10] This is not enough to specify a probability (→**1.A.20**), but we proceed naively here.

[11] There is a very good account of this problem on p21-2 of N G van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, 1981). This is a nice book, but should be scanned quickly after learning the standard probability theory.

[12] Propensity interpretation by Popper, interpretation in terms of degree of confirmation by Carnap, etc.

Our starting point in **1.1.1** was probability = measure of confidence, so the reader might have thought that the lecturer was on the subjectivist side. However, this subjective measure must agree at least to some extent with the actual experience in our lives. Thus, the relation with relative frequency was immediately discussed there. The lecturer's point of view is that probability is a measure of confidence inductively supported by relative frequency. 'Non-frequentists' forget that our mental faculty being used for deduction is a product of evolution; our confidence may not be tied to our own experience, but is tied to the result of our phylogenetic learning = the totality of experiences by our ancestors during the past 4 billion years.[13]

Laplace is the founder of probability theory.[14] His basic idea was: the relative number of equally probable elementary events in a given event is the probability of the event. Here, 'equally probable' implies that there is no reason to have difference. This is exactly the interpretation of probability used by Boltzmann when he discussed entropy (→**2.4.2**).

If we feel that there is no reason to have difference between event A and event B, then we feel A and B are equally likely. This feeling is often in accordance with our experience thanks to the evolution process that has molded us. That is, both our feeling and the 'logic' behind Laplace's and Boltzmann's fundamental ideas have the basis supported by phylogenetic learning.

Probability is always disguised relative frequency. Purely deductive probability concept is meaningless.

Discussion 1. Read the article by N. Williams, "Mendel's demon" in Current Biology **11**, R80-R81 (2001) and discuss the positive meaning of randomness.[15]

□

1.2 Conditional Probability and Bayesian Statistics

1.2.1 Conditional probability: elementary definition

Suppose we know for sure that an elementary event in B has occurred. Under this condition

---

[13] Wittgenstein thought deeply on related problems. Let us read some passages of *Philosophical Investigations* [There is a 50th anniversary commemorative edition: the German text with a revised English translation (by G. E. M. Anscombe), Third Edition; Blackwell, Oxford, 2001].

"472. The character of the belief in the uniformity of nature can perhaps be seen most clearly in the case in which we fear what we expect. Nothing could induce me to put my hand into a flame — although after all it is *only in the past* that I have burnt myself." (p114)

"481. If anyone said that information about the past could not convince him that something would happen in the future, I should not understand him. One might ask him: What do you expect to be told, then? What sort of information do you call a ground for such a belief? What do you call " conviction"? In what kind of way do you expect to be convinced? —If *these* are not grounds, then what are grounds? — $\cdots$." (p115)

"483. A good ground is one that looks *like this*." (p115)

[14] The *Laplace demon* that can predict everything deterministically was introduced to indicate its absurdity and to promote probabilistic ideas in the introduction to his *Essai philosophique sur les probabilitès* (1814, this year Napoleon abdicated; Fraunhofer lines; Berzelius introduced the modern chemical symbols now in use).

[15] This article is an introduction to the following book: M. Ridley, *The Cooperative Gene— How Mendel's Demon explains the evolution of complex beings* (The Free Press, 2001), "Mendel's demon is the executive of gene justice, and we all depend on it for our existence." Mendel's demon is five years elder to Maxwell's demon.

what is the probability of the occurrence of another event A? This question leads us to the concept of *conditional probability*. We write this conditional probability as $P(A|B)$, and define it through

$$P(A \cap B) = P(A|B)P(B). \tag{1.2.1}$$

1.2.2 Independence

If two events A and B satisfy

$$P(A \cap B) = P(A)P(B), \tag{1.2.2}$$

these two events are said to be (statistically) *independent*.

This definition is intuitively consistent with the definition of the conditional probability in **1.2.1**. If B has no information about A, then $P(A|B) = P(A)$ is plausible. Thus, (1.2.1) reduces to (1.2.2).

Logically, we must define the independence of countably many events such as $P(A \cap B \cap C) = P(A)P(B)P(C)$, because pairwise independence (1.2.2) does not imply such a relation among many events as the reader sees in Exercise 1.

Exercise 1.[16] Let X_1, X_2 and X_3 be independent random variables with $P(X_i = 0) = P(X_i = 1) = 1/2$. Consider the following three events: $A_1 = \{X_2 = X_3\}$, $A_2 = \{X_3 = X_1\}$ and $A_3 = \{X_1 = X_2\}$.
(1) Show that these events are pairwisely independent.
(2) $P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3)$.□

Kolmogorov proved:
There are infinitely many independent random variables on an appropriate probability space.

1.2.3 Borel-Cantelli lemma

Let $A_1, A_2, \cdots$ be an infinite sequence of events.
(1) If $\sum_{i=1}^{\infty} P(A_n) < \infty$, then with probability 1 only finitely many events among $\{A_i\}$ occur.
(2) If all the events A_k are independent and if $\sum_{i=1}^{\infty} P(A_n) = \infty$, then with probability 1 infinitely many of the events in $\{A_i\}$ occur.

[Demo] Here, a detailed (pedestrian) demonstration is given, because its logic is quite typical.
(1) To demonstrate probabilistic statements, explicitly writing down the event or its complement (negation) is often the best starting point. The negation of 'only finitely many' means 'infinitely many.' If there occur infinitely many of $\{A_k\}$, for any n some A_j $(j \geq n)$ must occur. That is, $B_n \equiv \cup_{j \geq n} A_j$ must occur for any n. That is, all B_n must occur, or $\cap_{n \geq 1} B_n$ must occur. Therefore, to demonstrate (1) we must show

$$P\left(\cap_{n \geq 1} \left(\cup_{j \geq n} A_j\right)\right) = 0. \tag{1.2.3}$$

This can be shown as follows:

$$P\left(\cap_{n \geq 1} \left(\cup_{j \geq n} A_j\right)\right) \leq \lim_{n \to \infty} P\left(\cup_{j \geq n} A_j\right) \leq \lim_{n \to \infty} \sum_{j \geq n} P\left(A_j\right) = 0, \tag{1.2.4}$$

because the series $\sum_{i=1}^{\infty} P(A_n)$ converges.
(2) Again, the negation is easier to prove: the probability of the occurrence of only finitely

---

[16] Taken from Durrett Example 4.1.

many of $\{A_k\}$ is zero. If only finitely many of them occur, then all of $\Omega\backslash A_k$ must occur for $k \geq n$ for some n. That is, $\cap_{k\geq n}(\Omega\backslash A_k)$ must occur for some n. That is, $\cup_{n\geq 1}\left(\cap_{k\geq n}(\Omega \setminus A_k)\right)$ must occur. However, this is the negation of the statement we wish to demonstrate. Therefore, we wish to show that $\cup_{n\geq 1}\left(\cap_{k\geq n}(\Omega \setminus A_k)\right)$ never occurs: we must show

$$P\left(\cup_{n\geq 1}\left(\cap_{k\geq n}(\Omega \setminus A_k)\right)\right) = 0. \tag{1.2.5}$$

This can be demonstrated as follows:

$$P\left(\cup_{n\geq 1}\left(\cap_{k\geq n}(\Omega \setminus A_k)\right)\right) \leq \lim_{n\to\infty} P\left(\cap_{k\geq n}(\Omega \setminus A_k)\right) = \lim_{n\to\infty} \prod_{k\geq n} P(\Omega \setminus A_k). \tag{1.2.6}$$

The last equality is due to the independence of all $\Omega \setminus A_k$. Since $\sum_{k\geq n} P(A_k)$ diverges for any n, $\prod_{k\geq n}(1 - P(A_k)) = 0$ for any n.[17] This implies the product in (1.2.6) vanishes.

1.2.4 Conditional probability: mathematical comment

(1.2.1) has no difficulty at all so long as we study a discrete probability space. However, there are many events that can change continuously. Suppose the event B is the event that a molecule in a cube of edge 1m has the speed infinitessimally close to 1m/s (event B). The reader might say such an event is meaningless because it is 'measure zero' or probability zero. However, it is conceivable that the particle has such a speed and in a ball of radius 10cm in the cube (event A). The conditional probability $P(A|B)$ should be intuitively meaningful.

According to (1.2.1), it is 0/0, so its definition requires some care; since the ratio $P(A\cap B)/P(B)$ is 0/0, conceptually it is close to a derivative. Obviously, $P(A\cap B) < P(B)$, so this derivative is always well-defined. Such a derivative is called the *Radon-Nikodym derivative*. We will discuss this when we really need such sophistication.[18]

1.2.5 Bayes' formula

Suppose we know the conditional probability $\{P(x|\theta)\}$ of X with the condition that Θ takes the value θ. Then,

$$P(x, \theta) = P(x|\theta)P(\theta) = P(\theta|x)P(x). \tag{1.2.7}$$

Therefore, we have

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{\sum_\theta P(x|\theta)P(\theta)}. \tag{1.2.8}$$

This is called *Bayes' formula*.

This formula is used, for example, as follows. Θ is a set of possible parameter values that describe (or govern) the samples X we actually observe. Thus, $P(x|\theta)$ is the probability we observe x, if the parameter value is θ. However, we do not know the actual θ. From the actual empirical result we can estimate $P(x|\theta)$, assuming a model of the observation. Our aim is to estimate θ that is actually behind what we observe; you could interpret this

---

[17] A clever way to show this is to use

$$e^{-x} \geq 1 - x.$$

This implies

$$e^{-\sum x_k} \geq \prod (1 - x_k).$$

[18] If you are impatient, A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis* (Revised English edition, Englewood Cliffs,1970) is recommended.

as a problem selecting a model describing the observed phenomenon. A sensible idea is the maximum *a posteriori* probability estimation (*MAP estimation*): Choose θ that maximize $P(\theta|x)$.

However, to use (1.2.8) we need $P(\theta)$: the *a priori* probability. The usual logic is that if we do not know anything about Θ, $P(\theta)$ is uniform. Of course, this is a typical abuse of Occam's razor.[19]

1.2.6 Linear regression from Bayesian point of view

Let us assume that the independent variable values $\{t_k\}$ (($k = 1, \cdots, N$) are obtained without any error. For t_k the observed dependent variable is y_k but this is not without error:

$$y_k = at_k + b + \epsilon_k. \tag{1.2.9}$$

The problem is to estimate a and b. In the resent problem ϵ_k are assumed to be iid[20] Gaussian variables with mean zero and variance V. Then, under the given parameters $\{a, b\}$

$$P(\{y_k\}|\{a,b\}) = \frac{1}{\sqrt{2\pi V}^N} \exp\left(-\frac{1}{2V}\sum(y_k - at_k - b)^2\right). \tag{1.2.10}$$

To use (1.2.8), we need $P(\{a, b\})$ the prior probability. If we assume that the prior is uniform, then the *a posteriori* probability is

$$P(\{a,b\}|\{y_k\}) \propto \frac{1}{\sqrt{2\pi V}^N} \exp\left(-\frac{1}{2V}\sum(y_k - at_k - b)^2\right). \tag{1.2.11}$$

Therefore, the MAP estimate dictates the least square scheme of determining a and b.

1.2.7 Problem of *a priori* probability

Bayesian statistics requires *a priori* probability. Sometimes, as have seen in **1.2.6**, we could disregard it but the reason for ignoring the a priori probability is not very clear.

The reader must have thought about the principle of equal probability. This is the *a priori* distribution. Notice that the principle is made of two parts:
(1) Designation of microstates (or elementary events).
(2) Probability assignment to each elementary event.

In a certain sense, (1) is not a serious problem thanks to quantum mechanics. Since very high energy excitations cannot occur, we may totally ignore the structure of the world at extremely small scales. Thus, Nature dictates what we may ignore.

However, (2) does not disappear even with quantum mechanics as already discussed. Thus, the problem of *a priori* distribution is a real issue.

If we wish to establish a general statistical framework for nonequilibrium states, (1) may well be the same as in equilibrium statistics. Thus, (2) becomes the key question of nonequilibrium statistical mechanics.

---

[19] "Occam's razor, in its legitimate application, operates as a logical principle about the complexity of argument, not as an empirical claim that nature must be maximally simple. ⋯ Whereas Occam's razor holds that we should not impose complexities upon nature from non-empirical sources of human argument, the factual phenomena of nature need not be maximally simple — and the Razor does not address this completely different issue at all." (S. J. Gould, *The Structure of Evolution Theory* (Harvard UP, 2002) p552). That is, questions about actual phenomena (e.g., whether two events are equally probable or not) should not invoke the razor. However, it seems empirically true that the abuse of the razor often works. The most natural explanation is that our reasoning utilizes our brain, a product of natural selection.

[20] A standard abbreviation for "independently and identically distributed."

1.3 Information

This section is not included in the 'minimum' to understand rudiments of nonequilibrium statistical mechanics. Therefore, the reader could skip this section. However, it is a good occasion to have some familiarity to the topic. An excellent general introduction to the topic is:
T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, 1991)

1.3.1 Probability and information

Suppose there are two sets of symbols L_n and L_m that contain n and m symbols, respectively ($n > m$). Let us assume that all the symbols appear equally probably. Now, we just received two messages with 20 symbols. Which do we expect to have more information, the message in symbols in L_m or that in L_n?

Here, the reader must 'naively' interpret the word 'information.' For example, compare the English alphabet and the Chinese characters. We would certainly say 20 Chinese characters carry more information; with only 20 Chinese characters the world greatest poems were written. Therefore, we expect the message written in the symbols in L_n should have more information than that in L_m. Let us introduce (the average) *information* per symbol $H_n = H(L_n)$[21] in the message written in the symbols in L_n as an increasing function of n.

What we are trying to do now is this: it is very hard to characterize 'information,' so let us at least try to quantify it, just as we do for energy. If the probability of encountering a particular symbol (for L_n this is $p = 1/n$ for any symbol) is small, the information carried by the symbol is large. This is the fundamental relation between information and probability.

Now, let us use two sets of symbols L_n and L_m to send a message. We could make compound symbols by juxtaposing them as ab for $a \in L_n$ and $b \in L_m$ (just as many Chinese characters do). The information carried by each compound symbol should be $I(mn)$, because there are mn symbols. To send the message written by these compound symbols, we could send all the left half symbols first and then the right half symbols later. The amount of information sent by these methods must be equal, so we must conclude that

$$H_{mn} = H_m + H_n. \tag{1.3.1}$$

Here, H_n is defined on positive integers n, but we can extend its definition to all positive real numbers as

$$H_n = c \log n, \tag{1.3.2}$$

where $c > 0$ is a constant. Its choice is equivalent to the choice of unit of information and corresponds to the choice of the base of the logarithm in the formula.

If $c = 1$ we say we measure information in *nat*; if we choose $c = 1/\log 2$, in *bit*. One bit is an amount of information obtained by the answer to a single yes-no question.
Remark. However, we do not discuss the 'value' of information. □

---

[21] In the notation $H(Q)$, Q is a random variable, or a probability space.

1.3.2 Shannon's information formula

Although we have assumed above that all the symbols are used evenly, such uniformity does not occur usually. What is the most sensible generalization of (1.3.2)?

A hint is already given in **1.3.1**. If we write $H_n = -\log_2(1/n)$ bits, $-\log_2(1/n)$ is the expectation value[22] of information carried by a single symbol chosen from L_n. $1/n$ is the occurrence probability of a particular symbol. Then, for the case with nonuniform occurrence $\{p_1, \cdots, p_n\}$, the expectation value of the information carried by the i-th symbol should be defined as $-\log_2 p_i$ bits. Therefore, the average information in bits carried by a single symbol taken from L_n should be defined by[23]

$$H(L_n) = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{1.3.3}$$

This is called the *Shannon information formula*.[24]

$-\log_2 p_i$ is sometimes called the *surprisal* of symbol i, because it measures how much we are surprised by encountering this symbol. Expected surprisal is the information.

Exercise 1.

(1) Given a fair die, we are told that the face with 1 is up. What is the information in bit carried by this message? [The information of a message is measured by the change of information (decrease of entropy, because often gain of information is equated with decrease of entropy) one has due to the knowledge in the message.] (2.58 bits)

(2) We are told that the sum of the faces of two fair dice is 7. What is the information carried by this message? ($2 \times 2.58 - 2.58 = 2.58$ bits). □

Exercise 2. Suppose we have a sentence consisting of 100 letters. We know that the information per letter is 0.7 bit. How many ye-no question s do you need to completely determine the sentence? □

Discussion 1. If the symbols of English alphabet (+ blank) appear equally probably, what is the information carried by a single symbol? (4.58 bits) In actual English sentences, it is about 1.5(?) bits. What happens? Also think about how to determine this actual value.[25,26] □

1.3.3 Information for continuous distribution

Let f be the density distribution function on an event space (probability space) that is continuous. It is customary to define the Shannon entropy of this distribution as

$$H = -\int dx\, f(x) \log f(x). \tag{1.3.4}$$

This formula was originally introduced by Gibbs as we see in **2.1.5**.

Warning. Notice that the formula has a very serious defect. The formula is not invariant under the coordinate transformation. Further worse, it is not invariant even under the

---

[22] Here, 'expectation value' may be understood intuitively and in an elementary way, but for a more mathematical definition see **1.4.3**.

[23] Notice that the symbol set is actually a probability space $(L_n, \{p_j\})$.

[24] for an uncorrelated (or Bernoulli) information source. About Shannon himself, see S. W. Golomb et al., "Claude Elwood Shannon (1916-2002)," Notices AMS **49**, 8 (2002), and J. F. Crow, "Shannon's brief foray into genetics," Genetics **159**, 915 (2001).

[25] Cf. Cover and Thomas, Chapter 6 "Gambling and Data Compression."

[26] There is a related 'fun' paper, Benedetto et al., "Language tree and zipping," Phys. Rev. Lett. **88**, 048702-1 (2002). This uses Lempel-Ziv algorithm LZ77 (used in gzip, for example) to measure information (or rather mutual information).

change of the unit to measure the coordinates, because the scaling transformation of the independent variable $x \to \alpha x$ alters the value of H. Generally speaking, no dimensionful quantity should appear inside the logarithmic function. □

The reader must have realized that f inside the logarithm should be a ratio. See **1.6.9**. In short, (1.3.4) may be used only under the assumption that x is *a priori* equally weighted.

1.3.4 Information and entropy

$H(A)$ is the expected information we obtain from one sample from A. The information she can gain on the average from the answer that the symbol is a particular one is given by $H(A)$. Thus, as we have already seen in Exercise 2 in **1.3.2**, if we obtain $nH(A)$ information, we can determine a sentence s_n made of n symbols taken from A.[27]

Suppose we already know the sentence s_n. Then, there is no surprise at all, even if we are shown s_n: the information carried by s_n is zero. This means that the information $nH(A)$ has reduced our ignorance from the initial high level to the final perfectly informed level. A convenient measure of the extent of ignorance is entropy in equilibrium statistical mechanics. Therefore, we could say that $nH(A)$ of information has reduced the entropy to zero. It is sensible to guess that the initial entropy is $nH(A)$. This is the Gibbs-Shannon entropy-information relation seen in (1.3.3) or (1.3.4).

Thus, information and entropy are often used interchangeably. If there is a macroscopic state whose entropy is S, to specify a single microstate (= elementary event) we need (on the average) information $H = S$.

Exercise 1. When reaction occurs, the entropy of the reacting chemicals change. This is usually measured in en (*entropy unit*) = cal/mol·K. How many bits per molecule does 1 en correspond? Is the reader's answer reasonable?[28] How about in pNnm/K?[29] □

Discussion 1. How much energy is required to transfer 1 bit between two communicators?[30] □

1.3.5 Information minimization principle

Let us consider the problem of estimating the 'most likely' distribution (density distribution

---

[27] On the average. However, if n is very large, the accuracy (= density of correctly identified letters) asymptotically reaches unity.

[28] If the entropy is measured in eu as S eu, then it is $0.72S$ bits/molecule.

[29]

$$\begin{aligned} k_B &= 1.381 \times 10^{-23} \text{ J/K}, \\ &= 1.381 \times 10^{-2} \text{ pN} \cdot \text{nm/K}, \\ &= 8.617 \times 10^{-5} \text{ eV/K} \end{aligned}$$

The *gas constant* R is given by

$$R \equiv N_A k_B = 8.314 \text{ J/mol} \cdot \text{K} = 1.986 \text{ cal/mol} \cdot \text{K} \quad (1.3.5)$$

Here, 1 cal = 4.18605 J. It is convenient to remember that at room temperature (300K)

$$\begin{aligned} k_B T &= 4.14 \text{ pN} \cdot \text{nm} \\ &= 0.026 \text{ eV} \\ RT &= 0.6 \text{ kcal/mol}. \end{aligned}$$

[30] R Landauer, "Energy requirements in communication," Appl. Phys. Lett. **51**, 2056-2059 (1987).

function) of a variable describing a system under some known conditions or knowledges, say, the average and the variance.

To answer this question we must interpret the requirement 'most probable' appropriately. One interpretation is that the most probable distribution is the one with the least biased under the extra conditions or knowledges we are at hand. A quantitative version of this interpretation uses the Shannon information as follows.

Compare the *a priori* distribution and the distribution we can infer with the extra knowledge. The most unbiased distribution we should infer must correspond to the distribution of the symbols with the least decrease of information per symbol compatible with the extra condition. Therefore, the distribution we wish to infer must be obtained by the maximization of the Shannon formula (1.3.4) or (1.3.3) under the extra conditions imposed. The decrease in H may be interpreted as the information gained from the extra conditions. Therefore, the prescription is to minimize the information gain from the newly acquired conditions. This is the so-called Jaynes' principle. The principle is useful, if we do not stretch it, or if we use it as a shortcut to known results.

The astute reader should have already realized that the principle works only when we know that the *a priori* distribution is uniform (= the principle of equal probability). Furthermore, from the warning in **1.3.3**, if the event is described by continuous variables, a further caution is required to use this principle. A typical abuse is the 'derivation of statistical mechanics' **1.3.6**.

Discussion 1. The distribution available by the information minimization principle and the true distribution are very likely different under nonequilibrium conditions.[31] □

1.3.6 Information-theoretical foundation of statistical mechanics is illusory.

The information H is maximized when all the symbols (or events) occur equally probably. Some people wish to utilize this property to justify (or derive) the principle of equal probability (→**2.6.2**) (the so-called information theoretical foundation of statistical mechanics[32]).

Let us maximize (1.3.3) under the condition that we know the expectation value of energy $\sum E_i p_i = E$.

Exercise 1. Using Lagrange's multipliers, find p_i in terms of E_i. □

As the reader has seen, the canonical distribution

$$p_i \propto e^{-\beta E_i} \tag{1.3.6}$$

is obtained, where β is determined to give the correct expectation value of the energy.[33] The logic is that H measures the amount of the lack of our knowledge, so this is the most unbiased distribution (→**1.3.5**). □

But does Nature care for what we know? Even if She does, why do we have to choose Shannon's information formula to measure the amount of our knowledge?

The formula (1.3.3) is so chosen that if all the symbols appear equally probably, then it depends only on the total number of elements (→**1.3.1**). Therefore, logically, there is a vicious circle. The reader will encounter a similar quantity, the Kullback-Leibler entropy later (→**1.6.9**), and will clearly recognize the fundamental logical defect in the so-called

---

[31] For example, see a recent paper: H.-D. Kim and H. Hayakawa, "Test of Information Theory on the Boltzmann Equation," J. Phys. Soc. Japan, 72 2473 (2003).

[32] E T Jaynes, Phys. Rev. **106**, 620 (1959).

[33] Boltzmann's argument summarized in Exercise of **2.4.11** just derives Shannon's formula and uses it. A major lesson is that before we use the Shannon formula important physics is over. The reader can use the Shannon formula as a shortcut, but can never use it to justify what she does.

information theoretical foundation of statistical mechanics.

Exercise 2. Is (1.3.6) really correct? [Suppose we do not know what the elementary event is in statistical mechanics and (erroneously) assume that elementary states are defined by their energy (that is, however many microstates we have for a given energy, they are assumed to be a single state).] □

Exercise 3. The use of the least square fit assumes that the error obeys a Gaussian distribution. Suppose we wish to fit the data $g(s)$ ($s \in [0,1]$) with a function $f(s,\alpha)$ with a parameter, choosing the parameter through minimizing the "L_2-error":

$$\int_0^1 ds |g(s) - f(s,\alpha)|^2. \tag{1.3.7}$$

We could also use an appropriate function M and minimize

$$\int_0^1 ds |M(g(s)) - M(f(s,\alpha))|^2. \tag{1.3.8}$$

Show that the answers cannot generally be identical. □

The reader must clearly recognize that information theory can never be used to justify or derive any fundamental statistical hypothesis.

1.3.7 Joint entropy and conditional entropy

Let $X = \{x_1, \cdots, x_n\}$ and $Y = \{y_1, \cdots, y_m\}$ be the letter sets. If $p(x,y)$ is the probability for a pair of letters $\{x,y\}$ ($x \in X$, $y \in Y$).

$$H(X,Y) = -\sum_{x\in X, y\in Y} p(x,y)\log p(x,y) \tag{1.3.9}$$

is the *joint entropy* of X and Y.

Consider $H(X,Y) - H(X)$. This must be the information (per letter pair) we can gain from $\{x,y\}$, when we already know x, Therefore, this is called the *conditional entropy* of Y. Thus, the following is a reasonable definition of the conditional entropy:

$$H(Y|X) = H(X,Y) - H(X) \tag{1.3.10}$$

Notice that

$$\begin{aligned} H(Y|X) &= -\sum_{x,y} p(x,y)\log p(x,y) + \sum_x p(x)\log p(x), && (1.3.11)\\ &= -\sum_{x,y} p(x,y)\log p(x,y) + \sum_{x,y} p(x,y)\log p(x), && (1.3.12)\\ &= -\sum_{x,y} p(x,y)\log p(y|x) = \sum_x p(x)\left[-\sum_{x,y} p(y|x)\log p(y|x)\right]. && (1.3.13) \end{aligned}$$

Here, $p(y|x)$ is the conditional probability of y under condition x. If we introduce

$$H(Y|x) = -\sum_y p(y|x)\log p(y|x), \tag{1.3.14}$$

we can write

$$H(Y|X) = \sum_x p(x) H(Y|x). \tag{1.3.15}$$

We can write

$$H(X,Y) = H(X) + H(Y|X). \tag{1.3.16}$$

This is sometimes called the *chain rule*. For example,

$$H(X,Y,Z) = H(Y,Z) + H(X|Y,Z) = H(Z) + H(Y|Z) + H(X|Y,Z). \tag{1.3.17}$$

It should be intuitively obvious that

$$H(X|Y) \le H(X). \tag{1.3.18}$$

This can be shown algebraically with the aid of the *log sum inequality*.

1.3.8 Log sum inequality

For non-negative numbers $a_1, \cdots, a_n$ and $b_1, \cdots, b_n$

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \ge \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{1.3.19}$$

with equality iff a_i/b_i is constant.
[Demo] Let $f(t) = t \log t$. The Jensen[34] implies for a stochastic vector $\boldsymbol{\alpha}$

$$\sum \alpha_i f(t_i) \ge f\left(\sum \alpha_i t_i \right). \tag{1.3.20}$$

Now, set $t_i = a_i/b_i$ and $\alpha_i = b_i / \sum b_i$. □

1.3.9 Mutual information

The *mutual information* $I(X,Y)$ between X and Y is defined as

$$I(X,Y) = H(X) + H(Y) - H(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{1.3.21}$$

Obviously, this is symmetric. Intuitively, this must be positive, because non-independence of X and Y should reduce the information that could have been carried by independent X and Y.
Exercise 1. Demonstrate that $I(X,Y) \ge 0$. □

1.3.10 Asymptotic equipartition property

Let $\{X_i\}$ be iid variables with probability $p(x)$. Then, in probability

$$-\frac{1}{n} \log P(X_1, \cdots, X_n) \to H(X). \tag{1.3.22}$$

This is just the weak law of large numbers (→**1.5.4**).

---

[34] If f is a convex function, $\langle f(x) \rangle \ge f(\langle x \rangle)$ (*Jensen's inequality*).

1.3.11 Typical set of sequences

The *typical set* $A_\epsilon^{(n)}$ wrt $p(x)$ is the set of $\{x_i\}_{i=1}^n$ such that

$$2^{-nH-n\epsilon} \le p(\{x_i\}) \le 2^{-nH+n\epsilon} \tag{1.3.23}$$

for small positive ϵ.
(1) If $\{x_i\} \in A_\epsilon^{(n)}$, then $H - \epsilon \le -(1/n)\log p(\{x_i\}) \le H + \epsilon$.
(2) For sufficiently large n $P(A_\epsilon^{(n)}) \ge 1 - \epsilon$. That is, the typical set is almost probability 1.
(3)

$$(1-\epsilon)2^{nH-\epsilon} \le |A_\epsilon^{(n)}| \le 2^{nH+\epsilon}. \tag{1.3.24}$$

Here | | denotes the cardinality of the set. The number of typical elements is 2^{nH} and they are all equally probable.
[Demo] (1) is the definition itself. For any $\delta > 0$, there is an N such that $\forall n > N$

$$P\left(\left|-\frac{1}{n}\log p(x_i) - H(X)\right| < \epsilon\right) > 1 - \delta. \tag{1.3.25}$$

This is asymptotic equipartition or weak law of large numbers. We may set $\delta = \epsilon$. This is (2). Since

$$1 = \sum p(\{x_i\}) \le \sum_{q \in A_\epsilon^{(n)}} p(q) \le \sum_{q \in A_\epsilon^{(n)}} 2^{-nH-n\epsilon} = |A_\epsilon^{(n)}|2^{-nH-n\epsilon}, \tag{1.3.26}$$

we obtain the upper bound in (3). From (2)

$$1 - \epsilon < P(A_\epsilon^{(n)}) \le \sum_{q \in A_\epsilon^{(n)}} 2^{-nH+n\epsilon} = |A_\epsilon^{(n)}|2^{-nH+n\epsilon} \tag{1.3.27}$$

we obtain the lower bound.□

1.3.12 Implication of typical set

1.3.11 implies (roughly):
If we produce uncorrelated sequences[35] of n ($\gg 1$ in practice) letters with entropy H, then only 2^{nH} of them appear. This typical set is usually vastly smaller than the totality of the possible sentences and can be written in nH 0 and 1's.

This is the principle of data compression,[36] and is fully used in, e.g., jpeg. Of course, occasionally sentences not in the typical set show up. In such cases we simply accept the situation as an unfortunate case and send or store such exceptional sequences as they are without compression. In any case, they are asymptotically very rare.

1.3.13 Statistical mechanics and law of large numbers revisited

The fourth law of thermodynamics implies that a macroscopic system may be understood as a collection of statistically independent macroscopic subsystems. The states of each subsystem may be characterized by the set of states of subsystems $\{x_i\}$. We know with almost probability one what we observe is from the asymptotic equiprobability states. Therefore, almost all the states of a macroscopic system have the same probability. This is an explanation of the principle of equal probability.

Thus, essentially the law of large numbers and the fourth law of thermodynamics are expected to be the two key elements of statistical mechanics.

Then, what is the role of mechanics?

---

[35] If there are correlations, we could exploit it to convert the original sentences shorter.

[36] A good exposition (not introductory) can be found in D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data Compression and Harmonic Analysis," IEEE Transactions on Information Theory, 44, 2435 (1998).

1.4 Random Variables

Basic tools such as characteristic functions, generating functions, etc., are introduced. The reader must be familiar with an elementary technique to compute probability densities using the δ-function (→**1.4.5**).

1.4.1 Random variables

Let P be a probability (= probability measure →**1.1.5**) defined on a set Ω = totality of elementary events under consideration.[37] A function defined on Ω is called a *random variable*.[38] A point in Ω (i.e., an elementary event) is often specified by a parameter ω called the *stochastic parameter*. Thus, a random variable can be written as $f(\omega)$.

Two random variables X and Y are said to be *independent*, if for any two real value sets A and B[39]

$$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\}). \tag{1.4.1}$$

1.4.2 Stochastic parameter and stochastic process

Roughly speaking, a *stochastic process* is a function of time parameterized by a stochastic parameter. That is, a stochastic process is a functional on a probability space: a map from a probability space to a space of functions of time. Analogously, we can define a stochastic field as a map from a probability space to a function defined on a space. ω specifies each sample = a particular realization.

1.4.3 Expectation value

Let $X(w)$ be a random variable on Ω and μ a probability on it [40]. The expectation value $E[f]$ of the random variable f is defined as

$$E(X) = \int_\Omega X(\omega) d\mu(\omega). \tag{1.4.2}$$

For a set (= event) $A \subset \Omega$, the following function χ_A is called its *indicator* (or indicator function):

$$\chi_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A. \end{cases} \tag{1.4.3}$$

That is, the indicator for an event A is 1, if the event occurs, and 0, otherwise.

The probability of an event $A \subset \Omega$ is $\mu(A)$, so

$$\mu(A) = \int_\Omega \chi_A(\omega) d\mu(\omega) = E(\chi_A). \tag{1.4.4}$$

---

[37] As seen in Appendix to this chapter, not all the events have their probabilities (measurable), so we must also specify a family of measurable events = the events whose probability we can discuss, but we will not explicitly write the so-called 'triplet' (→**1.A.20**).

[38] Again, we must discuss functions whose expectation values are meaningful, so we must require that random variables must be P-measurable (→**2.8.6**). However, we will not explicitly write such a condition.

[39] $A, B \subset \boldsymbol{R}$, but again not all the sets are measurable (have lengths), so we need some disclaimer here; we omit it as usual.

[40] We say $X(\omega)$ is a random variable on the probability space (Ω, μ).

That is, the expectation value of the indicator of an event is the probability of the event.

If Ω is a discrete set, then the above integral reduces to an appropriate summation.

$E((X - E(X))^n)$, if exists, is called the *n-th moment* (around the mean). The second moment around the mean is called the *variance* and is often denoted as

$$V(X) \equiv E((X - E(X))^2) = E(X^2) - E(X)^2. \tag{1.4.5}$$

A necessary and sufficient condition for two square-integrable random variables (i.e., random variables having the second moments) X and Y to be independent is: for any bounded functions[41] f and g

$$E(f(X)g(Y)) = E(f(X))E(g(Y)). \tag{1.4.6}$$

That is, we can average independent variables separately. □

Exercise 1. If X and Y have second moments, and are independent,

$$V(X + Y) = V(X) + V(Y). \tag{1.4.7}$$

□

1.4.4 Distribution functions

Let $X(\omega)$ be a real valued random variable.[42]

$$F(x) = P(\{X(\omega) \le x\}) \tag{1.4.8}$$

is called the *probability distribution function* for X. If we can write

$$f(x)dx = F(x + dx) - F(x), \tag{1.4.9}$$

f is called the *probability density distribution function.*[43] Thus, $f(x)dx = P(\{X(\omega) \in (x, x + dx]\})$.

If a function F on $\boldsymbol{R}$ satisfies:
(i) F is non-decreasing,
(ii) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$,
(iii) F is right continuous, i.e., $\lim_{x\to a+0} F(x) = F(a)$,
F is a *distribution function* of a random variable.

Exercise 1. The Cauchy distribution whose density is proportional to $1/(x^2 + a^2)$ does not have the second moment.□

Exercise 2. "If n-th moment does not exit, $m(> n)$-th moments do not exist, either." Is this statement correct? □

1.4.5 How to compute density

We have

$$P(A) = E(\chi_A), \tag{1.4.10}$$

---

[41] Again, we omit the measurability condition.

[42] Often the probability space is implicit.

[43] The Radon-Nikodym derivative of F with respect to the Lebesgue measure is called, if exists, the probability density function.

where χ_A is the indicator of A. If $A \in \boldsymbol{R}^n$, we can write

$$\chi_A(x) = \int_A \delta(x-y)\, d^n y, \tag{1.4.11}$$

where δ is the (n-dimensional) δ-function.

For physicists the density distribution is often more familiar with than distributions: the density distribution function $f(x)$ of a random variable X is given by

$$f(x) = E(\delta(X(\omega) - x)). \tag{1.4.12}$$

There are two major methods to compute this expectation. One is to use Fourier transformation which is equivalent to computing characteristic functions (→**1.4.7**). The other is to compute the integral directly. A typical problem is: Obtain the density distribution $\psi(y)$ of $y = g(x)$, given the density distribution of $\rho(x)$ of x. This is to compute

$$\psi(y) = E(\delta(g(x) - y)) = \int dx\, \rho(x)\, \delta(g(x) - y) = \sum_{x_i} \frac{\rho(x_i)}{|g'(x_i)|}, \tag{1.4.13}$$

where $g(x_i) = y$ (the sum is over all real zeros of $g - y$) (→**1.4.6**).

An example follows:

Suppose x is uniformly distributed on $[-1/2, 1/2]$. Find the density distribution function $\psi(y)$ of $y = x^2$. We obtain for $y \in (0, 1/4]$

$$\psi(y) = \int_{-1/2}^{1/2} dx\, \delta(y - x^2) = \frac{1}{\sqrt{y}}. \tag{1.4.14}$$

Exercise 1. Derive the energy density distribution function of a d-dimensional ideal gas in equilibrium. □

Exercise 2. Suppose x is distributed uniformly on $[0, 1]$. Let $F(x)$ be a monotone increasing smooth function from $\boldsymbol{R}$ to $[0, 1]$. Find the density distribution ϕ of $F^{-1}(x)$ (the inverse function of F). □

The result of this exercise is a standard method to convert uniform random number on $[0, 1]$ into the random number whose distribution function is given by F.

1.4.6 Computation of δ-function of complicated variable

To compute $E(\delta(g(x) - y))$, the basic formula we use is

$$\delta(g(x) - y) = \sum_{x_i} \frac{1}{|g'(x_i)|} \delta(x - x_i), \tag{1.4.15}$$

where $g(x_i) = y$ (the sum is over all real zeros of $g - y$).

Let us demonstrate (1.4.15). The delta function does not contribute anything to the integral, if its variable is not zero, so we have only to pay attention to the values of x that give zero of $g(x) - y$. Let x_1 satisfy $g(x_1) = y$. Then,

$$g(x) - y = g'(x_1)(x - x_1) + \cdots. \tag{1.4.16}$$

We must recall that $\delta(ax) = (1/|a|)\delta(x)$. Therefore, the contribution from x near x_1 must be $(1/|g'(x_1)|)\delta(x - x_1)$. There are other zeros as well. We should sum up all the contributions from the zeros of $g - y$. This outcome is (1.4.15).

1.4.7 Characteristic Function

The Fourier transform of a probability measure μ is called its *characteristic function*:

$$\varphi_\mu(\xi) \equiv \int e^{i\xi x}\mu(dx). \tag{1.4.17}$$

For a random variable X, the characteristic function of its distribution = probability measure μ_X is denoted by φ_X. The most important and useful property of this function is: If X and Y are independent random variables (cf. **1.4.1**),

$$\varphi_{X+Y}(\xi) = \varphi_X(\xi)\varphi_Y(\xi). \tag{1.4.18}$$

Exercise 1.
(1) Find the characteristic function of the Gaussian distribution $N(m, V)$ (with mean m and variance V).
(2) Find the characteristic function of the n-dimensional Gaussian distribution with mean $\boldsymbol{m}$ and covariance matrix V.
(3) Find the characteristic function of the *Poisson distribution* with average λ (i.e., $\mu(\{k\}) = (\lambda^k/k!)e^{-\lambda}$).
(4) Find the characteristic function of the *Cauchy distribution* whose density is given by $1/\pi(1+x^2)$. □
Exercise 2.
(1) Find the characteristic function of the *binary distribution* (for $k \in \{0, 1, \cdots, n\}$)

$$\mu(\{k\}) = \binom{n}{k} p^k(1-p)^{n-k}. \tag{1.4.19}$$

(2) Let $\lambda = np$, and take the limit $n \to \infty$, $p \to 0$ with fixed λ. The outcome is the characteristic function of the Poisson distribution just discussed (cf. **??**). □

1.4.8 Properties of characteristic functions

The characteristic function of a distribution is *positive definite*, i.e., for any positive integer n, any complex numbers $\{z_j\}_{j=1}^n$ and real numbers $\{\xi_j\}_{j=1}^n$

$$\sum_{j,k=1}^n z_j \overline{z}_k \varphi_\mu(\xi_j - \xi_k) \geq 0. \tag{1.4.20}$$

It is clear that

$$\varphi_\mu(0) = 1. \tag{1.4.21}$$

Also φ is continuous at the origin.[44] The converse is also true (→**1.4.9**), so these three properties characterize the characteristic function.

If it is sufficiently smooth, we can compute moments (around the origin) by differentiating it:

$$E(X^n) = (-i)^n \frac{d^n}{d\xi^n}\varphi_X(\xi)\bigg|_{\xi=0}. \tag{1.4.22}$$

---

[44] due to an elementary theorem of Lebesgue integration.

1.4.9 Bochner's theorem.

A positive definite function (in the sense of **1.4.8**) φ on $\boldsymbol{R}^n$ that is continuous at the origin and $\varphi(0) = 1$ is a characteristic function of some probability measure on $\boldsymbol{R}^n$.□

However, it is not always easy to check that a given function can be a characteristic function of a certain distribution. If ϕ is an even function, convex on $(0, \infty)$, and $\phi(0) = 1$, then ϕ is a characteristic function (*Polya's criterion*;[45] note that this criterion cannot be applied to the Gaussian case). For example, $\exp(-|\xi|^\alpha)$ for $\alpha \in [0, 1]$ is a characteristic function.

Exercise 1. Let $\{x(t, \omega)\}$ be a continuous stochastic process ($x(t, \omega)$ is continuous for almost all ω as a function of t), and $\langle x(0, \omega)^2\rangle_\omega = 1$. Also, let us assume that it is stationary (i.e., various expectation values do not depend on the absolute time). Demonstrate that its correlation function

$$C(t) = \langle x(t, \omega)x(0, \omega)\rangle_\omega \tag{1.4.23}$$

is a characteristic function of a probability distribution. This is related to an important topic of spectral analysis of stochastic processes (→**??**) □

1.4.10 Laplace transformation and generating function

The Laplace transform of the measure μ, if exists,

$$Z(\lambda) = \int \mu(dx)e^{\lambda x} \tag{1.4.24}$$

is called the *generating function* of μ, and λ is its parameter. It may be more appropriate to call it the partition function.

Just as we have done for the characteristic function (→**1.4.8**), we can compute moments with the aid of differentiation wrt λ. Its logarithmic differentiation is useful as the reader expects from Gibbs' work (→**2.1.4**). For example,

$$\left.\frac{d\log Z}{dlam}\right|_{\lambda=0} = \langle x\rangle. \tag{1.4.25}$$

Here (and henceforth), $\langle\ \rangle$ implies the expectation value (averaging). In the case of canonical distribution we do not set $\lambda = 0$. We also have

$$\left.\frac{d^2\log Z}{d\lambda^2}\right|_{\lambda=0} = \langle(x - \langle x\rangle)^2\rangle. \tag{1.4.26}$$

1.4.11 Cumulant expansion

The derivatives of the generating function at the origin of the parameter (→**1.4.10**) suggest that the Taylor expansion of $\log Z($
$lam)$ around $\lambda = 0$ would give useful quantities. This is the idea of *cumulant expansion*:

$$\log Z(lam) = \lambda C_1 + \frac{1}{2}\lambda^2 C_2 + \cdots + \frac{1}{n!}\lambda^n C_n + \cdots, \tag{1.4.27}$$

---

[45] Durrett p87.

where C_n is called the *n-th cumulant*, and is very often written as

$$C_n = \langle X^n \rangle_C \,. \tag{1.4.28}$$

It is easy to see that $C_1 = E(X)$, and $C_2 = V(X)$.

With the aid of this notation (1.4.27) is succinctly rewritten (symbolically) as

$$\log Z(\lambda) = \left\langle e^{\lambda X} - 1 \right\rangle_C \tag{1.4.29}$$

With the aid of $\log(1+x) = -\sum(-1)^n x^n/n$ and the multinomial expansion, cumulants are related to the ordinary moments $\langle X^n \rangle$ (around the origin) as

$$\langle X^n \rangle_C = -n! \sum_{\boldsymbol{K}\cdot\boldsymbol{N}=n} (\boldsymbol{K}\cdot\mathbf{1} - 1)!(-1)^{\boldsymbol{K}\cdot\mathbf{1}} \frac{1}{\boldsymbol{K}!}\left(\frac{\langle X^{\boldsymbol{N}}\rangle}{\boldsymbol{N}!}\right)^{\boldsymbol{K}} . \tag{1.4.30}$$

Here, $\boldsymbol{N}$ is a nonnegative integer vector $\boldsymbol{N} = (N_1, N_2, \cdots)$ such that $N_1 \geq N_2 \geq \cdots \geq 0$ (of various dimensions), and the summation is over all possible choices of $\boldsymbol{K} = (k_1, k_2, \cdots)$, where k_i specifies the number of N_i (i.e., the summation is over all the possible partition of n). $\mathbf{1} = (1, 1, \cdots)$. (Slightly extended) Hadamard's notation is adopted: for a vector $\boldsymbol{A} = (a_1, a_2, \cdots)$, $\boldsymbol{A}! = \prod a_i!$, $\boldsymbol{x}^{\boldsymbol{A}} = \prod x_i^{a_i}$; $x^{\boldsymbol{A}} = (x^{a_1}, x^{a_2}, \cdots)$; $\boldsymbol{A}/\boldsymbol{N}! = (a_1/N_1!, a_2/N_2!, \cdots)$.

For example, to compute $C_4 = \langle X^4 \rangle_C$, we first make the partitions of 4 = 3+1 = 2+2 = 2+1+1 = 1+1+1+1, or (4) ; (3)+(1); 2(2); (2)+2(1); 4(1). Therefore, if we write $\boldsymbol{K} = (k_4, k_3, k_2, k_1)$, where k_j denotes the number of j appearing in the partition $\boldsymbol{N}$, we have

partition	$\boldsymbol{K}$	$\boldsymbol{K}!$	$\boldsymbol{K}\cdot\mathbf{1}$
(4)	$(1,0,0,0)$	1	1
(3)+(1)	$(0,1,0,1)$	1	2
2(2)	$(0,0,2,0)$	2	2
(2)+2(1)	$(0,0,1,2)$	2	3
4(1)	$(0,0,0,4)$	4!	4

, (1.4.31)

so (1.4.30) tells us that

$$\begin{aligned}\langle X^4 \rangle_C &= -4!\left[0!(-1)^1\frac{1}{1!}\left(\frac{\langle X^4\rangle}{4!}\right) + 1!(-1)^2\frac{1}{1!1!}\left(\frac{\langle X^3\rangle}{3!}\right)\left(\frac{\langle X\rangle}{1!}\right) + 1!(-1)^2\frac{1}{2!}\left(\frac{\langle X^2\rangle}{2!}\right)^2\right.\\ &\qquad\left. +2!(-1)^3\frac{1}{1!2!}\left(\frac{\langle X^2\rangle}{2!}\right)\left(\frac{\langle X\rangle}{1!}\right)^2 + 3!(-1)^4\frac{1}{4!}\left(\frac{\langle X\rangle}{1!}\right)^4\right] \qquad (1.4.32)\\ &= \langle X^4\rangle - 4\langle X^3\rangle\langle X\rangle - 3\langle X^2\rangle^2 + 12\langle X^2\rangle\langle X\rangle^2 - 6\langle X\rangle^4. \qquad (1.4.33)\end{aligned}$$

Notice that the sum of all the numerical factors is zero. This may be used to check our result.

As seen from the general formula,

$$\langle X^n \rangle_c = \langle X^n \rangle + \text{ terms with extra cuts } \rangle\langle. \tag{1.4.34}$$

Exercise 1. Obtain the formula for C_6, assuming that all the odd order moments vanish. Then, compute the 6th moment of a Gaussian random variable centered at the origin in terms of its second moment. □

Exercise 2. A necessary and sufficient condition for the distribution to be Gaussian is that all the cumulants C_3 and beyond vanish. □

Exercise 3. Characterize the Poisson distribution in terms of the cumulants. □

1.4.12 Gibbs-Helmholtz relation and generating function

The internal energy is identified with (or is interpreted as) the average $\langle H \rangle$ of the Hamiltonian over the distribution P on the phase space (cf. **2.5.3**). Then, the Gibbs-Helmholtz relation

$$E = \left.\frac{\partial(A/T)}{\partial(1/T)}\right|_V = -\left.\frac{\partial(-A/k_BT)}{\partial(1/k_BT)}\right|_V, \tag{1.4.35}$$

where A is the Helmholtz free energy, may be interpreted as: $-A/k_BT$ is the logarithm of the generating function of energy Z called the *canonical partition function* (→**2.1.5**),[46] and $\beta = 1/k_BT$ is the parameter of the generating function:

$$Z = e^{-\beta A} = \sum e^{-\beta\epsilon} w(\epsilon), \tag{1.4.36}$$

where $w(\epsilon)$ is the statistical weight of the states with energy ϵ and the sum is over possible energies. What Gibbs demonstrated (→**2.1.4**) is that if $w(E)dE$ is the phase volume of the constant energy shell between the energy surfaces with energy E and $E + dE$, then indeed A satisfies the ordinary thermodynamic relation thanks to Liouville's theorem (→**2.7.2**).

However, as we have already seen, there is no way to impose this condition based on mechanics. Thus, a needed extra assumption is the *principle of equal probability* (→**2.6.2**) which is equivalent to the requirement that $w(E)dE$ is the phase volume of the energy shell between energy E and $E + dE$.

1.5 Law of Large Numbers and Ergodic Theorem

THE Law of large numbers is the most important theorem of probability theory for statistical mechanics. If we do not pay any attention to fluctuations, this (and Chebyshev's inequality **1.5.5**) is the only probability theory we need to understand macroscopic systems. Monte Carlo methods (→**1.5.6**) are direct application of this theorem. Ergodic theorems (→**1.5.8**) may be interpreted as its generalization to correlated variables.

1.5.1 Law of large numbers: an illustration

Let X_i be 1 if the i-th coin tossing of a fair coin gives a head, and 0, otherwise. $S_n = \sum_{i=1}^n X_i$ implies the number of heads we get with n tossing experiments. Then, everybody knows that surely

$$S_n/n \to 1/2 \tag{1.5.1}$$

as $n \to \infty$. Here, S_n/n is called the *sample mean* or *empirical average.* This 'law of large numbers' was first recognized by Jacob Bernoulli (1645-1705).[47]

---

[46] However, we do not set $\beta = 0$ after differentiation.

[47] The statement given here is much stronger than Bernoulli's version. His law is the *weak law.* That is, if n is large enough, then the average deviation of S_n/n from m becomes less and less (→ 0). This statement does not guarantee that for almost all the sample sequence S_n/n converges to m (for a particular sample sequence S_n/n may occasionally become big). What we intuitively know is that for almost all sample sequences $S_n/n \to m$. This is the *strong law* (→**1.5.4** footnote).

Discussion 1. Make an example of A stochastic sequence such that $P(|A_n| > 0) \to 0$, but $A_n \to 0$ never holds. The existence of such a sequence implies that the meaning of the convergence (1.5.1) is delicate. □

Exercise 1. For a finite n S_n/n is not exactly equal to 1/2 (that is, it is also a random variable), but fluctuates around it. Sketch the probability density distribution function for S_n/n for several n, say, 100, 10,000, and 1,000,000.□

From this exercise, the reader should have realized that the density distribution function of S_n/n seems to converge to $\delta(x - 1/2)$, that is, the distribution converges to an atomic measure concentrated at $x = 1/2$.

1.5.2 Law of large numbers for iid random variables

Let $\{X_i\}$ be a sequence of random numbers identically and independently distributed (*iid* is the standard abbreviation) on a probability space. Let us assume that $E(X_1) = m$ and $V(X_1) = C(< \infty)$. The law of large numbers tells us

$$S_n/n \to m \tag{1.5.2}$$

in a certain sense. As the reader expects, there are different ways to interpret this limit. Accordingly, there are different versions of law of large numbers.

1.5.3 Intuitive approach to law of large numbers

The simplest version of the law of large numbers may be to claim that the fluctuation vanishes in the $n \to \infty$ limit:

$$E((S_n/n - m)^2) \to 0. \tag{1.5.3}$$

This is called the L^2 *weak law.* This is obvious, because the LHS is equal to C/n.

What the reader has seen in Exercise 2 in **1.5.1** is a statement about the distribution of S_n/n:

$$P(S_n/n \in (x, x + dx)) \to \delta(x - m)dx. \tag{1.5.4}$$

Let us try to compute the (density) distribution f_n for S_n/n (cf. **1.4.5**):

$$f_n(x) = E\left(\delta\left(\frac{1}{n}\sum_{i=1}^{n} X_i - x\right)\right). \tag{1.5.5}$$

Its characteristic function $g_n(\xi)$[48] may be given by (cf. **1.2.2**)

$$g_n(\xi) \equiv \int dx\, e^{i\xi x} f_n(x) = E\left(\prod_{i=1}^{n} e^{i\xi X_i/n}\right) = E(e^{i\xi X_1/n})^n. \tag{1.5.6}$$

Now, let us write the characteristic function (→**1.4.7**) of X_1 as $\omega(\xi)$. Then, we have

$$g_n(\xi) = \omega(\xi/n)^n. \tag{1.5.7}$$

For very large n, the variable in ω is very close to zero. Since the distribution has a finite mean and variance, ω must be twice differentiable, with the aid of Taylor's formula, we may write $\omega(\xi/n) = 1 + im\xi/n + O(1/n^2)$. Putting this into (1.5.7), we obtain

$$\lim_{n\to\infty} g_n(\xi) = e^{im\xi}. \tag{1.5.8}$$

---

[48] We freely exchange the order of integrations. All of them can be justified a la Fubini (sometimes with some additional technical conditions).

This implies that the density distribution of S_n/n (weakly[49]) converges to $\delta(x-m)$. This is exactly what the reader has expected.

We may say that S_n/n converges to m *in law* (in distribution).

1.5.4 More standard formulation of law of large numbers

If for all $\epsilon > 0$, $P(|Y_n - Y| > \epsilon) \to 0$, we say Y_n converges *in probability* to Y.

Thus, we expect that S_n/n converges to m in probability. This is called the *weak law of large numbers*:

[**Weak law of large numbers**]. Let $\{X_i\}$ be iid random variables with $E(|X_1|) < \infty$ and $E(X_1) = m$. Then, S_n/n converges to m in probability.

Notice that the weak law does not hold, if $\{X_i\}$ obeys a Cauchy distribution (certainly, $E(|X_1|) < \infty$ is violated).[50]

1.5.5 Chebyshev's inequality and demonstration of weak law

To demonstrate the weak law of large numbers the most elementary way may be to use *Chebyshev's inequality*:

$$a^2 P(|X - m| \geq a) \leq V(X). \tag{1.5.9}$$

This can be shown as follows (let us write $X - m$ as X):

$$V(X) = \int x^2 \mu(dx) \geq \int_{|x|>a} x^2 \mu(dx) \geq a^2 \int_{|x|>a} \mu(dx). \tag{1.5.10}$$

Exercise 1. A more general inequality (in these days this is called Chebyshev's inequality) is: Let f be a nonnegative monotone increasing function on $[0, \infty)$. Then,

$$P(X \geq a) \leq \frac{E[f(\max(X, 0))]}{f(a)}. \tag{1.5.11}$$

□

Exercise 2. Apply (1.5.9) to S_n/n and demonstrate the weak law. □

1.5.6 Monte Carlo integration

Suppose we wish to compute the following integral over a very high dimensional cube ($N \gg 1$):

$$I = \int_{[0,1]^N} d^N x\, f(x). \tag{1.5.12}$$

If we wish to sample two values for each coordinate, we need 2^N sampling points. If $N = 20$ we need more than a million points, so we realize that numerical computation of a high

---

[49] Weak convergence $f_n \to f$ means $\int f_n \varphi\, dx \to \int f \varphi\, dx$ for 'any' φ taken from an appropriate function set. In the present case we may take the set of smooth functions with bounded support, that is, the functions that vanish outside some interval.

[50] There is a stronger version due to Kolmogorov (the *strong law of large numbers*):
Let $\{X_i\}$ be iid random numbers with $E(|X_1|) < \infty$ and $E(X_1) = m$. Then, S_n/n almost surely converges to m.

If the distribution of S_n/n converges weakly, we say S_n/n converges *in law* (or in distribution). Almost convergence ⇒ convergence in probability ⇒ convergence in the mean (if the variance exists). Convergence in the mean ⇒ convergence in probability ⇒ convergence in law. The converses of these ⇒ relations do not hold.

dimensional integral using numerical quadrature methods is prohibitively difficult.

(1.5.12) may be interpreted as the computation of $E(f(X))$, where X is a random variable on $\{\mu, [0,1]^N\}$ with μ being the Lebesgue measure (the ordinary volume →**1.A.13**) on $\boldsymbol{R}^N$. Then, the law of large numbers tells us

$$I_n = \frac{1}{n}\{f(X_1) + \cdots + f(X_n)\} \to I. \tag{1.5.13}$$

How large n is required? If $f(x) \in [0,1]$ (if the integrand is bounded, this can always be satisfied by appropriate shifting and scaling), then $V(f(X)) \leq I(1-I)$, so Chebyshev's inequality tells us

$$P(|I_n - I| > \epsilon) \leq \frac{I(1-I)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}. \tag{1.5.14}$$

Thus, if we wish to have an error less than $\epsilon = 1/20$ with probability at least 0.99, then $100/n < 0.01$. Therefore, $n = 10,000$ is needed. This is certainly much less than the number of points needed in the regular sampling methods.[51]

A remarkable feature of the Monte Carlo integration is that the cost does not depend on the dimensionality of the integral.

Discussion 1. Do we really need not to pay any price for high dimensionality? □

1.5.7 Importance sampling

(1.5.12) may be rewritten as

$$I = \int_{[0,1]^N} d^N x f(x) = \int_{[0,1]^N} d^N x\, w(x)\frac{f(x)}{w(x)} \tag{1.5.15}$$

with a positive function $w(x)$. We may freely choose this function with the normalization condition $\int d^N x\, w(x) = 1$. If we use the random number whose density distribution is given by w (→Exercise of **1.4.7** for how to make such random numbers), the law of large numbers tells us that

$$J_n = \frac{1}{n}\left(\frac{f(X_1)}{w(X_1)} + \cdots + \frac{f(X_n)}{w(X_n)}\right) \to I. \tag{1.5.16}$$

Exercise 1.
(1) Compute the variation of J_n. (Ans., $(1/n)\left\{\int (f^2/w) d^N x - I^2\right\}$)
(2) Find w that minimizes the variation just obtained. Notice that if f is of constant sign, this choice makes the variation vanish. □

1.5.8 Ergodic theorem as a law of large numbers

So far we have discussed the law of large numbers for independent random variables. There are many cases in which we wish to compute the sum of statistically not independent random variables. There are various extensions of the law of large numbers to such cases, but an ultimate version may be ergodic theorems. We will come to this topic much later (in Part III, not yet written), but it may be a good occasion to learn the meaning of *ergodicity*, which was first conceived by Maxwell and Boltzmann to found equilibrium statistical mechanics

---

[51] However, this estimate is not an almost sure estimate. The error for each sample sequence (i.e., the realization of $\{I_n\}$) decays as $\sqrt{32 \log\log n/n}$. Therefore, the worst case scenario is that the error decays roughly as $1/\sqrt{n}$.

(→**2.4.1**).

Let $x(t)$ be a sample time evolution of a dynamical system on a phase space Γ with the time evolution operator T_t (i.e., $T_t x(s) = x(t+s)$) and μ be an invariant measure (= time-independent distribution on the phase space Γ →**2.7.2**).[52] Then, the time average of an observable $f(T_t x_0)$ converges to a value dependent only on the initial condition:[53]

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t f(T_s x_0)\,ds = f^*(x_0) \tag{1.5.17}$$

for almost all $x_0 \in \Gamma$ with respect to the measure μ.[54] Further more,

$$\int f^*(x)d\mu(x) = \int f(x)d\mu(x). \tag{1.5.18}$$

This is the famous *Birkhoff ergodic theorem*.[55] It should be obvious that the essence of the ergodic theorem is quite analogous to the law of large numbers.

1.5.9 Ergodicity implies time average = phase average.

Intuitively speaking, a dynamical system is said to be *ergodic*, if "for any x_0, x_t can visit everywhere in Γ." (We will see, however, this careless statement contains a fatal flaw. →**1.5.10**)

Take a set A ($\subset \Gamma$). x_t starting from 'any' point x_0 in Γ can visit it, if the dynamical system is ergodic. Therefore, if A is an invariant set,[56] and has some positive weight ($\mu(A) > 0$), A and $T_t^{-1}A$ (= the set of points that visit A in time t) must agree for any t. But since the points can visit 'everywhere' in Γ, we must conclude that A with a positive weight (i.e., $\mu(A) > 0$) is actually Γ. Therefore, if A is a nonempty invariant set, $\mu(A) = \mu(\Gamma)$ ($= 1$).[57]

The function f^* in (1.5.17) is time independent, so a set $\{x : f^*(x) < c\}$ for any number c must be an invariant set. Therefore, if the system is ergodic, it must be a constant. Therefore,

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t f(T_s x_0)\,ds = \int f^*(x)\mu(dx) = \int f(x)\mu(dx). \tag{1.5.19}$$

If the volume of an energy shell corresponding to the microcanonical ensemble is ergodic, what Boltzmann wanted is true: time average = ensemble average.

1.5.10 Is ergodic theory really relevant to us?

The discussion in **1.5.9** must be made precise, and the reader will realize that the ergodic problem is almost irrelevant to equilibrium statistical mechanics.

In **1.5.9** it is written that a dynamical system is said to be ergodic, if for 'any' x_0, x_t

---

[52] Let $\{T_t\}$ be a one parameter family of maps $\Gamma \to \Gamma$ such that $T_0 = 1$, $T_t T_s = T_{t+s}$. Then, $\{T, \Gamma\}$ is called a *dynamical system*. If μ is an invariant measure, $\{T_t, \Gamma, \mu\}$ is called a *measure theoretical dynamical system.*

[53] for almost all initial conditions wrt μ.

[54] That is, (1.5.17) does not hold only on μ-measure zero set. Notice, however, that for a dynamical system $\{T_t, \Gamma\}$ the invariant measure is not unique. Often, there are uncountably many invariant measures. For any of them, Birkhoff's theorem holds!

[55] For a proof, see, for example, Section 1.6 of P. Walters, *An Introduction to Ergodic Theory* (Springer, 1982). This is an excellent (and probably a standard) introduction to the topic.

[56] That is, $\mu(T_t^{-1}A \triangle A) = 0$ for any t. Here, $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

[57] A precise definition of μ being ergodic is that $\mu(A) = 0$ or 1 for μ-invariant set A.

can visit 'everywhere' in Γ. Precisely speaking, there are exceptional points that cannot visit many places. Therefore, we must specify how to measure the amount of these exceptional points. That is, the concept of ergodicity is a measure-theoretical concept. Thus, precisely speaking, we must say:
A measure theoretical dynamical system $\{T, \mu, \Gamma\}$[58] is *ergodic*, if μ-almost all points can visit μ-everywhere on Γ.[59]

The most important message here is that the concept of ergodicity is the concept of measure theoretical dynamical systems = dynamical systems with *given* invariant measures. The most fundamental question of equilibrium statistical mechanics is the justification of the principle of equal probability. As the reader trivially realizes, ergodic theory presupposes an invariant measure, and the selection of a particular invariant measure is not in the scope of the theory.

However, the reader might say that she accepts that ergodic theory has nothing to say about the fundamental principle of statistical mechanics, but that after an invariant measure (= stationary state distribution) is given, it is important for the measure to be ergodic. However, we must not forget that macroscopic observables are extremely special observables (= measurable functions in mathematical terms[60]). That is, the equality of the time average and the phase average, even if required, need not be true for most measurable functions. Therefore, mathematical ergodicity is not required.

Generally speaking, to understand that the macroscopic observable as the time average is a wrong point of view, because, time average does not converge very quickly for general measurable functions; it is highly questionable to believe that the time average converges within the microscopic time scale. It is close to the truth that macroscopic observables are instantaneous values. This point of view will be elaborated in **??**. The crucial point is that phase functions whose averages give macroscopic observables have special properties such that they are often sums of functions depending on a few particle coordinates. Needless to say, such functions do not change in time very much, so even their short time average agree with their phase average. However, this agreement is not due to the ergodicity of the system.

Thus, there are crucial reasons for ergodicity to be irrelevant to equilibrium statistical mechanics.[61,62]

1.5.11 Two refinements of law of large numbers

There are at least two ways to refine the weak law of large numbers **1.5.4**. One is to study small fluctuations around the expectation value for finite n: Is there any power α that makes the following limit

$$\lim_{n\to\infty} P(|S_n/n - m| < \epsilon n^{\alpha}) \tag{1.5.20}$$

a nonsingular function of ϵ? This is the question of the central limit theorems and has a close relation to renormalization group theory (→**2.6**).

---

[58] A measure theoretical dynamical system $\{T, \mu, \Gamma\}$ is a dynamical system on Γ with an invariant measure μ.

[59] A non-measure theoretical counterpart is the *topological transitivity*: there is a trajectory which is dense in Γ. Generally speaking, the latter concept is much more strict than ergodicity, but may be closer to the original idea of Maxwell and Boltzmann.

[60] 'Measurable' means that any 'level set' $\{x : a \le f(x) \le b\}$ for any real a, b is measurable, that is, it has a probability with respect to μ.

[61] Or we might be able to say as follows: even if time averages and phase averages should agree, for the agreement ergodicity is not the only reason.

[62] More radically speaking, invariance of measure is not required even for equilibrium statistical mechanics, either, because we need stationarity for very special observables only.

The other question is to ask the speed of convergence of S_n/n to m: How does $P(|S_n/n - m| > \epsilon)$ go to zero as a function of n? This is the question of large deviation (→**2.5**). For statistical mechanics, large deviation is more important than central limit theorems (→**2.8.5**; **1.6.4**).

1.6 Large Deviation Theory

This is a study of macroscopic fluctuations. Its importance to us is obvious, because the study of fluctuations is one of the keys of nonequilibrium statistical thermodynamics (→**4.1.2**, **5.2.5**). We will later see its close relation to Langevin equations and principle of linear nonequilibrium statistical mechanics (→**X1.1.3**). Notice also a close relation between variational principles and large deviations (→**1.6.3**, **X1.1.10**).

There is not a very accessible reference for this topic. For statistical physicists the explanation given in this section is almost enough, but mathematically, a more careful statement is needed. A standard textbook the lecturer likes is:[63]

J D Deuschel and D W Strooke, *Large Deviation* (Academic Press, 1989).

1.6.1 Large deviation: a simple example

Let $\{X_i\}$ be iid random variables, and $S_n = x_1 + \cdots +_n$ be the empirical (sample) sum of n values. We know from the weak law of large numbers **1.5.4** that

$$P(|S_n/n - m| > \epsilon) \to 0 \tag{1.6.1}$$

in the $n \to \infty$ limit for any positive ϵ. An interesting question is the speed of this convergence to zero; if the fluctuation is large, the decay should be slow. Therefore, studying the decay rate should give us information about fluctuations.

Cramér first considered this type of problem for the coin tossing, that is, we have a sequence of iid (identically and independently distributed →**1.5.2**)) random numbers $\{X_i\}$ such that $P(X_1 = 1) = P(X_1 = -1) = 1/2$. Let us estimate the LHS of (1.6.1). Using Stirling's formula $m! \sim (m/e)^m$, we may approximate (taking the leading terms only)

$$\begin{aligned}
P(|S_n/n| > \epsilon) &= \sum_{S_n \ge n\epsilon} \frac{n!}{\left(\frac{n+S_n}{2}\right)!\left(\frac{n-S_n}{2}\right)!}2^{-n} + \sum_{S_n \le -n\epsilon} \frac{n!}{\left(\frac{n+S_n}{2}\right)!\left(\frac{n-S_n}{2}\right)!}2^{-n}, && (1.6.2)\\
&\sim 2^{-n+1}\frac{n!}{\left(\frac{n+n\epsilon}{2}\right)!\left(\frac{n-n\epsilon}{2}\right)!}, && (1.6.3)\\
&\sim \exp\left[n\log n - \frac{n(1+\epsilon)}{2}\log\frac{n(1+\epsilon)}{2} - \frac{n(1-\epsilon)}{2}\log\frac{n(1-\epsilon)}{2}\right], && (1.6.4)\\
&= \exp\left[-\frac{n}{2}[(1+\epsilon)\log(1+\epsilon) + (1-\epsilon)\log(1-\epsilon)]\right]. && (1.6.5)
\end{aligned}$$

---

[63] For an outline, Appendix to Y Oono, "Large deviation and Statistical Mechanics," Prog. Theor. Phys. Supple **99**, 165 (1989) may be short.

Or, we have

$$P(|S_n/n| > \epsilon) \sim \exp\left(-\frac{n}{2}\epsilon^2\right). \tag{1.6.6}$$

Thus, we have found that the decay is exponential. It is known that exponential decay is the rule.

Exercise 1. If somehow the coin has a strange memory and has a tendency to give the same surface consecutively, what happens to the above exponent ? Guess the answer and give a reason for it (no calculation required).[64] □.

1.6.2 Level-1 large deviation principle

Consider the empirical expectation value S_n/n. The theoretical framework to study the convergence rate of this value to the mean is called the *level-1 large deviation theory.* We say that the set of random variables $\{X_i\}$ satisfies the (level-1) *large deviation principle*, if the following holds:

$$P(S_n/n \sim x) \sim e^{-nI(x)}, \tag{1.6.7}$$

where $I(x)$ is called the *rate function* (*large deviation function* or *Cramér's function*)[65] and satisfies:

(i) $I(x)$ is a convex nonnegative function,

(ii) $I(m) = 0$ is the unique minimum.

For an iid random variable sequence with a finite variance, the principle certainly holds.

1.6.3 Large deviation principle and variational principle.

Notice that if we know the rate function I, it gives a minimum principle for the mean. That is, large deviation and variational principles are closely related; generally, we may claim that there is a large deviation principle behind a variational principle (at least in statistical physics).

The reader would then naturally ask whether the important principle of equilibrium statistical mechanics, the Gibbs variational principle stating that the equilibrium state is variationally defined by the minimization principle of the free energy (→**2.1.5**), has an underlying large deviation theoretical framework. The answer is affirmative, and the framework is thermodynamic fluctuation theory (**1.6.4**, →**2.8.5**).

Furthermore, since we have learned that the Birkhoff ergodic theorem is a kind of weak law of large numbers (→**1.5.8**), the reader must have expected the corresponding large deviation framework and its accompanying variational principle for dynamical systems (chaos). This framework is the so-called *thermodynamic formalism* for Axiom A dynamical systems.[66]

1.6.4 Thermodynamic fluctuation theory as large deviation principle

Let us look at the fundamental principle of thermodynamic fluctuation (→**2.8.5**). We know entropy is extensive, so introducing the entropy per particle $s = S/N$, we have

$$P(\{\delta x\}) = e^{N\delta^2 s/k_B}. \tag{1.6.8}$$

This implies two things:

(i) $-\delta^2 s/k_B$ is the rate function for equilibrium thermodynamic observables.

(ii) the large deviation theory for iid random variables works.

We will see that this picture is correct away from the critical point (→**2.8.3**; see also **??**).

---

[64] See Prog. Theor. Phys. Suppl. **99**, 165 (1989) for graphs.

[65] $A \sim B$ implies $\log A/\log B \to 1$ asymptotically.

[66] A standard textbook for this is D Ruelle, *Thermodynamic Formalism* (Addison-Wesley).

1.6.5 How to compute rate function: Gärtner-Ellis theorem

The relation of the rate function to the fluctuation theory tells us how to compute the rate function. Let us compute the partition function = generating function.

$$\begin{aligned} Z_N(\lambda) &= \int dx\, P\left(\frac{1}{N}\sum_{i=1}^{N} X_i \simeq x\right) e^{N\lambda x}, & (1.6.9)\\ &= E\left(\int dx\, \delta\left(\frac{1}{N}\sum_{i=1}^{N} X_i - x\right) e^{N\lambda x}\right), & (1.6.10)\\ &= E\left(e^{\lambda \sum X_i}\right). & (1.6.11)\end{aligned}$$

Here, we have used (1.4.11).

On the other hand,[67]

$$\begin{aligned} Z_N(\lambda) &= \int dx\, e^{-NI(x)+N\lambda x}, & (1.6.12)\\ &\sim e^{N \sup_x[\lambda x - I(x)]}. & (1.6.13)\end{aligned}$$

Therefore, we obtain the 'thermodynamic potential' $A(\lambda)$ as

$$\frac{1}{N}\ln Z_N(\lambda) \to A(\lambda) = \sup_x[\lambda x - I(x)]. \tag{1.6.14}$$

Or, inverse Legendre transformation (→**2.6.5**) gives

$$I(x) = \sup_\lambda[\lambda x - A(\lambda)]. \tag{1.6.15}$$

Thus, we have formally shown the following:
if the thermodynamic potential exists, large deviation principle holds (*Gärtner-Ellis's theorem*).

1.6.6 Empirical distribution function

We can study not only empirical averages but also empirical (density) distribution functions: Let $\{X_i\}$ be a sequence of iid random variables. Then, the empirical density $f_N(x)$ is given by

$$f_N(x) = \frac{1}{N}\sum_{i=1}^{N} \delta(X_i - x). \tag{1.6.16}$$

Notice that this is an empirical average of $\delta(X_i - x)$ (cf. **1.4.5**), so there must be a corresponding version of the weak law of large numbers (→**1.5.4**). Indeed, this converges (weakly) to the (density) distribution function f of the random variables.[68] Then, it is natural to consider the large deviation theory for the empirical distribution.

---

[67] thanks to WKB approximation or Varadhan's theorem

[68] Around here, the word 'function' is loosely used; generalized functions such as the δ-function are also called functions.

1.6.7 Level-2 large deviation principle and Sanov's theorem

The large deviation theory for measures (distributions) is called the *level-2 large deviation theory.* The *level-2 large deviation principle* reads

$$P(f_N \sim g) \sim e^{-NI^{(2)}(g)}, \tag{1.6.17}$$

where the rate function(al) is a function on the space of distributions, and satisfies:
(i) $I(g)$ is a convex nonnegative function,
(ii) $I(f) = 0$ is the unique minimum.

For iid random variables $\{X_i\}$ with the density distribution f, *Sanov's theorem* tells us that[69]

$$I^{(2)}(g) = \int dx\, g(x) \log \frac{g(x)}{f(x)}. \tag{1.6.18}$$

'Theoretical physics style' demonstration of this is not hard, if we mimic the level-1 calculation in **1.6.5**.

1.6.8 'Demonstration' of Sanov's theorem

To estimate $P(f_N \sim g)$ we use the characteristic function(al) technique (→**1.6.5**). First, notice that

$$P(f_N \sim g) = E(\Delta(f_N - g)), \tag{1.6.19}$$

where Δ is the functional delta function.[70] Therefore, we get the 'partition function'

$$\begin{aligned} Z_N(\xi) &= \int \delta[g]\, P(f_N \sim g) e^{N\int \xi(x)g(x)dx} = E\left[\exp\left(\sum \xi(X_i)\right)\right] && (1.6.20)\\ &= E\left(e^{\xi(X_i)}\right)^N = \left[\int dx f(x) e^{\xi(x)}\right]^N. && (1.6.21)\end{aligned}$$

On the other hand, by definition,

$$Z_N(\xi) = \int \delta[g]\, e^{N\left[\int \xi(x)g(x)dx - I^{(2)}(g)\right]}, \sim e^{N\sup_g\left[\int \xi(x)g(x)dx - I^{(2)}(g)\right]}. \tag{1.6.22}$$

Therefore, introducing the 'free energy' (thermodynamic potential),[71] we obtain

$$A(\xi) = \lim_{N\to\infty} \frac{1}{N} \log Z_N(\xi) = \log \int dx f(x) e^{\xi(x)} = \sup_g \left[\int dx\, \xi(x) g(x) - I^{(2)}(g)\right]. \tag{1.6.23}$$

This implies

$$I^{(2)}(g) = \sup_\xi \left[\int dx\, \xi(x) g(x) - A(\xi)\right]. \tag{1.6.24}$$

---

[69] If we write the probability measure for the random variable as μ and the empirical measure to be ν, then

$$I^{(2)}(\nu) = \int \nu(dx) \log \frac{d\nu}{d\mu}.$$

[70] A function $f(\alpha)$ may be interpreted as 'the α-component of the vector f.' Functional analysis may be informally interpreted as the analysis of functions regarded as such vectors. The n-dimensional δ-function is defined as $\delta(\boldsymbol{v}) = \delta(v_1)\cdots\delta(v_n)$. Analogously, $\Delta(f) = \cdots\delta(f(\alpha))\cdots$, where the 'product' is the 'continuous' product.

[71] If the limit does not exist, replace it with lim sup.

Let us calculate (1.6.24). To compute the supremum, we functional-differentiate [ ]:

$$\frac{\delta}{\delta\xi(x)}\left[\int dx\,\xi(x)g(x) - A(\xi)\right] = g(x) - e^{\xi(x)}f(x)\Big/\int dx\, e^{\xi(x)}f(x) \tag{1.6.25}$$

Equating this with zero, we obtain an equation for ξ that gives the supremum of [ ]: $g(x) = e^{\xi(x)}f(x)/e^{A(\xi)}$. Notice that this implies $\int dx\, g(x) = 1$. From this, we see $\xi = \log(g/f) + A$, so putting this ξ into (1.6.24), we obtain (1.6.18).

1.6.9 Kullback-Leibler entropy

The rate function appearing in Sanov's theorem (→**1.6.7**)

$$K(g, f) = \int dx\, g(x)\log\frac{g(x)}{f(x)} \tag{1.6.26}$$

is called the *Kullback(-Leibler) entropy.* Notice that if f is constant, this is essentially the Gibbs' entropy (→**2.1.5**) or Shannon's information formula (→**1.3.2**). This becomes minimum if and only if $g = f$.[72]

The most important property of $K(g, f)$ is its non-negativity. This is obvious from the log sum inequality **1.3.8**. Or, from the scratch with the aid of the convexity[73] of $x\log x$:

$$\begin{aligned}\int dx\, g(x)\log\frac{g(x)}{f(x)} &= \int dx\, f(x)\frac{g(x)}{f(x)}\log\frac{g(x)}{f(x)} &(1.6.27)\\ &\geq \left(\int dx\, f(x)\frac{g(x)}{f(x)}\right)\log\left(\int dx\, f(x)\frac{g(x)}{f(x)}\right) = \log 1 = 0. &(1.6.28)\end{aligned}$$

Therefore, $I(g)$ vanishes if and only if g is equal to the true density distribution.

1.6.10 Information gain

The meaning of the Kullback entropy (→**1.6.9**) may be made intuitive by the following consideration.[74]

Suppose a macroscopic state α consists of microstates $\{i\}$, and all the microstates belonging to it is equally likely: if the macrostate α contains n_α microstates, their probability is p_α/n_α, where p_α is the probability for a microstate to be compatible with macrostate α.[75] Let us compare the entropy difference between a macrostate specified by $\{p_\alpha\}$ and the macrostate specified by $\{p_\alpha^0\}$ whose microstates are all equally probable (i.e., $p_\alpha^0/n_\alpha = \text{const}$ independent of α, certainly this is a characteristics of an equilibrium distribution):

$$\begin{aligned}\Delta S &= -\sum_m (p_\alpha/n_\alpha)\log(p_\alpha/n_\alpha) + \sum_m (p_\alpha^0/n_\alpha)\log(p_\alpha^0/n_\alpha), &(1.6.29)\\ &= -\sum_m (p_\alpha/n_\alpha)\log(p_\alpha/p_\alpha^0) + \sum_m [(p_\alpha^0/n_\alpha) - (p_\alpha/n_\alpha)]\log(p_\alpha^0/n_\alpha) &(1.6.30)\\ &= -\sum_\alpha p_\alpha \log(p_\alpha/p_\alpha^0). &(1.6.31)\end{aligned}$$

---

[72] Precisely speaking, the equality is in the almost sure sense.

[73] If f is a convex function, $\langle f(x)\rangle \geq f(\langle x\rangle)$ (*Jensen's inequality*).

[74] due to F Schlögl

[75] that is, a macrostate is a bunch of equally likely microstates under the condition that macroscopic description is meaningful.

Here $\sum_m$ means the sum over all the microstates, and $\sum_\alpha$ means the sum over all the macrostates. We have used $p^0_\alpha/n_\alpha =$ constant (equidistributed), and $\sum_m[(p^0_\alpha/n_a)-(p_\alpha/n_\alpha)] = 1-1 = 0$ due to normalization of probabilities.

Therefore, the Kullback entropy $K(p, p^0)$ measures the gain of information due to the non-even distribution at the coarse-grained level. Therefore, sometimes the Kullback entropy is called the *information gain*.

1.6.11 From level 2 to level 1: contraction principle.

Let the level-2 large deviation function be $I^{(2)}(g)$ for the (density) distribution g for a random variable X. We must be able to know more from the level-2 rate function (→**1.6.7**) than we can know from the level-1 rate function $I(x)$ for X. Then, we should be able to obtain I from $I^{(2)}$. This is possible by the *contraction principle*:

$$I(x) = \inf_{g:\int dy\, yg(y)=x} I^{(2)}(g). \tag{1.6.32}$$

This implies that the value of the level-1 rate function at x is obtain by minimizing the level-2 rate function over the distributions having the same expectation value x. In other words, the conditional minimization of $I^{(2)}(g)$ under $\int y\, g(y)\, dy = x$ gives $I(x)$.

The reader may have already guessed this result. It is not hard to demonstrate it. Notice, first, that

$$\frac{1}{N}\sum_i X_i = \int dy\, y\, f_N(y), \tag{1.6.33}$$

where f_n is the empirical distribution given in (1.6.16). Therefore,

$$\begin{aligned}
P(S_N/N \sim x) &= E\left(\delta(S_N/N - x)\right) = E\left[\delta\left(\int dy\, y f_N(y) - x\right)\right], && (1.6.34)\\
&= E\left[\int \delta g\, \Delta(g - f_N)\delta\left(\int dy\, yg(y) - x\right)\right], && (1.6.35)\\
&= \int \delta g\, P(f_N \sim g)\delta\left(\int dy\, yg(y) - x\right) = \int \delta g\, e^{-NI^{(2)}(g)}\delta\left(\int dy\, yg(y) - x\right), && (1.6.36)\\
&\sim \exp\left[-N\left(\inf_{g:\int dy\, yg(y)=x} I^{(2)}(g)\right)\right]. && (1.6.37)
\end{aligned}$$

This implies the desired asymptotic relation (1.6.32).

A useful theorem is:

Let $\{X_n\}$ be iid sequence obeying a common law μ. Suppose that for each n f_n is a symmetric function of $X_1, \cdots, X_n$ and the strong law of large numbers $f_n \to f(\mu)$ holds.[76] Then, the large deviation principle for f_n holds with the rate function I given by the contraction principle:

$$I(y) = \inf_{\nu:\int dy\, yf(\nu)=y} I^{(2)}(\nu). \tag{1.6.38}$$

---

[76] N. O'Connell, "A large deviations heuristic made precise," Math. Proc. Camb. Phil. Soc. **128**, 561-569 (2000).

1.6.12 Derivation of canonical distribution by contraction

Now, let us apply the result (1.6.32) to iid random variables X with distribution f. The level-2 rate function is given by Sanov's theorem (1.6.18), so to obtain the level-1 rate function according to (1.6.32), we must minimize the Kullback entropy (→**1.6.9**) conditionally.

With the aid of a Lagrangian multiplier, we minimize

$$\int dx\; g(x) \log \frac{g(x)}{f(x)} + \alpha \int dx\, g(x) + \beta \int dx\, x g(x). \tag{1.6.39}$$

This is minimized, if

$$g(x) = f(x) e^{-\beta x}/Z \tag{1.6.40}$$

where the numerical constant β is determined to give the average y, and Z is the normalization constant (that corresponds to the canonical partition function). In other words, if the average is known for an ensemble that is obtained by restricting the original ensemble whose distribution is f, it is given by the 'canonical distribution' (1.6.40).

The usual derivation of the canonical ensemble from the microcanonical ensemble is to select the local energy fluctuations compatible with the temperature (or, equivalently, compatible with a give expectation value of energy). Therefore, we obtain the canonical distribution with the aid of contraction principle (→**1.6.11**). That is, the above derivation of the canonical distribution is equivalent to the standard derivation explained in textbooks (→**2.5.2**).

It is clear that we must specify f, and this specification is the principle of equal probability (→**2.6.2**). The method to derive the canonical distribution through maximizing Shannon's information formula conditionally is only a corollary to Sanov's theorem with the principle of equal probability (cf. **1.3.6**).

1.6.13 Contraction may be the key to nonequilibrium statistical mechanics

The idea of sampling right fluctuations to prepare the right ensemble is a fundamental idea, and is the essence of the regression hypothesis (→**4.1.2**): even nonequilibrium states are realized by equilibrium fluctuations, so if we collect right fluctuations, we should be able to prepare statistical ensembles that can describe nonequilibrium phenomena. The strategy certainly works for linear nonequilibrium phenomena, and is expected to be applicable more widely.

1.6.14 How do we justify ensemble theory?

When we observe a macroscopic system, usually we study a single sample. Therefore, there is a criticism that ensemble theories are unphysical.[77] Those who criticize the ensemble theory use time average to justify statistical mechanics.

When we measure a thermodynamic observable, we measure it for much less than a second or so. During such short observation times the system can cover only an insignificant portion of the whole phase space (→**1.5.10**), so the idea is not at all convincing that ergodicity of the Liouville measure justifies the ensemble.

---

[77] Ma's textbook, *Statistical Mechanics* (World Scientific) expels ensemble altogether from statistical mechanics, and interprets all the averages as time averages. This is due to a fundamental misunderstanding of the nature of (equilibrium) statistical mechanics.

One possibility to rescue this idea of time average might be to pay attention to the observation that thermodynamic observables are very special observables.[78] Macroobservables are the sums of functions that depend only on the extremely small (say, 10^6-dimensional) subspaces of the whole phase space. For each function of the small subspace the conclusion of the ergodic theory applies: time average during the macroobservation time = its phase space average. However, notice that the Poincaré time of this small sub-phase space is still astronomical. Therefore, time average theory is hardly relevant. To use time average to justify the ensemble theory is hopeless.

A much better possibility is based on the idea that these functions on small subspaces are virtually statistically independent, so their averages = macroobservables almost surely assume their average values thanks to the (strong) law of large numbers. Statistical mechanics works due to strong law of large numbers.

1.6.15 Large deviation point of view of nonequilibrium statistical mechanics

The reader may have wondered why we discuss the general framework of probability theory in this nonequilibrium course. Here, a preview is given that illustrates why large deviation theory is crucial. When we wish to discuss a statistical framework, there must be a clear distinction between the microscopic (mechanical) level and phenomenological (macroobservable) level.[79] For example, we need a clear separation of time scale between the microscopic time scale t_m and the macroscopic time scale t_M. That is, $t_m \ll t_M$. This implies that we could introduce the mesoscopic time scale τ such that $t_m \ll \tau \ll t_M$.

Let us consider an observable x whose long-time average $\langle x \rangle$ is a macroscopic observable. The time derivative dx/dt averaged over the mesoscopic level should read

$$\frac{\Delta x(t)}{\Delta t} = \frac{1}{\tau}\int_t^{t+\tau} \frac{dx}{dt} = \frac{x(t+\tau)-x(t)}{\tau}. \tag{1.6.41}$$

We assume that this obeys the large deviation principle (Onsager's principle):

$$P\left(\frac{\Delta x}{\Delta t} \sim \dot{x}\right) \sim e^{-\tau I(\dot{x})}. \tag{1.6.42}$$

Notice, that the above average is under the condition that $x(t)$ is given, so $I(\dot{x})$ should be written as $\dot{x}|x(t))$ more precisely. The unique zero of I gives the hydrodynamic (macroscopic) law; at this time scale we should interpret $\Delta x/\Delta t$ as the time derivative, so the unique zero of I is a function $f(x)$ of x:

$$\frac{dx}{dt} = f(x). \tag{1.6.43}$$

We may assume that I is well-behaved near its zero, we may approximate as

$$I(\dot{x}|x) = \frac{1}{2V}(\dot{x} - f(x))^2. \tag{1.6.44}$$

This implies that if we study the time derivative at the time scale of τ

$$\frac{dx}{dt} = f(x) + \nu, \tag{1.6.45}$$

---

[78] This idea was first emphasized by Khinchine: A. Ya. Khinchine, *Mathematical Foundation of Statistical Mechanics* (1943, Moskow) (Dover, translated by G Gamov).

[79] It is often called the hydrodynamic level.

where the term ν is a noise with $\langle \nu(t)\nu(0)\rangle = (V/\tau)\delta(t)$ (Gaussian white noise), and this equation is called a *Langevin equation.*

Because $x(t+\tau) = x(t) + \tau\Delta x/\Delta t$, $P\left(\frac{\Delta x}{\Delta t} \sim \dot{x}\right)$ may be interpreted as $P(x(t+\tau)|x(t))$; as is noted for I P in (1.6.42) is actually a conditional probability. That is, the transition probability can be written in terms of the rate function:

$$P(x(t+\tau)|x(t)) \sim e^{-\tau I((x(t+\tau)-x(t))/\tau)}. \tag{1.6.46}$$

If the reader is familiar with the Feynman path integral, then she would immediately see that the rate function is just the Lagrangian for a path integral (called the *Onsager-Machlup path integral*). Just the the Feynman path integral gives a Green's function for the Schrödinger equation, we can expect that our path integral solves a partial differential equation (called the *Fokker-Planck equation*).

If $f(x)$ is linear as $f(x) = -\zeta x$, then there must be a definite relation between V and ζ. This is the *fluctuation-dissipation relation* and linear nonequilibrium thermodynamics results.

1.7 Central Limit Theorems

The basic idea of central limit theorems is the coarse-graining + scaling of renormalization group theory, which may be interpreted as an attempt to generalize the central limit theorem to strongly correlated random variables.

The classic reference for this topic is

W Feller, *An introduction to probability theory and its applications*, vol. II (Wiley, 1971). Durrett is also excellent.

1.7.1 Central limit theorem

Let $\{X_i\}$ be iid (= independently and identically distributed) random variables with $E(X_1) = m$ and $V(X_1) = \sigma^2$ (→**1.4.3**). Then, the distribution of $Y_n \equiv (S_n - nm)/\sigma\sqrt{n}$ converges weakly to $N(0,1)$ (= the Gaussian or normal distribution with mean 0 and variance unity).

A formal demonstration of this theorem is not hard, if we use the characteristic function (→**1.4.7**) ω of X_1. Since we can always make the expectation vanish by shifting the variables, we may assume $m = 0$. First, note that for small ξ

$$\omega(\xi) = 1 - \frac{\sigma^2\xi^2}{2} + o(\xi^2). \tag{1.7.1}$$

This follows from the twice differentiability of the characteristic function (→**1.4.8**). Let us compute the characteristic function ω_n of Y_n:

$$\omega_n(\xi) = E(e^{i\xi Y_n}) = E\left(\prod_{j=1}^{n} e^{i\xi X_j/\sqrt{n}\sigma}\right) = E\left(e^{i\xi X_1/\sqrt{n}\sigma}\right)^n = \omega(\xi/\sqrt{n}\sigma)^n. \tag{1.7.2}$$

Therefore,

$$\lim_{n\to\infty} \omega_n(\xi) = \lim_{n\to\infty}\left(1 - \frac{\xi^2}{2n\sigma^2} + o(\xi^2/n)\right)^n = e^{-\xi^2/2\sigma^2}. \tag{1.7.3}$$

The most classic form of this theorem is the *de Moivre-Laplace theorem*:
Let $\{X_i\}$ be iid random variables with $P(X_1 = 1) = P(X_1 = -1) = 1/2$. Then,

$$P(a \le S_n/\sqrt{n} \le b) \to \int_a^b (2\pi)^{-1/2} e^{-x^2/2} dx. \tag{1.7.4}$$

In this case the reader should check (1.7.1) explicitly. The result may be interpreted as the distribution of the position of a random walker along a line (→**3.1.8**).

1.7.2 Central limit theorem vs. large deviation

If you look at the large deviation theoretical result such as (1.6.7) and if you assume that the rate function I is differentiable, then $I(x) = ax^2/2$ is a good approximation.[80] Therefore, this is identical to the central limit theorem. However, this coincidence is not the rule.

In the framework of large deviation theory the deviation is fixed and then the large sample number limit $N \to \infty$ is taken. In contrast, in the central limit theorem the size of the deviation is not fixed but scaled with an appropriate power of N. Roughly speaking, the central limit theorem tells us the width of the 'main part' of the density distribution. On the other hand, the rate function tells us how much is outside the boundary you specify. Therefore, if the distribution is with a fat tail the both estimates need not compliment each other.

1.7.3 Necessary and sufficient condition for Gaussianness

If the variance is not finite (i.e., the fluctuation is too large), generally speaking, the above theorem does not hold (→**1.7.4**). However, even in that case a careful choice of the scaling factor b_n instead of $\sqrt{n}\sigma$ could in some cases still make the distribution of $(S_n - nm)/b_n$ converge to $N(0,1)$. The last word for this is the following theorem by Lévy:
Let $\{X_i\}$ be iid random variables. For the distribution of $(S_n - a_n)/b_n$ with appropriate choice of a_n and b_n to converge to $N(0,1)$, it is necessary and sufficient that

$$\lim_{a\to\infty} \frac{a^2 P(|X_1| > a)}{E(|X_1|; |X_1| \le a)} = 0. \tag{1.7.5}$$

That is, the tail of the distribution must not be too fat.

For critical fluctuations, and many other phenomena the tail is too fat, and Gaussian statistics does not hold.

1.7.4 Central limit theorem for distributions with fat tails

If the distribution has a fat tail (i.e., large fluctuations from the mean can occur rather frequently), then (1.7.5) would certainly be violated. Even in such cases the distribution of the sample mean S_n/b_n with an appropriate choice of b_n still converges to a certain limit distribution (we have assumed that the distribution has the mean 0), called a *stable distribution*. This is a generalization of the standard central limit theorem. See Feller.

---

[80] Actually, as we will see, almost all elementary theories of nonequilibrium statistical mechanics are understood within this approximation.

1.7.5 Central limit theorem and renormalization group theory

An important point is that if a sequence of iid random variables $\{X_i\}$ obey such a stable distribution (with mean 0), S_n/b_n is distributed according to the same stable distribution for any n. In other words, the stable distribution is a fixed point of the procedure $\{X_i\}_{i=1}^n \to S_n/b_n$.

If the reader is already familiar with the *renormalization group theory*, she should immediately see that this procedure is a renormalization transformation: constructing S_n by summing corresponds to coarse-graining (= making a block spin) and dividing with b_n corresponds to scaling.

This structure can easily be seen for the standard central limit theorem **1.7.1**: if X_i obeys a distribution with a finite variance, then by this renormalization procedure the distribution 'flows into' $N(0,1)$. That is, a distribution with a finite variance is in the attracting domain of $N(0,1)$.

Such a renormalization group structure was clearly recognized by mathematicians well before the advent of renormalization group theory. Renormalization group theory may be interpreted as a(n attempt for) vast generalization of central limit theorems to non-independent (often strongly correlated) random variables.[81]

We will look at the Wiener process from this point of view (→**X1.4.11**). Donsker's invariance principle (→**X1.4.12**) is the outcome.

Remark. Coarse-graining discards information, but in a certain sense, if the distribution is a fixed point of an RG, then the information lost by coarsening is exactly compensated by scaling. Therefore, if we appropriately define the entropy production rate of the renormalization-group dynamical system, then it is positive away from the fixed point but is zero at the fixed point. That is, information can be a Lyapunov function of the RG dynamics. This idea has been used to understand the central limit theorem.[82] □

---

[81] This view point was clearly recognized first by G Jona-Lasinio, "The renormalization group: a probabilistic view," Nuovo Cim. **268**, 99-119 (1975).

[82] E A Carlen and A Soffer, "Entropy production by block variable summation and central limit theorems," Commun. Math. Phys. **40**, 339-371 (1991).

Appendix: What is Measure?

In this Appendix the general theory of the Lebesgue measure is outlined. Without measure theory proper understanding of statistical mechanics and dynamical systems is impossible.

1.A.1 Reader's guide to this appendix.
(1) + (3) is the minimum of this appendix:
(1) The ordinary Lebesgue measure = volume is explained up to **1.A.9**. These entries should be very easy to digest. Remember that Archimedes reached this level of sophistication more than 2000 years ago.
(2) General Lebesgue measure is outlined in **1.A.10-1.A.12**. This is an abstract repetition of (1), so the essence should be already obvious.
(3) Lebesgue integral is defined in terms of the Lebesgue measure in **1.A.16** with a preparation in **1.A.15**. This leads us naturally to the concept of functional and path integrals (**1.A.17**).
(4) Probability is a measure with total mass 1 (i.e., normalized) (**1.A.20**).
(5) If we read any probability book, we encounter the triplet $(\Omega, \mathcal{B}, P)$. The reason why we need such a nonintuitive device is explained in **1.A.21-1.A.22**.

1.A.2 What is volume? — An example of conceptual analysis
For simplicity, we confine our discussion to 2-space, but our discussion can easily be extended to higher dimensional spaces. The question is: what is 'area' ? It is not easy to answer this question for an arbitrary shape.[83] Therefore, we should start with a seemingly obvious example.

The area of a rectangle $[0,a]\times[0,b]$ ($a\geq 0$, $b\geq 0$) in $\boldsymbol{R}^2$ is ab. Do we actually know this? Why can we say the area of the rectangle is ab without knowing what area is? To be logically conscientious we must accept:
Definition. The area of a rectangle which is congruent[84] to $\langle 0,a\rangle\times\langle 0,b\rangle$ (Here $\langle$ is [or (and $\rangle$ is] or)) is defined to be ab. Notice that area is defined so that it is not affected by whether the boundary is included or not.[85]

1.A.3 Area of fundamental set.
A set which is a direct sum (disjoint union) of finite number of rectangles is called a *fundamental set*. The area of a fundamental set is defined by the sum of the areas of constitutive rectangles.

It should be intuitively obvious that the join and the common set of fundamental sets are again fundamental.

1.A.4 Heuristic consideration.
For an arbitrary shape, the strategy for defining its area should be to approximate the figure with a sequence of fundamental sets. We should use the idea going back to Archimedes; we must approximate the figure from its inside and from its outside. If both sequences converge to the same area, we should define the area to be the area of the figure.

1.A.5 Outer measure, measure zero.
Let A be a set. We consider a cover of A with finite number of rectangles P_k (inclusion or exclusion of their boundaries can be chosen conveniently →**1.A.2**), and call it a rectangular cover $P=\{P_k\}$ of A. Let us denote the area of a rectangle P_k by $m(P_k)$. The *outer measure* $m^*(A)$ of A is defined by

$$m^*(A)\equiv \inf\sum_k m(P_k), \tag{1.7.1}$$

where the infimum is taken over all the finite or countable rectangular covers of A.

If $m^*(A)=0$, the set A is called a *measure zero set* (or a *null set*).

---

[83] As we will see soon in **1.A.22**, if we stick to our usual axiomatic system of mathematics ZF+C, there are figures without area.

[84] This word is defined by the superposability.

[85] Thus, topological concepts are much more detailed and difficult than measure-theoretical concepts in the theory of dynamical systems.

1.A.6 Inner measure.
For simplicity, let us assume that $A \subset E \equiv [0,1] \times [0,1]$. Then, the inner measure $m_*(A)$ of A is defined by

$$m_*(A) = 1 - m^*(E \setminus A). \tag{1.7.2}$$

Obviously,

$$m^*(A) \geq m_*(A) \tag{1.7.3}$$

for any figure A.

1.A.7 Measurable set, area = Lebesgue measure.
Let A be a subset of E.[86] If $m^*(A) = m_*(A)$, we say A is *measurable* (in the sense of Lebesgue), and $m^*(A)$ written as $\mu(A)$ is called its area (= *Lebesgue measure*).

1.A.8 Additivity.
Assume that all the sets here are in a bounded rectangle, say, E above. The join and the common set of finitely many measurable sets are again measurable. This is true even for countably many measurable sets. The second statement follows from the preceding statement thanks to the finiteness of the outer measure of the join or the common set.

1.A.9 σ-additivity.
Let $\{A_n\}$ be a family of measurable sets satisfying $A_n \cap A_m = \emptyset$ for $n \neq m$. Let $A = \cup_n A_n$. Then,

$$\mu(A) = \sum_n \mu(A_n). \tag{1.7.4}$$

This is called the σ-*additivity* of the Lebesgue measure. ⊡
[Demo] A is measurable due to **1.A.8**. Since $\{A_n\}$ covers A, $\mu(A) \leq \sum \mu(A_n)$. On the other hand $A \supset \cup_{n=1}^N A_n$, so that for any N $\mu(A) \geq \sum_{n=1}^N \mu(A_n)$.

1.A.10 Measure, general case.
A map from a family of sets to $\boldsymbol{R}$ is called a *set function.* A set function m satisfying the following three conditions is called a *measure.*
(1) m is defined on a semiring[87] $\mathcal{S}$. [Note that the set of all the rectangles is a semiring.]
(2) $m(A) \geq 0$.
(3) m is an additive function: If A is direct-sum-decomposed in terms of the elements of $\mathcal{S}$ as $A = \cup_{k=1}^n A_k$, then $m(A) = \sum_{k=1}^n m(A_k)$.

Therefore, the area μ defined in **1.A.7** is a measure on the set of all the rectangles. In the case of area, the definition of area is extended from rectangles to fundamental sets (→**1.A.3**). This is the next step:

1.A.11 Minimum algebra on $\mathcal{S}$, extension of measure.
The totality of sets A which is a finite join of the elements in $\mathcal{S}$ is called the *minimum algebra* generated by $\mathcal{S}$. Notice that the totality of fundamental sets in **1.A.3** is the minimum algebra of sets generated by the totality of rectangles. Just as the concept of area could be generalized to the area of a fundamental set, we can uniquely extend m defined on $\mathcal{S}$ to the measure defined on the algebra generated by $\mathcal{S}$.

1.A.12 Lebesgue extension.
We can repeat the procedure to define μ from m^* and m_* in **1.A.7** for any measure m on $\mathcal{S}$ (in an abstract fashion). We define m^* and m_* with the aid of the covers made of the elements in $\mathcal{S}$. If $m^*(A) = m_*(A)$,

---

[86] It should be obvious how to generalize our argument to a more general bounded set in $\boldsymbol{R}^2$.

[87] If a family of sets $\mathcal{S}$ satisfies the following conditions, it is called a *semiring* of sets:
(i) $\mathcal{S}$ contains $\emptyset$,
(ii) If $A, B \in \mathcal{S}$, then $A \cap B$ and $A \cup B$ are in $\mathcal{S}$,
(iii) if A_1 and A are in $\mathcal{S}$ and $A_1 \subset A$, then $A \setminus A_1$ can be written as a direct sum (the join of disjoint sets) of elements in $\mathcal{S}$.

we define the *Lebesgue extension* μ of m with $\mu(A) = m^*(A)$, and we say A is *μ-measurable.*

1.A.13 Remark.

When we simply say the Lebesgue measure, we usually mean the volume (or area) defined as in **1.A.7**. However, there is a different usage of the word. μ constructed in **1.A.12** is also called a Lebesgue measure. That is, a measure constructed by the Lebesgue extension is generally called a Lebesgue measure. This concept includes the much narrower usage common to physicists.

1.A.14 σ-additivity.

(3) in **1.A.10** is often replaced by the following *σ-additivity* condition: Let A be a sum of countably many disjoint μ-measurable sets $A = \cup_{n=1}^{\infty} A_n$. If

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n), \tag{1.7.5}$$

we say μ is a σ-additive measure.

The Lebesgue measure defined in **1.A.7** is σ-additive. Actually, if m is σ-additive on a semiring of sets, its Lebesgue extension is also σ-additive.

1.A.15 Measurable function.

A real function defined on a set D is called a *μ-measurable function* for a Lebesgue measure μ on the set, if any 'level set' $\{x \,|\, f(x) \in [a, b]\} \cap D$ is μ-measurable. When we simply say a function is measurable, it means that any level set has a well defined volume in the ordinary sense.

1.A.16 Lebesgue integral with measure μ.

Let μ be the Lebesgue measure on $\boldsymbol{R}^n$. Then the Lebesgue integral of a μ-measurable function on $U \subset \boldsymbol{R}^n$ is defined as

$$\int_U f(x) d\mu(x) = \lim_{\epsilon \to 0} \sum a\, \mu(\{x \,|\, f(x) \in [a - \epsilon/2, a + \epsilon/2)\} \cap U), \tag{1.7.6}$$

where the sum is over all the disjoint level sets of 'thickness' ϵ (> 0).[88]

1.A.17 Functional integral.

As the reader has seen in **1.A.16**, if we can define a measure on a set, we can define an integral over the set. The set need not be an ordinary finite-dimensional set, but can be a function space. In this case the integral is called a *functional integral.* If the set is the totality of paths from time $t = 0$ to T, that is, if the set is the totality of continuous functions: $[0, T] \to \boldsymbol{R}^d$, we call the integral over the set a *path integral.* The Feynman-Kac path integral is an example (→**X1.10.2**).[89]

1.A.18 Uniform measure.

The Lebesgue measure defined in **1.A.17** is uniform in the sense that the volume of a set does not depend on its absolute location in the space. That is, the measure is translationally invariant (see **1.A.21** below for a further comment). However, there is no useful uniform measure in infinite dimensional spaces. Thus, every measure on a function space or path space must be non-uniform (cf., **X1.10.5**).

1.A.19 Borel measure.

Usually, we mean by a *Borel measure* a measure which makes measurable all the elements of the smallest algebra (→**1.A.11**) of sets containing all the rectangles.

---

[88] The measures m satisfying $\mu(A) = 0 \Rightarrow m(A) = 0$, where μ is a Lebesgue measure, is said to be *absolutely continuous* with respect to μ. If a measure is absolutely continuous with respect to the Lebesgue measure (volume), it has a density as an ordinary function.

[89] However, the definition of the Feynman path integral is too delicate to be discussed in the proper integration theory.

1.A.20 Probability space.

A (Lebesgue) measure P with the total mass 1 is called a *probability measure*. To compute the expectation value with respect to P is to compute the Lebesgue integral w.r.t. the measure P.

When we read mathematical probability books, we always encounter the 'triplet' $(\Omega, \mathcal{B}, P)$ (= a *probability space*), where P is a probability measure, Ω is the totality of elementary events (the event space; $P(\Omega) = 1$) and $\mathcal{B}$ is the algebra of measurable events.

Discussion 1. If the base space is not bounded, sometimes the choice of $\mathcal{B}$ seems to matter. Read and summarize T P Hill, *The Significant-digit Phenomenon*, Am. Math. Month. April 1995, p322. □

This specification is needed, because if we assume that every composite event has a probability, we have paradoxes. This was realized through the characterization of 'uniform measure' in a finite dimensional Euclidean space:

1.A.21 Lebesgue's measure problem.

Consider d-Euclidean space $\boldsymbol{R}^d$. Is it possible to define a set function (→**1.A.10**) m defined on every bounded set $A \in \boldsymbol{R}^d$ such that

(1) The d-unit cube has value 1.

(2) Congruent sets have the same value,

(3) $m(A \cup B) = m(A) + m(B)$ if $A \cap B = \emptyset$, and

(4) σ-additive

?

This is called *Lebesgue's measure problem.*

1.A.22 Hausdorff and non-measurable set.

Hausdorff demonstrated in 1914[90] for any d there is no such m satisfying (1)-(4) of **1.A.21**. Then, Hausdorff asked what if we drop the condition (4). He showed that m does not exist for $d \geq 3$.[91] He showed this by constructing a partition of a 2-sphere into sets A, B, C, D such that A, B, C and $B \cup C$ are all congruent and D is countable. Thus, if m existed, we had to conclude $1 = 2$. Therefore, we must admit non-measurable sets.[92]

---

[90] WWI began, Panama canal officially opened.

[91] Banach demonstrated in 1923 that there is a solution for $d = 1$ and for $d = 2$.

[92] under the current popular axiomatic system ZF + C. The reader might think such a set is far from any field of physics, but a popular characterization of chaos (Li -Yorke's theorem) in dynamical systems uses sets that are not measurable, if it is not measure zero (Baba-Kubo-Takahashi).

Chapter 2

Equilibrium Statistical Mechanics Revisited

In this second introductory chapter, we wish to 'dust' our equilibrium statistical thermodynamics. It is not interesting to do so by going through the standard textbooks, so let us try to trace the birth of statistical mechanics. Thus, the main purpose of this chapter is to give a (hopefully) readable story of statistical thermodynamics. The reader will clearly see that it was born of kinetic theory of gases = study of nonequilibrium processes. Elementary kinetic approaches are covered in this chapter. Also she will recognize that statistical mechanics cannot be justified without thermodynamics.

For those who still feel that they need a standard statistical mechanics course, the lecturer recommends them to study

R Kubo, *Statistical Mechanics, an advanced course with problems and solutions* (North Holland, 1965)

and its sister volume:

R Kubo, *Thermodynamics, an advanced course with problems and solutions* (North Holland, 1968).[1]

Solve all the problems at least up to level B (these books are problem books with *full detailed solutions*; even 'passively' reading the solutions is useful, but a more active learning is recommended). No course of elementary equilibrium statistical mechanics is needed, if the reader studies these books seriously.

The sister lecture notes, *Introductory Equilibrium Statistical Mechanics*, are available from http://www.rinst.org. The level of these notes is around Kubo quoted above. A rather extensive set of problems with complete solutions can be found.

---

[1] Actually, the first half of the original book in Japanese, *Thermodynamics and Statistical Mechanics* (Shokabo, 1951), is "Thermodynamics" and the second half "Statistical Mechanics." It is designed for undergraduate students and not at all advanced.

2.1 Gibbs's Statistical Mechanics

We are all told that equilibrium statistical mechanics we know now was established by J. W. Gibbs (1839-1903).[2] Using the framework, we can compute the Helmholtz free energy A as

$$A = -k_B T \ln Z, \tag{2.1.1}$$

where k_B is the Boltzmann constant, T is the absolute temperature, and Z is the (canonical) partition function defined by

$$Z = \sum_{\text{microstates}} e^{-\beta H} \tag{2.1.2}$$

with $\beta \equiv 1/k_B T$, and H being the Hamiltonian (= energy function) of the system.

To understand the nature and the scope of statistical mechanics, let us browse through Gibbs's famous book.[3]

2.1.1 Gibbs, "Elementary Principles in Statistical Mechanics" (1902[4])[5]

The problem Gibbs wished to solve was: "not to follow a particular system through its succession of configurations, but to determine how the whole number of systems in a given ensemble will be distributed among the various conceivable configurations and velocities at any required time, when the distribution has been given for some one time." (Gibbs pvii).

However, "in the present state of science, it seems hardly possible to frame a dynamic theory of molecular action which shall embrace the phenomena of thermodynamics, of radiation and of the electrical manifestations which accompany the union of atoms." This difficulty "deterred the author from attempting to explain of mystery of nature, and have forced him to be contented with the more modest aims of deducing some of the more obvious branch of mechanics."(px).

The outline of his book[6] is as follows: To begin with Gibbs explains the concept of ensemble: a collection of clones of a system satisfying some common conditions (say, identical energy).[7] He then considers "especially ensembles of systems in which the logarithm of

---

[2] M J Klein, "The physics of J. Willard Gibbs in his time," Physics Today, September 1990, p40-48; This is based on the first paper by the same author in *Proceedings of the Gibbs Symposium, Yale University, May 15-17, 1989* (edited by D G Caldi and G D Mostow) (AMS & APS, 1990).

[3] A much more careful summary may be found in the already cited Gibbs Symposium (1989) Proceedings: A S Wightman, "On the Prescience of J. Willard Gibbs," p23-38.

[4] What sort of year was it? In these notes, to have a glimpse of the cultural background or the Zeitgeist, the contemporary historical events are mentioned in the footnote. In 1902, Britain and Japan concluded a naval alliance (the Anglo-Japanese Alliance (mainly against Russia)); The Second Boer War ended; Ibn Saud recovered Riyadh; V. Pareto's *The Socialist Systems*, a refutation of the Marxist economics. As we will see Einstein should have been completing his system of statistical mechanics.

[5] M J Klein loc cit writes, "It is not impossible that the pressure to contribute to the Yale Bicentenial series propelled Gibbs into completing a book that he might otherwise have considered not quite ready to publish." According to Klein, Gibbs asked himself questions like, "What is the entropy of a system not in equilibrium?" at several places in his notes.

[6] Its full title read, *Elementary Principles in Statistical Mechanics developed with especial reference to the Rational Foundation of Thermodynamics* (Yale UP 1902, reprinted from Dover).

[7] In the standard formulation of statistical mechanics, 'ensemble' is only a device to visualize probability distribution (→**2.4.1**). In this interpretation, do not imagine we actually have that many copies of macroscopic systems. Those who feel this approach unphysical, go to **1.6.14**.

probability of phase is a linear function of energy. This distribution, on account of its unique importance in the theory of statistical equilibrium, I have ventured to call *canonical*."(pxi). From this he finds "a differential equation relating to average values in the ensemble which is identical in form with the fundamental differential equations of thermodynamics, the average logarithm of probability of phase, with change of sign, corresponding to entropy, and the modulus to temperature" (→**2.1.4**). These are the contents of his book up to Chapter XIII.[8] He discusses microcanonical ensemble and its equivalence to the canonical ensemble in the thermodynamic limit in Chapter XIV. The last chapter deals with multicomponent systems.

Preface tells us that Gibbs regards statistical mechanics as exact, but thermodynamics as the (approximate and empirical) laws of mechanics of systems of a great number of particles. This point of view is diametrically different from that of the lecturer's, but let us follow his logic for the time being.

2.1.2 Gibbs introduces canonical ensemble

Let P be the probability density on the phase space defining a given ensemble.[9] Then, P for an equilibrium ensemble must satisfy (→**2.7.7**)

$$[H, P]_{PB} = 0, \tag{2.1.3}$$

where H is the system Hamiltonian, and $[\ ,\ ]_{PB}$ is the Poisson bracket (→**2.7.1**). If P is a function of energy (i.e., H), this is automatically satisfied. Gibbs introduces the canonical distribution:

$$P = e^{(\psi - H)/\Theta}, \tag{2.1.4}$$

where ψ is determined by normalization (in the original book ϵ is used for energy):

$$e^{-\psi/\Theta} = \int d\Gamma e^{-H/\Theta} \tag{2.1.5}$$

with $d\Gamma$ being the volume element of the phase space. Θ is a positive numerical constant.

According to Gibbs this is "the most simple case conceivable since it has the property that when the system consists of parts with separate energies, the laws of the distribution in phase of the separate parts are of the same nature."(p33). He notes, "Θ has properties analogous to those of temperature in thermodynamics."(p35); his reasoning is based on the compound system: combining two systems at the same temperature to make a single system; if we can ignore the interaction energy between system 1 and 2 (because both are macroscopic and the surface contribution should be negligible), the canonical distribution of the compound system reads $e^{(\psi-H)/\Theta} = e^{(\psi_1-H_1)/\Theta}e^{(\psi_2-H_2)/\Theta}$ with $\psi = \psi_1 + \psi_2$ and $H = H_1 + H_2$.

2.1.3 Computing differential of ψ

Let a_i be "certain coordinates of bodies which we call *external*, meaning by this simply that

---

[8] Chapter XII is about the approach to equilibrium. There, he points out that spreading of the nearby trajectories is needed for the approach to equilibrium (p150).

[9] We must clearly recognize that probability theory was not at all popular in those days. Hilbert's problem number 6 (1900) was meant to be an axiomatization of probability theory with special reference to kinetic theory (Wightman, loc cit). In modern terms, we consider a probability space (Γ, P), where Γ is the phase space of the system. See **1.1.5**.

they are not to be regarded as forming any part of the system, although their positions affect the forces which act on the system." Then, from (2.1.5) we get

$$-d\psi + \frac{\psi}{\Theta}d\Theta = \frac{1}{\Theta}d\Theta \int d\Gamma H e^{(\psi-H)/\Theta} + \sum_i da_i \int d\Gamma A_i e^{(\psi-H)/\Theta}, \tag{2.1.6}$$

where

$$A_i \equiv \frac{\partial H}{\partial a_i}. \tag{2.1.7}$$

That is,

$$d\psi = \frac{\psi - \langle H \rangle}{\Theta} d\Theta - \sum_I \langle A_i \rangle da_i, \tag{2.1.8}$$

where $\langle \ \rangle$ implies the ensemble average.

If we introduce

$$\eta = \log P = \frac{\psi - H}{\Theta}, \tag{2.1.9}$$

(2.1.8) may be rewritten as

$$d\psi = \langle \eta \rangle d\Theta - \sum_i \langle A_i \rangle da_i. \tag{2.1.10}$$

2.1.4 Thermodynamics analogue

From (2.1.10), we have $d\psi = d\langle H \rangle + \Theta d\langle \eta \rangle + \langle \eta \rangle d\Theta$, so

$$d\langle H \rangle = -\Theta d\langle \eta \rangle - \sum_i \langle A_i \rangle da_i. \tag{2.1.11}$$

"This equation, if we neglect the sign of averages, is identical in form with the thermodynamic equation"(Gibbs p44)

$$dE = TdS + \sum_i X_i dx_i, \tag{2.1.12}$$

where E is the internal energy, T the absolute temperature, S the entropy, etc. "We may also compare (2.1.10) with the thermodynamic equation,"

$$dA = -SdT + \sum_i X_i dx_i. \tag{2.1.13}$$

"How far, or in what sense, the similarity of these equations constitutes any demonstration of the thermodynamic equations, or accounts for the behavior of material systems, as described in the theorems of thermodynamics"(Gibbs p45) is the crucial question.

2.1.5 Gibbs' entropy formula and thermodynamics-mechanics correspondence

The above analogy suggests the entropy formula (cf., **1.3.2**):

$$S \text{ `='} \langle \eta \rangle = -\int d\Gamma P \log P. \tag{2.1.14}$$

Gibbs demonstrates that under the constant Θ condition the canonical distribution minimizes $\langle \eta \rangle + \langle H \rangle/\Theta$ (Gibbs p161). "We have thus precisely defined quantities, and rigorously demonstrated propositions, which hold for any number of degrees of freedom, and which, when the number of degrees of freedom is enormously great, would appear to human faculties as the quantities and propositions of empirical thermodynamics." (Gibbs p169).

This is Gibbs' 'justification' of $-\psi/\Theta = \log Z$, or in the modern notation

$$A = -k_B T \log Z \tag{2.1.15}$$

with

$$Z = \int d\Gamma e^{-\beta H}. \tag{2.1.16}$$

Exercise 1. Demonstrate that $\langle \eta \rangle + \langle H \rangle/\Theta$ is minimum for the canonical distribution. □

2.1.6 Summary of Gibbs' logic

As the reader has realized, Gibbs' logic is based on
(1) correspondence 'in form' and
(2) the demonstration that general thermodynamic properties are satisfied by the analogue quantities.

However, is this really satisfactory?[10] Gibbs claims that he demonstrated the perfect parallelism between his statistical mechanical system and thermodynamics. However, for example, there is no counterpart of heat in his framework, so even the famous equilibrium relation $dS = dQ/T$ has no statistical mechanical interpretation (also see **2.7.14**).[11]

Gibbs' canonical ensemble may be a solution to his problem of finding microscopic expressions for thermodynamic quantities, but how unique is the solution? Even Gibbs knew that microcanonical ensemble would do as well (→**2.6.8**).

There are a lot of conceptual holes in Gibbs' statistical mechanics, and we inherit many of them.

2.1.7 What made Gibbs conceive his equilibrium statistical mechanics?

To see this we must go back almost to the starting point of the modern natural science. We will eventually realize how modern Einstein is. Learning statistical mechanics, the reader will recognize that we have hardly progressed since Einstein with respect to fundamental issues.

2.2 Prehistory of Statistical Mechanics

To understand the prehistory of equilibrium statistical mechanics, we must go almost all the way back to the birth of the modern science. We will see how equilibrium statistical

---

[10] L Navarro, "Gibbs, Einstein and the Foundations of Statistical Mechanics," Arch. Hist. Exact Sci. **53**, 147-180 (1998) [kindly informed by R Clegg] criticizes Gibbs almost bitterly, saying "This foundation is however confined to the establishment of analogies between two different frameworks (mechanical and thermodynamics), and therefore its reductionist value is dubious." The lecturer is fully sympathetic to this criticism, but see the last footnote in **2.5.2**; he is also very sympathetic to Gibbs.

[11] It is worth comparing this with what Einstein did. →**2.5.3-**.

mechanics originated from the struggle to understand the second law. However, the success of equilibrium statistical mechanics left many questions about nonequilibrium phenomena unanswered.

Recommended reading:

S G Brush, *Statistical Physics and the Atomic Theory of Matter, from Boyle and Newton to Landau and Onsager* (Princeton UP, 1983) Chapter 1.[12]

D Lindley, *Boltzmann's Atom, The great debate that launched a revolution in physics* (The Free Press, New York, 2001).

The second book may be more readable and sometimes more detailed than Brush's (perhaps with too many gossips).[13]

2.2.1 Everything started with Descartes and Newton

One of the main goals of *Principia* was to demonstrate the inadequacy of Descartes' theory of the universe. 'Therefore', Newton adopted Boyle's conception of air pressure due to the springiness of air particles instead of Descartes' swirling atoms. This lead to a static model of gases in which the pressure was attributed to interatomic repulsive forces and hampered the development of the kinetic theory (Brush p22).

On the other hand Newton seems to have accepted the atomism in the Greek sense also: "the Changes of corporeal Things are to be placed only in the various Separations and new Associations, and Motions of these permanent Particles $\cdots$" (*Optics* quoted in Brush p23). However, the static theory dominated the scene. Even Laplace was under the strong influence of static repulsion theory of pressure.[14]

2.2.2 'Modern kinetic theory' of D. Bernoulli

Daniel Bernoulli's gas model[15] (1738[16]) is the first fully kinetic model: there are indefinitely large number of similar particles of diameter d and speed v moving in a closed container of volume V supporting a piston with pressure p by their impacts. The average distance between particles is D, and the temperature T is a function of v.

His conclusions are (Brush p30; as the reader can see these are quite modern →**2.3.5**):
(1) Boyle's law $pV =$ const. in the $d/D \to 0$ limit.
(2) The deviation from Boyle's law is given as a function of d/D.
(3) If v is varied at constant volume, $p \propto v^2$.
(4) A temperature scale may be defined by using $pV = CT$, where C is a positive constant, so $T \propto v^2$.

It is noteworthy that Bernoulli did not give his particles any properties other than size; in particular no mass is specified.

---

[12] In Japanese, T Yamamoto, *Historical Development of Thoughts about Heat* (Gendai Sugakusha, 1987) is an excellent book (more detailed in certain respects than Brush).

[13] The reader will see why Maxwell was not so impressed by Boltzmann's kinetic theory; why Thomson and Planck disliked kinetic theory; what the real issue was of the difference between Mach and Boltzmann, etc.

[14] Laplace got interested in the properties of gases due to the necessity of accurate measurement of the positions of stars through the atmosphere (Yamamoto p189). Gay-Lussac was his chief experimentalist partner after Revolution. Note that Gay-Lussac was one of the first generation of scientists who were educated by systematic science education and who earned their living doing science.

[15] This was in his book on hydrodynamics.

[16] This year J S Bach composed Mass in B minor (BWV 232).

However, the kinetic theory of the 18th century was totally forgotten by the 19th century (→**2.2.4**).

2.2.3 Modern atomism and molecular theory

Between Daniel Bernoulli and the modern kinetic theory was the general acceptance of chemical atomic theory and the birth of physics in the *modern sense*. The most important result is Dalton(1766-1844)'s *law of partial pressure*: the total pressure of a gas mixture is simply the sum of the pressures each kind of gas would exert if it were occupying the space by itself. This is naturally understood with the aid of kinetic theory.[17]

Gay-Lussac (1778-1850) then established three important laws (ca 1810[18]):
(i) the law of thermal expansion of gases (also called Charles' law).
(ii) the law of constant temperature under adiabatic expansion: if a gas is suddenly allowed to occupy a much larger space by displacing a piston, there is practically no temperature change.
(iii) the law of combining volume: in gas phase reactions the volumes of reactants and products are related to each other by simple rational ratios.
(iii) implied that 'particles' cannot generally be atoms, but Dalton rejected this interpretation (suggesting, e.g., that Gay-Lussac results were not accurate, etc.)

At last in 1811[19] Avogadro proposed *Avogadro's hypothesis*: every gas contains the same number of *molecules* at the same pressure, volume, and temperature.

However, the molecular theory was not generally accepted until 1860, when Cannizzaro presented it to the Karlsruhe Congress (However, Clausius (→**2.2.5**) accepted this by 1850 (Brush p51); actually, Cannizzaro mentioned Clausius.[20]).

2.2.4 What is heat?

To arrive at the modern kinetic theory, there was one more obstacle: the *caloric theory of heat*. Even though the experiments of Rumford (1753-1814) and Davy were well known in the 1820s, no one considered them to be crucial; the heat generated by friction could be ascribed to caloric squeezed out of the bodies.

Rumford stressed that heat was not conserved: in his memoir of 1804[21] he said that he only found that heat source was inexhaustible. Thus, heat cannot be matter, because the latter is conserved. In contrast, the emphasis of the caloric theory was on the conservation law; this was reinforced by the theory of latent heat: whenever the conservation law looked violated, they generalized the concept of latent heat, and the theory became more versatile. Only after Joule(1818-1889)'s work in 1840[22] were Rumford's experiments correctly interpreted (Yamamoto p241).

The caloric theory was replaced in the 1830s by the wave theory of heat. The identity of heat and light favored the same theory for both. Its definite formulation was by Ampère

---

[17] Dalton arrived at his atomic theory not very inductively as is stressed by Brush on p32; Dalton's writings are sometimes hard to comprehend due to arbitrary thoughts and their outcomes being nebulously mixed up with real experimental results (Yamamoto p194).

[18] Napoleon married Marie Louise of Austria.

[19] This year J Austine published *Sense and Sensibility.*

[20] C. Cercignani, "The rise of statistical mechanics," in *Chance in Physics*, Lect. Notes Phys. 574 (edited by J. Bricmont, D. Dürr, M. C. Galavotti, G. C. Ghirardi, F. Petruccione and N. Zanghi) p25 (2001). This article gives a good summary of Boltzmann's progress.

[21] This year Code Napoléon established the principle of equal citizenship; Haiti gained independence.

[22] Opium Wars; Feuerbach's *The essence of Christianity* would be published next year.

(1832,[23] 1835[24]): heat flows between vibrating atoms by means of ether vibration. This theory provided a fairly smooth transition from the caloric theory of the early 1800s to the kinetic theory of the 1850s.[25]

In the 1840s the first law of thermodynamics was established by Mayer (1814-1878),[26] Joule, Helmholtz (1921-1894), and others. Clausius began developing its application to interconversion of heat and mechanical work (1850[27] and on).

2.2.5 Clausius, "The kind of motion we call heat" (1857[28])

Since Clausius (1822-88) had been working on interconversion of mechanical work and heat, it was very natural that he wished to know the 'real picture' : He stated in this paper that he had been thinking about molecular motion even before his first article on thermodynamics in 1850.[29] He gave a clear molecular picture of the three phases of matter, and explained latent heat, evaporation due to fluctuation, etc.

Exercise 1. The reader must have a clear molecular picture of the so-called three phases of matter. (Sample questions: what is the major distinction between gas and liquid?; what is the difference between the molecular pictures of volume expansion upon heating in solids and liquids? [Hint. Pay attention to the coordination number.]) □

He assumed that particles may contain more than one atoms (before Cannizzaro →**2.2.3**) and that molecules were extremely small: "the space actually filled by the molecules of the gas must be infinitesimal in comparison to the whole space occupied by the gas itself." He assumed also that molecular forces could be ignored, and derived the equation of state of the ideal gas (→**2.3.5**).

2.2.6 Mean free path

At the end of the 1857 paper Clausius calculated the speed of molecules at 0°C: oxygen 461m/s, nitrogen 492m/s, and hydrogen 1,844m/s.

Exercise 1. Are these numbers accurate? □

Dutch meteorologist C. H. D. Buys-Ballot noticed that if the molecules of gases really moved that fast, the mixing of gases by diffusion should have been much faster than we observed it to be.

Upon this criticism, Clausius (1858[30]) had to assume that the gas molecules have large

---

[23] This year Faraday established the laws of electrolysis; Goethe died.

[24] The Boers began the Great Treck; the second Seminor War.

[25] However, J. Herapath (1790-1868) was an exception in the heyday of caloric theory, proposing kinetic theory of gases; he arrived at his theory through realizing the intrinsic difficulty of caloric theory. But for him (since he did not know Bernoulli →**2.2.2**) Newton's springy particle theory was an obstacle (Yamamoto p245). His paper was rejected by the Royal Society in 1820 (Davy rejected it, saying that it was too speculative), but appeared elsewhere in 1821. His theory almost completely covered the elementary kinetic theory (→**1.3**), but it took 30 years for the paper to be recognized properly. However, he was much more fortunate than Waterston (1811-83) whose 1845 paper was also rejected by the Royal Society and was not published for half a century. The reader can find more about Waterston in Lindley's book.

[26] However, it is noteworthy that Mayer explicitly rejected to regard heat as a kind of motion (Yamamoto p240).

[27] Telegraph cable connection across the English Channel.

[28] The Indian Mutiny by the Sepoy troops; Universities were established in Calcutta, Bombay and Madras; C. Baudraire, *Les Fleurs du Mal.*)

[29] This is the paper establishing the second law with the concept of entropy.

[30] This year, the Lincoln-Douglas debate, the Government of India Act. However, the most important event was that the idea of natural selection was officially published by Darwin and Wallace. Physicists

enough diameters that a molecule cannot move very far without colliding with another one. In this way Clausius defined a new parameter called the *mean free path* ℓ of gas. We can obtain it by dimensional analysis as

$$\ell = cV/Nd^2, \tag{2.2.1}$$

where V is the volume, N is the number of molecules in the volume, and c is a numerical constant (whose precise value was a matter of dispute for some time →**2.3.6**).
Exercise 1. Obtain the above formula. □

Clausius did not have any independent method to estimate N and d. Then came Maxwell, who showed that ℓ could be related to transport properties of gases, and so could be determined experimentally (→**2.3.6**).

2.3 Maxwell and Introduction of Statistics

When Maxwell was 19 years old, he read an article introducing the continental statistical theory into British science (e.g., Gauss's theory, cf. **2.3.2**, **1.7.1**), and was really fascinated. He wrote to his friend: "the true logic for this world is the Calculus of Probabilities ···" (Brush p59)

For the elementary theory of gases

J Jeans, *An Introduction to The Kinetic Theory of Gases* (Cambridge UP 1952)
may be of some use.

2.3.1 J. C. Maxwell, "Illustrations of the dynamical theory of gases" (1860[31])

First, Maxwell (1831-79) demonstrates[32] that the velocity of the gas particle obeys the *Maxwell distribution* (→**2.3.2**). Based on this distribution, he gave the formula for the mean free path (→**2.3.6**):

$$\ell = \frac{1}{\sqrt{2}\pi n d^2}. \tag{2.3.1}$$

Also he derived the equation of state of ideal gas (→**2.3.5**) and gave an explanation of Avogadro's law.

Next, he computed the shear viscosity of gases (→**2.3.7**). Comparing his result with Stokes' experimental result, he concluded that the mean free path was 1/447000 inch and a particle collides with other particles 8,077,200,000 times in a second. He demonstrated that the viscosity was independent of density. He commented that this was remarkable, and that the existing experiments did not confirm this. Actually, his letter to Stokes (May, 1859) reads, "This is certainly very unexpected, that the friction should be as great in a rare as in a

---

should recognize that Boltzmann called the 19th century the century of Darwin (not of Maxwell) (see E Broda, *Ludwig Boltzmann, Mensch·Physiker·Philosoph* (F Deuticke, 1955) Part III). The actual chronology is as follows: In June 1858 Darwin received a thin well-wrapped package from Ternate sent by Wallace. The papers by Darwin and Wallace on natural selection were read at the Linnean Society on July 1, 1858. For more about this meeting see the footnote for Brown in **3.1.1**.

[31] This year, Lincoln was elected president; the British raided Beijing.

[32] The incomplete nature of this derivation was later remedied by himself in 1866 (→**2.3.3**).

dense gas. The reason is, that in the rare gas the mean path is greater, so that the frictional action extends to greater distances. Has the reader the means of refuting this result of the hypothesis?" Did Maxwell believe in the kinetic theory at that time?

It is a good occasion to learn (or review) elementary kinetic theory of gases. Let us see more details in **2.3.7** and on, but in these lecture notes no stress will be put on the topic.

2.3.2 Maxwell's velocity distribution

In his "Illustrations of the dynamical theory of gases" (1860) Maxwell introduced the density distribution function $f(\boldsymbol{v})$ of the velocity of gas particles.

Maxwell assumed that the orthogonal components of the velocity are statistically independent (→**1.2.2**). This implies that we may write

$$f(\boldsymbol{v}) = \phi_x(v_x)\phi_y(v_y)\phi_z(v_z), \tag{2.3.2}$$

where ϕ_x, etc., are density distribution functions for individual components. Maxwell also assumed the isotropy, so f is a function of $v^2 \equiv |\boldsymbol{v}|^2$, $f(\boldsymbol{v}) \equiv F(v^2)$ and ϕ_x, etc., do not depend on the suffices specifying the coordinates: $\psi(s^2) \equiv \phi_x(s) = \cdots$. Therefore,

$$F(x+y+z) = \psi(x)\psi(y)\psi(z). \tag{2.3.3}$$

If we assume that F and ψ are once differentiable, we obtain

$$F'(x+y+z) = \psi(x)\psi(y)\psi'(z). \tag{2.3.4}$$

Setting $y = z = 0$, we obtain

$$F(x) = \psi(x)\psi(0)\psi(0), \;\; F'(x) = \psi(x)\psi(0)\psi'(0). \tag{2.3.5}$$

Therefore, $F'(x)/F(x)$ must be a constant. This implies a Gaussian distribution[33]

$$f(\boldsymbol{v}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^3 e^{-\boldsymbol{v}^2/2\sigma^2}. \tag{2.3.6}$$

Discussion 1. The reader may have wondered whether there is no other solution, if we drop the differentiability assumption. Try to demonstrate that the Gaussian form is the unique outcome of (2.3.3), assuming only the continuity of F.[34]□

Discussion 2. We know from the equilibrium statistical mechanics that the distribution function can be obtained as $e^{-\beta E}$, where E is the energy of the particle. This was obtained from the consistency with thermodynamics (→**2.1.5**). In the above derivation of the Maxwell distribution, we have used only independence and isotropy, and no mechanics at all. Therefore, these two results seem to suggest that thermodynamics dictates the form of kinetic energy. Is this really true?[35] □

Maxwell did not like the independence assumption, so he rederived his distribution a few years later (in 1866[36]). The argument relies on detailed balance (→**2.7.6**), reminiscent of Boltzmann's argument (→**2.4.9**).

---

[33] The easy way to compute the normalization is: $\left[\int dx\, e^{-x^2/2\sigma^2}\right]^2 = 2\pi \int_0^\infty e^{-r^2/2\sigma^2} r dr = 2\pi\sigma^2$.

[34] Certainly, if we do not assume that F is continuous, there are uncountably many solutions.

[35] Considering the relativistic case may clarify the situation. Although most modern textbooks do not cast any doubt about independence, Maxwell's doubt about it indicates his correct intuition.

[36] This year Boltzmann published his first attempt to demonstrate the second law from mechanics. This is one year after Clausius introduced the concept of entropy,

2.3.3 Maxwell's rederivation of Maxwell's distribution

Suppose a binary collision between particle 1 with velocity $\boldsymbol{v}_1$ and particle 2 with $\boldsymbol{v}_2$ produces particle 1 with velocity $\boldsymbol{v}'_1$ and particle 2 with velocity $\boldsymbol{v}'_2$. Then, the *detailed balance* (→**2.3.4**) requires[37]

$$f(\boldsymbol{v}_1)f(\boldsymbol{v}_2) = f(\boldsymbol{v}'_1)f(\boldsymbol{v}'_2) \tag{2.3.7}$$

for any choice of velocities compatible with the collision. Energy conservation implies

$$\boldsymbol{v}_1^2 + \boldsymbol{v}_2^2 = \boldsymbol{v}'^2_1 + \boldsymbol{v}'^2_2. \tag{2.3.8}$$

Therefore, if we assume the isotropy and the continuity of the density distributions, these relations imply that $\log f$ is a linear function of $\boldsymbol{v}^2$.[38]

2.3.4 Cyclic balance vs. detailed balance

In equilibrium, we do not observe any macroscopic change (time dependence). For this to be possible, for example, *cyclic balance* would do: suppose we have three microscopic states, a, b, and c; if the positive cycle a→b→c→a→b→c→··· is equally likely as the negative cycle c→b→a→c→b→a→···, we would not see any macroscopic change, even if a→b is not as likely as b→a. However, in equilibrium, there is a *detailed balance* (→**2.7.6**): a→b and b→a are equally likely. That is, a microscopic process and its (time) reversed process occur equally likely.

We will learn that detailed balance is crucial in understanding near equilibrium irreversible processes (→**4.1.1**).

2.3.5 Pressure of ideal gas - Bernoulli's formula

The (kinetic interpretation of) pressure on the wall is the average momentum given to the wall per unit time and area. Consider the wall perpendicular to the x-axis. Then, the number of particles with its x-component of the velocity being v_x that can collide the unit area on the wall per unit time is given by $nv_xf(\boldsymbol{v})$,[39]where n is the number density of the gas molecules. Each particle gives the momentum $2mv_x$ upon collision to the wall, so

$$p = \int_{v_x \geq 0} d\boldsymbol{v}\, 2mnv_x^2 f(\boldsymbol{v}) = \frac{1}{3}mn\langle \boldsymbol{v}^2 \rangle. \tag{2.3.9}$$

Or,

$$pV = \frac{2}{3}N\langle K \rangle, \tag{2.3.10}$$

where K is the kinetic energy of the single gas particle, and N is the number of particles in the volume V. This equation is called *Bernoulli's equation* (→**2.2.2**).

Comparing this with the equation of state of an ideal gas $pV = Nk_BT$, we obtain the well-known relation:

$$\langle \boldsymbol{v}^2 \rangle = 3k_BT/m. \tag{2.3.11}$$

---

[37] If the reader wishes to be more elementary, the phase volume element for the two particles must be written. The reader can rely on Liouville's theorem (for the two particle system under consideration) (→**2.7.7**) to ignore these volume elements.

[38] Proving this is not so trivial as is quoted in **2.4.8**.

[39] Precisely speaking, we must speak about the number of particles in the (velocity) volume element $d^3\boldsymbol{v}$, that is, $nv_xf(\boldsymbol{v})d^3\boldsymbol{v}$, but throughout these notes, such obvious details may be omitted.

If we use (2.3.6) to compute the average, we can fix the parameter σ as $\sigma^2 = k_BT/m$, and we obtain the *Maxwell distribution*:[40]

$$f(\boldsymbol{v}) = \left(\frac{m}{2\pi k_BT}\right)^{3/2} e^{-m\boldsymbol{v}^2/2k_BT}. \tag{2.3.12}$$

Remark. Although the Boyle-Charles law is obtained, this does not tell us anything about molecules, because we do not know N. We cannot tell the mass of the particle, either. Remember that k_B was not known when the kinetic theory of gases was being developed. A notable point is that with empirical results that can be obtained from strictly equilibrium studies, we cannot tell anything about molecules. □

Exercise 1. The *virial theorem* tells us

$$\overline{K} = -\frac{1}{2}\overline{\sum_i q_iF_i}, \tag{2.3.13}$$

where q_i is the position coordinates and F_i is the force (i.e., $\dot{p}_i = F_i$, where p_i is the conjugate momentum of q_i), and the overline is time average.[41]

(i) Demonstrate (2.3.13) (Hint: $d(\sum p_iq_i)/dt = 2K + \sum_i q_iF_i$).

(ii) F_i consists of the force F_{0i} due to the particle-particle interactions and the force F_{wi} due to the wall. If the interaction with the wall is short ranged, near the wall perpendicular to the x-axis at the x-coordinate L_x we obtain $\overline{\sum_i q_iF_{wi}} = L_x \sum_i \overline{F_{wi}}$. Show that $\sum_i \overline{F_{wi}} = -pA$, where A is the area of the wall.

Thus, we have arrived at the *virial equation of state* ($V = L_xA$):

$$pV = \frac{2}{3}\overline{K} - \frac{1}{3}\sum_i \overline{q_iF_{0i}}. \tag{2.3.14}$$

Notice that this is correct both classically and quantum mechanically, irrespective of the phase (not necessarily for the gas phase), and both in and out of equilibrium (if time-independent or change is sufficiently slow). □

2.3.6 Mean free path: more details

We already know that the mean free path $\ell \propto 1/nd^2$, where n is the number density and d the diameter of the molecule (→**2.2.6**). If the molecule we study were the only moving particle in a gas,

$$\ell\pi d^2 n = 1 \tag{2.3.15}$$

should have been true. That is, $\ell = 1/\pi nd^2$. However, all the molecules are moving.

The velocities of different molecules are statistically independent, so on the average the velocities are orthogonal, but the averages of the speed must be identical. When they collide, the average relative speed must be $\sqrt{2}$ times the mean velocity. Therefore,

$$\ell = \frac{1}{\sqrt{2}\pi nd^2} \tag{2.3.16}$$

---

[40] About experimental verification, see Jeans, p124 ∼.

[41] With an appropriate reinterpretation of the overline, the theorem holds for quantum mechanical cases as well.

must be the true mean free path length.

Exercise 1. Let $\overline{w}$ be the average relative speed. Then, the number z of collisions per unit time is given by

$$z = \pi d^2 n\overline{w}. \tag{2.3.17}$$

Show that the (density) distribution function for w is

$$f(w) \propto e^{-m\boldsymbol{w}^2/4k_BT}, \tag{2.3.18}$$

so $\overline{w} = \sqrt{2}\overline{v}$, where v is the speed of a molecule. This is the relation we discussed above. Now,

$$\ell = \frac{\text{mean displacement per unit time}}{\text{mean number of collision per unit time}} = \frac{\overline{v}}{z}. \tag{2.3.19}$$

This is (2.3.16). □

Exercise 2. An explicit calculation of z with the aid of the distribution function is a good practice. □

Exercise 3. Compare the mean distance δ between the molecules and the mean free path.[42] □

Exercise 4. If a particle moves very slowly, its free path length must be short. Therefore, the mean free path is a function of the particle speed v: $\ell = \ell(v)$. What can the reader guess about this function? Is it a bounded function from above? [43] □.

Exercise 5. The free path length = the distance traversed by a molecule between two successive collisions is not constant. Its distribution function $f(x)$ does not have a simple analytic expression but is very accurately approximated by (Jeans p145)

$$f(x) \propto e^{-1.04x/\ell}. \tag{2.3.22}$$

Show that the free path length distribution f_v for a particle of speed v is given by[44]

$$f_v(x) \propto e^{-x/\ell(v)}. \tag{2.3.23}$$

□

2.3.7 Shear viscosity and mean free path

We consider gases that are dilute enough to obey the ideal gas law, but with mutual collisions occurring sufficiently frequently.

Suppose we have a shear flow with the velocity $\boldsymbol{V}$ in the x-direction and the velocity gradient in the z-direction; we assume $\partial V_x(z)/\partial z$ is constant. We further assume that the particle velocity obeys the "local equilibrium distribution": this implies that $\boldsymbol{v} - \boldsymbol{V}(\boldsymbol{r})$ obeys

---

[42] Roughly, $\ell/\delta = (\ell/d)^{2/3}$.

[43] The mean free path of a gas molecule running at speed v is given by (Jeans p141)

$$\ell(v) = \frac{\beta m v^2/2}{\sqrt{\pi} n d^2 \psi(v\sqrt{\beta m/2})}, \tag{2.3.20}$$

where

$$\psi(x) = xe^{-x^2} + (2x^2+1)\int_0^x e^{-y^2}dy. \tag{2.3.21}$$

[44] If this is averaged over v, we obtain (2.3.22).

the Maxwell distribution at each point $\boldsymbol{r}$,[45] where $\boldsymbol{v}$ is the particle velocity.

On the average, the particle reaching the plane $z = 0$ comes from distance ℓ away from the point where the particle crosses this plane. The mean momentum carried by this particle is

$$m\boldsymbol{V}(-\ell\cos\theta) \simeq m\boldsymbol{V}(0) - m\ell\cos\theta\frac{\partial \boldsymbol{V}}{\partial z}, \tag{2.3.24}$$

where θ is the angle the particle makes with the z-direction. The number of particles with this angle that goes through the plane $z = 0$ in unit time must be

$$n\overline{v}\cos\theta\left(\frac{1}{2}\sin\theta d\theta\right). \tag{2.3.25}$$

The total momentum flux $J = -\eta\partial V_x/\partial z$ must be (2.3.24) times (2.3.25) averaged over angle θ from 0 to π:

$$J = -\frac{mn}{2}\overline{v}\ell\frac{\partial V_x}{\partial x}\int_0^\pi \cos^2\theta\sin\theta d\theta = -\frac{1}{3}mn\overline{v}\ell\frac{\partial V_x}{\partial z}. \tag{2.3.26}$$

Therefore, the shear viscosity η is given by

$$\eta = \frac{1}{3}mn\overline{v}l. \tag{2.3.27}$$

With the already obtained estimate of ℓ (2.3.16) and $\overline{v} = \sqrt{8k_BT/\pi m}$, we obtain[46]

$$\eta = \frac{2}{3d^2}\sqrt{\frac{mk_BT}{\pi^3}}. \tag{2.3.28}$$

This is independent of the density n as noted by Maxwell (→**2.3.1**). Furthermore, the viscosity increases with temperature.

Other transport quantities such as heat conductivity may be obtained similarly, but we will do this in a slightly streamlined fashion later (→**2.3.9**).

Discussion 1. One of the arguments for the caloric theory of heat was as follows: if heat were a form of motion of molecules, then in solids its propagation should be wave like and should be much faster than we actually observe. Diffusion like behavior of heat (or temperature) is a good evidence of a fundamental conserved substance called *caloric* whose density determines temperature. □

2.3.8 Estimating Avogadro's constant from viscosity

We have noted in **2.3.5** that it is impossible to determine the properties of particles through equilibrium studies alone.

---

[45] More accurately, in the volume element around each point $\boldsymbol{r}$; we assume that we can choose a volume macroscopically sufficiently small (so, for example, in our present case $\boldsymbol{V}$ does not change inside), but contains sufficiently many microscopic objects that we can consider statistics meaningfully. For local equilibrium to hold, we need sufficiently many collisions among particles.

[46] If we assume that the particle mass, the cross section (d^2) and the particle thermal velocity are only relevant quantities, dimensional analysis gives essentially this result. Even if we try to take the density of the gas into account, it automatically drops out of the formula.

If we consider a cylinder of diameter d and height ℓ, it should contain on the average a single molecule. Therefore, Clausius guessed (as we have already seen in **2.2.6**)

$$\pi d^2 \ell = 1/n, \tag{2.3.29}$$

where n is the number density, but no quantity in this formula was known. mn is measurable, and from the ideal gas law we can estimate $\langle \boldsymbol{v}^2 \rangle$ (→(2.3.9)) or $\overline{v}$. Now, the viscosity is measurable, so we can determine ℓ (→(2.3.27)). Still, without any microscopic parameter directly related to the molecule (say, its mass m or its size d), there is no way to obtain n (or the Avogadro's constant N_A/molar volume of a gas).

In 1873[47] van der Waals proposed his equation of state of imperfect gases:

$$p(V - V_0) = Nk_BT - \frac{\alpha}{V}(1 - V_0/V). \tag{2.3.30}$$

Here, $\alpha > 0$ is a constant describing the effects of intermolecular attractive interactions, and V_0 corresponds to the actual volume occupied by the molecules: $V_0 \simeq bN\pi d^3/6$, where b is a geometrical constant of order unity. This equation allows us to determine d. Thus, we can compute N_A. The method gives $N_A \simeq (4 \sim 6) \times 10^{23}$.[48]

Exercise 1. Compute ℓ, $\overline{v}$ and the number of collisions per molecule in unit time (or the reciprocal of the mean free time $\ell/\overline{v}$) for appropriate gas samples.□

2.3.9 General formula of transport coefficients in terms of mean free path

Let X be a physical quantity carried by a volume element (for example, $X = \sum_{dV} mv_x$ is the quantity considered in **2.3.7**, where the sum $\sum_{dV}$ is taken over all the particles in the volume element dV around $\boldsymbol{r}$). We assume a local equilibrium, so the average density of X, indicated by the overline as $\overline{X}$, depends on the position $\boldsymbol{r}$:

$$\langle X \rangle_{\boldsymbol{r}}/dV = \overline{X}(\boldsymbol{r}), \tag{2.3.31}$$

where dV is the volume of the volume element (we are computing the local density). The z-component of the flux J_X of X at $\boldsymbol{r}$ may be computed as

$$J_X(\boldsymbol{r}) = \langle [\overline{X}(\boldsymbol{r} - \boldsymbol{\ell}) - \overline{X}(\boldsymbol{r})] v_z \rangle_\theta, \tag{2.3.32}$$

where $\boldsymbol{\ell}$ is the free path vector and $\langle \ \rangle_\theta$ implies the average over the direction of the particle motion (that is parallel to $\boldsymbol{\ell}$). Expanding the expression, we obtain

$$J_X(\boldsymbol{r}) = -\langle v_z \boldsymbol{\ell} \rangle_\theta \cdot grad\overline{X}(\boldsymbol{r}). \tag{2.3.33}$$

Since, the gradient has only the z-component, we need[49]

$$\langle v_z \ell_z \rangle_\theta \simeq \frac{1}{3}\overline{v}\ell. \tag{2.3.34}$$

Thus,

$$J_X(\boldsymbol{r}) = -\frac{1}{3}\overline{v}\ell\frac{\partial \overline{X}}{\partial z}. \tag{2.3.35}$$

---

[47] Maxwell, *A Treatise on Electricity and Magnetism* was published this year.

[48] The most accurate method to estimate Avogadro's constant is to use the X-ray analysis of crystals. See, Mana, Rivista Nuovo Cim 18(3) (1995); P. Beker, "History and progress in the accurate determination of the Avogadro constant," Rep Prog Phys **64** 1945 (2001).

[49] Note that $\boldsymbol{v}$ and $\boldsymbol{\ell}$ are parallel.

2.3.10 Viscosity, heat conductivity, and diffusion constant for dilute gas

For viscosity $X = \sum_{dV} mv_x$ in **2.3.9**, so $\overline{X}(\boldsymbol{r}) = nmV(\boldsymbol{r})$, and we recover η (2.3.27) from (2.3.35).

To obtain the heat conductivity λ, X must be the energy contained in the volume element: $X = \sum_{dV} m\boldsymbol{v}^2/2$, so $\overline{X}(\boldsymbol{r}) = 3nk_BT(\boldsymbol{r})/2$. Since $\boldsymbol{J} = -\lambda\nabla T$, we obtain

$$\lambda = \frac{1}{2}nk_B\ell\overline{v}. \tag{2.3.36}$$

For self-diffusion (i.e., isotope diffusion) $X = \sum_{dV} 1$, so $\overline{X}(\boldsymbol{r}) = n(\boldsymbol{r})$. $\boldsymbol{J} = -D\nabla n$ implies

$$D = \frac{1}{3}\overline{v}\ell = \frac{\ell^2}{3\tau}, \tag{2.3.37}$$

where τ is the *mean free time* $\tau = \ell/\overline{v}$.

Notice that $\eta = nmD$, $\eta/\lambda = 2m/3k_B$ and $\lambda/D = 3nk_B/2$. The last relation tells us $\lambda = c_VD$, where c_V is the specific heat per molecule of gas under constant volume. Again, we should note that these relations do not tell us anything about the microscopic properties of the gas particles.

2.3.11 What happens at wall?

In 1860 Helmholtz and Piotrowski showed that for a small tube the flow behavior of liquid appreciably deviates from Poiseuille's flow,[50] and later (in 1875) Kundt and Warburg showed the same for gases.[51].

If the collision with the wall is specular (that is, the momentum component parallel to the wall is maintained and that perpendicular to the wall flips its sign), we expect a slippery boundary condition for the velocity at the wall. However, if the wall is microscopically rough, the reflected momentum is randomized and the average of its components parallel to the wall may be close to zero. This case we expect a sticky boundary condition we usually use for hydrodynamics. The above experiments clearly say that sometimes randomization is not complete.

Maxwell used such experimental results to estimate the ratio of specular and random reflections at the wall, and concluded that their contributions could become comparable for some surfaces.[52]

2.3.12 Knudsen flow

If the gas density is low and the mean free path becomes comparable to the representative size of the system (e.g., the diameter of a pipe), the flow in a pipe behaves like a plug flow.

Consider a thin pipe whose diameter is smaller than the mean free path length. The local average velocity becomes independent of the distance from the wall and is constant u in this case. The gas particles colliding with the tube wall would lose momentum mu on the

---

[50] For a modern demonstration using a thin gap, see Granick et al., *Limits of the Hydrodynamic No-Slip Boundary Condition*, Phys. Rev. Lett. **88**, 106102 (2002).

[51] Read Jeans p168.

[52] The reader might think it is only a mere detail, but if we wish to study nonequilibrium states with the aid of molecular dynamics, this is one of the sources of headache.

average. The number of collisions is proportional to $n\bar{v}/4$ per unit time.[53] Therefore, along the tube of radius R and length L, the lost momentum per unit time must be

$$2\pi RL \times mu \times \frac{1}{4}n\bar{v}. \tag{2.3.38}$$

This must balance with the force due to the pressure difference: $\pi R^2 \Delta p$. Therefore, we can compute the average flow velocity

$$u = \frac{2R\Delta p}{Lnm\bar{v}}. \tag{2.3.39}$$

The total mass flux J is given by

$$J = nmu\pi R^2 = \frac{2R^3\Delta p}{L\bar{v}}. \tag{2.3.40}$$

Notice that this decreases with an increase of temperature. Its intuitive reason should be obvious to the reader.

Exercise 1. The boundary layer is the layer where the effect of the wall is significantly felt by the flow velocity field. Estimate its thickness and argue that it is of the order of the mean free path for the gas where the kinetic theory works (use the results in **2.3.7**). □

Exercise 2. Discuss the pressure balance between the two vessels at different temperatures connected by a very thin tube. □

Exercise 3. If there is no gas, obviously there should not be any flow nor any viscous drag. Discuss how the gas viscosity behaves as a function of density. □

2.4 Boltzmann and Modern Kinetic Theory

2.4.1 Maxwell-Boltzmann distribution

In 1868 Boltzmann (24 years old; 1844-1906) extended Maxwell's distribution theory to the cases with external forces, and introduced the so-called *Boltzmann factor* e^{-V/k_BT}. Notice that the probability of a microscopic state is explicitly defined by Boltzmann as the ratio of the sojourn time of the system in that state and the total time of observation.

Exercise 1. Can the reader explain to her lay friend that the Boltzmann factor is indeed plausible? [Consider the balance between the pressure and the force acting on the gas particles for an ideal gas.] □

Later (1879), Maxwell rederived the result in a much more general fashion; his motivation was to show that Boltzmann's result was more general than was suggested by the theoretical method used by Boltzmann applicable only to dilute gases. Maxwell realized that the only needed assumption is that sooner or later all the phase points in an energy shell is visited by the trajectory of the system ('*ergodicity*' in Boltzmann's terminology →**1.5.9**). To this end Maxwell introduced the concept of *ensemble* = a set of numerous systems that are different only for their initial conditions (but with the same total energy).[54]

Exercise 2. Consider a single component monatomic gas consisting of N atoms of mass

---

[53] 1/2 comes from the fact that the one half of molecules are going away from the wall at any moment; the other 1/2 factor comes from the average of the cos factor.

[54] Notice that Maxwell introduced the concept of ensemble only to visualize probability.

m. Let $\{q, p\}$ denote collectively the phase coordinates of all the atoms. We wish to trace Maxwell's logic.
(1) Obtain the phase volume $\Gamma(\{q\})$ of the state with total energy E and the spatial positions of atoms being in the volume element dq around $\{q\}$ (this is the collective expression of all the spatial coordinates). Assume that $V(\{q\})$ be the potential energy of interatomic interactions.
(2) Suppose V is smaller than $E \equiv 3NK$, where K is the average kinetic energy per particle, demonstrate that $\Gamma(\{q\}) \propto e^{-V/2K}$.□

By around 1870 it was clearly recognized that:
Thermal equilibrium condition in mechanics terms is the equality of average kinetic energy per atom.

2.4.2 Why and how did Boltzmann reach statistical mechanics?

Since Boltzmann started his physicist career trying to demonstrate the second law mechanically, it was natural for him to ask how the equilibrium Maxwell distribution was reached: Maxwell demonstrated that once the Maxwell distribution was established, it did not change (→**2.3.3**), but he did not show that the distribution was unique nor was reached from any initial distribution.

To complete Maxwell's program, Boltzmann decided to study the time evolution of the probability density function $f(q, p)$ of the one particle state of a (monatomic) molecule in a gas specified by its position q and momentum p (the probability measure on the *μ-space*[55]), and proposed his famous equation, the *Boltzmann equation* (in 1872,[56] →**2.4.7**). The equation has the following structure

$$\frac{\partial f}{\partial t} + \frac{p}{m}\frac{\partial f}{\partial q} = \text{effect of binary collisions.} \tag{2.4.1}$$

Boltzmann introduced a quantity corresponding to (negative) entropy (per particle) called H-function defined by

$$H = \int f \log f \, dpdq, \tag{2.4.2}$$

and demonstrated that $dH/dt < 0$ (= only for the Maxwell distribution, if f obeys the Boltzmann equation. This assertion is called the *H-theorem* →**2.4.10**). In other words, from any initial condition, the ultimate equilibrium distribution is reached that is given by Maxwell's distribution. Thus, Boltzmann believed that he could demonstrate the second law of thermodynamics starting from mechanics.

Then, his colleague Loschmidt commented that mechanics was time-reversal symmetric, so the asymmetry of the Boltzmann equation, if claimed as a result of pure mechanics, was a paradox (the *Loschmidt paradox*). Boltzmann realized that an extra assumption was needed, and clearly understood the meaning of the second law: the time evolution of the system to the most probable state. Now, he also realized that then, the most probable distribution had to be the equilibrium distribution. This is the birth of equilibrium statistical mechanics: we can study the statistical behavior of a macroscopic system without explicit use of mechanics (→**2.4.11**).

---

[55] μ for 'molecule'; Γ is actually for 'gas.'

[56] This year, Yellowstone was established as the first NP; the Challenger expedition began (one year before Maxwell's Treatise on electromagnetism); G Elliot, *Middlemarch*.

If the reader is interested in more details about the story outlined here, read the following entries (**2.4.4 -2.4.11**).[57]

2.4.3 Maxwell's reaction to attempts to derive the second law

Maxwell introduced his demon (*Maxwell's demon*) that can segregate molecules with more kinetic energies (faster molecules). A container is divided into two chambers H and C by a wall with a hole. The hole is equipped with a frictionless sliding door. The Demon sits besides the door, watching the gas molecules in both chambers. When a molecule in H (resp., in C) whose speed is slower (resp., faster) than the average in H (resp., in C) comes to the hole, he opens it and let it go into the other room. This way the chamber H becomes hotter and C colder, apparently violating the second law.

From his point of view the second law is equivalent to a denial of our power to perform the operation that the Demon can. "Hence, if the heat consists in the motion of its parts, the separate parts which move must be so small that we cannot in any way lay hold of them to stop them."[58] Thus, the second law cannot be a mere theorem of mechanics. Naturally, Maxwell was not thrilled by the German/Austrian development.

2.4.4 Elementary derivation of Boltzmann's equation: streaming term

The Maxwell distribution (→**2.3.2**) is the distribution of the velocity of a single particle. Therefore, Boltzmann wished to study the single-body distribution function $f(\boldsymbol{r}, \boldsymbol{v}, t)$, where $f(\boldsymbol{r}, \boldsymbol{v}, t) d\boldsymbol{r} d\boldsymbol{v}$ is the probability of finding a particle whose state specified by its position $\boldsymbol{r}$ and velocity $\boldsymbol{v}$ in a 6-dimensional volume element $d\boldsymbol{v} d\boldsymbol{r}$ centered at $(\boldsymbol{r}, \boldsymbol{v})$. That is, f is the probability density on the μ-space.[59]

If the particles do not interact with each other, then

$$\frac{\partial}{\partial t} f + \boldsymbol{v} \cdot \frac{\partial}{\partial \boldsymbol{r}} f = 0 \tag{2.4.3}$$

in the homogeneous space. The second term on the LHS describes the motion of the particle 'flowing with the stream with velocity $\boldsymbol{v}$,' so it is called the *streaming term.*

Exercise 1. This describes a ballistic motion. Its solution with the initial condition $f(\boldsymbol{r}, \boldsymbol{v}, 0) = f_0(\boldsymbol{r}, \boldsymbol{v})$ reads $f(\boldsymbol{r}, \boldsymbol{v}, t) = f_0(\boldsymbol{r} - \boldsymbol{v}t, \boldsymbol{v})$. Check this statement. □

Exercise 2. This is the Liouville equation (→**2.7.7**) for a single body. □.

If the particle interact with other particles (if there are collisions), the above equation is not true and we must write

$$\frac{\partial}{\partial t} f + \boldsymbol{v} \cdot \frac{\partial}{\partial \boldsymbol{r}} f = \left(\frac{\partial f}{\partial t} \right)_{col}, \tag{2.4.4}$$

where the subscript *col* indicates the contribution of collisions. The reader should immediately realize that the streaming term does not contribute irreversibility at all.

---

[57] In these notes the Boltzmann equation will not be discussed in detail; this may be historically important, and may still be relevant to gases, but we are much more interested in condensed phases, for which the equation is irrelevant. Also it should be irrelevant to the foundation of statistical mechanics, because the equation applies only to rather dilute gases.

[58] J. C. Maxwell, *Theory of Heat* (Longmans, London 1871).

[59] The 'true' μ-space is spanned by the position coordinates and the momentum coordinates instead of velocities.

2.4.5 Two-body collision term

To obtain the collision term, we assume the following:
(1) There are only two-body collisions.
(2) "Stosszahlansatz":[60] colliding particles are statistically uncorrelated.
(3) f does not change as a function of spatial position within the range of the particle-particle interactions.
According to the assumption (2) the probability of collisions between the particles with velocity $\boldsymbol{v}$ and velocity $\boldsymbol{v}'$ around $\boldsymbol{r}$ may be written as $nf(\boldsymbol{r},\boldsymbol{v})f(\boldsymbol{r},\boldsymbol{v}')d\boldsymbol{v}d\boldsymbol{v}'d\boldsymbol{r}d\boldsymbol{r}$, where n is the number density of the particles.

Suppose the relative velocity of the two colliding particles (far before the collision) is $\boldsymbol{g}$, and the collision parameter is $\boldsymbol{b}$. The collision parameter is the distance vector between the two colliding particles projected to the subspace orthogonal to $\boldsymbol{g}$. More precisely, let us write the ith particle coordinates be $(\boldsymbol{r}_i,\boldsymbol{v}_i)$. Then,

$$\begin{aligned}\boldsymbol{g} &= \boldsymbol{v}_2-\boldsymbol{v}_1, &(2.4.5)\\ \boldsymbol{b} &= (\boldsymbol{r}_2-\boldsymbol{r}_1)(1-\boldsymbol{g}\boldsymbol{g}^T/g^2), &(2.4.6)\end{aligned}$$

where $g=|\boldsymbol{g}|$. The number of particles coming with the collision parameter $\boldsymbol{b}$ per unit time can be written as

$$2\pi bdb\, nf(\boldsymbol{r}_2,\boldsymbol{v}_2,t)g, \tag{2.4.7}$$

where $b=|\boldsymbol{b}|$. Therefore, the expected number of collisions with the particle with velocity $\boldsymbol{v}_2$ per unit time experienced by the particle with velocity $\boldsymbol{v}_1$ at $\boldsymbol{r}_1$ may be expected to be

$$2\pi bdb\, nf(\boldsymbol{r}_1,\boldsymbol{v}_1,t)f(\boldsymbol{r}_1,\boldsymbol{v}_2,t)g\, d\boldsymbol{v}_1d\boldsymbol{v}_2d\boldsymbol{r}_1. \tag{2.4.8}$$

Here, we have used (3) to identify all the spatial coordinates in the density distribution functions.

2.4.6 Contribution of collisions to time evolution

The above collision process describes the collision $\{\boldsymbol{v}_1,\boldsymbol{v}_2\}\to\{\boldsymbol{v}_1',\boldsymbol{v}_2'\}$, where the primed velocities after the collision are determined according to the energy and momentum conservation. Since mechanics is time reversal symmetric, there must be the collision process $\{\boldsymbol{v}_1',\boldsymbol{v}_2'\}\to\{\boldsymbol{v}_1,\boldsymbol{v}_2\}$ as well.[61] Therefore, the expectation value of the number of collisions that give velocity $\boldsymbol{v}_1$ after collision must be given by

$$2\pi b'db'nf(\boldsymbol{r}_1,\boldsymbol{v}_1',t)f(\boldsymbol{r}_1,\boldsymbol{v}_2',t)g'd\boldsymbol{v}_1d\boldsymbol{v}_2d\boldsymbol{r}_1. \tag{2.4.9}$$

Combining the above results, we may write

$$\begin{aligned}\left(\frac{\partial f}{\partial t}\right)_{col} d\boldsymbol{r}_1d\boldsymbol{v}_1 &= n\int 2\pi b'db'f(\boldsymbol{r}_1,\boldsymbol{v}_1',t)f(\boldsymbol{r}_1,\boldsymbol{v}_2',t)g'd\boldsymbol{v}_1'd\boldsymbol{v}_2'd\boldsymbol{r}_1\\ &\quad -n\int 2\pi bdbf(\boldsymbol{r}_1,\boldsymbol{v}_1,t)f(\boldsymbol{r}_1,\boldsymbol{v}_2,t)gd\boldsymbol{v}_1d\boldsymbol{v}_2d\boldsymbol{r}_1. \qquad (2.4.10)\end{aligned}$$

Now, we use the equality

$$b'db'g'd\boldsymbol{v}_1'd\boldsymbol{v}_2'd\boldsymbol{r}_1'=bdb\,g\,d\boldsymbol{v}_1d\boldsymbol{v}_2d\boldsymbol{r}_1 \tag{2.4.11}$$

---

[60] hypothesis of collision numbers

[61] Here, more precisely, we use the detailed balance (→**2.7.6**).

that holds thanks to Liouville's theorem (→**2.7.2**) applied to the two-body phase volume. Thus, we finally obtain

$$\left(\frac{\partial f}{\partial t}\right)_{col} = n\int 2\pi b db d\boldsymbol{v}_2 g\left[f(\boldsymbol{r}_1,\boldsymbol{v}_1',t)f(\boldsymbol{r}_1,\boldsymbol{v}_2',t) - f(\boldsymbol{r}_1,\boldsymbol{v}_1,t)f(\boldsymbol{r}_1,\boldsymbol{v}_2,t)\right]. \tag{2.4.12}$$

After the collision the relative velocity changes its direction to the direction specified by the azimuthal angle θ measured from the direction of $\boldsymbol{g} = \boldsymbol{v}_2 - \boldsymbol{v}_1$.

$$2\pi b db = \sigma(\theta) d\Omega, \tag{2.4.13}$$

where Ω is the solid angle of the annule between θ and $\theta + d\theta$, and $\sigma(\theta)$ is called the *differential scattering cross section.*[62]

$$\left(\frac{\partial f}{\partial t}\right)_{col} = n\int \sigma d\Omega d\boldsymbol{v}_2\, g\left[f(\boldsymbol{r}_1,\boldsymbol{v}_1',t)f(\boldsymbol{r}_1,\boldsymbol{v}_2',t) - f(\boldsymbol{r}_1,\boldsymbol{v}_1,t)f(\boldsymbol{r}_1,\boldsymbol{v}_2,t)\right]. \tag{2.4.14}$$

2.4.7 Boltzmann equation

Thus, we have arrived at the Boltzmann equation

$$\frac{\partial}{\partial t}f + \boldsymbol{v}\cdot\frac{\partial}{\partial \boldsymbol{r}}f = n\int \sigma d\Omega g\left[f(\boldsymbol{r}_1,\boldsymbol{v}_1',t)f(\boldsymbol{r}_1,\boldsymbol{v}_2',t) - f(\boldsymbol{r}_1,\boldsymbol{v}_1,t)f(\boldsymbol{r}_1,\boldsymbol{v}_2,t)\right]. \tag{2.4.15}$$

Here, $\boldsymbol{v}_1 + \boldsymbol{v}_2 = \boldsymbol{v}_1' + \boldsymbol{v}_2'$ (and the energy conservation law) must be satisfied when the integration is performed.

Exercise 1. If there is an external force field $\boldsymbol{F}$, the streaming term (→**2.4.4**) must be modified. Give a necessary modification and write down the most general Boltzmann equation allowing spatial nonuniform equilibrium distributions. Of course, the external field must change spatially sufficiently gently to satisfy our assumption (3) in **2.4.5**. □.

A more careful (but still formal) derivation will be given later from the BBGKY hierarchy (→**2.7.10**).

The Boltzmann equation is an equation for a one-body distribution function. The actual system is an N ($\to \infty$) body system, so the justification of the Boltzmann equation is equivalent to justification of the reduction of N body description to the single body description. This has been accomplished by Lanford[63] for the hard core gas case for a very short time ($\sim 1/4$ of the mean free time) in the so-called *Boltzmann-Grad limit*: $N \to \infty$, $Nd^2 \to 1$, where d is the hard core diameter). That is, the reliability of the Boltzmann equation has been demonstrated only for this microscopic time scale.

Remark. If the limit $N \to \infty$, $Nd^3 \to 1$ is considered, then the *Enskog equation* that replaces the collision term of the Boltzmann equation with the one that takes into account the spatial difference of the distribution at the scale of d is expected.[64] ,[65] □

---

[62] There are other definitions as well.

[63] O E Lanford, III, "Time evolution of large classical systems," Lecture Notes in Physics **38**, 1-111 (1975).

[64] The Enskog equation is known to be a good model even for liquid. For example, see T. Scopigno, R. Di Leonardo, L. Comez, A. Q. R. Baron, D. Fioretto, and G. Ruocco, " Hard-Sphere-like Dynamics in a Non-Hard-Sphere Liquid," Phys. Rev. Lett., **94**, 155301 (2005).

[65] See a related paper: F Rezakhanlou, "A stochastic model associated with Enskog equation and its kinetic limit," Commun. Math. Phys. **232**, 327-375 (2003).

2.4.8 Collision invariants

Let φ be a function of single particle coordinates. Let us write φ_i be the quantity associated with particle i before the collision and φ'_i after the collision. Suppose particles 1 and 2 undergo a collision. If

$$\varphi_1 + \varphi_2 = \varphi'_1 + \varphi'_2, \tag{2.4.16}$$

we say φ is a *collision invariant.*

The collision invariant, if continuous, must have the following form

$$\varphi = A + \boldsymbol{B} \cdot \boldsymbol{v} + C\boldsymbol{v}^2, \tag{2.4.17}$$

where A, C are constants and $\boldsymbol{B}$ is a constant vector (actually, they can be functions of the position $\boldsymbol{r}$). This is not very trivial to show.[66]

If φ is a collision invariant, then

$$I = \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g\, \varphi(\boldsymbol{v}_1) \left[f(\boldsymbol{r}_1, \boldsymbol{v}'_1, t) f(\boldsymbol{r}_1, \boldsymbol{v}'_2, t) - f(\boldsymbol{r}_1, \boldsymbol{v}_1, t) f(\boldsymbol{r}_1, \boldsymbol{v}_2, t) \right] = 0. \tag{2.4.18}$$

This follows from the following identity that is valid for any function of the single particle coordinates

$$I = \frac{1}{4} \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g\, (\varphi_1 + \varphi_2 - \varphi'_1 - \varphi'_2) \left[f(\boldsymbol{r}_1, \boldsymbol{v}'_1, t) f(\boldsymbol{r}_1, \boldsymbol{v}'_2, t) - f(\boldsymbol{r}_1, \boldsymbol{v}_1, t) f(\boldsymbol{r}_1, \boldsymbol{v}_2, t) \right]. \tag{2.4.19}$$

Here, an abbreviation $\varphi'_2 = \varphi(\boldsymbol{r}'_2, \boldsymbol{p}'_2)$, etc., is adopted.

A demonstration follows:

$$\begin{aligned} I &= \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g\, (f_{(1)} f_{(2)} - f_{(1')} f_{(2')}) \varphi_1, & (2.4.20) \\ &= \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g\, (f_{(1)} f_{(2)} - f_{(1')} f_{(2')}) \varphi_2. & (2.4.21) \end{aligned}$$

Here, we have used the abbreviation $f_{(1)} = f(\boldsymbol{r}_1, \boldsymbol{v}_1, t)$, etc. Now, we change the variable as $(\boldsymbol{v}_1, \boldsymbol{v}_2) \to (\boldsymbol{v}'_1, \boldsymbol{v}'_2)$ to obtain (with the aid of (2.4.11))

$$\begin{aligned} I &= \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g\, (f_{(1')} f_{(2')} - f_{(1)} f_{(2)}) \varphi'_1, & (2.4.22) \\ &= \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g\, (f_{(1')} f_{(2')} - f_{(1)} f_{(2)}) \varphi'_2. & (2.4.23) \end{aligned}$$

Combining all of them, we reach the desired result.

2.4.9 Local equilibrium distribution

A density distribution function satisfying

$$\left(\frac{\partial f}{\partial t} \right)_{col} = 0 \tag{2.4.24}$$

---

[66] A proof can be found in C Cercignani, *The Boltzmann Equation and Its Applications* (Springer-Verlag, New York, 1988), p74-78. If the reader accepts that there are only four additive constants of motion, the equation (2.4.17) looks trivial, but here, we are actually demonstrating that the additive invariants are only linear momenta and kinetic energy (in our case, we do not take into account of the angular momenta of the particles).

is called a *local equilibrium distribution.* A local equilibrium distribution function must have the following form

$$\log f = A + \boldsymbol{B}\cdot\boldsymbol{v} + C\boldsymbol{v}^2, \tag{2.4.25}$$

where A, C are functions of spatial coordinates and $\boldsymbol{B}$ is a vector dependent only on the spatial coordinates. Therefore, notice that a local equilibrium distribution is *not* generally a stationary distribution function (cf., **5.1.8**).

To demonstrate this, we demonstrate the following fundamental inequality

$$J = \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g \log f(\boldsymbol{r}_1,\boldsymbol{v}_1,t)\,[f(\boldsymbol{r}_1,\boldsymbol{v}_1',t)f(\boldsymbol{r}_1,\boldsymbol{v}_2',t) - f(\boldsymbol{r}_1,\boldsymbol{v}_1,t)f(\boldsymbol{r}_1,\boldsymbol{v}_2,t)] \le 0 \tag{2.4.26}$$

for any density distribution f. Thanks to (2.4.19) we obtain (for the abbreviations see above (2.4.22))

$$J = \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g \log\frac{f_{(1)}f_{(2)}}{f_{(1')}f_{(2')}}(f_{(1')}f_{(2')} - f_{(1)}f_{(2)}). \tag{2.4.27}$$

Let $\lambda = f_{(1)}f_{(2)}/f_{(1')}f_{(2')}$. Then, this can further be rewritten as

$$J = \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g\,(1-\lambda)\log\lambda\, f_{(1')}f_{(2')}. \tag{2.4.28}$$

Since $(1-\lambda)\log\lambda \le 0$,[67] we are done.

Thus, we have demonstrated that (2.4.24) implies that $\lambda = 1$. That is, $\log f$ must be an additive invariant of the collision. This concludes the demonstration of (2.4.25) (with the aid of **2.4.8**). That is, for a spatially uniform case, the Maxwell distribution (→**2.3.2**) is the unique invariant solution of the (spatially uniform) Boltzmann equation.

Remark. Often the Boltzmann equation is approximated with finitely many choices of velocity values instead of continuous velocities. If the Galileian invariance is violated as such cases, the equation is not generally compatible with thermodynamics. See Cercignani.[68]

2.4.10 *H*-theorem

We know that if there is no external force, the Maxwell distribution is the unique stationary solution of the Boltzmann equation as can be seen from **2.4.9**). Boltzmann's original aim was to show not only that this is the unique stationary state but also that any initial condition would reach this state eventually.

To this end Boltzmann devised a Lyapunov function called *H-function*

$$H = \int f \log f d\boldsymbol{r} d\boldsymbol{v}. \tag{2.4.29}$$

Let us study its time evolution

$$\frac{dH}{dt} = \int (1+\log f)\frac{\partial f}{\partial t} d\boldsymbol{r} d\boldsymbol{v} = \int \log f \frac{\partial f}{\partial t} d\boldsymbol{r} d\boldsymbol{v}. \tag{2.4.30}$$

The streaming term (→**2.4.4**) does not change H

$$\left(\frac{dH}{dt}\right)_{str} = \int \log f \left(\frac{\partial f}{\partial t}\right)_{str} d\boldsymbol{r} d\boldsymbol{v}, \tag{2.4.31}$$

$$= -\int \log f\, \boldsymbol{v}\frac{\partial}{\partial \boldsymbol{r}} f d\boldsymbol{r} d\boldsymbol{v} = 0. \tag{2.4.32}$$

---

[67] $f(\lambda) = (1-\lambda)\log\lambda$ takes the unique maximum 0 at $\lambda = 1$; it is convex upward.

[68] C. Cercignani, "Temperature, entropy, and kinetic theory," J. Stat. Phys. **87**, 1097-1109 (1997).

The collision term does change H as we have just seen in (2.4.26)

$$\left(\frac{dH}{dt}\right)_{col} = \int \sigma d\Omega d\boldsymbol{v}_1 d\boldsymbol{v}_2\, g \log f_{(1)} \left[f_{(1')}f_{(2')} - f_{(1)}f_{(2)}\right] \le 0. \tag{2.4.33}$$

That is,

$$\frac{dH}{dt} \le 0. \tag{2.4.34}$$

This vanishes if and only if $\log f$ is a sum of additive invariant as we have already seen in **2.4.9**. That is, H is a Lyapunov function whose unique minimum is given by the Maxwell distribution.

Thus, Boltzmann concluded that his equation describes the irreversible process whose unique destination is the equilibrium state (for a mathematical statement see **2.4.13**).

2.4.11 Boltzmann reaches $S \propto$ log(Komplexionszahl).

Thus, Boltzmann thought he demonstrated the second law purely mechanically. Almost immediately, his colleague Loschmidt[69] pointed out that without approximation time-reversal symmetric mechanics could not conclude such an irreversible equation.

Boltzmann realized the probabilistic element had been smuggled into his argument[70] and wrote a fully statistical paper:[71] basically, he asserts that the mechanical content of the second law was that the macroscopic state evolves to the direction that increases the number of Komplexions = microscopic states compatible with the macroscopic state.

The most important conclusion is that the equilibrium state is the most probable macroscopic state allowed to the system, so it can be studied without explicit dynamical consideration. This way, the study of nonequilibrium states receded from the main stage of statistical mechanics.

Exercise 1. I wish you to trace Boltzmann's logic, answering the following questions.

Suppose there are N gas particles. Let w_n be the number of particles with the energy between $(n-1)\epsilon$ and $n\epsilon$ ($\epsilon > 0$).

(1) Obtain the number of Komplexions (ignoring the energy constraint), and (with the aid of Stirling's formula) show that its maximization condition is equivalent to the minimization condition of

$$M = \sum w_i \log w_i. \tag{2.4.35}$$

(2) Write $w_i = w(x)\epsilon$ and simultaneously take the $n \to \infty$ and $\epsilon \to 0$ limits, maintaining $x = n\epsilon$ finite. Show that minimizing M is equivalent to minimizing

$$M' = \int w(x) \log w(x) dx. \tag{2.4.36}$$

(3) We should not ignore the constraints that the total number of particles is N and the total energy is K. Under this condition, derive Maxwell's distribution in 3-space by minimizing M'. [In the original paper Boltzmann first reduced the problem to a single particle problem,

---

[69] About him and his relation to Boltzmann see the book by Lindley.

[70] This will be discussed when we rederive the Boltzmann equation from the BBGKY hierarchy (→**2.7.10**).

[71] "Über der Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Warscheinlichkeitsrechnung respective den Sätzen über des Wärmegleichgewicht," Wiener Ber. **76**, 373-435 (1877) [On the relation between the second principle of mechanical heat theory and probability theory with respect to theorems of thermal equilibrium].

and wrote $w(x)$ as $f(u,v,w)$, where u, v, w are three components of the velocity. Also Boltzmann understands w to be the probability density distribution function.]
(4) Now, Boltzmann realized that M' gives the entropy of the ideal gas. Show this.
Based on this finding, he proposed

$$S \propto \log(\text{Number of Komplexions}). \tag{2.4.37}$$

At the end of this same paper he says, although the relation could be demonstrated rigorously (sic) with the aid of kinetic gas theory, he cannot do so for other phases. However, he also claims that the relation is likely. □

2.4.12 What is statistical mechanics?: Fundamental problem

We have just witnessed the moment of the birth of statistical mechanics in **2.4.11**.

Since the fundamental assumption is that any macroscopic system is governed by mechanics, in principle, we only need mechanics to understand its state. The essence of statistical mechanics is to get rid of mechanic with the aid of a statistical principle. In modern terms, what Boltzmann found is: the probability of a state with a given energy is proportional to the number of microscopic states. That is, all the microscopic states with the same energy have the same statistical weight.

In this sense nonequilibrium statistical mechanics general enough to be a counterpart of equilibrium statistical mechanics does not yet exist. The *fundamental problem of nonequilibrium statistical mechanics* is to find a statistical principle as general as possible to remove explicit mechanical considerations.

Equilibrium thermodynamics is the key to establish a statistical principle as we see in Gibbs's work (→**2.1.5**) or in Einstein's work (→**2.5.3**). This tells us that if one wishes to establish nonequilibrium statistical mechanics, one must first establish a nonequilibrium phenomenology. If there is no general phenomenological framework, there is no general statistical framework. Therefore, the only thing those who are interested in extending the realm of statistical mechanic must do at present is to try to find such a phenomenological framework.[72]

2.4.13 What did actually Boltzmann demonstrate?

In the course of replying to Loschmitd's criticism Boltzmann realized that the choice of the initial condition is very crucial. Irreversibility is not derived but produced or smuggled into the result through the statistical assumption that the correlation produced by collision is always ignored. This point will be much clearer to the reader later (→**2.7.10**).

Logically speaking, even without such extra assumption, since we can justify the Boltzmann equation only for a very short time (→**2.4.7**), we are not entitled to discuss any long time asymptotic behavior with the aid of the Boltzmann equation. Therefore, demonstration of irreversibility from mechanics is out of question. Only when we accept the Boltzmann equation as a model of the gas dynamics can we discuss the irreversibility of gases in terms of the Boltzmann equation.

---

[72] Certainly, the lecturer has been trying this; Y Oono and M Paniconi, "Steady state thermodynamics," Prog. Theor. Phys. Suppl. **130**, 29-44 (1998). See also T. Hatano and S. Sasa, "Steady-State Thermodynamics of Langevin Systems," Phys. Rev. Lett. **86**, 346 (2001). At least the generalized second law seems experimentally verified very recently, E. Trepagnier, C. Jarzynski, F. Ritort, G. E. Crooks, C. Bustamante, and J. Liphardt, "Experimental test of Hatano and Sasa's nonequilibrium steady-state equality," Proc. Nat. Acad. Sci., **101**, 15038 (2004).

Even in this case the H-theorem can imply something meaningful only when the Boltzmann equation has a solution. The initial value problem of the Boltzmann equation is so far proved to be solvable only when the initial distribution is close to the equilibrium distribution. Therefore, even under the assumption that Boltzmann equation can describe a system, we can say that the irreversible behavior is demonstrated only for near equilibrium states.

2.5 Einstein's Equilibrium Statistical Mechanics

It is well known that in 1902-4 Einstein wrote three papers to found thermodynamics on mechanics.[73] These are virtually his first papers. From the physicist's point of view his approach, albeit not polished, may be more appealing than Gibbs's approach. As the reader will realize, the current standard way of teaching statistical mechanics (as can be seen in Kubo's problem book) is closer to Einstein's logic (although without one very crucial physics as will be seen) than Gibbs' supposedly orthodox framework.

Let us see an elementary introduction to equilibrium statistical mechanics roughly following Einstein.[74]

2.5.1 Einstein's statistical mechanics: microcanonical ensemble as starting point Einstein's starting point is the *principle of equal probability* on the energy shell. In the first paper Einstein 'derives' this with the aid of Liouville's theorem (→**2.7.2**). In the second paper Einstein 'derives' this from the generalized Liouville equation (→**3.5.1**) and the condition that only energy is the constant of motion.[75] His interpretation of probability is the relative sojourn time: the probability of a system being in a phase volume $d\Gamma \propto$ (time spent in the phase volume). That is, he is requiring that the phase space average is identical to the time average (this is the ergodic hypothesis due to Boltzmann).[76]

In short, Einstein assumes that the microscopic states (corresponding to each phase point) of an isolated system in equilibrium are all equally probable. Such a set of states with uniform probability is called a *microcanonical ensemble*.[77]

2.5.2 Introduction of canonical ensemble = canonical measure To study a system at constant temperature, Einstein regards the total system = system + thermostat as an isolated system in equilibrium.

Let H be the system Hamiltonian and H_T the Hamiltonian of the thermostat (= heat bath). Since we consider a macroscopic system, we ignore the interaction energy.[78] Let

---

[73] Ann. Phys. **9** 417 (1902); "Eine Theorie der Grundlagen der Thermodynamik," **11**, 170 (1903) and **14**, 354 (1904); One of its culminations is his Brownian motion paper →**3.1.3**.

[74] This is the closest to the best introduction, although the ergodicity requirement must be weakened considerably (→**1.5.10**).

[75] As to the impossibility of deriving the principle from mechanics, see **2.6.2**.

[76] Ergodicity is neither necessary nor sufficient condition for statistical mechanics to hold. See **1.5.9**∼. Also we now know that to interpret macroscopic observables as time averaged quantities is not a good way of understanding equilibrium statistical mechanics.

[77] It is much better to say that the set of microscopic states is equipped with a microcanonical measure (or at least microcanonical distribution).

[78] This is allowed if the molecular interaction potentials decays faster than $1/r^d$.

$\Omega(E)dE$ be the number of microstates for the system with energy in the energy shell $E \sim E + dE$, and $\Omega_T(E)$ the analogous quantity for the thermostat. Then, the the probability $p(E)$ of the system energy E to be in the energy shell[79] is proportional to

$$W(E)W_T(E_0 - E) = W(E) \exp \log W_T(E_0 - E), \tag{2.5.1}$$

where E_0 is the total energy of the system + heat bath. To evaluate the second factor we Taylor-expand the exponent as

$$\log W_T(E_0 - E) = \log W_T(T_0) - E\frac{d}{dE_0} \log W_T(E_0) + \cdots. \tag{2.5.2}$$

Thermodynamically, we know that E_0 is proportional to the size of the thermostat, and that $\log W_T$ is bounded by at most $N \log N$,[80] so the higher order derivatives are much smaller than order unity. Therefore, we may write

$$p(E) \propto W(E)e^{-\beta E}, \tag{2.5.3}$$

where $\beta = d \log W_T(E_0)/dE_0$ is a constant determined by the thermostat. We can identify it with $1/\Theta$ of Gibbs or $1/k_BT$, if we refer to the ideal gas. This is the *canonical ensemble* defined by Gibbs (→**2.1.2**).[81]

Exercise 1. Let us consider a system which is enclosed in a diathermal freely deformable membrane and which is placed in a large reservoir system. The whole system is isolated in a rigid container. Find the simultaneous distribution function of the energy and the volume of the system, mimicking Einstein. □

2.5.3 Relation between thermodynamics and statistical mechanics

An interesting point is beyond **2.5.2**. Gibbs 'did not derived' the relation between thermodynamics and statistical mechanics; he 'demonstrated' a parallelism (→**2.1.5**): 'A' $= -k_BT \log Z$ was shown to be consistent with the Helmholtz free energy A. In contrast, Einstein 'derived' $A = -k_BT \log Z$.

To make a contact between statistical mechanics and thermodynamics we need two ingredients. One is the identification of internal energy with the expectation value of the system energy. This is straightforward and not at all controversial. Since thermodynamics discusses heat, we must have a certain interpretation of heat or related quantities in terms of mechanical quantities (in terms of microscopic quantities); the reader can equivalently say that we need a microscopic interpretation of entropy.

In thermodynamics heat cannot be defined without the concept of adiabatic walls. Therefore, Einstein proposes a mechanical interpretation of adiabaticity.

Suppose the system Hamiltonian contains parameters $\{\lambda\}$ that can be controlled externally (these correspond to a_i of Gibbs in **2.1.3**). The total change of energy may be written as

$$dE = \sum_i \frac{\partial E}{\partial \lambda_i} d\lambda_i + \sum_i \frac{\partial E}{\partial y_i} dy_i, \tag{2.5.4}$$

---

[79] More precisely, $p(E)dE$ is the probability we discuss.

[80] If we properly take into account the indistinguishability of identical particles, this is bounded by N.

[81] However, Gibbs noted (on p183) that a small portion of a large microcanonical ensemble is canonical; It is likely that Gibbs knew everything Einstein wrote but refrained from writing 'theoretical physics-style hand-waving arguments.'

where y_i are the canonical coordinates. The first sum is the change of energy when the control parameters are varied, so this can be done adiabatically. Therefore, Einstein concludes that

$$dQ \equiv \sum_i \frac{\partial E}{\partial y_i} dy_i = 0 \tag{2.5.5}$$

is the adiabatic condition.[82] Then, this dQ must be the *heat*, and the first term in (2.5.4) must be the work.[83]

2.5.4 Derivation of $A = -k_B T \log Z$

Let $P = e^{c-\beta E}$ be the canonical distribution (→**2.5.2**), where c is determined by the normalization condition. Let us vary the temperature (i.e., β) and $\{\lambda\}$:

$$0 = \delta \int P dy = \int dy \left[\delta c - E\delta\beta - \beta \frac{\partial E}{\partial \lambda_i} \delta\lambda_i \right] P. \tag{2.5.6}$$

Using the microscopic expression of heat in (2.5.5), we may rewrite this as (we will write $\langle Q \rangle$ again as Q)

$$\beta dQ = -dc + d(\beta \langle E \rangle). \tag{2.5.7}$$

We identify $\langle E \rangle$ with the internal energy E, so (let us ignore the constant k_B) according to thermodynamics

$$dS = \beta dQ = d\left(\frac{E - Tc}{T} \right) \tag{2.5.8}$$

or $dc = d(A/T)$. Integrating this and ignoring the arbitrary integration constant,[84] we arrive at

$$A = -k_B T \log Z. \tag{2.5.9}$$

2.5.5 How satisfactory is Einstein's derivation of equilibrium statistical mechanics?

The crucial point of Einstein's approach **2.5.4** is his mechanical interpretation of adiabaticity in **2.5.3**. Certainly, the wall satisfying Einstein's requirement is adiabatic. However, as can be seen from its consequent definition of mechanical work, the characterization of adiabaticity is still excessively microscopic. As we will see later (→**2.7.19**), the conclusion derived from this interpretation is still contradictory to thermodynamics.

We will have to conclude (→**2.7.18**) that there is still a grave incompleteness in our understanding of the "kind of motion we call heat."

---

[82] We should regard this equivalence as the fundamental identification. Notice the vital importance of the concept of adiabatic condition.

[83] The splitting (2.5.4) is actually not required in Einstein's derivation of $A = -k_B T \log Z$. What we need is that inside (2.5.6) δE is the work part of the total energy change. No explicit form of heat as (2.5.5) is required.

[84] This can be set zero by appropriately choosing the origin of energy.

2.5.6 What can we learn from history?

(1) Initially, they wished to understand thermodynamics in terms of mechanics. However, Boltzmann clearly realized that no truly dynamical study was needed to understand thermodynamic equilibrium (→**2.4.2**). This line was perfected by Einstein and Gibbs as we have seen in **2.1.4** and up to **2.5.4** of this section.

(2) As is clear from the preface of Gibbs' book, he thought mechanics was much better established than thermodynamics that is "an approximate empirical result." However, what actually Gibbs did is to find a system compatible with mechanics and thermodynamics simultaneously. It is clear that even though Gibbs claimed mechanics was more fundamental (→**2.1.1**), without thermodynamics he could not build his statistical theory.

This is clearer from Einstein's derivation of equilibrium statistical mechanics with the aid of thermodynamics (→**2.5.3**, **2.5.4**). The subsequent history eloquently tells us that mechanics was less reliable than thermodynamics (cf. **2.7.15**). Thus, it is fair to say that although the founding fathers of statistical thermodynamics wished to construct the whole theory exclusively on mechanics, they failed (cf. **2.4.2**). This failure was actually an impetus for Boltzmann's statistical mechanics.

(3) The success of equilibrium statistical mechanics that is virtually independent of mechanics left many basic questions untouched or at best incompletely answered. For example, no one has been successful in computing thermodynamic entropy change due to irreversible processes using mechanics except for dilute gases with the aid of the Boltzmann equation (→**2.4.10**) (if we could trust it →**2.4.7**). We do not have a good microscopic expression for entropy; Einstein's expression in **2.5.1** is hardly usable.

There are many physicists who confess that they have really understood entropy after learning statistical mechanics. However, as we have seen, we still do not have a good understanding of entropy. Only those who do not carefully think about entropy may be able to feel that statistical mechanics really makes us understand entropy.

2.6 Equilibrium Statistical Thermodynamics — Brief Summary

This section summarizes important rudiments of equilibrium statistical thermodynamics as we now know it. However, the reader should not regard this section as a substitute for an introduction to the topic. The emphasis here is on the topics more or less directly related to fluctuations and dynamics. As already seen, to understand statistical mechanics, the reader needs good understanding of thermodynamics. Recommended references are repeated here.

Recommended thermodynamics references:[85]
H B Callen, *Thermodynamics* (Wiley, 1960)[86]
R Kubo, *Thermodynamics* (North-Holland, original 1951).

---

[85] In Japanese, H. Tasaki, *Thermodynamics* (Baifu-kan, 2002) is recommended; better than the books mentioned here.

[86] Do NOT confuse this with H B Callen, *Thermodynamics and an introduction to thermostatistics* (Wiley, 1985).

Recommended statistical mechanics references:

R Kubo, *Statistical Mechanics* (North-Holland, original 1951),

Landau-Lifshitz, *Statistical Mechanics* (Pergamon),

M Toda, R Kubo, and N Saito, *Statistical Physics I* (Springer Series in Solid-State Science 30, 1985)[87]

2.6.1 Fundamental postulate of statistical description

The assumption that mechanical description is the most detailed description of any system is the most fundamental assumption of statistical mechanics.
Discussion 1. If one can demonstrate that the assumption is true, then it is no more an assumption. Try to demonstrate that the above assumption is actually true. (Can the reader demonstrate that a piece of chalk is governed by mechanics, say, quantum mechanics?) Clearly recognize its metaphysical nature (= beyond any empirical demonstration).□

2.6.2 Principle of equal probability

If we could derive the fundamental relation between statistical mechanics and thermodynamics stated at the beginning of **2.1** purely in terms of mechanics, the microscopic world view could be said to be complete. However, this dream has never come true; recall that what Gibbs really showed was only that statistical mechanics is apparently consistent with thermodynamics (→**2.1.4**, **2.1.5**). Actually, this dream cannot in principle come true, because the equation of motion is not enough to describe the mechanical motion of any system; we need an initial condition that is not governed by the equation of motion. Recall that Boltzmann himself recognized the importance of initial conditions in 'demonstrating' the second law (→**2.4.2**, **2.7.10**).

Thus, we need an extra assumption that is independent of mechanics laws (cf., **2.4.12**). Usually, it is the *principle of equal probability* for an isolated mechanical system:

> On a constant energy surface $H = E$, the *a priori* probability distribution is uniform with respect to the Liouville measure, restricted to $H = E$.

Here, H is the system Hamiltonian, and the Liouville measure means the phase volume (→**2.7.2**). This principle means that the equilibrium probability of a classical state with energy E is proportional to

$$\delta(H - E)dqdp. \tag{2.6.1}$$

Here, $dqdp$ is, as usual in these notes, a short-hand notation of $\prod_i dq_i dp_i$, where i runs over all the degrees of freedom of the system (canonical coordinate pairs).
Exercise 1. Let $d\sigma$ be the surface element of the constant energy surface. Then, demonstrate that

$$\mu_E(dqdp) \propto \frac{d\sigma}{|grad\, H|}, \tag{2.6.2}$$

where μ_E is the uniform probability measure (i.e., the microcanonical measure; for a measure in general see Appendix to Chapter 1) on the constant energy surface, and *grad* is taken with respect to the phase coordinates. □

If a system is not isolated but in contact with a heat bath at constant temperature T,

---

[87] Not *Statistical Physics II*.

the Gibbs postulate is that the equilibrium probability for the system to be in the phase volume element $dqdp$ is given by[88]

$$\frac{1}{ZN!}e^{-\beta H}dqdp, \tag{2.6.3}$$

where $\beta \equiv 1/k_BT$ and the canonical partition function Z is defined by

$$Z = \frac{1}{N!}\int e^{-\beta H}d\Gamma, \tag{2.6.4}$$

with $d\Gamma = dqdp$ being the volume element of the phase space.

The relation between the canonical and the microcanonical ensembles is discussed in **2.5.2**, but rigorizing it, even if possible, is not trivial, so it is not a bad attitude to accept this as another principle.[89]

We must clearly recognize that these postulates transcend mechanics (any mechanics); they dictate the way we observe Nature and even transcend physics.[90]

However, some authors, especially authors of popular statistical mechanics textbooks (e.g., Huang), tend to assert, explicitly or implicitly, that 'from a purely logical point of view there is no room for an independent postulate.' For the reader of this notes, it should be clear that this is a statement due to a fundamental misunderstanding of nature of mechanics.

Remark. However, there is a serious attempt to derive the canonical distribution from quantum mechanics by Tasaki.[91]□

2.6.3 Postulate about expression of free energy.

Usually, we say the principle of equal probability (→**2.6.2**) is the fundamental postulate, and tend to regard it is the only postulate. However, the purpose of statistical mechanics is to study macroscopic properties of matter in terms of microscopic states, so we must specify how to interpret expectation values obtained by statistical mechanics. First of all, the identification of the expectation value of energy with the internal energy in thermodynamics must be required. This is, so to speak, the relation corresponding to the first law of thermodynamics.[92]

We need a relation between statistical mechanics and the aspect of thermodynamics related to the second law as well. As we have learned in **2.5.3**, Einstein obtained this

---

[88] As was noted in **2.5.2** this can be derived from the microcanonical ensemble in an asymptotic sense. Cf., the contraction principle **1.6.11**.

[89] There is an approach from the so-called zeroth law of thermodynamics (i.e., the transitivity of equilibrium relations): A Frigeiro, V Gorini and M Verri, "The zeroth law of thermodynamics," Physica **137**A 573-602 (1986). See **2.7.19**.

[90] If we accept quantum mechanics, the uncertainty principle seems to suggest that $\delta p\delta q \geq \hbar$ is not compatible with the states confined on a Lebesgue measure zero set (i.e., not compatible with a measure not absolutely continuous with respect to the Riemann volume of the phase space). Thus, the absolute continuity of the invariant measure with respect to the Riemann volume seems to be a consequence of quantum mechanics. This sounds like a very appealing proposal, but we must recall the meaning of the uncertainty principle: if p and q are measured for a given state of a particle, the product of the variance of observed p and that of q cannot be smaller than some positive number. This does not mean anything about the nature of the support of the distribution of p; it could be an everywhere dense measure zero set.

[91] H Tasaki, "From quantum dynamics to the canonical distribution: general picture and a rigorous example," Phys. Rev. Lett. **80**, 1373-1376 (1998).

[92] In the accompanying equilibrium lecture notes (section **1.2**), it is clearly stated that we need two principles: interpretation and statistical principles to establish statistical mechanics.

relation through microscopic identification of heat. This is based on an interpretation of adiabatic processes. This interpretation is not without problems (→**2.7.18**) and the subsequent derivation is not easily regorizable, either. Therefore, mathematically, it is safe to postulate the outcome: the Helmholtz free energy in thermodynamics is written in terms of the canonical partition function as

$$A = -k_BT\log Z. \tag{2.6.5}$$

2.6.4 Quantum equilibrium statistical mechanics

If we interpret that the internal energy is the average of the eigenvalues of the system Hamiltonian, we readily arrive at a quantum counterpart of (2.6.4):

$$Z = Tre^{-\beta H}. \tag{2.6.6}$$

This is, strictly speaking, a postulate (with (2.6.5)).

To obtain the relation between this and (2.6.4), perhaps the most elegant method is to use the path integral expression of (2.6.6) (see Feynman p77). A more elementary method may be as in Landau-Lifshitz. The correction should start with $\hbar^2$, because $\hbar$ always appears with i in quantum mechanics.

Exercise 1. $A \geq A_{\text{classic}}$? (Feynman and Landau give opposite inequalities!) □

2.6.5 Laplace transformation vs. Legendre transformation

We know that the Gibbs free energy G is obtained from A as[93]

$$G = \inf_V(A + pV) = -\sup_V((-p)V - A). \tag{2.6.7}$$

This is called *Legendre transformation.*[94] Recall that A is the *maximum work* we can gain from the system (or the *minimum work* we need to make a desired change in the system) under constant temperature.[95] Now, G is the maximum work we can gain from the system under constant temperature and pressure.

The Gibbs free energy is given by the *pressure ensemble*:

$$G = -k_BT\log\gamma(p), \tag{2.6.8}$$

where

$$\gamma(p) = \int Z(V)e^{(-p)V/k_BT}dV. \tag{2.6.9}$$

Here, $Z(V)$ is the canonical partition function with fixed volume V. Thus, we see that in statistical thermodynamics Legendre transformation corresponds to *Laplace transformation.*

Let us see why there is such a relation: (2.6.9) can be written as

$$\gamma(p) = \int dVe^{(-pV-A)/k_BT}. \tag{2.6.10}$$

---

[93] sup implies *supremum* = the smallest value that is not smaller than any values allowed to the set; inf implies *infimum* = the largest value that is not larger than any values allowed to the set. The open interval (a, b) does not have any maximum or minimum, but its infimum is a and supremum is b.

[94] Large deviation theorists call it the *Legendre-Fenchel transformation.*

[95] Never forget this constant temperature condition. The same should be true for ΔG; never forget the constancy of T and p.

The integrand is dominated by its maximum value, so we may approximate this as

$$\gamma(p) \simeq e^{\sup_V(-pV-A)/k_BT} = e^{-\inf_V(pV+A)/k_BT}. \tag{2.6.11}$$

Taking the logarithm of this formula, we obtain (2.6.7).

2.6.6 Laplace-Legendre correspondence: summary and practice

The correspondence between Laplace transformation in statistical mechanics and Legendre transformation in thermodynamics is quite a general relation that may be summarized in the following commutative diagram:

$$\begin{tikzcd}
J \arrow[r, leftrightarrow, "J=-k_BT\log Q(x)"] \arrow[d, "\mathrm{Lg}"] & Q(x) \arrow[d, "\mathrm{Lp}"] \\
K = \inf_x(J+xX) \arrow[r, leftrightarrow, "K=-k_BT\log R(X)"] & R(X) = \int dx\, Q(x)e^{-xX/k_BT}
\end{tikzcd}$$

Here, Lg implies the Legendre transformation, and Lp implies the Laplace transformation.

A practical way to find an appropriate partition function for a given thermodynamic potential may be as follows: Let us take G as an example. We know $A = G - pV$ and $Z = e^{-\beta A}$, so $Z(V) = e^{-\beta(G-pV)}$, where we clearly recognize that Z is under a fixed V. Therefore, formally, $Z(V)e^{-\beta pV} = e^{-\beta G}$, but this must be the result of the maximum term approximation of the integral $\gamma(p) \equiv \int dV Z(V)e^{-\beta pV}$. We have obtained (2.6.9).

Exercise 1. If the system is connected not only to a heat bath but to a chemostat, the appropriate thermodynamic potential is obtained by the following Legendre transformation

$$q = \inf_N(G - \mu N) = -\sup_N(\mu N - G). \tag{2.6.12}$$

(1) What is the physical (thermodynamic) meaning of q?
(2) Construct an appropriate partition function (which is called the *grand* (*canonical*) *partition function*). □

Exercise 2. We wish to study a magnet under constant magnetic field. What is a convenient partition function to study the system statistical mechanically? □

Exercise 3. The Helmholtz free energy A is obtained by a Legendre transformation from the internal energy E. How can one understand the relation between A and the canonical partition function with the aid of the above commutative diagram? □

Exercise 4. Let us consider an ideal rubber band[96] whose number of microscopic states is given by $N(L)$ when its length is L. The first law may be written as $dE = TdS + FdL$, where F is the tensile force. Under a constant force we wish to obtain the average length of the rubber band. Construct the most convenient partition function. □

We will learn that large deviation theory in probability theory is a closely related topic (→**2.5**). If the reader knows *convex analysis*, she would immediately realize that statistical mechanics is an elementary application of large deviation theory.

---

[96] Here, 'ideal' means that the total energy of the band does not depend on its length (just as the energy of an ideal gas does not depend on its volume). We can imagine an ideal rubber band as a very flexible polymer chain. Can the reader demonstrate that under an adiabatic condition relaxing the band decreases its temperature?

2.6.7 Ensemble equivalence is always correct in the thermodynamic limit

There are numerous ensembles, and it is practically important to choose the most convenient one when we study a particular system. However, the choice is strictly a matter of convenience and the obtained thermodynamics is independent of the choice of the ensemble. This is called the *equivalence of ensembles*.

The precise meaning of the equivalence of ensembles is:
In the thermodynamic limit (or for sufficiently large N) any thermodynamic observable is obtained correctly (up to $\log N$ correction to the extensive quantities; $\log N/N$ correction to the intensive quantities) irrespective of the ensemble used to compute it.

For a demonstration and details read **1.8.16-1.8.18** in IESM.

2.6.8 Orthodicity

Gibbs realized (cf. **2.1.1**) that the ensembles that are compatible with thermodynamics are not unique. We learn in elementary statistical mechanics that in the thermodynamic limit microcanonical, canonical, and grand canonical ensembles all give the same thermodynamics. Thus, their weighted averages are also compatible with thermodynamics: there are infinitely many ensembles that are compatible with thermodynamics.

Boltzmann posed the question: when can an ensemble = stationary probability distribution on the phase space (= invariant measure, we will come back to this later →**2.7.4**) be interpreted as a macroscopic state of a system governed by equilibrium thermodynamics? This problem is called the *orthodicity problem*, because Boltzmann called such ensembles *orthodic ensembles*.

We do not know what characterizes 'orthodicity'; the problem is widely open.[97]

---

[97] The problem is stressed by G Gallavotti, *Statistical Mechanics, a short treatise* (Springer, 1999). 'Complete' understanding of this question could reveal itself equivalent to the dynamical foundation of thermodynamics: the very problem that one is hoping to circumvent by deciding to only "build a mechanical model of thermodynamics, i.e., an orthodic ensemble."

Appendix: Convex Analysis

The standard textbook is:
R. T. Rockafeller, *Convex Analysis* (Princeton UP 1970)[98]

2.A.1 Convex set and convex hull

A subset C of $\boldsymbol{R}^n$[99] is a *convex set*, if

$$x \in C, y \in C \Rightarrow (1-\lambda)x + \lambda y \in C \tag{2.A.1}$$

for any $\lambda \in [0,1]$.

If $A, B \subset \boldsymbol{R}^n$ are convex, then $A \cap B$ is convex.

The *convex hull* of a set $A \subset \boldsymbol{R}^n$ is the smallest convex set containing A, and is denoted by $\mathrm{conv}A$.

Let $x_1, \cdots, x_n \in \boldsymbol{R}^n$. The convex hull of the finite set $\{x_1, \cdots, x_n\}$ is called the *convex combination* of the points in the set, and may be written as $\{\lambda_1 x_1 + \cdots + \lambda_n x_n \,|\, \lambda_1 + \cdots + \lambda_n = 1, \lambda_i \geq 0\}$.

Exercise 1. Let C be a convex set. Then, the convex combination of any n points in C is contained in C. The converse also holds.□

Exercise 2. $\mathrm{conv}A$ contains all the convex combination of the elements of A. □

Exercise 3. If A and B are convex, then $\{x + y \,|\, x \in A, y \in B\}$ is convex. □

2.A.2 Supporting hyperplane and supporting half space

Let C be an open convex set in $\boldsymbol{R}^n$.[100] If there is an affine set M such that $C \cap M = \emptyset$, then there is a hyperplane H containing M such that C is contained on one side of H. This is the Hahn-Banach theorem.

Let C be a closed convex set. Then, it is a common set of all the closed half spaces containing C.[101] This should be intuitively obvious.

A hyperplane H is called a *supporting hyperplane* of a convex set C, if the open kernel of C is contained in one side of this hyperplane and $H \cap C \neq \emptyset$ (that is, if H is 'touching C'). The closed half space defined by a supporting hyperplane and containing C is called a *supporting half space.*

2.A.3 Epigraph and convex function

Let $f : S(\subset \boldsymbol{R}^n) \to [-\infty, +\infty]$.[102] The *epigraph* of f denoted by $\mathrm{epi}f$ is defined as

$$\mathrm{epi}f = \{(x, \mu) \,|\, x \in S, \mu \geq f(x)\}. \tag{2.A.2}$$

A function on S is a *convex function*, if $\mathrm{epi}f$ is a convex set. A convex function f on S is regarded as a convex function on $\boldsymbol{R}^n$ by setting $f(x) = +\infty$ for $x \notin S$. Therefore, here, following Rockafeller, we suppress the domain of a convex function.

---

[98] reprinted as a book in the series Princeton Landmarks in Mathematics in 1997.

[99] This can be any affine space. An affine space is a space containing the line passing through any pair of the points in it.

[100] To assert this in the general affine space, the openness of the set must be more carefully defined. See Rockafeller p44.

[101] Rockafeller, Theorem 11.5.

[102] Here, $\pm\infty$ are allowed as the values of f.

Exercise 1. Let f_1 and f_2 be convex. Then $f_1 + f_2$ is convex. □
Exercise 2. Let $f : S \to \boldsymbol{R}$ be convex and $\varphi : \boldsymbol{R} \to \boldsymbol{R}$ be a non-decreasing function. Then, $\varphi \circ f$ is convex. □.

2.A.4 Bottom.[103]

Let F be a subset of $\boldsymbol{R}^{n+1}$, and S be its projection to $\boldsymbol{R}^n$. The bottom of F over S denoted by bot$F$ is a function on S defined as

$$\mathrm{bot}_x F = \mathrm{bot}F(x) = \inf\{\mu \,|\, (x, \mu) \in F, x \in S\} \tag{2.A.3}$$

If f is a convex function, then

$$f(x) = \mathrm{bot}_x(\mathrm{epi}f). \tag{2.A.4}$$

The *convex hull* of a function g denoted by conv$g$ is a function defined by

$$\mathrm{conv}g(x) = \mathrm{bot}_x(\mathrm{conv}(\mathrm{epi}g)). \tag{2.A.5}$$

Exercise 1. Let $\{f_i\}$ be a family of convex functions. Then, $f = \sup_i f_i$ is a convex function. □

2.A.5 Convex function is continuous

Let us consider a typical case (proper convex function): f is a *proper convex function*, if it is a finite convex function on an open domain C that is convex. Then, f is continuous on C.[104]

To demonstrate this, first we show that f is *lower semicontinuous*: f is lower semicontinuous at x, if $\liminf_{y \to x} f(y) = f(x)$.

The closure of a function f, denoted as cl$f$, is defined by

$$\mathrm{cl}f(x) = \mathrm{bot}_x([\mathrm{epi}f]), \tag{2.A.6}$$

where [ ] denotes the closure of the set. It is easy to see that cl$f$ is lower semicontinuous.

If f is as assumed above (proper convex function), then on C cl$f = f$, because epi$f$ is a closed set. Therefore, f is lower semicontinuous.[105]

The remaining thing is to demonstrate that for any α $\{x \,|\, f(x) < \alpha\}$ is an open set. Then, f is upper semicontinuous, so f must be continuous. To show the closedness of $\{x \,|\, f(x) \geq \alpha\}$ we note that the common set of epi$f$ and $\{(x, \mu) \,|\, \mu \geq \alpha\}$ is a closed set. Therefore, its complement must be open.

Remark. Actually, a convex function is Lipschitz continuous on a closed bounded set: f is Lipschitz continuous on any closed bounded subset in the domain of f.[106] Notice that Lipschitzian implies uniform continuity as well. □.

---

[103] This terminology is not generally used, but is convenient.

[104] For a precise statement see Rockafeller Theorem 10.1.

[105] If you wish to have a formal proof, see Rockafeller Theorem 7.1.

[106] Rockafeller Theorem 10.4; The basic idea of the proof is as follows: Let S be a closed set in the open kernel of the domain of f. Then, we can find $\epsilon > 0$ such that for any ϵ ball B $S + B$ is again in the domain. Let $x, y \in S$ and take z on the small extension of the segment connecting x to y: $z = y + \epsilon(y - x)/|y - x|$ which is in some $S + B$. Since f is convex on this $S + B$, we have

$$f(y) \leq f(x) + |y - x| \frac{f(z) - f(x)}{\epsilon + |y - x|}.$$

This inequality is valid for any x and y in S. Since f is bounded on $S + B$, this implies the Lipschitz continuity of f on S.

2.A.6 Pointwise convergence preserves convexity

Let $\{f_i\}$ be a sequence of finite convex functions on a relative open convex set C, and is pointwise convergent on a dense subset C'. Then, the sequence converges pointwisely on C to a convex function.[107]

We first note that on any closed subset S of C, the set $\{f_i\}$ is equi-Lipschitzian:[108] that is, we can find $\alpha > 0$ such that

$$|f_i(x) - f_i(y)| < \alpha|x - y| \tag{2.A.7}$$

for any $x, y \in S$. This follows from $f \equiv \sup_i f_i$ majorizing all the functions is convex, and $g(x) = \text{bot}_x\{\text{conv}(\cup \text{epi} f_i)\}$ minorizing all the functions is also convex.

Now, we make an $\epsilon/3\alpha$ net N of S (that is, any point of S is within $\epsilon/3\alpha$ of N). Then, thanks to the equi-Lipschitzian nature of $\{f_i\}$, for any $x \in S$ and $z \in C'$ within $\epsilon/3\alpha$ of x

$$|f_i(x)-f_j(x)| < |f_i(x)-f_i(z)|+|f_i(z)-f_j(z)|+|f_j(z)-f_j(x)| < 2\epsilon/3+|f_i(z)-f_j(z)|. \tag{2.A.8}$$

By assumption we may find $k \in \boldsymbol{N}$ such that $|f_i(z) - f_j(z)| < \epsilon/3$ for $i, j > k$. Therefore, for each x $\{f_i\}$ is a Cauchy sequence, so $\{f_i\}$ converges on C. Clearly, this convergence is uniform on S. The convexity inequality is preserved, so the limit is convex.

Remark. It should be clear from this that if $\{f_i\}$ is a sequence of convex functions majorized by a bounded function on a relatively open convex set C, then we may choose a uniformly converging subsequence to a convex function on closed bounded subsets of C.[109]

2.A.7 One-sided differentiability

The *one-sided directional derivative* of f at x in the direction of y is defined as

$$f'(x; y) = \lim_{\lambda \to 0+} \frac{f(x + \lambda y) - f(x)}{\lambda}, \tag{2.A.9}$$

if it exists. The two-sided directional derivative exists if and only if

$$f'(x; y) = -f'(x; -y). \tag{2.A.10}$$

Notice that if f is differentiable, then[110]

$$f'(x; y) = \nabla f(x) \cdot y. \tag{2.A.11}$$

For a convex function, the one-sided derivative exists in its domain everywhere, and

$$-f'(x; -y) \le f'(x; y) \tag{2.A.12}$$

for all y.[111]

To show the existence everywhere, we first note that the ratio defining the directional derivative is a convex function of y. Let $h(y) = f(x + y) - f(x)$. Then, we may write

$$\frac{f(x + \lambda y) - f(x)}{\lambda} = \lambda^{-1} h(\lambda y) = \text{bot}_y(\lambda^{-1}\, \text{epi}\, h). \tag{2.A.13}$$

---

[107] Rockafeller Theorem 10.8.

[108] Rockafeller Theorem10.6.

[109] Rockafeller Theorem 10.9.

[110] We do not discuss *subgradient*, so ∇ always implies the usual gradient.

[111] Rockafeller Theorem 23.1.

Since epi h contains the origin, $\lambda^{-1}\text{epi}h$[112] is an increasing set as $\lambda \to 0$. This implies that $\lambda^{-1}h(\lambda y)$ is monotone decreasing for any y as $\lambda \to 0$. This is obviously bounded from below,[113] so the directional derivative exists everywhere for any y.

Notice that $f'(x;y)$ is a convex function of y, and by definition $f'(x;0) = 0$. Therefore,

$$\frac{1}{2}f'(x;-y) + \frac{1}{2}f'(x;y) \geq 0. \tag{2.A.14}$$

This is (2.A.12).

Exercise 1. Let f be a bounded convex function on a closed interval. Let us write its right derivative as f'_+ and the left derivative f'_-.[114] Then, for any interior point of the interval $z_1 < x < z_2$, we have[115]

$$f'_+(z_1) \leq f'_-(x) \leq f'_+(x) \leq f'_-(z_2). \tag{2.A.15}$$

Moreover,

$$\lim_{z \searrow x} f'_+(z) = f'_+(x), \ \lim_{z \nearrow x} f'_+(z) = f'_-(x), \tag{2.A.16}$$

$$\lim_{z \searrow x} f'_-(z) = f'_+(x), \ \lim_{z \nearrow x} f'_-(z) = f'_-(x). \tag{2.A.17}$$

□

2.A.8 Differentiability except on null set

The fundamental theorem of the gradient of a convex function is as follows:[116]

Let f be a bounded convex function on a bounded closed set C (i.e., a proper convex function). Then, f is differentiable on the set $C' \subset C$ with $C \setminus C'$ being a null set. Furthermore, ∇f is continuous on C'.

The basic assertion is that for any vector y the directional derivative exists on $C'' \subset C$ with $C \setminus C''$ being a null set.[117] In essence all the assertions are reduced to the one-dimensional case along the line in the y direction.

First, note that when $f'(x;y)$ is continuous, $f'(x;y) = f'(x;-y)$ or

$$\liminf_{z \to x} f'(z;y) = -f'(x;-y). \tag{2.A.18}$$

This follows from: the existence of the directional derivatives implies the continuity of the one-sided directional derivatives. The computation of directional derivative for a fixed y is just that for $g(\lambda) = f(x + \lambda y)$. Therefore, Exercise in **2.A.7** implies the assertion. $C \setminus C''$ consists of points x such that $q(x) \equiv f'(x;y) + f'(x;-y) > 0$. That is, this set is a union of

$$S_k\{x \,:\, q(x) \geq 1/k, x \in C\}. \tag{2.A.19}$$

However, S_k must be a finite set.

---

[112] λA for a set $A = \{x\}$ means $\lambda A = \{\lambda x\}$. That is, uniform linear scaling of the space around the origin.

[113] because there is a hyperplane going through the origin, below which $\lambda\text{epi}h$ cannot invade

[114] $f_+(x) = f'(x;+1)$ and $f_-(x) = f'(x;-1)$.

[115] Rockafeller Theorem 24.1.

[116] Rockafellar Theorem 25.5.

[117] Rockafellar Theorem 25.4.

2.A.9 Convergence of derivative

Let f be a convex function which is finite and differentiable on an open convex set C. Let $\{f_i\}$ be a sequence of convex functions which are finite and differentiable on C such that $f_i \to f$ pointwisely on C. Then,

$$\lim_i \nabla f_i(x) = \nabla f(x) \tag{2.A.20}$$

on C. That is, if the limit is differentiable, its derivative can be computed from the limit of derivatives.[118]

We know the derivative is continuous.

2.A.10 Conjugate function and Legendre transformation[119]

Let $f : \boldsymbol{R}^n \to \boldsymbol{R}$ be a convex function. Then, epi f is a convex set, so there is a hyperplane $y = bx - \beta$, where $b \in \boldsymbol{R}^n$ and $\beta \in \boldsymbol{R}$, such that epi$f$ is above it (→**2.A.2**) That is, obviously for any x, we may choose b and β such that

$$f(x) \geq bx - \beta. \tag{2.A.21}$$

Since epi f is convex, it can be reconstructed from the totality of such hyperplanes (we know that supporting hyperplanes will do). That is,

$$F^* \equiv \{(x^*, \mu^*) \,|\, f(x) \geq x^*x - \mu* \ \text{ for } \forall x \in \boldsymbol{R}^n\} \tag{2.A.22}$$

contains the same information as f. The inequality in the definition of F^* may be rearranged as

$$\mu^* \geq x^*x - f(x), \tag{2.A.23}$$

so we can define a function f^* of x^* as

$$f^*(x^*) \equiv \sup_x [xx^* - f(x)]. \tag{2.A.24}$$

Notice that epi $f^* = F^*$. Since F^* is a convex set, f^* is a convex function, and is called the *conjugate function* of f. By construction f^* and f have exactly the same information: one can be reconstructed from the other. Indeed,

$$f(x) = \sup_{x^*} [xx^* - f^*(x^*)] \tag{2.A.25}$$

obviously from the symmetry. f and f^* are called the *conjugate pair.* Notice that

$$f_1 \leq f_2 \iff f_1^* \geq f_2^*. \tag{2.A.26}$$

and

$$f^{**} = f. \tag{2.A.27}$$

Exercise 1. Show:
(1) For $f(x) = e^x$

$$f^*(x^*) = \begin{cases} x^* \log x^* - x^* & \text{if } x^* > 0, \\ 0 & \text{if } x^* = 0, \\ +\infty & \text{if } x^* < 0. \end{cases} \tag{2.A.28}$$

(2) For $f(x) = |x|^p/p$ $(p > 1)$ $f^*(x^*) = |x^*|^q/q$ ($q > 1$ and $1/p + 1/q = 1$).
(3) Let Q be a regular matrix. Then, for $f(x) = x'Qx/2$ $f^*(x^*) = x^{*\prime}Q^{-1}x^*/2$.□

---

[118] Rockafeller Theorem 25.7.

[119] This transformation = the Legendre transformation is called the *Legendre-Fenchel transformation* in the large deviation theory community, because Legendre never considered non-smooth cases.

2.A.11 Fenchel's inequality.
Let f and f^* be a conjugate pair. Then,

$$f(x) + f^*(x^*) \geq xx^*. \tag{2.A.29}$$

This is immediate from (2.A.24). The equality holds if and only if $f(x) = xx^*/2$. This can be explicitly checked.

2.7 Mechanics vs. Thermodynamics

Let us summarize basic analytical mechanics and quantum mechanics relevant to statistical mechanics. Needless to say, this section cannot substitute any introductory course of the topic. It is a good lesson to learn that although the founding fathers of statistical mechanics thought mechanics was fundamental and reliable, it was thermodynamics that was much more reliable.

Recommended textbooks of mechanics are
Landau-Lifshitz, *Mechanics*,
J J Sakurai or G Baym, *Quantum Mechanics*.

2.7.1 Classical mechanics
Let q, p collectively denote the positions and momenta of (point) particles in a system, and the system Hamiltonian $H = K + U$, where K is the kinetic energy and U the potential energy. Then, the equation of motion of particles is written as

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}, \ \frac{dp}{dt} = -\frac{\partial H}{\partial q}. \tag{2.7.1}$$

This is called the *canonical equation of motion*. These equations can be written in a symmetric form with the aid of the *Poisson bracket*:

$$[A, B]_{PB} \equiv \sum_i \left(\frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} - \frac{\partial B}{\partial q_i}\frac{\partial A}{\partial p_i} \right) \tag{2.7.2}$$

as

$$\frac{dq}{dt} = [q, H]_{PB}, \ \frac{dp}{dt} = [p, H]_{PB}. \tag{2.7.3}$$

In general, if F is a *phase function* (a function of the canonical coordinates q, p, and time t) and is differentiable,

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + [F, H]_{PB}. \tag{2.7.4}$$

This is the *Hamilton's equation of motion* for a phase function (dynamical observable) F.

2.7.2 Liouville's theorem

Let $\{q(t), p(t)\}$ be a solution to the equation of motion (2.7.1) whose initial condition is given by $\{q(0), p(0)\}$. Then, the following Jacobian is unity:

$$\frac{\partial(q(t),p(t))}{\partial(q(0),p(0))} = \begin{vmatrix} \partial q(t)/\partial q(0) & \partial q(t)/\partial p(0) \\ \partial p(t)/\partial q(0) & \partial p(t)/\partial p(0) \end{vmatrix} = 1. \tag{2.7.5}$$

This is called *Liouville's theorem* (see Exercise 1 below).[120] Note that even if the Hamiltonian is time-dependent, the assertion is true.

This may also be interpreted that the Hamiltonian flow is incompressible.

Exercise 1. Demonstrate the incompressibility of the Hamiltonian flow. (This is a demonstration of Liouville's theorem.)□

Exercise 2. Suppose the system is not Hamiltonian, but its equations for the momenta are modified as

$$\frac{dp}{dt} = [p, H] - \alpha p, \tag{2.7.6}$$

where $\alpha > 0$ is a constant. What happens to the phase volume? □

2.7.3 Time evolution of distribution on phase space

Suppose A_0 is a set with a finite volume in the phase space of a Hamiltonian mechanics system. Studying the solutions of the equation of motion with the initial conditions chosen in A_0, we may study the fate of this volume flowing in the phase space riding on the dynamics.

Let μ_0 be a measure[121] assigned to the phase space at time $t = 0$. That is, the reader may regard $\mu_0(A)$ be the number of systems whose states are in A for the initial ensemble.

Let us introduce the *time evolution operator* T_t such that $T_t x(0) = x(t)$, where $x(0)$ is the initial coordinates of the system under consideration in its phase space, and $x(t)$ those at time t. For a set A, we can introduce $T_t^{-1}A \equiv \{y : T_t y \in A\}$, that is, the set of all the points that visit A in time t.

The number of systems in the ensemble does not change, so in order to determine the time evolution of the measure we may require

$$\mu_0(T_t^{-1}A) = \mu_t(A) \tag{2.7.7}$$

for any set A of finite measure in the phase space. This is the equation of motion for the distribution (= ensemble).

Exercise 1. Why does the invariance condition (2.7.7) use T_t^{-1} instead of T_t itself to write $\mu_0(A) = \mu_t(T_t A)$? [Hint: Consider the case when the trajectories can merge. It is true that this never happens within a finite time for a system described by a differential equation, but if we consider reduced systems (e.g., dynamics projected on a subspace of its phase space), we often encounter such cases.]□

---

[120] This implies that the Lebesgue measure of the phase space is an invariant measure of a Hamiltonian system. →**2.7.4**

[121] See Appendix to Chapter 1, if the reader wishes to have some notion of this concept. However, here the reader may simply understand it as a weighted volume, or weight; $\mu(A)$ implies the (statistical) weight of the set A.

2.7.4 Invariant distribution = invariant measure

If a measure μ (= weight of sets) satisfies

$$\mu(T_t^{-1}A) = \mu(A) \tag{2.7.8}$$

for any set A, this weight does not change in time. If μ satisfies (2.7.8), we say μ is an *invariant measure* (stationary distribution, stationary measure) for the time evolution T_t.

Liouville's theorem **2.7.2** tells us that the ordinary phase volume (= Lebesgue measure of the phase volume) is an invariant measure of any Hamiltonian system. If there were only one such measure, then statistical mechanics could have been founded firmly on mechanics, but generally speaking there are infinitely many (uncountably many) invariant measures, so we must confront the selection problem. The selection is done by the way we set up the set of initial conditions. This is the reason why we need an extra-mechanical assumption (→**2.6.2**).

Exercise 1. Suppose $\dot{\boldsymbol{y}} = \boldsymbol{G}$ such that its i-component G_i does not depend on y_i. Then, the Lebesgue measure of the phase space $dy_1 \cdots dy_n$ is obviously invariant. □.[122]

2.7.5 Time reversal symmetry of mechanics

In classical mechanics, we know that even if the time is reversed: $(t, q, p) \to (-t, q, -p)$, the equation of motion is intact. Therefore, a movie of a mechanical motion played backward is again a solution to the equation of motion; to obtain it we have only to flip the directions of all the momenta in the final state of the original trajectory and solve the equation of motion with this initial condition. That is, there is a one-to-one correspondence between a trajectory and its time reversed counterpart. This is the implication of the *time reversal symmetry* in classical mechanics.

Consider a bunch of trajectories (for a time span t) starting from a set V. Let the set of the totality of the end points of this bunch of trajectories be W. We can also imagine the time reversed bunch of trajectories.

Intuitively speaking, the number of trajectories in both bunches are identical due to the one-to-one correspondence between a trajectory and its time reversed counterpart. Therefore, if each trajectory is evenly weighted when being counted, then according to the natural motion of the system

$$P(V \to W) = P(\tilde{W} \to \tilde{V}), \tag{2.7.9}$$

where P is the measure of the initial conditions for the bunches and the tilde implies the time reversal operation (i.e., $p \to -p$; formally $\tilde{W} = \{(q, -p) : (q, p) \in W\}$). The principle of equal probability (→**2.6.2**) just requires an even weighting of trajectories, so in equilibrium states (2.7.9) holds.

In quantum mechanics, the transition probability from state $|a\rangle$ to $|b\rangle$ is written as $P_{a\to b} = |\langle b|T_t|a\rangle|^2$, where T_t is the time evolution operator. T_t is a unitary operator, so $P_{a\to b} = |\langle a|T_{-t}|b\rangle|^2$ which is indeed the transition probability from $|b\rangle$ to $|a\rangle$ with the time reversed motion.

2.7.6 Detailed balance

Let A and B be two sets in the phase space of a Hamiltonian system. Also let us assume

---

[122] An application may be found in F Legoll and R Monneau, "Designing reversible measure invariant algorithm with applications to molecular dynamics," J. Chem. Phys. **117**, 10452-10464 (2002).

these sets are invariant under time reversal symmetry. That is, if (q, p) is in A (in B), $(q, -p)$ is also in A (in B).[123] Let A_B be the set of all the points in A that visit B in t, and B_A the set of all the points in B that visit A in t (What are they in terms of T_t?).

Let the time reversed set of A_B be $\tilde{A_B}$. In t this set goes into B, so $\tilde{A_B}$ is T_tB_A, where T_t is the time evolution operator.

Let us reverse all the momenta in A_B and make $\tilde{A_B}$(→**2.7.5**). Since A is invariant under time reversal, $\tilde{A_B}$ is also in A. In t this set goes into B, so $\tilde{A_B}$ is T_tB_A, where T_t is the time evolution operator. Let μ be a stationary distribution (invariant measure) on the phase space. Then, obviously, $\mu(A_B) = \mu(\tilde{A_B}) = \mu(B_A)$.

If we use the conclusion of **2.7.5**, $\mu(A_B) = \mu(\tilde{B_A})$.

If we make the ratio $\mu(A_B)/\mu(A)$, this is the transition probability from A to B in t. Let us write

$$T(A \to B) = \mu(A_B)/\mu(A),\ T(B \to A) = \mu(B_A)/\mu(B). \tag{2.7.10}$$

Then, $\mu(A_B) = \mu(B_A)$ can be rewritten as

$$\mu(A)T(A \to B) = \mu(B)T(B \to A). \tag{2.7.11}$$

If this holds for any t, it is called the *detailed balance condition* for μ. For microcanonical and canonical distributions, this holds. Its essence is just $\mu(A_B) = \mu(B_A)$. However, do not forget that A and B must be time reversal symmetric.[124]

The detailed balance condition is extremely important in equilibrium.[125] For example, if a chemical reaction system is in equilibrium, all the reactions are canceled in detail by their reverse reactions. A quasiequilibrium process is retraceable thanks to this detailed balance condition. The importance of detailed balance was first emphasized by Onsager.[126] However, detailed balance alone does not imply that the distribution is in equilibrium.

2.7.7 Liouville's equation

If the measure μ_t (= probability distribution) has a density $f(q, p, t)$ such that

$$f(q, p, t)dqdp = \mu_t(dqdp), \tag{2.7.12}$$

then, thanks to *Liouville's theorem* (cf. **2.7.2**), we have

$$f(q, p, t) = f_0(T_t^{-1}q, T_t^{-1}p). \tag{2.7.13}$$

---

[123] If we are interested in a set corresponding to a macroscopic state, then this is not a very strong constraint, especially in equilibrium.

[124] A more general version is in **X1.11.1**.

[125] The mechanical detailed balance implies the detailed balance at the macroscopic level. However, even if there is no mechanical (microscopic) detailed balance, there may still be a detailed balance condition for macroscopic processes.

[126] L Onsager, "Reciprocal relations in irreversible processes I," Phys. Rev. **37**, 405-426 (1931).

Now, let us assume the infinitesimal time evolution (→(2.7.3))

$$f(q,p,t+dt) = f(T_{dt}^{-1}q, T_{dt}^{-1}p, t) = f(q-[q,H]_{PB}dt, p-[p,H]_{PB}dt, t), \tag{2.7.14}$$

$$= f(q,p,t) - \frac{\partial f(q,p,t)}{\partial q}[q,H]_{PB}dt - \frac{\partial f(q,p,t)}{\partial p}[p,H]_{PB}dt, \tag{2.7.15}$$

$$= f(q,p,t) - [f(q,p,t),H]_{PB}dt. \tag{2.7.16}$$

That is,

$$\frac{\partial f(t)}{\partial t} = [H, f(t)]_{PB}, \tag{2.7.17}$$

which is called *Liouville's equation.*[127]

If the ensemble is time-independent, the corresponding measure (= weight) $\mu = \mu_t$ must be time-independent (i.e., an invariant measure →**2.7.4**)), and its density f must also be time independent: $[f,H]_{PB} = 0$.

2.7.8 Active and passive time evolutions

Compare the Liouvilles equation (→**2.7.7**) and the Hamilton's equation of motion (→**2.7.1**). If we use the time evolution T_t introduced in **2.7.3**, then we realize (see (2.7.13))

$$f(q,p,t) = f(q(-t),p(-t),0) = T_{-t}f(q,p,0). \tag{2.7.18}$$

Therefore, the ensemble average at time t of an observable F denoted by $\langle F\rangle_t$ can be written as

$$\langle F\rangle_t = \int dqdp\, F(q,p)f(q,p,t) = \int dqdp\, F(q,p)f(T_{-t}q, T_{-t}p, 0). \tag{2.7.19}$$

Using Liouville's theorem, we can rewrite this as

$$\langle F\rangle_t = \int dqdp\, F(q(t),p(t))f(q,p,0) = \langle F_t\rangle_0, \tag{2.7.20}$$

where F_t is the observable F at time t and $\langle\ \rangle_0$ is the average over the initial ensemble.

In (2.7.19) the dynamical observable F does not evolve in time. That is, the values of F are averaged over the time-evolving distribution. This is the *passive* description (passive picture) of the system dynamics.

In contrast, in (2.7.20) the dynamical observable evolves in time, but the probability measure to average it is as given at the initial time. Therefore, this description of dynamics is called the *active* description (active picture).

As the reader realizes the passive picture corresponds to the Schrödinger picture and the active to the Heisenberg picture in quantum mechanics (→**2.7.12**). As she will learn later, in the theory of stochastic processes Kolmogorov's forward equation (Fokker-Planck equation →**??**) corresponds to the passive picture and Kolmogorov's backward equation (→**X1.9.2**) to the active one.

---

[127] This can be written as $df/dt = 0$.

2.7.9 BBGKY hierarchy

The Liouville equation (→**2.7.7**) is for all the particles in the system. Suppose the system consists of N identical particles of mass m interacting with each other through binary interactions whose potential may be written as $\phi(|\boldsymbol{q}|)$, where $\boldsymbol{q}$ is the vector connecting two interacting particles. Its Hamiltonian is

$$H = \sum_i \frac{p_i^2}{2m} + \frac{1}{2}\sum_{i \neq j} \phi(|\boldsymbol{q}_i - \boldsymbol{q}_j|). \tag{2.7.21}$$

Therefore, the Liouville equation (→**2.7.7**) for the whole N particle system reads

$$\frac{\partial f_N}{\partial t} = -\sum_i \frac{\boldsymbol{p}_i}{m} \cdot \frac{\partial}{\partial \boldsymbol{q}_i} f_N + \frac{1}{2}\sum_{i \neq j} \nabla_i \phi \cdot \left(\frac{\partial}{\partial \boldsymbol{p}_i} - \frac{\partial}{\partial \boldsymbol{p}_j} \right) f_N, \tag{2.7.22}$$

where f_N is the probability density distribution on the phase space of the N particles.

On the other hand, we know that the Boltzmann equation (→**2.4.7**) is an equation for a single body density distribution function (an equation on the μ-space). The single body distribution may be derived as the following marginal distribution, because the first particle may be used as a representative:

$$\text{‘}f_1(q_1, p_1, t)\text{’} = \int dq_2 dp_2 dq_3 dp_3 \cdots f_N(\{p, q\}, t). \tag{2.7.23}$$

However, in the thermodynamic limit $V \to \infty$ (with $N/V = n$ being kept constant), this quantity vanishes.[128] Therefore, it is convenient to define the single body distribution with a multiplication of V:

$$f_1(q_1, p_1, t) \equiv V \int dq_2 dp_2 dq_3 dp_3 \cdots f_N(\{p, q\}, t). \tag{2.7.24}$$

Let us integrate out the variables for $i \geq 2$ from the Liouville equation to derive the equation for f_1.

$$\frac{\partial}{\partial t} f_1 = -\frac{\boldsymbol{p}_1}{m} \cdot \frac{\partial}{\partial \boldsymbol{q}_1} f_1 + \sum_{j(\neq 1)} \int \nabla_1 \phi(|\boldsymbol{q}_1 - \boldsymbol{q}_j|) \cdot \left(\frac{\partial}{\partial \boldsymbol{p}_1} - \frac{\partial}{\partial \boldsymbol{p}_j} \right) \frac{1}{V} f_2(\boldsymbol{q}_1, \boldsymbol{p}_1, \boldsymbol{q}_j, \boldsymbol{p}_j, t) d\boldsymbol{q}_j d\boldsymbol{p}_j, \tag{2.7.25}$$

where

$$f_2(\boldsymbol{q}_1, \boldsymbol{p}_1, \boldsymbol{q}_j, \boldsymbol{p}_j, t) \equiv V^2 \int \overbrace{dq_2 dp_2 dq_3 dp_3 \cdots dq_N dp_N}^{\text{without } dq_j dp_j} f_N(\{p, q\}, t). \tag{2.7.26}$$

Using the permutation symmetry wrt the particles, we arrive at

$$\frac{\partial}{\partial t} f_1 = -\frac{\boldsymbol{p}_1}{m} \cdot \frac{\partial}{\partial \boldsymbol{q}_1} f_1 + n \int d\boldsymbol{q}_2 d\boldsymbol{p}_2 \, \nabla_1 \phi(|\boldsymbol{q}_1 - \boldsymbol{q}_2|) \cdot \left(\frac{\partial}{\partial \boldsymbol{p}_1} - \frac{\partial}{\partial \boldsymbol{p}_2} \right) f_2(\boldsymbol{q}_1, \boldsymbol{p}_1, \boldsymbol{q}_2, \boldsymbol{p}_2, t), \tag{2.7.27}$$

---

[128] See, for example, for an ideal gas $\int d\boldsymbol{p} f_1 = 1/V$.

where n is the particle number density.

Since (2.7.27) is not a closed equation for f_1, we need an equation for f_2. As the reader easily guesses, the equation for f_2 will not be closed, because it depends on the three-body distribution function f_3 defined analogously as (2.7.26). This way we obtain a hierarchy of equations for reduced distribution functions. The hierarchy is called the *BBGKY-hierarchy.*[129]

2.7.10 Derivation of Boltzmann equation from BBGKY hierarchy

To study f_1 to the lowest order in n, we need not worry about the higher order contribution to f_2, so we may regard the equation for f_2 be closed. Let $T_t^{(2)}$ be the time evolution operator for the two body dynamics. Then,

$$f_2(\{q,p\},t) = T_{-t}^{(2)} f_2(\{q,p\},0), \tag{2.7.28}$$

because in the Liouville equation the Poisson bracket with H appears as $[H,f]_{PB}$ in contrast to $[f,H]_{PB}$ in Hamilton's equation of motion (→**2.7.8**).

Assume that the initial two-body distribution is given by

$$f_2(\{q,p\},0) = f_1(\boldsymbol{q}_1,\boldsymbol{p}_1,0) f_1(\boldsymbol{q}_2,\boldsymbol{p}_2,0). \tag{2.7.29}$$

This is to assume that colliding particles are statistically independent (the so-called Stosszahlansatz →**2.4.5**) at the initial time.[130] Now, (2.7.27) reads

$$\frac{\partial}{\partial t} f_1 = -\frac{\boldsymbol{p}_1}{m}\cdot\frac{\partial}{\partial \boldsymbol{q}_1} f_1 + n\int d\boldsymbol{q}_2 d\boldsymbol{p}_2\, \nabla_1\phi(|\boldsymbol{q}_1-\boldsymbol{q}_2|)\cdot\left(\frac{\partial}{\partial \boldsymbol{p}_1}-\frac{\partial}{\partial \boldsymbol{p}_2}\right) T_{-t}^{(2)} f_1(\boldsymbol{q}_1,\boldsymbol{p}_1,0) f_1(\boldsymbol{q}_2,\boldsymbol{p}_2,0), \tag{2.7.30}$$

This is equivalent to the Boltzmann equation. The rest is to rewrite this in a more appealing form.

2.7.11 Derivation of Boltzmann equation: technical part

Thanks to the 2-body Liouville equation, notice that

$$\begin{aligned} \nabla_1\phi(|\boldsymbol{q}_1-\boldsymbol{q}_2|)\cdot\left(\frac{\partial}{\partial \boldsymbol{p}_1}-\frac{\partial}{\partial \boldsymbol{p}_2}\right) T_{-t}^{(2)} &= \left(\boldsymbol{v}_1\frac{\partial}{\partial \boldsymbol{q}_1}+\boldsymbol{v}_2\frac{\partial}{\partial \boldsymbol{q}_2}\right) T_{-t}^{(2)}, && (2.7.31) \\ &= (\boldsymbol{v}_1-\boldsymbol{v}_2)\cdot\frac{\partial}{\partial \boldsymbol{q}_{12}} T_{-t}^{(2)}, && (2.7.32) \end{aligned}$$

where $\boldsymbol{q}_{12} = \boldsymbol{q}_1 - \boldsymbol{q}_2$. Therefore, the collision terms reads

$$-n\int d\boldsymbol{q}_2 d\boldsymbol{p}_2\, \boldsymbol{g}\cdot\frac{\partial}{\partial \boldsymbol{q}_{12}} T_{-t}^{(2)} f_1(\boldsymbol{q}_1,\boldsymbol{p}_1,0) f_1(\boldsymbol{q}_2,\boldsymbol{p}_2,0). \tag{2.7.33}$$

If we perform the integration over $\boldsymbol{q}_2$, we can rewrite this expression with the aid of Gauss's theorem as

$$n\int d\boldsymbol{p}_2 d\boldsymbol{S}\cdot \boldsymbol{g} T_{-t}^{(2)} f_1(\boldsymbol{q}_1,\boldsymbol{p}_1,0) f_1(\boldsymbol{q}_2,\boldsymbol{p}_2,0). \tag{2.7.34}$$

---

[129] Bogoliubov-Born-Green-Kirkwood-Yvon hierarchy.

[130] This probabilistic assumption smuggles in irreversibility.

Here, $d\boldsymbol{S}$ is the surface element of a sphere centered at particle 1 with a radius small compared with the mean free path length but large enough to include the interaction range of φ.

Up to the surface from outside, the two-body dynamics is a simple ballistic motion, so we may write the above formula as

$$n \int d\boldsymbol{p}_2 d\boldsymbol{S} \cdot \boldsymbol{g} f_1(\boldsymbol{q}_1, \boldsymbol{p}_1, t) f_1(\boldsymbol{q}_2, \boldsymbol{p}_2, t). \tag{2.7.35}$$

There are two contributions from the surface integral $\boldsymbol{g} \cdot d\boldsymbol{S} <$ or > 0. They correspond, respectively, to pre and post collision configurations. Thus, we have arrived at the Boltzmann equation (→**2.4.7**).

2.7.12 Quantum mechanics

The equation of motion corresponding to Hamilton's equation of motion (→**2.7.1**) in quantum mechanics is *Heisenberg's equation of motion*:

$$i\hbar \frac{dF_t}{dt} = [F_t, H], \tag{2.7.36}$$

where [,] is the commutator, and H is the energy operator (Hamiltonian). This may be obtained by the canonical quantization from (2.7.4): the commutator is $i\hbar$ times the result of Poisson bracket (→**2.7.1**), if we interpret classical phase functions as observables.

The time evolution operator U_t is defined by the following operator equation (called the *Schrödinger equation*)

$$i\hbar \frac{dU_t}{dt} = HU_t \tag{2.7.37}$$

with the initial condition $U_0 = 1$. Then, the solution to (2.7.36) with the initial condition F_0 at $t = 0$ can be written as[131]

$$F_t = U_t^{-1} F_0 U_t. \tag{2.7.38}$$

If we introduce the state ket at time t as $|t\rangle$, we have $U_t|0\rangle = |t\rangle$. Therefore,

$$\langle t|F(0)|t\rangle = \langle 0|U_t^{-1} F(0) U_t|0\rangle, \tag{2.7.39}$$

so[132] (cf. (2.7.20) for the classical counterpart)

$$\langle t|F(0)|t\rangle = \langle 0|F(t)|0\rangle. \tag{2.7.40}$$

2.7.13 Density operator and von Neumann's equation

Let ρ_t be an operator, called a *density operator*, that gives the expectation value of an observable F at time t through

$$\langle F\rangle_t = Tr\, \rho_t F. \tag{2.7.41}$$

---

[131] If we introduce $A^\times B = [A, B]$, then $dF/dt = A^\times F$ can be formally solved as $F(t) = \exp(tA^\times)F(0)$. Assuming that t is not large, we can expand this as

$$F(t) = 1 + tA^\times F(0) + \cdots = (1 + tA + \cdots)F(0)(1 - tA + \cdots).$$

This gives the desired result.

[132] This is the relation between the *Schrödinger picture* and the *Heisenberg picture*.

If we require (cf. **2.7.8**)

$$\langle F\rangle_t = Tr\,\rho_0 F_t \tag{2.7.42}$$

for ρ_0 for any observable F,[133]

$$\rho_t = U_t\rho_0 U_t^{-1}. \tag{2.7.43}$$

Therefore, the time dependence of the density operator must be governed by

$$i\hbar\frac{d\rho_t}{dt} = [H,\rho_t]. \tag{2.7.44}$$

This is the quantum counterpart of Liouville's equation (→**2.7.7**) and is called *von Neumann's equation.*

2.7.14 Gibbs' entropy is invariant

Gibbs proposed the following microscopic interpretation of entropy as we have seen in **2.1.5**

$$S = -k_B\int \rho\log\rho\, d\Gamma, \tag{2.7.45}$$

where ρ is the density distribution function of microscopic states on the phase space. We know in the thermodynamic limit that this is in agreement with the Boltzmann entropy (→**2.4.2**; as we have seen, the formula was fist written down by Boltzmann).

An interesting feature of this formula is that this entropy is a constant of motion: $dS/dt = 0$. Let us check this invariance.

$$\frac{dS}{dt} = \int d\Gamma(1+\log\rho)\frac{\partial\rho}{\partial t}, \tag{2.7.46}$$

so with the aid of Liouville's equation (→**2.7.7**) and an integration by parts we obtain

$$\frac{dS}{dt} = \int d\Gamma(1+\log\rho)[\rho,H]_{PB} = \int d\Gamma\frac{\partial\rho}{\partial t} = 0. \tag{2.7.47}$$

What does this mean?

Exercise 1. Demonstrate the quantum version. □

Exercise 2. The reader must immediately see that the Boltzmann entropy $S_B = k_B\log\Omega$ is also invariant. (If the phase volume Ω is replaced by the number of states in quantum mechanics, the assertion does not change.)[134] □

Comment. In classical mechanics suppose the measure does not have a density.[135] Then, neither Gibbs nor Boltzmann entropy can be defined. For a closed (isolated) classical mechanical system, unless the initial measure is singular, this never happens.

---

[133] If (2.7.43) is assumed, certainly (2.7.42) holds. Is the choice unique? We need more conditions such as the KMS condition (→**X2.4.11**) that is assumed to be true in equilibrium (or is used to characterize the equilibrium state).

[134] However, if Ω is replaced by $\Omega(E_t)$, that is, the phase volume of the set of phase points whose energy does not exceed E_t at time t, this entropy increases under nonequilibrium adiabatic condition. Then, is this a better microscopic interpretation of entropy? Another possibility is to interpret $\Omega(E)$ as the phase volume of the energy shell of energy E.

[135] i.e., the measure is not absolutely continuous wrt the phase volume

2.7.15 Thermodynamics as precursor of quantum mechanics

We know that Quantum Revolution left thermodynamics intact. It is fair to say that thermodynamics paved the road to Quantum Revolution as can be seen from Planck's derivation of black-body radiation law. Gibbs' paradox[136] may be considered as an example of the incompatibility of classical picture of atoms and thermodynamics, if we stick to his statistical framework.

Titeica (1955[137]) demonstrated the following. Let $p(E)$ be the statistical weight of states up to the energy E with the lowest possible energy being taken as its origin. Then, unless

$$\lim_{E\to 0} p(E) > 0, \tag{2.7.48}$$

the entropy is not bounded from below in the $T \to 0$ limit. That is, classical mechanics contradicts thermodynamics, because classically the above limit must vanish. The zero-point energy represents the discontinuous onset of the energy spectrum of the whole system.

Exercise 1. The energy spectrum of a single particle may not have this discontinuity at $E = 0$. Still, the energy spectrum of the whole many-body system must have this discontinuity. Check that this is indeed the case for fermions or bosons (for weakly interacting quantum gases).□

2.7.16 Second law and equilibrium statistical mechanics

The driving motivation for Boltzmann when he was around 20 years old was to understand the second law in terms of mechanics.[138] Boltzmann wished to have a clear picture of how a system reaches its equilibrium state (→**2.4.2**).

However, to demonstrate the second law we need not worry about how a system relaxes to its equilibrium state. We have only to show, e.g., Thomson's principle:

$$\Delta A \le W, \tag{2.7.49}$$

that is, the work needed to change a system from an equilibrium state a to another equilibrium state b cannot be smaller than the free energy difference $\Delta A = A_b - A_a$, where A_x is the free energy of state x.

Exercise 1 [Adiabatic small systems are never reversible]. Read K Saito et al, "Irreversibility resulting from contact with a heat bath caused by the finiteness of the system," Phys. Rev. E **66**, 016119 (2002), and summarize their argument to demonstrate that for almost all small systems adiabatic processes cannot be reversible.

However, thermodynamic adiabaticity and the adiabatic condition used in this paper are different, so the conclusion of the paper should be taken with caution. □

2.7.17 Equilibrium statistical mechanics implies second law

Consider a quantum system with the Hamiltonian $H(\lambda)$ with a parameter. The initial state is with $\lambda = 0$ and the final one is with $\lambda = 1$. $\{|i_\lambda\rangle\}$ is the totality of the eigenstates of $H(\lambda)$. We assume $H(\lambda)|i_\lambda\rangle = E_i(\lambda)|i_\lambda\rangle$ and $E_i(\lambda) \le E_{i+1}(\lambda)$ for all $\lambda \in [0, 1]$.

---

[136] Distinguishability of particles violates extensivity of entropy.

[137] Einstein died (Mar 14, 1879 ∼ Apr 18).

[138] His first long paper (his second paper) "Über die mechanische Bedeutung des zweiten Hauptsatzes der Wärmetheorie" (Wiener Berichte **53**, 195-220 (1886)) was written when he was 22. This year Liszt died; Jeronimo surrendered; Statue of Liberty erected (what an irony!).

If we change the parameter sufficiently slowly (adiabatic change), the adiabatic theorem tells us that there would be no level crossing, so $|i_0\rangle$ would end up with $|i_1\rangle$ for all i. Suppose the initial distribution is given by the canonical distribution (i.e., the density matrix is given by $\rho(0) = e^{-\beta H(0)}/Z$). Let U be the time evolution operator corresponding to the parameter change from $\lambda = 0$ to 1. The density matrix at the end of this process is given by $\rho_U = U\rho(0)U^{-1}$. Then,[139]

$$\langle H\rangle_U \geq \langle H\rangle_{ad}, \tag{2.7.50}$$

where $\langle\ \rangle_U$ is the average over the final density matrix ρ_U and $\langle\ \rangle_{ad}$ is the average after the adiabatic evolution.

If the initial average of energy $Tr\rho(0)H(0)$ is subtracted from this, we obtain

$$W \geq \Delta A. \tag{2.7.51}$$

Therefore, the second law is a consequence of equilibrium statistical mechanics.[140]

2.7.18 Physical vs. mechanical adiabatic condition

However, there is one thing we have not understood. For the quantum adiabatic theorem to hold, the necessary time for the time evolution is of order $\hbar/\delta E$, where δE is the suitably averaged spacing of the adjacent energy levels. This spacing is exponentially small for many-body systems $\sim 10^{-N}$, where N is the number of degrees of freedom. That is, we can never realize such a slow change for reasonably macroscopic systems.

Still, in reality, we know the inequality in (2.7.51) holds for practically slow processes, so to really demonstrate the second law, we must be able to demonstrate that the equality in (2.7.51) holds for much faster changes than required for the adiabatic theorem to hold. We have no clear theory for this. In short, we have not yet understood the nature of the motion we call heat.

2.7.19 Canonical ensemble is 'implied' by the second law[141]

Conversely, Lenard demonstrated that essentially the second law and the fourth law[142] of thermodynamics imply the canonical ensemble.

More precisely, the assumptions are:

(1) *Passivity*: This is essentially the second law. Let $H(t)$ be a time-dependent Hamiltonian and the density operator $\rho(t)$ obeys von Neumann's equation with this Hamiltonian with no time dependence for t outside $[0,1]$ and $H(0) = H(1)$. Then, the work needed for this cycle is nonnegative:

$$W(K,\rho_0) = \int_0^1 dt\, Tr\rho(t)\frac{dH(t)}{dt} \geq 0 \tag{2.7.52}$$

(2) Structural stability: any small perturbation of the Hamiltonian does not destroy the system passivity.

(3) Let a compound system consist of two subsystems and in a passive structurally stable

---

[139] See, e.g., H Tasaki, *Statistical mechanical derivation of the second law of thermodynamics*, cond-mat/0009206. The statement is more general.

[140] However, obviously, we cannot take the thermodynamic limit before the proof. Therefore, we must take the statement with caution that equilibrium statistical framework implies the second law.

[141] A Lenard, "Thermodynamical proof of the Gibbs formula for elementary quantum systems," J. Statist. Phys. **19**, 575 (1978).

[142] "There are only extensive and intensive quantities in thermodynamics."

state. Its density operator is the product[143] of the density operators of the two subsystems.

(1) implies that H and ρ are commutative and the simultaneous eigenvalues have a consistent order: the eigenvalue of ρ is larger for the smaller energy eigenvalue. This is actually the converse of **2.7.16**.
(2) then implies that $\rho = f(H)$, i.e., the density operator is an operator function of H. Finally, (3) implies that $\log \rho$ is proportional to H.

Combining **2.7.16** and the above result, the reader might conclude that the orthodicity question (→**2.6.8**) is solved. This contradicts our knowledge that the canonical ensemble is not the only orthodic ensemble. What is wrong? A hint has already been discussed at the end of **2.7.16**. The passivity is unrealistically strict. Perhaps this is in accordance with the unrealistically strict adiabatic condition (→**2.7.18**). We need a much more realistic expression of the second law than (2.7.52).

Remark. The zeroth law can be used to obtain the canonical form as well.[144] The key idea is that the equilibrium condition is not destroyed by small perturbations of the system Hamiltonian.
(1) If the state ρ is stable under small perturbations of the system Hamiltonian, then $\rho = f(H)$ (ρ is an operator function of H).
(2) If the equilibrium contact condition is stable under small perturbations, then the eigenvalues have the consistent order (see (1) above; **2.7.16** implies that the systems are passive).
(3) If the equilibrium contact among three systems is stable under small perturbations, then $\rho \propto e^{-\beta H}$. □

2.7.20 Jarzynski's equality[145]

Another related theorem is due to Jarzynski. This is an equality that leads to the second law (2.7.49).

Suppose a thermostated system has a parameter $\lambda(t)$ that can be changed experimentally externally from $\lambda(0)$ to $\lambda(1)$. Let W be the work needed to perform this change. Then, the free energy change ΔA from the equilibrium state with $\lambda(0)$ and to that with $\lambda(1)$ may be obtained by

$$\langle e^{-\beta W}\rangle = e^{-\beta \Delta A}, \tag{2.7.53}$$

where the average is over many experiments with the same protocol. This is called *Jarzynski's equality.*[146,147]

More precisely, suppose a system is described classically with the Hamiltonian $H(x, \lambda(t))$ with a time-dependent parameter $\lambda(t)$, which may be externally controlled. The initial state a is distributed canonically at the temperature $T = k_B/\beta$ and with the Hamiltonian $H(x, \lambda_0)$. Let state b be an equilibrium state with

---

[143] More precisely, tensor product

[144] A Frigeiro, V Gorini and M Verri, "The zeroth law of thermodynamics," Physica **137**A 573-602 (1986)

[145] C. Jarzynski, "Nonequilibrium equality for free energy differences", Phys. Rev. Lett. **78**, 2690-2693 (1997). The latest note dispelling misunderstanding is: C. Jarzynski, J. Stat. Mech. 2004, P09005, "Nonequilibrium work theorem for a system strongly coupled to a thermal environment." Chris Jarzynsky calls his equality, 'nonequilibrium work theorem.' In this latest paper, Jarzynski discusses how H may be interpreted as the effective Hamiltonian of the system when the system interacts strongly with the heat bath; this is directly relevant to 'single molecule physiology.' How did he arrive at this remarkable formula? He realized the relation when he was analyzing a numerical experimental result. That is, the relation was suggested first empirically.

[146] The equality holds if the system obeys a Markov process as seen in **6.5.9**.

[147] For a cycle (i.e., $\Delta A = 0$), this was demonstrated by Bochkov and Kuzovlev in the 1980's (according to C. Jarzynski); G. N. Bochkov and Yu. E. Kuzovlev, Zh. Eksp. Teor. Fiz. **72**, 238 (1977) [Sov. Phys.-JETP **45**, 125 (1977)]; Physica **106A**, 443, 480 (1981).

the Hamiltonian $H(x,\lambda_1)$ at the same temperature. We fix the protocol to change the system by fixing the function $\lambda(t)$ ($\lambda(0)=\lambda_0$ and $\lambda(t_f)=\lambda_1$, where t_f is the last time to modify the parameter). We embed this system in a large thermostat that can be described by a Hamiltonian H_0. Let us write the Hamiltonian of the whole system as $H_w(z,\lambda(t))$, where z denotes the phase variables of the whole system collectively. This Hamiltonian includes the interaction Hamiltonian between the system and the thermostat. We assume this effect may be ignored.

Now, W is the work done on the system during the change under the condition that the whole system is isolated.[148] Then, the above equality holds.

Warning. The final state actually reached by the protocol at time t_f is usually very different from the final equilibrium state we are interested in. □

Exercise 1. There is a macromolecule whose end-to-end distance can be controlled. We change its length from L_0 to L_1 in one second (linearly in time). The free energy difference of the equilibrium states with parameters $L_0=1$nm and $L_1=3$nm is given by $\Delta A=40$pNnm (= pico newton × nanometer) (i.e., the stretching increases the free energy this much if one does this in a quasiequilibrium fashion). The experiments told us that the work done to the molecule W was distributed according to a Gaussian distribution, and the average work was $\langle W\rangle = 60$pNnm. Assuming that the temperature is 300K, find the variance of the work. (Check that $300k_B = 4$pNnm). □

Exercise 2. It is often hard to obtain equilibrium free energy change with actual experiments performed at finite speeds. Thus, sometimes Jarzynski's equality is practically meaningful. Read the paper: "Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality" by Liphardt et al.[149] in Science **296**, 1832-1835 (2002), and summarize the content.[150] □

Discussion 1. There is a box separated into two compartments with a separating wall. One compartment is empty but the other is filled with gas. Now, the separating wall is removed by sliding quickly without any work. In this case, the free energy decreases because entropy increases. However, there is definitely no external work. Therefore, Jarzynski's equality cannot hold. □

2.7.21 Demonstration of Jarzynski's equality

The work done on the system is given by $W = H_w(z(t_f),\lambda_b) - H_w(z(0),\lambda_a)$, because it is as a whole isolated. Therefore,

$$\begin{aligned}\langle e^{-\beta W}\rangle &= \frac{1}{Y_a}\int dz(0) e^{-\beta H(z(0),\lambda_a)-\beta W} && (2.7.54)\\ &= \frac{1}{Y_a}\int dz(0) e^{-\beta H(z(t_f),\lambda_b)}. && (2.7.55)\end{aligned}$$

Here, Y_a is the partition function for the system with the Hamiltonian $H(x,\lambda_a)$ + the thermostat. According to Liouville's theorem (→**2.7.2**), this reads

$$\langle e^{-\beta W}\rangle = \frac{1}{Y_a}\int dz(t_f) e^{-\beta H(z(t_f),\lambda_b)} = \frac{Y_b}{Y_a}, \tag{2.7.56}$$

---

[148] For the following equality to be correct, the temperature of the whole system must not change, but if the bath is large enough, this is not a constraint.

[149] by J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante.

[150] See also a recent paper by the same group: PNAS **99** 13544-13548 (2002), and Jensen et al. on GlpF PNAS **99** 6731 (2002).

where Y_b is the partition function for the system with the Hamiltonian $H(x, \lambda_b)$ + the thermostat. Since the Hamiltonian of the thermostat is intact, and we ignore the interaction Hamiltonian that may depend on the parameter λ, $Y_b/Y_a = Z_b/Z_a$, where Z_c is the canonical partition function for the system with the Hamiltonian $H(x, \lambda_c)$. Thus, we have demonstrated (2.7.53).[151]

Since e^{-x} is convex, (2.7.53) implies

$$\langle W \rangle \geq \Delta A. \tag{2.7.57}$$

This is Thomson's principle.
Warning. Notice that the whole system is isolated, so if the total work done to the system is large, then the final temperature would be definitely higher than the initial one. *Still* (2.7.53) holds. However, ΔA compared there is *not* the free energy change between the initial and the actual final equilibrium states experiments realize.

2.7.22 Jarzynski's equality and rare fluctuations

We know that the difference $W - \Delta A$ can be made large without bound. Still, Jarzynski's equality (2.7.53) holds. What does this mean?

Let us consider the simplest case; we make a cycle, and the initial and the final equilibria are identical. $\Delta A = 0$, so

$$\langle e^{-\beta W} \rangle = 1. \tag{2.7.58}$$

If we make a rapid change, very likely $W > 0$; maybe it is often very large. Still, the equality must hold, so we need fairly large negative W occasionally. That is, we must wait for a fairly large scale 'violation' of the second law (cf. **4.1.3**). This implies that we must wait for a very long time to use (2.7.53) for a macroscopic system or for a process with large dissipation, even if a system under study is not macroscopic.
Exercise 1. Imagine a cylinder with a piston containing an ideal gas. The whole system is in a thermostat. Let us pull out the piston extremely rapidly (ignore the mass of the piston), and then very slowly and gently we return the piston to the original position. [Do not believe that all the questions below can be answered with the conditions given here; Appropriate information must be added.]
(1) What is the total work needed on the average?
(2) What is the heat generated?
(3) What sort of fluctuations does one have to wait for (2.7.53) to hold?
(4) Estimate their probabilities.
(5) How long should we wait for such fluctuations? □

2.8 Equilibrium Fluctuation and Response

Einstein thought fluctuations were the window to the more microscopic level of the world below the thermodynamic level (→**3.1.3**). The study of fluctuations is a crucial key to statistical mechanics. Furthermore, violent fluctuations could relatively easily happen at smaller

---

[151] This is the derivation given in the original paper, but it is hardly justifiable, because 'no change' in the heat bath does not imply the ratio of its partition functions before and after the change of the heat bath is unity (it is ∞/∞). If we can model the whole system as a Markov process, we can justify the formula as we will see in **6.5.9**.

scales, so studying them may give considerable hints at nonequilibrium processes. This was the leitmotif of Onsager's study (→**4.1.2**).

For the study of fluctuations in equilibrium IESM Section **1.9** is more detailed than this section.

2.8.1 Response and second moment

Suppose we have an ensemble with constant intensive parameter x. Let X be its conjugate variable (wrt energy, i.e., $+xdX$ appears in the differential form dE). Then, the thermodynamic potential suitable for this ensemble is

$$\hat{A} = -k_BT \log \hat{Z} \tag{2.8.1}$$

with

$$\hat{Z} = \int dX\, Z(X) e^{xX/k_BT}, \tag{2.8.2}$$

where $Z(X)$ is the canonical partition function under the constant X (say, V) condition. Therefore,

$$\chi \equiv \left.\frac{\partial X}{\partial x}\right|_{T,V} = k_BT \frac{\partial^2 \log \hat{Z}}{\partial x^2} = \frac{1}{k_BT}(\langle X^2\rangle - \langle X\rangle^2) = \frac{1}{k_BT}\langle \delta X^2\rangle. \tag{2.8.3}$$

Here, $\delta X = X - \langle X\rangle$. The left-hand side is the susceptibility of X. For example, if $x = H$ (magnetic field) and $X = M$ (magnetization), then χ is the magnetic susceptibility, describing the response of the magnetization to the change of magnetic field.

We have found that susceptibility is essentially a second moment.

Warning. (2.8.3) is derived using classical statistical mechanics, so the result is generally not applicable to non-classical cases. The reason is seen from (2.8.6). There, the Taylor expansion formula $e^{a+\delta a} = e^a + \delta a\, e^a + \cdots$ is used, but this equation is not correct when a and δa do not commute. See **4.3.5**.□.

2.8.2 Fluctuation-response relation

More generally, the cross correlation should give the cross susceptibility (*no* summation convention):

$$\langle \delta X_i\rangle = \beta\langle \delta X_i \delta X_j\rangle x_j, \tag{2.8.4}$$

where x_i[152] is the conjugate variable of X_i (with respect to energy, not entropy), and δ implies the difference from the average before the perturbation. This is called the *fluctuation-response relation.* This can be obtained easily as follows (classically):

$$\begin{aligned}\langle X_i\rangle &= \int d\Gamma\, X_i e^{-\beta(H - x_jX_j)} / \int d\Gamma\, e^{-\beta(H-x_jX_j)} &(2.8.5)\\ &= \langle X_i\rangle_0 + \beta\langle X_iX_j\rangle_0 x_j - \beta\langle X_i\rangle_0\langle X_j\rangle_0 x_j + \cdots. &(2.8.6)\end{aligned}$$

Here $\langle\ \rangle_0$ implies the average without perturbation. (In the RHS of (2.8.4), $\langle\ \rangle_0$ and $\langle\ \rangle$ make only a higher order difference, so you need not write $\langle X_iX_j\rangle_0$.)

(2.8.4) is equivalent to

$$\langle \delta X_i \delta X_j\rangle = k_BT \frac{\partial X_i}{\partial x_j}. \tag{2.8.7}$$

---

[152] $= \delta x_i$; we assume the unperturbed state is with $x_i = 0$ for simplicity.

Notice that Maxwell relations follow from this.
Exercise 1. Obtain the magnetic susceptibility in terms of the fluctuation of the magnetization. □
Exercise 2. Obtain the isothermal compressibility in terms of fluctuation. □
Exercise 3. What fluctuation gives the constant volume specific heat? What does the reader think happens if the compressibility diverges? □
Exercise 4. Let us take the ideal rubber band discussed in **2.6.5**. Obtain the Hooke's constant in terms of the length fluctuation. □
Exercise 5. The above exercise is a prototype of an experimental technique to determine 'spring constants' of molecular machines. Suppose there is a torsional spring with the torsional stiffness K (i.e, $\tau = K\theta$, where τ is the torsion, and θ is the angle displacement). The mean square average of the angle is 44°. What is the torsional stiffness?[153]

2.8.3 Magnitude of fluctuation

Notice that χ in (2.8.3) is an extensive quantity, so we know that the fluctuation around an equilibrium state is of order $\sqrt{N}$, where N is the total number of particles in the system.

The reader may wonder what happens near the critical point, where fluctuations become large. As we can see from the central limit theorem (→**1.7.1**), the most natural interpretation of order $\sqrt{N}$ is that there are $O[N]$ independent stochastic variables in the region of space we study the fluctuation. This means that the correlation length of the fluctuation is much smaller than the region we observe fluctuations. Thus, it is clear that even if the system is close to a critical point, if the diameter of the region on which we observe fluctuations is much larger than (say, 5 times as large as) the correlation length, the $O[\sqrt{N}]$ behavior of fluctuations holds.

However, exactly at the critical point the correlation does not decay exponentially, so the correlation length is infinitely long. Still, the correlation decays algebraically, and the fluctuation should decay as the system size. It is no more as fast as $\sqrt{N}$. For example, for the magnetization at the critical point $\langle M^2\rangle = O[N^{1+(2-\eta)/d}]$, where $\eta > 0$ but small, so the fluctuation is significantly larger than the noncritical cases (*critical fluctuations*).

2.8.4 Importance of studying fluctuations

To study nonequilibrium processes the *time correlation function* of x, assuming that x is a function of time $x(t)$, defined as

$$C(t) = \langle (x(t) - \langle x\rangle)(x(0) - \langle x\rangle)\rangle \tag{2.8.8}$$

is a very important quantity; its Fourier transform is called the power spectrum (→ around **3.7.4**) and is often experimentally directly measurable. Note that the 'equal time correlation' (i.e., $C(0)$) gives equilibrium (or thermodynamic) response. Then, it is a natural guess that response to time-dependent perturbation must be closely related to time correlation functions. This is indeed the case as we will learn later (→**3.2.8**).

---

[153] This is the stiffness of the kinesin dimer neck taken from W Hua, J Chung and J Gelles, Science **295**, 844 (2002). Their estimate seems wrong. The assertion of the paper on kinesin has been totally disproved by many people in 2003; see, for example, A. Yildiz, M. Tomishige, R. D. Vale, and P. R. Selvin, "Kinesin walks hand-over-hand," Science **303**, 676 (2004).

2.8.5 Thermodynamic fluctuation theory

Einstein (in 1910[154] →**2.8.8**) inverted the Boltzmann formula (2.4.37) as

$$W(\{X\}) = e^{S(\{X\})/k_B}, \tag{2.8.9}$$

where $\{x\}$ collectively denotes extensive variables. Since we know the statistical weights, we can compute the probability of observing $\{X\}$ as

$$P(\{x\}) = \frac{W(\{x\})}{\sum_{\{x\}} W(\{x\})}. \tag{2.8.10}$$

The denominator may be replaced with the largest term in the summands, so we may rewrite the formula as

$$P(\{x\}) \simeq \frac{W(\{x\})}{W(\{x_{eq}\})} = e^{[S(\{x\})-S(\{x_{eq}\})]/k_B}, \tag{2.8.11}$$

where $\simeq$ implies the equality up to a certain unimportant numerical coefficient, and $\{x_{eq}\}$ is the value of $\{x\}$ that gives the largest W (maximizes the entropy), that is, the equilibrium value.

The reader must have wondered why (2.8.9) may be used, even though the system (or a small portion of a system) is not isolated; we should look at the general partition function dictated by the general relation **2.6.5**. Thus, the probability of deviations $\{x\}$ must read

$$P(\{x\}) = e^{[S(\{x\})-S(\{x_{eq}\})-\sum(x_j-x_{j,eq})F_j]/k_B}. \tag{2.8.12}$$

Here, we have assumed that the intensive variables F_j are maintained by appropriate reservoirs. Notice that $F_j = \partial S/\partial x_j$ (the conjugate intensive variable of x with respect to entropy).

To the second order in $\delta x = x - x_{eq}$, (2.8.11) and (2.8.12) give identical results, because

$$S(\{x\}) = S(\{x_{eq}\}) + \sum \delta x_j F_j + \frac{1}{2}\sum \frac{\partial^2 S}{\partial x_i x_j}\delta x_i \delta x_j + \cdots. \tag{2.8.13}$$

That is, both approaches give the following famous result

$$P(\{\delta x\}) = e^{\delta^2 S/2k_B}. \tag{2.8.14}$$

However, the agreement is up to this order, and (2.8.12) is more general (→**2.8.6**).

The second law implies that the quadratic form $\log P$ in this approximation is negative semidefinite. Notice that this is the *stability condition* of the equilibrium state; this stability implies that the fluctuations are bounded on the average.

Strictly speaking, since we are discussing deviations from equilibrium states, we cannot use equilibrium thermodynamics to derive (2.8.9). Therefore, if we wish to be logically fastidious, we must regard

$$P(\{\delta x\}) = Y(\{x\})/Y(\{x_{eq}\}) \tag{2.8.15}$$

to be the *fundamental principle of thermodynamic fluctuation theory*, where Y is the partition function for an appropriate ensemble corresponding to the intensive variables kept constant in the observation region in space of the fluctuation (cf. **2.6.5**).

---

[154] This year Cannizzaro died.

Exercise 1. Show that (2.8.15) may be interpreted as

$P \propto \exp(-$ minimum work needed to create the fluctuation$/k_BT)$.

This relation was also recognized by Einstein. □

Exercise 2. Compute the second moment of entropy fluctuation under constant volume and constant pressure. [A more systematic and practical formula is in **2.8.7**.] □

Exercise 3. Generally speaking, (2.8.3) is consistent with thermodynamic fluctuation theory. Discuss the relation between the fluctuation theory and (2.8.7). □

2.8.6 Massieu functions

The functions obtained from entropy S through Legendre transformations are generally known as *Massieu functions*. The thermodynamic fluctuation principle (2.8.15) is most conveniently written in terms of these functions. The principle says that for an isolated system (as Einstein said)

$$P(\{\delta x\} \simeq e^{\delta S/k_B}. \tag{2.8.16}$$

Here $\simeq$ implies 'modulo unimportant numerical factors.' If an intensive quantity X is fixed, and this quantity appears in dE as Xdx (x is its conjugate extensive quantity), (2.8.15) reads

$$P(\{\delta x\}) \simeq e^{\delta\Psi/k_B}, \tag{2.8.17}$$

where

$$\Psi = S + Xx/T. \tag{2.8.18}$$

This is called the (generalized) *Massieu function*.[155] Note that

$$d\Psi = \frac{1}{T}dE + xd\left(\frac{X}{T}\right). \tag{2.8.19}$$

Exercise 1. Obtain the Massieu function for the pressure ensemble (the ensemble with constant T and p). □

2.8.7 Practical formula for thermodynamic fluctuations[156]

Let $dE = TdS - pdV + \mu dN + Xdx$. From this

$$\delta S = \frac{1}{T}(\delta E + p\delta V - \mu\delta N - X\delta x). \tag{2.8.20}$$

To study the fluctuation we need $\delta^2 S$. This can be computed from the above formula as (this is the second order term of the Taylor expansion, so do not forget 1/2)

$$\begin{aligned} \delta^2 S &= \frac{1}{2}\left[\delta\left(\frac{1}{T}\right)(\delta E + p\delta V - \mu\delta N - X\delta x) + \frac{1}{T}(\delta p\delta V - \delta\mu\delta N - \delta X\delta x)\right], && (2.8.21) \\ &= -\frac{\delta T}{2T^2}T\delta S + \frac{1}{2T}(\delta p\delta V - \delta\mu\delta N - \delta X\delta x). && (2.8.22) \end{aligned}$$

Therefore, we arrive at the following useful expression worth remembering

$$\delta^2 S = \frac{-\delta T\delta S + \delta p\delta V - \delta\mu\delta N - \delta X\delta x}{2T}. \tag{2.8.23}$$

---

[155] The original Massieu function is for the temperature ensemble, so $\Psi = -A/T$.

[156] See IESM **1.9.13** may slightly be kinder.

As an example, let us study the fluctuations of V and T, assuming that N and x are fixed. To this end we express δS and δp in terms of δV and δT.

$$\delta S = \left.\frac{\partial S}{\partial T}\right|_V \delta T + \left.\frac{\partial S}{\partial V}\right|_T \delta V = \frac{C_V}{T}\delta T + \left.\frac{\partial p}{\partial T}\right|_V \delta V, \tag{2.8.24}$$

$$\delta p = \left.\frac{\partial p}{\partial T}\right|_V \delta T + \left.\frac{\partial p}{\partial V}\right|_T \delta V. \tag{2.8.25}$$

Therefore,

$$\delta T\delta S - \delta p\delta V = \frac{C_V}{T}(\delta T)^2 - \left.\frac{\partial p}{\partial V}\right|_T (\delta V)^2. \tag{2.8.26}$$

Hence, we find

$$\langle\delta T^2\rangle = \frac{T^2}{C_V},\ \langle\delta V^2\rangle = -T\left.\frac{\partial V}{\partial p}\right|_T,\ \langle\delta T\delta V\rangle = 0. \tag{2.8.27}$$

Note that $\langle\delta V^2\rangle = O[V]$ as expected in **2.8.3**.
Exercise 1. Again, let us assume N and x are kept constant. Study the fluctuation of S and p. □
Exercise 2. Study the energy fluctuation $\langle\delta E^2\rangle$. [Hint. We have already computed the fluctuations of V and T, so let us write $\delta E = T\delta S - p\delta V$ in terms of δV and δT.] □
Exercise 3. If an extensive variable x and an intensive variable Y are not conjugate, then $\langle\delta x\delta Y\rangle = 0$. □.
Exercise 4. What can we say generally about the sign of $\langle\delta X_i\delta x_i\rangle$? [The reader must explain why.] □

2.8.8 Einstein's motivation to study fluctuations

The purpose of Einstein to construct the theory of fluctuation was to understand critical fluctuations (critical opalescence) quantitatively. As a byproduct he explained why the sky was blue.

Let us humbly take note that it is still only less than 100 years since we first understood why the sky was blue.

Part II. Outline of Nonequilibrium Statistical Mechanics

Part II consists of two parts, the Outline part (Chapters 3-5) and more detailed auxiliary chapters (X1-X3). The Outline part may be used as an introductory (or undergraduate level) course of nonequilibrium statistical mechanics; the reader is expected to become acquainted with several narratives of nonequilibrium processes, and to pick up typical modes of thinking and rudimentary techniques of nonequilibrium statistical mechanics as well as some important keywords such as the fluctuation-dissipation relation. Chapters X1-X3 explain more advanced topics and technically complicated matter. These chapters tend to contain more typos, etc., than not-X-rated chapters. Please be careful.

Elementary kinetic theory will not be discussed (→**2.3.9**). The Boltzmann equation (→**2.4.7**, **2.7.10**) will not be discussed, either; in short, we do not emphasize topics that are not useful to understand nonequilibrium statistical physics in general. We wish to focus on the general theoretical framework applicable to condensed phases or systems immersed in condensed phases.

If the reader wishes to review equilibrium statistical mechanics briefly, go to Section **2.6**. If she wishes to do so at a leisurely pace, read the historical account in Chapter 2. If she wishes to review rudiments of probability theory, read Chapter 1. All the contents of these previous chapters will be extensively and in detail referred to at the places where we need them, so the reader need not review these materials before they become needed.

Chapter 3 outlines the Brownian motion. After describing the historical background, Einstein's work is explained, and its fundamental importance to nonequilibrium statistical thermodynamics is stressed. Then, the basic tools to study the Brownian motion (Langevin and Fokker-Planck equations) are explained. Several important consequences such as the fluctuation dissipation relations and the Green-Kubo relations are illustrated with simple examples. The relation between mechanics and stochastic models with the aid of projections is discussed. Also stochastic differential equations (Itô calculus) are intuitively introduced. The chapter concludes with the noise analysis (the Wiener-Khinchine theorem) and dynamical form factors (the van Hove theory).

Chapter 4 is an outline of Onsager's work, the regression principle and nonequilibrium thermodynamics. The Green-Kubo relations are derived from the study of fluctuations (both classical and quantum cases) and Onsager's reciprocity is explained. A general kinematic framework of linear response theory and an outline of the usual formal derivation of linear-response theory conclude the Chapter. Detailed examples, NMR (the Kubo-Tomita theory), electric conductivity (the Nakano theory) are relegated to Chapter X3.

Chapter 5 discusses the hydrodynamic limit and hydrodynamic equations. First, the problem is discussed from the conventional conservation law side. Then, the same problem is studied from Onsager's regression point of view.

Chapter 3

Brownian Motion

The standard rudiments such as Langevin equations and Fokker-Planck or Smoluchowski equations and their elementary applications (e.g., barrier crossing) are covered. (The entries for Brownian ratchets will be further augmented.) Advanced topics such as spectral analysis and stochastic calculus are also discussed at an elementary level.

3.1 Brownian Motion

First, the history of the topic is followed, and then the fundamental mode of thinking of nonequilibrium statistical mechanics due to Einstein is outlined (→**3.1.7**). The relation between diffusion and random walk is also discussed.[1]

3.1.1 What did Mr. Brown really do?

Robert Brown (1773-1858)[2](1773-1858) discovered in the summer of 1827[5] the *Brownian*

---

[1] F. Yonezawa, *Brownian Motion* (Kyoritsu, 1986) is an excellent and admirable book (in Japanese) on the Brownian motion in general.

[2] He was perhaps the greatest botanist (and a great microscopist; Alexander von Humboldt called him 'the glory of Great Britain') in the first half of the 19th century. He wrote (1810) a classic of systematic botany describing the Australian flora, following his expedition (1801-5). He recognized the nucleus of the cell[3] and so named it (1831; This name was later imported by Bohr to atomic physics). Before departing for his Beagle expedition (Dec., 1831 - Oct., 1836),[4] Darwin asked for Brown's advice in 1831, buying a portable dissecting microscope recommended by Brown; after returning to England, Brown encouraged Darwin to visit him every Sunday morning. Later, Brown was regularly invited to parties at Darwin's home. The participants of the now historical Linnean Society meeting, where the theory of natural selection was first read (July 1, 1858), were there mainly to listen to Lyell reading the eulogy for Brown and to praise his career. Cf. J. Browne, *Charles Darwin, voyaging* (Knopf, 1995), *Charles Darwin, the power of place* (Knopf, 2002).

[5] Beethoven died in March; Democratic party was founded.

motion of particles, coming from ruptured pollens, suspended in water.[6] We are usually given an impression that he simply observed the "Brownian motion." However, he did a very careful and thorough research to establish the *universal nature of the motion*.

Since the particles came from living cells, initially he thought that it was a vital phenomenon. Removing the effect of advection, evaporation, etc., carefully, he tested many flowers. Then, he tested old pollens in the British Museum (he was the director of the Botanical Division), and still found active particles. He conjectured that this was an organic effect, testing even coal with no exception found. This suggested him that not only vital but organic nature of the specimens were irrelevant. He then tested numerous inorganic specimens (including a piece of Sphinx; he also roasted his specimens).

He really deserves the name of the motion.
See Philip Pearle, Brian Collett, Kenneth Bart, David Bilderback, Dara Newman, and Scott Samuels, "What Brown saw and you can too," AJP 78, 1278 (2010).

3.1.2 Before Einstein

There was no work published on the Brownian motion between 1831 and 1857, but the phenomenon was well known. From 1850s new experimental studies began by Gouy and others. The established facts included:[7]
(1) Its trajectory is quite erratic without any tangent lines anywhere (→**3.4.1**).
(2) Two Brownian particles are statistically independent even when they come within their diameters.
(3) Smaller particles move more vigorously.
(4) The higher the temperature, the more vigorous the Brownian motion.
(5) The smaller the viscosity of the fluid medium, the more vigorous the motion (→**3.1.3**).
(6) The motion never dies out.
etc.
In the 1860s there were experimentalists who clearly recognized that the motion was due to the impact of water molecules. Even Poincaré mentioned this motion in 1900,[8] but somehow no founding fathers of kinetic theory and statistical mechanics paid any attention to Brownian motion.[9]

3.1.3 Einstein's paper[10] of 1905; Einstein's formula

Einstein's key idea was that a Brownian particle may be treated both as a large molecule and as a tiny macroscopic particle at the same time:
(a) its diffusion obeying Fick's law must be driven by the osmotic pressure gradient.
(b) its mean velocity under an external force is obtained with the aid of fluid dynamics

---

[6] The work was published the next year, but was communicated to Stokes very soon in August, 1827.

[7] Yonezawa p45-.

[8] His student Bachelier published in 1900 a paper on the dynamics of stock market in terms of numerous but small random variations. L Bachelier, "Théorie de la spéculation." Ann. Sci. de L'École Normal Supérier **17**, 21 (1900). This is the first paper of mathematical finance.

[9] The fact that the equipartition of energy did not seem to hold for Brownian particles seems to have been a big obstacle (See Editorial note p206 of *Collected Papers of Albert Einstein* vol.2); it was not hard to measure the mass of the particle (about 1pg), but to measure the velocity was the problem (→**3.1.5**). Only the average velocity was measured, so it was slow and the kinetic energy (that must be equal to $3k_BT/2$ (→**2.3.5**)) was grossly underestimated. Einstein was successful, because he avoided any direct discussion of the true velocity.

[10] "Über die von der molekularkinetischen Theorie der Wärme geforderten Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen," Ann. Phys. **17**, 549 (1905) [On the motion of suspended particles in stationary fluid required by the molecular kinetic theory of heat.]

(Stokes' law).

First of all, according to van't Hoff, the osmotic pressure Π of a solute is given by

$$\Pi = nk_BT, \tag{3.1.1}$$

where n is the number density of the suspended particles.[11] If there is a concentration gradient, there is an osmotic pressure gradient. If a stationary gradient is caused by an external force (e.g., gravity) $\boldsymbol{f}$ (per molecule), there must be a force balance:

$$-\nabla\Pi = n\boldsymbol{f}. \tag{3.1.2}$$

Therefore, the force per molecule due to the osmotic pressure gradient must be given by

$$\boldsymbol{f} = -k_BT\nabla\log n. \tag{3.1.3}$$

This is the consequence of (a).

On the other hand, the average molecular velocity $\boldsymbol{v}$ is given by $\zeta\boldsymbol{v} = \boldsymbol{f}$, where ζ is the friction constant that may be measured or computed for a given particle. If we use Stokes' law, $\zeta = 6\pi a\eta$, where a is the radius of the particle, and η is the viscosity of the fluid in which the particle is suspended.[12] Therefore, the consequence of (b) is

$$6\pi a\eta\boldsymbol{v} = \boldsymbol{f}. \tag{3.1.4}$$

$n\boldsymbol{v}$ must be the diffusion flux. This must be equal to the result given by Fick's law, so

$$-D\nabla n = n\boldsymbol{v}. \tag{3.1.5}$$

Therefore, we conclude

$$D = k_BT/6\pi a\eta. \tag{3.1.6}$$

This equality is called the *Einstein relation.* This equation allows us to obtain k_B or, since the gas constant R is known, to calculate Avogadro's constant.

3.1.4 Mean square displacement of Browning particle

Due to the bombardment of water molecules, the Brownian particle executes a zigzag motion, and on the average it is displaced.

Einstein derived the diffusion equation for the particle and computed the mean square displacement.

The diffusion flux $\boldsymbol{j}$ must be given by

$$\boldsymbol{j} = -Dgrad\, n. \tag{3.1.7}$$

Therefore, the conservation of the number of particles implies the *diffusion equation*:

$$\frac{\partial n}{\partial t} = -div\boldsymbol{j} = D\Delta n. \tag{3.1.8}$$

Its Green's function[13] describes the displacement distribution in d-dimensional space

$$G(\boldsymbol{r},t) = \left(\frac{1}{4\pi Dt}\right)^{d/2} e^{-\boldsymbol{r}^2/4Dt}. \tag{3.1.9}$$

---

[11] Einstein actually derived this from statistical mechanics in this 1905 paper

[12] For a proof see Landau-Lifshitz, *Fluid Dynamics.*

[13] on the whole 3-space with the boundedness condition everywhere.

Therefore, after t, the mean square displacement must be

$$\langle \boldsymbol{r}(t)^2 \rangle = 2dDt. \tag{3.1.10}$$

That is, if we observe the mean square displacement of a particle, then D of the ensemble of such particles may be measured.

The left are four sample paths and their average is on the right. [Courtesy of Prof. Nishizaka of Gakushuin Univ.]

3.1.5 Experimental verification of Einstein's theory

Now in the 21st century, we feel Einstein's argument is impeccable, but his theory was not immediately accepted, because of several logical or physical gaps in his argument outlined in **3.1.3**.

The ingredient (a) is not immediately clear, because it is not obvious whether such a large particle as a Brownian particle obeys exactly the same law as ordinary molecules (the formula for the osmotic pressure). The ingredient (b) is also not very obvious due to an opposite reason; we are not very sure whether the macroscopic fluid dynamics (i.e., Stokes' law) is applicable to the average velocity of such a small particle as a Brownian particle.[14] Perrin experimentally checked that these premises were right.

Then, Perrin used (3.1.10) and measured D. With the aid of (3.1.6) it is possible to determine k_B. That is, we may obtain Avogadro's constant N_A. Perrin performed such studies and consistently obtained N_A (various methods all gave $\sim 6.5 \times 10^{23}$). Thus, Perrin declared that it had become difficult to deny the existence of atoms.

Remark 1. However, do not naively conclude that the observation of erratic motion 'proves' the existence of atoms and molecules. It only tells us that around a Brownian particle is a pressure fluctuation. The consistency of several different measurements of Avogadro's constant is the key. □

---

[14] The reader may suggest a molecular dynamic simulation to check this and might add that there is already such a paper (e.g., M Vergeles, P Keblinski, K Koplik, and J R Banavar, "Stokes drag at the molecular level," Phys. Rev. Lett. **75**, 232 (1995); this paper suppresses the fluctuation of the Brownian particle by choosing the particle mass very large). However, since the hydrodynamic effect is long-ranged (→**5.2.6**), realistic simulation requires a huge system and is not easy.

Exercise 1. The diffusion constant of insulin is about $8.2 \times 10^{-11}\mathrm{m}^2/\mathrm{s}$. What is the typical time scale for the insulin molecule to diffuse across a typical cell? □
Exercise 2. How about ribosomes in a cell? How long will it take for a complete ribosome to diffuse across a eukaryotic cell or a prokaryotic cell?[15] □

3.1.6 Einstein's remark on instantaneous velocity of Brownian particle

In 1907 Einstein remarked: for a Brownian particle its average velocity may be obtained from its short-time displacement, but a very short time average does not make sense. An argument very similar to the one we have seen in (3.1.10) shows that the mean displacement during time t is proportional to $\sqrt{t}$. This implies that the instantaneous velocity must be undefinable; the motion is very erratic. Perrin commented that the Brownian path is everywhere nondifferentiable. See **3.4.1** (cf. **X1.4.6**).
Exercise 1. A Brownian particle of radius 0.1 μm is suspended in water. Estimate the time needed for the initial speed to decay to its 1/10. What does this tell you about the observability of the speed of the particle? □

3.1.7 Fundamental idea of Einstein

Einstein's fundamental ideas dominate nonequilibrium statistical mechanics:

> Microscopic fluctuations can build up (semi)macroscopic fluctuations whose dynamics is on the average governed by the laws of macroscopic time evolution.

Brownian motion is a semimacroscopic (mesoscopic) motion that is a result of building up of microscopic fluctuations. Its decay is described by Stokes' law, a macroscopic dissipative dynamics.

This was later more clearly stated by Onsager as the regression hypothesis (→**4.1.2**).

3.1.8 Random walk and diffusion

Due to the pressure fluctuation caused by microscopic motion of fluid molecules, a Brownian particle executes an erratic motion. We know that its velocity is hardly well-defined (→**3.1.6**). However, we can observe its displacement vectors for a fixed time interval, say τ (stroboscopic observation). Let $\boldsymbol{r}_i$ be the i-th total displacement between time $(i-1)\tau$ and $i\tau$. Thus, we may model the movement of a gas molecule by a *random walk*: for every τ the point makes a displacement $\boldsymbol{r}$

After N steps (after $N\tau$), the total displacement $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \boldsymbol{r}_1 + \cdots + \boldsymbol{r}_N. \tag{3.1.11}$$

For a Brownian particle we know that $\langle \boldsymbol{r}_i \rangle = 0$ (there is no systematic movement) and the steps may be regarded mutually statistically independent (→**1.2.2**). Therefore, the mean-square displacement is (notice that (3.1.11) is a sum of iid random variables, cf., **1.7.1**)

$$\langle \boldsymbol{R}^2 \rangle = \langle \boldsymbol{r}_1^2 \rangle + \cdots + \langle \boldsymbol{r}_N^2 \rangle = N\langle \boldsymbol{r}_1^2 \rangle = N\ell^2, \tag{3.1.12}$$

where $\langle \boldsymbol{r}^2 \rangle = \ell^2$ which may be observable. Comparing this with (3.1.10) with $t = N\tau$, we obtain

$$D = \ell^2/2d\tau. \tag{3.1.13}$$

Here, d is, as before, the spatial dimensionality.

---

[15] This tells us why we eukaryotes need molecular transport devices.

Actual observation results of a latex particle trajectory for 3.3 sec. The left every 1/8000 sec, and the right every 1/30 sec. [Courtesy of Prof Nishizaka of Gakushuin U]

3.1.9 Diffusion equation and central limit theorem

We can also apply the central limit theorem **1.7.1** to the stroboscopic observation in **3.1.8**: each component of $\boldsymbol{R}/\sqrt{N\ell^2/d} = \boldsymbol{R}/\sqrt{2Dt}$ obeys asymptotically $N(0,1)$ (here, D is given by (3.1.13))

$$f(\boldsymbol{R},t) = \left(\frac{1}{4\pi Dt}\right)^{d/2} e^{-\boldsymbol{R}^2/4Dt}. \tag{3.1.14}$$

This of course agrees with the Green's function of the diffusion equation (→(3.1.9)).

Let us more directly arrive at the diffusion equation in a theoretical physicists' fashion. Let us consider the characteristic function (→**1.4.7**) $\psi(\boldsymbol{k},N)$ of $\boldsymbol{R}$. Let $\omega(\boldsymbol{k})$ be the characteristic function of each step $\boldsymbol{r}$. Then,

$$\psi(\boldsymbol{k},N) = \omega(\boldsymbol{k})^N. \tag{3.1.15}$$

Assuming that N is very large, we can differentiate this as

$$\frac{\partial}{\partial N}\psi(\boldsymbol{k},N) = [\log\omega(\boldsymbol{k})]\,\psi(\boldsymbol{k},N). \tag{3.1.16}$$

We are interested in small $\boldsymbol{k}$, so $\omega(\boldsymbol{k}) = 1 - \ell^2 k^2/2d + \cdots$, and (3.1.16) reads

$$\frac{\partial}{\partial N}\psi(\boldsymbol{k},N) = -\frac{\ell^2}{2d}k^2\psi(\boldsymbol{k},N). \tag{3.1.17}$$

This is just parallel to the demonstration of the central limit theorem in **1.7.1**. Replacing N with t/τ, this is the spatial Fourier transform of the diffusion equation (3.1.8):

$$\frac{\partial}{\partial t}\psi(\boldsymbol{k},N) = -Dk^2\psi(\boldsymbol{k},N). \tag{3.1.18}$$

There is a way to make the above argument rigorous due to Lindeberg.

3.1.10 Chapman-Kolmogorov equation, Markov property

Suppose the probability to find the random walker in a volume element centered at position $\boldsymbol{r}'$ at time t is known as $P(\boldsymbol{r}',t)d^3\boldsymbol{r}'$. If we know the distribution p of the next one step $\boldsymbol{\ell}$, we should be able to obtain the probability to find the walker around a specified location $\boldsymbol{r}$ at time $t+\tau$. We can write

$$P(\boldsymbol{r},t+\tau) = \int d\boldsymbol{\ell}\, p(\boldsymbol{\ell})P(\boldsymbol{r}-\boldsymbol{\ell},t). \tag{3.1.19}$$

This type of equation is called the *Chapman-Kolmogorov equation.* It is nothing but a consistency condition among various probabilities and transition probabilities.[16]

The reason why we can write down the Chapman-Kolmogorov equation (3.1.19) is that what happens next depends only on the knowledge about the present (more precisely, what happens after time t_0 is dependent only on the state of the system at t_0 and not on the states at $t < t_0$).[17] Such a process is called a *Markov process* (→**6.1.3**).

3.1.11 From Chapman-Kolmogorov equation to diffusion equation

Since we assume $|\boldsymbol{\ell}|$ is small (or P does not change appreciably for this length scale), we should be able to reduce the Chapman-Kolmogorov equation to a differential equation. To this end it is convenient to rewrite (3.1.19) as

$$P(\boldsymbol{r},t+\tau) = \int d\boldsymbol{\ell}\, p(\boldsymbol{\ell})e^{-\boldsymbol{\ell}\cdot\frac{\partial}{\partial \boldsymbol{r}}}P(\boldsymbol{r},t). \tag{3.1.20}$$

Here, we have used the formal expression that can easily be guessed from the Taylor series[18]

$$\begin{aligned} e^{a(d/dx)}f(x) &= \left[1 + a\frac{d}{dx} + \frac{1}{2!}a^2\frac{d^2}{dx^2} + \cdots + \frac{1}{n!}a^n\frac{d^n}{dx^n} + \cdots\right] f(x), && (3.1.21)\\ &= f(x) + af'(x) + \frac{1}{2!}a^2 f''(x) + \cdots + \frac{1}{n!}f^{(n)}(x) + \cdots, && (3.1.22)\\ &= f(x+a). && (3.1.23) \end{aligned}$$

Or, this is equivalent to the statement that the momentum is the generator of translation in quantum mechanics.

Let us compute the average over ℓ. We assume that p is Gaussian.[19]

$$\int d\boldsymbol{\ell}\, p(\boldsymbol{\ell})e^{-\boldsymbol{\ell}\cdot\frac{\partial}{\partial \boldsymbol{r}}} = \exp\left[\frac{\langle \ell_i \ell_j\rangle}{2}\frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\right] = \exp\left[\frac{\langle \ell^2\rangle}{2d}\Delta\right]. \tag{3.1.24}$$

---

[16] p governs the transition from the position at present to the position at the next time step, so it is called the *transition probability.*

[17] Of course, the past influences the future in this case, but only through the state at time t_0.

[18] This is only formal; or more precisely, this is used only inside the integration symbol, we may use the theory of generalized function. For an operator A e^A is defined by the operator series

$$e^A \equiv 1 + A + A^2/2! + \cdots + A^n/n! + \cdots.$$

This is meaningful only when A is a bounded operator. Differential operators are unbounded, so our calculation is only formal.

[19] One plausibility argument for this is that this displacement is a result of numerous collisions, so the central limit theorem (→**1.7.1**) suggests the Gaussianness of p.

Therefore,

$$P(\boldsymbol{r},t)+\tau\frac{\partial}{\partial t}P(\boldsymbol{r},t)+\cdots=\exp\left[\frac{\langle\ell^2\rangle}{2d}\Delta\right]P(\boldsymbol{r},t)=\left(1+\frac{\ell^2}{2d}\Delta+\cdots\right)P(\boldsymbol{r},t). \tag{3.1.25}$$

That is, if we set $D=\ell^2/2d\tau$ as before (→(3.1.10)), we arrive at the diffusion equation again

$$\frac{\partial P}{\partial t}=D\Delta P. \tag{3.1.26}$$

3.2 Introduction to Langevin Equation

Langevin's original:
Langevinfs paper in translation: Am. J. Phys. 65 (11), November 1997, 1079.

3.2.1 Path space description of random phenomenon

Einstein's original treatment of Brownian particles is to study them through averages.[20] Description of Brownian particles in terms of distribution functions is the next step (the next theoretical refinement). We have already glimpsed this approach through the study of diffusion equation (→**3.1.8**): it is an equation governing the time evolution of the spatial density distribution of Brownian particles.[21] We will further discuss this approach later (→**3.3**).

The reader must have thought that the ultimate description of the Brownian motion is to study the 'trajectory' of each sample particle; that is, we should describe the time-dependent phenomenon in its *path space* = history space.[22] To this end we need a detailed modeling of the motion of each Brownian particle.

3.2.2 Langevin introduced path space description of Brownian motion

Such a description was proposed by Langevin (1872-1946): he tried to model the equation of motion of a Brownian particle:

$$m\frac{d\boldsymbol{v}}{dt}=-\zeta\boldsymbol{v}+\boldsymbol{F}(t)+\boldsymbol{w}(t), \tag{3.2.1}$$

where $\boldsymbol{v}$ is the velocity of the particle, m is its mass, ζ is the friction constant, $\boldsymbol{F}$ is a systematic mechanical force that may be time-dependent, and $\boldsymbol{w}$ is the noise corresponding to the impacts of water molecules on the particle. This is Newton's equation of motion, but the nature of the noise force $\boldsymbol{w}$ is very different from the ordinary forces in mechanics (→**3.2.3**; **3.7.6**).

In modern terms Langevin introduced a stochastic differential equation (→**3.4.4**. **6.5**)

$$m\frac{d\boldsymbol{v}(t,\omega)}{dt}=-\zeta\boldsymbol{v}(t,\omega)+\boldsymbol{F}(t)+\boldsymbol{w}(t,\omega), \tag{3.2.2}$$

---

[20] This is the 'level 1' in the terminology we have used in large deviation theory (→**1.6.2**).

[21] This is the 'level 2' in the terminology we have used in large deviation theory (→**1.6.7**).

[22] This is the 'level 3' in the terminology of large deviation theory.

where ω is the stochastic parameter (see **1.4.1**, **1.4.2**) . For each sample (for each observation) a different ω is chosen from a probability space.

To model a Brownian particle, we need a stochastic process (→**1.4.2**) $\boldsymbol{w}(t,\omega)$, a map from a probability space to a function of time that should not be differentiable.

3.2.3 How to specify noise driving force

We may assume that the random force is, on the average, zero: the average $\langle \boldsymbol{w} \rangle = 0$, where $\langle\ \rangle$ implies the ensemble average = average wrt the stochastic parameter ω. Therefore, at least we must specify its fluctuation (second moments).

Let us specify its *time correlation function* (in this case it is a matrix)

$$C(t,s) = \langle \boldsymbol{w}(t,\omega)\boldsymbol{w}^T(s,\omega)\rangle, \tag{3.2.3}$$

where $t \geq s$, ^{T} denotes the transposition of the vector, and the average is over ω.[23] Langevin's idea is as follows:[24] Even at the same time different components are statistically independent; the same component (say, x-component) at different times may be correlated, but its 'memory' is very short. Also he assumes that the noise is stationary. That is, statistical properties do not depend on the absolute time. Thus, $C(t,s) = C(t-s,0)$; We may simply write $C(t,0)$ as $C(t)$. The ultimate short memory model is defined by the following '*white noise*,'[25]

$$C(t) \equiv \langle \boldsymbol{w}(t)\boldsymbol{w}^T(0)\rangle = 2\alpha I\delta(t). \tag{3.2.4}$$

Here, α is a positive constant, and I is the 3×3 unit matrix.[26]

Although we have not completed the specification of the noise,[27] we do not need any more detail, if we wish to compute only two-time correlation functions.

3.2.4 Langevin model of free Brownian particle: introduction

Let us assume that there is no external force in (3.2.1):

$$m\frac{d\boldsymbol{v}}{dt} = -\zeta\boldsymbol{v} + \boldsymbol{w}(t). \tag{3.2.5}$$

With the initial condition $\boldsymbol{v}(0)$ we can solve it as

$$\boldsymbol{v}(t) = \frac{1}{m}\int_0^t \boldsymbol{w}(s)e^{-(\zeta/m)(t-s)}ds + \boldsymbol{v}(0)e^{-(\zeta/m)t}. \tag{3.2.6}$$

If a particle is with radius a and the viscosity of the fluid medium is η, $\zeta = 6\pi a\eta$ with a sticky particle-surface-fluid boundary condition (Stokes' law).[28]

Using this result, let us compute the mean square velocity:

$$\langle \boldsymbol{v}(t)\cdot\boldsymbol{v}(t)\rangle = \frac{3\alpha}{m\zeta} + \left(\langle \boldsymbol{v}(0)^2\rangle - \frac{3\alpha}{m\zeta}\right)e^{-2(\zeta/m)t}. \tag{3.2.7}$$

---

[23] If the existence of the stochastic parameter ω is obvious from the context, it will be suppressed.

[24] of course, in the modern notation

[25] It is said to be 'white', because the noise changes so rapidly in time that all the frequencies appear. We will outline harmonic analysis of Langevin equations soon (→**3.7.4**).

[26] The prefactor 2 is very convenient (cf. **X1.2.10**). 3 is the spatial dimensionality, so in d-space I should be the $d\times d$ unit matrix.

[27] Often we say we assume the Gaussian white noise. Then, the statistical specification of the noise is complete in terms of correlation functions, because any higher order moments are completely determined by the second moment for a Gaussian distribution.

[28] If we use a slippery boundary condition, it is $4\pi a\eta$. See Landau-Lifshitz, *Fluid Mechanics*.

Here, we have assumed that the noise and the initial velocity are statistically independent.

3.2.5 Fluctuation-dissipation relation of the second kind

We are studying a Brownian particle suspended in an equilibrium fluid. Therefore, after a sufficiently long time $\langle \boldsymbol{v}^2 \rangle$ must be time independent and the equipartition of energy $\langle \boldsymbol{v}^2 \rangle = 3k_BT/m$ must hold. Therefore, (3.2.7) implies $\alpha = k_BT\zeta$. That is,

$$\langle \boldsymbol{w}(t,\omega)\boldsymbol{w}^T(s,\omega)\rangle = 2\zeta k_B T I \delta(t) \tag{3.2.8}$$

is required. This is called a *fluctuation dissipation relation of the second kind* (for a general form see **X1.2.6**, **X2.3.15**): it relates the noise amplitude and the magnitude of the damping factor through the requirement that they are compatible with equilibrium statistics: if the damping factor is large, the random driving force must be large to maintain sufficiently vigorous thermal motion. Notice that this relation automatically makes $\langle \boldsymbol{v}(t)\cdot\boldsymbol{v}(t)\rangle$ time-independent as it should be.

3.2.6 Langevin model of Brownian particle: summary

Let us summarize the Langevin model of a particle subject to the equilibrium noise. The equation may be used to describe a suspended particle of radius a in a fluid of viscosity η and temperature T subject to an external force $\boldsymbol{F}$.[29] The equation is called a Langevin equation:

$$m\frac{d\boldsymbol{v}}{dt} = -\zeta\boldsymbol{v} + \boldsymbol{F}(t) + \boldsymbol{w}(t,\omega), \tag{3.2.9}$$

where $\zeta = 6\pi\eta a$ and the noise is a stationary Gaussian white noise:

$$\langle \boldsymbol{w}(t,\omega)\rangle = 0, \tag{3.2.10}$$

$$\langle \boldsymbol{w}(t,\omega)\boldsymbol{w}(0,\omega)^T\rangle = 2\zeta k_B T I\delta(t). \tag{3.2.11}$$

The last equation is the fluctuation dissipation relation of the second kind, and is required to maintain the Maxwell distribution of the particle velocity. In classical statistical mechanics, the velocity distribution is always the same independent of the external force. Therefore, our consideration for a free particle in **3.2.3** should work.

3.2.7 Equilibrium time correlation function of Brownian velocity

Even in equilibrium there are fluctuations that are explicitly time dependent. The study of time evolution of various quantities in equilibrium is an important topic of 'nonequilibrium' statistical mechanics (→Onsager's principle **3.2.10**).

Since we know $\boldsymbol{v}(t,\omega)$ for any initial condition (and for any noise sample specified by the stochastic parameter ω),[30] we can compute the *equilibrium time correlation function* of the velocity of a free (i.e., $\boldsymbol{F} = 0$) Brownian particle: for $t > 0$

$$\langle \boldsymbol{v}(t,\omega)\boldsymbol{v}^T(0,\omega)\rangle = \langle \boldsymbol{v}(0,\omega)\boldsymbol{v}^T(0,\omega)\rangle e^{-(\zeta/m)t} = \frac{k_BT}{m}Ie^{-(\zeta/m)t}. \tag{3.2.12}$$

---

[29] Here, $\boldsymbol{v}$ also depends on ω, but this dependence is not explicitly written.

[30] Here, the stochastic parameter ω only specifies a noise sample, but does not specify the initial condition for the velocity randomly sampled from the equilibrium ensemble. Mathematically, $\boldsymbol{v}$ is a sample path, so it may be better to let ω specify both the stochastic objects, the noise and the initial velocity. Here, mathematically-oriented readers may interpret the notation as follows. We could introduce two stochastic parameters for these two stochastic = random objects to make the situation explicit, but the stochastic parameter for the initial condition is suppressed, since $\boldsymbol{v}(0)$ appears explicitly.

Here, as above, I is the 3×3 unit matrix, and ^{T} denotes the transposition of vectors. In obtaining this we have used the statistical independence of noise and the initial velocity, and have assumed that the initial velocity distribution is Maxwellian (→**2.3.2**).

Equilibrium dynamics should be time reversal symmetric (cf. **3.7.4**, **X2.3.2**), so for all t, the Langevin model of the Brownian dynamics tells us that

$$\langle \boldsymbol{v}(t,\omega)\boldsymbol{v}^T(0,\omega)\rangle = \frac{k_BT}{m} I e^{-(\zeta/m)|t|}. \tag{3.2.13}$$

Notice that this is not differentiable at $t = 0$. We will discuss its implication later (→**3.3.3**), but the reader should start thinking about it (cf., **3.1.6**). Here, it is simply pointed out that the inertia should make the time correlation differentiable at $t = 0$. Thus, the Langevin equation should not be reliable for a very short time scale.

Discussion 1. Experimentally, we can obtain some information about dynamics from the time correlation. Read the following paper and summarize the content:
M. A. Digman, P. Sengupta, P. W. Wiseman, C. M. Brown, A. R. Horwitz, and E. Gratton, "Fluctuation Correlation Spectroscopy with a Laser-Scanning Microscope: Exploiting the Hidden Time Structure," Biophys. J, **88**, L33-L36 (2005). □

3.2.8 Fluctuation-dissipation relation of the first kind: a prototype

Now, let us apply an external force $\boldsymbol{F}$ as described by (3.2.1). We use the noise satisfying the fluctuation-dissipation relation of the second kind (3.2.8).

It is easy to solve the equation as

$$\boldsymbol{v}(t,\omega) = \frac{1}{m}\int_0^t (\boldsymbol{F}(s) + \boldsymbol{w}(s,\omega))\, e^{-(\zeta/m)(t-s)} ds + \boldsymbol{v}(0)e^{-(\zeta/m)t}. \tag{3.2.14}$$

Its noise ensemble average (this is the average over the stochastic parameter ω) reads

$$\langle \boldsymbol{v}(t,\omega)\rangle = \frac{1}{m}\int_0^t \boldsymbol{F}(s) e^{-(\zeta/m)(t-s)} ds + \boldsymbol{v}(0)e^{-(\zeta/m)t}. \tag{3.2.15}$$

Furthermore, if we average over the initial condition that is sampled from an equilibrium ensemble, we obtain the response of the velocity of the Brownian particle in equilibrium:

$$\langle \boldsymbol{v}(t,\omega)\rangle = \frac{1}{m}\int_0^t \boldsymbol{F}(s) e^{-(\zeta/m)(t-s)} ds. \tag{3.2.16}$$

This implies that the response has the following form:

$$\langle \boldsymbol{v}(t,\omega)\rangle = \int_0^t \phi(t-s)\boldsymbol{F}(s) ds, \tag{3.2.17}$$

where ϕ is called the *response function* (in our case it is a 3×3 diagonal matrix) and is given by (cf. (3.2.12), **X2.2.1**)

$$\phi(t) = \langle \boldsymbol{v}(t,\omega)\boldsymbol{v}^T(0)\rangle / k_BT. \tag{3.2.18}$$

Here, the average is over ω (thermal noise) and the initial velocity $\boldsymbol{v}(0)$. Such a relation expressing the response function in terms of the correlation function of the quantity under consideration is called the *fluctuation-dissipation relation of the first kind* (→**X2.3.11**).

The reader must have thought that she had already encountered an analogous relation in equilibrium (fluctuation-response relation →**2.8.2**).

3.2.9 Prototype of Green-Kubo relation

Let us study the diffusion constant D of the Brownian particle. It is defined by (see **3.1.4**, **3.1.8**)[31]

$$D = \lim_{t\to\infty} \langle \boldsymbol{r}(t)^2\rangle/6t, \tag{3.2.19}$$

where $\boldsymbol{r}(t)$ is the position of the particle at time t that started from the origin at $t = 0$. Since $\boldsymbol{r}(t) = \int_0^t \boldsymbol{v}(s)ds$,

$$D = \lim_{t\to\infty} \frac{1}{6t}\int_0^t ds \int_0^t ds' \langle \boldsymbol{v}(s)\cdot\boldsymbol{v}(s')\rangle. \tag{3.2.20}$$

With the aid of l'Hospital's rule,[32]

$$D = \frac{1}{3}\int_0^\infty \langle \boldsymbol{v}(s)\cdot\boldsymbol{v}(0)\rangle ds. \tag{3.2.21}$$

That is, transport coefficients can be computed by the time integral of correlation functions. Such a relation is called the *Green-Kubo relation* (→**4.1.12**, **4.1.14**, **X2.3.12**). Note an intimate relation between the response function and the transport coefficient (3.2.18).

Exercise 1. Compute D for the Brownian particle of radius a in a viscous fluid of viscosity η with the aid of the Green-Kubo relation. □

Discussion 1. The self-diffusion constant of a liquid is not easy to measure, but viscosity is much easier to measure. If Einstein's relation (3.1.6) holds for a single atom or molecule, we could easily obtain the diffusion constant. How reliable is the result? Study noble gas liquids and simple organic liquids (e.g., benzene, methane, etc.). □

Discussion 2. In Discussion 1 try the slippery fluid velocity boundary condition. For argon it is known that the relation with the slippery boundary condition is quantitative. In any case, as the reader has seen there, the order of the diffusion constant is roughly correct. A possible argument explaining this success is dimensional analysis: the relevant length scale is only a, so if we perform dimensional analysis taking into account k_BT, η, and the general belief that the numerical constant appearing in the dimensional analytic result is of order unity, we must conclude that the Einstein relation is correct even for atoms. Is this a sound argument? □

3.2.10 Large deviation, Onsager's principle, and Langevin equation

This entry is largely a repetition of **1.6.15**. In **3.2.7**, it is pointed out that the Langevin equation is not reliable for short times. This implies that the microscopic mechanics time scale and the time scale t_M we observe the Brownian motion must be well separated, and that the time scale of the Langevin equation must be the intermediate time scale. Let us write this time scale as τ.

Suppose dx/dt is the true derivative (in the sense used in mechanics). Then, it fluctuates violently. However, if we average it for τ,

$$\frac{\Delta x}{\Delta t} \equiv \frac{1}{\tau}\int_0^\tau ds\,\frac{dx(t+s)}{dt} = \frac{x(t+\tau)-x(t)}{\tau} \tag{3.2.22}$$

fluctuates much less violently. Its expectation value is, however, not constant at the scale of t_M, and obeys a certain macroscopic law of evolution (regression principle →**3.1.7** and

---

[31] Here, 6 is $2 \times d$.

[32] We must assume that the system is infinitely large. This is an example of noncommutativity of space and time infinity limits. That is, macroscopic limit is a singular limit. Cf. **X2.3.7**.

4.1.2) on the average:

$$\left\langle \frac{\Delta x}{\Delta t} \right\rangle = f(x) \quad \text{in the mean.} \tag{3.2.23}$$

Here, $\langle\ \rangle$ is the equilibrium ensemble average.[33] Notice that for us macroscopic observers the mesoscopic time scale τ may be interpreted as infinitesimal dt, so we may write $\Delta x/\Delta t = dx/dt$ in Langevin equations.[34]

For a Brownian particle, the expectation value (the equilibrium ensemble average) of $d\boldsymbol{v}/dt$ averaged over τ, that is, the ensemble average of $(1/\tau)(\boldsymbol{v}(t+\tau) - \boldsymbol{v}(t))$, is $-\zeta\boldsymbol{v}(t)$.

We assume that the large deviation principle (→**1.6.2**) holds for the short time average introduced just above (denoted by an overline):

$$P(\overline{dx/dt} \sim \dot{x}|x(t)) \simeq e^{-\tau I(\dot{x}|x(t))}, \tag{3.2.24}$$

where $x(t)$ is written after $|$ to denote explicitly that the noise is studied for a given 'macroscopic state,' and I is the rate function, whose form we will discuss later. Here, the $x(t)$ dependence of the rate function I is explicitly denoted with the added $x(t)$ after the vertical line; the rate function is under the condition that the initial condition is $x(t)$. Let us call this *Onsager's principle*.[35] Roughly, we may say (for a more detail →**X1.1.3**):

> *Onsager's Principle*: Large deviation principle holds for the time average over the mesoscopic time scale.

We know the value $\dot{x}(t)$ that minimizes I (that is, $I = 0$) must be the most probable value we observe. This should be identical with the macroscopically observable result. This is Onsager's regression hypothesis. Thus, it may be better to call the following slightly restricted version as Onsager's principle:

> *Onsager's Principle (more conventional)*: Large deviation principle holds for the time average over the mesoscopic time scale. The rate function vanishes for the macroscopically expected result.

The rate function I is convex and vanishes for the most probable value (expectation value) (→**1.6.2**). Therefore, I must vanish if and only if the phenomenological law holds. This is the above restricted version. I is usually differentiable, so we may assume that near the origin I has a quadratic form. Therefore, for the velocity of a Brownian particle in a viscous medium, we can write

$$I(\dot{\boldsymbol{v}}(t)|\boldsymbol{v}(t)) = \frac{1}{4\zeta k_B T}(\dot{\boldsymbol{v}} + \zeta\boldsymbol{v})^2. \tag{3.2.25}$$

This rate function is compatible with the Langevin equation. That is, the Langevin equation is an expression of Onsager's principle in the Gaussian approximation. Here, the overall coefficient of the quadratic term is chosen to be compatible with what we already know (the fluctuation dissipation relation of the second kind), we will discuss how to determine this

---

[33] under the condition that $x(t)$ is specified; precisely speaking, short time average of $x(t)$ is specified.

[34] Do not confuse this notation and the 'true' time derivative in the integrand of (3.2.22).

[35] Needless to say, the naming is anachronistic, because large deviation theory being used in these lecture notes is much newer than Onsager. Hashitsume named the Onsager's original proposal the *Onsager's principle* after correcting it (→**4.1.19**); the large deviation theoretical form given here is due to the lecturer, because he believes that if Onsager knew large deviation theory, he should have written down this form as his fundamental proposal.

within Onsager's principle later (→**4.1.18**).

Since $\dot{x}(t)\tau = x(t+\tau) - x(t)$, $P(\dot{x}(t)|x(t)) = P(x(t+\tau)|x(t))$. Therefore, (3.2.24) may be interpreted as the *transition probability* of a Markov process:[36]

$$P(x(t+\tau)|x(t)) \simeq e^{-\tau I(\dot{x}(t)|x(t))}. \tag{3.2.26}$$

From this we can define a path integral (→**4.1.16**) that describes the solution to the Fokker-Planck equation.[37]

3.3 Langevin Equation: relation to mechanics

3.3.1 How realistic is Langevin model?

What is $-\zeta \boldsymbol{v}$ in the Langevin equation (→**3.2.2**)? Is the $\boldsymbol{v}$ really instantaneous velocity in the genuine sense of the word in mechanics? Its square average satisfies the equipartition of energy,[38] so it looks as a microscopic true velocity. However, the friction constant ζ which is usually obtained as $\zeta = 6\pi a\eta$ is not applicable to such an instantaneous velocity (cf. **3.1.5**). This implies that there is a certain inconsistency (or at least a mismatch) in the Langevin equation. This is saved by the 'artificial choice' of the noise amplitude according to the fluctuation dissipation relation (→**3.2.5**). Therefore, the equation should not be trusted for very short time scales.

We have noticed that the time correlation function of the velocity at the origin is not differentiable (→**3.2.7**); there is a cusp as $|t|$ there. Since velocity is differentiable due to inertia, this is *inconsistent* with (classical) mechanics.

Therefore, again we reach the same conclusion: the Langevin model of the Brownian motion must not be trusted for very short times (a time scale much shorter than the macroscopic scale).

This feature is very clear from the large deviation interpretation of the Langevin equation (→**3.2.10**); the Langevin equation is a result of short-time average of the microscopic dynamics.

3.3.2 Realistic energy consideration is possible with Langevin equation

It is stressed that the Langevin equation is reliable only in the time scale longer than the noise memory time scale in **3.3.1**. However, beyond the memory time scale of the noise the Langevin equation or, more generally, the models with stochastic noises could be sufficiently reliable.

(3.2.1) is a Newton's equation of motion, so the RHS must be the force acting on the Brownian particle. $\boldsymbol{F}$ is a systematic force exerted by an external agent. Then, the remaining force

$$\boldsymbol{F}_B \equiv -\zeta \boldsymbol{v} + \boldsymbol{w} \tag{3.3.1}$$

must be the force due to the interaction of the particle and its surrounding environment (fluid).

---

[36] See **6.2.1**; we will discuss this in a much more detailed fashion in Chapter 6.

[37] The reader should have recalled the relation between the Schrödinger equation and the Feynman path integral. Formally, they are equivalent.

[38] This is NOT a consequence of the equation, but is imposed. However, we must recognize that the Langevin equation can accommodate this requirement.

What is the work done by the force $\boldsymbol{F}_B$ on the Brownian particle? The resultant energy flow must be heat: the work done by $\boldsymbol{F}_B$ may be consistently interpreted as heat imported to the particle from the surrounding fluid.[39] This is the theory of *stochastic energetics* due to Sekimoto (→**X1.6**).[40] With this framework we can discuss the second law of thermodynamics and efficiency of thermal ratchets, etc.

3.3.3 Mechanics and stochastic processes

The reader may have complaints about the nature of the Langevin equation, because it is written down apparently independent of mechanics, although it is said that it is a Newton's equation of motion. As she has worried, it is even inconsistent with mechanics (→**3.3.1**). We certainly have the definite microscopic mechanical picture of a Brownian particle and its surrounding liquid medium. What is the relation between this true (many-body) mechanics and the Langevin equation that contradicts mechanics?

In this section one strategy (the 'projection strategy') is outlined. The form of the Langevin equation can be derived (or the equation of motion can be recast in this form; *Mori theory* →**3.3.6**). The basic idea is to single out the slow (or systematic) coordinates describing the center of mass of the Brownian particle and derive a closed equation for them with the aid of the formal solution to the equation of motion for the remaining variables (*Nakajima-Zwanzig projection method* →**3.3.4**). After this formal reduction we introduce an approximation that the time correlation function of the noise decays to zero instantaneously (white noise approximation or Markovian approximation) to reach the Langevin equation introduced in **3.2.2**. Thus, we see that the result may be reliable for longer time scales, but not so for microscopic time scales as expected.

Whenever we wish to derive a stochastic process from mechanics, we need some extra probabilistic assumptions (→**X1.2.10**; recall that even the Boltzmann equation needed such an assumption →**2.4.11**, **2.7.10**). However, we cannot freely assume an arbitrary stochastic process. In equilibrium the most important requirement is the consistency with the equilibrium distribution: the solution to the Langevin equation must be compatible with the equilibrium distribution after a sufficiently long time. We have just seen such a typical argument (*fluctuation-dissipation relation* →**3.2.5**).

The lesson we must learn is that a reasonable stochastic model can be constructed only when we know the equilibrium (or more generally, steady state) distribution, because we have no argument based solely on mechanics to model the noise.

This implies that if we are away from equilibrium, we have no firm basis for any stochastic model that is guaranteed by mechanics. No stochastic model can be used to answer fundamental questions of nonequilibrium statistical mechanics.

3.3.4 Idea of projection operator

We have learned that the essence of the Brownian motion is the possibility of separating two time scales, microscopic and macroscopic, clearly. Therefore, if we can remove rapid fluctuating motions from the true microscopic mechanical equation of motion of the system consisting of the Brownian particle and the surrounding water molecules, then the remaining dynamics would describe the systematic part (long time and large length scale motion) of the Brownian particle.

A natural idea is to project out the noise part of dynamics to obtain the systematic long time global dynamic.[41] This can be formally accomplished by applying a *projection* P

---

[39] As usual, the sign convention is: importing is positive.

[40] K. Sekimoto, *Stochastic Energetics* (Iwanami, Tokyo, 2004) [in Japanese].

[41] The idea of projection was introduced by Nakajima and (later) by Zwanzig and was utilized by Zwanzig and Mori (→**X1.2.6**) to derive mesoscopic dynamic equations.

to the equation of motion. P is a linear operator acting on a linear space of functions or vectors such that

$$PP = P. \tag{3.3.2}$$

If P is a projection, then $Q \equiv 1 - P$ is also a projection. Of course, $PQ = 0$.

Notice that local time averaging is a projection.

3.3.5 Kawasaki identity

We can write the equation of motion as

$$\frac{d}{dt}A = iLA \tag{3.3.3}$$

for both classical and quantum mechanics. L is often called the *Liouvillian* (Liouville operator). Here, A may be a phase function (classically) or an operator (quantum-mechanically).
Exercise 1. What is iL? (cf., **2.7.1**,**2.7.12**) □
Let P be a time-independent projection operator. Then, the following identity called the *Kawasaki identity*[42] holds:

$$\frac{d}{dt}e^{iLt} = e^{iLt}PiL + \int_0^t ds\, e^{iL(t-s)}PiLe^{QiLs}QiL + e^{QiLt}QiL. \tag{3.3.4}$$

where $Q = 1 - P$.
[Demo] The equation of motion for the evolution operator (propagator) is

$$\frac{d}{dt}e^{iLt} = e^{iLt}PiL + e^{iLt}QiL. \tag{3.3.5}$$

We wish to separate the 'noise' part $e^{QiLt}QiL$:

$$\frac{d}{dt}e^{iLt} = e^{iLt}PiL + (e^{iLt} - e^{QiLt})QiL + e^{QiLt}QiL \tag{3.3.6}$$

Now, we use the identity:

$$e^{iLt} - e^{QiLt} = \int_0^t ds\, e^{iL(t-s)}PiLe^{QiLs}, \tag{3.3.7}$$

and we get (3.3.4).

The easiest way to demonstrate (3.3.7) is to use Laplace transformation: $\mathcal{L}e^{iLt} = (s - iL)^{-1}$. Then, (3.3.7) is nothing but the so-called *resolvent identity*:

$$(s - A)^{-1} - (s - B)^{-1} = (s - A)^{-1}(A - B)(s - B)^{-1}. \tag{3.3.8}$$

Exercise 2. Demonstrate the above identity (note that A and B may not commute). □

By choosing an appropriate projection P, the Kawasaki identity becomes various known useful equations.[43]

---

[42] K. Kawasaki, J. Phys. A **6**, 1289 (1973).

[43] Here, the identity is introduced abruptly, but a more natural derivation may be found in Chapter X1.

3.3.6 Mori's generalized Langevin equation

Mori[44] cast Heisenberg's equation of motion (3.3.3) to quantum Langevin equation.

First, the projection operator P must be chosen physically correctly. Because (→**4.3.5**)

$$e^{-\beta(H-x_jX_j)} = e^{-\beta H}\left[1+\int_0^\beta ds\, e^{sH}x_je^{-sH}X_j+\cdots\right], \tag{3.3.9}$$

the effect of perturbation can be written as

$$\langle \delta x_i\rangle = \left\langle \int_0^\beta ds\, e^{sH}\delta x_j e^{-sH}\delta x_i\right\rangle X_j. \tag{3.3.10}$$

Thus, it is natural to introduce the following 'scalar product' symbol (called the *canonical correlation*)[45]

$$\langle A;B\rangle = \beta^{-1}\int_0^\beta ds\, Tr(\rho_e e^{sH}Ae^{-sH}B). \tag{3.3.11}$$

Here, ρ_e is the equilibrium distribution (density operator).

Let us define P as

$$PB = A\langle A;B\rangle/\langle A;A\rangle. \tag{3.3.12}$$

This may be expressed more conveniently as

$$P|B\rangle = |A\rangle\langle A|B\rangle/\langle A|A\rangle \tag{3.3.13}$$

with the interpretation of the bracket product as the canonical correlation.

Using this projection the Kawasaki identity reads *Mori's generalized Langevin equation*

$$\frac{d}{dt}|A(t)\rangle = i\Omega|A(t)\rangle + \int_0^t ds\, K_{AA}(t-s)|A(t)\rangle + |f(t)\rangle, \tag{3.3.14}$$

where

$$|f(t)\rangle = e^{itQLQ}QiL|A\rangle, \tag{3.3.15}$$

which is supposedly the noise term, and

$$K_{AA}(t) = \langle f(0)|f(t)\rangle/\langle A|A\rangle,\ i\Omega = \langle A|iL|A\rangle/\langle A|A\rangle. \tag{3.3.16}$$

The $i\Omega$ term represents the reversible term. The first relation is the *fluctuation-dissipation relation of the second kind* (→**3.2.5**) generalized to the case with memory.

3.3.7 Markov approximation of Mori's Langevin equation

(3.3.14) or

$$\frac{d}{dt}A(t) = i\Omega A(t) + \int_0^t ds\, K_{AA}(t-s)A(t) + f(t), \tag{3.3.17}$$

has a memory, so dA/dt is not determined by $A(t)$ (and $f(t)$) alone. That is, the equation does not describe a Markovian stochastic process. Actually, the equation is obtained by a

---

[44] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).

[45] That this correlation defines a scalar product may be seen from the general properties summarized in **X2.5.7**.

formal transformation of the equation of motion, so it is still a mechanical equation and has no direction of time. By discarding the memory, (3.3.17) becomes a (Markovian) stochastic equation. This procedure is called the *Markov approximation*. The result reads

$$\frac{d}{dt}A(t) = i\Omega A(t) - \zeta_A A(t) + f(t), \tag{3.3.18}$$

where

$$\zeta_A = \int_0^\infty K_{AA}(t)dt \tag{3.3.19}$$

and

$$\langle f(t)f(0)\rangle = 2k_BT\zeta_A\delta(t). \tag{3.3.20}$$

Notice that discarding short=time detailed information is crucial to derive stochastic equations.

3.3.8 Limitations of Mori's Langevin equation and projection onto space spanned by gross variables

Mori's Langevin equation projects the motion onto the space spanned by $|A\rangle$ (i.e., by an observable A). Therefore, e.g., A^2, which is not expressible as a constant multiple of A, is treated as nonsystematic noise. However, A^2 surely contains a systematic part that does not average out even if A averages out. Therefore, Mori's Langevin equation (3.3.14) is reliable only when $|A| \gg A^2$.

If we really wish to study a nonlinear system with noise, we must project the dynamics onto the space spanned by all the functions of macrovariables = gross variables $\boldsymbol{A}$. Let us discuss the classical cases only.

For the linear case the projection operator we use should reduce to Mori's choice (its classical version). Therefore, as its natural extension we introduce the set of (real) orthonormal polynomials $\phi_n(\boldsymbol{A})$ of gross variables $\boldsymbol{A}$ ($\phi_n(\boldsymbol{A})$ are multivariate polynomials of $\boldsymbol{A}$) such that

$$\int d\Gamma\rho_e(\Gamma)\phi_n(\boldsymbol{A}(\Gamma))\phi_m(\boldsymbol{A}(\Gamma)) = \delta_{nm}, \tag{3.3.21}$$

where ρ_e is the equilibrium distribution function on the phase space. The gross variables are considered as functions of microscopic variables Γ in the calculation of projection; $\boldsymbol{A}$ without explicit specification of Γ-dependence means gross variables as macroscopic variables. The projection operator $\mathcal{P}$ is defined as

$$\mathcal{P}X = \sum_n \phi_n(\boldsymbol{A})\int d\Gamma\rho_e(\Gamma)\phi_n(\boldsymbol{A}(\Gamma))X(\Gamma) = \sum_n \phi_n(\boldsymbol{A})\langle\phi_n(\boldsymbol{A}(\Gamma))X(\Gamma)\rangle, \tag{3.3.22}$$

where $\langle\ \rangle$ is the equilibrium average. On the right-hand side X is expressed in terms of mechanical (microscopic) variables, but after projection $\mathcal{P}X$ is a function of $\boldsymbol{A}$.

3.3.9 Zwanzig's nonlinear Langevin equation[46]

We use the projection introduced in **3.3.8** in the Kawasaki identity **3.3.5**. The result, called *Zwanzig's nonlinear Langevin equation*, reads (for a derivation, see **X1.3.5**)

$$\frac{d}{dt}A_j = v_j(\boldsymbol{A}(t)) + \int_0^t ds\frac{1}{P_e(\boldsymbol{A}(t-s))}\sum_k \frac{\partial}{\partial A_k(t-s)}\left[K_{kj}(s;\boldsymbol{A}(t-s))P_e(\boldsymbol{A}(t-s))\right] + f_j(t), \tag{3.3.23}$$

---

[46] R Zwanzig, *Memory effects in irreversible thermodynamics*, Phys. Rev. **124**, 983-992 (1961).

where

$$P_e(\boldsymbol{A}) = \int d\Gamma \delta(\boldsymbol{A} - \boldsymbol{A}(\Gamma))\rho_e(\Gamma). \tag{3.3.24}$$

and

$$\begin{aligned} v_j(\boldsymbol{A}) &= \langle iLA_j|\boldsymbol{A}\rangle, & (3.3.25)\\ K_{ij}(t;\boldsymbol{A}) &= \langle f_i f_j(t)|\boldsymbol{A}\rangle, & (3.3.26) \end{aligned}$$

and the conditional average is defined as

$$\langle B|\boldsymbol{A}\rangle = P_e(\boldsymbol{A})^{-1}\int \delta(\boldsymbol{A} - \boldsymbol{A}(\Gamma))B(\Gamma)\rho_e d\Gamma. \tag{3.3.27}$$

v_j is called the *streaming term* and K_{ij} is called the *memory kernel.* (3.3.26) is the fluctuation dissipation relation of the second kind (for a prototype see **3.2.5**). Very short time dynamics is governed by the streaming term, and is the instantaneous time-reversal part of dynamics. The streaming term has an important property that it does not affect the equilibrium distribution (→**X1.3.6**, cf. **X1.11.6**).

3.3.10 Markov approximation of Zwanzig's nonlinear Langevin equation

If the noise is close to the white noise, we may use the 'Markov approximation'[47] (see cautionary remarks in **X1.3.9**). If we write the equilibrium distribution as $P_e(\boldsymbol{A}) \propto e^{-\beta F}$, where F is the free energy, the nonlinear Langevin equation reads, ignoring the explicit $\boldsymbol{A}$-dependence of the Onsager coefficients, (3.3.23) reads

$$\begin{aligned} \frac{d}{dt}A_i &= v_i(\boldsymbol{A}) - \sum_j L_{ij}\frac{\partial F}{\partial A_j} + f_i, & (3.3.28)\\ &\langle f_i(0)f_j(t)\rangle = 2L_{ij}k_BT\delta(t). & (3.3.29) \end{aligned}$$

This is the starting point of the study of critical dynamics (mode coupling theory initiated by Fixman and fully developed by Kawasaki →**X1.3.8**).

3.4 Brownian Path and Stochastic Differential Equation

3.4.1 How zigzag is Brownian path?

As we have seen, Perrin said that the Brownian path is nowhere differentiable, and Einstein remarked that the instantaneous velocity of a Brownian particle is not definable (→**3.1.5**). However, since a Brownian particle is a classical object, if we magnify the time scale sufficiently, we should observe inertial effects. Then, the trajectory must be twice differentiable with respect to time.

The point is that there are two time scales: t_m the microscopic time scale of the molecules, and t_M the time resolution of our macroscopic observation (such as done by

---

[47] dA/dt is governed by the state of the system at time t; the effect of the past is only through the state at time t. There is no memory.

Perrin). For $t \sim t_m$ we should see full Newtonian mechanics behavior, but for $t > t_M$ we see what Perrin saw. Usually, $t_M/t_m \gg 1$, so mathematically perhaps it is better to idealize the situation by taking the $t_m \to 0$ limit.

This implies that we ignore the short-time inertial effect, and as we have already observed, the time correlation function of velocity loses differentiability at $t = 0$ (→**3.2.7**). In this limit the noise is accurately modeled by a white noise (→**3.7.3**). The velocity $\boldsymbol{v}$ of a particle obeying classical mechanics (i.e., no damping $\zeta = 0$ in (3.5.8)) driven by this white noise is the *mathematical Brownian motion* (*Wiener process* →**X1.4.4**; an intuitive introduction is in **3.4.2**).

How rugged is the graph of, say, $v_x(t)$ for this mathematical Brownian motion? As Perrin said it is nowhere differentiable with probability one (→**X1.4.6**). An interesting property is: if the reader lands on this graph, 'she will surely land on a peak'. That is, for any t we can take two times t_1 and t_2 on both sides of t : $t_1 < t < t_2$ within any positive ϵ (i.e., $t - t_1 < \epsilon$, $t_2 - t < \epsilon$) such that $v_x(t_1) < v_x(t) > v_x(t_2)$ (with probability one).[48]

A Brownian path is almost surely nondifferentiable and its Hausdorff dimension is 2.

3.4.2 Intuitive introduction to Wiener process

Let us start with the Langevin equation summarized in **3.2.6** without any systematic force. We can write it as

$$\frac{m}{\zeta}\frac{d^2\boldsymbol{r}}{dt^2} + \frac{d\boldsymbol{r}}{dt} = \boldsymbol{\nu}(t), \tag{3.4.1}$$

where $\boldsymbol{\nu}$ is a Gaussian stationary noise with $\langle \boldsymbol{\nu} \rangle = 0$ and $\langle \boldsymbol{\nu}(t)\boldsymbol{\nu}^T \rangle = (2k_BT/\zeta)\delta(t)$. If the mass is small and the viscosity is large enough, we may ignore the inertial effect.[49] That is, we have only to study

$$\frac{d\boldsymbol{r}}{dt} = \boldsymbol{\nu}. \tag{3.4.2}$$

This is a model of what Brown and Perrin observed (see an illustration in **3.1.8**); they saw only the position of the particle at long time scales. Notice that without inertia the velocity need not be differentiable. Solving this equation is easy:

$$\boldsymbol{r}(t) = \boldsymbol{r}(0) + \int_0^t ds\, \boldsymbol{\nu}(s). \tag{3.4.3}$$

From this, we obtain (let us assume that the initial position is the origin)

$$\langle \boldsymbol{r}(t)^2 \rangle = \int_0^t ds \int_0^t ds' \,\langle \boldsymbol{\nu}(s) \cdot \boldsymbol{\nu}(s') \rangle = \frac{2k_BT}{\zeta} d\,t = 2dDt. \tag{3.4.4}$$

Here, $D = k_BT/\zeta$ (Einstein's relation →**3.1.3** with $\zeta = 6\pi a\eta$). Einstein concluded from this that the Brownian trajectory was not differentiable (its velocity was not well defined) (→**3.1.6**). Thus, the differential equation (3.4.2) is self-contradictory.

To clarify the problem further, let us consider $B(t)$ formally introduced as

$$\frac{dB(t)}{dt} = w(t), \tag{3.4.5}$$

---

[48] Adelman, Israel J. Math. **50**, 189 (1985) is a good summary; the statement is due to Dvoretsky, Erdös and Kakutani.

[49] If you are interested in the time scale beyond m/ζ, this can be justified.

where w is a Gaussian noise[50] with mean zero and

$$\langle w(t)w(s)\rangle = \delta(t-s). \tag{3.4.6}$$

We already know that (3.4.5) does not make sense. How about, then

$$B(t) = \int_0^t w(s)ds\ ? \tag{3.4.7}$$

We can formally calculate $\langle B(t)\rangle = 0$ and

$$\langle B(t)B(s)\rangle = \min\{t,s\}. \tag{3.4.8}$$

We can define a Gaussian process $B(t)$ by these expectation values. This is called the *Wiener process* (→**X1.5.7**).

We will see later that w may be understood as white noise. Intuitively,

$$w(t) = \sum_\omega c_\omega e^{i\omega t}, \tag{3.4.9}$$

where c_ω obeys $N(0,1)$ and statistically independent for different ω's.[51] We have indeed

$$\langle w(t)w(s)\rangle = \sum_\omega\sum_{\omega'}\langle c_w c_{\omega'}\rangle e^{i\omega t+i\omega' s} = \sum_\omega e^{-i\omega(t-s)} \propto \delta(t-s). \tag{3.4.10}$$

Thus, intuitively, the Wiener process $B(t)$ is an integral of white noise; it is continuous but not differentiable. As already mentioned in **3.4.1** the sample path is extremely rugged.

3.4.3 How to interpret Langevin equation for velocity

We have seen that (3.4.2) is not well defined but its formal integrated version can make sense. Then, we should follow the same idea to make sense of the Langevin equation summarized in **3.2.6**:

$$\boldsymbol{v}(t) = \boldsymbol{v}(0)e^{-t/\zeta} + \int_0^t e^{-(t-s)/\zeta}\sqrt{2k_BT\zeta}\,d\boldsymbol{B}(s). \tag{3.4.11}$$

Here, in the last term a formal calculation

$$\int_0^t e^{-(t-s)/\zeta}\boldsymbol{w}(s)ds = \int_0^t e^{-(t-s)/\zeta}d\boldsymbol{B}(s). \tag{3.4.12}$$

is used. $\boldsymbol{B}$ is a 3D version of the Wiener process.[52] It is not hard to show that this is mathematically respectable.[53]

Our trouble with w and B is due to its white-noise limit; this causes extremely rapid violent changes in functions, and we must stretch the use of the ordinary calculus as we have managed up to this point. We could do this because the noise amplitude is constant. If not, we are in real trouble; if we wish to take the white noise limit *before* solving the differential equations, we have a severe difficulty.

---

[50] Precisely speaking, that $A(t)$ is a Gaussian process means that for any countably many time points $\{t_k\}$, the simultaneous distribution of the sample values $\{A(t_k)\}_{k=1}^N$ is a N-dimensional Gaussian distribution.

[51] Here, c_w is a complex number, so $\langle c_w c_{\omega'}\rangle = \delta_{\omega,-\omega'}$.

[52] If we have 1D Wiener process, then we prepare three statistically independent 1D Wiener processes and regard them as components of the vector $\boldsymbol{B}(t)$.

[53] For $\int f(t)dg(t)$ to exist for any continuous function f g must be of bounded variation. Unfortunately, $B(t)$ is not of bounded variation, but if f is of bounded variation, we may formally use an integration by parts to write $\int_0^t f(s)dg(s) = f(t)g(t) - f(0)g(0) - \int_0^t g(s)df(s)$ that makes sense.

3.4.4 What is stochastic or Itô calculus?

As we have discussed in **3.4.3** a mathematically respectable interpretation of the Langevin equation (3.2.1) (here we pay attention only to one component) is in terms of a *stochastic differential equation* (the stochastic parameter ω is suppressed)

$$dv = -\zeta v dt + \sqrt{2k_BT\zeta} dB, \tag{3.4.13}$$

that is interpreted as a short-hand notation of the following integral equation:

$$v(t) = v(0) - \zeta \int_0^t ds\, v(s) + \sqrt{2k_BT\zeta} B(t), \tag{3.4.14}$$

where B is the Wiener process.

Mathematically, (3.4.13) is an example of the general stochastic differential equation (SDE)

$$dx = f(x,t)dt + \sigma(x,t)dB, \tag{3.4.15}$$

that is understood as a short-hand notation of the following integral equation:

$$x(t) = x(0) + \int_0^t f(x(s),s)ds + \int_0^t \sigma(x(s),s)dB(s). \tag{3.4.16}$$

The problem now is that σ is no more constant, and contains nondifferentiable x, so the (formal) Riemann-Stieltjes integral in (3.4.16) (the last term) does not make sense. If the noise is physical (without strong high frequency components), then we may use the ordinary calculus to compute (3.4.16), but we know memory terms are the source of nuisance, so we wish to take the memoryless white-noise limit before any calculation. This simplification cannot be obtained for free.

The minimum price we must pay may be *stochastic calculus.* It is defined so that the above diagram becomes commutative. The last integral in (3.4.16) with respect to B is defined as an *Itô integral.* That is, a new calculus (*Itô calculus* →**X1.5.6**) is needed.[54] However, practically (at least for physicists and other practitioners), its intuitive understanding and mastering one or two calculation rules must be enough (i.e., the rest of this entry, **3.4.5**, **3.4.6** and a summary **3.4.7**).

The most notable feature of (3.4.15) is that the noise increment dB is statistically independent of x up to time t (i.e, dB is *non-anticipating*). The physical noise with finite correlation time cannot have this property, so the calculation of the cross terms such as $f(x)dB$ is a source of nuisance; its effect remains even after the white noise limit.[55] With the Itô calculus, however, this problem can be avoided completely.

---

[54] The reason is, as already noted in other entries, that this formal Stieltjes integral is not uniquely defined in the usual Riemann-Stieltjes fashion, because B is not of bounded variation.

[55] This is a big reason why the white noise limit and the calculus do not commute.

3.4.5 Itô's lemma, a preview

Perhaps, the most conspicuous feature of Itô calculus is the following rule (*Itô's lemma* →**X1.5.10**). Let x be a solution to (3.4.15) and u a smooth function of x and t. Then,[56]

$$du(x,t) = \frac{\partial u}{\partial t}dt + \frac{\partial u}{\partial x}dx + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}dx^2 \tag{3.4.17}$$

with the interpretation $dx^2 = \sigma^2 dt$ (i.e. $dB^2 = dt$). That is, (3.4.17) is the following stochastic differential equation:

$$du(x,t) = \left(\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}f(x,t) + \frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2}\right)dt + \frac{\partial u}{\partial x}\sigma(x,t)dB. \tag{3.4.18}$$

This is physically very appealing, because the small displacement dx is of order $\sqrt{dt}$ as Einstein pointed out (→**3.1.6**).

3.4.6 Physical and mathematical Langevin equations

As we have discussed, physical noises are all of bounded variation before idealization. Mathematicians write physical Langevin equations with physical noise as

$$dx = a(x,t)dt + \sigma(x,t)\circ dB, \tag{3.4.19}$$

where $\circ$ is used to denote clearly that the term is NOT interpreted in Itô's sense, but in the usual calculus sense. Since Itô's lemma tells us

$$d\int_0^{B(t)}\sigma(x)dx = \sigma(B(t))dB(t) + \frac{1}{2}\left.\frac{d\sigma}{dx}\right|_{x=B}dt, \tag{3.4.20}$$

and since the LHS can be calculated as the ordinary calculus formula, the $\circ$ product must be interpreted as

$$\sigma(B(t))\circ dB = \sigma(B(t))dB(t) + \frac{1}{2}\sigma'(B(t))dt. \tag{3.4.21}$$

This implies that, if we wish to apply Itô calculus to (3.4.19), we must interpret (→**X1.5.13**)

$$\sigma(x,t)\circ dB = \sigma(x,t)dB + \frac{1}{2}\frac{\partial\sigma}{\partial x}dxdB = \sigma(x,t)dB + \frac{1}{2}\sigma\frac{\partial\sigma}{\partial x}dt. \tag{3.4.22}$$

Thus, the physical Langevin equation (3.4.19) with the noise term dependent on x must be mathematically interpreted as

$$dx = \left(a(x,t) + \frac{1}{2}\sigma\frac{\partial\sigma}{\partial x}\right)dt + \sigma(x,t)dB. \tag{3.4.23}$$

The relation between (3.4.19) and (3.4.23) is called *Wong-Zakai's theorem*.

---

[56] The equality is in the almost sure sense. That is, with probability 1.

3.4.7 Practical summary of Itô calculus

Let

$$dx = f(x,t)dt + \sigma(x,t)dB. \tag{3.4.24}$$

Then,
(1) dB is statistically independent of $x(t)$ (consequently, independent of $\sigma(x(t),t)$), and
(2) when you calculate $du(x,t)$, regard $dx \sim \sqrt{dt}$[57] and truncate at $O[dt]$. Then, replace $(dx)^2$ with $\sigma^2 dt$.

To apply Itô calculus, the Langevin equation must be rewritten as a proper stochastic differential equation. See **3.4.6** or **X1.5.13**.

Some examples follow:

$$d(x^2) = 2xdx + (dx)^2 = (2xf + \sigma^2)dt + 2x\sigma dB. \tag{3.4.25}$$

This implies that $d\langle x^2\rangle/dt = \langle 2xf + \sigma^2\rangle$.

$$de^{B(t)-t/2} = e^{B(t)-t/2}\left(dB(t) - \frac{1}{2}dt\right) + \frac{1}{2}e^{B(t)-t/2}dB(t)^2 = e^{B(t)-t/2}dB(t). \tag{3.4.26}$$

This implies $\left\langle e^{B(t)-t/2}\right\rangle = 1$ or $\left\langle e^{B(t)}\right\rangle = e^{t/2}$. This is, of course, directly obvious from the definition of $B(t)$.

3.4.8 Simple example of Wong-Zakai's theorem

Let us consider an oscillator whose frequency fluctuates: we introduce $z = x + iy$, where (x,y) is the (appropriately normalized) canonical coordinates. The equation of motion may be modeled (physically) as

$$\frac{dz}{dt} = i\left(1 + \sqrt{2\zeta}w\right)z. \tag{3.4.27}$$

This is supposedly a physical model, so should be interpreted as a physical Langevin equation:

$$dz = izdt + i\sqrt{2\zeta}z \circ dB. \tag{3.4.28}$$

With the aid of Wong-Zakai's theorem (→**3.4.6**, or **X1.5.13**), we have

$$dz = (i - \zeta)zdt + i\sqrt{2\zeta}zdB. \tag{3.4.29}$$

Therefore,

$$d\langle z\rangle = (i - \zeta)\langle z\rangle dt. \tag{3.4.30}$$

Notice that the average decays to zero. Its implication to irreversibility will be discussed in **3.5.12**.

Exercise 1. Show that $d|z|^2 = 0$. □

---

[57] Actually, it may be written as $\sqrt{dt}\chi$, where χ obeys $N(0,1)$ and is statistically independent from any time increment at other instant.

3.4.9 Numerical computation of Langevin equations: simplest scheme

If (3.2.1) is not solvable analytically as is often the case with a force $\boldsymbol{F}$, we need a numerical scheme to solve this equation. A modification of the simple Euler scheme works that is also helpful in building our intuition. For simplicity, we consider the following one dimensional equation

$$\frac{dx}{dt} = -\gamma x + \phi(x) + w. \tag{3.4.31}$$

Integrating this from t to $t + \delta t$, we obtain

$$x(t+\delta t) - x(t) = -\gamma \int_0^{\delta t} d\delta s\, x(t+\delta s) + \int_0^{\delta t} d\delta s\, w(t+\delta s). \tag{3.4.32}$$

The expectation value of the last term is zero, but its variance is of order δt:

$$\left\langle \left[\int_0^{\delta t} d\delta s\, w(t+\delta s) \right]^2 \right\rangle = 2\alpha\delta t. \tag{3.4.33}$$

This implies that the 'instantaneous value' of the last term of (3.4.32) is of order $\sqrt{\delta t}$, which is much bigger than δt. We may replace this with (approximation in law!)

$$\int_0^{\delta t} d\delta s\, w(t+\delta s) = \sqrt{2\alpha\delta t}\chi, \tag{3.4.34}$$

where χ is a Gaussian random variable obeying $N(0,1)$,[58] and is uncorrelated with increments at different times.

This implies that $x(t+\delta s) - x(t)$ is not of order δs but of order $\sqrt{\delta s}$, but still we may ignore it in the first term of (3.4.32): we have arrived at the following difference scheme[59]

$$x(t+\delta t) = x(t) - gx(t)\delta t + \sqrt{2\alpha\delta t}\chi. \tag{3.4.35}$$

Exercise 1. The above equation is correct to order δt, so the error should not accumulate to order 1 after integration. However, this is not always good enough. For simplicity, set $g = 0$ and study the time evolution of $x^2(t)$ with this difference scheme.
(1) What do you have to conclude?
(2) You might say that such a pathology is not harmful, if $g > 0$. Really?[60] □

3.5 Introduction to Fokker-Planck Equation

The ensemble of Brownian particles may be described by their density distribution function. Its time evolution is governed by the Fokker-Planck equation (or by the Smoluchowski equation when overdamped).

---

[58] Mnemonics: the square of a random variable obeying $N(0,1)$ obeys the χ^2-distribution.

[59] For a crude simulation, this may be OK, but this scheme has a defect of systematic increase of $\langle x^2\rangle$, if the systematic part is not linear. For the simple equation without external force, this is fine.

[60] Sekimoto's comment (1999): The wrong difference scheme produces its own heat, so even if the solution may look all right with damping, the resultant state may not realize an equilibrium state, but a driven steady state.

3.5.1 Generalized Liouville equation

To study the time evolution of the density distribution function of the Brownian particle position, first, we discuss the so-called generalized Liouville's equation approach.

Suppose an equation of motion is given as a differential equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)). \tag{3.5.1}$$

Instead of following each trajectory (solution curve), we could follow the cloud of points; this is the basic idea of ensemble. This would give us an equation governing the density distribution on the phase space (= the space spanned by the basic dynamical variables $\boldsymbol{x}$).

Suppose μ is a measure (= distribution) on the phase space. Since no sample disappears (→**2.7.7**), the equation of motion of the measure reads

$$\mu_t(d\boldsymbol{x}(t)) = \mu_{t+dt}(d\boldsymbol{x}(t+dt)). \tag{3.5.2}$$

Here, $d\boldsymbol{x}$ denotes symbolically the volume element around $\boldsymbol{x}$ riding on the solution curves (on the flow determined by the equation of motion). If we introduce its density[61] $\phi(t, \boldsymbol{x})$, this equation reads

$$\phi(t+dt, \boldsymbol{x}(t+dt))\frac{\partial(\boldsymbol{x}(t+dt))}{\partial(\boldsymbol{x}(t))} = \phi(t, \boldsymbol{x}(t)), \tag{3.5.3}$$

where $\partial(\boldsymbol{x}(t+dt))/\partial(\boldsymbol{x}(t))$ is the Jacobian and is given by (to the first order in dt)[62]

$$\frac{\partial(\boldsymbol{x}(t+dt))}{\partial(\boldsymbol{x}(t))} = 1 + div\boldsymbol{f}dt. \tag{3.5.4}$$

In our case we usually do not have an analogue of Liouville's theorem ($div\boldsymbol{f} = 0$ for a Hamiltonian dynamical system, →**2.7.2**), so we cannot ignore the second term.

Combining (3.5.3) and (3.5.4), we obtain

$$\frac{\partial \phi}{\partial t} + div(\boldsymbol{f}\phi) = 0. \tag{3.5.5}$$

This is the *generalized Liouville equation.*

Exercise 1. Show that (3.5.5) may be rewritten as

$$\frac{d\log \phi}{dt} = -div\boldsymbol{f}. \tag{3.5.6}$$

$div\boldsymbol{f}$ is called the *phase compression factor.* [Notice that the time derivative on the LHS is not a partial derivative.]□

Exercise 2. For $dq/dt = [q, H]$, $dp/dt = [p, H] - \alpha p$, where H is the system Hamiltonian, find the phase compression factor. □

---

[61] As usual, we allow densities in the distribution sense, so even if μ_t is not absolutely continuous wrt the Riemann volume of the phase space (i.e., no Radon-Nikodym derivative exists in the conventional sense), we freely introduce the density.

[62] Note that $det(1 + \epsilon A) = 1 + \epsilon TrA + \cdots$, and that the trace of the Jacobi matrix is the divergence.

3.5.2 Stochastic Liouville equation

The position $\boldsymbol{r}$ of a Brownian particle is described by (→**3.2.2**)[63]

$$\frac{d\boldsymbol{r}(t)}{dt} = \boldsymbol{v}(t), \tag{3.5.7}$$

$$\frac{\partial \boldsymbol{v}(t)}{\partial t} = -\frac{\zeta}{m}\boldsymbol{v}(t) + \frac{\boldsymbol{F}}{m} + \frac{\boldsymbol{w}(t,\omega)}{m}. \tag{3.5.8}$$

Here, $\boldsymbol{F}$ is the external force, and as before, ζ is the friction constant, m is the particle mass, and $\boldsymbol{w}$ is the fluctuating force (noise). Let us apply the generalized Liouville equation (→**3.5.1**) straightforwardly to this coupled equation: our phase space is spanned by $\boldsymbol{r}$ and $\boldsymbol{v}$, so the probability density function ϕ is a function of these variables and time. The obtained equation[64]

$$\frac{\partial \phi}{\partial t} + \frac{\partial}{\partial \boldsymbol{v}} \cdot \left[\left(-\frac{\zeta}{m}\boldsymbol{v} + \frac{\boldsymbol{F}}{m}\right)\phi\right] + \frac{\partial}{\partial \boldsymbol{v}} \cdot \left[\frac{\boldsymbol{w}}{m}\phi\right] + \frac{\partial}{\partial \boldsymbol{r}} \cdot [\boldsymbol{v}\phi] = 0 \tag{3.5.9}$$

is called a *stochastic Liouville equation*. Notice that the equation depends on the stochastic parameter (→**1.4.2**) through $\boldsymbol{w}$.

3.5.3 Derivation of Fokker-Planck equation

We wish to average (3.5.9) over the stochastic parameter (average over the noise ensemble). Since ϕ is a solution to this equation, it also depends on ω, so the technical point is how to compute the cross term $\langle \boldsymbol{w}\phi\rangle$.[65] An important point to realize is that $\boldsymbol{w}$ is, intuitively, of order $\sqrt{dt}$ as can be seen from $\langle(\int_t^{t+\Delta t}\boldsymbol{w}dt)^2\rangle \sim \Delta t$.[66]

Perhaps the most transparent method (if we avoid Itô calculus; if not →**3.5.4**) to compute the noise average is to use the formal solution for ϕ at time $t + dt$ in terms of that at time t written as

$$\phi|_{t+dt} = e^{dt\mathcal{L}_0 + dt\mathcal{L}_1}\phi|_t, \tag{3.5.10}$$

where the differential operators are defined as

$$\mathcal{L}_0 = -\frac{\partial}{\partial \boldsymbol{v}} \cdot \left(-\frac{\zeta}{m}\boldsymbol{v} + \frac{\boldsymbol{F}}{m}\right) - \frac{\partial}{\partial \boldsymbol{r}}\boldsymbol{v},\ \mathcal{L}_1 = -\frac{\partial}{\partial \boldsymbol{v}} \cdot \frac{\boldsymbol{w}}{m}. \tag{3.5.11}$$

Although these two operators do not commute, since dt is infinitesimal, we may approximate $\exp(dt\mathcal{L}_0 + dt\mathcal{L}_1) = \exp(dt\mathcal{L}_0)\exp(dt\mathcal{L}_1) + o(dt)$.[67] We obtain[68]

$$\langle e^{dt\mathcal{L}_1}\rangle = \exp\left(\frac{1}{2}\left\langle (dt\mathcal{L}_1)^2\right\rangle\right), \tag{3.5.12}$$

---

[63] Since $\boldsymbol{w}$ depends on the stochastic parameter ω, $\boldsymbol{r}$ and $\boldsymbol{v}$ must also depend on ω, but ω is not explicitly written in them.

[64] $\frac{\partial}{\partial \boldsymbol{x}}$ is used interchangeably with $\nabla_{\boldsymbol{x}}$.

[65] The reader might say that there are other terms such as $\boldsymbol{v}\phi$, but in this partial differential equation $\boldsymbol{v}$ is an independent variable, not a function of time or the stochastic parameter. Incidentally, the calculation of the cross terms becomes transparent with the aid of stochastic calculus (→**3.4.5**, **X1.5.14**).

[66] We have already encountered this in **3.4.5**. If you are not satisfied with an intuitive explanation, see **X1.5.9**.

[67] Although obvious, it is a very important observation basic to Trotter's formula, important in constructing path integrals (→**4.3.5**).

[68] We need an extracondition to guarantee that the tail of the noise is not very fat, and the moments (cumulants, more precisely, →**1.4.11**) higher than the second order must be assumed to be ignored; if we assume the noise to be Gaussian, there is no problem at all.

$$= \exp\left(\frac{1}{2m^2}\frac{\partial}{\partial \boldsymbol{v}}\langle \boldsymbol{w}dt\boldsymbol{w}^T dt\rangle \frac{\partial}{\partial \boldsymbol{v}}\right) = \exp\left(dt\frac{\zeta k_B T}{m^2}\frac{\partial^2}{\partial \boldsymbol{v}^2}\right). \tag{3.5.13}$$

Here, $\langle \boldsymbol{w}\boldsymbol{w}^T\rangle dt = 2\zeta k_B T I$ (→(3.2.8)) has been used (i.e., $\delta(t)dt = 1$, formally). Therefore, if we introduce $P(\boldsymbol{r}, \boldsymbol{v}, t) = \langle \phi(\boldsymbol{r}, \boldsymbol{v}, t)\rangle$, the outcome of averaging reads

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial \boldsymbol{r}}\boldsymbol{v}P - \frac{\partial}{\partial \boldsymbol{v}}\left[-\frac{\zeta}{m}\boldsymbol{v} + \frac{\boldsymbol{F}}{m} - \frac{k_B T\zeta}{m^2}\frac{\partial}{\partial \boldsymbol{v}}\right]P, \tag{3.5.14}$$

which is called the *Fokker-Planck equation* (or this particular equation for the Brownian particle is sometimes called the *Kramers equation*).

Exercise 1. The more standard way to derive the Fokker-Planck equation is to use Chapman-Kolmogorov equation as we did in **3.1.10**:

$$P(\boldsymbol{r}, \boldsymbol{v}, t + dt) = \int d\boldsymbol{\ell} d\boldsymbol{u}\, \phi(\boldsymbol{\ell}, \boldsymbol{u}|\boldsymbol{r} - \boldsymbol{\ell}, \boldsymbol{v} - \boldsymbol{u})P(\boldsymbol{r} - \boldsymbol{\ell}, \boldsymbol{v} - \boldsymbol{u}, t), \tag{3.5.15}$$

where ϕ is the transition probability between t and $t + dt$. This transition probability may be obtained from the stochastic equations (3.5.7) and (3.5.8) (Hint: **3.4.9**). Derive (3.5.14) with the aid of (3.5.15).[69]□

Exercise 2. Demonstrate that the Maxwell-Boltzmann distribution is the equilibrium solution for (3.5.14) (cf. **3.5.9**). Discuss that this condition implies the fluctuation-dissipation relation of the second kind (→**3.2.5**). [That is, assume that the noise is generally defined as **3.2.3**, and then impose that the Maxwell-Boltzmann distribution is the stationary solution. What is the condition imposed on the noise amplitude?] □

3.5.4 Derivation of Fokker-Planck equation via Itô calculus

The most standard derivation is in **3.5.3**, but if we use the Itô calculus (especially, Itô's lemma →**3.4.5**), then the derivation becomes very transparent (easy). We start with the definition of the density distribution (→**1.4.5**)

$$P(\boldsymbol{r}, \boldsymbol{v}, t) = \langle \delta(\boldsymbol{r} - \boldsymbol{r}(t))\delta(\boldsymbol{v} - \boldsymbol{v}(t))\rangle, \tag{3.5.16}$$

where the average is over the noise; $\boldsymbol{r}$ and $\boldsymbol{v}$ without t dependence are the variables in the density distribution and $\boldsymbol{r}(t)$ and $\boldsymbol{v}(t)$ are dynamical variables obeying the equation of motion (3.5.7) and (3.5.8). Thus, $\boldsymbol{r}(t)$ and $\boldsymbol{v}(t)$ depend on the stochastic parameter ω. $\boldsymbol{w}dt$ is mathematically $\sqrt{2\zeta k_B T}d\boldsymbol{B}(t)$, where $\boldsymbol{B}$ is a three dimensional vector whose components are independent Wiener processes (i.e., 3D Wiener process). We use Itô's lemma as

$$d[\delta(\boldsymbol{r} - \boldsymbol{r}(t))\delta(\boldsymbol{v} - \boldsymbol{v}(t))] = \left[-d\boldsymbol{r}(t)\frac{\partial}{\partial \boldsymbol{r}} - \boldsymbol{v}(t)\frac{\partial}{\partial \boldsymbol{v}} + (d\boldsymbol{v}(t))^2\frac{1}{2}\frac{\partial^2}{\partial \boldsymbol{v}^2}\right][\delta(\boldsymbol{r} - \boldsymbol{r}(t))\delta(\boldsymbol{v} - \boldsymbol{v}(t))]. \tag{3.5.17}$$

Here, $\boldsymbol{r}(t)$ is at least once differentiable, so we may ignore $(d\boldsymbol{r}(t))^2$. However, $\boldsymbol{v}(t)$ is not smooth, so we need the second order term:

$$dv_i(t)dv_j(t) = \frac{2\zeta k_B T}{m^2}\delta_{ij}dt. \tag{3.5.18}$$

---

[69] A more general discussion may be found in **6.2.16**.

Precisely speaking, the third term in the square brackets must be interpreted as

$$(d\boldsymbol{v}(t))^2\frac{1}{2}\frac{\partial^2}{\partial \boldsymbol{v}^2} = \frac{1}{2}dv_i(t)dv_j(t)\frac{\partial}{\partial v_i}\frac{\partial}{\partial v_j}. \tag{3.5.19}$$

Thus, we see

$$d[\delta(\boldsymbol{r}-\boldsymbol{r}(t))\delta(\boldsymbol{v}-\boldsymbol{v}(t))] = \frac{\partial}{\partial \boldsymbol{r}}\left(-\boldsymbol{v}(t)dt\right)[\delta(\boldsymbol{r}-\boldsymbol{r}(t))\delta(\boldsymbol{v}-\boldsymbol{v}(t))]$$

$$+\frac{\partial}{\partial \boldsymbol{v}}\left[-\frac{\zeta}{m}\boldsymbol{v}(t)dt + \frac{\boldsymbol{F}}{m}dt + \frac{\sqrt{2\zeta k_BT}}{m}dB(t)\right][\delta(\boldsymbol{r}-\boldsymbol{r}(t))\delta(\boldsymbol{v}-\boldsymbol{v}(t))] + \frac{\partial^2}{\partial \boldsymbol{v}^2}\left[\frac{\zeta k_BT}{m^2}dt\right]\delta[(\boldsymbol{r}-\boldsymbol{r}(t))\delta(\boldsymbol{v}-\boldsymbol{v}(t)$$

$$. \qquad (3.5.2$$

Taking the expectation value of (3.5.20), we obtain the Fokker-Planck equation (3.5.14); notice that the average of the term containing dB simply disappears due to its non-anticipating nature.

3.5.5 Smoluchowski equation

In many cases, the effect of inertia is unimportant,[70] so we may ignore $\dot{\boldsymbol{v}}(t)/\zeta$ in (3.5.8) (for a detail →**X1.7.4**). Thus, we obtain the following reduced equation that is also called a Langevin equation — physicists call all the equations with noise terms (especially additive noise terms) Langevin equations:

$$\frac{d\boldsymbol{r}}{dt} = \frac{1}{\zeta}(\boldsymbol{F}+\boldsymbol{w}) = \frac{1}{\zeta}\boldsymbol{F}+\boldsymbol{\nu}, \tag{3.5.21}$$

where we have introduced the new noise $\boldsymbol{\nu}$ which has mean zero and the time correlation function (cf. (3.2.8))

$$\langle \boldsymbol{\nu}(t)\boldsymbol{\nu}^T(0)\rangle = (2k_BT/\zeta)I\delta(t). \tag{3.5.22}$$

Notice that the fluctuation-dissipation relation of the second kind holds (→**3.2.5**).[71]

Now, we can apply the stochastic Liouville equation (→**3.5.2**) to obtain the equation for the position distribution of the particles

$$\frac{\partial P(\boldsymbol{r})}{\partial t} = \frac{\partial}{\partial \boldsymbol{r}}\left(-\frac{\boldsymbol{F}}{\zeta}+\frac{k_BT}{\zeta}\frac{\partial}{\partial \boldsymbol{r}}\right)P. \tag{3.5.23}$$

This reduced Fokker-Planck equation is often called the *Smoluchowski equation*.[72]

3.5.6 Does Smoluchowski equation have a solution?

Generally, a constant diffusion coefficient equation

$$\frac{d\phi}{dt} = D\Delta\phi - \nabla\cdot(\boldsymbol{v}\phi), \tag{3.5.24}$$

where $\boldsymbol{v}$ is a vector field on an oriented manifold has a solution for initial value problems. It is shown that if $\boldsymbol{v}$ is C^m ($m \geq 1$), then the equation has a unique steady state (of smoothness C^{m-1}) to which all the solutions tend.[73]

---

[70] For example, for proteins the dynamics slower than the time scale of 1ps is surely overdamped.

[71] In this case the relation may also be interpreted as a fluctuation-dissipation relation of the first kind.

[72] There is a way to reduce the Fokker-Planck equation directly to the Smoluchowski equation (→**X1.7.5**).

[73] C. S. C. Lin, B. Yang, and F. Chen, "On stability of a dynamical system," SIAM J. Math. Anal. **26**, 428-435 (1995).

3.5.7 Thermodynamic interpretation of Smoluchowski equation

(3.5.23) can be rewritten as

$$\partial P/\partial t = -div\boldsymbol{J} \tag{3.5.25}$$

with

$$\boldsymbol{J} = \left[\frac{\boldsymbol{F}}{\zeta} - \left(\frac{\partial}{\partial \boldsymbol{r}}\frac{k_BT}{\zeta}\log P\right)\right]P, \tag{3.5.26}$$

which is the *probability flux.*

Look at the quantity in the square brackets above. $k_BT\log P$ is the chemical potential for an ideal gas, if we interpret P as the particle density. $\boldsymbol{F} - \nabla k_BT\log P$ is the total 'thermodynamic' force on the ideal gas particle in this interpretation. This produces the velocity

$$\boldsymbol{u} = (\boldsymbol{F} - \nabla k_BT\log P)/\zeta \tag{3.5.27}$$

under the overdamping condition. Thus, $\boldsymbol{J} = P\boldsymbol{u}$ is the probability flux. The conservation law of probability (3.5.25) is the Smoluchowski equation. That is, if we interpret the ensemble as an ideal gas of sample points, we can introduce a thermodynamic driving force that is the gradient of the chemical potential for this ideal gas.

Exercise 1. A similar interpretation is possible for Fokker-Planck equations.□

3.5.8 Detailed balance and equilibrium condition

In equilibrium we know the system must satisfy the detailed balance condition (→**2.7.6**).

For the Smoluchowski equation the state of the system described by it is specified by $\boldsymbol{r}$. Any state (that is, $\boldsymbol{r}$) is trivially invariant under time reversal, so the detailed balance condition must hold between any pair of states in equilibrium. If we write the probability flux as $\boldsymbol{J} = P\boldsymbol{u}$ (→**3.5.7**), $\boldsymbol{u}$ describes the imbalance of transitions between infinitesimally different adjacent states (i.e., between $\boldsymbol{r}$ and $\boldsymbol{r} + d\boldsymbol{r}$). Therefore, $\boldsymbol{u} = 0$ or $\boldsymbol{J} = 0$ everywhere is a necessary and sufficient condition for an equilibrium density distribution P. See also **X1.11.3**.

Exercise 1. What happens, if $\boldsymbol{F}$ is not conservative (that is, does not have a potential)?[74] □

3.5.9 *H*-theorem for Smoluchowski equation

We are discussing irreversible processes, so we should have a guarantee that the solution of the Smoluchowski equation for any time-independent potential force field can reach its equilibrium distribution (that certainly exists →**3.5.6**).

First of all, we must obtain the equilibrium distribution. Let us assume that the force $\boldsymbol{F}$ is a potential force $\boldsymbol{F} = -\nabla V$. Then, the probability flux (3.5.26) for the Smoluchowski equation can be written as

$$\boldsymbol{J} = -\frac{k_BT}{\zeta}e^{-\beta V}\nabla(Pe^{\beta V}). \tag{3.5.28}$$

Therefore, in the equilibrium state $J = 0$ and $P_{eq} \propto e^{-\beta V}$ as we expect. Can all the initial conditions eventually reach this distribution? Remember that this was the question Boltzmann asked as to the Maxwell distribution (→**2.4.2**).

We know the Kullback-Leibler entropy (→**1.6.9**)

$$K(P, P_{eq}) = \int dx P\log(P/P_{eq}) \geq 0, \tag{3.5.29}$$

---

[74] Far away from equilibrium, detailed balance does not generally hold, and an interesting thing could happen (→**6.7.5**).

where the integral is all over the domain under consideration. K vanishes if and only if $P = P_{eq}$. Therefore, if we could show that its time derivative is negative definite for $P \neq P_{eq}$, we are done.[75]

Let us compute dK/dt:

$$\frac{dK}{dt} = -\frac{k_BT}{\zeta}\int \frac{1}{P}[\beta P\nabla V + \nabla P]^2. \tag{3.5.30}$$

This is negative except when [ ] vanishes,[76] but the latter condition is the equilibrium condition. Such a theorem guaranteeing the relaxation to the unique equilibrium is generally called an *H-theorem* after the celebrated original H-theorem by Boltzmann for the Boltzmann equation (→**2.4.10**).

Discussion 1. What happens if $\boldsymbol{F}$ does not have any potential? □

Exercise 1. Find the corresponding statement for the Fokker-Planck equation and demonstrate it.[77] □

3.5.10 Passivity of Smoluchowski equation

Consider a system described by the Smoluchowski equation (3.5.23). Here, we interpret $\boldsymbol{r}$ to be generalized coordinates (we imagine that the system has many parts each of which may have its own coordinates) on some orientable manifold M. The net power W done to the system by the force $\boldsymbol{F}$ may be written as

$$W = \int_M \boldsymbol{F}\cdot\boldsymbol{u}Pd\boldsymbol{r}, \tag{3.5.31}$$

where $\boldsymbol{u}$ is the velocity (generalized velocity) that has appeared in **3.5.7**: $P\boldsymbol{u} = \boldsymbol{J}$. In the steady state $div\,\boldsymbol{J} = 0$. Since $\boldsymbol{F} = (\zeta\boldsymbol{J} + k_BT\nabla P)/P$, we can compute

$$W = \zeta\int_M \frac{\boldsymbol{J}\cdot\boldsymbol{J}}{P}d\boldsymbol{r} + k_BT\int_M div(\boldsymbol{J}\log P)d\boldsymbol{r}. \tag{3.5.32}$$

The last term is zero, so $W \geq 0$. That is, we cannot obtain any work from the steady state.[78]

The above argument was used to demonstrate that Maxwell's demon could not violate the second law of thermodynamics.[79] However, what we have to show is for any time-dependent choice of $\boldsymbol{F}$, we cannot gain any free lunch. Thus, the above calculation is not the demonstration of the assertion.

3.5.11 No demon can gain net work in the long run

Let us change the force by controlling it. Now, the system is governed by the Smoluchowski equation with $\boldsymbol{F}(t)$, and P is of course not a steady state. Still, we should not be able to

---

[75] This is the so-called *Lyapunov's second method* to prove the stability of P_{eq}. Although it is said here "we are done," needless to say, we must show that the solution does exist. In contrast to the Boltzmann equation (→**2.4.13**), this is not a problem in this case (→**3.5.6**).

[76] The reader might wonder whether P is everywhere nonzero. This is actually guaranteed for any positive t even if the initial P is delta-function like, thanks to the parabolic nature of the equation.

[77] In the case of the Smoluchowski equation, the H-theorem is directly related to the work, but this case it does not seem so.

[78] For the Fokker-Planck equation, see **X1.6.8**.

[79] M. O. Magnasco, "Szilard's heat engine," Europhys. Lett. **33**, 583-588 (1996).

gain work from the system. Let $W(t)$ be the net power at time t. What we should assert is that the long time average of $W(t)$ cannot be negative. That is, even if the demon has a high intelligence and control the force (or the system) in a clever fashion, no net work can be gained in the long run. Let us demonstrate this.

$$\int_0^t W(t)dt = \int_0^t dt \int \boldsymbol{F}(t) \cdot \boldsymbol{J} d\boldsymbol{r}, \tag{3.5.33}$$

$$= \int_0^t dt \left[\zeta \int \frac{\boldsymbol{J} \cdot \boldsymbol{J}}{P} d\boldsymbol{r} + k_B T \int \boldsymbol{J} \cdot \nabla \log P d\boldsymbol{r} \right] \tag{3.5.34}$$

$$= \int_0^t dt \left[\zeta \int \frac{\boldsymbol{J} \cdot \boldsymbol{J}}{P} d\boldsymbol{r} + k_B T \int div(\boldsymbol{J} \log P) d\boldsymbol{r} - k_B T \int \log P div \boldsymbol{J} d\boldsymbol{r} \right] \tag{3.5.35}$$

$$= \int_0^t dt \left[\zeta \int \frac{\boldsymbol{J} \cdot \boldsymbol{J}}{P} d\boldsymbol{r} + k_B T \int \log P \frac{\partial P}{\partial t} d\boldsymbol{r} \right]. \tag{3.5.36}$$

Now, integrating by parts with respect to time, we know that the last term can be rewritten as the difference of entropy $S = -k_B \int P \log P d\boldsymbol{r}$:

$$k_B T \int P(t) \log P(t) d\boldsymbol{r} - k_B T \int P(0) \log P(0) d\boldsymbol{r}, \tag{3.5.37}$$

so the time average of the power reads

$$\frac{1}{t} \int_0^t \zeta \int \frac{\boldsymbol{J} \cdot \boldsymbol{J}}{P} d\boldsymbol{r} - \frac{T}{t} \left[S(t) - S(0) \right]. \tag{3.5.38}$$

Therefore, if the entropy is bounded, then no work can be obtained in the long run. See also **X1.6.7**.

3.5.12 Where does irreversibility come from?

Consider a particle going around the unit circle with a slightly fluctuating angular velocity. Its (x, y) coordinate may be expressed in terms of a complex number $z = x + iy$, and the evolution equation is

$$\frac{dz}{dt} = i(1 + \omega(t))z, \tag{3.5.39}$$

where ω is a stationary Gaussian white noise with mean zero and $\langle \omega(t)\omega(0) \rangle = 2\zeta\delta(t)$. Then, as already discussed in **3.4.8**,

$$\langle z(t) \rangle = z(0)e^{it - \zeta t}. \tag{3.5.40}$$

Therefore, the average eventually vanishes, never to recover any finite value. This is nothing but clear irreversibility.

Now, look at individual oscillators. There is no damping of its oscillation at all; we only see dephasing. The reader may have thought about the spin echo experiment.

Exercise 1. If we have only a finite number of such oscillators, we do not have an irreversible behavior. What is the expected 'recovery time' as a function of the number of oscillators? □

3.6 Applications of Smoluchowski Equation

3.6.1 Sedimentation or centrifugation equilibrium

Consider a column of liquid along the z-axis in which small particles (macromolecules or colloidal particles) are suspended. We assume that the bottom of the column is closed and is at the origin. Applying a uniform gravitational or centrifugal force f in the negative z-direction, we wish to study the distribution of these particles along the z-axis.

We may assume that the viscosity is sufficiently high, so we may use the Smoluchowski equation (→**3.5.5**):

$$\frac{\partial P}{\partial t} = \frac{mg}{\zeta}\frac{\partial P}{\partial z} + \frac{k_BT}{\zeta}\frac{\partial^2 P}{\partial z^2}, \tag{3.6.1}$$

where m is an effective mass with the buoyancy effect taken into account.

Notice that mg/ζ is the sedimentation speed — if there were no bottom of the column (but if some how the fluid column is kept stationary), the particle reaches this average speed (in the negative direction of z) eventually.

The particle flux is proportional to (→(3.5.27))

$$J = \frac{1}{\zeta}\left(-mg - k_BT\frac{\partial}{\partial z}\log P\right)P. \tag{3.6.2}$$

In equilibrium we may assume $J = 0$ (why? →**3.5.8**), so we obtain a familiar Boltzmann result: $P(z) = P(0)e^{-\beta mgz}$ (immediate from the form of $\boldsymbol{J}$ given in **3.5.9**).

To solve (3.6.1) we need an initial and a boundary condition. The initial condition is the initial distribution. For a column of liquid with a solid bottom the upper boundary condition is that P should be zero at infinity,[80] and at the bottom, no particle can go though the bottom, so the flux must vanish there for $t > 0$:

$$-mgP(0) - k_BT\left.\frac{\partial P}{\partial z}\right|_{z=0} = 0. \tag{3.6.3}$$

The rest is technicality.

Exercise 1. *Sedimentation coefficient* s is defined by $v = s\alpha$ where v is the speed of the sedimenting particle, and α is the acceleration $\alpha = g$ in sedimentation or in the case of ultracentrifugation $\alpha = r\omega^2$, where ω is the rotor angular velocity and r is the distance between the particle and the center of the rotor axis. For proteins s is measured in the unit of 10^{-13}s and is called *Svedberg* (S). For example, eukaryote ribosome has 80S (4.2 $\times 10^6$ D).
(1) Derive

$$s = D\frac{m(1-\overline{v}\rho)}{k_BT}, \tag{3.6.4}$$

where D is the diffusion constant, m is the molecular mass ($=M/N_A$), $\overline{v}$ is the specific volume (1/density)[81] of the particle, and ρ is the density of the fluid medium.
(2) An eukaryotic ribosome consists of two particles with 30S and 50S. Why is its sedimentation coefficient only 70S? □

---

[80] Actually, assuming the boundedness at infinity is enough. For a general discussion of auxiliary conditions, see **X1.8.1** and on.

[81] For proteins this is about 0.7 cm^3/g.

3.6.2 Escape through potential barrier — Kramers theory[82]

Suppose there is a constant probability flux across a potential barrier along the x-axis (= reaction coordinate). See the figure.

Integrating (3.5.28) $\times e^{\beta V}$ from x_A to x_B, we get

$$J\int_{x_A}^{x_B} e^{\beta V} dx = \frac{k_BT}{\zeta}\left[Pe^{\beta V}\right]_{x_B}^{x_A}. \tag{3.6.5}$$

Therefore,

$$J = \frac{k_BT}{\zeta}\frac{P(x_A)e^{\beta V(x_A)} - P(x_B)e^{\beta V(x_B)}}{\int_{x_A}^{x_B} e^{\beta V} dx}. \tag{3.6.6}$$

If we prepare the particle in state x_A and remove all the particles from x_B, then

$$J = \frac{k_BT}{\zeta}P(x_A)e^{\beta V(x_A)}\Big/\int_{x_A}^{x_B} e^{\beta V} dx. \tag{3.6.7}$$

We wish to make this formula easier to use when the barrier height $E^{\neq}$ is significant. If n_A is the total probability of finding the particle in the potential well around x_A, then $P_{esc} = J/n_A$ is the escape probability per unit time. This n_A may be computed as

$$n_A = \int P(x_A)e^{-\beta V} dx, \tag{3.6.8}$$

where the integration is over the potential well around x_A. If we approximate the potential well with a harmonic potential $m\omega_A^2 x^2/2$ (we may identify $m\omega_A^2$ with the potential curvature around x_A: $m\omega_A^2 = V''(x_A)$), then

$$n_A = \frac{P(x_A)}{\omega_A}\sqrt{\frac{2\pi k_BT}{m}}. \tag{3.6.9}$$

Therefore,

$$P_{esc} = \frac{k_BT}{\zeta}\omega_A\sqrt{\frac{m}{2\pi k_BT}}\Big/\int_{x_A}^{x_B} e^{\beta V} dx. \tag{3.6.10}$$

Now, the denominator must be dominated by the contribution from the potential barrier whose height we assume to be $E^{\neq}$ (see the figure). Then, near this peak we may approximate $V = E^{\neq} - m\omega_E^2(x - x^{\neq})^2/2$

---

[82] Cf. M Dresden, *Kramers's contributions to statistical mechanics*, Physics Today, September 1988 p26-33.

($V(x_A) = 0$ is assumed for convenience), where $x^{\neq}$ is the location of the barrier peak ($m\omega_E^2 = -V''(x^{\neq})$). Then,

$$\int_{x_A}^{x_B} e^{\beta V} dx \simeq e^{\beta E^{\neq}} \int dy\, e^{\beta m\omega_E y^2/2} = e^{\beta E^{\neq}} \frac{1}{\omega_E}\sqrt{\frac{2\pi k_B T}{m}}. \tag{3.6.11}$$

Combining this result and (3.6.10), we finally obtain *Kramers' escape probability formula*:[83]

$$P_{esc} = \frac{\omega_A \omega_E}{2\pi(\zeta/m)} e^{-\beta E^{\neq}} = \frac{\sqrt{V''(x_A)|V''(x^{\neq})|}}{2\pi\zeta} e^{-\beta E^{\neq}} \tag{3.6.12}$$

By the way, if the friction is small, the particle crosses the barrier and reach the lower potential well without losing energy very much, so the particle may be bounced back to the original metastable state. This implies that there is a right friction that can maximize the rate of successful transition. Let ω be the frequency in the metastable well ($1/\omega$ is the number of trials per unit time to cross the barrier). The time constant is m/ζ, where m is the mass of the particle, so dimensional analytically $\zeta = cm\omega$, where c is a numerical constant, is the optimal friction.[84]

3.6.3 Brownian particle in periodic potential

Let us consider an overdamped Brownian particle in a periodic potential V such that $V(r + L) = V(r)$. The Langevin equation reads as usual with an equilibrium noise ν satisfying the fluctuation-dissipation relation:

$$\frac{dr}{dt} = -\frac{1}{\zeta}\frac{dV}{dr} + \nu. \tag{3.6.13}$$

The corresponding Smoluchowski equation is

$$\frac{\partial P}{\partial t} = -\frac{1}{\zeta}\frac{\partial}{\partial r}\left(-\frac{dV}{dr}P - k_B T\frac{\partial P}{\partial r}\right). \tag{3.6.14}$$

Since the space is infinite, no normalizable steady (= time-independent) state can exist (steady solutions are spatially extended). To study a steady state, we have only to solve (3.6.14) on one period with a periodic boundary condition: P and J are continuous ($P(0) = P(L)$, $J(0) = J(L)$). The continuity of J may be replaced with the continuity of $\nabla\left(Pe^{\beta V}\right)$ thanks to (3.5.28). Notice that in the steady state J must be constant. This constant must be zero for a periodic system, so no particle net flow is possible.

The *particle current* is defined as $\langle \dot{r} \rangle$. The simplest way to obtain this is to average (3.6.13)

$$\langle \dot{r} \rangle = -\frac{1}{\zeta}\langle V'(r) \rangle = -\frac{1}{\zeta}\int_0^L dr\, V'(r)P(r,t). \tag{3.6.15}$$

Since $J = uP$ and u is the particle velocity, we can also write

$$\langle \dot{r} \rangle = \int_0^L dr\, J(r,t). \tag{3.6.16}$$

---

[83] There is another derivation based on the waiting time for escape. See **X1.9.6**.

[84] H. Frauenfelder and P. G. Wolynes, *Rate theories and puzzles of hemoprotein kinetics*, Science **229**, 337 (1985) around page 341 may be readable. Also see, G. R. Fleming and P. G. Wolynes, *Chemical dynamics in solution*, Physics Today, May 1990, p36-43 (esp., Figure 5).

That is, generally, we can write

$$\langle \dot{r} \rangle = LJ^{st}, \tag{3.6.17}$$

where J^{st} is the steady state flux.

If the system is periodic, then the distribution is an equilibrium distribution in each periodic portion of the space, so there cannot be any flux. That is. the particle current must be zero.

3.6.4 Brownian particle in tilted periodic potential

Let us tilt the ratchet.

$$\frac{dx}{dt} = -\frac{1}{\zeta}V'(x(t)) + F + \nu(t), \tag{3.6.18}$$

The steady state[85] can be obtained as

$$\begin{aligned} \hat{P}^{st}(x) &= \mathcal{N}\frac{\zeta}{k_BT}e^{-\beta V(x)+\beta Fx}\int_x^{x+L} dy\, e^{\beta V(y)-\beta Fy}, && (3.6.19)\\ \langle \dot{x} \rangle &= L\mathcal{N}\left[1 - e^{-\beta LF}\right], && (3.6.20) \end{aligned}$$

where $\mathcal{N}$ is the normalization constant for $\hat{P}_{st}$.
i) It can be proved that the current is monotonic in F.
ii) The current is maximum for a fixed F when $V =$ const.
iii) Linear response is symmetric wrt F.

If F is not very small, the system is far away from equilibrium, and the Einstein relation

$$D_{eff} = k_BT\frac{d\langle \dot{x} \rangle}{dF} \tag{3.6.21}$$

does not hold.

If the thermal energy is sufficiently smaller than the barrier height, we can apply the saddle point method to evaluate (3.6.19), etc. However, the reader should have realized that that is nothing but an application of the Kramers theory of barrier crossing (→**3.6.2**). We obtain

$$\begin{aligned} \langle \dot{x} \rangle &= L(k_+ - k_-), && (3.6.22)\\ k_+ &= \frac{|V''_{eff}(x_M)V''_{eff}(x_m)|^{1/2}}{2\pi\zeta}e^{-\beta\Delta V_{eff}}, && (3.6.23)\\ k_- &= k_+e^{-\beta FL}. && (3.6.24) \end{aligned}$$

3.6.5 Brownian ratchet

We have learned that no particle net motion is possible on a periodic structure, if the system is in a steady state (→**3.6.3**). Actually, there is no steady state other than the equilibrium state, so to have a steady particle current, we need a condition that prevent the system from reaching an equilibrium state. A system with spatial periodicity (on the average) that

---

[85] This must be maintained by the source and the sink of the particles at infinity.

exhibits a nonzero particle current is called a *Brownian ratchet*.[86]

To realize a Brownian ratchet we must
1) break thermal equilibrium (detailed balance),
2) break the spatial inversion symmetry.
Let us write down a general Langevin equation (overdamped):[87]

$$\frac{dx}{dt} = -\frac{\partial}{\partial x}W(x,t) + F(x,t), \tag{3.6.25}$$

Here, W is the potential that can change in time (maintaining the periodicity on the average), and F is the external force including the noise. A Brownian ratchet may be realized by
1) Changing force: W is fixed but without inversion symmetry. F is a noise violating the fluctuation-dissipation relation, or a non-thermal chaotic force.[88]
2) Changing potential: F is the usual noise satisfying the fluctuation-dissipation relation, but W changes.

There are two typical cases:
Pulsating ratchet W oscillates between two periodic shapes W_1 and W_2. The transition may be periodic or stochastic.
Tilting ratchet W oscillates between two shapes $W_\pm = W_0 \pm Fr$ tilted in the opposite directions (on the average, they are periodic), where W_0 lacks inversion symmetry. This case may be an example of 1).

Curie's principle tells us that what is not forbidden by symmetry principles will occur. Thus, nonequilibrium spatially noninvertible system should exhibit nonzero particle current in general.

3.6.6 How does pulsating ratchet work?

The following figure tells us why a periodically pulsating ratchet can produce a non-zero particle current. Notice that the direction of the current is given by the sign of the steeper slope. This is true for saw-tooth potential; generally the relation between the potential shape and the sign of the current is highly delicate.

---

[86] F. Jülicher, A. Ajdari and J. Prost, "Modeling molecular motors," Rev. Mod. Phys. 69, 1269 (1997); P. Reimann, "Brownian motors: noisy transport far from equilibrium," Phys. Rep. 361, 57-265 (2002). The former is an introductory review by the group initiated the Brownian ratchet study and the latter is an encyclopedic review including many realization experiments.

[87] Usually, we do not discuss the inertial effect. Although the Brownian ratchet has no direct relevance to molecular motors, these motors are overdamped; also many transport experiments discuss overdamped situations.

[88] A beautiful example is due to Hondou: T. Hondou, and S. Sawada, "Dynamical behavior of a dissipative particle in a periodic potential subjected to chaotic noise: Retrieval of chaotic determinism with broken parity," Phys. Rev. Lett. 75, 3269 (1995).

The potential oscillates periodically between W_1 (flat) and W_2 (asymmetric saw-tooth). When the potential is flat, the particle diffuses from the sharper spatial distribution to a broader one above. The right tail of the distribution will fall into the next potential minimum.

Even if the transition between the potentials is stochastic, qualitatively there is no difference.

3.6.7 How does tilting ratchet work?

We know that if there is no tilt, there is no net flux. That is, in (3.6.13) if $F = -dV/dr$ is zero on the spatial average, we cannot have any systematic flux. In the tilting ratchet, we periodically change the tilt as a function of time (but its time average is still zero).[89] This ratchet was inspired by the ratchet model that changes V periodically in time due to Ajdari and Prost:[90]

The following figure tells us why a periodically tilting ratchet can produce a non-zero particle current. Notice that the opposite direction of the current is given by the sign of the steeper slope. This is true for saw-tooth potential; generally the relation between the potential shape and the sign of the current is highly delicate.

Red: $+Fx$ tilt. The distribution stays sharply localized in the potential minimum.
Green: $-Fx$ tilt. In this case free diffusion is possible, and a significant portion of the distribution could be in the next potential minimum, almost never to return.

---

[89] M. O. Magnasco, “Forced thermal ratchet,” Phys. Rev. Lett. **71**, 1477 (1993).

[90] A. Ajdari and J. Prost, CR Acad Sci Paris **315**, Ser II 1635 (1992).

3.7 Analyzing Fluctuations and Noises

In this section harmonic analysis of noises is introduced. The relation between the correlation function and the power spectrum is experimentally vital. As a related topic dynamical scattering factor $S(\omega, \boldsymbol{k})$ is also briefly discussed.

3.7.1 Importance of time correlation functions

We have already encountered time correlation functions of various observables. We have learned that the friction constant is determined by the noise correlation function (→**3.2.5**). The system response is governed by the time correlation function (→**3.2.8**). Transport coefficients and kinetic coefficients are given in terms of time correlation functions (the Green-Kubo formulas →**3.2.9**). Fourier analysis is a very useful method to study correlation functions.

3.7.2 Time correlation is maximum at $t = 0$

Let $C(t) = \langle q(t)q(0)\rangle$ for some observable q. Then, for a stationary system

$$|C(t)| \leq |C(0)|. \tag{3.7.1}$$

Exercise 1. Demonstrate this, using $\langle [q(t) \pm q(0)]^2 \rangle \geq 0$ and stationarity. □.

3.7.3 Fourier transform of noise

When we introduced the Langevin equation (→**3.2.4**), it was mentioned that the noise is 'white.' Such a statement is best understood by analyzing the frequency spectrum of the noise. To this end we must Fourier transform the noise.

Let us define the (time-domain) *Fourier transform* of a function $f(t)$ by[91]

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} dt\, f(t) e^{-i\omega t}. \tag{3.7.2}$$

Its inverse is

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \hat{f}(\omega) e^{i\omega t}. \tag{3.7.3}$$

$|\hat{f}(\omega)|^2$ is related to the *power spectrum* of $f(t)$. There is a device called the *spectrum analyzer* that gives us the power spectrum of the signal.[92]

Let us compute the power spectrum of the noise appearing explicitly in the Langevin equation (here, ω is NOT the stochastic parameter; we suppress it). Let w be, say, the x-component of $\boldsymbol{w}$ in **3.2.4**.

$$\hat{w}(\omega) = \int_{-\infty}^{\infty} dt\, w(t) e^{-i\omega t}. \tag{3.7.4}$$

---

[91] There are many different conventions, but the choice in these notes is the standard one in statistical mechanics.

[92] $|\hat{f}(\omega)|$ is called the *voltage spectrum*, and is also often considered, but as the reader will learn soon, the power spectrum is more fundamental.

Then,

$$\begin{aligned}
\langle\hat{w}(\omega)\hat{w}(\omega')\rangle &= \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} ds\,\langle w(t)w(s)\rangle e^{-i\omega t - i\omega' s}, &(3.7.5)\\
&= 2\zeta k_B T \int_{-\infty}^{\infty} dt\, e^{-i(\omega+\omega')t} &(3.7.6)\\
&= 4\pi\zeta k_B T\delta(\omega+\omega'), &(3.7.7)\\
&\equiv 2\pi\sigma(\omega)\delta(\omega+\omega'). &(3.7.8)
\end{aligned}$$

Here, (the multiplicative factor for $2\pi\delta(\omega+\omega')$)

$$\sigma(\omega) = 2\zeta k_B T \tag{3.7.9}$$

is called the *spectral density* or *power spectrum* of w.[93]

The power spectrum $\sigma(\omega)$ of our noise appearing in the Langevin equation is constant (independent of ω). Thus, the Langevin noise is deservingly called a *white noise*. However, this is an idealization. We have already discussed that there is a time scale τ that corresponds to the 'cutoff' frequency (→**3.4.1**); for shorter time scales the noise cannot change appreciably, so $\sigma(\omega)$ for $\omega > 1/\tau$ must decay to zero (→**3.7.6**).

3.7.4 Relation between power spectrum and correlation function: an illustration

As an example of 'harmonic analysis of stochastic systems' let us study the power spectrum of a free Brownian particle discussed in **3.2.4**.

Let $C(t) = \langle v(t)v(0)\rangle$ be the correlation function of the x-component of the velocity $\boldsymbol{v}$ of a Brownian particle governed by (3.5.8). Then, (again ω is NOT a stochastic parameter but the angular frequency)

$$v(\omega) = \frac{w(\omega)}{i\omega m + \zeta}. \tag{3.7.10}$$

Therefore,

$$\langle v(\omega)v(\omega')\rangle = \frac{2\pi\sigma_w(\omega)}{\omega^2 m^2 + \zeta^2}\delta(\omega+\omega'), \tag{3.7.11}$$

where $\sigma_w(\omega)$ is the spectral density of the (x-component of the) noise w, given by $\sigma_w(\omega) = 2\zeta k_B T$ (cf. (3.7.8)). Therefore, the power spectrum of the velocity $\sigma_v(\omega)$ reads

$$\sigma_v(\omega) = \frac{2\zeta k_B T}{\zeta^2 + m^2\omega^2}. \tag{3.7.12}$$

Let us Fourier-transform this back to the time domain: with the aid of the residue theorem we get

$$\frac{1}{2\pi}\int d\omega\,\sigma_v(\omega)e^{i\omega t} = \frac{k_B T}{m}e^{-\zeta|t|/m}. \tag{3.7.13}$$

Compare this with the result in **3.2.7**. We have

$$\langle v(t)v(0)\rangle = \frac{1}{2\pi}\int d\omega\,\sigma_v(\omega)e^{i\omega t}, \tag{3.7.14}$$

or

$$\sigma_v(\omega) = \int dt \langle v(t)v(0)\rangle e^{-i\omega t}. \tag{3.7.15}$$

That is, the Fourier transform of the power spectrum is the time correlation function, and vice versa. This is actually a general theorem. See **3.7.5**.

---

[93] Again, other definitions choose different numerical factors.

3.7.5 Wiener-Khinchine theorem

Let $q(t)$ be a stochastic process with integrable stationary time correlation function $C(t) = \langle q(t)q(0)\rangle$. Let us compute its power spectrum $\sigma(\omega)$:

$$\begin{aligned} \langle \hat{q}(\omega)\hat{q}(\omega')\rangle &= \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} ds\, C(t-s)e^{-i\omega t - i\omega' s}, & (3.7.16)\\ &= \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\tau\, C(\tau)e^{-i(\omega+\omega')t - i\omega'\tau}, & (3.7.17)\\ &= 2\pi\delta(\omega+\omega')\int_{-\infty}^{\infty} dt\, C(t)e^{-i\omega t}. & (3.7.18) \end{aligned}$$

That is (→**3.7.3**),

$$\sigma(\omega) = \int_{-\infty}^{\infty} dt\, C(t)e^{-i\omega t}, \tag{3.7.19}$$

or

$$C(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt\, \sigma(\omega)e^{i\omega t}. \tag{3.7.20}$$

This is called *Wiener-Khinchine's theorem.*

This is a practically very important theorem, because the equality (3.7.20) is the most convenient method to compute the correlation function: the power spectrum is easy to compute numerically from the original data thanks to the fast Fourier transformation (FFT).

Discussion 1 If you have never heard of the FFT, study the Cooly-Tukey algorithm.[94] □

Exercise 1. Show that the time correlation function is positive definite in the sense appearing in Bochner's theorem (→**1.4.8**). This implies that if the correlation function is continuous at the origin and if normalized as $C(0) = 1$, it is a characteristic function of a certain probability distribution. We know that this 'certain function' is the corresponding power spectrum. Therefore, the power spectrum may be interpreted as a probability distribution function (of what?). Then, the information minimization technique (→**1.3.5**) may be used to infer or model the time correlation function. Suppose the power spectrum is centered around the origin, and its second moment $\langle\omega^2\rangle$ is given. What kind of time correlation function do you get? □

Exercise 2. There is a small rotator whose equation of motion may be written as

$$I\frac{d^2\theta}{dt^2} = -\zeta\frac{d\theta}{dt} - k\theta + w, \tag{3.7.21}$$

where I, ζ and k are positive constants (the inertial moment, the viscous damping factor, and the spring constant, if you wish to interpret them). w is an appropriate equilibrium thermal noise. Assume that the damping factor is not large, and we can observe a damped oscillatory behavior.

(1) What is the average of θ^2 in equilibrium?

(2) Find the solution to the initial value problem for the equation of motion, assuming that there is no noise, and the initial angular velocity is zero.

(3) Find the autocorrelation function $\langle\theta(t)\theta(0)\rangle$.

(4) What is the amplitude of the white noise w?

(5) Find the power spectrum of θ.

(6) Are the reader's answer to (3) and (5) consistent? □

Exercise 3. The noise current being observed consists of uncorrelated pulse of intensity I_0

---

[94] As usual, Gauss used this method long before Cooly and Tukey.

and duration τ. It is known that λ pulses occur on the average per unit time. You may assume the signal is statistically stationary.
(1) Find the correlation function of the current.
(2) Find the power spectrum of the current.
(3) Confirm the Wiener-Khinchine theorem. □

3.7.6 Long correlation means sharp spectrum

As we have learned, the power spectrum is easy to observe, so it is advantageous to have some 'feeling' about the relation between correlation functions and power spectra. A summary statement is: "A signal that has a short time correlation has a broad band." This is intuitively obvious, because short time correlation means rapid changes that must have high frequency components.

There may be several mathematical statements that are related to the above assertion. Perhaps the most elementary is to look at the result in **3.7.4**, or

$$C(t) = e^{-\alpha|t|} \ \leftrightarrow \ \sigma(\omega) = \frac{2\alpha}{\omega^2+\alpha^2}. \tag{3.7.22}$$

That is, the half-width of the power spectrum is the reciprocal of the 'life-time' of the signal. (The power spectrum with the Cauchy distribution is often called the *Lorentzian spectrum*; for its natural meaning see **X2.6.8**)

The result can be understood intuitively with the aid of dimensional analysis $\tau\omega \sim 1$.

Exercise 1. The *Riemann-Lebesgue lemma* tells us that $\lim_{|\omega|\to\infty}\sigma(\omega) = 0$ for time-continuous signals. Learn about this lemma. This tells us that white noise (→**3.7.3**) we have discussed is NOT continuous.[95]□

Exercise 2. The Riemann-Lebesgue lemma also tells us that the decay rate of the power spectrum for large $|\omega|$ gives information about the smoothness of the signal.
(1) Demonstrate that if the signal is m-times continuously differentiable with respect to time, $\sigma(\omega) = o(\omega^{-2m})$ for large ω.
(2) What is the general behavior of the power spectrum of the velocity of a Brownian particle in the high frequency limit? Is (3.7.12) physically sensible? Discuss in terms of time scales used in **3.4.1**.
(3) What can you say about the signal whose power spectrum decays exponentially? □.

3.7.7 Thermal noise in circuits: Nyquist's theorem

The idea of the regression hypothesis that goes back to Einstein's Brownian motion theory (→**3.1.7**) implies that if a system dissipates energy due to a decay of some (macroscopic) flow $\boldsymbol{J}$, then the reverse process, that is, creation of $\boldsymbol{J}$ from thermal noise can occur by fluctuation, albeit the created flow is at most mesoscopic.

Thus, if we have an electric circuit with a resistance R,[96] the thermal noise in the resistor (the noise voltage $V(t)$ in the resistor) should create a current. This must be detected as noise; this was observed by Johnson, and is called the *Johnson noise.* Nyquist explained this noise along the line of the idea just discussed.[97]

---

[95] However, its integral called *Wiener process* (→**X1.4.4**) is continuous with probability one.

[96] One must have an element that destroys macroscopic effects, if one wishes to observe their spontaneous creation. This is quite general; think about the role of mutation in evolution, or the role of viscosity in Helmholtz's vortex theorem.

[97] Johnson and Nyquist were friends, Swedish immigrants to the US. See D. Abbott, B. R. Davis, N. J. Phillips, and K. Eshraghian, "Simple derivation of the thermal noise using window-limited Fourier transforms," IEEE Trans. Educ. **39**, 1 (1996).

Let us consider an RC circuit. The macroscopic equation (that may be obtained by averaging a more microscopic equation over noise) reads

$$\frac{dQ}{dt} = -\frac{1}{RC}Q, \tag{3.7.23}$$

where Q is the charge stored in the capacitor of capacitance C. What actually happens may be modeled in terms of the noise voltage as

$$\frac{dQ}{dt} = -\frac{1}{RC}Q + \frac{V(t)}{R}. \tag{3.7.24}$$

Due to thermal noise, Q does not settle down to zero, but fluctuates around 0. This fluctuation changes the energy stored in the capacitor $Q^2/2C$. From **2.8.1** (or from the equipartition of energy), we obtain

$$\langle Q^2 \rangle = k_B T C. \tag{3.7.25}$$

The fluctuation-dissipation relation of the second kind (→**3.2.5**) should determine the voltage noise. Let us compute the power spectrum σ_Q of Q in terms of that of V (→**3.7.4**):

$$\sigma_Q(\omega) = \sigma_V(\omega)/(R^2\omega^2 + 1/C^2). \tag{3.7.26}$$

We expect that the correlation function for Q is given by $\langle Q^2 \rangle_{eq} e^{-|t|/RC}$, so (3.7.22) gives its power spectrum immediately, and we conclude

$$\sigma_V(\omega) = k_B T R. \tag{3.7.27}$$

This is called *Nyquist's theorem.*

Exercise 1. What is the current fluctuation for the above circuit? □
Exercise 2. If a resistor R is short circuited, what can you imagine to happen? □
Exercise 3. A resistor of resistance R and the capacitor of capacitance C are connected in series.
(1) What is the voltage fluctuation between the two ends of the open circuit?
(2) Suppose the input is the voltage across the open ends, and the response the current through R. Find the response function.
(3) State the fluctuation-dissipation relation for the current. □
Exercise 4. There is a circuit consisting of an inductor L and a resistor R.
(1) Write down the Langevin equation governing the equilibrium thermal noise current. (You need not determine the noise.)
(2) What is the variation of the equilibrium current? (Think about the average energy stored in L; this is an equilibrium statistical mechanics question.)
(3) What is the average Joule heat loss at the resistor?
(4) Determine the white noise in your Langevin equation.
(5) We know the answer to (3) is positive. Does not this violate the second law? □

3.7.8 Nyquist's theorem: quantum correction

Nyquist's theorem (3.7.27) is not correct for very low temperatures or very high frequencies, because the thermodynamic fluctuation theory does not hold under such conditions. The most natural guess is that k_BT should be replaced by the average energy of a quantum harmonic oscillator

$$E_\omega(T) = \frac{1}{2}\hbar\omega \cot\left(\frac{\beta\hbar\omega}{2}\right). \tag{3.7.28}$$

Notice that this asymptotically agrees with $1/\beta$ in the $\beta\hbar\omega \to 0$ limit. Therefore, the most general Nyquist's theorem applicable to any temperature and frequency reads

$$\sigma_V(\omega) = RE_\omega(T). \tag{3.7.29}$$

Of course, the constancy of R is an idealization: we should replace it with T and ω dependent R.

We will see that the replacement of k_BT with $E_\omega(T)$ is widely applicable in linear response theory (cf **4.4.15**).

3.7.9 Space-time correlation functions and its Fourier transform

Even if the system is spatially uniform, the fluctuations of various quantities are generally not uniform in space. This implies that we should study not only the time-correlation function but also *space-time correlation functions.* We will encounter such correlation functions in the calculation of transport coefficients (→**4.1.14**).

As a concrete example, let $\rho(\boldsymbol{r}, t)$ be the density field.

$$C(\boldsymbol{r}, t) = \langle \rho(\boldsymbol{r}, t)\rho(\boldsymbol{0}, 0)\rangle \tag{3.7.30}$$

is the space-time correlation function of ρ, where $\langle\ \rangle$ implies the ensemble average.[98] Here, we have assumed that the system is in equilibrium and spatially uniform on the average.

$C(\boldsymbol{r}, 0)$ is an equilibrium density correlation function, so its Fourier transform is the *form factor* $S(\boldsymbol{k})$ that can be observed by a scattering experiment. In such an experiment transfer of momentum of particles to the scatterer (elastic scattering) gives us the information about $S(\boldsymbol{k})$.

The space-time Fourier transform of $C(\boldsymbol{r}, t)$ is called *dynamical form factor*:

$$S(\boldsymbol{k}, \omega) = \int d^3\boldsymbol{r} \int dt\, C(\boldsymbol{r}, t)e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)}. \tag{3.7.31}$$

The reader may expect that this can also be obtained by some sort of scattering experiment. This time, perhaps the transfer of energy in the scattering process (inelastic scattering) gives the time-dependent information (*van Hove theory*). This is exactly the case, and such scattering experiment is called a *dynamical scattering* experiment.

Exercise 1. Let $\rho(\boldsymbol{r}, t)$ be the number density field of a monatomic dense fluid.

$$S(\boldsymbol{r}, t) = \langle \rho(0, 0)\rho(\boldsymbol{r}, t)\rangle \tag{3.7.32}$$

is observable as the space-time Fourier inverse transform of the dynamical form factor. Explain (illustrate) the time dependence of $S(\boldsymbol{r}, t)$. □

3.7.10 Van Hove theory: dynamical scattering experiment[99]

Dynamical form factors can be directly observed experimentally. To show this, we must

---

[98] In practice, this is the space-time average over a sample.

[99] L van Hove, Phys. Rev. **95**, 249 (1954).

analyze the scattering process: a monochromatic projectile of energy $\hbar\omega$ and momentum $\hbar\boldsymbol{k}$ hits the target whose state is specified by its eigenstate with energy E_n. The state of the whole system is specified by $(E_n, \boldsymbol{k}, \omega)$. The outcome of this collision is assumed to be $(E_m, \boldsymbol{k}', \omega')$.

Let us assume that the density distribution function of the scatterers in the system be $\psi(\boldsymbol{r})$, and the interaction potential between the scatterer and the projectile be V: the interaction Hamiltonian h between the projectile and the system may be written as

$$h(\boldsymbol{r}) = \int d\boldsymbol{r}'\, V(\boldsymbol{r}-\boldsymbol{r}')\psi(\boldsymbol{r}'). \tag{3.7.33}$$

The state ket of state $(E_n, \boldsymbol{k}, \omega)$ may be written as $|E_n\rangle|\boldsymbol{k}, \omega\rangle$.

With the aid of the perturbation theory (the Fermi golden rule) the scattering transition rate may be written as

$$P = \frac{2\pi}{h}|\langle \boldsymbol{k}', \omega'|\langle E_m|h|E_n\rangle|\boldsymbol{k}, \omega\rangle|^2\delta(\hbar(\omega-\omega') + E_n - E_m). \tag{3.7.34}$$

If we introduce $\Delta\boldsymbol{k} = \boldsymbol{k} - \boldsymbol{k}'$ and $\Delta\omega = \omega - \omega'$, and write the above P as $P(\Delta\boldsymbol{k}, \Delta\omega)$, then

$$P(\Delta\boldsymbol{k}, \Delta\omega) \propto S(\Delta\boldsymbol{k}, \Delta\omega). \tag{3.7.35}$$

That is, studying the scattering rate with momentum and energy transfer of $\hbar\Delta\boldsymbol{k}$, and $\hbar\Delta\omega$ directly gives the dynamical form factor.

A demonstration is as follows. Since the plane wave is given by

$$\langle \boldsymbol{r}|\boldsymbol{k}, \omega\rangle = \left(\frac{1}{2\pi}\right)^{3/2} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{3.7.36}$$

$$\begin{aligned}
\langle \boldsymbol{k}', \omega'|\langle E_m|h|E_n\rangle|\boldsymbol{k}, \omega\rangle &= \frac{1}{(2\pi)^3}\int d\boldsymbol{r}\int d\boldsymbol{r}'\, e^{i\boldsymbol{k}\cdot\boldsymbol{r} - i\boldsymbol{k}'\cdot\boldsymbol{r}'}\delta(\boldsymbol{r}-\boldsymbol{r}')\langle E_m|h(\boldsymbol{r})|E_n\rangle && (3.7.37)\\
&= \frac{1}{(2\pi)^3}\int d\boldsymbol{r}\, e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}}\langle E_m|\int d\boldsymbol{x}\, V(\boldsymbol{r}-\boldsymbol{x})\psi(\boldsymbol{x})|E_n\rangle && (3.7.38)\\
&= \frac{1}{(2\pi)^3}\int d\boldsymbol{\rho}\, e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{\rho}}V(\boldsymbol{\rho})\int d\boldsymbol{x}\, \langle E_m|e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}\psi(\boldsymbol{x})|E_n\rangle && (3.7.39)
\end{aligned}$$

Let us define (here, $\Delta\boldsymbol{k} = \boldsymbol{k} - \boldsymbol{k}'$)

$$W(\Delta\boldsymbol{k}) = \left|\int d\boldsymbol{\rho}\, e^{-i\Delta\boldsymbol{k}\cdot\boldsymbol{\rho}}V(\rho)\right|^2. \tag{3.7.40}$$

Then, the transition rate reads

$$\begin{aligned}
P &= \frac{2\pi}{h}W(\Delta\boldsymbol{k})\frac{1}{(2\pi)^6}\int d\boldsymbol{x}\int d\boldsymbol{x}'\langle E_n|e^{-i\Delta\boldsymbol{k}\cdot\boldsymbol{x}}\psi(\boldsymbol{x})|E_m\rangle \\
&\qquad \times \langle E_m|e^{i\Delta\boldsymbol{k}\cdot\boldsymbol{x}'}\psi(\boldsymbol{x}')|E_n\rangle\delta(\hbar(\omega-\omega') + E_n - E_m). && (3.7.41)\\
&= \frac{2\pi}{h}W(\Delta\boldsymbol{k})\frac{1}{(2\pi)^6}\int d\boldsymbol{x}\int d\boldsymbol{x}'\, e^{-i\Delta\boldsymbol{k}\cdot\boldsymbol{x} + i\Delta\boldsymbol{k}\cdot\boldsymbol{x}'}\langle E_n|\psi(\boldsymbol{x})|E_m\rangle \\
&\qquad \times \langle E_m|\psi(\boldsymbol{x}')|E_n\rangle\delta(\hbar(\omega-\omega') + E_n - E_m). && (3.7.42)
\end{aligned}$$

Now, we rewrite the delta function as (here, $\Delta\omega = \omega - \omega'$)

$$\delta(\hbar\Delta\omega + (E_n - E_m)) = \frac{1}{h}\int dt\, e^{-i(\Delta\omega + (E_n - E_m)/\hbar)t} \tag{3.7.43}$$

With this we may rewrite (here, H is the system Hamiltonian)

$$\begin{aligned}&\langle E_n|\psi(\boldsymbol{x})|E_m\rangle\langle E_m|\psi(\boldsymbol{x}')|E_n\rangle\delta(\hbar(\omega-\omega') + E_n - E_m) \\ &\qquad = \frac{1}{h}\int dt\, e^{i\Delta\omega t}\langle E_n|\psi(\boldsymbol{x})|E_m\rangle\langle E_m|e^{iHt/\hbar}\psi(\boldsymbol{x}')e^{-iHt/\hbar}|E_n\rangle.\end{aligned} \tag{3.7.44}$$

Now, recall the Heisenberg picture $e^{iHt/\hbar}\psi(\boldsymbol{x})e^{-iHt/\hbar} = \psi(\boldsymbol{x}, t)$. We arrive at

$$P \propto W(\Delta\boldsymbol{k})\int d\boldsymbol{x}\int d\boldsymbol{x}'\, e^{i\Delta\boldsymbol{k}\cdot(\boldsymbol{x}'-\boldsymbol{x})}\int dt\, e^{-i\Delta\omega t}\langle E_n|\psi(\boldsymbol{x})|E_m\rangle\langle E_m|\psi(\boldsymbol{x}', t)|E_n\rangle \tag{3.7.45}$$

What we observe is the average of P over the equilibrium distribution of the system initial energy E_n

$$\frac{1}{Z}Tre^{-\beta E_n}\langle E_n|\psi(\boldsymbol{x})|E_m\rangle\langle E_m|\psi(\boldsymbol{x}', t)|E_n\rangle = \frac{1}{Z}Tre^{-\beta H}\psi(\boldsymbol{x})\psi(\boldsymbol{x}', t), \tag{3.7.46}$$

but this is the definition of the space-time correlation function

$$C(\boldsymbol{r}, t) = Tr\,\rho\,\psi(0, 0)\psi(\boldsymbol{r}, t), \tag{3.7.47}$$

so we have (for neutron scattering W is almost constant)

$$P \propto \int d\boldsymbol{r}\int dte^{i(\Delta\boldsymbol{k}\cdot\boldsymbol{r} - \Delta\omega t)}C(\boldsymbol{r}, t). \tag{3.7.48}$$

The calculation may look acrobatic, but it is a rather routine calculation in linear response theory (→**X2.4.2**).

3.7.11 Long-range correlation under nonequilibrium conditions[100]

A typical dynamics of the fluctuation of a conserved quantity obeys

$$\frac{d\rho(\boldsymbol{k}, t)}{dt} = -Dk^2q(\boldsymbol{k}, t) + i\boldsymbol{k}\cdot\boldsymbol{w}(\boldsymbol{k}, t). \tag{3.7.49}$$

Here, we have already space-Fourier transformed the Langevin equation. If we write

$$\langle\boldsymbol{w}(\boldsymbol{k}, t)\boldsymbol{w}(\boldsymbol{k}', s)^T\rangle = 2(2\pi)^3R\delta(\boldsymbol{k}+\boldsymbol{k}')\delta(t-s), \tag{3.7.50}$$

we may solve this just as we have done in **3.2.4**, and after the initial memory is lost, we have

$$\langle q(\boldsymbol{k}, t)q(\boldsymbol{k}', 0)\rangle = (2\pi)^3\frac{\boldsymbol{k}R\boldsymbol{k}'}{Dk^2}e^{-Dk^2t}\delta(\boldsymbol{k}+\boldsymbol{k}'). \tag{3.7.51}$$

The fluctuation-dissipation relation (the condition to guarantee the stationary state to be the equilibrium state) implies $\boldsymbol{k}R\boldsymbol{k}'/Dk^2 = S(\boldsymbol{k})$, the static form factor; if we coarse-grain the system spatially, $S(\boldsymbol{k}) \sim \chi$, where χ is the compressibility.[101] so there is no long-range

---

[100] P. L. Garrido, J. L. Lebowitz, C. Maes, and H. Spohn, "Long-range correlation for conserved dynamics," Phys. Rev. A **42**, 1954-1968 (1990).

[101] Just as Wiener-Khinchine (→**3.7.5**), $S(\boldsymbol{k})$ is the spatial Fourier transform of the spatial correlation function $\langle q(\boldsymbol{r})q(0)\rangle$, so $\langle q(0)^2\rangle = (1/2\pi)^3\int S(\boldsymbol{k})d^3\boldsymbol{k} = V\chi$, where χ is the compressibility, so $\langle q(\boldsymbol{k})q(-\boldsymbol{k})\rangle \propto \chi$.

correlation; the spatial correlation is $\sim \delta(\boldsymbol{r})$, very short-ranged.

However, if the system is not isotropic, or not in equilibrium, $\boldsymbol{k}R\boldsymbol{k}'/Dk^2$ gives a long range correlation $\sim 1/r^d$ (like an electrostatic potential due to a quadrupole).[102].

This result may be understood easily, if we accept the long time tail for conserved quantities (→**5.2.6**) (actually, this can be derived from (3.7.51)). Due to the long time tail, two points at a distance r apart are under the influence of a common disturbance that happened $t \sim r^2$ ago. This effect decays as $t^{-d/2}$, so the correlation between these two points should be of order $(r^2)^{-d/2} = r^{-d}$.

Exercise 1. Show that $\langle q(r,t)q(0,0)\rangle$ for fixed r behaves as $t^{-d/2}$, using (3.7.51). □

Warning. The generality of the above result is not known. There are different nonequilibrium systems. For example, experimentally, the system driven at the boundary (the system wall) is definitely different from the system considered above. This entry will hopefully be expanded in the near future.

---

[102] Remember the Poisson equation. It is easy to see the correspondence $k^2 \iff 1/r^{d-2}$, so the Fourier transform of $\boldsymbol{k}R\boldsymbol{k}'/k^2 \iff \nabla R \nabla (1/r^{d-2})$.

Chapter 4

Fluctuations and Phenomenology

Let us pursue the line initiated by Einstein (→**3.1.7**): we can understand phenomenological laws through the study of fluctuations. Onsager's reciprocity, linear nonequilibrium thermodynamics, and the Green-Kubo relations are typical outcomes of this line of thoughts. The reader will realize how crucial the study of fluctuations is (and large deviation theory is).

4.1 Dynamics of Fluctuations and Phenomenological Laws

4.1.1 Large fluctuations in equilibrium

Big fluctuations can eventually be observed if we are patient enough. Its probability is controlled by the thermodynamic theory of fluctuation (→**2.8.5** = large deviation theory for an equilibrium ensemble **1.6.4**).

Exercise 1. There is a macromolecular chain that may be regarded as a harmonic spring with the spring constant $k = 10$pN/nm. What is the probability of its reaching the displacement of 20nm? Can you tell how long, on the average, we have to wait to encounter such a fluctuation? [You may have to supply needed information to answer the questions.] □

If we are patient enough, extremely big fluctuations may be observed. If we sample all the members of the equilibrium ensemble with such extreme fluctuations (i.e., a conditional ensemble), then the resultant ensemble (= a conditional ensemble) is certainly a nonequilibrium ensemble. A natural generalization of Einstein's fundamental idea (→**3.1.7**) is that such an ensemble may be used to study nonequilibrium phenomena. We do not know yet how legitimate this idea is. We know, as we will see soon, that Onsager's articulation of Einstein's idea near equilibrium (regression hypothesis →**4.1.2**) seems to hold.

Discussion 1. F_1ATPase is a rotary molecular motor that rotates stepwisely $2\pi/3$ per one ATP molecule when it is immersed in a solution of ATP.[1] The molecule contains a central

---

[1] K. Kinoshita, Jr., K. Adachi and H. Ito ,"Rotation of F_1 -ATPase: How an ATP-driven molecular machine may work" Annu. Rev. Biophys. Biochem. Struct., **33** 245-268 (2004). The original paper demon-

shaft that rotates when ATP is consumed in the surrounding ring. A molecule of ATP goes into a catalytic pocket and is hydrolyzed. The released energy is used to rotate the shaft. In equilibrium in the ATP solution there are appropriate amounts of ADP molecules and phosphate ions as well. Now, from this equilibrium ensemble construct a collection of F_1-ATPases that just received ATP in their pockets. Is it an equilibrium ensemble? □

Notice that equilibrium dynamics is time reversible (due to detailed balance →**2.7.6**), so on the average such extreme fluctuations should be produced quickly by the time reversal of the decay process, once the building up process really starts.[2] This symmetric quick building up of large fluctuations can be illustrated nicely in an experiment on an electronic circuit.[3]

Discussion 2. If the building up process is so quick, then why is the second law of thermodynamics not violated often? [We have almost thought about this in Discussion 1]. □

4.1.2 Large fluctuations and regression principle

Einstein's theory of Brownian motion (→**3.1.3**) assumes, as a key idea, that the mesoscale dynamics driven by fluctuations decay following macroscopic laws (in his case Fick's law). This already contains the idea that spontaneously formed fluctuations and the 'same' fluctuations created through applying certain external perturbations are indistinguishable. More precisely speaking, a local spontaneous fluctuation described by the deviation of thermodynamic quantities $\{\delta x\}$ is indistinguishable mesoscopically from the same deviation created by an external means (say, by modifying conjugate thermodynamic fields).[4]

Then, the instantaneous time evolution (i.e., the differential equation governing the time evolution) of the spontaneous fluctuation when it decays should be, on the average, identical to the macroscopic equation describing the decay of the macroscopically prepared initial condition. This idea is called Onsager's *regression hypothesis* (1931[5]).[6]

We could state Onsager's regression hypothesis as follows:

> In order to specify the macroscopic state of a system uniquely we can find a set of macroscopic state variables $\{\alpha_i\}$ the decay behaviors of whose spontaneous fluctuations are on the average described by the macroscopic phenomenological laws governing the time derivatives of $\{\alpha_i\}$.

This should be a principle. Accordingly, Onsager's original hypothesis should be called *Onsager's regression principle.* A much more direct and transparent statement will be given in the large deviation framework later (→**4.1.15**).

---

strating the rotation unequivocally is an many-vba award-winning paper by R. Yasuda, H. Noji, and K. Kinosita, Jr, "F_1-ATPase is a highly efficient molecular motor that rotates with discrete 120° steps," Cell **93**, 1117 (1998).

[2] See **4.1.3**

[3] D. G. Luchinsky and P. V. E. McClintock, "Irreversibility of classical fluctuations studied in analogue electrical circuit," Nature **389**, 463-466 (1997). As demonstrated in this paper under the nonequilibrium driven condition, the formation process and the decay process of a large fluctuation is not symmetric in time, because detailed balance is violated.

[4] There may be microscopic differences, but such differences are irrelevant at the larger scale we observe the system.

[5] This year, Japan occupied Manchuria; nylon invented by Carothers; Empire State Bldg completed.

[6] L. Onsager, "Reciprocal relations in irreversible processes. II." Phys. Rev. **38**, 2265-2279 (1931). Section 3 explains this idea.

4.1.3 Probability of violation of the second law

We have seen that fluctuation can go to a considerable extent against the second law (→**4.1.1**). After all, fluctuation decreases (local) entropy and without such fluctuations Jarzynski's equality (→**2.7.20**, also **4.1.4**) cannot hold.

Let the classical system[7] under study be described by its canonical coordinates y and externally controllable parameters λ. The system history may be expressed by $\{y(t), \lambda(t)\}$. We assume that the system satisfies the following *microscopic reversibility*:

$$P(\{x,\lambda\}_0^t | x(0))/P(\{\overline{x},\overline{\lambda}\}_t^0 | x(t)) \equiv P(\rightarrow)/P(\leftarrow) = e^{-\beta Q(\{x,\lambda\}_0^t)}, \tag{4.1.1}$$

where $\{x,\lambda\}_0^t$ is the path = history with a given protocol $\lambda(t)$, and the overline indicates the time reversal operation.[8] Q is the energy absorbed by the system from the heat bath during the forward process.

If the initial and the final states are in equilibrium, the difference of the surprisal (→**1.3.2**) $-\log\rho$ is interpreted as the entropy change. Therefore, we may assume[9]

$$\Sigma = \log \rho(x(0))/\rho(x(t)) - \beta Q, \tag{4.1.2}$$

where Σ is the entropy production, $\rho(x(t))$ is the distribution of canonical coordinates at time t.[10] Notice that the probability of the entropy production to be Σ by the forward dynamics is given by

$$P_F(\Sigma) = \sum \rho(x(0)) P(\rightarrow), \tag{4.1.3}$$

where the sum is over all the transitions compatible with the entropy production Σ. An analogous formula can be written for the backward dynamics P_R

$$P_R(-\Sigma) = \sum \rho(x(t)) P(\leftarrow). \tag{4.1.4}$$

There is a one to one correspondence between the forward and reversed paths, so the sum above is over exactly the same paths (although reversed in time) in the sum for P_F. We wish to compute

$$\frac{P_F(+\Sigma)}{P_R(-\Sigma)} = \frac{\sum \rho(x(0))P(\rightarrow)}{\sum \rho(x(t))P(\leftarrow)}. \tag{4.1.5}$$

Combining (4.1.1) and (4.1.2), we obtain for each path

$$\frac{\rho(x(0))P(\rightarrow)}{\rho(x(t))P(\leftarrow)} = e^{\Sigma} \tag{4.1.6}$$

Now, take the following trivial identity:

$$\frac{a}{b} = \frac{c}{d} \Rightarrow \frac{a}{b} = \frac{c}{d} = \frac{a+c}{b+d}. \tag{4.1.7}$$

---

[7] not necessarily a deterministic system, but must be Markovian. Cf. **6.5.9**.

[8] The conditional probability implies that we do not take into account the (equilibrium or stationary) distribution of the states into account.

[9] The original paper: G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relations from free energy differences," Phys. Rev. E **60**, 2721 (1999) claims that the result is applicable far away from equilibrium, but it is questionable simply because we have no clear definition of entropy away from equilibrium.

[10] Precisely speaking, this ratio must be the reciprocal of the Radon-Nikodym derivative $d\rho(x(t))/d\rho(x(0))$.

Therefore, (4.1.5) implies

$$P_F(\Sigma)/P_R(-\Sigma) = e^{\Sigma}. \tag{4.1.8}$$

This is called the *fluctuation theorem.* This type of theorems hold for sufficinetly chaotic dynamical systems (→**7.6.3**).[11]

Exercise 1. The fluctuation theorem may be experimentally checked with a small system. Read Wang et al. Phys. Rev. Lett. **89**, 050601-1 (2002)[12] and summarize its content. □

Remark. There have been numerical and empirical attempts to check the fluctuation theorem: M. Dolowschi(a)k and Z. Kovács, "Fluctuation formula for nonreversible dynamics in the thermostated Lorentz gas," Phys. Rev. E **65** 066217-1-4 (2002). This is for a driven periodic Lorentz gas, and is consistent. □

4.1.4 Relation of fluctuation theorem to Jarzynski's equality

Let W_d be the dissipated part of the work done to the system. Then,

$$\Sigma = \beta W_d \tag{4.1.9}$$

Therefore, from (6.7.22) we have

$$\langle e^{-\beta W_d}\rangle_F = 1. \tag{4.1.10}$$

We know $W_d = W - \Delta F$, this is Jarzynski's equality (→**2.7.20**).

Discussion 1.[13] We can also derive

$$\langle e^{\Sigma}\rangle_R = 1. \tag{4.1.11}$$

This is, so to speak, an anti-Jarzynski equality. Can we use it? □

4.1.5 Decay dynamics and time reversal symmetry

The regression principle (→**4.1.2**) asserts that the 'macroscopic decay dynamics' is on the average identical to the decay dynamics of fluctuations. The dynamics of fluctuation may contain non-decaying components, so we must separate thermodynamic effects and purely mechanical (time reversible) effects.

Let a mesoscopic observable[14] X be zero on the average in equilibrium, but be displaced from zero due to fluctuation. Its time derivative may be described by a sum of the systematic part (with suffix s) and noise w as in the Langevin equation (→**3.2.2**):

$$\frac{dX}{dt} = \left(\frac{dX}{dt}\right)_s + w. \tag{4.1.12}$$

The systematic part may be defined by averaging the microscopic time derivative over the ensemble whose members are specified by a particular initial condition for macrovariables

---

[11] This type of relation was first found numerically by D. J. Evans, E. G. D. Cohen and G. P. Morriss, Phys. Rev. Lett. **71**, 2401 (1993), and later proved for Anosov dynamical systems by G. Gallavotti and E. G. D. Cohen, J. Stat. Mech. **80**, 931 (1995) (see **7.6.3**). The form given here is due to G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[12] by G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles and D. J. Evans, "Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales."

[13] from the question by Mr Y. Liu

[14] here, mesoscopic means the scale between microscopic and macroscopic and well-separated from both; recall that this is the condition that Langevin equations may be used.

and various realizations (samples) of the noise.[15]

The systematic part may be further classified into *reversible* and *irreversible* parts. Under the time reversal transformation (→**4.1.6** more systematically), the former should behave as X times t (or as X/t or as computed according to pure mechanics), and the latter behaves oppositely. For example, consider the velocity: $d\boldsymbol{r}/dt = \boldsymbol{v}$ is a mechanical result, and the both sides behave similarly, so there is no irreversible component. Or, look at the Smoluchowski equation (→**3.5.5**) or the corresponding Langevin equation. Its systematic part is $\boldsymbol{F}/\zeta$ and is invariant under time reversal (behaves differently from $d\boldsymbol{r}/dt$), so this term describes an irreversible effect as we know.

4.1.6 Time reversal symmetry operator

The *time reversal symmetry* operation $\mathcal{I}$ is formally defined by $\mathcal{I}(q,p,t) = (q,-p,-t))$. Thus, if everything is purely macroscopically mechanical, we should not detect the effect of this operation.[16] The reversible and irreversible parts of a microscopically expressible observable X are defined formally as

$$\left(\frac{dX}{dt}\right)_r = \frac{1}{2}\left[\left(\frac{dX}{dt}\right)_s - \epsilon_X \mathcal{I}\left(\frac{dX}{dt}\right)_s\right], \tag{4.1.13}$$

$$\left(\frac{dX}{dt}\right)_{ir} = \frac{1}{2}\left[\left(\frac{dX}{dt}\right)_s + \epsilon_X \mathcal{I}\left(\frac{dX}{dt}\right)_s\right]. \tag{4.1.14}$$

We have assumed that $\mathcal{I}X = \epsilon_X X$.

Notice that the above equations are to be used to judge the reversible and irreversible parts in the equation for X. Therefore, the force $\boldsymbol{F}$ is a reversible part in the Langevin equation for $\boldsymbol{v}$, but not in the equation for $\boldsymbol{x}$. For the Langevin equation (3.2.1) for $\boldsymbol{v}$, we know $(d\boldsymbol{v}/dt)_s = -\zeta\boldsymbol{v} + \boldsymbol{F}$, so (4.1.13) reads as

$$\frac{1}{2}[(-\zeta\boldsymbol{v} + \boldsymbol{F}) + \mathcal{I}(-\zeta\boldsymbol{v} + \boldsymbol{F})] = \frac{1}{2}[(-\zeta\boldsymbol{v} + \boldsymbol{F}) + (\zeta\boldsymbol{v} + \boldsymbol{F})] = \boldsymbol{F}. \tag{4.1.15}$$

Therefore, as we easily guess, $\boldsymbol{F}$ is the reversible part and $-\zeta\boldsymbol{v}$ is the irreversible part. However, for the Langevin equation for $\boldsymbol{x}$ $(d\boldsymbol{x}/dt)_s = \zeta\boldsymbol{F}$. Therefore, (4.1.13) reads in this case as

$$\frac{1}{2}[\boldsymbol{F} - \mathcal{I}\boldsymbol{F}] = 0. \tag{4.1.16}$$

That is, the term containing $\boldsymbol{F}$ is the irreversible part.

4.1.7 Implication of regression principle

The regression principle (→**4.1.2**) implies that on the average (denoted by the overline) the dynamics of fluctuations is governed by

$$\overline{\frac{dX}{dt}} = \left(\frac{d\overline{X}}{dt}\right)_s \tag{4.1.17}$$

with the right hand side being identical with that given by the macroscopic phenomenological laws (the law governing the dynamics of small macroscopic deviation from equilibrium).[17]

---

[15] However, as we will learn soon, this interpretation is not quite right (→**4.1.13**); only if we ignore short time dynamics (of the time scale of t_m in **3.4.1**), the argument is right. Notice that the initial ensemble may be obtained by selecting samples with extra conditions just as we do in the contraction principle (→**1.6.11**).

[16] This is not the same as the backward play of a movie. In that case $(q,p,t) \to (q,-p,t)$ occurs, because observer's time still flows in the ordinary direction.

[17] A more precise statement may be as follows: make a subensemble of an equilibrium ensemble with the condition that the mesocopic variable X be close to a particular value x. Then, the expectation value of dX/dt for this ensemble is given by the macroscopic phenomenological law.

Therefore, the overline denoting averaging.

The irreversible part of this macroscopic dynamics (the phenomenological law) is driven by the thermodynamic potential $\delta^2 S$[18](→**2.8.5**). The dynamics may be assumed to be linear, so we may write the irreversible phenomenological law as (henceforth $\overline{X}$ will be written as x; we assume that the equilibrium values are defined to be zero)

$$\frac{dx}{dt} = L\nabla_x S, \quad \left(\text{that is, } \frac{dx_i}{dt} = \sum_j L_{ij}\frac{\partial S}{\partial x_j}\right) \tag{4.1.18}$$

where L is a positive semidefinite (→**4.1.8**) constant matrix whose elements are called *kinetic coefficients*.

This macroscopic law is assumed as given in *linear irreversible thermodynamics* (→**4.2.2**). Some examples follow later (→**4.2.4** and on). We will learn that fluid dynamic laws are also examples (→**5.2.5**).

Warning. Up to **4.1.14** we discuss only spatially uniform systems. Thus, no spatial coordinates appear in the argument. □

How reliable is the regression principle? We can compute the dynamical form factor with the aid of hydrodynamic fluctuation, and the result can be compared with experimental results.

Is this a principle of nonequilibrium statistical mechanics? Yes, it is. As we have clearly learned from the history and the logic of equilibrium statistical mechanics, we need some extra mechanical principle (→**2.6.2**). However, no one knows how general the principle is; at least very close to equilibrium it seems to be correct.

4.1.8 Second law implies positive semidefiniteness of kinetic coefficient matrix

Let us compute the change of entropy due to the decay of fluctuation. Since the starting point is a large fluctuation, the entropy of the initial state must be low. Therefore, the dynamics of fluctuation must increase the entropy on the average. That is, $dS/dt \geq 0$.

Let us compute the *entropy production* σ

$$\sigma = \frac{dx}{dt}\cdot\frac{\partial S}{\partial x} = L\frac{\partial S}{\partial x}\cdot\frac{\partial S}{\partial x} \geq 0. \tag{4.1.19}$$

Therefore, the second law implies that the kinetic coefficient matrix L must be positive semidefinite.

4.1.9 Expression of kinetic coefficients in terms of correlation functions[19]

Integrating (4.1.12) with (4.1.18) for a short time Δt, we obtain

$$x(t+\Delta t|t) - x(t) = Lk_B\frac{\partial \log P}{\partial x(t)}\Delta t + \Delta w. \tag{4.1.20}$$

---

[18] Although we say ‘potential’, in contrast to the ordinary potential it drives the system to the direction to ‘increase it.’ That is, the driving force due to this ‘potential’ is given by the + gradient of the ‘potential,’ and *not* by the negative gradient of the potential.

[19] This argument is in L. Onsager, “Reciprocal relations in irreversible processes. I.” Phys. Rev. **37**, 405 (1931).

We have used $\delta^2 S/k_B = \log P+$ const. Here, $x(t+\Delta t|t)$ is the value of x at time $t+\Delta t$ under the condition that x is given at time t. If we multiply $x^T(t)$ (transposition) and average over the initial condition at time t and over the noise, we obtain

$$\langle x(t+\Delta t|t)x^T(t)\rangle - \langle x(t)x^T(t)\rangle = \int dx(t)\, LPk_B \frac{\partial \log P}{\partial x(t)} x^T(t)\Delta t, \tag{4.1.21}$$

$$= k_B L \int dx(t)\, \frac{\partial P}{\partial x(t)} x^T(t)\Delta t, \tag{4.1.22}$$

$$= -k_B L \int dx(t)\, P\frac{\partial x^T(t)}{\partial x(t)}\Delta t. \tag{4.1.23}$$

We have used integration by parts. We conclude

$$L = -\frac{1}{k_B \Delta t}[\langle x(\Delta t)x^T\rangle - \langle xx^T\rangle], \tag{4.1.24}$$

where $x(\Delta t)$ is the value of deviations after Δt with the initial condition x.

4.1.10 Onsager's reciprocity relation

Let us apply the time reversal operation to the correlation function. We assume that each variable has a definite parity $\epsilon = \pm 1$ under this operation:

$$\mathcal{I}x_i(t) = \epsilon_i x_i(-t). \tag{4.1.25}$$

If the system is under the influence of magnetic field $\boldsymbol{B}$ or is being rotated with angular velocity $\boldsymbol{\omega}$, we must take their time reversal into account $\mathcal{I}\boldsymbol{B} = -\boldsymbol{B}$ and $\mathcal{I}\boldsymbol{\omega} = -\boldsymbol{\omega}$.[20]

Let us study the time reversal symmetry of the time correlation function

$$\langle x_i(t)x_j(0)\rangle_{\boldsymbol{B},\boldsymbol{\omega}} = \epsilon_i\epsilon_j\langle x_i(-t)x_j(0)\rangle_{-\boldsymbol{B},-\boldsymbol{\omega}}. \tag{4.1.26}$$

If we assume classical statistical mechanics,[21] all the quantities commute, so we obtain

$$\langle x_i(t)x_j(0)\rangle_{\boldsymbol{B},\boldsymbol{\omega}} = \epsilon_i\epsilon_j\langle x_j(t)x_i(0)\rangle_{-\boldsymbol{B},-\boldsymbol{\omega}}. \tag{4.1.27}$$

That is,

$$L_{ij\,\boldsymbol{B},\boldsymbol{\omega}} = \epsilon_i\epsilon_j L_{ji\,-\boldsymbol{B},-\boldsymbol{\omega}}. \tag{4.1.28}$$

This relation is called *Onsager's reciprocity relation*.[22]

---

[20] Onsager clearly recognized this comment as a trivial side remark, saying "Barring certain exceptional cases* which can readily be recognized and sorted out, $\cdots$," and with a footnote at *[Coriolis forces, external magnetic fields (and permanent magnets).] He discussed the system under magnetic field briefly in Section 7.

[21] Even for quantum cases essentially the same proposition holds, cf., **4.4.13**, **X2.5.7**.

[22] For experimental verification of this symmetry, see, e.g., D. D. Fitts, *Nonequilibrium Thermodynamics* (McGraw-Hill, 1962). However, originally such a symmetry was semiempirically recognized by Thomson and others as summarized in Onsager's paper I.

4.1.11 What happens to quantum systems? Nakajima's Theory[23]

In the above derivation of the reciprocity relation, the commutativity of dynamical variables has been used. Furthermore, honestly speaking, thermodynamic fluctuation theory cannot be used generally for quantum systems, so (4.1.20) cannot be used as a starting point. Thus, a slightly more careful approach is needed, which will be discussed in Section **4.3**.

Nakajima's approach is to start with (4.1.20) and iterpret it as

$$Tr\,[\rho(t+\Delta t|t)x] - x(t) = L\frac{\partial S}{\partial x}\Delta t = -L\frac{X}{T}\Delta t, \tag{4.1.29}$$

where $\rho(t+\Delta t|t)$ is the conditional density matrix conditioned at time t. The, "It seems to us that the most natural answer for the form of" this density matrix is

$$\rho(t+\Delta t|t) = \exp[\beta(F-H)+\beta xX], \tag{4.1.30}$$

where X is the conjugate variable to x (wrt energy).[24] To compute (4.1.30) the following formula is used (→**3.3.6**, **4.3.5**)

$$\exp[-\beta(H-xX)] = e^{-\beta H}\left(1+\int_0^\beta Xx(-i\hbar\lambda)d\lambda+\cdots\right). \tag{4.1.31}$$

Therefore, the LHS of (4.1.29) reads

$$Tr\,[\rho(t+\Delta t|t)x] = \langle x\rangle + X\int_0^\beta \langle x(-i\hbar\lambda)x(t+\Delta t)\rangle d\lambda + o[X]. \tag{4.1.32}$$

Setting $\Delta t = 0$, we get

$$Tr\,[\rho(t|t)x] = x(t) = \langle x\rangle + X\int_0^\beta \langle x(-i\hbar\lambda)x(t)\rangle d\lambda + o[X]. \tag{4.1.33}$$

Combining these two, we obtain

$$Tr\,[\rho(t+\Delta t|t)x] - x(t) = X\int_0^\beta \langle x(-i\hbar\lambda)[x(t+\Delta t)-x(t)]\rangle d\lambda \tag{4.1.34}$$

Therefore,

$$-k_BL\Delta t = \frac{1}{\beta}\int_0^\beta \langle x(-i\hbar\lambda)[x(\Delta t)-x(0)]\rangle d\lambda. \tag{4.1.35}$$

That is, corresponding to (4.1.24), we obtain

$$L = -\frac{1}{k_B\Delta t}[\langle x; x(\Delta t)^T\rangle - \langle xx^T\rangle]. \tag{4.1.36}$$

Practically, if the correlation functions are replaced by the canonical correlation functions (→**3.3.6**, **4.3.5**), many classical results survive.

---

[23] The correct quantum formula for thermal distrubances was obtained by Nakajima: R. Kubo, M. Yokota and S. Nakajima, "Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance," J. Phys. Soc. Jpn., **12**, 1203 (1957).

[24] In the original paper, this form is thought to be natural from the information theoretical argument (maximization of entropy).

4.1.12 Green-Kubo formula for kinetic coefficients

We are studying equilibrium fluctuations, so time correlations depend only on the relative time. Therefore,

$$\frac{d^2}{dt^2}\left\langle x(t)x^T(0)\right\rangle = \frac{d}{dt}\left\langle \frac{dx(t)}{dt}x^T(0)\right\rangle, \tag{4.1.37}$$

$$= \frac{d}{dt}\left\langle \frac{dx(0)}{dt}x^T(-t)\right\rangle, \tag{4.1.38}$$

$$= -\left\langle \frac{dx(0)}{dt}\frac{dx^T(-t)}{d(-t)}\right\rangle, \tag{4.1.39}$$

$$= -\left\langle \frac{dx(t)}{dt}\frac{dx^T(0)}{dt}\right\rangle. \tag{4.1.40}$$

Integrating this formula twice with respect to t, we obtain (remember that ^{T} denotes transposition; the dot $\dot{}$ denotes the time derivative)

$$\langle x(\Delta t)x^T\rangle - \langle xx^T\rangle = -\int_0^{\Delta t} ds \int_0^s ds^T \langle \dot{x}(s')\dot{x}^T(0)\rangle. \tag{4.1.41}$$

That is (→**4.1.9**),

$$L = \frac{1}{k_B\Delta t}\int_0^{\Delta t} ds \int_0^s ds' \langle \dot{x}(s')\dot{x}^T(0)\rangle. \tag{4.1.42}$$

Here, Δt is macroscopically small, but compared with the noise correlation time scale it is very long. Therefore, we may take $\Delta t \to \infty$ limit, keeping the values of macrovariables.[25] Using l'Hospital's rule, we get

$$L = \frac{1}{k_B}\int_0^{\infty} ds \langle \dot{x}(s)\dot{x}^T(0)\rangle. \tag{4.1.43}$$

This is called the *Green-Kubo formula* (for a prototype see **3.2.9**).

Exercise 1. Let us interpret the formula for the diffusion constant as (4.1.43). According to Einstein $D = k_BT/\zeta$ (→(3.1.6)), so $\boldsymbol{v} = \boldsymbol{F}/\zeta = \beta D\boldsymbol{F}$. $\beta\boldsymbol{F} = \nabla_{\boldsymbol{x}} \log P$, where $P \propto e^{-\beta V}$ with V being the potential of the force $\boldsymbol{F}$. That is, we may identify D/k_B and L, so (4.1.43) implies (3.2.21).□

4.1.13 Warning: noise ensemble average is not enough

We have interpreted $\langle\ \rangle$ up to this point in this section as a noise ensemble average. Let us demonstrate that this interpretation contradicts equilibrium statistical mechanics.

Let us consider a simple situation that only x_1 deviates from equilibrium. Then,

$$\langle \dot{x}_1 x_1\rangle = \left\langle L_{11}\frac{dS}{dx_1}x_1\right\rangle = -k_BL_{11} < 0, \tag{4.1.44}$$

so something must be wrong.

The reader should have realized that in the derivation of the Green-Kubo relation we

---

[25] If you wish to mathematize the procedure we are using, we must take that ratios of time scales meso/micro and macro/meso and take both to infinity.

have ignored the correlations such as $\langle \dot{x}_i(0)x_j(0)\rangle$. This is not the problem so long as $\dot{}$ is interpreted as the true time derivative: it is easy to demonstrate that $\langle \dot{x}_1 x_1\rangle = 0$.[26]

The time derivative in the phenomenological linear law is actually calculated, according to (4.1.20), as a short time average of the true time derivative:

$$\frac{A(t+\Delta t)-A(t)}{\Delta t} = \frac{1}{\Delta t}\int_t^{t+\Delta t} ds\dot{A}(s). \tag{4.1.45}$$

The paradox above is caused by careless identification of macroscopic time derivatives and mechanically true time derivatives.

What is the lesson to learn? Although we introduced the Langevin equation as a true mechanical equation of motion, if the driving force is thermodynamic, it is extremely dangerous to interpret it this way. It should be interpreted as a semimacroscopic equation whose time derivative is not the true mechanical time derivative but its short time average as (4.1.45).

4.1.14 Green-Kubo formula for transport coefficients

We have given a general expression for the Onsager coefficients L in **4.1.12**, but for spatially uniform systems. We have already discussed spatial fluctuations and transport in nonuniform systems, so let us derive a more convenient formula for general nonuniform systems.

Probably, the easiest approach is to divide the space into cells and the variables in each cell $\{x\}$ are denoted with the representative cell position as $\{x(\boldsymbol{r})\}$. Then, we interpret (4.1.18) as (we write $\partial S/\partial x = F$)

$$\frac{\partial x(\boldsymbol{r})}{\partial t} = \sum_{\boldsymbol{r}'} L(\boldsymbol{r},\boldsymbol{r}')F(\boldsymbol{r}'). \tag{4.1.46}$$

Using (4.1.43), let us first proceed quite formally without paying attention to the convergence of the time integral:

$$\begin{aligned}
\sum_{\boldsymbol{r}'} L(\boldsymbol{r},\boldsymbol{r}')F(\boldsymbol{r}') &= \frac{1}{k_B}\int_0^\infty dt \sum_{\boldsymbol{r}'}\left\langle \frac{\partial x(\boldsymbol{r},t)}{\partial t}\frac{\partial x(\boldsymbol{r}')^T}{\partial t}\right\rangle F(\boldsymbol{r}'), && (4.1.47)\\
&= \frac{1}{k_B}\int_0^\infty dt \sum_{\boldsymbol{r}'}\langle div\, \boldsymbol{j}(\boldsymbol{r},t)\, div\, \boldsymbol{j}(\boldsymbol{r}')^T\rangle F(\boldsymbol{r}'), && (4.1.48)\\
&= -div_{\boldsymbol{r}}\int_0^\infty dt \sum_{\boldsymbol{r}'}\langle \boldsymbol{j}(\boldsymbol{r},t)\boldsymbol{j}(\boldsymbol{r}')^T\rangle \nabla_{\boldsymbol{r}'}F(\boldsymbol{r}') && (4.1.49)
\end{aligned}$$

Here, $\boldsymbol{j}$ is the fluctuating part of the flux (non-systematic part).

On the other hand, the phenomenological law has the form in **4.2.1**, i..e.,

$$\frac{\partial x(\boldsymbol{r})}{\partial t} = -div(L\nabla F). \tag{4.1.50}$$

If F changes slowly compared with the correlation length of fluctuating flux, we may set $\boldsymbol{r}' = \boldsymbol{r}$ in F, and we arrive at

$$L = \frac{1}{k_B}\int_0^\infty dt \int d\boldsymbol{r}\left\langle \boldsymbol{j}(\boldsymbol{r},t)\boldsymbol{j}(0)^T\right\rangle. \tag{4.1.51}$$

---

[26] Obviously, $\langle \dot{x}_1 x_1\rangle + \langle x_1\dot{x}_1\rangle = 0$. Also, $\langle x_1(t)x_1(0)\rangle = \epsilon_1^2\langle x_1(0)x_1(t)\rangle$, so even quantum mechanically, this average vanishes.

If we introduce

$$\boldsymbol{J}(t) = \int d\boldsymbol{r}\, \boldsymbol{j}(\boldsymbol{r},t), \tag{4.1.52}$$

then

$$L = \frac{1}{Vk_B}\int_0^\infty dt\, \left\langle \boldsymbol{J}(t)\boldsymbol{J}(0)^T \right\rangle. \tag{4.1.53}$$

Here, V is the volume of the system (or the domain being paid attention). This is almost the final result.

However, the time integral in (4.1.53) may not exist, because $\lim_{t\to\infty}\boldsymbol{J}$ may not be zero. To avoid this, we can simply subtract this value from $\boldsymbol{J}$, and define $\boldsymbol{J}^- = \boldsymbol{J} - \lim_{t\to\infty}\boldsymbol{J}$ to obtain the final result:

$$L = \frac{1}{Vk_B}\int_0^\infty dt\, \langle \boldsymbol{J}^-(t)\boldsymbol{J}^-(0)^T\rangle. \tag{4.1.54}$$

This is the *Green-Kubo formula* for transport coefficients.

4.1.15 Onsager's principle

This is a continuation of **3.2.10**. Onsager's principle takes the following large deviation principle form:

$$P(\Delta x/\Delta t \sim \dot{x}|x(t)) \simeq e^{-\tau I(\dot{x}|x(t))}. \tag{4.1.55}$$

Here, $x(t)$ is explicitly written to denote the starting point of the change $\Delta x(t)$. Combining this with the regression principle (→**4.1.2**), we conclude that the rate function is written in terms of macroobservables.

The basic idea of Onsager's principle (interpreted in the large deviation principle sense) has a clear intuitive and computational meaning:
If we wish to study a macroscopic evolution law, take statistics of a macroscopically short time (= mesoscopic time) changes; their average provides the most probable time derivative = the macroscopic evolution law. The distribution of the fluctuation around the most probable evolution law is governed by the rate function that dictates the kinetic level stochastic behavior of the system. Thus, there is a possibility that we can use short time simulations to obtain coarse-grained evolution rule to accelerate simulations.

As noted in **1.6.3** some sort of variational principle is also expected for the mesoscopic dynamics (→**4.1.22**).

4.1.16 Onsager-Hashitsume path integral

As already noted in **3.2.10**, Onsager's principle gives the transition probability:

$$P(x(t+\tau)|x(t)) \simeq e^{-\tau I(\dot{x}(t)|x(t))}. \tag{4.1.56}$$

Therefore,

$$P(x(t+2\tau)|x(t)) \simeq \sum_{x(t+\tau)} e^{-\tau[I(\dot{x}(t+\tau)|x(t+\tau))+I(\dot{x}(t)|x(t))]}, \tag{4.1.57}$$

where the summation (that may be an integration) over $x(t+\tau)$ means the summation for all the values of x at the intermediate time step $t+\tau$ between t and $t+2\tau$.

Repeating this procedure, we can bridge the macroscopic time span $[t_0, t_1]$ as

$$P(x(t_1)|x(t_0)) \simeq \sum_{x_1,x_2,\cdots,x_{n-1}} e^{-\tau\sum_{i=1}^{n-1} I(\dot{x}_i|x_i)}, \tag{4.1.58}$$

where $x_i = x(t_0 + i\tau)$ and we assume $t_1 = t_0 + n\tau$. Since τ is macroscopically very small (in the mathematical idealization, infinitesimal and may be regarded macroscopically as dt), (4.1.58) is understood as a *path integral*, and may be formally rewritten as (→**X1.10.5**)

$$P(x_F, t_1|x_I, t_0) \simeq \int_{x(t_0)=x_I}^{x(t_1)=x_F} \mathcal{D}[\{x(t)\}] \exp\left[-\int_{t_0}^{t_1} ds\, I(\dot{x}(s)|x(s))\right], \tag{4.1.59}$$

This is often called the *Onsager-Machlup path integral*, but should better be called the *Onsager-Hashitsume path integral.*[27]

Thus, Onsager's principle may be restated as follows:
The Lagrangian is a rate function: the Lagrangian for the path integral representation of the time evolution of probability of observing a macroscopic state is given by the rate function for the microscopic time scale average of the time derivative of macroscopic quantities specifying the state.

4.1.17 Gaussian approximation and Langevin equation

The rate function I is convex and vanishes for the most probable value (expectation value) (→**1.6.2**). Therefore, I must vanish if and only if the phenomenological law

$$\dot{x} = f(x) \tag{4.1.60}$$

holds. I is usually differentiable, so we may assume that near the origin I has a quadratic form:[28]

$$I(\dot{x}(t)|x(t)) = \frac{\Gamma}{4}(\dot{x} - f(x))^2. \tag{4.1.61}$$

This rate function is compatible with

$$\dot{x} = f(x) + w \tag{4.1.62}$$

where w is a Gaussian noise such that $\langle w \rangle = 0$ and

$$\langle w(t)w(s)\rangle = \frac{2}{\Gamma}\delta(t-s), \tag{4.1.63}$$

where $\langle\ \rangle$ is the ensemble average (= average over the stochastic parameter →**1.4.1**). That is, the Langevin equation (4.1.62) is another expression of Onsager's principle in the Gaussian approximation.

The numerical factor Γ may be related to the variance of $\dot{x}$. Let us compute it. Obviously,

$$2/\Gamma = \tau\langle \dot{x}^2\rangle - \tau\langle \dot{x}\rangle^2. \tag{4.1.64}$$

Notice that the first term on RHS is of order 1 and the second term of order τ.[29] Therefore, we may write for the use at our time scale

$$\frac{1}{\Gamma} = \frac{1}{2}\tau\langle \dot{x}^2\rangle = \frac{1}{2\tau}\int_0^\tau ds \int_0^\tau ds' \left\langle \frac{dx}{dt}(s)\frac{dx}{dt}(s')\right\rangle \simeq \int_0^\infty ds \left\langle \frac{dx}{dt}(s)\frac{dx}{dt}(0)\right\rangle. \tag{4.1.65}$$

---

[27] The transition probability (4.1.56) was essentially proposed by Onsager. Hashitsume introduced the path integral in 1951; Onsager introduced the same path integral with his student Machlup in 1953. Thus, it is fair to call this path integral the Onsager-Hashitsume path integral. We will learn its relation to the *Feynman-Kac formula* later (→**X1.10.5**).

[28] Of course, more generally, I is a general quadratic functional of $\dot{x} - f(x)$. This is equivalent to assuming a noise with memory as in **3.3.6**.

[29] Roughly speaking, $\langle \dot{x}\rangle = O[1]$, and $\langle \dot{x}^2\rangle = O[1/\tau]$.

Here, dx/dt denotes the true mechanical time derivative, and we have assumed that the noise correlation decays sufficiently fast in the time scale of t_M, so we have replaced the upper limits of the integrals with infinity.

4.1.18 Onsager's principle near equilibrium and fluctuation dissipation relation

The original Onsager principle was proposed for the linear nonequilibrium regime. Here, the phenomenological law reads (→**4.1.7**)

$$\dot{x} = L\frac{\partial S}{\partial x}, \tag{4.1.66}$$

so that the Langevin equation is

$$\dot{x} = L\frac{\partial S}{\partial x} + w, \tag{4.1.67}$$

where the correlation function of the noise is given in (4.1.63). The steady distribution of this Langevin equation may be obtained as

$$P(x) \propto e^{L\Gamma S} \tag{4.1.68}$$

as can be seen from the result in **3.5.5** (a good exercise[30]).

The equilibrium distribution of x is given by $\propto e^{S/k_B}$ (→**2.8.5**), so

$$\Gamma = 1/k_B L. \tag{4.1.69}$$

Combining this and (4.1.65), we obtain the *Green-Kubo relation* (→**4.1.12**):

$$L = \frac{1}{k_B}\int_0^\infty ds \left\langle \frac{dx}{dt}(s)\frac{dx}{dt}(0)\right\rangle. \tag{4.1.70}$$

Thus, Onsager's principle stated in the large deviation theoretical form + thermodynamic fluctuation theory can construct a self-contained nonequilibrium statistical theoretical framework.

4.1.19 Original Onsager's principle

The transition probability in **4.1.15** close to equilibrium reads with the aid of **4.1.18**

$$P(x(t+\tau)|x(t)) \simeq \exp\left[-\frac{\tau}{4Lk_B}\left(\dot{x} - L\frac{\partial S}{\partial x}\right)^2\right], \tag{4.1.71}$$

---

[30] The Fokker-Planck equation reads

$$-\frac{\partial}{\partial x}\left(L\frac{\partial S}{\partial x}P\right) + \frac{1}{\Gamma}\frac{\partial^2}{\partial x^2}P = 0.$$

The equilibrium state satisfies

$$L\frac{\partial S}{\partial x}P - \frac{1}{\Gamma}\frac{\partial P}{\partial x} = 0,$$

which implies (4.1.68).

so the joint probability at two different times reads (see thermodynamic fluctuation theory **2.8.5**)

$$P(x(t+\tau),x(t)) \propto P(x(t+\tau)|x(t))e^{S(t)/k_B}. \tag{4.1.72}$$

To study its structure further, let us consider the exponents of (4.1.72):

$$\frac{1}{k_B}\left\{S(t)-\frac{\tau}{4L}\left(\dot{x}-L\frac{\partial S}{\partial x}\right)^2\right\}=\frac{S(t)}{k_B}-\frac{\tau}{2k_B}\left[\frac{1}{2}L^{-1}\dot{x}^2+\frac{1}{2}L\left(\frac{\partial S}{\partial x}\right)^2\right]+\frac{\tau}{2k_B}\frac{dS}{dt}. \tag{4.1.73}$$

Now, $\tau dS/dt = S(t+\tau)-S(t)$, so we arrive at

$$P(x(t+\tau),x(t)) \simeq \exp\frac{1}{2k_B}\left\{[S(t+\tau)+S(t)]-\tau\left[\frac{1}{2}L^{-1}\dot{x}^2+\frac{1}{2}L\left(\frac{\partial S}{\partial x}\right)^2\right]\right\}. \tag{4.1.74}$$

This is the original Onsager's principle (with Hashitsume's correction).

The transition probability reads

$$P(x(t+\tau)|x(t)) \simeq \exp\frac{1}{2k_B}\left\{[S(t+\tau)-S(t)]-\tau\left[\frac{1}{2}L^{-1}\dot{x}^2+\frac{1}{2}L\left(\frac{\partial S}{\partial x}\right)^2\right]\right\}. \tag{4.1.75}$$

In these formulas the overall factor $1/2k_B$ instead of $1/k_B$ is due to Hashitsume.

4.1.20 Onsager principle for many variables near equilibrium

Let us summarize what we have learned, while generalizing the formulas to the many variable case.

The phenomenological law is (corresponding to (4.1.66))

$$\dot{x}_i=\sum_j L_{ij}\frac{\partial S}{\partial x_j}, \tag{4.1.76}$$

where L_{ij} are Onsager coefficients (a good occasion to review **4.1.12**). The rate function for $\dot{x}_i$ is (corresponding to (4.1.61))

$$I(\{\dot{x}_i(t)\}|\{x_i(t)\})=\frac{1}{4}\sum_{ij}\Gamma_{ij}\left(\dot{x}_i(t)-\sum_{i'}L_{ii'}\frac{\partial S(t)}{\partial x_{i'}(t)}\right)\left(\dot{x}_j(t)-\sum_{j'}L_{jj'}\frac{\partial S(t)}{\partial x_{j'}(t)}\right), \tag{4.1.77}$$

where

$$k_B\Gamma_{ij}=(L^{-1})_{ij} \tag{4.1.78}$$

is the fluctuation dissipation relation with an explicit form corresponding to (4.1.70):

$$L_{ij}=\frac{1}{k_B}\int_0^\infty ds\left\langle\frac{dx_i}{dt}(s)\frac{dx_j}{dt}(0)\right\rangle. \tag{4.1.79}$$

The Langevin equation is (corresponding to (4.1.67)

$$\dot{x}_i=\sum_j L_{ij}\frac{\partial S}{\partial x_j}+w_i. \tag{4.1.80}$$

with the noise satisfying $\langle w_i \rangle = 0$ and

$$\langle w_i(t) w_j(s) \rangle = 2L_{ij} k_B \delta(t-s). \tag{4.1.81}$$

The transition probability reads (corresponding to (4.1.75))

$$\begin{aligned} & P(\{x_i(t+\tau)\} | \{x_i(t)\}) \\ \simeq \ & \exp \frac{1}{2k_B} \left\{ [S(t+\tau) - S(t)] - \tau \left[\frac{1}{2} \sum_{ij} (L^{-1})_{ij} \dot{x}_i \dot{x}_j + \frac{1}{2} \sum_{ij} L_{ij} \left(\frac{\partial S}{\partial x_i} \right) \left(\frac{\partial S}{\partial x_j} \right) \right] \right\}. \end{aligned} \tag{4.1.82}$$

This is the original Onsager's principle.

These formulas can easily be guessed from the single variable case (4.1.75) discussed in detail above.

4.1.21 Variational principles near equilibrium

A large deviation principle implies a variational principle (→**1.6.3**). In our case, the variational principle implies that what we actually observe phenomenologically is the most probable behavior of the system. The most probably behavior corresponds to the path from $x(t)$ to $x(t+\tau)$ that maximizes the transition probability. Therefore, $x(s)$ that maximizes the formula inside the curly brackets in (4.1.82) should correspond to the phenomenology. This is a tautology for us.

The term inside the curly brackets in (4.1.82) may be rewritten as follows:

$$\tau \left\{ \sigma(t) - [\Phi(\{\dot{x}_i\}) + \Psi(\{F_i\})] \right\}, \tag{4.1.83}$$

where σ is the *entropy production rate*, and

$$\Phi(\{\dot{x}_i\}) \equiv \frac{1}{2} \sum_{ij} (L^{-1})_{ij} \dot{x}_i \dot{x}_j, \tag{4.1.84}$$

$$\Psi(\{F_i\}) \equiv \frac{1}{2} \sum_{ij} L_{ij} F_i F_j \tag{4.1.85}$$

with $\partial S / \partial x_i = F_i$ (the conjugate driving force for x_i). Φ is called the *dissipation function.*[31]

4.1.22 Onsager's principle of minimum energy dissipation

If the thermodynamic forces F_i are given (i.e., x_i are fixed), $\dot{x}_i$ are determined by the condition of maximizing (4.1.83), or equivalently, by the minimization of

$$\Phi(\{\dot{x}_i\}) - \sigma(t). \tag{4.1.86}$$

This is called *Onsager's principle of minimum energy dissipation.* Differentiating the above formula with respect to $\dot{x}_i$, we obtain

$$\sum_i (L^{-1})_{ij} \dot{x}_j = \frac{\partial \sigma}{\partial \dot{x}_i}. \tag{4.1.87}$$

---

[31] One thing we must be careful about is that in the above formalization F is a function of $\{x_i(t)\}$ that are given as the state from which the transition occurs. However, F evaluated at $\{x_i(t+\tau)\}$ (the outcome of the transition) and the original F differs by a quantity only of order some positive power of τ (likely to be $O[\tau^{1/2}]$), so we may interpret all the variables in (4.1.83) are after the transition.

The entropy production rate can be written as

$$\sigma(t) = \frac{dS}{dt} = \sum_i \frac{\partial S}{\partial x_i}\dot{x}_i = \sum_i F_i \dot{x}_i. \tag{4.1.88}$$

Therefore, (4.1.87) indeed implies the phenomenological law $\dot{x}_i = \sum_j L_{ij}F_j$ (→**4.1.7**).

4.2 Linear Irreversible Thermodynamics

4.2.1 Macroscopic phenomenological laws

To discuss heat and electric currents, we must discuss the spatially nonuniform fluctuations.

For the deviation of a quantity $q(\boldsymbol{r})$ at $\boldsymbol{r}$, the thermodynamic driving force is $F(\boldsymbol{r}) = \partial S/\partial q$ evaluated at $\boldsymbol{r}$; here for simplicity we assume only one quantity q fluctuates. This driving force tries to change (actually to diminish) the deviation q at $\boldsymbol{r}$. Thus, we can imagine at each point there is $F(\boldsymbol{r})$, and, if quantity q is conserved (like energy or matter →**5.2.4**), its flow occurs only when F is not uniform; its gradient is the actual driving force of the flow (flux) $\boldsymbol{J}$ of q:

$$\boldsymbol{J} = L_{qq}\nabla \left.\frac{\partial S}{\partial q}\right|_{\boldsymbol{r}}. \tag{4.2.1}$$

There is no negative sign, because the driving direction is opposite to the ordinary potential as discussed in **4.1.7** (footnote).

Generalizing this to cases with simultaneously fluctuating quantities, the basic phenomenological law reads

$$\boldsymbol{J}_i = \sum_j L_{ij}\nabla\frac{\partial S}{\partial x_j}. \tag{4.2.2}$$

Here, ∇ is the gradient with respect to the actual spatial coordinates, and L is the Onsager's kinetic coefficient matrix introduced in **4.1.7**.[32] The expression of L_{ij} in terms of flux correlations is called the *Green-Kubo formula* for transport coefficients (→**4.1.14**).

4.2.2 Linear irreversible thermodynamic framework

As we have seen in Chapter 1, thermodynamics was the key in constructing equilibrium statistical mechanics. Then, it should be useful to summarize nonequilibrium phenomenology at least for the cases close to equilibrium. This is the *linear irreversible thermodynamics.*

The standing assumption is:

[**0**] The macroscopic state of a system is described by the field of equilibrium thermodynamic variables. At each space-time point ordinary thermodynamic relations (e.g., the Gibbs relation) hold.

That is, if the equilibrium state is described by a thermodynamic space spanned by extensive quantities $\{x_i\}$, then the macroscopic state of the system is uniquely described by

---

[32] Not all the quantities couple each other; for example a scalar quantity does not couple with a vector quantity, IF a system is isotropic (the so-called *Curie's theorem*). This is understood easily by considering the rotational symmetry of the quantities.

the field of the corresponding densities. This is called the *local equilibrium assumption*.[33]
[**1**] Macroscopic phenomenology holds: i.e.,

$$\boldsymbol{J}_i = \sum_j L_{ij} \nabla F_i, \tag{4.2.3}$$

where $\boldsymbol{J}_i$ is the flux of x_i, and F_i is the conjugate (intensive) variable (wrt entropy). This becomes simpler without ∇, if we do not pay attention to the spatial nonuniformity.
[**2**] Onsager's reciprocity relation holds: $L_{ij} = L_{ji}$.[34]

This general theoretical framework is illustrated with thermoelectric effects in **4.2.4**-.

4.2.3 Minimum entropy production in linear regime

Let us differentiate σ, assuming that L is symmetric and constant. With the summation convention, we have

$$\frac{d\sigma}{dt} = 2L_{ij} \frac{\partial S}{\partial x_i} \frac{\partial^2 S}{\partial x_j x_k} \frac{dx_k}{dt} = 2\frac{dx_j}{dt} \frac{\partial^2 S}{\partial x_j x_k} \frac{dx_k}{dt}. \tag{4.2.4}$$

but this is non-positive, because of the stability of equilibrium (which is again a consequence of the second law):

$$\frac{d\sigma}{dt} \le 0. \tag{4.2.5}$$

This implies that if there is an entropy production minimum state, it is stable.[35] This is called the *principle of minimum entropy production* (due to Prigogine). Notice that constancy of L is crucial for this principle.

Thomson's *principle of minimum heat production* is a classic example.

4.2.4 Thermoelectric effects: setting up theoretical framework

Consider a system that allows heat and electrical current conductions. Here, with the aid of the thermoelectric effect as an example, a step-by-step explanation of how to use linear irreversible thermodynamics is given.
(1) First, we must identify the thermodynamic variables whose change cause currents. For heat, it is energy (density) u, and for electric current its carrier density n.
(2) We must know the dependence of entropy (density) on these variables:

$$ds = \frac{1}{T}du - \frac{\mu}{T}dn + \cdots. \tag{4.2.6}$$

Here, μ is the electrochemical potential of the charge carrier (that is, the ordinary chemical potential + $e\phi$, where ϕ is the electric potential, and e is the charge carried by the carrier

---

[33] The word 'local equilibrium' appears in different contexts with different meanings. In statistical mechanics 'local equilibrium' distribution (→**5.2.3**) implies the equilibrium distribution function with space-time dependent parameters as we have already seen in the context of the Boltzmann equation (→**2.4.9**). This 'local equilibrium state' cannot properly describe transport properties (→**4.3.3**). In contrast, 'local equilibrium state' in linear irreversible thermodynamics describes transport phenomena. Thus, thermodynamic local equilibrium states *cannot* be described by the local equilibrium distribution functions of statistical mechanics. Mori's theory of thermal transport properties (→**4.3.4**) clearly tells us the difference.

[34] Here, we simplify this, assuming that there is no magnetic field nor Coriolis force (see **4.1.10** for the detail).

[35] Again, we follow Lyapunov's argument.

under consideration).
(3) Now, identifying F, we can write down the linear phenomenological law:

$$\boldsymbol{J}_u = L_{uu}\nabla\frac{1}{T} + L_{un}\nabla\left(-\frac{\mu}{T}\right), \tag{4.2.7}$$

$$\boldsymbol{J}_n = L_{nu}\nabla\frac{1}{T} + L_{nn}\nabla\left(-\frac{\mu}{T}\right). \tag{4.2.8}$$

Here, all the kinetic coefficients are assumed to be constant.
(4) Onsager's reciprocity relation (→**4.1.10**) implies $L_{un} = L_{nu}$,[36] and the positive semidefiniteness of L (→**4.1.8**) implies $L_{uu} > 0$, $L_{nn} >$, and $L_{uu}L_{nn} - L_{nu}^2 \geq 0$.

4.2.5 Thermoelectric effect in terms of more convenient fluxes

The remaining task is to make the phenomenological relation easier to use. Perhaps the first thing to do is to introduce more practical fluxes.

To consider an ordinary metal, we assume the charge carriers are electrons, and let us introduce the charge flux (current density) $\boldsymbol{i}$ and heat flux $\boldsymbol{j}$ as

$$\boldsymbol{i} = -e\boldsymbol{J}_n, \quad \boldsymbol{j} = \boldsymbol{J}_u - \mu\boldsymbol{J}_n. \tag{4.2.9}$$

Here, $Tds = du - \mu dn$ has been used to obtain the second relation. e is the elementary charge (> 0). We can rewrite (4.2.8) as

$$\boldsymbol{i} = L_{11}\frac{1}{T}\nabla\mu + L_{12}\nabla\frac{1}{T}, \tag{4.2.10}$$

$$\boldsymbol{j} = L_{21}\frac{1}{T}\nabla\mu + L_{22}\nabla\frac{1}{T}. \tag{4.2.11}$$

Exercise 1. As is warned in (4) of **4.2.4**, there is no automatic guarantee of $L_{12} = L_{21}$ for such a 'non-thermodynamic' choice of driving forces. However, in this case we have $L_{12} = L_{21}$. Demonstrate this symmetry. Also show that this is a positive definite matrix. □

4.2.6 Thermoelectric effects: practical results

Now, our starting point is (4.2.10) and (4.2.11).
(1) First, we should write *electrical conductivity* σ and *thermal conductivity* κ in terms of L_{ij}. σ is measured under constant temperature, and the measurement is done for a uniform material $\nabla\mu = -e\nabla\phi = e\boldsymbol{E}$, where $\boldsymbol{E}$ is the electric field. Therefore,

$$\boldsymbol{i} = L_{11}\frac{e}{T}\boldsymbol{E} \;\Rightarrow\; \sigma = \frac{eL_{11}}{T}. \tag{4.2.12}$$

To study thermal conductivity, there is no current $\boldsymbol{i} = 0$. However, we cannot say μ is constant, because temperature changes: $\boldsymbol{i} = 0$ implies $\nabla\mu = (L_{12}/TL_{11})\nabla T$. Putting this into (4.2.11), we obtain

$$\boldsymbol{j} = -(|L|/TL_{11})\nabla T, \;\Rightarrow\; \kappa = |L|/T^2L_{11}, \tag{4.2.13}$$

---

[36] This relation is automatically guaranteed only when we correctly choose thermodynamic driving force and so forth as we have done.

where $|L|$ is the determinant of L and is positive.

Interesting phenomena are due to the cross effects among various currents.

(2) If there is no electrical current,

$$L_{11}T\nabla\mu = L_{12}\nabla T, \tag{4.2.14}$$

so $\Delta\mu = (L_{12}/L_{11})\Delta\log T$. This allows us to translate the temperature difference into the electrochemical potential difference, which can be measured as an electrical potential difference, if we make an electrical cell connecting two different metals. This is the principle of *thermocouple* and is called the *Seebeck effect*.

(3) Suppose we keep T constant everywhere and maintain the electrical current. Then, we have

$$\boldsymbol{j} = (L_{21}/L_{11})\boldsymbol{i}. \tag{4.2.15}$$

If two different materials are connected, the kinetic coefficients are generally different, so if $\boldsymbol{i}$ is constant, $\boldsymbol{j}$ cannot be constant. Thus, we may make a cooling device. This effect is called the *Peltier effect*. You can generate electricity by maintaining $\boldsymbol{j}$ as well.

Exercise 1. The thermoelectric effect is essentially due to the change of electrochemical potential depending on temperature. Therefore, temperature gradient can induce material diffusion as well. This is called *thermal diffusion*. Discuss the *Soret effect* and *Dufour effect*. What are they? □

Exercise 2. Two chemical fluxes or two chemical reactions couple. Discuss typical examples. □

4.3 Fluctuation and Response: Introduction to General theory

If we wish to have a general theory of fluctuation and response, purely mechanical theory cannot be constructed, because not all the perturbations can naturally be treated as mechanical perturbations and because even a purely mechanical perturbation usually induces various thermal effects.

Therefore, in these introductory lecture notes a physically very natural approach due to Mori is adopted to explain the general framework. The usual formally mechanical approach is given only for the completeness of the notes later in Section **4.5**.

4.3.1 Impossibility of purely mechanical study

The reader may suggest that since we know the fundamental equation of motion of the distribution function (or the density operator), the Liouville (→**2.7.7**) or the von Neumann equation (→**2.7.13**), we have only to solve the equation under an appropriate initial condition and boundary condition.[37] One choice of the initial condition is a canonical distribution imposed at $t = -\infty$ (infinite past). The Liouville operator is a linear operator, so we can write down the formal solution, if we know the Hamiltonian and the boundary conditions for all t.

---

[37] Studying an infinite system is out of any serious question.

The reader may assert the system Hamiltonian is not hard to write down. The Hamiltonian includes the effect of external mechanical perturbations. Due to the passivity (→**2.7.19**) the system becomes hotter unless the heat is discarded. This must be accomplished by the interaction with the external world. It may be possible to model the environment (heat bath)-system interaction in terms of a certain interaction Hamiltonian. However, the variables describing the external thermal environment is not governed by the system and interaction Hamiltonians. Therefore, purely mechanical description of the system turns out to be difficult. Just as in the case of equilibrium statistical mechanics what is outside the description of mechanics dictates the ensemble (= distribution).

4.3.2 Local equilibrium state

We have realized that pure mechanical study is out of question in **4.3.1**. The easiest conceivable strategy may be to use the values of macroobservables at t to construct the most probable distribution, following the strategy outlined in **1.6.12**. The resultant distribution is called the *local equilibrium distribution*.[38] It has the following form (→**5.1.4**):

$$\rho_L = \frac{1}{Z}\exp\left\{-\sum_i \beta(\boldsymbol{x}_i)\left[\frac{m(\boldsymbol{v}_i-\boldsymbol{v}(\boldsymbol{x}_i))^2}{2}+\frac{1}{2}\sum_j \phi(\boldsymbol{x}_i-\boldsymbol{x}_j)-\mu(\boldsymbol{x}_i)\right]\right\}, \tag{4.3.1}$$

where $\boldsymbol{x}_i$ is the position of the i-th particle, $\boldsymbol{v}_i$ is the velocity of the i-th particle, β (local inverse temperature), μ (local chemical potential) and $\boldsymbol{v}$ (local flow velocity) are functions of positions (i.e., they are classical fields) that change very slowly in space (→**5.1.4**, i.e., microscopically almost constant); ϕ is the pairwise particle-particle interaction potential that is assumed to be short-ranged (i.e., at the length scale where ϕ changes appreciably, the fields introduced above, i.e., $\boldsymbol{v}$, β and μ, almost do not change).[39]

4.3.3 Local equilibrium state does not describe transport phenomena

Unfortunately, this distribution cannot describe transport phenomena (→**5.1.8**). At each point $\boldsymbol{x}$, it is just the ordinary equilibrium distribution, so if we compute fluxes, they are fluxes of ideal fluid. Or, intuitively speaking, each particle cannot tell whether the ‘local’ equilibrium it is in is local or global, so there is no irreversible process (→**5.1.10**).

For example, the energy flux $\hat{\boldsymbol{j}}_E$ is microscopically defined as

$$\hat{\boldsymbol{j}}_E(\boldsymbol{x}) = \sum_i\left[\frac{m(\boldsymbol{v}_i-\boldsymbol{v}(\boldsymbol{x}_i))^2}{2}+\frac{1}{2}\sum_j \phi(\boldsymbol{x}_i-\boldsymbol{x}_j)\right]\boldsymbol{v}_i. \tag{4.3.2}$$

Here, the summation over i is just for particles in the volume element in $d\boldsymbol{x}$ around $\boldsymbol{x}$.[40] If we average this with the local equilibrium distribution, we obtain

$$\boldsymbol{j}_E(\boldsymbol{x}) = u(\boldsymbol{x})\boldsymbol{v}(\boldsymbol{x}), \tag{4.3.3}$$

---

[38] In the framework of linear nonequilibrium thermodynamics (→**4.2.2**), it is assumed that the system is in local equilibrium. Its meaning is that the Gibbs relation holds in each volume element. This assumption is not the same as the local equilibrium assumption here in the statistical mechanics sense (→**5.1.8**). In the former, the thermodynamic observables are assumed to satisfy the usual thermodynamic relations, but this does not mean that the distribution is the local equilibrium distribution being discussed here. As we will see soon, Mori constructs the distribution that is not a local equilibrium distribution but whose thermodynamic variables take the same values dictated by the local equilibrium distribution.

[39] The binary interaction need not be assumed; any type of interactions are admissible if short ranged.

[40] **Notation**: Henceforth, the quantities with hats imply their microscopic expressions, and their averages (macroscopic observables) will be without hats.

where $u(\boldsymbol{x})$ is the internal energy density (the total energy seen from the co-moving or the center of mass coordinates), because the product of the fluctuation of energy (even under spatial inversion) and that of the velocity (odd under spatial inversion) vanishes by symmetry. Therefore, there is no conduction of energy (only advection) is possible; →**5.1.9**). That is, there is no heat conduction.

The same is true for the particle flux, so there is no diffusion process (again, only advection is possible). With a similar argument we can show that the momentum flux has only the advection terms. Therefore, there is no viscosity.

In short, local equilibrium distribution does not describe irreversible processes.

Notice that this distribution is not even a stationary solution of the Liouville equation (→**2.7.7**).

4.3.4 Mori's approach from local equilibrium: idea

We have seen that the local equilibrium distribution cannot describe nonequilibrium states properly (→**4.3.3**). However, the distribution can give correct expectation values of static observables at each time t of a system that is not in equilibrium. Therefore, if the thermodynamic fields (= the intensive variable fields) are correct, the local equilibrium distribution must be in a certain sense close to the true nonequilibrium distribution.

If we use the local equilibrium distribution at time t as the initial condition and solve the Liouville (→**2.7.7**) or von Neumann equation (→**2.7.13**) for a very short time (for a mesoscopic time scale which is a macroscopically negligibly short time), we should obtain the true nonequilibrium distribution compatible with the static observables at present. This is Mori's fundamental idea.[41]

4.3.5 Technical interlude: canonical correlation, etc., revisited

Because of the noncommutativity of operators, various calculations become complicated, so here some useful formulas are summarized.

Since generally $e^{A+B} \neq e^A e^B$, we cannot write $e^{A+\epsilon B} = e^A(1+\epsilon B + \cdots)$, even if ϵ is a small number, so even the differentiation of e^{A+xB} with respect to x becomes nontrivial. We have

$$\frac{d}{dx} e^{A+xB} = \int_0^1 d\lambda\, e^{(1-\lambda)(A+xB)} B e^{\lambda(A+xB)} = \int_0^1 d\lambda\, e^{\lambda(A+xB)} B e^{(1-\lambda)(A+xB)}. \tag{4.3.4}$$

An intuitive way to guess this is to use *Trotter's formula*[42]

$$e^{A+B} = \lim_{n\to\infty} \left(e^{A/n} e^{B/n}\right)^n. \tag{4.3.5}$$

Therefore, approximately

$$\left.\frac{d}{dx} \epsilon^{A+xB}\right|_{x=0} \simeq \sum_j \frac{1}{n} \left(e^{A/n} e^{B/n}\right)^{n-j} B \left(e^{A/n} e^{B/n}\right)^j. \tag{4.3.6}$$

---

[41] H Mori, Phys. Rev. **112**, 1829 (1958).

[42] This formula holds if at least A or B is a bounded operator. See J. Glimm and A. Jaffe, *Quantum Mechanics, a functional integral point of view*, second edition, Section 3.2 (Springer, 1987). The idea of Trotter's formula is very important in numerical solutions of (nonlinear) partial differential equations.

This is a Riemann sum formally converging to the above formula. A more respectable demonstration is to check the following formal relation:

$$\frac{d}{dt}\left\{e^{-t(A+xB)}\frac{d}{dx}e^{t(A+xB)} - \int_0^t d\lambda\, e^{-\lambda(A+xB)}Be^{\lambda(A+xB)}\right\} = 0. \tag{4.3.7}$$

Integrating this from $t = 0$ to 1, we obtain the desired formula.

With the aid of (4.3.6) we can Taylor expand as

$$e^{A+xB} = e^A\left(1 + x\int_0^1 d\lambda\, e^{-\lambda A}Be^{\lambda A} + \cdots\right). \tag{4.3.8}$$

From this we have a more convenient formula for statistical physics:

$$e^{-\beta(A+B)} = e^{-\beta A}\left(1 - \int_0^\beta d\lambda\, e^{\lambda A}Be^{-\lambda A} + \cdots\right). \tag{4.3.9}$$

We have already used this (→**3.3.6**). Notice that

$$e^{\beta H}Ae^{-\beta H} = A(-i\beta\hbar). \tag{4.3.10}$$

This is a pure imaginary time evolution of the observable (Recall the Heisenberg picture).

Therefore, if we wish to compute the perturbation effect on an equilibrium system whose density operator is $\rho_e \propto e^{-\beta H}$, we have

$$Tr\left(e^{-\beta(H+h)}B\right) = Tr\left\{e^{-\beta H}\left(1 - \int_0^\beta d\lambda\, h(-i\lambda\hbar)\right)B\right\} = Tre^{-\beta H}(B - \beta\langle h; B\rangle), \tag{4.3.11}$$

i.e., for an observable B whose equilibrium average vanishes,[43]

$$\langle B\rangle = -\beta\langle h; B\rangle, \tag{4.3.12}$$

where we have used the *canonical correlation* $\langle X; Y\rangle$ defined as

$$\langle X; Y\rangle \equiv k_BT\int_0^\beta d\lambda\langle X(-i\lambda\hbar)Y\rangle_e. \tag{4.3.13}$$

This was introduced in **3.3.6**.

4.3.6 Time evolution of local equilibrium state

The local equilibrium distribution $\rho_L(t)$ may be written in the following form:[44]

$$\rho_L(t) = \frac{1}{Z(t)}e^{-\beta H - \beta TF(t)B}. \tag{4.3.14}$$

If we may assume that the system is sufficiently close to equilibrium, we may use the linear approximation (4.3.9):

$$\rho_L(t) = \rho_e\left(1 - TF(t)\int_0^\beta d\lambda\, e^{\lambda H}Be^{-\lambda H} + \phi(t) + \cdots\right), \tag{4.3.15}$$

---

[43] Notice that under this condition the effect of the perturbation on the normalization constant does not appear in the final result.

[44] F is the conjugate intensive variable with respect to entropy, so we need T.

where ρ_e is the unperturbed canonical density operator $\rho_e = (1/Z)e^{-\beta H}$, and $\phi(t)$ is a numerical factor coming from the mismatch between the normalization constant $Z(t)$ for $\rho_L(t)$ and Z.

Now, we use this as the initial condition and solve the von Neumann equation:

$$\frac{\partial \rho}{\partial t} = -iL\rho = -i[H,\rho]/\hbar = \frac{1}{i\hbar}H^\times \rho. \tag{4.3.16}$$

The Hamiltonian in the above is the system Hamiltonian without any external perturbation, because we are performing a perturbative calculation.

The symbol $^\times$ implies the following: for an operator A, $A^\times B = [A,B]$. We have the following formula:

$$e^{A^\times}B = e^A B e^{-A}. \tag{4.3.17}$$

This may be shown by expanding the exponential functions.[45]

The solution we want reads at time $t+\tau$ (t is the starting time of our natural motion, and τ is a mesoscopic time)

$$\begin{aligned} \rho(t+\tau) &= \rho_e\left(1 - TF(t)\int_0^\beta d\lambda\, e^{\tau H^\times/i\hbar}e^{\lambda H}Be^{-\lambda H} + \phi(t) + \cdots\right), & (4.3.18)\\ &= \rho_e\left(1 - TF(t)\int_0^\beta d\lambda\, B(-\tau - i\lambda\hbar) + \phi(t) + \cdots\right). & (4.3.19)\end{aligned}$$

Here, note that in the von Neumann equation and in the Heisenberg's equation of motion the place of H in the commutator is different. Thus, we need $-$ in front of τ in B (cf. **2.7.10**).
Remark. Notice that the proper choice of the variables (or rather, variable pairs $\{x, X\}$ (wrt to energy) or $\{x, F\}$ (wrt entropy)) relies heavily on thermodynamics. □

4.3.7 Mori's derivation of quantum Green-Kubo relation: prototype

Let $\hat{J}(t)$ be a flux of a certain physical observable. Then, its average over $\rho(t+\tau)$ in (4.3.19) is regarded as the average over the true distribution at time t (that is, τ is macroscopically ignored).

$$J(t) = -TF(t)\int_0^\beta d\lambda\, \langle B(-\tau - i\lambda\hbar)\hat{J}\rangle_e. \tag{4.3.20}$$

This may be rewritten as (assuming $B(0) = 0$)[46]

$$\begin{aligned} J(t) &= -TF(t)\int_0^\beta d\lambda\, \langle B(-\tau - i\lambda\hbar)\hat{J}\rangle_e, & (4.3.21)\\ &= -TF(t)\int_0^\beta d\lambda \int_0^\tau ds \left\langle \frac{d}{ds}B(-s - i\lambda\hbar)\hat{J}\right\rangle_e. & (4.3.22)\end{aligned}$$

If we introduce the flux for B

$$\hat{J}_B = \frac{dB(t)}{dt}, \tag{4.3.23}$$

---

[45] Or, in a smarter way by the following 'Heisenberg's equation of motion'

$$\frac{d}{dt}e^{tA^\times} = [A, e^{tA^\times}] = A^\times e^{tA^\times}.$$

[46] The equilibrium average of J vanishes, so even $\phi(t)$ disappears from the result.

(4.3.22) reads

$$
\begin{aligned}
J(t) &= TF(t)\int_0^\beta d\lambda \int_0^\tau ds\,\left\langle \hat{J}_B(-s-i\lambda\hbar)\hat{J}\right\rangle_e, && (4.3.24)\\
&= TF(t)\int_0^\tau ds \int_0^\beta d\lambda\,\left\langle \hat{J}_B(-i\lambda\hbar)\hat{J}(s)\right\rangle_e && (4.3.25)\\
&= \frac{1}{k_B}\int_0^\tau ds\,\langle \hat{J}_B;\hat{J}(s)\rangle F(t). && (4.3.26)
\end{aligned}
$$

Since τ is the time scale longer than the microscopic relaxation time, we may assume that the correlation function decays by then, so we may set $\tau\to\infty$. Thus, we have arrived at the Green-Kubo formula for the transport coefficient for the flux J when there is a driving force F that produces the flux of B:

$$
\sigma = \frac{1}{k_B}\int_0^\infty ds\,\left\langle \hat{J}_B(0);\hat{J}(s)\right\rangle. \tag{4.3.27}
$$

If we compare this result with the formulas obtained in **4.1.14**, we see that the quantum version can simply be obtained by replacing the correlation in the classical version with the canonical correlation.

4.3.8 Quantum Green-Kubo formula for transport coefficients

Up to this point we have not assumed that the auxiliary field F to be space-dependent. Let us generalize the simple case discussed up to **4.3.7** to this case (→**4.1.14**). Actually, we do not have anything extra to do than the classical case we have already discussed (→**4.1.14**). For convenience, let us virtually repeat the argument.

To this end we first replace

$$
BF(t)\to\int d\boldsymbol{r}\sum_j F_j(\boldsymbol{r},t)\hat{A}_j(\boldsymbol{r}), \tag{4.3.28}
$$

where $\hat{A}_j$ is a microscopic description of the density of a macroscopic observable A_j, and F_j is its conjugate (with respect to entropy) intensive variable (a classical field, not an operator).

Let us compute the flux density $\hat{\boldsymbol{j}}_i(\boldsymbol{r},t)$ of $\hat{A}_i$. We must identify the counterpart of dB/dt. It must be related to the derivative of (4.3.28). If we may assume all $\hat{A}_j$ are conserved quantities, then, the conservation law

$$
\frac{\partial}{\partial t}\hat{A}_j = -\nabla\cdot\hat{\boldsymbol{j}}_j \tag{4.3.29}
$$

tells us that the substitute of dB/dt must be

$$
\frac{dB}{dt}F(t)\to-\int d\boldsymbol{r}\sum_j F_j\,\nabla\cdot\hat{\boldsymbol{j}}_j. \tag{4.3.30}
$$

Therefore, (4.3.22) reads[47]

$$
\begin{aligned}
\boldsymbol{j}_i(\boldsymbol{r},t) &= -\frac{1}{k_B}\sum_j\int d\boldsymbol{r}'\int_0^\infty ds\,\langle\nabla\cdot\hat{\boldsymbol{j}}_j(\boldsymbol{r}',0);\hat{\boldsymbol{j}}_i(\boldsymbol{r},s)\rangle F_j(\boldsymbol{r}',t), && (4.3.31)\\
&= \frac{1}{k_B}\sum_j\int d\boldsymbol{r}'\int_0^\infty ds\,\langle\hat{\boldsymbol{j}}_j(\boldsymbol{r}',0);\hat{\boldsymbol{j}}_i(\boldsymbol{r},s)\rangle \nabla F_j(\boldsymbol{r}',t). && (4.3.32)
\end{aligned}
$$

---

[47] The components are not carefully or explicitly written, but the outcome should be clear.

Therefore, we obtain the formula we can easily guess from the classical case (→**X2.3.13**, **4.1.14**):

$$\sigma_{ij} = \frac{1}{k_B}\int d\boldsymbol{r}\int_0^\infty dt\,\langle \hat{\boldsymbol{j}}_j(0,0);\hat{\boldsymbol{j}}_i(\boldsymbol{r},t)\rangle. \tag{4.3.33}$$

Again, the replacement of the correlation with the canonical correlation can be seen; it is not the replacement with the symmetrized correlation function initially guessed even by Kubo.

4.4 General Framework of Linear Response Theory

Recommended References:
R. Kubo, Rep. Prog. Phys. **29**, 255 (1966),
H. Nakano, Int. J. Mod. Phys. B **7**, 2397 (1993).[48]

R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II* (Springer, 1985) is perhaps a kinder introduction, although it is not critical. Not recommended.[49] The book covers the materials in Chapters X1 and X2 of these notes.

4.4.1 Fluctuation and Response: what we already know
We have already seen in the static case that the response of the system is governed by the fluctuation (→**2.8.1**)

$$\langle \delta x_i\rangle = \beta\langle \delta x_j \delta x_i\rangle X_j \tag{4.4.1}$$

or in the quantum case (→**3.3.6**)

$$\langle \delta x_i\rangle = \left\langle \int_0^\beta d\lambda\, e^{\lambda H}\delta x_j e^{-\lambda H}\delta x_i\right\rangle X_j = \beta\langle \delta x_j;\delta x_i\rangle X_j, \tag{4.4.2}$$

where we have used the canonical correlation (→**3.3.6**; summary in **4.3.5**). This suggests that if the perturbation X_j is time-dependent, then

$$\langle x_i(t)\rangle = \beta\int_0^t ds\,\langle \delta\dot{x}_j(s);\delta x_i(t)\rangle X_j(s). \tag{4.4.3}$$

This is true as we will see later (→**4.4.12**).[50] Also we have learned that the *response function* is given by the *fluctuation-dissipation relation* of the first kind (→**3.2.8**). This has just the form guessed above.

More generally, we have learned that the *Onsager coefficients* can be calculated in terms of the time correlation of fluctuations (→**4.1.12**). Transport coefficients can be written in terms of the correlation function of fluxes (the *Green-Kubo relation* →**4.1.14**). Also we have learned that the Green-Kubo relation can be obtained directly from the Onsager principle (→**4.1.18**).

The reader will see that all the important key concepts and results have already been discussed at least in their simplest forms in Chapter 3.

---

[48] This contains an episode of Feynman on electric conductivity.

[49] However, the first volume of this book is strongly recommended.

[50] The dot is the time derivative; its necessity is easily guessed from the dimensional reason.

4.4.2 Response function.

We believe that if the external perturbation of the system parameter or the change of the thermodynamic force $x(\boldsymbol{r},t)$,[51] which may depend on space and time coordinates, is small, the system response described by the change $G(\boldsymbol{r},t)$ of thermodynamic densities (= the extensive variables per volume)[52] should also be small, and G is linearly related to X:

$$G(\boldsymbol{r},t) = L(\{X(\boldsymbol{r},t)\}). \tag{4.4.4}$$

Here, $G(\boldsymbol{r},t)$ could depend on the values of X at the point away from $\boldsymbol{r}$ at a different time from t (in the past only; this is causality →**4.4.8**). In these notes, we assume X is the thermodynamic force (defined with respect to energy) and G is the conjugate density (extensive quantities conjugate to x per unit volume). They may be vectors or tensors.

According to the Riesz representation theorem,[53] such a linear functional can always be written in the following form

$$G(\boldsymbol{r},t) = \int d\boldsymbol{r}' \int_{-\infty}^{t} dt'\, \phi(\boldsymbol{r}-\boldsymbol{r}', t-t')X(\boldsymbol{r}',t'). \tag{4.4.5}$$

Here, we have assumed that the system is space-time uniform (no absolute position in time and space matters) for simplicity. The function ϕ is called the *response function* in general (see for a prototype **3.2.8**).

4.4.3 Admittance

The space-time Fourier transform of the response function (→**3.2.8**, **4.4.2**) is called the *admittance.*

To define this unambiguously we must specify the convention of space-time Fourier transformation. We adopt the following definition (consistent with the time Fourier transformation already introduced):

$$f(\boldsymbol{k},\omega) = \int d\boldsymbol{r} \int dt\, e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} f(\boldsymbol{r},t). \tag{4.4.6}$$

The corresponding inverse transformation reads

$$f(\boldsymbol{r},t) = \frac{1}{(2\pi)^4} \int d\boldsymbol{k} \int d\omega\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} f(\boldsymbol{k},\omega). \tag{4.4.7}$$

Notice that each harmonic component behaves as a plane wave[54]

$$e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} \tag{4.4.8}$$

with the velocity $\omega/|\boldsymbol{k}|$ in the direction of $\boldsymbol{k}$.

Thus, the *admittance* χ is defined as

$$\chi(\boldsymbol{k},\omega) = \int d\boldsymbol{r} \int dt\, e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} \phi(\boldsymbol{r},t). \tag{4.4.9}$$

---

[51] In these notes, X is used for the thermodynamic intensive parameter appearing in the ordinary Gibbs relation. The intensive variable in the entropy formula will be denoted by F. That is $dE = \cdots + Xdx + \cdots$, and $dS = \cdots + Fdx + \cdots$. This choice is different from IESM Lecture notes.

[52] Here, we define thermodynamic densities as the conjugate of x with respect to energy.

[53] e.g., K. Yosida, *Functional Analysis* (Springer, Berlin, 1965; Sixth edition 1980), p90. Here, it is a theorem on the continuous (= bounded) linear map on the Hilbert space.

[54] We choose our convention so that the plane wave can be written as $e^{ix_j k^j}$ in terms of 4-vectors and the relativistic scalar product with the usual indefinite metric $(+,-,-,-)$.

Therefore,

$$G(\boldsymbol{k},\omega)=\chi(\boldsymbol{k},\omega)X(\boldsymbol{k},\omega). \tag{4.4.10}$$

4.4.4 Real and imaginary parts of admittance

Since X and G in our case are observable quantities, they must be real, and so must be the response function. Therefore,

$$\overline{\chi(\boldsymbol{k},\omega)}=\int d\boldsymbol{r}\int dt\, e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\phi(\boldsymbol{r},t)=\chi(-\boldsymbol{k},-\omega). \tag{4.4.11}$$

We introduce the real and imaginary parts of the admittance as (the names will become clear in **4.4.5**)

$$\begin{aligned}\chi'(\boldsymbol{k},\omega) &= Re\,\chi(\boldsymbol{k},\omega) \quad \text{called } \textit{dispersive} \text{ or } \textit{storage part}, &(4.4.12)\\ \chi''(\boldsymbol{k},\omega) &= -Im\,\chi(\boldsymbol{k},\omega) \quad \text{called } \textit{dissipative} \text{ or } \textit{loss part}, &(4.4.13)\end{aligned}$$

or

$$\chi(\boldsymbol{k},\omega)=\chi'(\boldsymbol{k},\omega)-i\chi''(\boldsymbol{k},\omega). \tag{4.4.14}$$

Here, do not forget the inclusion of $-$ in the definition of χ''. This negative sign is needed to make this quantity positive definite as we will see later (→**4.4.5**).

We may write

$$\begin{aligned}\chi'(\boldsymbol{k},\omega) &= \int d\boldsymbol{r}\int dt\, \cos(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})\phi(\boldsymbol{r},t), &(4.4.15)\\ \chi''(\boldsymbol{k},\omega) &= \int d\boldsymbol{r}\int dt\, \sin(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})\phi(\boldsymbol{r},t). &(4.4.16)\end{aligned}$$

(4.4.11) now reads in terms of χ' and χ'' as

$$\begin{aligned}\chi'(\boldsymbol{k},\omega) &= \chi'(-\boldsymbol{k},-\omega), &(4.4.17)\\ \chi''(\boldsymbol{k},\omega) &= -\chi''(-\boldsymbol{k},-\omega). &(4.4.18)\end{aligned}$$

That is, the real part of the admittance is an even function, and the imaginary part an odd function.

Warning about conventions: If the definition of the space-time Fourier transformation is different, then the signs may be flipped. If

$$f(\boldsymbol{k},\omega)=\int d\boldsymbol{r}\int dt\, e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}f(\boldsymbol{r},t), \tag{4.4.19}$$

then, $\chi=\chi'+i\chi''$ is the definition of χ' and χ'' (but (4.4.12) and (X2.2.10) are intact; that is, their definitions are consistent in different conventions). For example, the convention of D. Forster, *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions* (Benjamin, 1975) and L. P. Kadanoff and P. C. Martin, "Hydrodynamic Equations and Correlation Functions," Ann. Phys. **24**, 419-469 (1963) is different from ours.

4.4.5 Dissipation and admittance

The power $p(\boldsymbol{r}, t)d\boldsymbol{r}$ supplied by the force X to the volume element $d\boldsymbol{r}$ around $\boldsymbol{r}$ of the system is the product of X and the rate of change of the conjugate density G. Here, for simplicity, let us assume that X is a scalar (for many-variable cases →**X2.2.8**). Therefore,

$$p(\boldsymbol{r}, t) = X(\boldsymbol{r}, t)\frac{\partial}{\partial t}G(\boldsymbol{r}, t). \tag{4.4.20}$$

Let us compute the time average of the power for the following special 'monochromatic' perturbation

$$X(\boldsymbol{r}, t) = Re\left(e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\tilde{X}\right). \tag{4.4.21}$$

Here, $\tilde{X}$ is called the *phasor representation* of the monochromatic force X (the definition is the same as in the AC circuit theory). $\tilde{X}$ may be a complex number due to the phase.

The time average per unit volume $\langle p\rangle$ of the power $p(\boldsymbol{r}, t)$ is, as shown in **4.4.6**, given by

$$\langle p\rangle = \frac{1}{2}\omega\chi''(\boldsymbol{k}, \omega)|\tilde{X}|^2. \tag{4.4.22}$$

In the steady state, the power supplied by the perturbation of X must be dissipated as heat to the heat bath. Therefore, $\langle p\rangle$ must be the energy dissipation rate by the system per unit volume. Therefore, the second law implies that χ'' must be nonnegative definite, or more generally,

$$\omega\chi'' \geq 0. \tag{4.4.23}$$

4.4.6 Computation of average power loss

Postponing a smart derivation to **4.4.7**, let us compute $p(t)$ according to its definition (4.4.20). Since

$$G(\boldsymbol{r}, t) = \frac{1}{2}\left[\chi(\boldsymbol{k}, \omega)e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\tilde{X} + \chi(-\boldsymbol{k}, -\omega)e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\tilde{X}^*\right] \tag{4.4.24}$$

according to (4.4.5), we can obtain its time derivative easily. The time averaged power per unit volume $\langle p\rangle$ may be computed as

$$\langle p\rangle = \frac{1}{V}\frac{\omega}{2\pi}\int_0^{2\pi/\omega} dt \int_V d\boldsymbol{r}\, p(\boldsymbol{r}, t). \tag{4.4.25}$$

Here, V is the system volume.[55]

Putting (4.4.24) and the formula for X into (4.4.25), we obtain

$$\langle p\rangle = \frac{1}{4}\left[i\omega\chi(\boldsymbol{k}, \omega) - i\omega\chi(-\boldsymbol{k}, -\omega)\right]|\tilde{X}|^2. \tag{4.4.26}$$

This is what we wished to demonstrate (do not forget χ'' is defined with the negative sign in (4.4.14)).

---

[55] We take the $V \to \infty$ limit after all the calculations are over.

4.4.7 Phasor representation and dissipation

Let the plane wave be given as

$$A(t) = Re\left(e^{i\omega t}\tilde{A}\right). \tag{4.4.27}$$

We call the complex number $\tilde{A}$ the *phasor representation* of A. We have introduced the phasor representation of the thermodynamic force $\tilde{X}$ in **4.4.5**. The time average of the product of two plane waves in the form of (4.4.27) is given by (the overline implies time average)

$$\overline{AB} = \frac{1}{2} Re\, \tilde{A}^* \tilde{B}. \tag{4.4.28}$$

The space-time average is just a simple generalization of the above result.

The perturbation result G has the following phasor representation as clearly seen from (4.4.24):

$$\tilde{G} = \chi(\boldsymbol{k}, \omega)\tilde{X}. \tag{4.4.29}$$

Therefore, (the overline implies time and space average; not complex conjugate)

$$\langle p \rangle = \overline{X \frac{\partial G}{\partial t}} = \frac{1}{2} Re\, \tilde{X}^*(i\omega\tilde{G}) = \frac{1}{2} Re\left(i\omega\chi(\boldsymbol{k}, \omega)|\tilde{X}|^2\right). \tag{4.4.30}$$

This is exactly (4.4.22).

Remark. The many variable case is discussed in **X2.2.8**.

4.4.8 Causality and analyticity of admittance

Although the perturbation result $G(\boldsymbol{r}, t)$ may depend on $X(\boldsymbol{r}, s)$ at time s different from t, it cannot depend on X for $s > t$. This is *causality.*

Therefore, in terms of the response function $\phi(\boldsymbol{r}, t)$ (→**4.4.2**), the causality can be written as

$$\phi(\boldsymbol{r}, t) = 0 \text{ if } t < 0. \tag{4.4.31}$$

This may be written in terms of the admittance as

$$0 = \int d\boldsymbol{k} \int d\omega\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\chi(\boldsymbol{k}, \omega), \tag{4.4.32}$$

if $t < 0$.

For $t < 0$ $e^{i\omega t}$ decays exponentially for $|\omega| \to \infty$ if $Im\, \omega < 0$ (that is, on the lower half complex ω-plane). Therefore, (4.4.32) implies that

$$\oint_C d\omega\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\chi(\boldsymbol{k}, \omega) = 0. \tag{4.4.33}$$

Since the exponential factor is analytic[56] in the lower half plane, this implies that χ must be analytic in the lower half complex ω plane.[57] Conversely, if χ is analytic on the lower half complex ω-plane, then the causality holds.

In short:

<u>Causality $\Longleftrightarrow$ Analyticity of admittance on the lower half complex ω-plane.</u>[58]

---

[56] More precisely, we should use the term, *holomorphic.*

[57] This is not a trivial statement. This depends on the definition of the Fourier transformation. It is the lower half plane with our convention.

[58] This was first recognized by H. Takahashi (1942).

4.4.9 Kramers-Kronig relation — dispersion relation

Since the admittance is analytic on the lower half ω-plane, Cauchy's theorem tells us

$$0 = \oint_C \frac{\chi(\boldsymbol{k}, z)}{z - \omega} dz, \tag{4.4.34}$$

where the integration contour C is the infinite semicircle on the lower half plane with the straight line portion coinciding with the real axis. If ω is real, then there is a simple pole on the contour, so we may use the Plemelij formula (→**4.4.10**) to obtain

$$P \int_{-\infty}^{\infty} \frac{\chi(\boldsymbol{k}, z)}{z - \omega} dz + i\pi\chi(\boldsymbol{k}, \omega) = 0. \tag{4.4.35}$$

That is,

$$\chi(\boldsymbol{k}, \omega) = \frac{i}{\pi} P \int_{-\infty}^{\infty} \frac{\chi(\boldsymbol{k}, z)}{z - \omega} dz. \tag{4.4.36}$$

Now, we split this into real and imaginary parts of χ.

$$\chi'(\boldsymbol{k}, \omega) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\chi''(\boldsymbol{k}, x)}{x - \omega} dx, \tag{4.4.37}$$

$$\chi''(\boldsymbol{k}, \omega) = -\frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\chi'(\boldsymbol{k}, x)}{x - \omega} dx. \tag{4.4.38}$$

Here, P denotes the *Cauchy principal value*. These relations are known as the *Kramers-Kronig relation* (*dispersion relation*) to physicists.[59] As we will see soon, the dissipative part of χ is experimentally easier to access (that is, χ'' is easier to observe), so these relations may be of practical importance to obtain χ experimentally.[60]

4.4.10 Plemelij formula

The following formula is called the *Plemelij formula*

$$\frac{1}{x \pm i\epsilon} = P\frac{1}{x} \mp \delta(x), \tag{4.4.39}$$

where ϵ is an infinitesimal positive number. The equality holds inside the integral with respect to x.[61]

There are several ways to demonstrate this. Perhaps the most elementary one is to use the complex function theory. A simple pole on the integral path contributes one half of its residue to the integral, and the remaining integral becomes the Cauchy principal value.

4.4.11 Spectral representation

Let us reconstruct χ from its imaginary part χ'' with the aid of the Kramers-Kronig relation **4.4.9**:

$$\chi(\boldsymbol{k}, \omega) = \frac{1}{\pi} P \int_{-\infty}^{\infty} dx \, \frac{\chi''(\boldsymbol{k}, x)}{x - \omega} - i\chi''(\boldsymbol{k}, \omega). \tag{4.4.40}$$

---

[59] Well known formulas for the Hilbert transformation.

[60] However, to use these formulas the frequency range must be sufficiently widely observed.

[61] so, the limit $\epsilon \to +0$ is the weak limit.

With the aid of the Plemelij formula (→**4.4.10**), this can be rewritten as

$$\chi(\boldsymbol{k},\omega) = \lim_{\epsilon\to+0}\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\chi''(\boldsymbol{k},x)}{x-(\omega-i\epsilon)}. \tag{4.4.41}$$

Let us define the analytical continuation of χ'' that is analytic on the lower half ω-plane to the upper half ω-plane as

$$\hat{\chi}(\boldsymbol{k},z) = \frac{1}{\pi}\int_{-\infty}^{\infty}dx\frac{\chi''(\boldsymbol{k},x)}{x-z}. \tag{4.4.42}$$

This has no singularity if $Im\, z \neq 0$. That is, the real axis is a branch cut. (4.4.43) implies that

$$\chi(\boldsymbol{k},\omega) = \lim_{\epsilon\to+0} = \hat{\chi}(\boldsymbol{k},\omega-i\epsilon). \tag{4.4.43}$$

That is, the continuation from the lower half ω-plane gives the physically meaningful value. Therefore, we call the lower half ω-plane as the *physical sheet*. Its analytical continuation beyond the real axis gives the Riemann surface that contains the physically observable values for real z.

The representation of χ in terms of χ'' is called the *spectral representation.*

4.4.12 Response function

Let us consider a system under a perturbation that perturb $\hat{A}$ directly. For simplicity, we assume that the system is spatially uniform. Let us assume the perturbative Hamiltonian to be

$$h = -X(t)\hat{A}. \tag{4.4.44}$$

We can perturbatively solve the von Neumann equation to obtain the perturbed density operator as

$$\rho(t) = \rho_e\left[1+\int_{-\infty}^{0}du\, X(t+u)\int_0^{\beta}d\lambda\, \dot{\hat{A}}(u-i\lambda\hbar)\right]. \tag{4.4.45}$$

Here, we set $-TF = X$. The average of another observable B is given by (let us assume, for simplicity, that its average is zero in equilibrium (without the perturbation))

$$\begin{aligned}\langle B\rangle(t) &= \int_{-\infty}^{0}du\, X(t+u)\int_0^{\beta}d\lambda\, \langle\dot{\hat{A}}(u-i\lambda\hbar)B\rangle_e, & (4.4.46)\\ &= \int_{-\infty}^{t}du\, X(s)\beta\langle\dot{\hat{A}}(s-t);B\rangle_e. & (4.4.47)\end{aligned}$$

Therefore, the response function ϕ_{BA} reads

$$\phi_{BA}(t) = \beta\langle\dot{\hat{A}};B(t)\rangle_e. \tag{4.4.48}$$

This can be further transformed to our familiar form in terms of the commutator (→**X2.4.8**) with the aid of *Kubo's identity*

$$[\rho_e,\hat{A}] = i\hbar\int_0^{\beta}d\lambda\,\rho_e\dot{\hat{A}}(-i\lambda\hbar). \tag{4.4.49}$$

To show (4.4.53) we start with[62]

$$\frac{d}{d\beta}[\hat{A}, \rho_e] = -i\hbar \dot{\hat{A}} \rho_e - H[\hat{A}, \rho_e]. \tag{4.4.50}$$

Now, let us write

$$[\hat{A}, \rho_e] = \rho_e C(\beta). \tag{4.4.51}$$

Put this into (4.4.50), and we obtain

$$\dot{C}(\beta) = -i\hbar \hat{\dot{A}}(-i\beta\hbar). \tag{4.4.52}$$

This is

$$[\rho_e, \hat{A}] = i\hbar \int_0^\beta d\lambda \, \rho_e \dot{\hat{A}}(-i\lambda\hbar) \tag{4.4.53}$$

This is called *Kubo's identity.*

Now, we can rewrite (4.4.48) as

$$\begin{aligned} \phi_{BA}(t) &= \frac{1}{i\hbar} Tr\left([\rho_e, \hat{A}(-t)]B\right), \qquad (4.4.54)\\ &= \frac{1}{i\hbar} \int_0^\beta d\lambda \, Tr(\rho_e[\hat{A}(-t), B]). \qquad (4.4.55) \end{aligned}$$

Therefore, we have arrived at

$$\phi_{B\hat{A}}(t) = \frac{i}{\hbar} Tr \rho_e [B(t), \hat{A}]. \tag{4.4.56}$$

It should be almost obvious that

$$\phi_{B\hat{A}}(\boldsymbol{r}, t) = \frac{i}{\hbar} Tr \rho_e [B(\boldsymbol{r}, t), \hat{A}(0, 0)]. \tag{4.4.57}$$

for the spatially nonuniform case.

4.4.13 Canonical correlation function: summary

We have already introduced the canonical correlation function (→**4.3.5**)[63]

$$\langle X; Y \rangle = k_B T \int_0^\beta d\lambda \, Tr(\rho_e X(-i\lambda\hbar) Y). \tag{4.4.58}$$

In terms of this the response function may be written, as we know well by now (→**4.4.12**)

$$\phi_{B\hat{A}}(t) = \beta \langle \hat{\dot{A}}(0); B(t) \rangle = -\beta \langle \hat{A}; \dot{B}(t) \rangle. \tag{4.4.59}$$

We have already seen how natural the canonical correlation function is in various contexts.[64] Let us summarize its properties.

---

[62] This can be obtained by a straightforward differentiation and Heisenberg's equation of motion.

[63] Here, for simplicity, we discuss only the cases with X and Y being self-adjoint.

[64] It is the retarded two time Green's function in the field theoretical approach to many body theory.

[A] $\langle X;Y\rangle$ is a scalar product. That is,
  (i) $\overline{\langle X;Y\rangle} = \langle Y;X\rangle$,
  (ii) bilinear,
  (iii) $\langle X;X\rangle \geq 0$ and $X=0$ iff $\langle X;X\rangle = 0$.
[B] Stationarity: $\langle X(t);Y(t+s)\rangle = \langle X;Y(s)\rangle$.
[C] $\langle X;Y\rangle$ is real.[65]
[D] Symmetry: $\langle X;Y\rangle = \langle Y;X\rangle$.
[E] Let L be the Liouvillian (→**3.3.5**) of the natural motion. Then

$$\langle LX;Y\rangle = \langle X;LY\rangle. \tag{4.4.60}$$

[F] Reciprocity:

$$\langle X;Y(t)\rangle_{\boldsymbol{H},\boldsymbol{\omega}} = \epsilon_X\epsilon_Y\langle Y;X(t)\rangle_{-\boldsymbol{H},-\boldsymbol{\omega}}. \tag{4.4.61}$$

Here, $\boldsymbol{H}$ is the magnetic field and $\boldsymbol{\omega}$ is the rotation angular velocity.

Using [F] and (4.4.59), we conclude

$$\phi_{BA}(t)_{\boldsymbol{H},\boldsymbol{\omega}} = \epsilon_A\epsilon_B\phi_{AB}(t)_{-\boldsymbol{H},-\boldsymbol{\omega}}. \tag{4.4.62}$$

It is a good exercise to demonstrate [A] - [F].

4.4.14 Symmetrized correlation functions

We know that replacing the correlation function with the canonical correlation function, we can generalize the classical results to their quantum counterparts. However, initially it was thought that symmetrization of correlations $\langle AB\rangle \to \langle AB+BA\rangle/2$ was the right procedure for this generalization. Let us define the *symmetrized correlation function*

$$C(X;Y) \equiv \frac{1}{2}Tr\rho_e[XY+YX]. \tag{4.4.63}$$

Its salient features may be summarized as follows:
[A] $C(X;Y)$ is a scalar product:
  (i) $\overline{C(X;Y)} = C(Y;X)$,
  (ii) bilinear,
  (iii) $C(X;X) \geq 0$ and $X=0$ iff $C(X;X)=0$.
[B] Stationarity: $C(X(t);Y(t+s)) = C(X;Y(s))$.
[C] $C(X;Y)$ is real.[66]
[D] Symmetry: $C(X;Y) = C(Y;X)$.
[E] Let L be the Liouvillian (→) of the natural motion. Then

$$C(LX;Y) = C(X;LY). \tag{4.4.64}$$

[F] Reciprocity:

$$C(X;Y(t))_{\boldsymbol{H},\boldsymbol{\omega}} = \epsilon_X\epsilon_Y C(Y;X(t))_{-\boldsymbol{H},-\boldsymbol{\omega}}. \tag{4.4.65}$$

Here, $\boldsymbol{H}$ is the magnetic field and $\boldsymbol{\omega}$ is the rotation angular velocity.

Thus, there is a very good parallelism between the canonical and symmetrized correlation functions. Their precise relation is the fluctuation dissipation relation originally obtained by Callen and Welton.[67]

---

[65] if X and Y are self-adjoint, but here we assume this.

[66] if X and Y are self-adjoint, but throughout this section we assume dynamical operators are all self-adjoint.

[67] H. B. Callen and T. A. Welton, Phys. Rev. **83**, 34 (1951).

4.4.15 Relation between canonical and symmetrized correlations

The canonical correlation is closely related to responses, and (symmetrized) correlations are more directly related to observable quantities such as scattering functions and power spectra. Therefore, it should be very convenient, if we have a direct relation between them. Actually, this direct relation is the essence of fluctuation dissipation relations.

The relation we seek is

$$\frac{\beta\hbar\omega}{2}\coth\frac{\beta\hbar\omega}{2}\int dt\,\langle X_{-\boldsymbol{k}}(0);Y_{\boldsymbol{k}}(t)\rangle e^{-i\omega t} = \int dt\,C(X_{-\boldsymbol{k}};Y_{\boldsymbol{k}}(t))e^{-i\omega t}. \tag{4.4.66}$$

Let us start from the canonical correlation: Since $Y\rho_e = \rho_e Y(-i\beta)$ (→(**??**))

$$\begin{aligned}\int dt\,e^{-i\omega t}Tr(\rho_e X_{-\boldsymbol{k}}(0)Y_{\boldsymbol{k}}(t)) &= \int dt\,e^{-i\omega t}Tr(X_{-\boldsymbol{k}}(0)\rho_e Y_{\boldsymbol{k}}(t-\beta i\hbar)), &(4.4.67)\\ &= \int dt\,e^{-i\omega t}Tr(\rho_e Y_{\boldsymbol{k}}(t-i\beta\hbar)X_{-\boldsymbol{k}}(0)). &(4.4.68)\end{aligned}$$

Now, we assume the KMS condition (→**X2.4.12**):

$$\begin{aligned}&= \int dt\,e^{-i\omega(t+i\beta\hbar)}Tr[\rho_e Y_{\boldsymbol{k}}(t)X_{-\boldsymbol{k}}(0)), &(4.4.69)\\ &= e^{\beta\hbar\omega}\int dt\,e^{-i\omega t}\langle Y_{\boldsymbol{k}}(t)X_{-\boldsymbol{k}}(0)\rangle_e. &(4.4.70)\end{aligned}$$

Therefore, with the aid of symmetry, from this we obtain

$$\int dt\,e^{-i\omega t}C(X_{-\boldsymbol{k}};Y_{\boldsymbol{k}}(t)) = \frac{1}{2}(1+e^{\beta\hbar\omega})\int dt\,e^{-i\omega t}\langle X_{-\boldsymbol{k}}Y_{\boldsymbol{k}}(t)\rangle_e. \tag{4.4.71}$$

On the other hand, we have

$$\int dt\,e^{-i\omega t}Tr(\rho_e[X_{-\boldsymbol{k}},Y_{\boldsymbol{k}}(t)]) = (1-e^{\beta\hbar\omega})\int dt\,e^{-i\omega t}\langle X_{-\boldsymbol{k}}Y_{\boldsymbol{k}}(t)\rangle_e. \tag{4.4.72}$$

Now, we use the two expressions of the response function (→(4.4.56), (4.4.59))

$$\frac{i}{\hbar}Tr(\rho_e[X_{-\boldsymbol{k}},Y_{\boldsymbol{k}}(t)]) = -\beta\langle X_{-\boldsymbol{k}};\dot{Y}_{\boldsymbol{k}}(t)\rangle. \tag{4.4.73}$$

Therefore, (4.4.72) reads

$$\int dt\,e^{-i\omega t}\langle X_{-\boldsymbol{k}};\dot{Y}_{\boldsymbol{k}}(t)\rangle = \frac{i}{\beta\hbar}(1-e^{\beta\hbar\omega})\int dt\,e^{-i\omega t}\langle X_{-\boldsymbol{k}}Y_{\boldsymbol{k}}(t)\rangle_e. \tag{4.4.74}$$

Combining this and (4.4.71), we obtain the desired relation (X2.5.40).

4.4.16 How to calculate response functions

The typical examples can be found in **X2.6** (Kubo-Tomita's spectral line shape of NMR) and in **X2.7** (Nakano's electric conductivity formula; that is, the famous formula erroneously called the Kubo formula).

Chapter 5

Hydrodynamics

Standard references for this topic are

D. Forseter, *Hydrodynamic Fluctuations Broken Symmetry and Correlation Functions* (Addison-Wesley, 1989).
H. Spohn, *Large Scale Dynamics of Interacting Particles* (Springer-Verlag, 1991).

5.1 Review of Hydrodynamics

5.1.1 What is hydrodynamic description?

The reader may feel that 'hydrodynamics' is a word used as an old name for fluid dynamics. In statistical mechanics, this word is used to denote a level of description of macroscopic dynamics in terms of quantities averaged over the macroscopic volume elements. Indeed, the usual fluid dynamics is in this sense a kind of hydrodynamic description.

If a macroscopic system is away from critical point (or phase transition point), the atomic scale (the correlation scale) is microscopic (say, ∼1nm). On the other hand macroscopic observable, say, the fluid velocity field, is at our space-time scale. We have already known that between these two levels, atomic and hydrodynamic levels (= micro and macro scales), is the level described in terms of coarse-grained distribution functions or in terms of stochastic processes (the mesoscopic scales). This is called the *kinetic level.* In the fluid system Langevin equations (→**3.2.2**) are the tool to describe this level, and for gases, the Boltzmann equation (→**2.4.7**). It is a traditional wisdom of nonequilibrium statistical mechanics that there are these three levels of describing nonequilibrium systems as has been stressed already.

If the correlation length is too large, we must use raw mechanics. If this length is very small, then the law of large numbers (→**1.5.4**) tells us the average is everything. When we cannot ignore the scale of correlation, we must respect fluctuations. Therefore, we need large deviation theory (→**1.6.2**).

5.1.2 Why only three levels?

Why is it usually said that there are only three distinct levels to describe nonequilibrium systems?

In equilibrium, there are three levels of descriptions, microscopic (mechanical or kinetic), mesoscopic (fluctuational), and macroscopic (thermodynamic). The length scales characterizing these levels are the size of molecules or atoms a, the correlation length ξ, and our scale (perhaps larger than 1 μm). At our scale the world is not homogeneous, but in equilibrium we regard the world consists of piecewise homogeneous equilibrium macroscopic systems. Thus, there are only three crucial levels that may be clearly distinguished.

It is natural to expect three description levels for a nonequilibrium system near equilibrium. There may be a larger scale structure (e.g., the Karman vortex train) produced by a large scale fluid dynamics, but if we do not pay attention to such large scales, then it is plausible that a system that may be described by a local equilibrium state has only three levels of description.

Therefore, the answer to the question 'why three?' may be that if we can describe a system with a local phenomenological framework (e.g., local equilibrium), then there is only one length scale ξ between the macroscopic and the microscopic descriptions. If the system is driven by an external perturbation sufficiently away from equilibrium, then perhaps new dynamical scales (dynamical correlation length) appear between the microscopic and macroscopic levels. In this case the materials properties would be different from the ones we know. However, it seems such drastic nonequilibrium processes are rare.[1]

5.1.3 What is hydrodynamic limit?

Let us introduce the scaling parameter ϵ

$$\epsilon = \frac{\text{mesoscopic length scale}}{\text{macroscopic length scale}}, \tag{5.1.1}$$

where the mesoscopic length scale is of the order of the correlation length ξ (usually a few to a few tens of molecular diameters), and the macroscopic scale is the 'size of the volume element,' i.e., the scale where the macroscopic observable does not change appreciably.[2]

The limit $\epsilon \to 0$ is called the *hydrodynamic limit*. That is, in this limit we consider a macroscopic system whose macroscopic observables[3] change only gently with space.

If $\epsilon \ll 1$, then locally around a point $\boldsymbol{x}$ (in the volume element around this point), the system looks to be in an equilibrium state with the macroscopic (thermodynamic) parameters sampled in the volume element. That is, the microscopic distribution function in this volume element must be close to a local equilibrium distribution (→**2.4.9**).

The law of large numbers (→**1.5.4**) tells us that the averaged observables in the volume element tend to lose fluctuations, and should behave as deterministic variables.

---

[1] Perhaps because the equilibrium fluctuation is sufficiently fast from our macroscopic standard.

[2] Spohn defines as

$$\epsilon = \frac{\text{variation of average density over one interparticle distance}}{\text{average density}}$$

[3] Macroscopic observables are the average of microscopic mechanical observables over the scale of ξ/ϵ.

5.1.4 Local equilibrium distribution: explicit form

Let us consider a classical particle system whose Hamiltonian may be written as (as in **2.7.9**)

$$H = \sum_i \frac{m\boldsymbol{v}_i^2}{2} + \frac{1}{2}\sum_{i \neq j} \phi(|\boldsymbol{x}_i - \boldsymbol{x}_j|). \tag{5.1.2}$$

The local equilibrium distribution function around $\boldsymbol{x}$ may be written as

$$P \propto \exp\left(-\int \beta(\boldsymbol{x})u(\boldsymbol{x})d^3\boldsymbol{x}\right), \tag{5.1.3}$$

where β is the 'local temperature,' and u is the internal energy density defined as

$$u(\boldsymbol{x}) = \sum_i \left(\frac{1}{2}m(\boldsymbol{v}_i - \boldsymbol{v}(\boldsymbol{x}))^2 + \frac{1}{2}\sum_{j(\neq i)} \phi(\boldsymbol{r}_i - \boldsymbol{r}_j)\right)\delta(\boldsymbol{r}_i - \boldsymbol{x}), \tag{5.1.4}$$

where $\boldsymbol{v}(\boldsymbol{x})$ is the local velocity of the volume element, and ϕ is the two-body interaction potential.

Notice that the behavior of $\beta(\boldsymbol{x})$ and that of $u(\boldsymbol{x})$ are diametrically different. The former does not change appreciably within the volume element, but the latter is a wildly varying function of position $\boldsymbol{x}$. Therefore, if we regard $\boldsymbol{x}$ as a macroscopic coordinate, then a very small change of $|\boldsymbol{x}|/\epsilon \sim 1$ would significantly alter the local energy density u. If we wish to express this property explicitly, it is better to write $u(\boldsymbol{x}) \to u(\boldsymbol{x}/\epsilon)$, etc. Thus, the local equilibrium distribution (5.1.3) may be rewritten as

$$P \propto \exp\left(-\int \beta(\boldsymbol{x})u(\boldsymbol{x}/\epsilon)d^3\boldsymbol{x}\right), \tag{5.1.5}$$

Let us choose a volume Λ_ϵ centered at $\boldsymbol{x}$ such that in the $\epsilon \to 0$ limit $|\Lambda_\epsilon| \to \infty$ but $\epsilon^3\Lambda_\epsilon \to 0$. That is, from the macroscopic scale it is tiny enough, but is big enough from the microscopic point of view. Then, the hydrodynamic observables around $\boldsymbol{x}$ is defined as (cf. **4.3.3**)

$$\overline{\varphi}(\boldsymbol{x}) = \lim_{\epsilon \to 0} \frac{1}{|\Lambda_\epsilon|} \int_{\Lambda_\epsilon(\boldsymbol{x})} d\boldsymbol{x}' \varphi(\boldsymbol{x}'/\epsilon). \tag{5.1.6}$$

Thanks to the law of large numbers, this is not a stochastic variable.

Henceforth, we will not explicitly scale the microscopic variable.

5.1.5 Local form of conservation laws and balance equation

Suppose φ is a hydrodynamic observable defined as (5.1.6) (we omit the overline for simplicity). If it is a density of a conserved quantity, then

$$\frac{\partial}{\partial t}\varphi = -div\, \boldsymbol{j}_\varphi \tag{5.1.7}$$

must hold, where $\boldsymbol{j}_\varphi$ is the *flux* of φ. This equation is called the *conservation equation.*

If the quantity whose density is φ may not be conserved but could be produced, then we introduce the *source* σ_φ, and write

$$\frac{\partial}{\partial t}\varphi = -div\, \boldsymbol{j}_\varphi + \sigma_\varphi. \tag{5.1.8}$$

This is called the *balance equation.* The source term is the production rate per unit volume.

5.1.6 Microscopic expression of conserved densities

To describe a fluid, we pay attention to conserved quantities, the particle number, momentum, and energy.[4]

The *particle number density* n_0 has the following expression

$$n_0(\boldsymbol{x}) = \sum_i \delta(\boldsymbol{x}_i - \boldsymbol{x}), \tag{5.1.9}$$

where $\boldsymbol{x}_i$ is the position vector of the i-th particle.

The *momentum density* n_α ($\alpha = 1, 2, 3$, 1 corresponds to the x-component, 2 to the y, and 3 the z-component.)

$$n_\alpha(\boldsymbol{x}) = \sum_i p_{\alpha i} \delta(\boldsymbol{x}_i - \boldsymbol{x}), \tag{5.1.10}$$

where $p_{\alpha i}$ is the α-component of the momentum of the i-th particle. The *total energy density* n_4 is defined as

$$n_4(\boldsymbol{x}) = \sum_i \left(\frac{p_i^2}{2m} + \frac{1}{2} \sum_{j(\neq i)} \phi(\boldsymbol{x}_i - \boldsymbol{x}_j) \right) \delta(\boldsymbol{x}_i - \boldsymbol{x}). \tag{5.1.11}$$

5.1.7 Microscopic expression of flux

We expect the macroscopic conservation law in **5.1.5** is the averaged version of microscopic counterpart. Let us compute the time derivative of the densities.

$$\begin{aligned} \frac{\partial}{\partial t} n_0 &= \sum_i \left(-\nabla_{\boldsymbol{x}} \delta(\boldsymbol{x}_i - \boldsymbol{x}) \right) \frac{d\boldsymbol{x}_i}{dt}, & (5.1.12) \\ &= -div \left(\sum_i \frac{\boldsymbol{p}_i}{m} \delta(\boldsymbol{x}_i - \boldsymbol{x}) \right), & (5.1.13) \\ &= -div\, \boldsymbol{j}_0, & (5.1.14) \end{aligned}$$

where

$$\boldsymbol{j}_0 = \frac{1}{m} \sum_i \boldsymbol{p}_i \delta(\boldsymbol{x}_i - \boldsymbol{x}), \tag{5.1.15}$$

which is the momentum density/m. That is, the particle number flux is the velocity field.

More complicated but analogous computations give a microscopic expression of conservation of momentum

$$\frac{\partial n_\alpha}{\partial t} = -div\, \boldsymbol{j}_\alpha, \tag{5.1.16}$$

where the β component of $\boldsymbol{j}_\alpha$ is given by

$$\begin{aligned} j_{\alpha\beta} &= \sum_i \frac{1}{m} p_{i\alpha} p_{i\beta} \delta(\boldsymbol{x}_i - \boldsymbol{x}) \\ &\quad + \frac{1}{2} F_\alpha(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_{i\beta} - \boldsymbol{x}_{j\beta}) \int_0^1 d\lambda\, \delta(\lambda \boldsymbol{x}_i + (1-\lambda)\boldsymbol{x}_j - \boldsymbol{x}), & (5.1.17) \end{aligned}$$

---

[4] The reason why we pay attention to these quantities is discussed in **5.2.4**.

where $\boldsymbol{F} = -\nabla\phi$.

Also for the energy density, we have

$$\frac{\partial}{\partial t} n_4 = -div\, \boldsymbol{j}_4, \tag{5.1.18}$$

where

$$\begin{aligned} \boldsymbol{j}_4 &= \sum_i \frac{1}{m}\boldsymbol{p}_i \left(\frac{\boldsymbol{p}_i^2}{2m} + \frac{1}{2}\sum_{J(\neq i)} \phi(\boldsymbol{x}_i - \boldsymbol{x}_j) \right) \delta(\boldsymbol{x}_i - \boldsymbol{x}) \\ &\quad + \frac{1}{2}\sum_{i,j} \frac{1}{2m} \Big[(\boldsymbol{p}_i + \boldsymbol{p}_j) \cdot \boldsymbol{F}(\boldsymbol{x}_i - \boldsymbol{x}_j) \Big] (\boldsymbol{x}_i - \boldsymbol{x}_j) \int_0^1 d\lambda\, \delta(\lambda \boldsymbol{x}_i + (1-\lambda)\boldsymbol{x}_j - \boldsymbol{x}). \end{aligned} \tag{5.1.19}$$

5.1.8 Local equilibrium state and local distribution function

We have noted that the local equilibrium state is the state that can be uniquely specified by the thermodynamic fields (→**4.2.2**), but the state is not perfectly described by the local equilibrium distribution function. The difference between the true distribution function for the local equilibrium state and the local equilibrium distribution function should be of order ϵ, because the difference is related to the gradient of the thermodynamic fields. Therefore, if we ignore the $O[\epsilon]$ corrections, we can use the local equilibrium distribution to average the microscopic quantities to obtain hydrodynamic descriptions.

Let us write this average as $\langle\ \rangle_0$. Let the microscopic expression of the density of an observable to be $\varphi(\boldsymbol{x}, t)$. Precisely speaking, for the above procedure to be consistent, the average $\langle\varphi(\boldsymbol{x}, t)\rangle_0$ over the initial condition with respect to the local equilibrium distribution at time $t = 0$ must be equal (to $O[\epsilon]$) to the average of $\varphi(\boldsymbol{x})$ with respect to the local distribution function at time t (cf. **2.7.8**):

$$\langle\varphi(\boldsymbol{x}, t)\rangle_0 = \langle\varphi(\boldsymbol{x})\rangle_t + O[\epsilon]. \tag{5.1.20}$$

Here, $\langle\ \rangle_t$ is the average with respect to the local equilibrium distribution at time t:

$$P \propto \exp\left(-\int \beta(\boldsymbol{x}, t) u(\boldsymbol{x}) d^3\boldsymbol{x} \right), \tag{5.1.21}$$

Notice that the time dependence of thermodynamic fields is slow; we must observe the system for sufficiently long time to detect the time dependence.

5.1.9 Euler's equation

If we average the conservation laws with respect to the local equilibrium distribution, we obtain *Euler's equation* (= the conservation equation for the momentum) and associating equations for the number density and energy. By construction, the resultant macroscopic equations are without dissipation (describe only advection).

The resultant set of equations is:

$$\frac{\partial}{\partial t}\rho = -div(\rho \boldsymbol{v}), \tag{5.1.22}$$

$$\frac{\partial}{\partial t} m\rho v_\alpha = -div(m\rho v_\alpha \boldsymbol{v}) - grad\, p, \tag{5.1.23}$$

$$\frac{\partial}{\partial t} e = -div\{(e+p)\boldsymbol{v}\}. \tag{5.1.24}$$

Here, the pressure p is given as a a function of u and ρ (the equilibrium equation of state is used, consistent with the local equilibrium assumption), and $e = u + \rho \boldsymbol{v}^2/2$, the total energy density.

5.1.10 Origin of irreversibility

We see clearly that to have irreversible effects in fluids, we must take into account the order ϵ corrections to the local equilibrium distribution (cf **4.3.3**). Then, the equation for the conserved quantities should have the following form

$$\frac{\partial}{\partial t}\varphi = -div[\text{reversible flux} + \epsilon(\text{irreversible flux})]. \tag{5.1.25}$$

The time scale of $1/\epsilon$ describes the reversible hydrodynamic flux. Therefore, the above equation implies that the dissipation sets in in the time scale of $1/\epsilon^2$.

5.1.11 Navier-Stokes equation

To augment Euler's equation with the dissipative fluxes (the order ϵ terms, at the phenomenological level), we use the linear irreversible laws. That is, we adopt linear irreversible thermodynamics (→**4.2.2**) in the local form. The resultant equation is called the *Navier-Stokes equation* (→**5.2.5**).[5]

Thus, it is clear that the supposedly fundamental equation of fluid dynamics is not derived, but proposed as a plausible model of fluid motion at our scale. The present theoretical status of the Navier-Stokes equation may be summarized as follows:

Microscopic derivation. As just said, there is no general derivation. There is a very serious attempt for a system with some stochasticity (i.e., not pure classical mechanics) by Yau and coworkers. A summary review is available.[6] The Navier-Stokes equation is derived from the Boltzmann equation, but this is unsatisfactory, because it is only for dilute gases even if the starting point is reliable, which is not justifiable beyond microscopic time scales (→**2.4.13**).

Empirical check. We could solve the Navier-Stokes equation numerically, and compare the result with the actual flow experiments. For slow flows numerical analysis is reliable, and we may safely conclude that empirically the Navier-Stokes equation is reliable for slow flows (for small Reynolds numbers). However, for larger Reynolds numbers, there is no guarantee that the numerical schemes actually solve the Navier-Stoke equation, so even if the numerical results agree with experiments (and such is often the case), this does not mean the correctness of the original Navier-Stokes equation.

Mathematical studies. If the Reynolds number is sufficiently small, then the unique existence of the solution to the Cauchy problem is proven. However, generally speaking, the existence of the solution is still questionable for the 3D Navier-Stokes equation.[7]

In short, the Navier-Stokes equation has never been derived nor empirically verified, and we do not know whether it has a physically meaningful solution or not.

---

[5] Navier derived it in 1824, but the meaning of η was clarified by Stokes in 1845. See M. Cannone and S. Friedlander, "Navier: Blow-up and Collapse," Notices AMS **50**, 7-13 (2003).

[6] H.-T. Yau, "Asymptotic solutions to dynamics of many-body systems and classical continuum equations," in *Current Development in Mathermatics, 1998*, p155-236.

[7] For the 2D case, the unique existence of the solution is demonstrated, but as we see in **5.2.6**, 2D fluid dynamics is physically meaningless. For 3D the condition that guarantees the existence of the solution is known, but the uniqueness condition is not known. So far, all the attempts to impose uniqueness destroys all the solutions. Thus, it is a "million dollar" prize problem offered by the Clay Mathematics Institute. See http://www.claymath.org/prizeproblems/navier_stokes.pdf.

5.2 Hydrodynamics from Onsager's Point of View

5.2.1 Fluctuation of velocity field

In Chapter 4 we have considered fluctuations of quantities usually considered as thermodynamic quantities. The key is the regression hypothesis or principle (→**4.1.7**), asserting that the starting point of a nonequilibrium process may be created spontaneously by equilibrium fluctuations.

As a Brownian particle can have nonzero velocity, fluctuation can spontaneously generate systematic (at least mesoscopic scale) velocity. Then, the study of velocity fluctuations should give us the ordinary hydrodynamics outlined in **5.1**, if we believe in Einstein and Onsager.

To consider spontaneous creation of velocity field, the internal energy is inconvenient. We should use the total energy per unit volume (total energy density) $e = u + \rho\boldsymbol{v}^2/2$ as the fundamental quantity.[8] The local Gibbs relation reads (if you are not very sure, see **5.2.2**)

$$\begin{aligned} de &= du + \boldsymbol{v}\cdot d(\rho\boldsymbol{v}), && (5.2.1)\\ &= Tds + \mu dn + \cdots + \boldsymbol{v}\cdot d(\rho\boldsymbol{v}). && (5.2.2)\end{aligned}$$

Here, $\cdots$ denotes the usual terms coming from the ordinary Gibbs relation. In other words,

$$ds = \frac{1}{T}de - \frac{\mu}{T}dn - \frac{\boldsymbol{v}}{T}\cdot d(\rho\boldsymbol{v}). \tag{5.2.3}$$

This equality clearly tells us that if a mesoscopic velocity is organized (i.e., $\boldsymbol{v} \neq 0$), certainly entropy goes down. The driving force for the dissipating dynamics of velocity is given by $-\boldsymbol{v}/T$. This is consistent with the systematic part of the Langevin equation (3.2.1).

Here, s is the entropy seen from the observer moving with the volume element that is riding on the local velocity field.

5.2.2 Local form of Gibbs relation

Let us review the Gibbs relation and derive its local form in terms of thermodynamic densities. The *Gibbs relation* is of the following form:

$$dU = TdS - pdV + \mu dN + Xdx. \tag{5.2.4}$$

Since extensive quantities are homogeneous functions of degree one of the number of particles, we have

$$U = TS - pV + \mu N + Xx. \tag{5.2.5}$$

Therefore, we obtain the *Gibbs-Duhem relation*

$$SdT - Vdp + Nd\mu + xdX = 0. \tag{5.2.6}$$

Let us introduce the densities as $u = U/V$, $s = S/V$, $n = N/V$, and $\rho_x = x/V$. Then, the Gibbs-Duhem relation and (5.2.5) tell us that

$$du = Tds + \mu dn + Xd\rho_x. \tag{5.2.7}$$

Now, we introduce the total energy density $e = u + \rho\boldsymbol{v}^2/2$. Then, we have

$$de = Tds + \mu dn + Xd\rho_x + \boldsymbol{v}\cdot d\rho\boldsymbol{v}. \tag{5.2.8}$$

This is the local form of the Gibbs relation with explicit velocity or momentum density as a thermodynamic variable.

---

[8] Y. Oono, "Physical meaning of $\delta^2 z$ of Glansdorff and Prigogine," Phys. Lett. **57A**, 207-208 (1976). The paper indicates that Prigogine did not pay serious attention to fluctuations.

5.2.3 Local equilibrium state

In a true equilibrium state, there is no flow field. In the above, we have assumed that equilibrium thermodynamic relation holds locally despite flows. If this holds throughout the system, we say the system is in a *local equilibrium state*. Thus, in the application of irreversible thermodynamics (→**4.2.2**), we assume that the system is in such a state. We have seen in **5.1** that the above assumption is asymptotically exact in the hydrodynamic limit.

Local equilibrium distributions are convenient to compute local thermodynamic quantities, but they cannot correctly describe nonequilibrium states. For a state described by the local equilibrium distribution the collisions of molecules cannot change the local equilibrium state (cf. **2.4.9**), so, for example, the temperature cannot change. From our point of view (at the scale of $1/\epsilon$) the system is never in equilibrium, so there must be irreversible time evolution of the system. The lesson is: to describe nonequilibrium processes we must pay careful attention to the deviation of microscopic states from the local equilibrium distribution (→**5.1.8**).

If the evolution equations such as (4.2.2) are assumed to be given, we can use the local equilibrium distribution, assuming that the parameters in it evolve according to these evolution equations. However, it must be stressed again that the evolution equations cannot be obtained with the aid of the local equilibrium distribution.

5.2.4 Decay dynamics of slow variables

When we look at fluctuations, there are a few long lasting ones that are governed by slow dynamics compared with majority of fluctuations. A typical example is a fluctuation of a conserved quantity. If q is conserved, it cannot be created nor annihilated, so it stays there unless it goes somewhere else. Moving things always takes time. This is the reason why the fluctuations of conserved quantities are slow.

Another important class of slow variables consists of variables associated with symmetry breaking.[9] For example, close to the phase transition point that breaks symmetry, a long spatial correlation arises. Due to the long correlation, it is hard to change things locally, so their dynamics become slow. *Order parameters* close to second order phase transitions are such variables.

Exercise 1. Suppose the correlation length is 100nm. Estimate how many molecules are correlated.[10] □

We will not discuss such slow modes in detail.

The dynamics of conserved quantities always take the form of conservation law (→**5.1.5**):

$$\frac{\partial x}{\partial t} = -div(\boldsymbol{J}_x + \boldsymbol{w}), \tag{5.2.9}$$

where $\boldsymbol{J}_x$ is the systematic part of flux of x, and $\boldsymbol{w}$ is its fluctuating part.[11]

The systematic part of flux consists of reversible and irreversible fluxes (cf. **4.1.5**, **4.1.6**). This can be done by studying the time reversal symmetry of various terms (if x has a definite parity). If the term J times x changes its sign under time reversal, it is a part of reversible flux. If the sign changes, then it is a part of irreversible flux.

---

[9] If the distribution function does not have the full symmetry of the system Hamiltonian, we say the state has a *spontaneously broken symmetry*. Most realistic systems break its symmetry spontaneously at low temperatures. A recommended reference is P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Benjamin/Cummings Publ. Co. Menlo Park, CA, 1984) Chapter 2.

[10] Critical fluctuations easily reach this correlation length.

[11] If the conserved quantity can be imported from outside, we must add terms describing it.

5.2.5 Outline of hydrodynamics from regression point of view

The main problem is to write down the irreversible and reversible parts of various fluxes in (5.2.9).

The reversible fluxes may be obtained by mechanical considerations. The representative reversible flux is the one due to *advection* having the form $q\boldsymbol{v}$, where q is the density under consideration. For mass this is $\rho\boldsymbol{v}$, for momentum $\rho\boldsymbol{v}\boldsymbol{v} + p\boldsymbol{v}$, where p is pressure, and for energy $e\boldsymbol{v}$. For momentum bodily forces (gravity, etc.) creates momentum, so we must add terms describing it (e.g., $\rho\boldsymbol{F}$) as a source term.

It is not very simple to write down irreversible parts, so here only an outline is given. The crucial point is that the regression hypothesis works; that is, the forms guessed from the consideration of fluctuations indeed agree with the phenomenological laws.

The irreversible flux can have a general form given in **4.2.2**. Therefore, for mass, there is no irreversible flux, because there is no corresponding term in (5.2.2). Therefore,

$$\frac{\partial \rho}{\partial t} = -div(\rho\boldsymbol{v}). \tag{5.2.10}$$

For momentum (its density is $\rho\boldsymbol{v}$), we have the following general form: the a-component of the flux for the b-component of momentum reads

$$(\boldsymbol{J}_{irr,\rho\boldsymbol{v}_b})_a = L_{ab,cd}\nabla_c\left(-\frac{v_d}{T}\right) + L_{ab,c}\nabla_c\frac{1}{T}. \tag{5.2.11}$$

The second term is the cross term between the energy flux and momentum flux.[12] There is a very strong constraint due to translational symmetry: the equation of motion for $\boldsymbol{v}$ must not depend on the absolute value of $\boldsymbol{v}$. This implies $-\sum_d L_{ab,cd}v_d + L_{ab,c} = 0$.[13] Assuming the isotropy of the system, this reduces to the usual form of the viscous stress tensor with only two scalar parameters corresponding to shear η and bulk ζ viscosities.

The irreversible flux for energy e is of the following form:

$$(\boldsymbol{J}_{irr,e})_a = L_{a,bc}\nabla_b\left(-\frac{v_c}{T}\right) + \lambda\nabla_a\frac{1}{T}. \tag{5.2.12}$$

We may use Onsager's reciprocity (→**4.1.10**) for the cross term coefficients $L_{a,bc}$.

This way, we can derive the *Navier-Stokes equation*

$$\frac{\partial \rho\boldsymbol{v}}{\partial t} = -div(\rho\boldsymbol{v}\boldsymbol{v}) + \eta\Delta\boldsymbol{v} + \left(\frac{\eta}{3} + \zeta\right)\nabla(div\,\boldsymbol{v}) - \nabla p. \tag{5.2.13}$$

and the equation describing the energy balance in fluid. Notice that the irreversible contribution in the Navier-Stokes equation comes from higher order spatial derivatives of the velocity field.

5.2.6 Long time tail

We have discussed the correlation function of the Brownian particle in **??**, but an important effect of the surrounding fluid medium is not included: the slowness of the dynamics of excess momentum.

---

[12] The reader might say that such coupling 'should not' appear according to the so-called Curie's theorem mentioned in the footnote of **4.2.2**. Notice that this 'theorem' applies only when no translational motion exists; this motion violates the isotropy of the system.

[13] Simply set $\boldsymbol{v}$ is spatially uniform.

When the Brownian particle decreases its speed, its momentum must be imparted to the surrounding fluid. If it returns to the particle, the speed does not decay, so for further decay, the momentum imparted to the fluid layer immediately surrounding the particle must be handed over to the surrounding outer fluid layer, and so on. The process is mediated by viscosity, which is a diffusion process of momentum (→**3.1.8**): the decay process of the particle speed is the diffusion process of momentum to a wider and wider domain surrounding the particle.

Let $R(t)$ be the domain the momentum is imparted by diffusion within time t. Its size increases as $t^{1/2}$ (think about the random walk →**3.1.8**, or use dimensional analysis), so the mass of this ball is $\propto t^{3/2}$. The momentum in the ball must be conserved, so the speed of the center of mass of the ball decays as $\sim t^{-3/2}$. The Brownian particle is riding on the ball, so its velocity correlation has a positive tail decaying as $t^{-3/2}$. This is called the *long time tail.*

Exercise 1. What happens in 2D? What is its grave consequence (see (3.2.21))? If a particle is floating on the surface of water, what happens? (Note that in the real world there is no two dimensional fluid with momentum conservation; physically; 2D Navier-Stokes equation is nonsensical.) □

In the actual liquid, its microscopic structure must also relax to its equilibrium structure. This process may not be very fast, even though there is no conservation law. Thus, this can cause a long time correlation that is often much more important than the above long time tail. Such long time correlation due to the sluggishness of the structural relaxation is called the *molasses tail.*

The reader may have wondered whether there is a long range spatial correlation as well, if we have conserved quantities. As we see in **3.7.11**, miraculously, this does not happen, if we have a fluctuation-dissipation relation (→**3.2.5**). However, under non-equilibrium conditions, generally we expect log-range spatial correlation as well.

5.2.7 Shear viscosity and thermal conductivity in Green-Kubo form

Let us give two examples of (4.1.54). The key is to identify the phenomenological law.

Let us consider the shear viscosity: Looking at the general form $\boldsymbol{j} = L\nabla F$, we must identify $\boldsymbol{j}$ that is proportional to η. Consider the transport of the x-component of the momentum to the y-direction by shear stress. The xy-component of the stress tensor σ^{xy} is the corresponding flux, so j is chosen to be the xy-component of the stress tensor. The driving force F in the present case is $\boldsymbol{v}/T$, so $L\partial(v_x/T)/\partial y = \eta\partial v_x/\partial y$, or $\eta = L/T$; that is, we must divide (4.1.54) by T:

$$\eta = \frac{1}{k_BT}\int_0^\infty dt \int d\boldsymbol{r}\, \langle \sigma^{xy}(\boldsymbol{r},t)\sigma^{xy}(0,0)\rangle. \tag{5.2.14}$$

For thermal conductivity, we must isolate the conducting energy flux. The driving force is $1/T$, so $\boldsymbol{q} = L\nabla(1/T) = -\kappa\nabla T$. That is, to obtain κ (4.1.54) must be divided with T^2:

$$\kappa = \frac{1}{k_BT^2}\int_0^\infty dt \int d\boldsymbol{r}\, \langle j_u^x(\boldsymbol{r},t) j_u^x(0,0)\rangle. \tag{5.2.15}$$

where j_u is the energy flux due to conduction (advection terms must be subtracted; if we do not, we must define j^- as in the previous entry, but the outcome is the same as this formula).

Part II Details for Chapters 3-5

Part II Details consists of three chapters about Brownian dynamics, Response theory and Modeling. They may be read largely independently of Chapters 3-5 (most topics are repeated with more details), butperhaps they contain too many materials for the first time reading.

Chapter X1 discusses the so-called Brownian dynamics. Langevin equations (stochastic equations) are frequently used modeling tools of nonequilibrium phenomena. Therefore, it may be convenient to summarize their derivation, general properties and associated partial differential equations (Fokker-Planck equations and the backward equations). Stochastic calculus and path integrals (Feynman-Kac formulas) are outlined as well. The chapter begins with an overview from the large deviation point of view.

Chapter X2 summarizes the (linear) response theory. Fluctuation dissipation relations and Green-Kubo formulas are discussed with representative examples (the NMR line shape and the electric conductivity formula by Nakano). This is a standard topic of the current nonequilibrium statistical mechanics, but to stress it as its main topic is a very biased point of view, deflecting our attention from the real problems of nonequilibrium statistical physics.

Chapter X1

Brownian Dynamics II

The main purpose of this chapter is to give a reasonably detailed theory and practical tips to study dynamics of fluctuations to deepen what we have learned in our introductory survey (Sections **3.2** and **3.3**) summarized in **X1.1.1**.

X1.1 is a unified overview from the modern statistical physics point of view (unified from the large deviation point of view). **X1.2** explains the projection technique to reduce systems to stochastic models. **X1.3** is its continuation to derive nonlinear Langevin equations that are the starting point of kinetic theory of phase transitions and critical phenomena. Since nonlinear Langevin equations have mathematically subtle problems, we then go on to define white noise properly in **X1.4**. This is the section about the idealization. The Wiener process is introduced and we will learn how to interpret nonlinear Langevin equations properly, and how to analyze them with the aid of stochastic calculus in **X1.5**.

With these preparations, we discuss stochastic energetics — how to discuss the energy balance of the system described by a Langevin equation in **X1.6**.

Beyond **X1.7** is the study of fluctuations in terms of partial differential equations and path integrals. For example, we will learn the famous Feynman-Kac formula.

X1.1 Onsager's Principle

The purpose of this section is to outline the theory of Brownian motion from a unified point of view: a *large deviation point of view* (→Section **2.4**). Although many statements are not yet fully justified mathematically, the lecturer believes that something close to the following scenario will be mathematically justified.

X1.1.1 Brownian motion: Review
In Chapter 3, we have learned the history and elementary theories of Brownian motion. It is essentially the motion driven by equilibrium thermal noise.

We have learned its model due originally to Langevin (→**3.2.4**)

$$m\frac{d\boldsymbol{v}}{dt} = -\zeta\boldsymbol{v} + \boldsymbol{F} + \boldsymbol{w} \tag{X1.1.1}$$

that allows us to describe each sample path (→**3.2.1**). The crucial problem in this modeling is to choose the correct noise $\boldsymbol{w}$ that drives the motion (→**3.2.5**). The requirement is that the noise has only a very short memory, and that the solution to the model equation to be compatible with equilibrium statistical mechanics: the fluctuation dissipation relation (of the second kind) (→**3.2.5**). We have also seen that the Brownian paths are likely to be extremely jagged (→**3.4.1**).

We have had a glimpse of important relations such as the Green-Kubo relation (→**3.2.9**) and the fluctuation dissipation relation of the first kind (→**3.2.8**).

The solution to a Langevin equation is a stochastic process, so its value at time t has a distribution. Its density distribution is governed by a parabolic equation called the Fokker-Planck equation (→**3.5.3**). Under the overdamping condition we may ignore the velocity and obtain the Smoluchowski equation (→**3.5.5**). We have studied its relaxational property to equilibrium (→**3.5.9**) and the relation to the second law (→**3.5.10**).

X1.1.2 Separation of time scales

As has already been discussed in **3.4.1**, the motion of a Brownian particle is characterized by the existence of two time scales that are well separated. They are the time scale t_M we observe the particle and the time scale t_m of microscopic noise correlation. If these two time scales are not separated, the Brownian particle behaves just as an ordinary mass point governed by a Newton's equation of motion. Thus, the most crucial property of a Brownian particle is the mesoscopic time scale $]tau$ such that $t_M \gg \tau \gg t_m$.

Mathematically, the simultaneous limits $t_m/\tau \to 0$ and $\tau/t_M \to 0$ are the idealizations describing the situation, but then, the noise becomes discontinuous (→ Exercise 1 of **3.7.6**), and dx/dt in the Langevin equation must be interpreted properly, because x becomes everywhere non-differentiable. This is the topic of the Wiener process (→Section **X1.4**) and stochastic calculus (→Section **X1.5**).

X1.1.3 Onsager's principle

Suppose dx/dt is the true derivative (in the sense used in mechanics). Then, it fluctuates violently. However, if we average it for τ,

$$\dot{x}(t) \equiv \frac{1}{\tau}\int_0^\tau ds\, \frac{dx(t+s)}{dt} = \frac{x(t+\tau)-x(t)}{\tau} \tag{X1.1.2}$$

fluctuates much less violently. Its ensemble expectation value[14] is, however, not constant at the scale of t_M, and obeys a certain macroscopic law of evolution (regression principle →**4.1.2**) asymptotically[15]

$$\dot{x} = f(x) \quad \text{in the mean.} \tag{X1.1.3}$$

Notice that for us macroscopic observers the mesoscopic time scale τ may be interpreted as infinitesimal time, dt.

We assume that the large deviation principle (→**1.6.2**) holds for the short time average introduced just above:

$$P(\dot{x}|x(t)) \simeq e^{-\tau I(\dot{x})}, \tag{X1.1.4}$$

where $x(t)$ is written after | to denote explicitly that the noise is studied for a given macroscopic state of the system specified by the hydrodynamic variable (→**5.1.4**) $x(t)$ from time t, and I is the rate function, whose form we will discuss later. Let us call this *Onsager's principle.*[16] Roughly, we may say:

---

[14] Notice that $\dot{x}$ still fluctuates. The ensemble expectation value here means the average of this over many samples

[15] This is in the $\tau \to \infty$ limit. Physically, it means that τ is much larger than the correlation time of the fluctuations (noise). To formulate this mathematically, we must introduce the ratio t_M/τ and take the limit of this ratio to go to infinity first, and then take the $\tau \to \infty$ limit.

[16] Needless to say, the naming is anachronistic, because large deviation theory being used in these lecture notes is much newer than Onsager. Hashitsume named the Onsager's original proposal the *Onsager's principle* after correcting it (→**X1.1.8**); the large deviation theoretical form given here is due to the lecturer, because he believes that if Onsager knew large deviation theory, he should have written down this form as his fundamental proposal.

> *Onsager's Principle*: Large deviation principle holds for the time average over the noise time scale around the phenomenological relation.

This already incorporates the regression principle (→**4.1.2**), and we conclude that the rate function is written in terms of macroobservables.

X1.1.4 Intuitive picture of Onsager's principle.

The basic idea of Onsager's principle (interpreted in the large deviation principle sense) has a clear intuitive meaning.

If we wish to study a macroscopic evolution law, take statistics of a macroscopically short time changes: their average provides the most probable time derivative = the macroscopic evolution law. The distribution of the fluctuation around the most probable evolution law is governed by the rate function that dictates the kinetic level (= mesoscopic) stochastic behavior of the system.

As noted in **1.6.3** some sort of variational principle is also expected for the mesoscopic dynamics (→**X1.1.11**).

X1.1.5 Onsager-Hashitsume path integral

Since $\dot{x}(t)\tau = x(t+\tau) - x(t)$, $P(\dot{x}(t)|x(t)) = P(x(t+\tau)|x(t))$. Therefore, (X1.1.4) may be interpreted as the *transition probability* of a Markov process:[17]

$$P(x(t+\tau)|x(t)) \simeq e^{-\tau I(\dot{x}(t)|x(t))}. \tag{X1.1.5}$$

Here, the $x(t)$ dependence of the rate function I is explicitly denoted with the added $x(t)$ after the vertical line; the rate function is under the condition that the initial condition is $x(t)$. Therefore,

$$P(x(t+2\tau)|x(t)) \simeq \sum_{x(t+\tau)} e^{-\tau[I(\dot{x}(t+\tau)|x(t+\tau))+I(\dot{x}(t)|x(t))]}, \tag{X1.1.6}$$

where the summation (that may be an integration) over $x(t+\tau)$ means the summation for all the values of x at the intermediate time step $t+\tau$ between t and $t+2\tau$.

Repeating this procedure, we can bridge the macroscopic time span $[t_0, t_1]$ as

$$P(x(t_1)|x(t_0)) \simeq \sum_{x_1,x_2,\cdots,x_{n-1}} e^{-\tau\sum_{i=1}^{n-1} I(\dot{x}_i|x_i)}, \tag{X1.1.7}$$

where $x_i = x(t_0 + i\tau)$ and we assume $t_1 = t_0 + n\tau$. Since τ is macroscopically very small (in the mathematical idealization, infinitesimal and may be regarded macroscopically as dt), (X1.1.7) is understood as a *path integral*, and may be formally rewritten as (→**X1.10.5**)

$$P(x_F, t_1|x_I, t_0) \simeq \int_{x(t_0)=x_I}^{x(t_1)=x_F} \mathcal{D}[\{x(t)\}] \exp\left[-\int_{t_0}^{t_1} ds\, I(\dot{x}(s)|x(s))\right], \tag{X1.1.8}$$

This is often called the *Onsager-Machlup path integral*, but should better be called the *Onsager-Hashitsume path integral.*[18]

---

[17] See **6.2.1**; we will discuss this in a much more detailed fashion in Chapter 8.

[18] The transition probability (X1.1.5) was essentially proposed by Onsager. Hashitsume introduced the path integral in 1951; Onsager introduced the same path integral with his student Machlup in 1953. Thus, it is fair to call this path integral the Onsager-Hashitsume path integral. We will learn its relation to the *Feynman-Kac formula* later (→**X1.10.5**).

Thus, Onsager's principle may be restated as follows:
The Lagrangian is a rate function: the Lagrangian for the path integral representation of the time evolution of probability of observing a macroscopic state is given by the rate function for the microscopic time scale average of the time derivative of macroscopic quantities specifying the state.

X1.1.6 Gaussian approximation and Langevin equation

The rate function I is convex and vanishes for the most probable value (expectation value) (→**1.6.2**). Therefore, I must vanish if and only if the phenomenological law

$$\dot{x} = f(x) \tag{X1.1.9}$$

holds. I is usually differentiable, so we may assume that near the origin I has a quadratic form:[19]

$$I(\dot{x}(t)|x(t)) = \frac{\Gamma}{4}(\dot{x} - f(x))^2. \tag{X1.1.10}$$

This rate function is compatible with

$$\dot{x} = f(x) + w \tag{X1.1.11}$$

where w is a Gaussian noise such that $\langle w \rangle = 0$ and

$$\langle w(t)w(s) \rangle = \frac{2}{\Gamma}\delta(t-s), \tag{X1.1.12}$$

where $\langle\ \rangle$ is the ensemble average (= average over the stochastic parameter →**1.4.1**). That is, the Langevin equation (X1.1.11) is another expression of Onsager's principle in the Gaussian approximation.

The numerical factor Γ may be related to the variance of $\dot{x}$. Let us compute it. Obviously,

$$2/\Gamma = \tau\langle \dot{x}^2 \rangle - \tau \langle \dot{x} \rangle^2. \tag{X1.1.13}$$

Notice that the first term on RHS is of order 1 and the second term of order τ.[20] Therefore, we may write for $t_M \gg \tau$

$$\frac{1}{\Gamma} = \frac{1}{2}\tau\langle \dot{x}^2 \rangle = \frac{1}{2\tau}\int_0^\tau ds \int_0^\tau ds' \left\langle \frac{dx}{dt}(s)\frac{dx}{dt}(s') \right\rangle \simeq \int_0^\infty ds \left\langle \frac{dx}{dt}(s)\frac{dx}{dt}(0) \right\rangle. \tag{X1.1.14}$$

Here, dx/dt denotes the true mechanical time derivative, and we have assumed that the noise correlation decays sufficiently fast in the time scale of t_M, so we have replaced the upper limits of the integrals with infinity.

X1.1.7 Onsager's principle near equilibrium and fluctuation dissipation relation

The original Onsager principle was proposed for the linear nonequilibrium regime. Here, the phenomenological law reads (→**4.1.7**)

$$\dot{x} = L\frac{\partial S}{\partial x}, \tag{X1.1.15}$$

---

[19] Of course, more generally, I is a general quadratic functional of $\dot{x} - f(x)$. This is equivalent to assuming a noise with memory as in **X1.2.6**.

[20] Roughly speaking, $\langle \dot{x} \rangle = O[1]$, and $\langle \dot{x}^2 \rangle = O[1/\tau]$.

so that the Langevin equation is

$$\dot{x} = L\frac{\partial S}{\partial x} + w, \tag{X1.1.16}$$

where the correlation function of the noise is given in (X1.1.12). The steady distribution of this Langevin equation may be obtained as

$$P(x) \propto e^{L\Gamma S} \tag{X1.1.17}$$

as can be seen from the result in **3.5.5** (a good exercise).

The equilibrium distribution of x is given by $\propto e^{S/k_B}$ (→**2.8.5**), so

$$\Gamma = 1/k_B L. \tag{X1.1.18}$$

Combining this and (X1.1.14), we obtain the *Green-Kubo relation* (→**4.1.12**):

$$L = \frac{1}{k_B}\int_0^\infty ds \left\langle \frac{dx}{dt}(s)\frac{dx}{dt}(0)\right\rangle. \tag{X1.1.19}$$

X1.1.8 Original Onsager's principle

The transition probability in **X1.1.3** close to equilibrium reads with the aid of **X1.1.7**

$$P(x(t+\tau)|x(t)) \simeq \exp\left[-\frac{\tau}{4Lk_B}\left(\dot{x} - L\frac{\partial S}{\partial x}\right)^2\right], \tag{X1.1.20}$$

so the joint probability at two different times reads (see thermodynamic fluctuation theory **2.8.5**)

$$P(x(t+\tau), x(t)) \propto P(x(t+\tau)|x(t))e^{S(t)/k_B}. \tag{X1.1.21}$$

To study its structure further, let us consider the exponents of (X1.1.21):

$$\frac{1}{k_B}\left\{S(t) - \frac{\tau}{4L}\left(\dot{x} - L\frac{\partial S}{\partial x}\right)^2\right\} = \frac{S(t)}{k_B} - \frac{\tau}{2k_B}\left[\frac{1}{2}L^{-1}\dot{x}^2 + \frac{1}{2}L\left(\frac{\partial S}{\partial x}\right)^2\right] + \frac{\tau}{2k_B}\frac{dS}{dt}. \tag{X1.1.22}$$

Now, $\tau dS/dt = S(t+\tau) - S(t)$, so we arrive at

$$P(x(t+\tau), x(t)) \simeq \exp\frac{1}{2k_B}\left\{[S(t+\tau) + S(t)] - \tau\left[\frac{1}{2}L^{-1}\dot{x}^2 + \frac{1}{2}L\left(\frac{\partial S}{\partial x}\right)^2\right]\right\}. \tag{X1.1.23}$$

This is the original Onsager's principle (with Hashitsume's correction).

The transition probability reads

$$P(x(t+\tau)|x(t)) \simeq \exp\frac{1}{2k_B}\left\{[S(t+\tau) - S(t)] - \tau\left[\frac{1}{2}L^{-1}\dot{x}^2 + \frac{1}{2}L\left(\frac{\partial S}{\partial x}\right)^2\right]\right\}. \tag{X1.1.24}$$

In these formulas the overall factor $1/2k_B$ instead of $1/k_B$ is due to Hashitsume.

X1.1.9 Onsager principle for many variables near equilibrium

Let us summarize what we have learned, while generalizing the formulas to the many variable case.

The phenomenological law is (corresponding to (X1.1.15))

$$\dot{x}_i = \sum_j L_{ij} \frac{\partial S}{\partial x_j}, \tag{X1.1.25}$$

where L_{ij} are Onsager coefficients (a good occasion to review **4.1.12**). The rate function for $\dot{x}_i$ is (corresponding to (X1.1.10))

$$I(\{\dot{x}_i(t)\}|\{x_i(t)\}) = \frac{1}{4}\sum_{ij}\Gamma_{ij}\left(\dot{x}_i(t) - \sum_{i'} L_{ii'}\frac{\partial S(t)}{\partial x_{i'}(t)}\right)\left(\dot{x}_j(t) - \sum_{j'} L_{jj'}\frac{\partial S(t)}{\partial x_{j'}(t)}\right), \tag{X1.1.26}$$

where

$$k_B\Gamma_{ij} = (L^{-1})_{ij} \tag{X1.1.27}$$

is the fluctuation dissipation relation with an explicit form corresponding to (X1.1.19):

$$L_{ij} = \frac{1}{k_B}\int_0^\infty ds \left\langle \frac{dx_i}{dt}(s)\frac{dx_j}{dt}(0)\right\rangle. \tag{X1.1.28}$$

The Langevin equation is (corresponding to (X1.1.16))

$$\dot{x}_i = \sum_j L_{ij}\frac{\partial S}{\partial x_j} + w_i. \tag{X1.1.29}$$

with the noise satisfying $\langle w_i \rangle = 0$ and

$$\langle w_i(t)w_j(s)\rangle = 2L_{ij}k_B\delta(t-s). \tag{X1.1.30}$$

The transition probability reads (corresponding to (X1.1.24))

$$\begin{aligned} & P(\{x_i(t+\tau)\}|\{x_i(t)\}) \\ \simeq\ & \exp\frac{1}{2k_B}\left\{[S(t+\tau)-S(t)] - \tau\left[\frac{1}{2}\sum_{ij}(L^{-1})_{ij}\dot{x}_i\dot{x}_j + \frac{1}{2}\sum_{ij}L_{ij}\left(\frac{\partial S}{\partial x_i}\right)\left(\frac{\partial S}{\partial x_j}\right)\right]\right\}. \end{aligned} \tag{X1.1.31}$$

This is the original Onsager's principle.

These formulas can easily be guessed from the single variable case (X1.1.24) discussed in detail above.

X1.1.10 Variational principles near equilibrium

A large deviation principle implies a variational principle (→**1.6.3**). In our case, the variational principle implies that what we actually observe phenomenologically is the most probable behavior of the system. The most probably behavior corresponds to the path from $x(t)$ to $x(t+\tau)$ that maximizes the transition probability. Therefore, $x(s)$ that maximizes the formula inside the curly brackets in (X1.1.31) should correspond to the phenomenology.

This is a tautology for us.

The term inside the curly brackets in (X1.1.31) may be rewritten as follows:

$$\tau \left\{\sigma(t) - [\Phi(\{\dot{x}_i\}) + \Psi(\{F_i\})]\right\}, \tag{X1.1.32}$$

where σ is the *entropy production rate*, and

$$\Phi(\{\dot{x}_i\}) \equiv \frac{1}{2}\sum_{ij}(L^{-1})_{ij}\dot{x}_i\dot{x}_j, \tag{X1.1.33}$$

$$\Psi(\{F_i\}) \equiv \frac{1}{2}\sum_{ij} L_{ij}F_iF_j \tag{X1.1.34}$$

with $\partial S/\partial x_i = F_i$ (the conjugate driving force for x_i). Φ is called the *dissipation function.*[21]

X1.1.11 Onsager's principle of minimum energy dissipation

If the thermodynamic forces F_i are given (i.e., x_i are fixed), $\dot{x}_i$ are determined by the condition of maximizing (X1.1.32), or equivalently, by the minimization of

$$\Phi(\{\dot{x}_i\}) - \sigma(t). \tag{X1.1.35}$$

This is called *Onsager's principle of minimum energy dissipation.* Differentiating the above formula with respect to $\dot{x}_i$, we obtain

$$\sum_i (L^{-1})_{ij}\dot{x}_j = \frac{\partial \sigma}{\partial \dot{x}_i}. \tag{X1.1.36}$$

The entropy production rate can be written as

$$\sigma(t) = \frac{dS}{dt} = \sum_i \frac{\partial S}{\partial x_i}\dot{x}_i = \sum_i F_i\dot{x}_i. \tag{X1.1.37}$$

Therefore, (X1.1.36) indeed implies the phenomenological law $\dot{x}_i = \sum_j L_{ij}F_j$ (→**4.1.7**).

X1.1.12 Derivation of Fokker-Planck equation via Chapman-Kolmogorov relation

We can write the Chapman-Kolmogorov equation (→**3.1.10**, **6.2.2**) in terms of the (density) distribution ψ of the increment as

$$P(x, t + dt) = \int d\Delta x\, \psi(\Delta x|x - \Delta x)P(x - \Delta x, t), \tag{X1.1.38}$$

where

$$\psi(\Delta x|x - \Delta x) \simeq e^{-dtI(x|x-\Delta x)} \tag{X1.1.39}$$

---

[21] One thing we must be careful about is that in the above formalization F is a function of $\{x_i(t)\}$ that are given as the state from which the transition occurs. However, F evaluated at $\{x_i(t+\tau)\}$ (the outcome of the transition) and the original F differs by a quantity only of order some positive power of τ (likely to be $O[\tau^{1/2}]$), so we may interpret all the variables in (X1.1.32) are after the transition.

with

$$I(x|x-\Delta x) = \frac{1}{4L}[\Delta x/dt - f(x-\Delta x)]^2. \tag{X1.1.40}$$

Let us expand (X1.1.38) as

$$\begin{aligned} & P(x,t) + \frac{\partial P}{\partial t}dt + \cdots \\ = & \int d\Delta x \left[\psi(\Delta x|x)P(x,t) - \frac{\partial}{\partial x}\psi(\Delta x|x)P(x,t)\Delta x + \frac{1}{2}\frac{\partial^2}{\partial x^2}\psi(\Delta x|x)P(x,t)\Delta x^2 + \cdots\right]. \end{aligned} \tag{X1.1.41}$$

We know (ignoring higher order terms)

$$\int d\Delta x\, \psi(\Delta x|x)\Delta x = f(x)dt,\ \int d\Delta x\, \psi(\Delta x|x)\Delta x^2 = 2Ldt, \tag{X1.1.42}$$

Therefore, calculating the averages, we obtain to order dt (be careful about the ordering of derivatives and averaging; here we assume L to be constant)

$$\frac{\partial P}{\partial t}dt + \cdots = -\frac{\partial}{\partial x}f(x)dtP(x,t) + Ldt\frac{\partial^2}{\partial x^2}P(x,t) + \cdots, \tag{X1.1.43}$$

or we have obtained the Fokker-Planck (or Smoluchowski) equation

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left(-f(x) + L\frac{\partial}{\partial x}\right)P. \tag{X1.1.44}$$

We have already encountered this equation (→**3.5.3**).

X1.2 Projection and Generalized Langevin Equation

X1.2.1 Idea of projection

We have learned that the essence of the Brownian motion is the possibility of separating two time scales, microscopic and macroscopic, clearly (→**X1.1.2**). Therefore, if we can remove rapid fluctuating motions from the true microscopic mechanical equation of motion of the system consisting of the Brownian particle and the surrounding water molecules, then the remaining dynamics would describe the systematic part (long time and large length scale motion) of the Brownian particle.

A natural idea is to project out the noise part of dynamics to obtain the systematic long time global dynamic. The idea of projection was introduced by Nakajima and (later) by Zwanzig and was utilized by Zwanzig (→**X1.3**) and Mori (→**X1.2.6**) to derive mesoscopic dynamic equations.

X1.2.2 Projection

A *projection* P is a linear operator acting on a linear space of functions or vectors such that (the property is called the *idempotency* of projections)

$$PP = P. \tag{X1.2.1}$$

If P is a projection, then $Q \equiv 1 - P$ is also a projection. Of course, $PQ = 0$. If P_1 and P_2 are projections and if they commute, then $P_1 P_2$ is a projection.

If the domain of P is a vector space, we can introduce the concept of orthogonal projection: If P is a projection, and is Hermitian $P^* = P$, then it is called an *orthogonal projection.* In this case PX and QX are orthogonal vectors, so an orthogonal projection defines an orthogonal decomposition of the vector space.

Suppose V is an n-dimensional vector space with an orthonormal basis $e_1, e_2, \cdots, e_n$. Then,[22]

$$P_i \equiv e_i e_i^\dagger \tag{X1.2.2}$$

is an orthogonal projection, and all the P_i's commute. For example, $P_1 + P_3$ is an orthogonal projection, and its range is the subspace spanned by e_1 and e_3.

X1.2.3 Separation of unity

Let an n-dimensional vector space V have an orthonormal basis $e_1, e_2, \cdots, e_n$. If we define the orthogonal projection (→**X1.2.2**) $P_i \equiv e_i e_i^\dagger$, we have the *separation of unity*:

$$\sum_{i=1}^{n} P_i = I. \tag{X1.2.3}$$

Notice that $X = \sum_i (e_i e_i^\dagger) X$ is a coordinate representation of X.

If we use Dirac's bra-ket notation and write the orthonormal basis of V as $\{|1\rangle, \cdots, |n\rangle\}$, then

$$P_i = |i\rangle\langle i|, \tag{X1.2.4}$$

and (X1.2.3) is the famous separation of unity:

$$\sum_{i=1}^{n} |i\rangle\langle i| = 1. \tag{X1.2.5}$$

X1.2.4 Projective separation of time evolution equation

We wish to 'derive' the Langevin equation from mechanics. The Langevin equation we know is the equation for an observable. Therefore, the starting equation of motion must be the equation of motion for an observable (Heisenberg's equation or Hamilton's equation →**2.7.1**, **2.7.12**). The equation is a linear equation of the following form

$$\frac{dA}{dt} = iLA, \tag{X1.2.6}$$

where $LA = [H, A]/\hbar$ for Heisenberg's equation, and $iLA = [H, A]_{PB}$ for Hamilton's equation with $[\ ,\ ]_{PB}$ the Poisson bracket. L is sometimes called the *Liouville operator* or *Liouvillian.*

We consider this equation as an equation of motion in the space of observables. Thus, the projection we wish to apply on the equation must act on this space.

Let $Q = 1 - P$ and assume that the projection is time independent. We have

$$\frac{d}{dt} QA(t) = QiLPA(t) + QiLQA(t) = QiLP[PA(t)] + QiLQ[QA(t)]. \tag{X1.2.7}$$

---

[22] $\dagger$ implies the Hermitian conjugate. That is, transposition + complex conjugation $a^\dagger = \overline{a^T}$.

If we wish to make PA (the systematic part of A), Q should be the projection extracting the erratic noisy component of dynamics. Therefore, the above equation is regarded to be the equation governing the Langevin noise. We assume that the initial condition for the equation (X1.2.6) is given macroscopically (i.e., $QA(0) = 0$). Solving (X1.2.7), we obtain

$$QA(t) = \int_0^t e^{(t-s)QiLQ} QiL[PA(s)]ds = \int_0^t e^{sQiLQ} QiL[PA(t-s)]ds. \tag{X1.2.8}$$

The term whose time evolution is governed by $QiLQ$ is the best candidate of noise in the Langevin equation. Therefore, let us try to extract such a term from (X1.2.6).

Differentiating $A = PA + QA$ with respect to time, we obtain with the aid of (X1.2.8) (note that $PA(0) = A(0)$)

$$\frac{dA}{dt} = \frac{dPA}{dt} + \int_0^t e^{isQLQ} QiL \frac{d}{dt}[PA(t-s)]ds + e^{itQLQ} QiLA(0). \tag{X1.2.9}$$

To cook this further, we need a more explicit specification of the projection P.

X1.2.5 Mori theory: choice of projection

The above separation holds for any choice of time-independent projection P. We must choose a physically useful one. Let us look closely the simplest Langevin equation in **3.2.4**. The friction term is related to the correlation function of the noise. Therefore, the scalar product we need to define the orthogonal projection in **X1.2.4** should be the correlation. We know that the correlation occurs naturally in the formula describing the response (e.g., **2.8.1**).

To have a quantum counterpart, we should have the quantum version of the cross susceptibility in **2.8.1**. The difference from the classical case is due to the fact that $e^{-\beta(H - x_j X_j)} \neq e^{-\beta H} e^{\beta x_j X_j}$ (here, X_j is a number)

$$e^{-\beta(H - x_j X_j)} = e^{-\beta H} \left[1 + \int_0^\beta ds\, e^{sH} x_j e^{-sH} X_j + \cdots \right]. \tag{X1.2.10}$$

Therefore,

$$\langle \delta x_i \rangle = \left\langle \int_0^\beta ds\, e^{sH} \delta x_j e^{-sH} \delta x_i \right\rangle X_j. \tag{X1.2.11}$$

Exercise 1. Derive the above formula. (Notice that if all the quantities commute, the result agrees with the one in **2.8.1**.)[23]□

Let us introduce the following 'scalar product' symbol (called the *canonical correlation*)[24]

$$\langle A; B \rangle = \beta^{-1} \int_0^\beta ds\, Tr(\rho_e e^{sH} A e^{-sH} B). \tag{X1.2.12}$$

Here, ρ_e is the equilibrium distribution (density operator).

Using this correlation, we may introduce the projection P as

$$PB = A\langle A; B \rangle / \langle A; A \rangle. \tag{X1.2.13}$$

This may be expressed more conveniently as

$$P|B\rangle = |A\rangle\langle A|B\rangle / \langle A|A\rangle \tag{X1.2.14}$$

with the interpretation of the bracket product as the canonical correlation.

---

[23] Actually, a detailed explanation is given in **4.3.5**.

[24] That this correlation defines a scalar product may be seen from the general properties summarized in **X2.5.7**.

X1.2.6 Mori's generalized Langevin equation

Using the projection introduced in **X1.2.5**, (X1.2.9) reads (for simplicity, let us assume $\langle A;A\rangle = 1$; we can always recover this multiplicative factor easily)

$$\frac{d}{dt}|A(t)\rangle = |A\rangle\frac{d}{dt}\langle A|A(t)\rangle + \int_0^s e^{isQLQ}QiL|A\rangle\frac{d}{dt}\langle A|A(t-s)\rangle ds + e^{itQLQ}QiL|A\rangle. \quad \text{(X1.2.15)}$$

Let us introduce

$$|f(t)\rangle = e^{itQLQ}QiL|A\rangle, \quad \text{(X1.2.16)}$$

which is supposedly the noise.

(X1.2.8) reads

$$Q|A(t)\rangle = \int_0^t ds\,|f(s)\rangle\langle A|A(t-s)\rangle, \quad \text{(X1.2.17)}$$

so we may write

$$|A(t)\rangle = |A\rangle\langle A|A(t)\rangle + \int_0^t ds\,|f(s)\rangle\langle A|A(t-s)\rangle. \quad \text{(X1.2.18)}$$

(X1.2.15) reads

$$\frac{d}{dt}|A(t)\rangle = |A\rangle\frac{d}{dt}\langle A|A(t)\rangle + \int_0^s ds\,|f(s)\rangle\frac{d}{dt}\langle A|A(t-s)\rangle + |f(t)\rangle. \quad \text{(X1.2.19)}$$

The strategy beyond this point is to derive the equation of motion for the time correlation $\langle A|A(t)\rangle$, and convert (X1.2.19) to an equation for $|A(t)\rangle$ (→**X1.2.8**).

The result is *Mori's generalized Langevin equation*

$$\frac{d}{dt}|A(t)\rangle = i\Omega|A(t)\rangle + \int_0^t ds\,K_{AA}(t-s)|A(t)\rangle + |f(t)\rangle, \quad \text{(X1.2.20)}$$

where (if we recover $\langle A|A\rangle$)

$$K_{AA}(t) = \langle f(0)|f(t)\rangle/\langle A|A\rangle,\ i\Omega = \langle A|iL|A\rangle/\langle A|A\rangle. \quad \text{(X1.2.21)}$$

The first relation is the *fluctuation dissipation relation of the second kind* (→**3.2.5**) generalized to the case with memory.

X1.2.7 Memory function equation

If we apply $\langle A|$ to (X1.2.20), we obtain

$$\frac{d}{dt}C_{AA}(t) = i\Omega C_{AA}(t) - \int_0^t ds\,K_{AA}(t-s)C_{AA}(s), \quad \text{(X1.2.22)}$$

where

$$C_{AA}(t) = \langle A;A(t)\rangle \quad \text{(X1.2.23)}$$

is the canonical correlation function for A (→**X1.2.5**). The above equation is called the *memory function equation.*

X1.2.8 Demonstration of Mori's generalized Langevin equation

If the reader is familiar with Laplace transformation, go immediately to **Exercise 1** below. To study the time derivative of the correlation function apply $\langle A|$ to the original equation of motion

$$\frac{d|A(t)\rangle}{dt} = iLP|A(t)\rangle + iLQ|A(t)\rangle \tag{X1.2.24}$$

and use (X1.2.8). We obtain

$$\frac{d}{dt}\langle A|A(t)\rangle = i\Omega\langle A|A(t)\rangle + \langle A|iL\int_0^s ds\, e^{isQLQ}QiL|A\rangle\langle A|A(t-s)\rangle. \tag{X1.2.25}$$

where $i\Omega = \langle A|iL|A\rangle$. This must be a pure imaginary number. Notice that the second term reads

$$\langle A|iL\int_0^s ds\, e^{isQLQ}QiLP|A(t-s)\rangle = -\int_0^s ds\, \langle f(0)|f(s)\rangle\langle A|A(t-s)\rangle, \tag{X1.2.26}$$

Here, $\langle A|iL\cdots = -\langle iLA|\cdots$ is used (integration by parts, essentially). Therefore, (X1.2.25) becomes

$$\frac{d}{dt}\langle A|A(t)\rangle = i\Omega\langle A|A(t)\rangle - \int_0^s ds\, \langle f(0)|f(s)\rangle\langle A|A(t-s)\rangle. \tag{X1.2.27}$$

Let us put this into (X1.2.19) to obtain

$$\begin{aligned}\frac{d}{dt}|A(t)\rangle &= |A\rangle\left(i\Omega\langle A|A(t)\rangle - \int_0^s \langle f(0)|f(s)\rangle\langle A|A(t-s)\rangle ds\right) &\text{(X1.2.28)}\\ &+ \int_0^t ds\, |f(s)\rangle\left(i\Omega\langle A|A(t-s)\rangle - \int_0^{t-s}\langle f(0)|f(s')\rangle\langle A|A(t-s-s')\rangle ds'\right)\\ &+ |f(t)\rangle. \\ & &\text{(X1.2.29)}\end{aligned}$$

We can rearrange this as

$$\begin{aligned}\frac{d}{dt}|A(t)\rangle &= i\Omega\left(|A\rangle\langle A|A(t)\rangle - \int_0^t ds\, |f(s)\rangle\langle A|A(t)\rangle\right)\\ &- \int_0^t ds\, \langle f(0)|f(s)\rangle\left(|A\rangle\langle A|A(t-s)\rangle + \int_0^{t-s}|f(s')\rangle\langle A|A(t-s-s')\rangle ds'\right)\\ &+ |f(t)\rangle.\end{aligned} \tag{X1.2.30}$$

Therefore, with the aid of (X1.2.18) we finally obtain Mori's *generalized Langevin equation* (X1.2.20).

Exercise 1. If the reader is familiar with Laplace transformation, the above derivation becomes considerably transparent. Define the Laplace transform of f as

$$\int_0^\infty e^{-st}f(t)dt = f[s]. \tag{X1.2.31}$$

Then, with

$$C(t) = \langle A|A(t)\rangle, \tag{X1.2.32}$$

(X1.2.18) reads

$$|A[s]\rangle = (|A\rangle + |f[s]\rangle)C[s], \tag{X1.2.33}$$

(X1.2.15) reads (this can be obtained directly from the time derivative of (X1.2.18))

$$s|A[s]\rangle - |A\rangle = iL(|A\rangle + |f[s]\rangle)C[s], \tag{X1.2.34}$$

and (X1.2.27) reads (this can be obtained directly from (X1.2.34) by multiplying $\langle A|$)

$$sC[s]-1=(i\Omega-K[s])C[s]. \tag{X1.2.35}$$

(The RHS must be divided with $\langle A|A\rangle$, if $|A\rangle$ is not normalized.) We can easily obtain the Laplace-transformed (X1.2.20) from these. If you have still trouble, read the original: H. Mori, Prog. Theor. Phys. **33**, 423 (1965). □

X1.2.9 Kawasaki identity

Let us streamline the projection method with the aid of the following *Kawasaki identity*:[25]

$$\frac{d}{dt}e^{iLt}=e^{iLt}PiL+\int_0^t ds\, e^{iL(t-s)}PiLe^{QiLs}QiL+e^{QiLt}QiL. \tag{X1.2.36}$$

The equation of motion for the evolution operator (propagator) is

$$\frac{d}{dt}e^{iLt}=e^{iLt}PiL+e^{iLt}QiL. \tag{X1.2.37}$$

We wish to separate the 'noise' part $e^{QiLt}QiL$:

$$\frac{d}{dt}e^{iLt}=e^{iLt}PiL+(e^{iLt}-e^{QiLt})QiL+e^{QiLt}QiL \tag{X1.2.38}$$

Now, we use the identity:

$$e^{iLt}-e^{QiLt}=\int_0^t ds\, e^{iL(t-s)}PiLe^{QiLs}, \tag{X1.2.39}$$

and we get (X1.2.36).

The easiest way to demonstrate (X1.2.39) is to use Laplace transformation: $\mathcal{L}e^{iLt}=(s-iL)^{-1}$. Then, (X1.2.39) is nothing but the so-called *resolvent identity*:

$$(s-A)^{-1}-(s-B)^{-1}=(s-A)^{-1}(A-B)(s-B)^{-1}. \tag{X1.2.40}$$

Exercise 1. Demonstrate the above identity (note that A and B may not commute). □
Exercise 2. Using the Kawasaki identity, derive Mori's Langevin equation (X1.2.20) with $P=|A\rangle\langle A|$ (for the definition of the bra-ket notation in this context, see **X1.2.5**). □

X1.2.10 Structure of generalized Langevin equation

Let us look at Mori's generalized Langevin equation (X1.2.20):

$$\frac{d}{dt}|A(t)\rangle=i\Omega|A(t)\rangle-\int_0^t ds\, K_{AA}(t-s)|A(t)\rangle+|f(t)\rangle, \tag{X1.2.41}$$

where

$$K_{AA}(t)=\langle f(0)|f(t)\rangle. \tag{X1.2.42}$$

---

[25] K Kawasaki, J Phys A **6**, 1289 (1973).

The term with $i\Omega$ is an explicitly reversible term (→**4.1.5**).[26]

Since the original equation is a microscopic equation of motion, and since the derivation of the above equation uses identities only, the result (X1.2.41) must also be time reversal symmetric. We know, however, that the Langevin equation is not time reversal symmetric (→**4.1.5** or **4.1.6**), so to derive the Langevin equation we must introduce some further approximations (= some more physics) into (X1.2.41).[27]

Whether we can utilize this equation to obtain a physical insight or not depends on how to choose the projection P. If $|f(t)\rangle$ still contains some systematic motion or correlations that do not decay quickly, (X1.2.41) is very different from the usual Langevin equation we are already familiar with (→**??**).

If we could find the projection P that gives the noise $|f(t)\rangle$ with a short time correlation, then we could replace the noise with a stochastic process with no correlation (white noise). With this approximation (X1.2.41) reduces to the usual Langevin equation. However, the reader may well say that this is not a *derivation* of a Langevin equation from microscopic mechanics. She is right. Discarding the microscopic mechanical correlation in $f(t)$ makes the system irreversible.

Exercise. If the reader wishes to justify a stochastic equation microscopically, what does she have to demonstrate (or what properties do you wish to demand)? (Cf. **X1.3.10**) □

X1.2.11 Brownian particle suspended in fluid

Consider a box of size L (with a periodic boundary condition) containing a single Brownian particle of mass M whose canonical coordinates are Q and P (3-vectors) and N fluid particles of mass m whose canonical coordinates are q_i and p_i. The Hamiltonian of the system reads

$$H = P^2/2M + \sum_i V(Q - q_i) + \sum_i p_i^2/2m + \Phi(\{q_i\}), \tag{X1.2.43}$$

where V is the potential of the force between each fluid particle and the Brownian particle, and Φ is the potential energy of the fluid molecular interactions. The equation of motion for the Brownian particle reads

$$\frac{dQ}{dt} = \frac{P}{M}, \tag{X1.2.44}$$

$$\frac{dP}{dt} = -\sum_i \nabla_Q V(Q - q_i). \tag{X1.2.45}$$

---

[26] The *time reversal operator* $\mathcal{I}$ that acts on an operator in the Heisenberg picture is defined as $\hat{A}(t) = \mathcal{I}A(t)\mathcal{I}^{-1} = \overline{A(-t)}$, where $\overline{\cdots}$ is complex conjugate (not Hermitian conjugate). This complex conjugation is required, because we wish to keep the canonical commutation relation intact (i.e., if $[A, B] = C$, then we demand $[\hat{A}, \hat{B}] = \hat{C}$). Note that, in particular, $\mathcal{I}i\mathcal{I}^{-1} = -i$. That is, $\mathcal{I}$ is *antilinear*: $\mathcal{I}\sum c_i A_i = \sum \overline{c}_i \mathcal{I}A_i$. We wish to keep $\langle\psi|\psi\rangle = Tr|\psi\rangle\langle\psi|$ intact under time reversal, so $Tr(\mathcal{I}|\psi\rangle\langle\psi|\mathcal{I}^\dagger) = \langle\psi|\mathcal{I}^\dagger\mathcal{I}|\psi\rangle = \langle\psi|\psi\rangle$. Therefore, we may demand $\mathcal{I}^\dagger\mathcal{I} = \mathcal{I}\mathcal{I}^\dagger = 1$. That is, the time reversal $\mathcal{I}$ may be defined as an *antiunitary operator* (= a linear operator that is unitary but antilinear).

For example, as we know well the time reversal parity of the momentum is −1:

$$\mathcal{I}p\mathcal{I} = \mathcal{I}(-i\hbar\nabla)\mathcal{I} = i\hbar\nabla = -p$$

The time reversal of the Heisenberg equation of motion is, if H is time-reversal symmetric as is often the case,

$$i\hbar\frac{dA}{dt} = [A, H] \;\rightarrow\; i\hbar\frac{d\hat{A}}{dt} = [\hat{A}, H].$$

According to our definition $\mathcal{I}iL\mathcal{I} = -iL$, so under time reversal $i\Omega$ must change its sign, and behaves as dA/dt. That is, this term is a reversible term (→**4.1.5**).

[27] Mori once told me that a theory without approximation is not a theory in physics.

The coordinates we wish to pay particular attention are Q and P, so the projection operator we choose must project the dynamics onto the space spanned by these two coordinates. Therefore, the projection we use is formally

$$\mathcal{P}f = Q\langle Qf\rangle/\langle Q^2\rangle + P\langle Pf\rangle/\langle P^2\rangle. \tag{X1.2.46}$$

Here, $\langle\ \rangle$ implies the equilibrium average. $\langle P^2\rangle = 3k_BT/M$. $\langle Q^2\rangle$ is a big quantity.

Let us compute the reversible part (i.e., $i\Omega$ in (X1.2.20); here, the dot implies the true time derivative in mechanics).

$$\begin{aligned} \mathcal{P}\dot{Q} &= P\langle P\dot{Q}\rangle/\langle P^2\rangle = P/M, & \text{(X1.2.47)}\\ \mathcal{P}\dot{P} &= = Q\langle Q\dot{P}\rangle/\langle Q^2\rangle = 0. & \text{(X1.2.48)} \end{aligned}$$

Next, let us compute the fluctuation force at $t = 0$ (i.e., $|f(0)\rangle$). For Q, the random force at $t = 0$ is given by

$$(1-\mathcal{P})\dot{Q} = \dot{Q} - P/M = 0. \tag{X1.2.49}$$

This must be obvious without any work, because the equation for Q is purely mechanical:

$$\frac{dQ}{dt} = \frac{P}{M}. \tag{X1.2.50}$$

The random force f_P for P at time $t = 0$ is

$$f_P(0) = (1-\mathcal{P})\dot{P} = \dot{P} - Q\langle Q\dot{P}\rangle/\langle Q^2\rangle = \dot{P} = -\sum_i \nabla V(Q-q_i). \tag{X1.2.51}$$

To compute $f_P(t)$ defined by (X1.2.16), we must apply $e^{t(1-\mathcal{P})iL(1-\mathcal{P})}$ to $f_P(0)$. We Taylor-expand the exponetiated operator in t and perform the calculation order by order. The first order term reads

$$\begin{aligned} t(1-\mathcal{P})iL(1-\mathcal{P})f_P(0) &= t(1-\mathcal{P})iLf_P(0) = t(1-\mathcal{P})\frac{d}{dt}[-\sum_i \nabla_Q V(Q-q_i)] \\ &= t(1-\mathcal{P})[-\sum_i \nabla_Q^2 V(Q-q_i)(\dot{Q}-\dot{q}_i)]. & \text{(X1.2.52)} \end{aligned}$$

We assume that Q and the velocities are uncorrelated. Then, we may replace $1-\mathcal{P}$ with $1 - P\langle P\,\cdot\rangle/\langle P^2\rangle$. This implies that the term containing $\dot{Q}$ disappears. Thus,

$$t(1-\mathcal{P})iL(1-\mathcal{P})f_P(0) = t\sum_i \nabla^2 V(Q-q_i)\dot{q}_i. \tag{X1.2.53}$$

The second order is obtained by applying $(t/2)(1-\mathcal{P})iL$ to this result. The result becomes quickly complicated, but notice that the term containing $\dot{Q}$ disappears. Therefore, under this projected time evolution Q is being fixed, and only q_i evolves in time:

$$f_P(t) = e^{t(1-\mathcal{P})iL(1-\mathcal{P})}f_P(0) = -\sum_i \nabla V(Q-q_i(t)). \tag{X1.2.54}$$

We need the time correlation function of $f_P(t)$. Notice that f_P is the force due to the fluid particles on the Brownian particle with Q being fixed at the initial position. Therefore,

$$\begin{aligned} \frac{dQ}{dt} &= \frac{P}{M}, & \text{(X1.2.55)}\\ \frac{dP}{dt} &= -\int_0^t ds\, K_{ff}(t-s)P(s) + f_P(t). & \text{(X1.2.56)} \end{aligned}$$

If the force correlation is very short lived, we may write $K_{ff}(t) = 2\zeta\delta(t)$, and the second equation above reads

$$\frac{dP}{dt} = -\frac{\zeta}{M}P + f_P, \text{ with } \langle f_P(t)f_P(0)\rangle = 2\zeta k_BT\delta(t). \tag{X1.2.57}$$

Here,

$$\zeta/M = \int_0^\infty K_{ff}(s)ds \tag{X1.2.58}$$

which is the fluctuation-dissipation relation of the second kind.[28] If we assume further that the noise is Gaussian, this Langevin equation defines the *Ornstein-Uhlenbeck process* (→**X1.8.9**).

We know, however, that the above reduction is not quite right, because it does not exhibit the long time tail (→**5.2.6**). What is wrong? Up to (X1.2.56), nothing is wrong. The assumption that the correlation K_{ff} is short-lived and/or that P, Q are the only slow variables must be wrong. Which or both?

X1.2.12 Further dissection of noise

In the Mori generalized Langevin equation (X1.2.20), the noise may still contain a systematic motion that should not be treated as noise. In such a case, we should extract systematic parts further from $|f(t)\rangle$. One possible method is to apply Mori's theory **X1.2.6** again to this "noise."

Let $P_1 = |f(0)\rangle\langle f(0)|$. Now, the time evolution operator of the system is $(1-P)iL \equiv iL_1$, the natural Liouvillian governing the evolution of f. Then,

$$\frac{d}{dt}|f(t)\rangle = i\Omega_1|f(t)\rangle - \int_0^t K_2(t-s)|f(t)\rangle + |f_2(t)\rangle, \tag{X1.2.59}$$

where

$$|f_2(t)\rangle = e^{t(1-P_1)iL_1(1-P_1)}(1-P_1)iL_1|f(0)\rangle, \tag{X1.2.60}$$

and

$$K_2(t) = \langle f_2(t)|f_2(0)\rangle. \tag{X1.2.61}$$

This should easily be guessed (and derived) from the result in **X1.2.6**.

Now, it is clear that we can repeat this procedure recursively:

$$\frac{d}{dt}|f_n(t)\rangle = i\Omega_n|f_n(t)\rangle - \int_0^t ds\, K_{n+1}(t-s)|f_n(t)\rangle + |f_{n+1}(t)\rangle \tag{X1.2.62}$$

with $|f_0(t)\rangle = |A(t)\rangle$, $|f_1(t)\rangle = |f(t)\rangle$ in **X1.2.6**. Here, (we assume that vectors are not normalized)

$$P_n = |f_n\rangle\langle f_n|/\langle f_n|f_n\rangle, \tag{X1.2.63}$$

$$L_n = (1-P_{n-1})L_{n-1}, \tag{X1.2.64}$$

$$i\Omega_n = \langle f_n|iL_n|f_n\rangle/\langle f_n|f_n\rangle, \tag{X1.2.65}$$

$$K_n(t) = \langle f_n|f_n(t)\rangle/\langle f_{n-1}|f_{n-1}\rangle, \tag{X1.2.66}$$

$$|f_n(t)\rangle = e^{iL_{n+1}t}iL_{n+1}|f_n\rangle. \tag{X1.2.67}$$

The reader may say that there are two questions remaining:

(i) Can we make the noise really short-time correlated without any trace of systematic motion after a finite number of iterations of the above procedure?

(ii) If we cannot remove systematic parts with a finite number of the above recursive procedure, is the infinite procedure meaningful?

There is actually a much more important question not yet asked: does the original equation for A properly capture the systematic motion? (→**X1.2.11**, **X1.3.1**)

---

[28] The relation between K_{ff} and f_P is $K_{ff}(t) = \langle f_p(0)f_P(t)\rangle/\langle P^2\rangle = \beta\langle f_P(t)f_P(0)\rangle/M$ (calculated componentwisely).

X1.2.13 Continued fraction expansion of correlation function

If we apply $\langle f_n|$ to (X1.2.62), we obtain the memory function equations (→**X1.2.7**)

$$\frac{d}{dt}K_n(t) = i\Omega_n K_n(t) - \int_0^t ds\, K_{n+1}(t-s)K_n(s). \tag{X1.2.68}$$

If we Laplace transform this with respect to time, we get

$$sK_n[s] - K_n(0) = i\Omega_n K_n[s] - K_{n+1}[s]K_n[s], \tag{X1.2.69}$$

or

$$K_n[s] = K_n(0)/(s - i\Omega_n + K_{n+1}[s]). \tag{X1.2.70}$$

For $n = 0$ this is

$$\langle A|A[s]\rangle = \langle A|A\rangle/(s - i\Omega + \langle f|f[s]\rangle). \tag{X1.2.71}$$

Therefore, combining this and the $n = 1$ result, we obtain

$$\langle A|A[s]\rangle = \cfrac{\langle A|A\rangle}{s - i\Omega + \cfrac{\langle f|f\rangle}{s - i\Omega_2 - \langle f_2|f_2[s]\rangle}}. \tag{X1.2.72}$$

That is, more generally,

$$\langle A|A[s]\rangle = \cfrac{\langle A|A\rangle}{s - i\Omega_1 + \cfrac{\langle f_1|f_1\rangle}{s - i\Omega_2 + \cfrac{\langle f_2|f_2\rangle}{s - i\Omega_3 + \cfrac{\langle f_3|f_3\rangle}{s - i\Omega_4 + \cfrac{\langle f_4|f_4\rangle}{\cdots}}}}}. \tag{X1.2.73}$$

This is called the *continued fraction expansion of the correlation function* by Mori. The beauty of this result is that all the quantities in this formula are computed with equilibrium statistical mechanics.

Needless to say, this formal result has exactly the same problems mentioned at the end of **X1.2.12**.

X1.2.14 Finite pole approximation

In (X1.2.56) suppose the correlation function of the noise decays exponentially:

$$\langle f_1(0)|f_1(t)\rangle = e^{-\gamma t}. \tag{X1.2.74}$$

Then, we know that we can model f_1 as

$$\frac{df_1}{dt} = -\gamma f_1 + f_2 \tag{X1.2.75}$$

with $\langle f_2(0)f_2(t)\rangle = 2\gamma k_B T\delta(t)$.

$$\langle v_x|v_x[s]\rangle = \cfrac{\langle v_x^2\rangle}{s + \cfrac{K_1(0)}{s + K_2[s]}} = \cfrac{\langle v_x^2\rangle}{s + \cfrac{K_1(0)}{s+\gamma}} \tag{X1.2.76}$$

This implies that depending on $K_1(0)$ and γ, the damping can become oscillatory. Notice that the poles of the correlation function governs the general behavior.

If the power spectrum of the velocity has three peaks as in the case of classical liquids (the central peak + Stokes and anti-Stokes lines), then we may model its dynamics with three poles:

$$\langle v_x | v_x[s] \rangle = \cfrac{\langle v_x^2 \rangle}{s + \cfrac{K_1(0)}{s + \cfrac{K_2(0)}{s+\gamma}}} \tag{X1.2.77}$$

X1.2.15 How reliable is the stochastic approximation?
We have already learned that there could be lots of complications due to long-lived correlations, but even in the ideal case in which noise correlation is short-lived, how good is the replacement of the microscopic erratic fast motions that are still deterministic with a noise = stochastic process? This question will be discussed briefly in **X1.3.10**.

X1.3 Zwanzig's Nonlinear Langevin Equation

X1.3.1 Limitations of Mori's Langevin equation
Mori's Langevin equation projects the motion onto the space spanned by $|A\rangle$ (i.e., by an observable A). Therefore, e.g., A^2, which is not expressible as a constant multiple of A, is treated as nonsystematic noise. However, A^2 surely contains a systematic part that does not average out even if A averages out. Therefore, Mori's Langevin equation (X1.2.20) is reliable only when $|A| \gg A^2$.

The continued fraction expansion can certainly remove systematic parts or not fast decaying parts from the putative noise term. However, the method does not extract the systematic contribution of, e.g., A^2 that is independent of A. Therefore, the time evolution of an arbitrary function of A cannot be studied. This statement is consistent with the condition for the Mori theory to be applicable: the magnitude of $|A| \gg A^2$.

X1.3.2 Projection onto space spanned by gross variables
If we really wish to study a nonlinear system with noise, we must project the dynamics onto the space spanned by all the functions of macrovariables = gross variables $\boldsymbol{A}$. Let us discuss the classical cases only.

For the linear case the projection operator should reduce to Mori's choice (its classical version). Therefore, as its natural extension we introduce the set of (real) orthonormal polynomials $\phi_n(\boldsymbol{A})$ of gross variables $\boldsymbol{A}$ (they are multivariate polynomials) such that

$$\int d\Gamma \rho_e(\Gamma)\phi_n(\boldsymbol{A}(\Gamma))\phi_m(\boldsymbol{A}(\Gamma)) = \delta_{nm}, \tag{X1.3.1}$$

where ρ_e is the equilibrium distribution function on the phase space. The gross variables are considered as functions of microscopic variables Γ in the calculation of projection; $\boldsymbol{A}$ without explicit specification of Γ-dependence means gross variables as macroscopic variables. The projection operator $\mathcal{P}$ is defined as

$$\mathcal{P}X = \sum_n \phi_n(\boldsymbol{A}) \int d\Gamma \rho_e(\Gamma)\phi_n(\boldsymbol{A}(\Gamma))X = \sum_n \phi_n(\boldsymbol{A})\langle \phi_n(\boldsymbol{A}(\Gamma))X\rangle, \tag{X1.3.2}$$

where $\langle\ \rangle$ is the equilibrium average. Notice that (separation of unity →**X1.2.3**)

$$\sum_n \phi_n(\boldsymbol{A})\phi_n(\boldsymbol{A}') = \delta(\boldsymbol{A}-\boldsymbol{A}')/P_e(\boldsymbol{A}), \tag{X1.3.3}$$

where $P_e(\boldsymbol{A})$ is the equilibrium distribution function of the gross variables collectively denoted as a vector $\boldsymbol{A}$

$$P_e(\boldsymbol{A}) = \int d\Gamma \delta(\boldsymbol{A}-\boldsymbol{A}(\Gamma))\rho_e(\Gamma). \tag{X1.3.4}$$

X1.3.3 Zwanzig's nonlinear Langevin equation[29]

We use the projection introduced in **X1.3.2** in the Kawasaki identity **X1.2.9**. The result, called *Zwanzig's nonlinear Langevin equation*, reads (for a derivation, see **X1.3.5**)

$$\frac{d}{dt}A_j = v_j(\boldsymbol{A}(t)) + \int_0^t ds \frac{1}{P_e(\boldsymbol{A}(t-s))}\sum_k \frac{\partial}{\partial A_k(t-s)}\left[K_{kj}(s;\boldsymbol{A}(t-s))P_e(\boldsymbol{A}(t-s))\right] + f_j(t), \tag{X1.3.5}$$

where

$$\begin{aligned} v_j(\boldsymbol{A}) &= \langle iLA_j|\boldsymbol{A}\rangle, & \text{(X1.3.6)}\\ K_{ij}(t;\boldsymbol{A}) &= \langle f_i f_j(t)|\boldsymbol{A}\rangle, & \text{(X1.3.7)}\end{aligned}$$

and the conditional average is defined as

$$\langle B|\boldsymbol{A}\rangle = P_e(\boldsymbol{A})^{-1}\int \delta(\boldsymbol{A}-\boldsymbol{A}(\Gamma))B(\Gamma)\rho_e d\Gamma. \tag{X1.3.8}$$

v_j is called the *streaming term* and K_{ij} is called the *memory kernel.* (X1.3.7) is the fluctuation dissipation relation of the second kind (for a prototype see **3.2.5**). Very short time dynamics is governed by the streaming term, and is the instantaneous time-reversal part of dynamics. The streaming term has an important property that it does not affect the equilibrium distribution (→**X1.3.6**, cf. **X1.11.6**).

X1.3.4 Markov approximation of Zwanzig's nonlinear Langevin equation

If the noise is close to the white noise, we may use the 'Markov approximation'[30] (see cautionary remarks in **X1.3.9**):

$$K_{ij}(t;\boldsymbol{A}) = 2k_BTL_{ij}(\boldsymbol{A})\delta(t). \tag{X1.3.9}$$

---

[29] R Zwanzig, *Memory effects in irreversible thermodynamics*, Phys. Rev. **124**, 983-992 (1961).

[30] dA/dt is governed by the state of the system at time t; the effect of the past is only through the state at time t. There is no memory.

L_{ij} is called the *bare Onsager coefficient*. Then, (X1.3.5) reads

$$\frac{d}{dt}A_j = v_j(A(t)) + P_e(\boldsymbol{A}(t))^{-1}\sum_k \frac{\partial}{\partial A_k(t)}\left[k_BTL_{kj}(\boldsymbol{A}(t))P_e(\boldsymbol{A}(t))\right] + f_j(t), \tag{X1.3.10}$$

where the fluctuation dissipation relation of the second kind reads

$$\langle f_i(0)f_j(t)|\boldsymbol{A}(t)\rangle = 2k_BTL_{ij}(\boldsymbol{A}(t))\delta(t). \tag{X1.3.11}$$

If we write the equilibrium distribution as $P_e(\boldsymbol{A}) \propto e^{-\beta F}$, where F is the free energy, the nonlinear Langevin equation reads

$$\frac{d}{dt}A_i = v_i(\boldsymbol{A}) + k_BT\sum_j \frac{\partial L_{ij}(\boldsymbol{A})}{\partial A_j} - \sum_j L_{ij}(\boldsymbol{A})\frac{\partial F(\boldsymbol{A})}{\partial A_j} + f_i, \tag{X1.3.12}$$

$$\langle f_i(0)f_j(t)|\boldsymbol{A}\rangle = 2k_BTL_{ij}(\boldsymbol{A})\delta(t). \tag{X1.3.13}$$

Often, we ignore the explicit $\boldsymbol{A}$-dependence of the Onsager coefficients. Then, (X1.3.12) reads

$$\frac{d}{dt}A_i = v_i(\boldsymbol{A}) - \sum_j L_{ij}\frac{\partial F}{\partial A_j} + f_i, \tag{X1.3.14}$$

$$\langle f_i(0)f_j(t)\rangle = 2L_{ij}k_BT\delta(t). \tag{X1.3.15}$$

This is the starting point of the study of critical dynamics (→**X1.3.8**).
Exercise. In (X1.3.12) assume that $A = x$ is the position of a Brownian particle in a box and its equilibrium distribution is uniform in the box (i.e., $P_e(A) =$ constant). Assume that there is no streaming term. The equation reads

$$\frac{d}{dt}x(t) = k_BTL'(x(t)) + f(t). \tag{X1.3.16}$$

This describes a system with the noise intensity dependent on the position, but still its equilibrium distribution is uniform in space. Explain why we need the force term which is proportional to the derivative of L. [Hint. Think of a nonuniformly vertically vibrating horizontal tray with sand grains on it.] Can you think of some relation of this to the ratchet problem (→**??**)? □

X1.3.5 Derivation of Zwanzig's nonlinear Langevin equation

The streaming term is obtained from

$$\begin{aligned} v_j(A) &\equiv \mathcal{P}iLA_j, && \text{(X1.3.17)}\\ &= \sum_n \phi_n(\boldsymbol{A})\langle\phi_n(\boldsymbol{A}(\Gamma))iLA_j\rangle, && \text{(X1.3.18)}\\ &= \langle iLA_j|\boldsymbol{A}\rangle. && \text{(X1.3.19)}\end{aligned}$$

The noise is given by $f_j(t) = e^{i(1-\mathcal{P})Lt}i(1-\mathcal{P})LA_j$, and $i\mathcal{P}Le^{i(1-\mathcal{P})Ls}i(1-\mathcal{P})LA_j = i\mathcal{P}Lf_j(t)$, where

$$i\mathcal{P}Lf_j(t) = \sum_n \phi_n(\boldsymbol{A})\langle\phi_n(\boldsymbol{A}(\Gamma))iLf_j(s)\rangle, \tag{X1.3.20}$$

$$= -\sum_n \phi_n(\boldsymbol{A})\langle\{iL\phi_n(\boldsymbol{A}(\Gamma))\}f_j(s)\rangle, \tag{X1.3.21}$$

$$= -\sum_n \phi_n(\boldsymbol{A})\sum_k\langle\{iLA_k(\Gamma)\}\frac{\partial\phi}{\partial A_k(\Gamma)}f_j(s)\rangle, \tag{X1.3.22}$$

$$= -\int d\Gamma\sum_k\{iLA_k(\Gamma)\}\frac{\partial}{\partial A_k(\Gamma)}\delta(\boldsymbol{A}-\boldsymbol{A}(\Gamma))P_e(\boldsymbol{A})^{-1}f_j(s)\rho_e(\Gamma), \tag{X1.3.23}$$

$$= -\int d\Gamma\sum_k\{iLA_k(\Gamma)\}f_j(s)\frac{\partial}{\partial A_k(\Gamma)}\delta(\boldsymbol{A}-\boldsymbol{A}(\Gamma))\rho_e(\Gamma)/P_e(\boldsymbol{A}), \tag{X1.3.24}$$

$$= P_e(\boldsymbol{A})^{-1}\sum_k\frac{\partial}{\partial A_k}\int d\Gamma\{iLA_k(\Gamma)\}f_j(s)\delta(\boldsymbol{A}-\boldsymbol{A}(\Gamma))\rho_e(\Gamma), \tag{X1.3.25}$$

$$= P_e(\boldsymbol{A})^{-1}\sum_k\frac{\partial}{\partial A_k}\{\langle f_kf_j(s)|\boldsymbol{A}\rangle P_e(\boldsymbol{A})\}. \tag{X1.3.26}$$

Combining all of these results, we obtain (X1.3.5).

X1.3.6 Streaming term does not affect equilibrium

The streaming term (X1.3.6) may be calculated further as

$$v_j(\boldsymbol{A}) = P_e(\boldsymbol{A})^{-1}\int d\Gamma\delta(\boldsymbol{A}-\boldsymbol{A}(\Gamma))[A_j(\Gamma),H]_{PB}\rho_e(\Gamma), \tag{X1.3.27}$$

$$= k_BTP_e(\boldsymbol{A})^{-1}\int d\Gamma\delta(\boldsymbol{A}-\boldsymbol{A}(\Gamma))[A_j(\Gamma),\rho_e(\Gamma)]_{PB}, \tag{X1.3.28}$$

$$= k_BTP_e(\boldsymbol{A})^{-1}\int d\Gamma\delta(\boldsymbol{A}-\boldsymbol{A}(\Gamma))\sum_k[A_j(\Gamma),A_k(\Gamma)]_{PB}\frac{\partial}{\partial A_k(\Gamma)}\rho_e(\Gamma), \tag{X1.3.29}$$

$$= k_BTP_e(\boldsymbol{A})^{-1}\sum_k\frac{\partial}{\partial A_k}\left(\langle[A_j(\Gamma),A_k(\Gamma)]_{PB}\boldsymbol{A}\rangle P_e(\boldsymbol{A})\right). \tag{X1.3.30}$$

Here, the conditional average in the last line is defined in (X1.3.8). To go from line 3 to 4, we have used integration by parts.

This term should not affect the equilibrium distribution function given by $e^{-\beta F}$. That is, (→**X1.7.1** or **??**)

$$\sum_j\frac{\partial}{\partial A_j}v_j(\boldsymbol{A})P_e(\boldsymbol{A}) = 0. \tag{X1.3.31}$$

This is immediate from the antisymmetry of the Poisson bracket in (X1.3.30).

X1.3.7 Mode coupling terms

If we use the expression of the projection in terms of the orthonormal polynomials, we have

$$v_j(\boldsymbol{A}) = \sum_n\phi_n(\boldsymbol{A})\int d\Gamma\,\phi_n(\boldsymbol{A}(\Gamma))[A_j(\Gamma),H]\rho_e(\Gamma), \tag{X1.3.32}$$

$$= -\sum_n \phi_n(\boldsymbol{A}) \int d\Gamma\, k_BT\phi_n(\boldsymbol{A}(\Gamma))[A_j(\Gamma), \rho_e(\Gamma)]_{PB}, \tag{X1.3.33}$$

$$= \sum_n \phi_n(\boldsymbol{A}) \int d\Gamma\, k_BT[A_j(\Gamma), \phi_n(\boldsymbol{A}(\Gamma))]_{PB}\rho_e(\Gamma). \tag{X1.3.34}$$

That is,

$$v_j(\boldsymbol{A}) = k_BT \sum_n \langle [\phi_n(\boldsymbol{A}(\Gamma)), A_j(\Gamma)]_{PB}\rangle_e \phi_n(\boldsymbol{A}) \equiv \sum_n v_v^{(n)}\phi_n(\boldsymbol{A}), \tag{X1.3.35}$$

where $\langle\ \rangle_e$ is the equilibrium average. In many applications we truncate this expansion at $n = 2$. The $n = 2$ terms are called *mode-coupling terms.*

X1.3.8 Mode-coupling equation for critical dynamics

If we wish to study nonlinear dynamics of fluctuations, P_e is given by the thermodynamic theory of fluctuation (→**2.8.5**) as

$$P_e(\boldsymbol{A}) \propto \exp(-\sum_j A_j^2/2\chi_j), \tag{X1.3.36}$$

where χ_j is the susceptibility (→**2.8.1**). Since this is Gaussian, the multidimensional Hermite polynomials are the most suitable orthonormal polynomials to construct projections (→**X1.3.2**):

$$\phi_i^{(1)} = A_i/\sqrt{\chi_i}, \tag{X1.3.37}$$

$$\phi_{ij}^{(2)} = (A_iA_j - \langle A_iA_j\rangle_e)/\sqrt{\chi_i\chi_j}. \tag{X1.3.38}$$

Here, $\langle\ \rangle_e$ is the equilibrium average. From this it is easy to compute $v_j^{(1)}$ and $v_j^{(2)}$. Thus, we obtain

$$v_j(\boldsymbol{A}) = \sum_k i\omega_{jk}A_k + \frac{i}{2}\sum_{kl} V_{jkl}[A_kA_l - \langle A_kA_l\rangle_e] \tag{X1.3.39}$$

with

$$\omega_{jk} = -ik_BT\langle [A_j, A_k]\rangle_e/\chi_k, \tag{X1.3.40}$$

$$V_{jkl} = -ik_BT\langle [A_j, A_kA_l]\rangle_e/\chi_k\chi_l. \tag{X1.3.41}$$

(X1.3.31) implies

$$\omega_{jk}/\chi_j = \omega_{kj}/\chi_k \tag{X1.3.42}$$

and

$$V_{jkl}/\chi_j + V_{klj}/\chi_k + V_{ljk}/\chi_l = 0. \tag{X1.3.43}$$

The Langevin equation

$$\frac{dA_j}{dt} = \sum_k i\omega_{jk}A_k + \frac{i}{2}\sum_{kl} V_{jkl}[A_kA_l - \langle A_kA_l\rangle_e] - \sum_k L_{jk}\frac{A_k}{\chi_k} + f_j \tag{X1.3.44}$$

with $\langle f_if_j(t)\rangle = 2k_BTL_{ij}\delta(t)$ is the starting point of the *mode coupling theory* for critical dynamics.[31]

---

[31] K. Kawasaki, Ann. Phys. **61**, 1 (1970). The author told me that if he had known beforehand that this paper was to be quoted so often, he should have written a better paper, but still this is a readable paper.

X1.3.9 Warning about white noise approximation

It has repeatedly been stressed that the essence of Brownian motion is that there are two disparate time scales, macroscopic t_M and mesoscopic τ[32] such that $\tau/t_M \ll 1$ (e.g., **X1.1.2**). Also we have discussed that $\tau/t_M \to 0$ is the limit to idealize this situation mathematically. However, we already know that in this limit the noise loses its continuity as a function of time (→**3.7.6**), so in (X1.3.12) $\boldsymbol{A}$ changes very rapidly, and deciding clearly whether a certain change in $\boldsymbol{A}(t)$ ($\boldsymbol{A}$ at time t) is 'before the delta function peak in (X1.3.13), after it, or with a more delicate relation to it' becomes an important issue.

For physicists, the most sensible attitude is the following: Nothing is discontinuous dynamically: the fundamental equation of motion is a differential equation, so the coordinates q and p are at least once differentiable. Therefore, the idealization $\tau/t_M \to 0$ must be performed after solving the differential equations (or after finishing all the integrations). Even quantum mechanically the basic equation of motion is a differential equation, so all the natural changes of expectation values must be smooth.

However, it is also true that if we take this idealization limit and can assume that the rapid increment of the noise term is always statistically independent of the $\boldsymbol{A}$ at the moment (this is equivalent to assuming the noise increment is independent; 'after the delta function peak in (X1.3.13),' so to speak), many calculations become simple.

In this idealization, the result of the calculation must agree with the physicists' way of calculation: the idealization limit is taken after everything is over. To this end, we must invent a different calculus (*Itô calculus*) as already explained briefly in **3.4.4**). This is the topic of Section **X1.5**.

X1.3.10 Reliability of nonlinear reduction: Khas'minskii's theorem

Consider the following equation with two time scales

$$\dot{x}(t) = f(x(t), \xi(t/\epsilon)), \tag{X1.3.45}$$

where f is a function, and ξ is a fast motion that does not depend on x. The reader may imagine this is an external noise. Let the following limit exist:

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T dt f(u, \xi(t/\epsilon)) = F(u), \tag{X1.3.46}$$

where the time average is for fixed u. Modulo technical assumptions, Khas'minskii[33] demonstrated the following:
Up to the time scale of $1/\epsilon^2$ $x^\epsilon(t)$ defined by the following set of equations (→**X1.5.7** for a precise definition) with the same initial condition as $x(t)$ deviates at most of order ϵ from $x(t)$:

$$\begin{aligned} dx^\epsilon(t) &= F(x^\epsilon(t)) + \sqrt{\epsilon}d\eta(t), && \text{(X1.3.47)} \\ d\eta(t) &= F'(x^\epsilon(t))\eta(t)dt + \sqrt{D(x^\epsilon(t))}dB(t), && \text{(X1.3.48)} \end{aligned}$$

where B is the Wiener process (→**X1.4.4**), and

$$D(x) = \lim_{T\to\infty}\frac{1}{T}\int_0^T dt \int_0^T ds\, (f(x,\xi(t/\epsilon)) - F(x))(f(x,\xi(s/\epsilon)) - F(x)). \tag{X1.3.49}$$

---

[32] Often it is called the microscopic timescale.

[33] R. Z. Khas'minskii, "A limit theorem for the solutions of differential equations with random right-hand sides," Theor. Prob. Appl. **11**, 390-406 (1966).

That is, the effect of a rapidly changing perturbation may be mimicked by a system with a noise that is an Ornstein-Uhlenbeck process (→**X1.8.9**). The reader would say this is not quite the assertion we wish to have, because in our case ξ depends on $x(t)$.

In this case we wish to have an assertion like the following:
Consider

$$\dot{x}(t) = f(x(t), y(t)), \ \ \dot{y}(t) = \frac{1}{\epsilon} g(x(t), y(t)). \tag{X1.3.50}$$

In the $\epsilon \to 0$ limit, y is the fast variable. If the second member of the equation for each x describes a sufficiently chaotic dynamics,[34] (X1.3.47) holds with F being defined as

$$F(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(x, y(t))dt \tag{X1.3.51}$$

and with the noise governed by (X1.3.48) with an appropriate replacement of D that can easily be guessed. This type of theorems are now being proved by Kifer.[35]

Notice that these theorems are up to the time scale of $1/\epsilon^2$. Beyond this scale we do not have any control. Hopefully, with the whole dynamics being sufficiently ergodic at least on the average the solution to the reduced equation mimics the original system.

X1.4 Idealization of Brownian Motion: Wiener Process

X1.4.1 Idealization of Brownian motion

Now, we wish to idealize the Brownian motion, taking the time scale ratio $\tau/t_M \to 0$ (fast noise limit), where t_M is the macroscopic time scale (the time scale of our observation; this may perhaps be 1μs) and τ the mesoscopic time scale (this may be 10ps,[36] so the ratio is 10^{-5}).

The simplest model of the Brownian motion in 1-space is

$$\frac{dx}{dt} = w(t), \tag{X1.4.1}$$

where x is the position vector of the center of mass of a Brownian particle, and w describes the effect of random noise; this equation describes an overdamped Brownian particle bombarded by molecular collisions in a uniform space (ignoring the hydrodynamic effect).

Its idealization in the fast noise limit is, intuitively at least, to interpret w as a white noise. However, if we carefully think over the problem, we realize that 'white noise' has not been defined; we assert $\langle w(t)w(s)\rangle \propto \delta(t-s)$ (→**3.2.4**), so the equal-time correlation is unthinkable. This implies that $w(t)$ is not a stochastic process in the usual sense = a map from a probability space to a set of functions of time (→**1.4.2**).

---

[34] A precise characterization of this requirement is, for example, that for each x, the second member has a unique SRB measure.

[35] Cf. Y. Kifer, "Averaging and climate models" in Progress in Probability, vol. 49 (P. Imkeller editor, Birkhäuser, Basel, 2000).

[36] The oscillation of the CH bond stretching is of the order of 1fs = 1 $\times 10^{-15}$s.

X1.4.2 Heuristic construction of white noise

Let us try to construct a white noise $\hat{w}(t,\omega)$. Here, the stochastic parameter ω is explicitly written. What we wish to require are:
(i) 'Whiteness': all the frequency components have the same amplitudes.[37]
(ii) 'Randomness': all the frequency components are statistically independent.
Let us make such a noise on the time span $[0,1]$.

If $\hat{w}(t,\omega)$ is expanded in a Fourier series

$$\hat{w}(t,\omega)=\sum_{k=-\infty}^{\infty}a_k(\omega)e^{i2\pi kt}, \tag{X1.4.2}$$

(i) and (ii) imply that $\{a_k(\omega)\}_{k=-\infty}^{\infty}$ must be modeled as a set of iid (identically and independently distributed, cf. **??**) stochastic variables. We choose they are Gaussian variables[38]

$$\langle a_k^\dagger(\omega)a_j(\omega)\rangle=\delta_{kj},\ \ \langle a_k(\omega)\rangle=0. \tag{X1.4.3}$$

$\hat{w}(t,\omega)$ is real, so $a_{-k}^\dagger=a_k$ must be demanded.[39]

The convergence of the sum in (X1.4.2) is a non-trivial question, so we proceed intuitively: the average must vanish

$$\langle\hat{w}(t,\omega)\rangle=0. \tag{X1.4.4}$$

The time correlation function reads[40]

$$\langle\hat{w}(s,\omega)\hat{w}(t,\omega)\rangle=\sum_{k=-\infty}^{\infty}e^{2\pi k(t-s)}=\delta(t-s). \tag{X1.4.5}$$

X1.4.3 Construction of idealized Brownian motion = Wiener process

Let us formally solve (X1.4.1) with the initial condition $x(0,\omega)=0$:[41]

$$x(t,\omega)=\int_0^t\hat{w}(s,\omega)ds=\sum_{k=-\infty}^{\infty}\frac{a_k(\omega)}{2\pi ik}(e^{i2\pi kt}-1). \tag{X1.4.6}$$

Here, the adverb 'formally' means that the summation and any operation are treated as commutative.[42] We see, then

$$\langle x(t,\omega)\rangle=0. \tag{X1.4.7}$$

---

[37] This implies that all the harmonic modes have the same energy, so the total energy of a white noise diverges; that is why white noise is unthinkable as an ordinary stochastic process.

[38] If the simultaneous distribution of any finite subset of $\{a_k(\omega)\}_{k=-\infty}^{\infty}$ is Gaussian, we say the original infinite set is Gaussian-distributed.

[39] Then, how can we claim that $\{a_i(\omega)\}_{k=-\infty}^{\infty}$ is iid? In this case the interpretation is that the real and the imaginary parts of $\{a_k\}_{k=0}^{\infty}$ are iid stochastic variables (note that $\langle a_k(\omega)^2\rangle=0$; what should $\langle(\mathrm{Re}\,a_k(\omega))^2\rangle$ be?).

[40] Recall the Poisson summation formula.

[41] For $t=1$ $x(1,\omega)=a_0(\omega)$.

[42] The reader who is familiar with the theory of Fourier series may suggest that we should use the Cesaro sum, following the idea of Fejér. A similar idea indeed works (a theorem by Paley and Wiener; see Ito-McKean p21-2).

The correlation function reads

$$\langle x(s,\omega)x(t,\omega)\rangle = \min(t,s). \tag{X1.4.8}$$

We have constructed x with the aid of linear operations (summation and integration), so we may intuitively expect that it is a Gaussian stochastic process.

The constructed process is in conformity with our intuitive picture of diffusing particles (cf. **3.1.8**):

$$\langle [x(t,\omega) - x(s,\omega)]^2\rangle = |t-s|. \tag{X1.4.9}$$

That is, the displacement during time t behaves as $\sqrt{t}$.

A Gaussian process[43] $x(t,\omega)$ satisfying (X1.4.7) and (X1.4.8) is called the *Wiener process.*

From the mathematician's point of view that such a Gaussian process is indeed well-defined is a major triumph. Our construction can be fully justified with the aid of a stochastic process in the distribution sense, but a more traditional approach is outlined below in **X1.4.5**.

X1.4.4 Wiener process: definition

Through our heuristic construction of the dynamics of a particle driven by a white noise, we have clearly grasped what sort of stochastic process we should study mathematically. Although we have only studied the process on $[0,1]$, it should not be hard to extend this for all nonnegative time.

An official definition of the *Wiener process* (mathematicians often call this *Brownian motion*) is:

Definition: Wiener process

The Wiener process B is a stochastic process $\{B(t,\omega) : t \in [0,\infty), \omega \in \Omega\}$[44] satisfying the following conditions:

(A) $B(0,\omega) = 0$ with probability one.

(B) for any positive integer n and $0 < t_1 < \cdots < t_n$ the increments

$$B(t_1), B(t_2) - B(t_1), \cdots, B(t_n) - B(t_{n-1}) \tag{X1.4.10}$$

are independently and normally distributed (i.e., Gaussian) with mean 0 and variance

$$t_1, t_2 - t_1, \cdots, t_n - t_{n-1} \tag{X1.4.11}$$

(C) $B(t,\omega)$ is continuous with probability 1.

The conditions are not minimal: the reader can say that if a process satisfies (A) and (B), then it is possible to construct a model with (C). Or, for physicists it should be appealing that (A) + (C) + independent increments (the first part of (B)) implies the second part of (B) (that is, if we assume continuity and independent increments, then the central limit theorem (→**1.7.1**) forces us to swallow Gaussianness).[45]

Giving a definition is one thing, and demonstrating the existence is another; we could

---

[43] A stochastic process whose any finite dimensional marginal distribution is Gaussian is called a *Gaussian process.*

[44] on a probability space $\{\Omega, \mathcal{F}, P\}$.

[45] See Durrett, p334.

define a unicorn! The existence was first shown by Wiener in 1923.[46] An argument probably most appealing to physicists (except for the renormalization group approach →**X1.4.10**) is outlined in **X1.4.5**.[47]

X1.4.5 Outline of mathematical construction of Wiener process.

This is essentially a rigorization of our intuitive construction: **X1.4.3** suggests

$$B_n(t,\omega) = \sum_{j=0}^{n} \eta_j(\omega)\Phi_j(t) \tag{X1.4.12}$$

converges to B in the $n \to \infty$ limit, where η_i are iid stochastic variables obeying $N(0,1)$ and

$$\Phi_j(t) = \int_0^t ds\, \phi_j(t) \tag{X1.4.13}$$

with $\{\phi_j(t)\}$ being a complete orthonormal function set on $[0,\infty)$. We can show that $B_n(t,\omega)$ almost surely converges to $B(t,\omega)$[48] with the aid of Lévy's *three series convergence theorem*:
Theorem [Lévy] Let $\{X_n\}$ be independently distributed. Then, for the sum $\sum X_n$, its convergence in probability, almost sure convergence, and convergence in law are equivalent.[49] □
Therefore, we have only to show the convergence in the mean (that implies convergence in probability) of B_n, but this is a mere exercise.
Exercise. Demonstrate that $E[B_n(t,\omega)^2] \le t$. □

X1.4.6 Important properties of Wiener process

We are interested in the properties of sample paths: the properties of $B(t,\omega)$ for a fixed ω.

By construction, it is almost surely[50] continuous. However,
Non-differentiability: $B(t,\omega)$ is nowhere differentiable with probability one.[51]
This is exactly what Perrin asserted (→**3.1.5**, **3.4.1**).
Stationarity (or translational symmetry): for any positive number a $\{B(t+a,\omega)-B(a,\omega), t \ge 0\}$ is again the Wiener process.
This is easy to check. This means that if we can show that some statement holds with probability one at one time, we may claim the same statement at any other time.
Scaling invariance (or self-similarity): for any positive number λ $\{\lambda^{-1}B(\lambda^2 t,\omega), t \ge 0\}$ is again the Wiener process.
This is also easy to check.
Conformal invariance: $\{tB(1/t,\omega), t \ge 0\}$ is again the Wiener process.
This is easy to check. The very short time behavior magnified appropriately looks like a long time behavior.

---

[46] This year: France occupied the Ruhr, München Putsch; Hubble discovered extragalactic spiral galaxies; the first issue of Time magazine.

[47] Perhaps the easiest proof may be found in D. Freedman, *Brownian Motion and Diffusion* (Springer, 1983). As always, a smart proof is in Durrett, p335.

[48] It is known that the convergence is actually uniform in t.

[49] *Convergence in law* may be understood as the convergence of characteristic functions.

[50] This is a statement equivalent to "with probability 1."

[51] Not Lipschitz at any point. However, it is Hölder continuous with exponent less than 1/2 with probability one. The Hausdorff dimension of a sample path is 2 with probability 1.

X1.4.7 Increments are roughly of order $\sqrt{t}$

The translational and scaling invariance (→**X1.4.6**) imply that statistically $B(t+\delta t,\omega) - B(t,\omega)$ and $\sqrt{\delta t}B(1,\omega)$ are indistinguishable. Therefore, short time increment behaves like $\sqrt{t}$. This is of course what Einstein pointed out (→**3.1.5**).

There are two related statements:

$$\sum_{m=1}^{2^n}\left[B\left(\frac{mt}{2^n}\right) - B\left(\frac{(m-1)t}{2^n}\right)\right]^2 \to t \tag{X1.4.14}$$

with probability one as $n \to \infty$. However,

$$\limsup_{t\to 0} B(t,\omega)/\sqrt{t} = \infty. \tag{X1.4.15}$$

That is, the amplitude 'overshoots' $\sqrt{t}$.

A more precise statement about this overshoot is the *law of iterated logarithm*:

$$\sup|B(t+h) - B(t)| \sim \sqrt{2h\log\log(1/h)} \tag{X1.4.16}$$

in the $h \to 0$ limit. Or, using the conformal invariance (→**X1.4.6**), almost surely

$$\limsup_{t\to\infty} B(t,\omega)/\sqrt{2t\log\log t} = 1. \tag{X1.4.17}$$

This is Khintchine's celebrated *law of iterated logarithm*.[52]

X1.4.8 Markov property of Wiener process

By definition (→**X1.4.4**) the Wiener process has independent increments, so the conditional probability of $B(t,\omega)$ under the condition that we know 'everything' about it up to time s must be identical to the conditional probability under the condition that we know only $B(s,\omega)$. This is mathematically expressed as

$$P(B(t,\omega)|\mathcal{F}_s) = P(B(t,\omega)|B(s,\omega)), \tag{X1.4.18}$$

where $\mathcal{F}_s$ is the set of all the events up to time s.[53]

X1.4.9 *n*-dimensional Wiener process

Prepare n statistically independent Wiener processes $B_1(t,\omega),\cdots,B_n(t,\omega)$. $\boldsymbol{B}(t,\omega) = (B_1(t,\omega),\cdots,B_n($ is called the n-dimensional Wiener process.

$$\langle \boldsymbol{B}(t,\omega)\cdot\boldsymbol{B}(t,\omega)\rangle = nt. \tag{X1.4.19}$$

---

[52] This is a refinement of the central limit theorem. For a proof see Durrett p392.

[53] Here, 'all' implies, as usual, all the probabilistically meaningful events. See **1.A.20**. The Markov process will be discussed later in Chapter 9.

X1.4.10 From random walk to Wiener process: idea

Consider an isotropic random walk (→**3.1.8**) on a line. At time k ($k = 0, 1, 2, 3, \cdots$) a walker at a lattice point moves to its left or right nearest neighbor lattice point with equal probability 1/2. Let $x(n)$ be the walker's position at time n, and a_k be the k-th step (i.e., the step between time $k-1$ and k). Then,

$$x(n) = \sum_{k=1}^{n} a_k. \tag{X1.4.20}$$

This exactly has the form of the empirical sum S_n in the central limit theorem (→**1.7.1**).

What does the walk look like if seen from distance, when the duration of the random walk is sufficiently long (that is, $n \gg 1$)? To answer this question, we mimic the observation by an observer with a limited eyesight. To mimic a limited resolution, we define a coarse-grained 'block step' S_n by connecting n consecutive steps. To mimic an observation from distance we scale it with $\sqrt{n}$, because the span of the random walk is of order $\sqrt{n}$ (→**3.1.8**). Finally, we take the $n \to \infty$ limit. The limit should be what we wish to have. This idea is followed in **X1.4.11**.

X1.4.11 Construction of coarse-grained and scaled random walk

We introduce a piecewise linear continuous time version of a discrete time random walk through linear interpolation:

$$S(nt) = a_1 + \cdots + a_{[nt]} + (nt - [nt])a_{[nt]+1}, \tag{X1.4.21}$$

where the i-th step a_i is ± 1, and $t \in [0, 1]$. These steps are identically and independently distributed. Now, we scale this to define

$$B_n(t) = \frac{1}{\sqrt{n}} S(nt). \tag{X1.4.22}$$

We may say this is a renormalized random walk.

Thanks to the central limit theorem (→**1.7.1**), for any fixed t, it is easy to show that the distribution of $B_n(t)$ converges to $N(0, t)$. To check this we have only to compute the variance. Obviously,

$$E(S([nt])^2) = [nt]. \tag{X1.4.23}$$

Due to the construction, for any finite time points $t_1, t_2, \cdots, t_N \in [0, 1]$, the simultaneous distribution of $\{B_n(t_1), \cdots, B_n(t_N)\}$ converges weakly to that of $\{B(t_1), \cdots, B(t_N)\}$, where B is the Wiener process (→**X1.4.4**). This is almost trivial. Thus, we expect B_n converges to B in a certain sense.

X1.4.12 Donsker's invariance principle (continuation of **X1.4.11**)

Theorem [Donsker (1951)] B_n converges to the Wiener process B in distribution (i.e., the weak convergence of the corresponding measure). □

This is called *Donsker's invariance principle.* Here, 'invariance' implies that "many different random walks converge invariably to the Wiener process." The procedure in **X1.4.11** is a *renormalization-group procedure*, so a better name may be 'Donsker's universality principle' and for physicists Donsker's theorem should be remembered as:

Theorem [Donsker restated] The fixed point of the renormalization transformation $\mathcal{R}B_n = B_{2n}$ (defined according to **X1.4.11**) is the Wiener Process (i.e., $\mathcal{R}B = B$). □

In **X1.4.11** the convergence of the any finite-dimensional marginal distribution is mentioned, but this does not necessarily imply the convergence in the function space, because simple counter examples can be found.[54] Therefore, the nontrivial part is to show the convergence in distribution of $\{B_n(t)\}$ to $\{B(t)\}$.

X1.4.13 Idea for demonstrating Donsker's theorem.

We need 'stochastic compactness', so to speak, to be able to choose a convergent sequence. In essence we need a probabilistic version of the Arzelà-Azcoli theorem called *Prohorov's theorem*:

Theorem [Prohorov] Let Π be a set of probability measures on a probability space. The set is relatively compact iff it is tight. □

Here, the technical terms are:

Relative compactness: every sequence has a weakly convergent subsequence.

Tightness: For any $\epsilon > 0$ there is a compact set K such that for any element μ in Π $\mu(K) > 1-\epsilon$.

The tightness of Π is characterized as follows:

(i) Probabilistic boundedness: For any η there is a such that $\mu_n(\{|x(0)| \geq a\}) < \eta$.

(ii) Probabilistic equicontinuity: For any ϵ and η, there is $\delta \in (0,1)$ such that $\mu_n(\{\sup_{|s-t|<\delta} |x(s) - x(t)| \geq \epsilon\}) \leq \eta$.

This is, so to speak, 'a probabilistic version of the Arzelà-Azcoli theorem.'

Note the following theorem:

Theorem A necessary and sufficient condition for P_n to converge to P weakly is that each subsequence $\{P_{n_i}\}$ contains a further subsequence converging weakly to P. □

Necessity is obvious. The other direction is also easy; if not, there must be a subsequence not converging to P, and this subsequence cannot have a further subsequence with the required property.

Therefore, if we accept Prohorov, we have

Theorem Let P_n and P be probability measures.[55] Suppose all the finite dimensional marginal distributions of P_n converge to those of P, and $\{P_n\}$ is tight, then P_n converges to P weakly. □

Since we know the existence of B, we have only to check the tightness of $\{B_n\}$.

X1.4.14 Global properties of random walks survive RG

If we compare the starting point (lattice random walk) and the final product (the Wiener process) of the 'renormalization-group (RG) construction' in **X1.4.11**, we clearly understand what RG does: it washes away details, and only the features we can observe from distance survive.

Look at the following table in which d is the spatial dimensionality.

Lattice Random Walk	Wiener Process
$d = 2$ recurrent $d > 2$ transient	$d = 2$ recurrent $d > 2$ transient
$\langle R^2 \rangle \propto N$	$\langle R^2 \rangle \propto t$
$\exists$ double points in any d For all d the numbers of $n(\geq 3)$-tuple points are proportional to N	$\exists$ double points only for $d \leq 3$ $\exists$ $n(\geq 3)$-tuple points only in $d = 2$.†

---

[54] Recall that the joint set of countably many finite sets is a countable set, so most points in $\boldsymbol{R}$ are always excluded.

[55] on $C([0,1])$ with the uniform topology and with the usual construction of the Borel family.

Here, 'recurrent' implies that the walker returns to its starting point within a finite time with probability one,[56] and 'transient' implies that this does not happen (the returning probability is less than 1). An n-tuple point is a point where the walker visits n times.[57]

From the table we see that the global features such as the overall size of the walk $\langle R^2 \rangle$ and recurrence are not altered by renormalization; this should be so: the properties common to the lattice walk and the Wiener process must be global features of the lattice random walk, so they are independent from details of the lattice structure. In 3-space we see only occurrence of double points is global. In other words we can have macroscopic loops. However, we do not have any higher order multiple points. This implies that higher order multiple points in 3-space are all due to microscopic tiny loops that cannot be observed from distance. In higher than 4-space, even double points are not global because the space is so big.
Exercise. Let us consider a d-dimensional Wiener process.
(1) Find the probability density function for its position at time t.
(2) Write down the Chapman-Kolmogorov equation and demonstrate that it holds for the d-dimensional Wiener process.
(3) Study the recurrence for d-dimensional Wiener process (i.e., check the table above). [$\int dtp(0,t) < \infty$ is necessary and sufficient for transience to the origin. Use this fact. This is discussed in **6.1.14**.] □

X1.4.15 How reliable is the asymptotic estimates?

As we have seen above the property around the fixed point of RG in our example corresponds to the global properties of an extremely long random walk. Therefore, you may well claim that such an asymptotic result is useful only to study very long random walks, and not useful for a walk with a few tens of steps. The next table gives the probability of finding a 1d 10 step random walk to come to an even lattice point (between -10 and 10). The corresponding asymptotic approximation is the normal distribution with $N(0, 10)$

r	Exact	Asymptotic
0	0.246	0.252
2	0.205	0.207
4	0.117	0.113
6	0.044	0.042
8	0.010	0.010
10	0.001	0.002

That is, if the number of steps is about 10, then the asymptotic result may be reliable. We see that RG fixed points could be useful even practically.[58]

---

[56] However, do not jump to the conclusion that the average recurrence time is finite; indeed for $d = 2$ this average is not finite.

[57] $*$ was proved in: A. Dvoretsky, P. Erdös, and S. Kakutani, Double points of Brownian motion in n-space, Acta Scientarum Matematicarum (Szeged) **12**, 75-81 (1950); A. Dvoretsky, P. Erdös, and S. Kakutani, Multiple points of paths of Brownian motion in the plane, Bull. Res. Council Israel, **3**, 364-371 (1954); A. Dvoretsky, P. Erdös, S. Kakutani, and S. J. Taylor, Triple points of Brownian paths in 3-space, Proc. Camb. Phil. Soc. **53**, 856-862 (1957).

[58] However, how easily the asymptopia is reached depends strongly on the problems. If we view the RG procedure as a dynamical system, then the reliability of the fixed point result is closely related to the strength of the fixed point as a global attractor. Therefore, to answer the question how useful the RG is is much harder than finding the fixed points themselves.

X1.5 Stochastic Calculus

This section is an informal introduction to stochastic calculus now famous thanks to the Black-Scholes equation. Motivations of needed ideas are explained rather slowly, so those who are impatient can read the summary **X1.5.1** and the quoted final results.

X1.5.1 Stochastic calculus: practical summary of this section

A nonlinear Langevin equation

$$\frac{dx}{dt} = a(x,t) + \sigma(x,t)w, \tag{X1.5.1}$$

where w is the normalized white noise is mathematically the stochastic differential equation

$$dx = a(x,t) + \sigma(x,t)dB, \tag{X1.5.2}$$

which is defined as a shorthand notation of the following integral equation (→**X1.5.7**):

$$x(t) - x(0) = \int_0^t ds\, a(x(s),s) + \int_0^t \sigma(x(s),s)dB(s). \tag{X1.5.3}$$

Here, the last term is defined as an Itô integral **X1.5.6**.

The difficulty is due to $dB \sim \sqrt{dt}$ (→**X1.4.7**). To use the above machinery first we must properly interpret the Langevin equation (→**X1.5.8**, **X1.5.13**). With stochastic calculus, we can have the most straightforward derivation of the Fokker-Planck equation (→**X1.5.14**). Stochastic calculus is a prerequisite for easy access to the Feynman-Kac formulas (→**X1.10.1**) and path integrals (→**X1.10.2**).

X1.5.2 How to rationalize idealized Langevin equation

We have rationalized the Wiener process (→**X1.4.4**) that is supposedly a solution to (X1.4.1), so the white noise in this equation is the 'derivative' of the Wiener process. However, we have seen that it is not differentiable anywhere with probability one (→**X1.4.6**). Therefore, for example,

$$\frac{dx}{dt} = -\eta x + `\frac{dB}{dt}' \tag{X1.5.4}$$

does not make sense yet. Its formal solution can be written as

$$x(t) = e^{-\eta t}x(0) + B(t) - \eta \int_0^t dse^{-\eta(t-s)}B(s)ds, \tag{X1.5.5}$$

if we use the following formal integration by parts:

$$\int_0^t f(s)dB(s) = f(t)B(t) - \int_0^t f'(s)B(s)ds. \tag{X1.5.6}$$

Its right-hand-side is mathematically well-defined, because B is continuous, so we wish to define the Riemann-Stieltjes integral (→**X1.5.4**) on the left-hand-side in terms of the right-hand side.

The solution (X1.5.5) agrees with the result obtained by the following 'non-ideal' calculation. Assuming that $w = d\hat{B}/dt$ with $\hat{B}$ being a noise that is differentiable (its correlation function is an ordinary continuous function at worst). Our familiar ordinary calculus can be used freely, and we obtain

$$x(t) = e^{-\eta t}x(0) + \hat{B}(t) - \eta \int_0^t dse^{-\eta(t-s)}\hat{B}(s)ds. \tag{X1.5.7}$$

Now, we take the white noise limit. The result is consistent with (X1.5.5).

Thus, when the noise amplitude does not depend on x (may depend on t), we interpret

$$\frac{dx}{dt} = a(x) + \sigma(t)`\frac{dB}{dt}' \tag{X1.5.8}$$

as

$$dx(t) = a(x(t))dt + \sigma(t)dB(t), \tag{X1.5.9}$$

which is interpreted as a short-hand notation of the following integral equation

$$x(t) - x(0) = \int_0^t ds\, a(x(s)) + \int_0^t \sigma(s)dB(s). \tag{X1.5.10}$$

The last term is interpreted as (X1.5.6).

That is, we *define* (X1.5.9) as a short hand notation of the integral equation (X1.5.10).

X1.5.3 How to rationalize nonlinear Langevin equation

We are tempted to interpret (define) the general nonlinear Langevin equation, e.g., Zwanzig's nonlinear Langevin equation (→**X1.3.3**), with the normalized white noise w

$$\frac{dx}{dt} = a(x,t) + \sigma(x,t)w \tag{X1.5.11}$$

as the following integral equation:

$$x(t) - x(0) = \int_0^t ds\, a(x(s), s) + \int_0^t \sigma(x(s), s)dB(s), \tag{X1.5.12}$$

interpreting the last integral as a Riemann-Stieltjes integral (→**X1.5.4**).

As we will see later (→**X1.5.7**), this idea works eventually, but there are obstacles. The difficulty is essentially the one we have already pointed out: it is not clear whether the noise-increment could affect $x(s)$ or not. If $dB(t) = B(t+dt) - B(t)$ were of order dt, we did not have to worry about such details at all; the difference is at worst dt^2. However, in our case this quantity is roughly of order $\sqrt{dt}$ (→**X1.4.7**), so its square cannot be ignored. In short, rapid change of noise causes a headache; if a system is sufficiently nonlinear, the noise effect does not simply average out.

Therefore, if we wish to invent a method of calculation that allows taking the white noise limit before any calculation, we must scrutinize the last term in (X1.5.12).

X1.5.4 Riemann-Stieltjes integration

The *Riemann-Stieltjes integral* is defined as the following limit

$$\int_a^b f(x)dg(x) = \lim_{\|\Delta\|\to 0} \sum_{\Delta} f(x_i)[g(b_i) - g(a_i)], \tag{X1.5.13}$$

where $x_i \in \Delta_i \equiv [a_i, b_i]$ and $\|\Delta\| = \max\{b_i - a_i\}$.[59] It is known that a necessary and sufficient condition[60] for this limit to exist is that g is of *bounded-variation*:

$$v(g) \equiv \sup_{\Delta} \sum_i |g(b_i) - g(a_i)| < \infty. \tag{X1.5.14}$$

Here, $v(g)$ is called the *total variation* of g on $[a, b]$. Here, sup is taken over all possible finite partition (chopping up) of $[a, b]$.

If g is not of bounded variation, the sum in (X1.5.13) depends on the position of the sampling points x_i on each Δ_i.

If the function g is Lipschitz continuous,[61] this is finite. However, mere continuity does not guarantee the boundedness of its total variation $v(g)$.

What happens in the case of the Wiener process B (→**X1.4.4**)? We can easily guess that $B(t, \omega)$ on $[0, 1]$ is not of bounded variation: Let us chop up $[0, 1]$ into small intervals of width Δt. Then, we have $1/\Delta t$ intervals on each of which B varies of order $\sqrt{\Delta t}$, so $v(B) \sim 1/\sqrt{\Delta t}$.

Therefore, the integral $\int_0^t \sigma(x(s))dB(s)$ must be defined more carefully. Indeed, as we seen in **X1.5.5** the integral depends on the sampling method of the representative values.

X1.5.5 Backward and forward integrals

Let us take a smooth function f and consider the following sums:

$$S_F(n) = \sum_{k=1}^{n} f\left(B\left(\frac{(k-1)t}{n}\right)\right)\left[B\left(\frac{kt}{n}\right) - B\left(\frac{(k-1)t}{n}\right)\right], \tag{X1.5.15}$$

$$S_B(n) = \sum_{k=1}^{n} f\left(B\left(\frac{kt}{n}\right)\right)\left[B\left(\frac{kt}{n}\right) - B\left(\frac{(k-1)t}{n}\right)\right], \tag{X1.5.16}$$

where B is the Wiener process (→**X1.4.4**). S_F is called the *forward sum*, because the noise increment goes ahead of the sampling time; S_B is called the *backward sum*.

It is easy to see that they give different values even if convergent: the expectation value of S_F vanishes because of the independence of the increment (→**X1.4.4**), but S_B is generally nonzero. We can easily guess that the difference is

$$\lim_{n\to\infty}(S_B(n) - S_F(n)) = \int_0^t ds f'(B(s)). \tag{X1.5.17}$$

It is clear that if we can define the 'stochastic integral' by the limit of the forward sum, then calculation of expectation values should be easy. The integral defined in this forward way is called *Itô integral*.

---

[59] Precisely speaking, if the limit exists for any choice of the partition sequence and the sampling points $\{x_i\}$.

[60] See, for example, P. Protter, *Stochastic Integration and Differential Equations* (Springer, 1990) Section 1.7.

[61] A function is *Lipshitz continuous* at x, if we can find a positive constant L such that

$$|f(x) - f(y)| < L|x - y|$$

for any y in the domain of f. For example, $|x|$ is Lipshitz continuous everywhere, but $\sqrt{|x|}$ is not so at $x = 0$.

X1.5.6 Definition of Itô integral

We *define* the stochastic integral (*Itô integral*) as the limit of the regular forward Riemann-Stieltjes sum (for a motivation →**X1.5.5**):

$$\int_0^t f(s)dB(s) \equiv \lim_{n\to\infty}\sum_{k=1}^n f\left(B\left(\frac{(k-1)t}{n}\right)\right)\left[B\left(\frac{kt}{n}\right)-B\left(\frac{(k-1)t}{n}\right)\right]. \tag{X1.5.18}$$

The convergence (that must be demonstrated for continuous functions f) is in the almost sure sense.

X1.5.7 Definition of stochastic differential equation

We *interpret* the nonlinear Langevin equation (X1.5.11) as a shorthand notation of (X1.5.12) with the last integral interpreted in Itô's sense. That is:
The differential relation (called a *stochastic differential equation*):

$$dx = a(x,t) + \sigma(x,t)dB \tag{X1.5.19}$$

is defined as a shorthand notation of the following integral equation:

$$x(t) - x(0) = \int_0^t ds\, a(x(s),s) + \int_0^t \sigma(x(s),s)dB(s), \tag{X1.5.20}$$

where the last term is defined as an Itô integral **X1.5.6**.

Notice that Zwanzig's Langevin equation must, by derivation (→**X1.3.5**),[62] be interpreted in Itô's sense.

X1.5.8 Physical equations with noise vs stochastic differential equations

For physicists physical noises are all of bounded variation (→**X1.5.4**), so Itô's specification is often unphysical. That is, the ideal white noise limit of the solution to the following 'physical' Langevin equation

$$\frac{dx}{dt} = a(x,t) + \sigma(x,t)\frac{d\hat{B}}{dt} \tag{X1.5.21}$$

and the solution to the stochastic differential equation

$$dx(t) = a(x,t)dt + \sigma(x,t)dB(t) \tag{X1.5.22}$$

generally do *not* agree.

Since mathematically and even from the practical computational point of view, Itô's interpretation is far more superior than other interpretations, and since for physicists this interpretation is not necessarily natural, we should have a way to translate different interpretations (→**X1.5.13**; **X1.5.9** is a preparation).
Warning: However, if a stochastic differential equation is derived through projection, you must scrutinize the derivation before deciding the interpretation. □

---

[62] Notice that the noise $\boldsymbol{f}(t)$ is always computed with the condition $\boldsymbol{A}(t)$ and then evolving in time, so the increment $d\boldsymbol{f}$ is not affecting $\boldsymbol{A}(t)$.

X1.5.9 Itô's lemma: heuristic consideration

Let u be a function of t and the Wiener process $B(t)$ (→**X1.4.4**). Taylor-expanding u, we have

$$\begin{aligned} & u(t+\Delta t, B(t+\Delta t)) - u(t, B(t)) \\ = \ & \Delta t \frac{\partial u}{\partial t} + \Delta B(t)\frac{\partial u}{\partial B} + \frac{1}{2}(\Delta t)^2 \frac{\partial^2 u}{\partial t^2} + \Delta t \Delta B \frac{\partial^2 u}{\partial t \partial B} + \frac{1}{2}(\Delta B)^2 \frac{\partial^2 u}{\partial B^2} + \cdots, \end{aligned} \tag{X1.5.23}$$

where $\Delta B(t) = B(t+\Delta t) - B(t)$. Among the formally second order terms, the dangerous term is only the last term, because $\Delta B \sim \sqrt{\Delta t}$ (→**X1.4.7**).

Summing this from $t = 0$ to T, we can express $u(T, B(T))$ as an integral of the following form

$$u(T, B(T)) - u(0,0) = \int_0^T \frac{\partial u}{\partial t} ds + \int_0^T \frac{\partial u}{\partial B} dB(s) + \frac{1}{2}\int_0^T \frac{\partial^2 u}{\partial B^2} dt. \tag{X1.5.24}$$

Here, the stochastic integral (the second term on the right hand side) is in Itô's sense (→**X1.5.6**). That is, almost surely

$$du(t, B) = \left(\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial^2 u}{\partial B^2}\right) dt + \frac{\partial u}{\partial B} dB. \tag{X1.5.25}$$

To demonstrate (X1.5.24), the nontrivial part is to show that almost surely

$$\sum \frac{1}{2}(\Delta B)^2 \frac{\partial^2 u}{\partial B^2} \to \frac{1}{2}\int_0^T \frac{\partial^2 u}{\partial B^2} dt. \tag{X1.5.26}$$

The first step is to show this is true in the mean (in the L^2-sense), and then we use the fact that we can choose a subsequence that almost surely converges.

X1.5.10 Itô's lemma: general case

If $dx = adt + \sigma dB$, it is easy to generalize (X1.5.25) as[63]

$$du(t, x) = \left(\frac{\partial u}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 u}{\partial x^2}\right) dt + \frac{\partial u}{\partial x} dx. \tag{X1.5.27}$$

This is called *Itô's lemma*. An easy way to memorize it is

$$du = \frac{\partial u}{\partial t} dt + \frac{\partial u}{\partial x} dx + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} dx^2 \tag{X1.5.28}$$

with the rule $dB^2 = dt$ (so $dx^2 = \sigma^2 dt$; other cross terms are of higher order).

Exercise.
(1) Compute $d \sin B(t)$. Then, check that $d\langle \sin B(t)\rangle = -\langle \sin B\rangle dt/2$.
(2) Compute $d\sin(\omega t + \epsilon B(t))$. Demonstrate that its average decays eventually to zero. (We already did this calculation!).
(3) Compute $d\exp(B(t) - t/2)$, and then find the general solution to $df = f dB(t)$ for f.
(4) Suppose $dx = -xdt + \sqrt{2}xdB$. Compute dx^2. Then, study the time-dependence of $\langle x^2\rangle$.[64]
□

---

[63] A fairly readable (but you must be patient when you read math books) proof can be found in Z. Brzeźniak and T. Zastawniak, *Basic Stochastic Processes* (Springer, 1998).

[64] This is a very pathological example. Think how pathological it is. Is there any stationary distribution?

X1.5.11 Itô integral: examples

Let us first compute the forward sum S_F (→**X1.5.5**) in the $n \to \infty$ limit for an integrable function f. According to Itô's lemma[65] (→**X1.5.10**)

$$d\left[\int_0^x f(x')dx'\right]_{x=B(t)} = f(B(t))dB(t) + \frac{1}{2}f'(B(t))dt. \tag{X1.5.29}$$

Here, the integral on the left-hand side is an ordinary integral. Thus, we have obtained

$$\lim_{n\to\infty} S_F(n) = \int_0^t f(B(s))dB(s) = \int_0^{B(t)} f(x)dx - \frac{1}{2}\int_0^t dsf'(B(s)). \tag{X1.5.30}$$

Notice that the average must be zero. Therefore,

$$\left\langle \int_0^{B(t)} f(x)dx \right\rangle = \frac{1}{2}\int_0^t ds\langle f'(B(s))\rangle. \tag{X1.5.31}$$

More concrete examples follow:

$$\int_0^t B(s)dB(s) = \frac{1}{2}(B(t)^2 - t). \tag{X1.5.32}$$

Indeed, its average vanishes by definition of the Winer process: $\langle B(t)^2\rangle = t$.

$$\int_0^t e^B dB = \int_0^B e^x dx - \frac{1}{2}\int_0^t e^{B(s)}ds. \tag{X1.5.33}$$

Since $B(t)$ is Gaussian $\langle e^B\rangle = e^{\langle B^2\rangle/2} = e^{t/2}$, we can explicitly check that this is correct.
Exercise.
(1) Compute $\int e^{3B}dB$ and demonstrate that its expectation value vanishes.
(2) Let $dp = -(\zeta/m)pdt + \sqrt{2k_BT\zeta}dB$. Compute dp^2, and then study the time dependence of its average $\langle p^2\rangle$.□

X1.5.12 Wong-Zakai's theorem: motivation

For physicists all the noises are of bounded variation (→**X1.5.4**), so

$$I = \int_0^t f(\hat{B}(s))d\hat{B}(s) = \int_0^{\hat{B}(s)} f(x)dx. \tag{X1.5.34}$$

Therefore, taking the ideal white noise limit of this integral, we must have

$$I \to \int_0^{B(t)} f(x)dx = \int_0^t f(B(s))dB(s) + \frac{1}{2}\int_0^t f'(B(s))ds. \tag{X1.5.35}$$

This follows from the usual Ito's lemma (→**X1.5.10**):

$$d\int_0^{B(t)} f(x)dx = f(B(t))dB(t) + \frac{1}{2}f'(B(t))dt, \tag{X1.5.36}$$

---

[65] Warning: If the integral inside the square brackets is understood as Itô's integral, *by definition* (→**X1.5.7**) $d\int f(x)dx = f(x)dx$. No extra term appears. Here, it is an ordinary integral; after finishing the integral we put B in the integral result (in the primitive function of f).

The short-hand notation for this integral relation is

$$f(B)\circ dB = f(B)dB + \frac{1}{2}f'(B)(dB)^2, \tag{X1.5.37}$$

where ∘ implies that we take the white noise limit *after* performing the integral, assuming that the noise of B is of bounded variation (→**X1.5.4**).

More, generally, if $dx = adt + \sigma dB$, then the idealization limit should read

$$f(x)\circ dB = f(x)dB + \frac{1}{2}f'(x)dxdB = f(x)dB + \frac{1}{2}f'(x)\sigma dt. \tag{X1.5.38}$$

X1.5.13 Wong-Zakai's theorem

Suppose we write down a stochastic differential equation (Langevin equation) physically, assuming that the noise is of bounded variation (→**X1.5.4**) as (→**X1.5.12** for ∘)

$$dx = a(x,t)dt + \sigma(x,t)\circ dB. \tag{X1.5.39}$$

Here, ∘ is used to denote clearly that the term is NOT interpreted in Itô's sense, but in the usual calculus sense.[66] If we wish to use the Itô calculus, we must rewrite this as (→(X1.5.38))

$$dx = \left(a(x,t) + \frac{1}{2}\sigma\frac{\partial\sigma}{\partial x}\right)dt + \sigma(x,t)dB. \tag{X1.5.40}$$

That is, we must augment the extra term in (X1.5.38). The relation between (X1.5.39) and (X1.5.40) is called *Wong-Zakai's theorem.*

Exercise 1. Suppose a variable is described by the following stochastic differential equation (in the standard sense)

$$dx = \sqrt{f(x)}dB. \tag{X1.5.41}$$

We know the average of x is time independent. That is, the noise cannot drive the particle on the average. Suppose we wish to model this with a physical noise. If we study

$$\frac{dx}{dt} = \sqrt{f(x)}w, \tag{X1.5.42}$$

where w is a random noise (not quite white), then we see that x is time dependent on the average.

(1) Find the equation for the average $\langle x(t)\rangle$.

(2) Therefore, the reader realizes that to model (X1.5.41) she must add a systematic drive. Find a physical model that mimics (X1.5.41). [This is a trivial question if the reader has answered (1).] □

X1.5.14 Derivation of Fokker-Planck equation: a smart way

Let $dx = adt+\sigma dB$. We know the density distribution function of x at time t reads (→**1.4.5**)

$$P(y,t) = \langle\delta(y - x(t))\rangle. \tag{X1.5.43}$$

---

[66] Often we say in *Stratonovich's sense.*

Let us derive the Fokker-Planck equation with the aid of Itô calculus (→**X1.5.10**):

$$d\delta(y - x(t)) = -\frac{\partial}{\partial y}\delta(y - x(t))dx(t) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\delta(y - x(t))\sigma^2(x(t), t)dt \tag{X1.5.44}$$

$$= -\frac{\partial}{\partial y}\delta(y - x(t))[a(x(t), t)dt + \sigma(x(t), t)dB(t)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}\delta(y - x(t))\sigma^2(x(t), t)dt. \tag{X1.5.45}$$

Let us take the average over the stochastic parameter:

$$dP(y, t) = -\frac{\partial}{\partial y}a(y, t)P(y, t)dt + \frac{1}{2}\frac{\partial^2}{\partial y^2}\sigma(y, t)^2 P(y, t)dt. \tag{X1.5.46}$$

Thus, we have arrived at

$$\frac{\partial}{\partial t}P(y, t) = -\frac{\partial}{\partial y}a(y, t)P(y, t) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\sigma(y, t)^2 P(y, t). \tag{X1.5.47}$$

This is the stochastic Liouville approach (→**??**) combined with the Itô's lemma (→**X1.5.10**); perhaps this is the easiest derivation of the Fokker-Planck (or Smoluchowski) equation.

X1.6 Stochastic Energetics

This section is largely dependent on Ken Sekimoto's lecture notes in 1999 for a graduate course at Kyoto University (but perhaps uses stochastic calculus more extensively).

X1.6.1 How good is Langevin equation?

We have 'derived' the Langevin equation through separating slow and fast variables with the aid of projection (→**X1.2.4**), and then approximating the fast variable with a white noise (→Section **6.2**, **6.3**). By this last step, the time reversal symmetry of the original mechanics is lost.

Consider

$$\frac{dp}{dt} = -\zeta\frac{p}{m} + \sqrt{2\zeta k_BT}w(t), \tag{X1.6.1}$$

where w is the normalized white noise constructed in **X1.4.2** (intuitively, the 'derivative' of the Wiener process, dB/dt; $\langle w\rangle = 0$, $\langle w(s)w(t)\rangle = \delta(t-s)$). Its coefficient is due to the fluctuation dissipation relation (→**3.2.5**).[67] If we average this over the stochastic parameters (i.e., the ensemble average), we obtain

$$\frac{\partial\langle p\rangle}{\partial t} = -\eta\frac{\langle p\rangle}{m}, \tag{X1.6.2}$$

so indeed the average decays irreversibly.

However, this irreversibility is caused by the choice of the 'unnatural' initial ensemble that is away from equilibrium. For an ensemble with $\langle p\rangle = 0$ there is no explicit irreversibility. Actually, the stationary correlation function of p is given by the Langevin equation realistically except for a very short time (cf. **3.2.7**) and is time reversal symmetric.

Exercise. Confirm the above assertion. Solve (X1.6.1) for $p(t)$ for very large t (sufficient to forget the initial condition) and then compute the correlation function $\langle p(t)p(s)\rangle$ (the answer can be read off from **3.2.4**). □

How natural is the motion described by the Langevin equation? From the large deviation approach explained in **X1.1.6**, we see that the Langevin equation describes the 'semimacroscopic' fluctuations faithfully. Therefore, we can expect that reversible fluctuations around the equilibrium state can be described accurately.

Sekimoto[68] has realized that even energetically the Langevin equation is a good description of mesoscopic scale dynamics.

X1.6.2 Langevin equation is a force-balance equation

Consider

$$\begin{aligned} \frac{dx}{dt} &= \frac{p}{m}, &\text{(X1.6.3)}\\ \frac{dp}{dt} &= -\zeta\frac{p}{m} + \sqrt{2\zeta k_BT}w - \frac{\partial U(x,\lambda)}{\partial x}, &\text{(X1.6.4)} \end{aligned}$$

---

[67] Mathematically, the equation is (→**X1.5.7**)

$$dp = -\zeta\frac{p}{m}dt + \sqrt{2\zeta k_BT}dB(t).$$

[68] K. Sekimoto, J. Phys. Soc. Jpn. **66**, 1234 (1997).

where w is the normalized white noise (→**X1.6.1**, **X1.4.2**), and λ is the control parameter just as in Einstein's statistical mechanics (→**2.5.3**). The second equation above is the equation of motion, so the right-hand side describes the total force acting on the particle.

What is the interaction with the heat bath? The damping term produces heat and is returned to the heat bath, so it is an interaction with the bath. The thermal noise is certainly the interaction with the heat bath. The potential term appearing in the Langevin equation is already the systematic part alone, because it is a function of the systematic gross variable(s) and the control parameter λ. Therefore,

$$f_{HB} = -\zeta \frac{p}{m} + \sqrt{2\zeta k_B T} w \tag{X1.6.5}$$

is the force exerted on the particle by the heat bath.

In the overdamped case (corresponding to the Smoluchowski equation **3.5.5**)

$$0 = -\zeta \frac{dx}{dt} + \sqrt{2k_B T\zeta} w - \frac{\partial U(x,\lambda)}{\partial x}, \tag{X1.6.6}$$

where w is the normalized white noise. The first two terms on the right-hand side must be f_{HB} in agreement with (X1.6.5) in form.

X1.6.3 Heat: with inertia

The work done by the force f when the displacement due to the force is dx is fdx in the usual mechanics. If we wish to take the white noise idealization limit, it must be interpreted as $f \circ dx$ (→**X1.5.13**).[69]Therefore, the work done by the heat bath on the particle must be written as

$$\Delta \hat{Q} = \int_0^t f_{HB} \circ dx(t), \tag{X1.6.7}$$

and is interpreted as the heat imported to the system. That is, we must interpret the stochastic integral in the Stratonovich sense *irrespective* of the interpretation of the Langevin equation being used to describe the stochastic motion.

For the case of (X1.6.4)

$$\Delta \hat{Q} = \int_0^t \left(-\zeta \frac{p}{m} + w\right) \circ dx \tag{X1.6.8}$$

Let us compute $d\hat{Q}$:

$$\begin{aligned} d\hat{Q} &= \left(-\zeta \frac{p}{m} + \sqrt{2\zeta k_B T}\hat{w}\right) \circ dx, & \text{(X1.6.9)} \\ &= -\zeta \frac{p}{m} dx + \sqrt{2\zeta k_B T} \frac{p}{m} \circ dB, & \text{(X1.6.10)} \\ &= -\zeta \left(\frac{p}{m}\right)^2 dt + \sqrt{2\zeta k_B T} \left(\frac{p}{m} dB + \frac{1}{2m} dp dB\right), & \text{(X1.6.11)} \\ &= -\frac{2\zeta}{m}\left(\frac{p^2}{2m} - \frac{k_B T}{2}\right) dt + \frac{\sqrt{2\zeta k_B T}}{m} p dB. & \text{(X1.6.12)} \end{aligned}$$

---

[69] Whether we choose fdx in the Itô sense or in the Stratonovich sense depends on the order of calculation and idealization. Since fdx is a formula of mechanics, the white-noise idealization must be done after all the calculations are finished. Therefore, $f \circ dx$ is the proper interpretation. An intuitive explanation might be give as follows as well. Mechanics assumes that the increment dx and f are simultaneous; the change in x and that in f are simultaneous and actually neither one is the cause of the other. Also mechanics is time reversal symmetric, so $f \circ dx$ must be the most natural interpretation.

We have used Wong-Zakai's theorem **X1.5.13**:

$$f(\boldsymbol{x},t)\circ dx_k = f(\boldsymbol{x},t)dx_k + \frac{1}{2}\sum_j \frac{\partial}{\partial x_j} f(\boldsymbol{x},t)dx_j dx_k \tag{X1.6.13}$$

with an appropriate interpretation of the formally second order terms (see Itô's lemma **X1.5.10**); in (X1.6.11) $dpdB = \sqrt{2\zeta k_B T}dt$.

The ensemble average gives

$$\frac{d\langle \hat{Q}\rangle}{dt} = -\frac{2\zeta}{m}\left(\frac{\langle p^2\rangle}{2m} - \frac{k_BT}{2}\right). \tag{X1.6.14}$$

That is, if the kinetic energy is excessive, it is discarded to the heat bath (recall that our sign convention is: import > 0; export < 0).

X1.6.4 Energy balance: with inertia

The energy balance equation can be written in a thermodynamically familiar form for each sample (the following argument is NOT ensemble-averaged).

$$\Delta\hat{Q} = \int_0^t \left(\frac{dp}{dt} + \frac{\partial U}{\partial x}\right)\circ dx(t) = \left[\frac{p^2}{2m} + U\right]_0^t - \int_0^t \frac{\partial U}{\partial \lambda} d\lambda(t). \tag{X1.6.15}$$

We have thus obtained

$$\Delta E \equiv \left[\frac{p^2}{2m} + U\right]_0^t = \Delta\hat{Q} + \int_0^t \frac{\partial U}{\partial \lambda} d\lambda(t). \tag{X1.6.16}$$

Therefore, we must conclude that the work given to the system ΔW must be

$$\Delta W = \int_0^t \frac{\partial U}{\partial \lambda} d\lambda(t) \tag{X1.6.17}$$

as Einstein did (→**2.5.3**).

X1.6.5 Energy balance: overdamped case

In the overdamped case we must use (X1.6.6), and

$$\Delta\hat{Q} = \int_0^t \left(-\zeta\frac{dx}{dt} + \sqrt{2k_BT\zeta}\hat{w}\right)\circ dx(t). \tag{X1.6.18}$$

With the aid of (X1.6.6) this implies

$$\Delta E = U(t) - U(0) = \Delta\hat{Q} + \int_0^t \frac{\partial U}{\partial \lambda} d\lambda(t). \tag{X1.6.19}$$

Since we assume that the velocity is totally damped, there is no kinetic energy. Again, this energy balance equation is not an averaged equation, but holds for each sample (for each choice of the stochastic parameter).

X1.6.6 Heat: overdamped case

Let us compute the ensemble average as well.

$$\begin{aligned} dQ &= \frac{\partial U}{\partial x} \circ dx, & \text{(X1.6.20)} \\ &= \frac{\partial U}{\partial x} dx + \frac{1}{2}\frac{\partial^2 U}{\partial x^2}(dx)^2 & \text{(X1.6.21)} \\ &= -\frac{1}{\zeta}\left[\left(\frac{dU}{dx}\right)^2 - k_BT\frac{\partial^2 U}{\partial x^2}\right] dt + \sqrt{\frac{2k_BT}{\zeta}}\frac{\partial U}{\partial x} dB. & \text{(X1.6.22)} \end{aligned}$$

Therefore, the ensemble average gives us

$$\frac{d\langle \hat{Q}\rangle}{dt} = -\frac{1}{\zeta}\left\langle \left(\frac{dU}{dx}\right)^2 - k_BT\frac{\partial^2 U}{\partial x^2}\right\rangle. \tag{X1.6.23}$$

For example, if the potential is harmonic $U = \alpha x^2/2$,

$$\frac{d\langle \hat{Q}\rangle}{dt} = -\frac{2\alpha}{\zeta}\left[\frac{1}{2}\alpha\langle x^2\rangle - \frac{1}{2}k_BT\right]. \tag{X1.6.24}$$

That is, the deviation from the equipartition of energy is adjusted as heat.

Exercise. Using the fact that in equilibrium $\langle x^2\rangle$ must be constant, compute $\langle x\partial U/\partial x\rangle$. (First, compute $d(x^2)$.) □

X1.6.7 Passivity revisited

We have discussed the passivity of the Smoluchowski equation in **3.5.10**. We can hope we will have a more direct understanding of passivity = the second law with stochastic energetics.

The Langevin equation we are interested in is

$$\zeta\frac{dx}{dt} = \sqrt{2\zeta k_BT}\hat{w} - \frac{\partial U}{\partial x} + F(t). \tag{X1.6.25}$$

Here, $F(t)$ is the external force we can change freely (that is why it is explicitly time-dependent), but is uncorrelated with noise (controllable only macroscopically)

$$\begin{aligned} dQ &= \left[\frac{\partial U}{\partial x} - F\right] \circ dx, & \text{(X1.6.26)} \\ &= \left[\frac{\partial U}{\partial x} - F\right] dx + \frac{1}{2}\frac{\partial^2 U}{\partial x^2}(dx)^2, & \text{(X1.6.27)} \\ &= -\frac{1}{\zeta}\left[\frac{\partial U}{\partial x} - F\right]^2 dt + \left[\frac{\partial U}{\partial x} - F\right]\sqrt{\frac{2k_BT}{\zeta}} dB + \frac{k_BT}{\zeta}\frac{\partial^2 U}{\partial x^2} dt. & \text{(X1.6.28)} \end{aligned}$$

Now, we must consider the long time average of this. If the long time average may be equated as the ensemble average, then it is easy to show that the second law holds (→**3.5.10**; we obtain $-W$): We can obtain

$$dQ = \int dx\left(\frac{\partial U}{\partial x}J\right) dt, \tag{X1.6.29}$$

where J is the probability flux introduced in **3.5.7**. This is just $-Wdt$ in **3.5.10**, so we are done. Or, more precisely,

$$dQ = -\int dx \frac{J^2}{P} dt + TdS, \tag{X1.6.30}$$

where S is the entropy defined by $S = -k_B \int dx P \log P$. That is, heat is on the average going out of the system; no perpetual motion of the second kind is possible.

X1.6.8 Passivity of Fokker-Planck equation

We have obtained (X1.6.14):

$$\frac{dQ}{dt} = -\frac{2\zeta}{m}\left(\frac{\langle p^2\rangle}{2m} - \frac{k_BT}{2}\right). \tag{X1.6.31}$$

This may be rewritten as

$$\frac{dQ}{dt} = \int d\boldsymbol{x}d\boldsymbol{p}\left[-\frac{\zeta}{m}\boldsymbol{p} - k_BT\zeta\frac{\partial}{\partial \boldsymbol{p}}\log P\right]\cdot \boldsymbol{J}_x. \tag{X1.6.32}$$

Here, $\boldsymbol{J}_x = \boldsymbol{v}P$ (the $\boldsymbol{x}$-component of the probability flux). Similarly, $\boldsymbol{J}_p$ appearing below is the $\boldsymbol{p}$-component and is given by

$$\boldsymbol{J}_p = (-\zeta\boldsymbol{p}/m + \boldsymbol{F} - k_BT\zeta\nabla_{\boldsymbol{p}}\log P)P. \tag{X1.6.33}$$

Exercise. Check the above equality. □

We may rewrite the above equation further into

$$\begin{aligned}\frac{dQ}{dt} &= \int d\boldsymbol{x}d\boldsymbol{p}\frac{[\boldsymbol{J}_p - \boldsymbol{F}P]\cdot[\boldsymbol{J}_p - \boldsymbol{F}P + k_BT\nabla_{\boldsymbol{p}}P]}{P}, && \text{(X1.6.34)}\\ &= \int d\boldsymbol{x}d\boldsymbol{p}\left\{\frac{[\boldsymbol{J}_p - \boldsymbol{F}P]^2}{P} + k_BT\nabla_{\boldsymbol{p}}\log P\cdot(\boldsymbol{J}_p - \boldsymbol{F}P)]\right\}, && \text{(X1.6.35)}\\ &= \int d\boldsymbol{x}d\boldsymbol{p}\frac{[\boldsymbol{J}_p - \boldsymbol{F}P]^2}{P} - k_BT\int d\boldsymbol{x}d\boldsymbol{p}\,[\log P\nabla_{\boldsymbol{p}}\cdot\boldsymbol{J}_p + \boldsymbol{F}\cdot\nabla_{\boldsymbol{p}}P]. && \text{(X1.6.36)}\end{aligned}$$

The last term with $\boldsymbol{F}$ vanishes, so

$$\begin{aligned}\frac{dQ}{dt} &= \int d\boldsymbol{x}d\boldsymbol{p}\frac{[\boldsymbol{J}_p - \boldsymbol{F}P]^2}{P} - k_BT\int d\boldsymbol{x}d\boldsymbol{p}\,\log P\nabla_{\boldsymbol{p}}\cdot\boldsymbol{J}_p, && \text{(X1.6.37)}\\ &= \int d\boldsymbol{x}d\boldsymbol{p}\frac{[\boldsymbol{J}_p - \boldsymbol{F}P]^2}{P} - k_BT\int d\boldsymbol{x}d\boldsymbol{p}\,\log P\left[\frac{\partial P}{\partial t} + \nabla_{\boldsymbol{x}}\cdot\boldsymbol{J}_x\right]. && \text{(X1.6.38)}\end{aligned}$$

The contribution of the second term in the last integral vanishes, and using the definition of entropy $S = -k_B\int d\boldsymbol{x}d\boldsymbol{p}\,P\log P$, we have

$$\frac{dQ}{dt} = T\frac{dS}{dt} - \int d\boldsymbol{x}d\boldsymbol{p}\frac{[\boldsymbol{J}_p - \boldsymbol{F}P]^2}{P}. \tag{X1.6.39}$$

That is, on the average the heat must be discarded.

X1.6.9 Contact with heat baths at different temperatures

Consider a system in contact with n heat baths at different temperatures T_i ($i = 1, \cdots, n$).

$$dx_i = -\frac{1}{\zeta_i}\frac{\partial}{\partial x_i}U + \sum_i \sqrt{\frac{2k_BT_i}{\zeta}}dB_i. \tag{X1.6.40}$$

The heat exchange with the i-th bath may be written as

$$dQ_i = \frac{\partial U}{\partial x_i} \circ dx_i \tag{X1.6.41}$$

Following the calculation in **X1.6.7**, it is easy to show

$$dS = \sum_i \frac{dQ_i}{T_i} + \sum_i \frac{1}{T_i}\int dx_i \frac{J_i^2}{P}dt, \tag{X1.6.42}$$

where $J_i = -(\partial U/\partial x_i)P - (k_BT_i/\zeta_i)\partial P/\partial x_i$. The net entropy production is $dS/dt - \sum_i (dQ_i/dt)/T_i$, and is nonnegative.

As a concrete example, let us consider the conduction of heat from a high temperature bath to the low temperature one. We have two baths 1 and 2. Using the above model with $U(x_1, x_2) = U(x_1 - x_2)$, and introducing $X = (x_1 + x_2)/2$, $x = x_1 - x_2$, we obtain

$$dQ_1 = U'(x) \circ dx_1 = \frac{1}{2}dU + U' \circ dX = \frac{1}{2}dU + U'dX + \frac{1}{2}U''dxdX \tag{X1.6.43}$$

$$dQ_2 = -U'(x) \circ dx_2 = \frac{1}{2}dU - U' \circ dX = \frac{1}{2}dU - U'dX - \frac{1}{2}U''dxdX. \tag{X1.6.44}$$

Immediately, we can confirm the conservation of energy:

$$dQ_1 + dQ_2 = dU. \tag{X1.6.45}$$

Taking the ensemble average, we obtain

$$d\langle Q_1\rangle = \frac{1}{2}d\langle U\rangle - \frac{1}{2}\langle U'^2\rangle\left(\frac{1}{\zeta_1} - \frac{1}{\zeta_2}\right)dt + \frac{1}{2}k_B\langle U''(x)\rangle\left(\frac{T_1}{\zeta_1} - \frac{T_2}{\zeta_2}\right)dt. \tag{X1.6.46}$$

Therefore,

$$J = \frac{1}{2\zeta_1}\left[k_BT_1\langle U''\rangle - \langle U'^2\rangle\right] - \frac{1}{2\zeta_2}\left[k_BT_2\langle U''\rangle - \langle U'^2\rangle\right]. \tag{X1.6.47}$$

To compute the entropy production in a steady state, we need the steady P.

Exercise 1. Assuming that U is harmonic, explicitly check that $J > 0$, if $T_1 - T_2 > 0$. Remember that heat coming into the system is +.

(1) Show that there is no heat flow when $T_1 = T_2$ (that is, the system is in equilibrium as should be).

(2) Assume that the temperature of the spring is at the mean temperature $(T_1 + T_2)/2$ (if ζ_1 is close to ζ_2, it is reasonable). Check Clausius is correct.

(3) The reader should have realized that she can generally check Clausius' principle. Do it.

(4) If we wish to be quantitative, how should we proceed? □

The flux itself must fluctuate. This must be obtained by the fluctuating part of $dQ_1 + dQ_2$. That is,

$$U'dX = U'\left(\frac{k_BT_1}{\zeta_1} + \frac{k_BT_2}{\zeta_2}\right)^{1/2} dB. \tag{X1.6.48}$$

Exercise 2. Calculate the fluctuation of the heat flux when $U = \alpha x^2/2$. □

X1.7 Fokker-Planck Equations

X1.7.1 From stochastic differential equation to Fokker-Planck equation[70]

We have already demonstrated in **X1.5.14** the following (here, the result is generalized for the case of vectors): Consider the following vector stochastic differential equation[71]

$$d\boldsymbol{x} = \boldsymbol{a}(\boldsymbol{x}, t)dt + \sigma(\boldsymbol{x}, t)d\boldsymbol{B}(t), \tag{X1.7.1}$$

where $\boldsymbol{B}$ is the d-dimensional Wiener process (→**X1.4.9**). In components this reads

$$dx_i = a_i(\boldsymbol{x}, t)dt + \sum_k \sigma_{ik}(\boldsymbol{x}, t)dB_k. \tag{X1.7.2}$$

The density distribution function $P(\boldsymbol{x}, t)$ of $\boldsymbol{x}$ is governed by the equation (symbolically; in terms of components see the next equation)

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{a}(\boldsymbol{x}, t)P(\boldsymbol{x}, t) + \frac{1}{2}\frac{\partial^2}{\partial \boldsymbol{x}^2}\sigma(\boldsymbol{x}, t)\sigma(\boldsymbol{x}, t)^T P(\boldsymbol{x}, t). \tag{X1.7.3}$$

Its meaning may be easier to understand in a componentwise expression:

$$\frac{\partial P(\boldsymbol{x}, t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}[a_i(\boldsymbol{x}, t)P(\boldsymbol{x}, t)] + \frac{1}{2}\sum_{ij}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\left[\left(\sum_k \sigma_{ik}(\boldsymbol{x}, t)\sigma_{jk}(\boldsymbol{x}, t)\right)P(\boldsymbol{x}, t)\right] \tag{X1.7.4}$$

Let the solution to this equation with the initial condition $P(\boldsymbol{x}, t_0) = \delta(\boldsymbol{x} - \boldsymbol{x}_0)$ be written as $P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0)$. It is clear that $P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0)d\boldsymbol{x}$ gives the *transition probability*: the probability of finding the sample process to be in the volume element $d\boldsymbol{x}$ at time t under the condition that at time t_0 it is at $\boldsymbol{x}_0$. Thus, we may understand that the Fokker-Planck equation governs the transition probabilities.

Exercise 1. Derive the above equation, following **X1.5.14**. It is convenient to use the summation convention of repeated indices. □

Exercise 2. Derive the Fokker-Planck equation corresponding to (X1.3.12):

$$\frac{dP}{dt} = \sum_i \frac{\partial}{\partial A_i}\left[\left(-v_i + \sum_j L_{ij}\frac{\partial \mathcal{F}}{\partial A_j}\right) + \sum_j k_BTL_{ij}\frac{\partial}{\partial A_j}\right]P. \tag{X1.7.5}$$

□

---

[70] A. D. Fokker, “Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld,” Ann. Phys. **43**, 810-820 (1914); later more generally by M. Planck, “Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie,” Sitz. der preuss. Akad. p324-341 (1917). A mathematically respectable derivation was first given by A. N. Kolmogorov, “Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung,” Math. Ann. **104**, 415-458 (1931).

[71] When we say a stochastic differential equation (SDE), it implies the SDE in Itô’s sense (→**X1.5.7**).

X1.7.2 Physical stochastic differential equation vs stochastic differential equation

In many cases, especially when we simply write down an equation with the aid of physical intuition, we assume the noise is of bounded variation (→**X1.5.4**), so when we write

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{a}(\boldsymbol{x}, t) + \sigma(\boldsymbol{x}, t)\boldsymbol{w}(t) \tag{X1.7.6}$$

it should be interpreted as (→**X1.5.13**)

$$d\boldsymbol{x}(t) = \boldsymbol{a}(\boldsymbol{x}, t)dt + \sigma(\boldsymbol{x}, t) \circ d\boldsymbol{B}(t). \tag{X1.7.7}$$

The Wong-Zakai's theorem **X1.5.13** tells us that this must be read as

$$dx_i = \left[\alpha_i(\boldsymbol{x}, t) + \frac{1}{2}\sum_{jk}\frac{\partial \sigma_{ij}}{\partial x_k}\sigma_{kj}\right]dt + \sum_j \sigma_{ij}(\boldsymbol{x}, t)dB_j(t), \tag{X1.7.8}$$

if we wish to use stochastic calculus (→**X1.5.7**) .

However, as warned before (→**X1.5.8**), not all the stochastic differential equations derived in physics are interpretable as physical (or Stratonovich) stochastic differential equations. Equations derived by the projection technique must be scrutinized before deciding its interpretation. A rule of thumb is: if the stochastic increment is treated as independent from the current state, then the obtained equation is in the standard (i.e., Itô) sense. A typical example is Zwanzig's Langevin equation (→**X1.3.3**) in the white noise limit.

X1.7.3 Kramers equation and Smoluchowski equation

We call the equation governing the density distribution function of the solution to a stochastic differential equation generally a *Fokker-Planck equation*.[72] For a Brownian particle, there seems to be a tradition to call the equation with the inertial effect the *Kramers equation* and the one without it the *Smoluchowski* equation. We have studied their properties relevant to physics in **3.5.9**, **3.5.10**, **X1.6.7**, **X1.6.8**, etc. The relation to detailed balance will be discussed in **6.11**.

As an exercise let us derive the Kramers and the Smoluchowski equations according to **X1.7.1**.

The stochastic differential equation describing a Brownian particle under a force $\boldsymbol{F}$ (that may depend on time, the particle position, and external control parameters) reads (cf. **3.5.2**, **??**)

$$\begin{aligned} d\boldsymbol{x} &= (\boldsymbol{p}/m)dt, && \text{(X1.7.9)} \\ d\boldsymbol{p} &= [-(\zeta/m)\boldsymbol{p} + \boldsymbol{F}(\boldsymbol{x}, t)]dt + \sqrt{2k_BT\zeta}d\boldsymbol{B}, && \text{(X1.7.10)} \end{aligned}$$

where $\boldsymbol{B}$ is the d-dimensional Wiener process. In this case, the noise term is with a constant coefficient, so this physical stochastic differential equation may be interpreted as a stochastic differential equation. Therefore, (X1.7.4) immediately gives us

$$\frac{\partial P(\boldsymbol{x}, \boldsymbol{p}, t)}{\partial t} = -(\partial_{\boldsymbol{x}}, \partial_{\boldsymbol{p}})^T(\boldsymbol{p}/m, -\zeta/m)\boldsymbol{p} + \boldsymbol{F})P + k_BT\zeta\frac{\partial^2}{\partial \boldsymbol{p}^2}P, \tag{X1.7.11}$$

---

[72] Mathematicians sometimes call it a Kolmogorov's forward equation. For the backward equation, see **X1.9.2**.

$$= -\frac{\partial}{\partial \boldsymbol{x}}\frac{\boldsymbol{p}}{m}P - \frac{\partial}{\partial \boldsymbol{p}}\left[-\frac{\zeta}{m}\boldsymbol{p} + \boldsymbol{F} - k_BT\zeta\frac{\partial}{\partial \boldsymbol{p}}\right]P, \tag{X1.7.12}$$

$$= -\sum_i \frac{\partial}{\partial x_i}\frac{p_i}{m}P - \sum_i \frac{\partial}{\partial p_i}\left[-\frac{\zeta}{m}p_i + F_i - k_BT\zeta\frac{\partial}{\partial p_i}\right]P. \tag{X1.7.13}$$

This is in agreement with (3.5.14). This is sometimes called the *Kramers equation.*

Without inertia, the starting point is (→**3.5.5**)

$$d\boldsymbol{x} = (\boldsymbol{F}/\zeta)dt + \sqrt{2k_BT/\zeta}d\boldsymbol{B}(t). \tag{X1.7.14}$$

Again, we may interpret this as a proper stochastic differential equation, so the corresponding Fokker-Planck equation reads, as we already know,

$$\frac{\partial P(\boldsymbol{x},t)}{\partial t} = \frac{\partial}{\partial \boldsymbol{x}}\left[-\frac{F(\boldsymbol{x})}{\zeta} + \frac{k_BT}{\zeta}\frac{\partial}{\partial \boldsymbol{x}}\right]P. \tag{X1.7.15}$$

This is often called the *Smoluchowski equation.*

X1.7.4 From Kramers to Smoluchowski equation: directly from Langevin equation

We simply reduced (6.3.7) to (6.3.11), saying that if the friction is large, then inertia should be ignorable (→**3.5.5**). We should be able to do this reduction more systematically.

To perform the reduction directly from the Langevin equation, we stretch the time as $\tau = \zeta t = t/\epsilon$; Let us write a small quantity $1/\zeta = \epsilon$. In terms of this new time (6.3.7) reads

$$d\boldsymbol{r} = \epsilon(\boldsymbol{p}/m)d\tau, \tag{X1.7.16}$$

$$d\boldsymbol{p} = (-\boldsymbol{p}/m + \boldsymbol{F}/\zeta)d\tau + \sqrt{2k_BT/\zeta}dB(\tau), \tag{X1.7.17}$$

where the time-scaled Wiener process $dB(\tau)$ is defined as $\langle dB(\tau)dB(\tau)\rangle = d\tau$; since this is ζdt, the multiplicative factor is scaled with ζ. We expand the solution as $\boldsymbol{r} = \boldsymbol{r}_0 + \epsilon\boldsymbol{r}_1 + \cdots$, $\boldsymbol{p} = \boldsymbol{p}_0 + \epsilon\boldsymbol{p}_1 + \cdots$.

To the zeroth order we have

$$d\boldsymbol{r}_0 = 0, \tag{X1.7.18}$$

$$d\boldsymbol{p}_0 = -(\boldsymbol{p}_0/m)d\tau + \sqrt{2k_BT/\zeta}d\boldsymbol{B}(\tau). \tag{X1.7.19}$$

Notice that the second equation is a closed equation for the momentum. In the $\zeta \to \infty$ limit, $\boldsymbol{p}_0$ relaxes to its equilibrium distribution extremely rapidly. This is what we can intuitively expect: the velocity field equilibrates very quickly around the Brownian particle.

Simple order-by-order calculation to the second order gives

$$d\boldsymbol{r} = \epsilon(\boldsymbol{p}_0(t)/m)d\tau + \epsilon^2(1 - e^{-\tau/m})\boldsymbol{F}d\tau, \tag{X1.7.20}$$

$$= (\boldsymbol{p}_0(t)/m)dt + (\boldsymbol{F}/\zeta)dt. \tag{X1.7.21}$$

In the last line we have discarded the transcendentally small term $e^{-\zeta t/m}$. The first term is a noise: its correlation function reads (let us consider only its x-component) (assuming that $\zeta \gg 1$)

$$\left\langle \frac{p_0(t)}{m}\frac{p_0(s)}{m}\right\rangle = \frac{k_BT}{m}e^{-\zeta|t-s|/m}. \tag{X1.7.22}$$

In the $\zeta \to \infty$ limit, this becomes proportional to the delta function, because $(1/2\tau)e^{-|t|/\tau}$ becomes very sharp in the $\tau \to 0$ limit, but its area below the graph is always unity. Thus, we know

$$\frac{k_BT}{m}e^{-\zeta|t-s|/m} \simeq \frac{2k_BT}{\zeta}\delta(t-s). \tag{X1.7.23}$$

We have arrived at

$$d\boldsymbol{x} = \frac{\boldsymbol{F}}{\zeta} + \sqrt{\frac{2k_BT}{\zeta}}d\boldsymbol{B}(t). \tag{X1.7.24}$$

X1.7.5 From Kramers to Smoluchowski equation: using Fokker-Planck equation

Let us reduce (6.3.10) to (6.3.12). We can easily guess to the lowest order approximation

$$P(\boldsymbol{x},\boldsymbol{p},t) \simeq P_e(\boldsymbol{p})\rho(\boldsymbol{x},t), \tag{X1.7.25}$$

where P_e is the Maxwell distribution (→**2.3.2**). This is easy to guess, but we should get this approximation systematically.

Our starting point is the Fokker-Planck equation (Kramers equation, →**X1.7.3**)

$$\frac{\partial P(x,p,t)}{\partial t} = -\frac{\partial}{\partial x}\frac{p}{m}P - \frac{\partial}{\partial p}\left[-\frac{\zeta}{m}p + F - k_BT\zeta\frac{\partial}{\partial p}\right]P. \tag{X1.7.26}$$

That is, in terms of the stretched time,

$$\frac{\partial}{\partial \tau}P = \frac{\partial}{\partial p}\left[\frac{1}{m}p + k_BT\frac{\partial}{\partial p}\right]P + \epsilon\left[-\frac{\partial}{\partial x}\frac{p}{m} - \frac{\partial}{\partial p}F\right]P. \tag{X1.7.27}$$

Expanding naively as $P = P_0 + \epsilon P_1 + \epsilon^2 P_2 +$, we can solve the equation order by order. P_0 has the structure $P_0 = f(x)p_0(p,t)$ We immediately realize that P_1 and P_2 have secular terms proportional to τ. After renormalizing $f(x)$ and absorbing the secular terms, we obtain, as the renormalization group equation, the Smoluchowski equation after integrating the momentum out.[73]

X1.8 Elementary methods to solve Fokker-Planck equations

This section briefly discusses how to solve the Fokker-Planck equation. It is a parabolic equation, so the general approaches for parabolic equations work. The relation to the (Wick rotated) Schrödinger equation (→**6.3.12**) is useful, because the latter is much better studied than the Fokker-Planck equation.

X1.8.1 Auxiliary conditions for Fokker-Planck equation

Let us consider the following general Fokker-Planck equation. The Fokker-Planck equation is a linear parabolic partial differential equation:

$$\frac{\partial P(\boldsymbol{x},t)}{\partial t} = -\nabla\cdot\left[\boldsymbol{a} - D\nabla\right]P(\boldsymbol{x},t), \tag{X1.8.1}$$

---

[73] This renormalization group approach is the best; see an introductory review "Renormalization and Asymptotics," Int. J. Mod. Phys. B **14**, 1327-1361 (2000).

where D is a symmetric positive semidefinite matrix. We consider it in a spatial domain $\mathcal{D}$ for $t \geq 0$. To solve (X1.8.1) we need *auxiliary conditions*: initial and boundary conditions.

We know that (X1.8.1) may be interpreted as the conservation law of probability and the probability flux $\boldsymbol{J}$ is given by (→**3.5.7**)

$$\boldsymbol{J} = \boldsymbol{a}P - D\nabla P. \tag{X1.8.2}$$

It is a first order differential equation in time, so the initial value of P on D is needed.

On the boundary $\partial\mathcal{D}$ of the domain $\mathcal{D}$, we need a boundary condition. An intuitive and practical way to understand what sort of boundary conditions we can impose is to discretize (X1.8.1) on a space-time lattice and ask whether we can uniquely obtain $P(\boldsymbol{x}, t)$. Typical boundary conditions are the *no flux condition* (→**X1.8.2**), the *absorbing boundary condition* (→**X1.8.3**), and the *periodic boundary condition* (→**X1.8.4**):[74] If the domain is not bounded, we need some condition to check the possible wild growth of the solution (→**X1.8.5**).

X1.8.2 No flux condition

That is, at the boundary, the outward normal component of the flux $\boldsymbol{J} = [\boldsymbol{a} - D\nabla]P$ vanishes. The physical meaning of the flux tells us that this condition implies that the particles cannot get out of the domain: the particles are reflected by the boundary. Therefore, this condition is also called the *reflecting boundary condition.*[75]

X1.8.3 Absorbing boundary condition

$P = 0$ at the boundary implies that there is no particle at the boundary (infinitesimally close inside the boundary). This implies that the particle is killed there (i.e., the boundary absorbs the coming particle). This condition destroys the conservation of probability, so if (X1.8.1) is interpreted as the equation governing the probability density, we cannot impose this condition.[76] However, if (X1.8.1) is interpreted as the equation governing the concentration (particle number density) of noninteracting particles (as in the ideal gas or solution) (particle interpretation), we may impose this condition.[77]

The crucial point is that the flux must be continuous where there is no annihilation or creation of particles in the particle interpretation of (X1.8.1) or the conservation of probability must not be violated in the probability interpretation.

X1.8.4 Periodic boundary condition

We must impose the continuity of the flux across the boundary (i.e., the normal component of $\boldsymbol{J}$ is continuous across the boundary). To have the continuity of the flux, we must also impose the continuity of P itself. These two conditions are the periodic boundary condition.
Exercise. Consider the 1D Fokker-Planck equation, and discretize it using the simplest Euler scheme.

---

[74] If the diffusion matrix D becomes singular at the boundary, the problem becomes complicated, so here we ignore such cases, and assume that D is positive definite.

[75] This is a special case of the so-called *Robin condition.* Neumann conditions are usually not discussed in our context. However, the no flux condition is nothing but a mere homogeneous Neumann condition, if the Fokker-Planck equation is conveniently rewritten as an equation for $Q = e^{\beta V}P$ (try to rewrite the Fokker-Planck equation as an equation for Q).

[76] except for the escape problem or the first passage problems. See, e.g., **X1.9.5**.

[77] This is a *homogeneous Dirichlet condition.* We may impose inhomogeneous conditions as well.

(1) Check that the above conditions imposed at time $t = 0$ indeed allows us to compute the solution at the next time step uniquely.
(2) Suppose the diffusion constant is discontinuous at $x = 0$ (e.g., $D = 1$ for $x < 0$ and $D = 2$ for $x > 0$). What kind of extra condition should we impose at this singular point? □

X1.8.5 Unbounded domain

If the domain $\mathcal{D}$ is not bounded, we must require that the solution is bounded.[78]

Exercise. The diffusion equation may be regarded as the simplest Fokker-Planck equation. Consider the equation in a semiinfinite domain $x \geq 0$ with the condition $P(0,t) = 0$ and $P(x,0) = 0$. There is a solution $P(x,t) \neq 0$ apparently violating the conservation law. That is why at least some condition must be imposed on an unbounded domain. Can you explain why such a bizarre thing is ever possible? □

X1.8.6 Change of spatial coordinates

Sometimes it is convenient to change the independent variables of the Fokker-Planck equation

$$\frac{\partial P(\boldsymbol{x},t)}{\partial t} = \left(-\sum_i \frac{\partial}{\partial x_i} a_i + \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}\right) P \tag{X1.8.3}$$

from $\boldsymbol{x}$ to $\boldsymbol{y}$. Through this coordinate change, the density distribution function P must also be transformed to P': $P(\boldsymbol{x},t)d\boldsymbol{x} = P'(\boldsymbol{y},t)d\boldsymbol{y}$, that is

$$P(\boldsymbol{x},t) = JP'(\boldsymbol{y},t), \tag{X1.8.4}$$

where J is the Jacobian: $J \equiv det(d\boldsymbol{y}/d\boldsymbol{x})$. The Fokker-Planck equation (X1.8.5) must be transformed to another Fokker-Planck equation:

$$\frac{\partial P'(\boldsymbol{y},t)}{\partial t} = \left(-\sum_i \frac{\partial}{\partial y_i} a'_i + \sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j} D'_{ij}\right) P'. \tag{X1.8.5}$$

To relate the coefficients in these two equations, the technically the smartest way is to go back to the corresponding stochastic differential equations. The stochastic differential equation corresponding to (X1.8.3) is (→**X1.7.1**)

$$dx_i = a_i dt + \sum_k \sigma_{ik} dB_k \tag{X1.8.6}$$

with $D_{ij} = \sum_k \sigma_{ik}\sigma_{jk}$. From this, with the aid of Itô's lemma (→**X1.5.10**)

$$dy_i = \sum_j \frac{\partial y_i}{\partial x_j} dx_j + \frac{1}{2}\sum_{jk} \frac{\partial^2 y_i}{\partial x_j \partial x_k} dx_j dx_k \tag{X1.8.7}$$

$$= \sum_j \frac{\partial y_i}{\partial x_j}[a_j dt + \sum_k \sigma_{jk} dB_k] + \frac{1}{2}\sum_{jk} \frac{\partial^2 y_i}{\partial x_j \partial x_k} \sigma_{jk} dt \tag{X1.8.8}$$

---

[78] Physically, it is sensible to require that P vanishes at infinity, but this follows if we require the boundedness of the solution. Actually, if the solution is required not to grow faster than some power of time, then the solution must be bounded. That is, a weaker condition may be imposed instead of the boundedness condition.

Therefore, we may conclude

$$a'_i = \sum_j \frac{\partial y_i}{\partial x_j} a_j + \frac{1}{2} \sum_{jk} \frac{\partial^2 y_i}{\partial x_j \partial x_k} \sigma_{jk}, \;\; D'_{ij} = \sum_{nm} \frac{\partial y_i}{\partial x_n} \frac{\partial y_j}{\partial x_m} D_{nm}. \tag{X1.8.9}$$

The last line is obtained from

$$\sigma'_{ij} = \sum_k (\partial y_i / \partial x_k) \sigma_{kj}. \tag{X1.8.10}$$

X1.8.7 Reduction to constant diffusion coefficient equation

If we look at (X1.8.10), sometimes we could make σ' constant by choosing $\boldsymbol{y}$ appropriately. Therefore, it is useful to consider a method to solve the Fokker-Planck equation with constant D_{ij}. For simplicity, let us consider a 1D Fokker-Planck equation:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x} a(x) P + \frac{\partial^2}{\partial x^2} D(x) P. \tag{X1.8.11}$$

Let us introduce a new spatial coordinate y so that (see (X1.8.10))

$$dy/dx = \sqrt{D/D(x)}. \tag{X1.8.12}$$

Then, the result of **X1.8.6** tells us that

$$\frac{\partial P'(y,t)}{\partial t} = -\frac{\partial}{\partial x} a'(y) P' + D \frac{\partial^2}{\partial y^2} P' \tag{X1.8.13}$$

with

$$a'(y) = \sqrt{\frac{D}{D(x)}} \left[a(x) - \frac{1}{2} \frac{dD(x)}{dx} \right]. \tag{X1.8.14}$$

The Jacobian J reads $J = \sqrt{D(x)/D}$.

X1.8.8 Equilibrium solution

If the diffusion matrix is a constant matrix $D(x) = DI$ and if $\boldsymbol{a}$ is a gradient field $\boldsymbol{a} = -\nabla U$, then the time independent solution must satisfy

$$\left[\frac{\partial}{\partial \boldsymbol{y}} U \right] P + D \frac{\partial}{\partial \boldsymbol{y}} P = \text{const.} \tag{X1.8.15}$$

If the problem is considered in the whole space, P and its derivatives must vanish at infinity, so the constant in the equation must vanish. Therefore,

$$P_{eq} \propto e^{-U/D} \tag{X1.8.16}$$

is the time independent (supposedly equilibrium) solution.

X1.8.9 Ornstein-Uhlenbeck process

There are not many examples of Fokker-Planck equations whose time-dependent solutions can be obtained analytically. One rare example is the Fokker-Planck equation corresponding to the Ornstein-Uhlenbeck process defined by

$$d\boldsymbol{p} = -(\zeta/m)\boldsymbol{p} + \sqrt{2k_BT\zeta}d\boldsymbol{B}. \tag{X1.8.17}$$

The corresponding Fokker-Planck equation is

$$\frac{\partial P}{\partial t} = \left[\frac{\partial}{\partial \boldsymbol{p}}\frac{\zeta}{m}\boldsymbol{p} + \zeta k_BT\frac{\partial^2}{\partial \boldsymbol{p}^2}\right] P. \tag{X1.8.18}$$

Let us impose the initial condition $P(\boldsymbol{p}, 0) = \delta(\boldsymbol{p} - \boldsymbol{p}_0)$ (That is, we wish to solve the transition probability (density) $P(\boldsymbol{p}, t|\boldsymbol{p}_0, 0)$.).

Perhaps the easiest method to solve the problem (without the use of the Langevin equation) is to use the Fourier transformation. We have the following correspondence

$$\partial/\partial\boldsymbol{p} \leftrightarrow i\boldsymbol{k},\ \boldsymbol{p} \leftrightarrow -i\partial/\partial\boldsymbol{k}. \tag{X1.8.19}$$

Therefore, we obtain

$$\frac{\partial \hat{P}}{\partial t} = \frac{\zeta}{m}\boldsymbol{k}\frac{\partial}{\partial \boldsymbol{k}}\hat{P} - \zeta k_BT\boldsymbol{k}^2\hat{P}. \tag{X1.8.20}$$

This is a quasilinear first order partial differential equation, so it can be solved by the method of characteristics.

The characteristic equation reads

$$\frac{dt}{1} = \frac{dk_i}{\zeta k_i/m} = \frac{dP}{\zeta k_BT\boldsymbol{k}^2P}. \tag{X1.8.21}$$

Therefore, the solution with the initial condition above reads

$$\hat{P} = \exp\{-i\boldsymbol{k}\cdot\boldsymbol{p}_0e^{-\zeta t/m} - \zeta k_BTk^2(1 - e^{-2\zeta t/m})/2(\zeta/m)\}. \tag{X1.8.22}$$

That is,

$$P(\boldsymbol{p}, t|\boldsymbol{p}_0, 0) = \left(\frac{1}{2\pi mk_BT(1 - e^{-2\zeta t/m})}\right)^{3/2} \exp\left[-\frac{(\boldsymbol{p} - e^{-\zeta t/m}\boldsymbol{p}_0)^2}{2mk_BT(1 - e^{-2\zeta t/m})}\right]. \tag{X1.8.23}$$

However, in this case, there is a much simpler method to obtain the solution as seen in **X1.8.10**.

X1.8.10 Ornstein-Uhlenbeck process: direct approach from stochastic differential equation

It would be illuminating to obtain this result without the use of partial differential equation. The stochastic differential equation (6.3.26) describing the Ornstein-Uhlenbeck process is a linear equation, so the process must be a Gaussian process. Therefore, we have only to compute the mean and the covariance.

From (6.3.26) we obtain

$$d\langle \boldsymbol{p}\rangle = -(\zeta/m)\langle \boldsymbol{p}\rangle \ \Rightarrow \langle \boldsymbol{p}(t)\rangle = \boldsymbol{p}_0 e^{-\zeta t/m}. \tag{X1.8.24}$$

Also for each component, say for the x-component, we can obtain

$$d\langle (p_x - p_{x0}e^{-\zeta t/m})^2\rangle = -2(\zeta/m)\langle (p_x - p_{x0}e^{-\zeta t/m})^2\rangle dt + \zeta k_B T dt. \tag{X1.8.25}$$

From this we can almost immediately read off the needed variance.

X1.8.11 Relation of Smoluchowski equation to Schrödinger equation

If a Smoluchowski equation can be mapped to a Schrödinger equation, then all the (well-developed) standard methods to solve the Schrödinger equation may be applied. They include eigenfunction expansion, variational methods, and WKB methods. These are very standard topics, so we will not discuss them here.

The Smoluchowski equation

$$\partial P/\partial t = \nabla \cdot [(\nabla U)P] + D\Delta P \tag{X1.8.26}$$

may be mapped to the following Schrödinger equation (with imaginary time) (→**X1.10.4**, **X1.10.5**)

$$\partial \psi/\partial t = D\Delta\psi + V\psi \tag{X1.8.27}$$

with

$$\psi(\boldsymbol{x},t) = P(\boldsymbol{x},t|\boldsymbol{x}_0)e^{[U(\boldsymbol{x})-U(\boldsymbol{x}_0)]/2D}, \ \ V = -\frac{1}{2}\left[\frac{1}{2D}\nabla U\cdot\nabla U - \Delta U\right]. \tag{X1.8.28}$$

Notice that the potential condition for the streaming term of the Smoluchowski equation is needed for this rewriting.

Although the above relation is discussed in conjunction to path integrals later (→**X1.10.5**), it is easy to check it directly by putting (6.3.33) into the Schrödinger equation.

X1.9 Backward Equations

X1.9.1 Transition probability and Fokker-Planck equation

Let us consider the stochastic process defined by the following Langevin equation:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{a}(\boldsymbol{x},t) + \sigma(\boldsymbol{x},t)\boldsymbol{w}, \tag{X1.9.1}$$

where $\boldsymbol{w}$ is a normalized Gaussian white noise (say, d-dimensional vector). Mathematically, we interpret this as (→**X1.5.7**)

$$d\boldsymbol{x} = \boldsymbol{a}(\boldsymbol{x},t)dt + \sigma(\boldsymbol{x},t)d\boldsymbol{B}, \tag{X1.9.2}$$

where $\boldsymbol{B}$ is the d-dimensional Wiener process (→**X1.4.9**). The corresponding Fokker-Planck equation is, with the definition of $D = \sigma\sigma^T$, (→**X1.7.1**)

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial \boldsymbol{x}}\left[\boldsymbol{a} - \frac{1}{2}\frac{\partial}{\partial \boldsymbol{x}}D\right]P. \tag{X1.9.3}$$

If the interpretation of the formula seems ambiguous, refer to its detailed expression (X1.7.4).

Let the initial condition given at time t_0 be $P(\boldsymbol{x}, t_0) = \delta(\boldsymbol{x} - \boldsymbol{x}_0)$, and the solution to (X1.9.3) with an appropriate boundary conditions be written as $P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0)$. It is clear that $P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0)d\boldsymbol{x}$ is the probability of finding the sample process to be in the volume element $d\boldsymbol{x}$ at time t under the condition that at time t_0 it started from $\boldsymbol{x}_0$. Thus, we may understand that the Fokker-Planck equation governs the transition probabilities as discussed already in **X1.7.1**.

X1.9.2 Kolmogorov's backward equation

Let $f(\boldsymbol{x})$ be a well-behaved function of $\boldsymbol{x}$. We wish to study its expectation value (average) $E_{\boldsymbol{x}_0}[f](t)$ at time t, when the initial value of $\boldsymbol{x}$ is $\boldsymbol{x}_0$:

$$E_{\boldsymbol{x}_0}[f](t) \equiv \int d\boldsymbol{x}\, f(\boldsymbol{x})P(\boldsymbol{x}, t|\boldsymbol{x}_0, 0). \tag{X1.9.4}$$

To obtain the equation governing $E_{\boldsymbol{x}_0}[f](t)$, it is convenient to find an equation that governs $P(\boldsymbol{x}, t|\boldsymbol{x}_0, 0)$ as a function of t and $\boldsymbol{x}_0$. We can use the Chapman-Kolmogorov equation (→**3.1.10**);

$$P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0) = \int d\Delta\boldsymbol{x} P(\boldsymbol{x}, t|\boldsymbol{x}_0 + \Delta\boldsymbol{x}, t_0 + \Delta t)P(\boldsymbol{x}_0 + \Delta\boldsymbol{x}, t_0 + \Delta t|\boldsymbol{x}_0, t_0). \tag{X1.9.5}$$

Here, $P(\boldsymbol{x}_0 + \Delta\boldsymbol{x}, t_0 + \Delta t|\boldsymbol{x}_0, t_0)$ corresponds to the probability density $\phi(\Delta\boldsymbol{x}|\boldsymbol{x}_0)$ for the transition step $\Delta\boldsymbol{x}$ we have encountered in **??** (Exercise 1):

$$P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0) = \int d\Delta\boldsymbol{x} P(\boldsymbol{x}, t|\boldsymbol{x}_0 + \Delta\boldsymbol{x}, t_0 + \Delta t)\phi(\Delta\boldsymbol{x}|\boldsymbol{x}_0). \tag{X1.9.6}$$

Taylor expanding $P(\boldsymbol{x}, t|\boldsymbol{x}_0 + \Delta\boldsymbol{x}, t_0 + \Delta t)$ in the above equation, we obtain

$$\begin{aligned} 0 &= \int d\Delta\boldsymbol{x}\,\phi(\Delta\boldsymbol{x}|\boldsymbol{x}_0)\left[\frac{\partial P}{\partial t_0} + \sum_i \Delta x_i \frac{\partial P}{\partial x_{0i}} + \frac{1}{2}\sum_{ij}\Delta x_i \Delta x_j \frac{\partial^2 P}{\partial x_{0i} x_{0j}}\right] \\ &= \frac{\partial P}{\partial t_0} + \sum_i a_i(\boldsymbol{x}_0)\frac{\partial P}{\partial x_{0i}} + \frac{1}{2}\sum_{ij} D_{ij}(\boldsymbol{x}_0)\frac{\partial^2 P}{\partial x_{0i}\partial x_{0j}}. \end{aligned} \tag{X1.9.7}$$

Therefore,

$$\frac{\partial P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0)}{\partial t_0} = -\sum_i a_i \frac{\partial P}{\partial x_{0i}} - \frac{1}{2}\sum_{ij} D_{ij}\frac{\partial^2 P}{\partial x_{0i}\partial x_{0j}}. \tag{X1.9.8}$$

This is the backward equation.[79] Note that all the independent variables are the variables describing the initial condition (so are the variables in a_i and D_{ij}).

Given a Fokker-Planck equation, it is easy to write down the corresponding backward equation: identifying the coefficient functions, the drift term $\boldsymbol{a}$ and the diffusion term D, immediately we can write down (X1.9.8) with switching the variables to the initial variables.

---

[79] (X1.9.8) is sometimes called *Kolmogorov's backward equation*, but more often (6.4.13) is called so.

X1.9.3 Equation for expectation values for time homogeneous process

Let $f(\boldsymbol{x})$ be a well-behaved function of $\boldsymbol{x}$ and

$$E_{\boldsymbol{x}_0}[f](t) \equiv \int d\boldsymbol{x}\, f(\boldsymbol{x}) P(\boldsymbol{x}, t|\boldsymbol{x}_0, 0). \tag{X1.9.9}$$

Its time derivative is

$$\frac{\partial}{\partial t} E_{\boldsymbol{x}_0}[f](t) = \int d\boldsymbol{x}\, f(\boldsymbol{x}) \frac{\partial}{\partial t} P(\boldsymbol{x}, t|\boldsymbol{x}_0, 0). \tag{X1.9.10}$$

Let us assume that the stochastic process governed by (X1.9.2) is time homogeneous (that is, the coefficients do not depend on time explicitly, and the system is independent of the absolute time). Notice that P is homogeneous in t and is a function of $t - t_0$, so

$$\frac{\partial}{\partial t} P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0) = -\frac{\partial}{\partial t_0} P(\boldsymbol{x}, t|\boldsymbol{x}_0, t_0). \tag{X1.9.11}$$

Therefore, (6.4.11) reads with the aid of the backward equation in **X1.9.2**

$$\frac{\partial}{\partial t} E_{\boldsymbol{x}_0}[f](t) = \int d\boldsymbol{x}\, f(\boldsymbol{x}) \left[\sum_i a_i \frac{\partial}{\partial x_{0i}} + \frac{1}{2} \sum_{ij} D_{ij} \frac{\partial^2}{\partial x_{0i} \partial x_{0j}} \right] P(\boldsymbol{x}, t|\boldsymbol{x}_0, 0). \tag{X1.9.12}$$

That is, the solution to the following equation called *Kolmogorov's backward equation*:

$$\frac{\partial}{\partial t} u(\boldsymbol{x}, t) = \left[\sum_i a_i(\boldsymbol{x}) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{ij} D_{ij}(\boldsymbol{x}) \frac{\partial^2}{\partial x_i \partial x_j} \right] u(\boldsymbol{x}, t). \tag{X1.9.13}$$

with the initial condition $u(\boldsymbol{x}, 0) = f(\boldsymbol{x})$ is given by

$$u(\boldsymbol{x}, t) = E_{\boldsymbol{x}}[f](t) = E_{\boldsymbol{x}}[f(\boldsymbol{x}(t))]. \tag{X1.9.14}$$

Here, t is the time to observe the expectation value (that is, *now*), and $\boldsymbol{x}$ is the *starting position* of all the paths at $t = 0$. This is the most important use of the backward equation. This is also a special case of the Feynman-Kac formula (→**X1.10.1**).
Exercise. Write down the backward equation for the Ornstein-Uhlenbeck process (→**X1.8.9**).
□

X1.9.4 Streamlined demonstration of (6.4.13)

Let $u(\boldsymbol{x}, t)$ be the solution to the backward equation (→**X1.9.3**)

$$\frac{\partial u}{\partial t} = \sum_i a_i \frac{\partial}{\partial x_i} u + \frac{1}{2} \sum_{ij} D_{ij} \frac{\partial^2}{\partial x_i \partial x_j} u \tag{X1.9.15}$$

with the initial condition $u(\boldsymbol{x}, 0) = f(\boldsymbol{x})$ (and with appropriate boundary conditions).[80] Then, for any fixed time $T > 0$ it is 'easy'[81] to demonstrate that the expectation value of

---

[80] Actually, if u grows only algebraically with $|\boldsymbol{x}|$, any solution with the initial condition $u(\boldsymbol{x}, 0) = f(\boldsymbol{x})$ may be used.

[81] Formally, straightforward as seen here. Mathematically, some care is needed, but essentially the formal calculation is justified.

$g(t) \equiv u(T-t, \boldsymbol{x}(t))$ is not t-dependent with the aid of Itô's lemma (→**X1.5.10**). Here, $\boldsymbol{x}(t)$ obeys the stochastic differential equation associated with the above backward equation (i.e., (X1.9.2)) with the initial condition $\boldsymbol{x}(0) = \boldsymbol{x}$.

Using Itô's lemma (→**X1.5.10**), we can differentiate g as follows.

$$\begin{aligned} dg(t) &= -\left.\frac{\partial u(s, x(t))}{\partial s}\right|_{s=T-t} dt + \sum_i \frac{\partial u}{\partial x_i} dx_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 u}{\partial x_i \partial x_j} dx_i dx_j, \quad \text{(X1.9.16)} \\ &= -\left[\sum_i a_i \frac{\partial}{\partial x_i} u + \frac{1}{2}\sum_{ij} D_{ij} \frac{\partial^2}{\partial x_i \partial x_j} u\right] dt \\ &\quad + \sum_i \frac{\partial u}{\partial x_i}(a_i dt + \sum_k \sigma_{ik} dB_k) + \frac{1}{2}\sum_{i,j} D_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j}, \\ &\qquad \text{(X1.9.17)} \\ &= \sum_{i,j} \frac{\partial u}{\partial x_i} \sigma_{ij} dB_j. \quad \text{(X1.9.18)} \end{aligned}$$

We have used $D_{ij} = \sum_k \sigma_{ik}\sigma_{jk}$. Therefore, its average vanishes and $E_{\boldsymbol{x}}[u(T-t, \boldsymbol{x}(t))]$ is time independent.

Setting $t = 0$, we have

$$E_{\boldsymbol{x}}[u(T-t, \boldsymbol{x}(t))] = u(T, \boldsymbol{x}). \tag{X1.9.19}$$

On the other hand, from the initial condition for u

$$E_{\boldsymbol{x}}[u(0, \boldsymbol{x}(T))] = E_{\boldsymbol{x}}[f(\boldsymbol{x}(T))]. \tag{X1.9.20}$$

Therefore, indeed,

$$u(\boldsymbol{x}, t) = E_{\boldsymbol{x}}[f(\boldsymbol{x}(t))] \tag{X1.9.21}$$

is the solution to the initial value problem of (6.5.1) as asserted in **X1.9.3**.

X1.9.5 First passage or exit time

Let U be a set in the domain of the stochastic process $\{\boldsymbol{x}(t)\}$ defined by the stochastic differential equation (X1.9.2). Let $P(\boldsymbol{y}, t|\boldsymbol{x}, 0)$ be the solution to the corresponding Fokker-Planck equation (the forward equation associated with (X1.9.8)) with the initial condition $P(\boldsymbol{y}, 0|\boldsymbol{x}, 0) = \delta(\boldsymbol{y} - \boldsymbol{x})$ and the absorbing boundary condition (= homogeneous Dirichlet condition, →**X1.8.1**) at ∂U. Let $\boldsymbol{x} \in U$ and

$$G(\boldsymbol{x}, t) \equiv \int_U d\boldsymbol{y} P(\boldsymbol{y}, t|\boldsymbol{x}, 0). \tag{X1.9.22}$$

This is the probability of finding $\boldsymbol{x}(t)$ that started at time $t = 0$ from $\boldsymbol{x}$ in U without ever getting out of U because of the absorbing boundary condition (→**X1.9.2**). Therefore, this is the probability for the exit time T from U not to be smaller than t:

$$P(T > t|\boldsymbol{x}(0) = \boldsymbol{x}) = G(\boldsymbol{x}, t). \tag{X1.9.23}$$

From the backward equation, we see that

$$\frac{\partial}{\partial t} G(\boldsymbol{x}, t) = \left[\sum_i a_i \frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{ij} D_{ij} \frac{\partial^2}{\partial x_i \partial x_j}\right] G(\boldsymbol{x}, t) \tag{X1.9.24}$$

with the initial condition $G(\boldsymbol{x},0) = 1_U(\boldsymbol{x})$ which is the indicator (→**??**) of U, and the boundary condition $G = 0$ at ∂U.

Exercise 1. Let $f(T)$ be a smooth function of the exit time for a particle starting from $\boldsymbol{x}$ in U as above. Show that its expectation value for U is given by

$$\langle f(T)\rangle_{\boldsymbol{x}} = -\int_0^\infty dt\, f(t)\frac{\partial}{\partial t}G(\boldsymbol{x},t). \tag{X1.9.25}$$

In particular, the mean exit time $\langle T\rangle_{\boldsymbol{x}}$ for $\boldsymbol{x}(t)$ starting from $\boldsymbol{x}$ in U is given by

$$\langle T\rangle_{\boldsymbol{x}} = \int_0^\infty G(\boldsymbol{x},t)dt. \tag{X1.9.26}$$

□

Exercise 2. Derive the differential equation that governs $\langle T\rangle_{\boldsymbol{x}}$ in Exercise 1 (note that $G(\boldsymbol{x},0) = 1$ and $G(\boldsymbol{x},\infty) = 0$):

$$\left[\sum_i a_i\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{ij} D_{ij}\frac{\partial^2}{\partial x_i\partial x_j}\right]\langle T\rangle_{\boldsymbol{x}} = -1. \tag{X1.9.27}$$

The boundary condition is a homogeneous Dirichlet condition on ∂U. □

Exercise 3. Let us consider the simplest case: the 1d Wiener process (but starting from some point x at time $t = 0$).

(1) Find the mean exit time from $[-1,2]$ of the Wiener process starting from x ($x \in [-1,2]$).

(2) Consider an interval $[-L,0]$. In the $L \to \infty$ limit, however close the starting point is to the origin, the mean exit time seems to diverge. Is this intuitively plausible?

(3) Now, let us add a drift term, so that the Langevin equation reads $dx = vdt + dB$. Find the mean exit time from $[-1,1]$, assuming the starting point x is in this interval. □

Exercise 4. For the following overdamped Brownian particle on the x-axis

$$dx = -(U'(x)/\zeta)dt + \sqrt{2k_BT/\zeta}dB(t) \tag{X1.9.28}$$

find the mean exit time from $[-1,1]$, assuming that the particle is initially at the origin. □

X1.9.6 Potential barrier crossing: Kramers formula revisited

We studied this problem already in **3.6.2**. Let us compute the mean exit time from a potential well. Our system is described by

$$dx = -(U'(x)/\zeta)dt + \sqrt{2k_BT/\zeta}dB(t), \tag{X1.9.29}$$

so $a = -U'/\zeta$ and $D = k_BT/\zeta$ in (X1.9.2).

Let us assume that the potential U has two wells around A and C with the separating barrier around B. We are interested in the exit time from the left well A into the right one C. Therefore, we consider the exit time from $(-\infty, X)$, where X is some point between B and C. Solving (6.4.21), we have

$$\langle T\rangle_A = \beta\zeta\int_A^X dx'\int_{-\infty}^{x'} dy e^{-\beta U(y)}e^{\beta U(x')}. \tag{X1.9.30}$$

We must introduce some approximations under the condition that the barrier crossing is not easy (the peak at B is high). We may use the same approximation as in **3.6.2**, and it is an exercise to recover Kramers' result there.

X1.10 Feynman-Kac formula

X1.10.1 Feynman-Kac formula

Let us try to extend **X1.9.4**. Instead of (X1.9.15), consider

$$\frac{\partial u}{\partial t} = \sum_i a_i \frac{\partial}{\partial x_i} u + \frac{1}{2} \sum_{ij} D_{ij} \frac{\partial^2}{\partial x_i \partial x_j} u + V(\boldsymbol{x}) u \tag{X1.10.1}$$

Here, V is a well-behaved function, $D_{ij} = \sum_k \sigma_{ik}\sigma_{jk}$, and we consider the following stochastic differential equation:

$$dx_i = a_i dt + \sum_j \sigma_{ij} dB_j. \tag{X1.10.2}$$

This stochastic differential equation is compatible with the backward equation (→**X1.9.3**) given by $V = 0$ in (X1.10.1). Let $\boldsymbol{x}(t)$ be the solution to this equation with the initial condition $\boldsymbol{x}(0) = \boldsymbol{x}$. Let u be a solution to (X1.10.1) with the initial condition $u(0, \boldsymbol{x}) = f(\boldsymbol{x})$.

Consider[82] (similar to **X1.9.4**)

$$\begin{aligned} & d\left[u(T-t, \boldsymbol{x}(t)) \exp\left(\int_0^t V(\boldsymbol{x}(s))ds\right)\right] \\ = & \left[-\frac{\partial}{\partial t}u + \sum_i a_i \frac{\partial}{\partial x_i}u + \frac{1}{2} D_{ij} \frac{\partial^2}{\partial x_i \partial x_j} u + Vu\right] \exp\left(\int_0^t V(\boldsymbol{x}(s))ds\right) dt + Q, \end{aligned} \tag{X1.10.3}$$

where Q is the term proportional to dB. Therefore,

$$dE_{\boldsymbol{x}}\left[u(T-t, \boldsymbol{x}(t)) \exp\left(\int_0^t V(\boldsymbol{x}(s))ds\right)\right] = 0. \tag{X1.10.4}$$

That is, this average is time independent. Therefore,

$$E_{\boldsymbol{x}}\left[u(0, \boldsymbol{x}(T)) \exp\left(\int_0^T V(\boldsymbol{x}(s))ds\right)\right] = E_{\boldsymbol{x}}[u(T, \boldsymbol{x}(0))], \tag{X1.10.5}$$

that is,

$$u(t, \boldsymbol{x}) = E_{\boldsymbol{x}}\left[f(\boldsymbol{x}(t)) \exp\left(\int_0^t V(\boldsymbol{x}(s))ds\right)\right]. \tag{X1.10.6}$$

This is the celebrated *Feynman-Kac formula.*[83]

The implication of the formula is this: prepare a sample path $\boldsymbol{x}(t, \omega)$ obeying the stochastic equation (X1.10.2) starting from $\boldsymbol{x}$. Compute for each path (here ω is the stochastic parameter specifying each sample)

$$f(\boldsymbol{x}(t, \omega)) \exp\left(\int_0^t V(\boldsymbol{x}(s, \omega)ds\right) \tag{X1.10.7}$$

and then make the ensemble average of this quantity. This procedure solves (X1.10.1) with the initial condition $u = f$ at $t = 0$.[84]

---

[82] Again, the following formal calculation as it is cannot be justified (why?), but 'essentially correct.'

[83] Mathematicians often call this the *Kac formula.* What Feynman suggested is not this formula but the one in **X1.10.3**. Kac realized the relation to the backward equation. Feynman discussed only the forward equations.

[84] If the boundary conditions are imposed, they must be interpreted in terms of the path.

X1.10.2 Feynman-Kac path integral: general case

The formula in **X1.10.1** is not so familiar to physicists. The idea of the backward equation is not popular in any case.

Let us consider the following differential with $\boldsymbol{x}(t)$ as in **X1.10.1**:

$$d\left[\delta(\boldsymbol{x}(t)-\boldsymbol{y})\exp\left(\int_0^t V(\boldsymbol{x}(s))ds\right)\right] = -\sum_i \frac{\partial}{\partial y_i}\delta(\boldsymbol{x}(t)-\boldsymbol{y})dx_i e^{\int V ds}$$
$$+\frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial y_i \partial y_j}\delta(\boldsymbol{x}(t)-\boldsymbol{y})dx_i dx_j e^{\int V ds} + V(\boldsymbol{x})\delta(\boldsymbol{x}(t)-\boldsymbol{y})dx_i e^{\int V ds}dt. \quad \text{(X1.10.8)}$$

Let us define

$$\phi(\boldsymbol{y},t|\boldsymbol{x}) \equiv E_{\boldsymbol{x}}\left[\delta(\boldsymbol{x}(t)-\boldsymbol{y})\exp\left(\int_0^t V(\boldsymbol{x}(s))ds\right)\right]. \quad \text{(X1.10.9)}$$

Then, averaging (X1.10.8) over the stochastic process, we get

$$\frac{\partial}{\partial t}\phi(\boldsymbol{y},t|\boldsymbol{x}) = \left[\sum_i \frac{\partial}{\partial y_i}(-a_i) + \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial y_i \partial y_j}D_{ij} + V\right]\phi(\boldsymbol{y},t|\boldsymbol{x}) \quad \text{(X1.10.10)}$$

with the initial condition $\phi(\boldsymbol{y},0|\boldsymbol{x}) = \delta(\boldsymbol{y}-\boldsymbol{x})$. That is, the solution to (X1.10.10) under the initial condition $\phi(\boldsymbol{y},0|\boldsymbol{x}) = \delta(\boldsymbol{y}-\boldsymbol{x})$ may be expressed as (X1.10.9), or more pictorially

$$\phi(\boldsymbol{y},t|\boldsymbol{x}) = \int_{\boldsymbol{x}(0)=\boldsymbol{x}}^{\boldsymbol{x}(t)=\boldsymbol{y}} \mathcal{D}_{FP}[\boldsymbol{x}(t)]\exp\left[\int_0^t ds\, V(\boldsymbol{x}(s))\right]. \quad \text{(X1.10.11)}$$

This is called the *Feynman-Kac path integral* over the stochastic process compatible with the Fokker-Planck equation that is (X1.10.10) without V. In this formula

$$\int_{\boldsymbol{x}(0)=\boldsymbol{x}}^{\boldsymbol{x}(t)=\boldsymbol{y}} \mathcal{D}_{FP}[\boldsymbol{x}(t)] \quad \text{(X1.10.12)}$$

symbolically expresses that we average the integrand with respect to all the sample paths of the stochastic process starting from $\boldsymbol{x}$ and ending up at $\boldsymbol{y}$ whose density distribution obeys the Fokker-Planck equation.

X1.10.3 Feynman-Kac path integral: diffusion case

The most basic Feynman-Kac path integral is: the solution to

$$\frac{\partial \psi}{\partial t} = D\Delta\psi + V\psi \quad \text{(X1.10.13)}$$

with the initial condition $\psi(\boldsymbol{x},0) = \delta(\boldsymbol{x}-\boldsymbol{x}_0)$ is given by

$$\psi(\boldsymbol{x},t) = \int_{\boldsymbol{x}(0)=\boldsymbol{x}_0}^{\boldsymbol{x}(t)=\boldsymbol{x}} \mathcal{D}_W[\boldsymbol{x}(t)]\exp\left[\int_0^t ds\, V(\boldsymbol{x}(s))\right], \quad \text{(X1.10.14)}$$

where $\boldsymbol{x}(t)$ is the solution to

$$d\boldsymbol{x} = \sqrt{2D}d\boldsymbol{B} \quad \text{(X1.10.15)}$$

with the initial condition $\boldsymbol{x}(0)=\boldsymbol{x}_0$.

$\mathcal{D}_W$ is called the *Wiener measure* and is a measure defined on a set of continuous functions:

$$\int_{\boldsymbol{x}(0)=\boldsymbol{x}}^{\boldsymbol{x}(t)=\boldsymbol{y}} \mathcal{D}_W[\boldsymbol{x}(t)] \tag{X1.10.16}$$

symbolically expresses that we average the integrand with respect to all the sample paths of the (d-dimensional) Wiener process (→**X1.4.4**) (scaled with $\sqrt{2D}$) starting from $\boldsymbol{x}$[85] and ending up at $\boldsymbol{y}$.

X1.10.4 Feynman-Kac path integral for Smoluchowski equation

Then, how can we devise a similar formula for the Smoluchowski equation (→**X1.7.3**, **3.5.5**)

$$\frac{\partial \phi}{\partial t} = D\Delta\phi + \nabla\cdot[(\nabla U)\phi] \tag{X1.10.17}$$

?

The most elementary way to answer this question is to note the following relation: If we identify

$$V = -\frac{1}{2}\left[\frac{1}{2D}(\nabla U)^2 - \Delta U\right]. \tag{X1.10.18}$$

Then,

$$\psi(\boldsymbol{x},t) = e^{[U(\boldsymbol{x})-U(\boldsymbol{x}_0)]/2D}\phi(\boldsymbol{x},t), \tag{X1.10.19}$$

where ψ is the solution to (X1.10.13) with the same initial condition. That is, the solution to (X1.10.17) with the initial condition $\phi(\boldsymbol{x},0)=\delta(\boldsymbol{x}-\boldsymbol{x}_0)$ is given by

$$\phi(\boldsymbol{x},t) = e^{[U(\boldsymbol{x})-U(\boldsymbol{x}_0)]/4D}\int_{\boldsymbol{x}(0)=\boldsymbol{x}_0}^{\boldsymbol{x}(t)=\boldsymbol{x}} \mathcal{D}_W[\boldsymbol{x}(t)]\exp\left[\int_0^t ds\, V(\boldsymbol{x}(s))\right] \tag{X1.10.20}$$

with V given by (X1.10.18).

Notice that, with the aid of Itô's lemma (→**X1.5.10**),

$$\begin{aligned} \frac{1}{2D}[U(\boldsymbol{x})-U(\boldsymbol{x}_0)] &= \frac{1}{2D}\int_0^t ds\nabla U(\boldsymbol{x}(s))\circ d\boldsymbol{x}(s), && \text{(X1.10.21)}\\ &= \frac{1}{2D}\left[\int_0^t \nabla U(\boldsymbol{x}(s))\cdot d\boldsymbol{x}(s) + D\int_0^t \Delta U(\boldsymbol{x}(s)ds\right]. && \text{(X1.10.22)} \end{aligned}$$

Here, the integral in the second line is interpreted as the Itô integral. That is,

$$\phi(\boldsymbol{x},t) = \int_{\boldsymbol{x}(0)=\boldsymbol{x}_0}^{\boldsymbol{x}(t)=\boldsymbol{x}} \mathcal{D}_W[\boldsymbol{x}(t)]\exp\left[-\frac{1}{4D}\int_0^t ds\,[\nabla U(\boldsymbol{x}(s))]^2 - \frac{1}{2D}\int_0^t \nabla U\cdot d\boldsymbol{x}(s))\right]. \tag{X1.10.23}$$

In **X1.10.5** we see this is naturally related to the Onsager-Hashitsume path integral. Thus, if the reader traces the logic backward, it gives a derivation of the relation between the Schrödinger equation and the Smoluchowski equation.

---

[85] The standard Wiener process is defined as starting from $\boldsymbol{x}=0$, but here we generalize its definition slightly by allowing the starting point to be any point.

X1.10.5 Onsager-Machlup-Hashitsume path integral

We know that Onsager's principle allows us to write down the solution to the Smoluchowski equation more directly in a path integral form (→**4.1.16**, **X1.1.8**). That is the Onsager-Machlup-Hashitsume path integral: For

$$d\boldsymbol{x} = -\nabla U dt + \sqrt{2D} d\boldsymbol{B} \tag{X1.10.24}$$

formally we have

$$\phi(\boldsymbol{x}, t) = \int_{\boldsymbol{x}(0)=\boldsymbol{x}_0}^{\boldsymbol{x}(t)=\boldsymbol{x}} \delta[\boldsymbol{x}(t)] \exp\left[-\frac{1}{4D}\int_0^t ds(\dot{\boldsymbol{x}}(s) + \nabla U(\boldsymbol{x}(s))^2\right], \tag{X1.10.25}$$

where $\delta[\boldsymbol{x}(t)]$ denotes the 'uniform measure' (even weighting) on the path space.[86] This result must be equal to (X1.10.23).

Let us formally compute the integral as the stochastic integral:

$$\begin{aligned} & -\frac{1}{4D}\int_0^t ds\,[\dot{\boldsymbol{x}}(s) + \nabla U(\boldsymbol{x}(s)]^2, \\ = & -\frac{1}{4D}\left[\int_0^t \dot{\boldsymbol{x}}(s)^2 ds + \int_0^t ds\,[\nabla U(\boldsymbol{x}(s))]^2 + 2\int_0^t \nabla U \cdot d\boldsymbol{x}(s)\right]. \end{aligned} \tag{X1.10.26}$$

Therefore, if we identify (→**X1.10.3**)

$$\mathcal{D}_W[\boldsymbol{x}(t)] = \delta[\boldsymbol{x}(t)] \exp\left[-\frac{1}{4D}\int_0^t \dot{\boldsymbol{x}}(s)^2 ds\right] \tag{X1.10.27}$$

and interpret $\int \nabla U \cdot \dot{\boldsymbol{x}} dt = \int \nabla U \cdot d\boldsymbol{x}(s)$ as the Itô integral, we can mathematically properly interpret the path integral in the white noise idealization.

The result may, with the aid of (X1.10.23), be interpreted as a derivation of (X1.10.19) with (X1.10.18) with the aid of the Feynman-Kac path integral (→**X1.10.3**), which is mathematically defined without any ambiguity.

X1.11 Detailed Balance and Fokker-Planck Operator

X1.11.1 Detailed balance revisited

We have discussed that in equilibrium the most important property of the transition probability between macroscopic states (that are invariant under time-reversal) is the detailed balance condition (→**2.7.6**).

A slightly more general version of detailed balance does not require the macroscopic states to be invariant under time-reversal operation. Let a tilde imply the time reversed state (e.g., classically, $\tilde{A}$ is obtained from A by flipping the sign of momenta; this corresponds to the movie played backward). Consider two states A and B and identify them as sets consisting of microscopic states. T_t be the evolution operator of the system and $A_B = A \cap T_t^{-1}B$ (as before, →**2.7.6**). Let $T_t(A_B) = C$. Then, this is a subset of B.

---

[86] The reader who studied Appendix to Chapter 2 immediately see that this concept is very tricky, if not outright nonsense.

Now, we play the movie describing the above transition backward. Then, due to the reversibility of mechanics, all the states in $\tilde{C}$ return to $\tilde{A}$; no other points in $\tilde{B}$ can go back to $\tilde{A}$. Therefore, $\tilde{C} = \tilde{B}_{\tilde{A}}$. This implies that for an invariant measure μ (= time independent distribution →**2.7.4**) that is time reversal symmetric

$$\mu(A_B) = \mu(\tilde{B}_{\tilde{A}}). \tag{X1.11.1}$$

This is true for equilibrium distribution. Since the equilibrium distribution is symmetric under time reversal, we have

$$P(A,t;B) = P(\tilde{B},t;\tilde{A}) \ \Rightarrow \ P(A,t|B)P_{eq}(B) = P(\tilde{B},t|\tilde{A})P_{eq}(A). \tag{X1.11.2}$$

This is the most general form of the *detailed balance condition*.

The detailed balance condition may be expressed as

$$\int d\boldsymbol{x}d\boldsymbol{y}\, f(\boldsymbol{x})P(\boldsymbol{x},t|\boldsymbol{y})P_{eq}(\boldsymbol{y})g(\boldsymbol{y}) = \int d\boldsymbol{x}d\boldsymbol{y}\, f(\boldsymbol{x})P(\tilde{\boldsymbol{y}},t|\tilde{\boldsymbol{x}})P_{eq}(\boldsymbol{y})g(\boldsymbol{y}), \tag{X1.11.3}$$

where f and g are arbitrary well-behaved functions (test functions; they must satisfy some general boundary conditions, e.g., fP_{eq} must vanish at infinity or at the boundary of the domain).

X1.11.2 Detailed balance condition for Fokker-Planck equation: abstract form

Even after coarse-graining the dynamics in equilibrium should satisfy the detailed balance condition (→**X1.11.1**). Therefore, the Fokker-Planck equation describing the equilibrium dynamics must satisfy the detailed balance condition.

Let us consider the general Fokker-Planck equation (→**X1.7.1**, **X1.8.6**) and write it in the operator equation form as

$$\frac{dP}{dt} = \mathcal{L}P, \tag{X1.11.4}$$

where $\mathcal{L}$ in the position representation reads

$$\mathcal{L} = -\sum_i \frac{\partial}{\partial x_i}a_i + \sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}D_{ij}. \tag{X1.11.5}$$

The transition probability can be formally written as (→**X1.9.1**)

$$P(\boldsymbol{x},t|\boldsymbol{y}) = e^{t\mathcal{L}_{\boldsymbol{x}}}\delta(\boldsymbol{x}-\boldsymbol{y}). \tag{X1.11.6}$$

Here, the variables on which $\mathcal{L}$ operates are explicitly shown as the subscript of the operator as $\mathcal{L}_{\boldsymbol{x}}$.

We can write the detailed balance condition (X1.11.3) as

$$\int d\boldsymbol{x}d\boldsymbol{y}\, f(\boldsymbol{x})e^{t\mathcal{L}_{\boldsymbol{x}}}\delta(\boldsymbol{x}-\boldsymbol{y})P_{eq}(\boldsymbol{y})g(\boldsymbol{y}) = \int d\boldsymbol{x}d\boldsymbol{y}\, f(\boldsymbol{x})e^{t\tilde{\mathcal{L}}_{\boldsymbol{y}}}\delta(\boldsymbol{x}-\boldsymbol{y})P_{eq}(\boldsymbol{x})g(\boldsymbol{y}). \tag{X1.11.7}$$

Here, we have used the invariance of the delta function under time reversal (because $\boldsymbol{x} = \boldsymbol{y}$ is equivalent to $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{y}}$).

Let us perform the integration. The left-hand side reads

$$\text{LHS} = \int d\boldsymbol{x}\, f(\boldsymbol{x})e^{t\mathcal{L}}P_{eq}(\boldsymbol{x})g(\boldsymbol{x}). \tag{X1.11.8}$$

This can be checked order by order in t. Here, we have used the fact that

$$\int d\boldsymbol{y}\,\mathcal{L}_{\boldsymbol{x}}\delta(\boldsymbol{x}-\boldsymbol{y})f(\boldsymbol{y}) = \mathcal{L}f(\boldsymbol{x}). \tag{X1.11.9}$$

We have analogously

$$\text{RHS} = \int d\boldsymbol{y}\,g(\boldsymbol{y})\left[\int d\boldsymbol{x}\,e^{t\tilde{\mathcal{L}}_{\boldsymbol{y}}}\delta(\boldsymbol{x}-\boldsymbol{y})P_{eq}(\boldsymbol{x})f(\boldsymbol{x})\right] = \int d\boldsymbol{x}\,g(\boldsymbol{x})e^{t\tilde{\mathcal{L}}}P_{eq}(\boldsymbol{x})f(\boldsymbol{x}). \tag{X1.11.10}$$

Therefore, (X1.11.7) implies

$$\int d\boldsymbol{x}\,f(\boldsymbol{x})e^{t\mathcal{L}}P_{eq}(\boldsymbol{x})g(\boldsymbol{x}) = \int d\boldsymbol{x}\,g(\boldsymbol{x})e^{t\tilde{\mathcal{L}}}P_{eq}(\boldsymbol{x})f(\boldsymbol{x}). \tag{X1.11.11}$$

(Note that operators act on everything on their right sides.)

If we introduce the adjoint of $\mathcal{L}$ by

$$\int d\boldsymbol{x} f(\boldsymbol{x})\mathcal{L}g(\boldsymbol{x}) = \int d\boldsymbol{x} g(\boldsymbol{x})\mathcal{L}^{\dagger}f(\boldsymbol{x}), \tag{X1.11.12}$$

(X1.11.13) implies

$$\int d\boldsymbol{x}\,f(\boldsymbol{x})e^{t\mathcal{L}}P_{eq}(\boldsymbol{x})g(\boldsymbol{x}) = \int d\boldsymbol{x}\,f(\boldsymbol{x})P_{eq}(\boldsymbol{x})e^{t\tilde{\mathcal{L}}^{\dagger}}g(\boldsymbol{x}). \tag{X1.11.13}$$

Therefore, the detailed balance condition may be written as[87]

$$\mathcal{L}P_{eq} = P_{eq}\tilde{\mathcal{L}}^{\dagger}. \tag{X1.11.14}$$

X1.11.3 Detailed balance condition in terms of coefficients

Let

$$\mathcal{L} = -\sum_i \frac{\partial}{\partial x_i}a_i(\boldsymbol{x}) + \sum_{ij}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}D_{ij}(\boldsymbol{x}). \tag{X1.11.15}$$

Its adjoint reads

$$\mathcal{L}^{\dagger} = \sum_i a_i(\boldsymbol{x})\frac{\partial}{\partial x_i} + \sum_{ij}D_{ij}(\boldsymbol{x})\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}. \tag{X1.11.16}$$

Notice that the backward equation (→**X1.9.2**) is just $\partial_t u = \mathcal{L}^{\dagger}u$.

Let us compute (X1.11.14) in terms of these explicit expressions. For simplicity, let us assume that all the variables have definite parity under time reversal: $\tilde{x}_i = \epsilon_i x_i$, where $\epsilon_i = \pm 1$. Applying the LHS of (X1.11.14) to a test function f, we have

$$\left[-\sum_i \frac{\partial}{\partial x_i}a_i(\boldsymbol{x}) + \sum_{ij}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}D_{ij}(\boldsymbol{x})\right]P_{eq}f$$
$$= -\sum_i a_i P_{eq}\frac{\partial}{\partial x_i}f + \sum_{ij}\left[2\frac{\partial}{\partial x_j}D_{ij}P_{eq}\right]\frac{\partial}{\partial x_i}f + \sum_{ij}D_{ij}P_{eq}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}f. \tag{X1.11.17}$$

---

[87] If we do not assume that the equilibrium state is not time reversal symmetric, e.g., due to the presence of the magnetic field, then this should read $\mathcal{L}P_{eq} = \tilde{P}_{eq}\tilde{\mathcal{L}}^{\dagger}$.

Here, the symmetry of D_{ij} and $\mathcal{L}P_{eq} = 0$ have been used.

Similarly, the RHS operated on f reads

$$\sum_i \epsilon_i \tilde{a}_i P_{eq} \frac{\partial}{\partial x_i} f + \sum_{ij} \epsilon_i \epsilon_j \tilde{D}_{ij} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f. \tag{X1.11.18}$$

Here, $\tilde{a}_i$, etc., imply that the variables are time-reversed. Therefore, comparing the coefficients of corresponding differential operators, we conclude that the detailed balance condition is equivalent to

$$\begin{aligned} -a_i P_{eq} + 2\sum_j \frac{\partial}{\partial x_j} D_{ij} P_{eq} &= \epsilon_i \tilde{a}_i P_{eq}, & \text{(X1.11.19)} \\ D_{ij} &= \epsilon_i \epsilon_j \tilde{D}_{ij}. & \text{(X1.11.20)} \end{aligned}$$

If the variables are all of even parity $\epsilon_i = +1$, then the above condition reduces to

$$-a_i P_{eq} + \sum_j \frac{\partial}{\partial x_j} D_{ij} P_{eq} = 0. \tag{X1.11.21}$$

This is exactly the condition that the probability flux vanishes in equilibrium (→**3.5.8**).

X1.11.4 Symmetric operator associated with $\mathcal{L}$

(X1.11.14) may be rewritten as

$$\mathcal{L}_s \equiv P_{eq}^{-1/2} \mathcal{L} P_{eq}^{1/2} = P_{eq}^{1/2} \tilde{\mathcal{L}}^\dagger P_{eq}^{-1/2}. \tag{X1.11.22}$$

That is,

$$\mathcal{L}_s = \tilde{\mathcal{L}}_s^\dagger. \tag{X1.11.23}$$

Therefore, if all the variables are time-reversal symmetric as is often the case, then $\mathcal{L}_s$ is a symmetric operator under the condition of detailed balance (cf. **??**).[88]

X1.11.5 Detailed balance condition for Zwanzig's nonlinear Langevin equation

To be concrete let us consider Zwanzig's Langevin equation (simplified version in **X1.3.3**) or rather the corresponding Fokker-Planck equation;

$$\frac{\partial P}{\partial t} = -\sum_i \frac{\partial}{\partial A_i} \left(v_i - \sum_j L_{ij} \left(\frac{\partial F}{\partial A_j} \right) \right) P + \sum_{ij} \frac{\partial}{\partial A_i} k_B T L_{ij} \frac{\partial}{\partial A_j} P. \tag{X1.11.24}$$

Here, we assume that the equilibrium distribution of the system is $\propto e^{-\beta F}$.

The symmetric operator associated with this $\mathcal{L}_s$ defined in (X1.11.22) reads

$$\mathcal{L}_s = \sum_i \left(\frac{\partial}{\partial A_i} - \frac{1}{2} \frac{\partial \beta F}{\partial A_i} \right) v_i + \sum_{ij} \left(\frac{\partial}{\partial A_i} - \frac{1}{2} \frac{\partial \beta F}{\partial A_i} \right) k_B T L_{ij} \left(\frac{\partial}{\partial A_j} + \frac{1}{2} \frac{\partial \beta F}{\partial A_j} \right). \tag{X1.11.25}$$

---

[88] It may not be self-adjoint; to show the self-adjointness we must demonstrate that the domains agree.

Notice that the combination

$$\frac{\partial}{\partial A_i}\frac{\partial}{\partial A_j}L_{ij} \tag{X1.11.26}$$

is time-reversal symmetric (→**4.1.10**), and so is F. Therefore, the second term of $\mathcal{L}_s$ satisfies the detailed-balance condition (X1.11.23).

For the first term to satisfy the detailed balance condition, we must require (this is an operator identity, so do not forget that it operates on a test function)

$$\left(\frac{\partial}{\partial A_i} - \frac{1}{2}\frac{\partial \beta F}{\partial A_i}\right) v_i = -\epsilon_i \tilde{v}_i \left(\frac{\partial}{\partial A_i} + \frac{1}{2}\frac{\partial \beta F}{\partial A_i}\right). \tag{X1.11.27}$$

This relation must hold for any choice of F (because perturbed system must also have the equilibrium state), so

$$\begin{aligned} v_i &= -\epsilon_i \tilde{v}_i, &\text{(X1.11.28)}\\ \frac{\partial v_i}{\partial A_i} - \frac{1}{2} v_j \frac{\partial \beta F}{\partial A_i} &= -\frac{1}{2}\epsilon_i \frac{\partial \beta F}{\partial A_i}\tilde{v}_i &\text{(X1.11.29)}\end{aligned}$$

are required. Therefore, the second condition implies

$$\sum_i \frac{\partial}{\partial A_i}(v_i P_{eq}) = 0. \tag{X1.11.30}$$

This is exactly what we have shown in **X1.3.6**.

X1.11.6 Relaxation to equilibrium

We expect that due to irreversibility the system governed by the Fokker-Planck equation satisfying the detailed balance condition starting from any initial distribution P_0 should relax to the equilibrium state. Formally, the solution has the form $e^{t\mathcal{L}}P_0$, so $\mathcal{L}$ should be a negative semidefinite operator. Indeed, this is the case. We can show

$$\int d\boldsymbol{A}\, f \mathcal{L}_s f \le 0. \tag{X1.11.31}$$

$\mathcal{L}_s$ consists of two different types of terms: the first term and the second term in (X1.11.25). The second law implies that the contribution to the second term is negative semidefinite. It is easy to show that the contribution of the first term vanishes under the detailed balance condition (→**X1.11.2**).

X1.11.7 H-theorem for Fokker-Planck equation: detailed study

X1.11.6 tells us that the Fokker-Planck equation drives any initial condition to the subspace where the equilibrium distribution function resides. However, without a further detailed study of the operator $\mathcal{L}$ we cannot assert that the destination (normalized) is definitely P_{eq}. The best way to demonstrate this is to construct a Lyapunov function whose unique minimum is the equilibrium distribution function. This is the basic spirit of the H-theorem due to Boltzmann (→**2.4.2**).

Let us study the time derivative of the following Kullback-Leibler entropy (→**1.6.9**).

$$K(P, P_{eq}) = \int d\boldsymbol{A}\, P(\boldsymbol{A}, t) \log \frac{P(\boldsymbol{A}, t)}{P_{eq}}. \tag{X1.11.32}$$

As we already know (→**1.6.9**), this is nonnegative and is zero if and only if $P = P_{eq}$. Therefore, if we can show that $dK/dt < 0$ for $P \neq P_{eq}$, we are done.

Let us again stick to Zwanzig's nonlinear Langevin equation (→**X1.3.3**). Its Fokker-Planck equation (X1.11.24) has the following structure: the streaming terms + diffusion terms

$$\frac{\partial P}{\partial t} = \left(\frac{\partial P}{\partial t}\right)_{str} + \left(\frac{\partial P}{\partial t}\right)_{dif}, \tag{X1.11.33}$$

where

$$\left(\frac{\partial P}{\partial t}\right)_{str} \equiv -\sum_i \frac{\partial}{\partial A_i} v_i P, \tag{X1.11.34}$$

$$\left(\frac{\partial P}{\partial t}\right)_{dif} \equiv \sum_{ij} \frac{\partial}{\partial A_i}\left[L_{ij}\frac{\partial F}{\partial A_j} + k_B T L_{ij}\frac{\partial}{\partial A_j}\right] P. \tag{X1.11.35}$$

We know that the streaming terms do not affect the equilibrium distribution (→**X1.3.6**), and its origin is an instantaneous reversible dynamics (→**X1.3.5**). Therefore, we expect this term should not produce any entropy:

$$\left(\frac{\partial P}{\partial t}\right)_{str} K(P, P_{eq}) = 0. \tag{X1.11.36}$$

This is indeed the case. The diffusion term represents irreversibility, so we expect

$$\left(\frac{\partial P}{\partial t}\right)_{dif} K(P, P_{eq}) < 0. \tag{X1.11.37}$$

This is also the case.

Therefore, under the detailed balance condition, we have demonstrated that the solution to the Fokker-Planck equation eventually settles down to the unique equilibrium distribution.

The case without detailed balance will be discussed in **8.4**.

Chapter X2

Response to Perturbation

Linear response theory (LRT) is perhaps the most cultivated subfield of nonequilibrium statistical physics. Some people even claim that LRT is the main pillar of nonequilibrium statistical mechanics. However, linear response functions are straightforward generalization of static susceptibilities (→**2.8.1**) to time dependent perturbations, and, needless to say, calculation of susceptibility is not a central problem of statistical mechanics at all. The central problem of statistical mechanics is to compute the partition function; there is no counter part (at least so far) in nonequilibrium statistical mechanics. This is the reason why such a peripheral question as LRT looks central.

Still, it is true that LRT is a major part of the current nonequilibrium statistical mechanics, so we wish to pay a reasonable respect. In this chapter LRT is discussed critically, and typical examples (Kubo-Tomita's spectral line shape of NMR **X2.6**, and Nakano's electric conductivity formula **X2.7**) will be outlined.

The structure of this chapter is as follows:

Kinematics of linear systems. In **X2.2** a general kinematics of linear systems is summarized (consequences of causality and passivity). This is a prerequisite and important.

Elementary approaches to LRT. Pure microscopic mechanical approaches to LRT are impossible (→**X2.4.4**), so to construct LRT mechanics and phenomenology (macroscopic laws) are skillfully (or nonchalantly) knit into a single texture. This makes the theory opaque (or erroneously transparent as if 'theorems' could be proved purely mechanically as in Kubo's famous paper[89]). Therefore, in this chapter, several routes to LRT are shown. In **X2.3** in the classical case it is shown that the basic theoretical framework of LRT (fluctuation dissipation relations and Green-Kubo relations) may be derived from the pure phenomenology + linear kinematics of **X2.2**. No microscopic input is used in this derivation, so there is a very strong reason to believe that LRT is rigorous. In the quantum case discussed in **X2.4**, unfortunately, this seems impossible, so the loss part of the admittance is computed with the aid of the Fermi golden rule. Then, with the aid of kinematics everything is derived.[90]

Computation of nonequilibrium distribution. There are two ways to perturb a system, mechanically and thermally. In the former case, we can easily describe the perturbation with a perturbative Hamiltonian added to the system Hamiltonian. Mechanical perturbations can at least formally be treated within the framework of the ordinary perturbation theory. In contrast, for thermal perturbations (e.g., temperature modification), there is no simple natural perturbative Hamiltonian, so LRT cannot be derived purely mechanically.[91] Therefore, in **X2.5** a unified approach is given that computes the evolution of the density operator (called the nonequilibrium statistical operator) in the spirit of statistical thermodynamics.

---

[89] R Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).

[90] A more phenomenological, i.e., a more reliable, derivation of quantum mechanical cases is possible, if we admit the KMS condition (→**X2.4.12**) as given. Since the KMS is regarded fundamental, also for quantum cases LRT is expected to be rigorous.

[91] With some contrivance, this is possible as can be seen in J. M. Luttinger, "Theory of Thermal Transport Coefficients," Phys. Rev. **135**, A1505-A1514 (1964). This is a very nice paper worth reading.

The reader can skip **X2.3** and **X2.4** to understand the general framework of LRT; from **X2.2** she may directly go to **X2.5**.

Those who are impatient should first go to **X2.1.3**.

Recommended References:

R. Kubo, Rep. Prog. Phys. **29**, 255 (1966),
H. Nakano, Int. J. Mod. Phys. B **7**, 2397 (1993).[92]

R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II* (Springer, 1985) is perhaps a kinder introduction, although it is not critical. The book covers the materials in Chapters X1 and X2 of these notes.

X2.1 Linear Response — Introduction

X2.1.1 Fluctuation and Response: what we already know

We have already seen in the static case that the response of the system is governed by the fluctuation (→**2.8.1**)

$$\langle \delta x_i \rangle = \beta \langle \delta x_j \delta x_i \rangle X_j \tag{X2.1.1}$$

or in the quantum case (→**X1.2.5**)

$$\langle \delta x_i \rangle = \left\langle \int_0^\beta d\lambda \, e^{\lambda H} \delta x_j e^{-\lambda H} \delta x_i \right\rangle X_j = \beta \langle \delta x_j ; \delta x_i \rangle X_j, \tag{X2.1.2}$$

where we have used the canonical correlation (→**X1.2.5**; summary in **4.3.5**). This suggests that if the perturbation X_j is time-dependent, then

$$\langle x_i(t) \rangle = \beta \int_0^t ds \, \langle \delta \dot{x}_j(s) ; \delta x_i(t) \rangle X_j(s). \tag{X2.1.3}$$

This is true as we will see later (→**X2.3.3**, **X2.5.3**).[93] Also we have learned that the *response function* is given by the *fluctuation-dissipation relation* of the first kind (→**3.2.8**). This has just the form guessed above.

More generally, we have learned that the *Onsager coefficients* can be calculated in terms of the time correlation of fluctuations (→**4.1.12**). Transport coefficients can be written in terms of the correlation function of fluxes (the *Green-Kubo relation* →**4.1.14**). Also we have learned that the Green-Kubo relation can be obtained directly from the Onsager principle (→**X1.1.7**).

The reader will see that all the important key concepts and results have already been discussed at least in their simplest forms in Chapter 3.

X2.1.2 Hydrodynamics and response

We have outlined the derivation of hydrodynamic equations from the study of fluctuation

---

[92] This contains an episode of Feynman on electric conductivity.

[93] The dot is the time derivative; its necessity is easily guessed from the dimensional reason.

dynamics (→**5.2.5**). Green-Kubo relations are derived almost phenomenologically with Onsager's regression hypothesis (→**4.1.2**; its more complete version is Onsager's principle **??**). Study of the hydrodynamic behavior of the system with the analysis of its fluctuation dynamics must allow us to understand almost everything about nonequilibrium properties of the system close to equilibrium. This is the approach summarized in **X2.3** and **X2.4**.

Even when we wish to start from microscopic mechanics, we need a selection principle of the distribution (measure) just as in the case of equilibrium statistical mechanics (→**2.7.4**). Thus, our guiding principle is always the same. Even in nonequilibrium, statistical mechanics is a framework to try to understand the effect of microscopic degrees of freedom with the theoretical framework compatible with phenomenology and mechanics. This is the approach in **X2.6**.

Without a reliable phenomenological framework, we cannot construct a reliable statistical framework.

X2.1.3 User's guide: practical summary

If the reader wishes a clean outline of the framework of linear response theory, read **X2.6**. All the important basic results already outline in Chapter 3 are derived quantum mechanically within the statistical thermodynamic framework. All the theoretical summaries are also there.

It is, however, important to understand the general kinematics of linear systems summarized in **X2.2**.

The consequence of the second law constrains the imaginary part (skew-Hermitian part) of admittance (→**X2.2.5**, **X2.2.8**). The relation between causality and analytic properties of admittance **X2.2.9** and its consequence (the Kramers-Kronig relation, dispersion relation **X2.2.10**) must clearly be understood.

Classical applications of the general framework are illustrated in **X2.7** and **X2.8**.

X2.2 Linear Irreversible Process — General Framework

Before going into dynamics, we wish to summarize the kinematics of small perturbations and responses to them. Clearly recognize the consequences of the second law (→**X2.2.7**) and causality (→**X2.2.9**).

X2.2.1 Response function.

We believe that if the external perturbation of the system parameter or the change of the thermodynamic force $X(\boldsymbol{r}, t)$,[94] which may depend on space and time coordinates, is small, the system response described by the change $G(\boldsymbol{r}, t)$ of thermodynamic densities (= the

---

[94] In these notes, X is used for the thermodynamic intensive parameter appearing in the ordinary Gibbs relation. The intensive variable in the entropy formula will be denoted by F. That is $dE = \cdots + Xdx + \cdots$, and $dS = \cdots + Fdx + \cdots$.

extensive variables per volume)[95] should also be small, and G is linearly related to X:

$$G(\boldsymbol{r},t) = L(\{X(\boldsymbol{r},t)\}). \tag{X2.2.1}$$

Here, $G(\boldsymbol{r},t)$ could depend on the values of X at the point away from $\boldsymbol{r}$ at a different time t (in the past only; this is causality →**X2.2.9**). In these notes, we assume X is the thermodynamic force (defined with respect to energy) and G is the conjugate density (extensive quantities conjugate to X per unit volume). They may be vectors or tensors.

According to the Riesz representation theorem,[96] such a linear functional can always be written in the following form

$$G(\boldsymbol{r},t) = \int d\boldsymbol{r}' \int_{-\infty}^{t} dt'\, \phi(\boldsymbol{r}-\boldsymbol{r}', t-t')X(\boldsymbol{r}',t'). \tag{X2.2.2}$$

Here, we have assumed that the system is space-time uniform (no absolute position in time and space matters) for simplicity. The function ϕ is called the *response function* in general (see for a prototype **3.2.8**).

X2.2.2 Admittance

The space-time Fourier transform of the response function (→**3.2.8**, **X2.2.1**, **X2.5.3**) is called the *admittance.*

To define this unambiguously we must specify the convention of space-time Fourier transformation. We adopt the following definition (consistent with the previous sections):

$$f(\boldsymbol{k},\omega) = \int d\boldsymbol{r} \int dt\, e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} f(\boldsymbol{r},t). \tag{X2.2.3}$$

The corresponding inverse transformation reads

$$f(\boldsymbol{r},t) = \frac{1}{(2\pi)^4} \int d\boldsymbol{k} \int d\omega\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} f(\boldsymbol{k},\omega). \tag{X2.2.4}$$

Notice that each harmonic component behaves as a plane wave[97]

$$e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} \tag{X2.2.5}$$

with the velocity $\omega/|\boldsymbol{k}|$ in the direction of $\boldsymbol{k}$.

Thus, the *admittance* χ is defined as

$$\chi(\boldsymbol{k},\omega) = \int d\boldsymbol{r} \int dt\, e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} \phi(\boldsymbol{r},t). \tag{X2.2.6}$$

Therefore,

$$G(\boldsymbol{k},\omega) = \chi(\boldsymbol{k},\omega)X(\boldsymbol{k},\omega). \tag{X2.2.7}$$

---

[95] Here, we define thermodynamic densities as the conjugate of X with respect to energy.

[96] e.g., K Yosida, *Functional Analysis* (Springer, Berlin, 1965; Sixth edition 1980), p90. Here, it is a theorem on the continuous (= bounded) linear map on the Hilbert space.

[97] We choose our convention so that the plane wave can be written as $e^{ix_jk^j}$ in terms of 4-vectors and the relativistic scalar product with the usual indefinite metric $(+,-,-,-)$.

X2.2.3 Real and imaginary parts of admittance

Since X and G in our case are observable quantities, they must be real, and so must be the response function. Therefore,

$$\overline{\chi(\boldsymbol{k},\omega)} = \int d\boldsymbol{r} \int dt\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\phi(\boldsymbol{r},t) = \chi(-\boldsymbol{k},-\omega). \tag{X2.2.8}$$

We introduce the real and imaginary parts of the admittance as (the names will become clear in **X2.2.5**)

$$\begin{aligned} \chi'(\boldsymbol{k},\omega) &= Re\,\chi(\boldsymbol{k},\omega) \quad \text{called } \textit{dispersive} \text{ or } \textit{storage part}, & \text{(X2.2.9)} \\ \chi''(\boldsymbol{k},\omega) &= -Im\,\chi(\boldsymbol{k},\omega) \quad \text{called } \textit{dissipative} \text{ or } \textit{loss part}, & \text{(X2.2.10)} \end{aligned}$$

or

$$\chi(\boldsymbol{k},\omega) = \chi'(\boldsymbol{k},\omega) - i\chi''(\boldsymbol{k},\omega). \tag{X2.2.11}$$

Here, do not forget the inclusion of $-$ in the definition of χ''. This negative sign is needed to make this quantity positive definite as we will see later (→**X2.2.5**).

We may write

$$\begin{aligned} \chi'(\boldsymbol{k},\omega) &= \int d\boldsymbol{r} \int dt\, \cos(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})\phi(\boldsymbol{r},t), & \text{(X2.2.12)} \\ \chi''(\boldsymbol{k},\omega) &= \int d\boldsymbol{r} \int dt\, \sin(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})\phi(\boldsymbol{r},t). & \text{(X2.2.13)} \end{aligned}$$

(X2.2.8) now reads in terms of χ' and χ'' as

$$\begin{aligned} \chi'(\boldsymbol{k},\omega) &= \chi'(-\boldsymbol{k},-\omega), & \text{(X2.2.14)} \\ \chi''(\boldsymbol{k},\omega) &= -\chi''(-\boldsymbol{k},-\omega). & \text{(X2.2.15)} \end{aligned}$$

That is, the real part of the admittance is an even function, and the imaginary part an odd function.

X2.2.4 Warning about conventions

If the definition of the space-time Fourier transformation is different, then the signs may be flipped. If

$$f(\boldsymbol{k},\omega) = \int d\boldsymbol{r} \int dt\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}f(\boldsymbol{r},t), \tag{X2.2.16}$$

then, $\chi = \chi' + i\chi''$ is the definition of χ' and χ'' (but (X2.2.9) and (X2.2.10) are intact; that is, their definitions are consistent in different conventions). For example, the convention of D. Forster, *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions* (Benjamin, 1975) and L. P. Kadanoff and P. C. Martin, "Hydrodynamic Equations and Correlation Functions," Ann. Phys. **24**, 419-469 (1963) is different from ours.

X2.2.5 Dissipation and admittance

The power $p(\boldsymbol{r},t)d\boldsymbol{r}$ supplied by the force X to the volume element $d\boldsymbol{r}$ around $\boldsymbol{r}$ of the system is the product of X and the rate of change of the conjugate density G. Here, for simplicity, let us assume that X is a scalar (for many-variable cases →**X2.2.8**). Therefore,

$$p(\boldsymbol{r},t) = X(\boldsymbol{r},t)\frac{\partial}{\partial t}G(\boldsymbol{r},t). \tag{X2.2.17}$$

Let us compute the time average of the power for the following special 'monochromatic' perturbation

$$X(\boldsymbol{r},t) = Re\left(e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\tilde{X}\right). \tag{X2.2.18}$$

Here, $\tilde{X}$ is called the *phasor representation* of the monochromatic force X (the definition is the same as in the AC circuit theory). $\tilde{X}$ may be a complex number due to the phase.

The time average per unit volume $\langle p\rangle$ of the power $p(\boldsymbol{r},t)$ is, as shown in **X2.2.6**, given by

$$\langle p\rangle = \frac{1}{2}\omega\chi''(\boldsymbol{k},\omega)|\tilde{X}|^2. \tag{X2.2.19}$$

In the steady state, the power supplied by the perturbation of X must be dissipated as heat to the heat bath. Therefore, $\langle p\rangle$ must be the energy dissipation rate by the system per unit volume. Therefore, the second law implies that χ'' must be nonnegative definite, or more generally,

$$\omega\chi'' \geq 0. \tag{X2.2.20}$$

X2.2.6 Computation of average power loss

Postponing a smart derivation to **X2.2.7**, let us compute $p(t)$ according to its definition (X2.2.17). Since

$$G(\boldsymbol{r},t) = \frac{1}{2}\left[\chi(\boldsymbol{k},\omega)e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\tilde{X} + \chi(-\boldsymbol{k},-\omega)e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\tilde{X}^*\right] \tag{X2.2.21}$$

according to (X2.2.2), we can obtain its time derivative easily. The time averaged power per unit volume $\langle p\rangle$ may be computed as

$$\langle p\rangle = \frac{1}{V}\frac{\omega}{2\pi}\int_0^{2\pi/\omega} dt \int_V d\boldsymbol{r}\, p(\boldsymbol{r},t). \tag{X2.2.22}$$

Here, V is the system volume.[98]

Putting (X2.2.21) and the formula for X into (X2.2.22), we obtain

$$\langle p\rangle = \frac{1}{4}\left[i\omega\chi(\boldsymbol{k},\omega) - i\omega\chi(-\boldsymbol{k},-\omega)\right]|\tilde{X}|^2. \tag{X2.2.23}$$

This is what we wished to demonstrate (do not forget χ'' is defined with the negative sign in (X2.2.11)).

X2.2.7 Phasor representation and dissipation

Let the plane wave be given as

$$A(t) = Re\left(e^{i\omega t}\tilde{A}\right). \tag{X2.2.24}$$

We call the complex number $\tilde{A}$ the *phasor representation* of A. We have introduced the phasor representation of the thermodynamic force $\tilde{X}$ in **X2.2.5**. The time average of the

---

[98] We take the $V \to \infty$ limit after all the calculations are over.

product of two plane waves in the form of (X2.2.24) is given by (the overline implies time average)

$$\overline{AB} = \frac{1}{2} Re\, \tilde{A}^* \tilde{B}. \tag{X2.2.25}$$

The space-time average is just a simple generalization of the above result.

The perturbation result G has the following phasor representation as clearly seen from (X2.2.21):

$$\tilde{G} = \chi(\boldsymbol{k}, \omega)\tilde{X}. \tag{X2.2.26}$$

Therefore, (the overline implies time and space average; not complex conjugate)

$$\langle p \rangle = \overline{X \frac{\partial G}{\partial t}} = \frac{1}{2} Re\, \tilde{X}^*(i\omega\tilde{G}) = \frac{1}{2} Re \left(i\omega\chi(\boldsymbol{k}, \omega)|\tilde{X}|^2 \right). \tag{X2.2.27}$$

This is exactly (X2.2.19).

X2.2.8 Dissipation and admittance in many-variable case.

Now, let us generalize what we have accomplished in **X2.2.5** to the case with many simultaneous perturbations. X is regarded as a vector, and so is G in **X2.2.5**. Therefore, the response function or the admittance must be a matrix.

The power $p(\boldsymbol{r}, t)d\boldsymbol{r}$ supplied by the force X to the volume element $d\boldsymbol{r}$ around $\boldsymbol{r}$ of the system now reads

$$p(\boldsymbol{r}, t) = X(\boldsymbol{r}, t)^T \frac{\partial}{\partial t} G(\boldsymbol{r}, t). \tag{X2.2.28}$$

Therefore, If we introduce the phasor representation (→**X2.2.7**) of the perturbation:

$$X(\boldsymbol{r}, t) = Re\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\tilde{X}, \tag{X2.2.29}$$

we have

$$\langle p \rangle = \frac{1}{2} Re \left(\tilde{X}^\dagger i\omega\chi(\boldsymbol{k}, \omega)\tilde{X} \right). \tag{X2.2.30}$$

Here, † is the Hermitian conjugate (transposition + complex conjugation). This should easily be guessed from (X2.2.27).

Since $i\tilde{X}\tilde{X}^\dagger$ is a skew-Hermitian matrix, only the skew-Hermitian part χ_- of the admittance matrix χ

$$\chi_- \equiv \frac{1}{2}(\chi - \chi^\dagger) \tag{X2.2.31}$$

contributes $\langle p \rangle$. Therefore,

$$\langle p \rangle = \frac{1}{2} \omega Re\, Tr \left(i\chi_- \tilde{X}\tilde{X}^\dagger \right). \tag{X2.2.32}$$

This must be non-negative definite due to the second law, but the constraints cannot be written in a compact fashion.

Exercise. Suppose we have a circularly polarized magnetic field. Its phasor representation can be $(H, -iH, 0)^T$. Show that the power absorption rate is given by (we assume a spatially uniform system)

$$p = \frac{1}{2}(\chi'_{xy} - \chi'_{yx} + \chi''_{xx} + \chi''_{yy})\omega H^2. \tag{X2.2.33}$$

□

X2.2.9 Causality and analyticity of admittance

Although the perturbation result $G(\boldsymbol{r},t)$ may depend on $X(\boldsymbol{r},s)$ at time s different from t, it cannot depend on X for $s > t$. This is *causality.*

Therefore, in terms of the response function $\phi(\boldsymbol{r},t)$ (→**X2.2.1**), the causality can be written as

$$\phi(\boldsymbol{r},t) = 0 \text{ if } t < 0. \tag{X2.2.34}$$

This may be written in terms of the admittance as

$$0 = \int d\boldsymbol{k} \int d\omega\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\chi(\boldsymbol{k},\omega), \tag{X2.2.35}$$

if $t < 0$.

For $t < 0$ $e^{i\omega t}$ decays exponentially for $|\omega| \to \infty$ if $Im\,\omega < 0$ (that is, on the lower half complex ω-plane). Therefore, (X2.2.35) implies that

$$\oint_C d\omega\, e^{i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\chi(\boldsymbol{k},\omega) = 0. \tag{X2.2.36}$$

Since the exponential factor is analytic[99] in the lower half plane, this implies that χ must be analytic in the lower half complex ω plane.[100] Conversely, if χ is analytic on the lower half complex ω-plane, then the causality holds.

In short:

Causality $\Longleftrightarrow$ Analyticity of admittance on the lower half complex ω-plane.[101]

X2.2.10 Kramers-Kronig relation — dispersion relation

Since the admittance is analytic on the lower half ω-plane, Cauchy's theorem tells us

$$0 = \oint_C \frac{\chi(\boldsymbol{k},z)}{z-\omega}dz, \tag{X2.2.37}$$

where the integration contour C is the infinite semicircle on the lower half plane with the straight line portion coinciding with the real axis. If ω is real, then there is a simple pole on the contour, so we may use the Plemelij formula (→**X2.2.11**) to obtain

$$P\int_{-\infty}^{\infty} \frac{\chi(\boldsymbol{k},z)}{z-\omega}dz + i\pi\chi(\boldsymbol{k},\omega) = 0. \tag{X2.2.38}$$

That is,

$$\chi(\boldsymbol{k},\omega) = \frac{i}{\pi}P\int_{-\infty}^{\infty} \frac{\chi(\boldsymbol{k},z)}{z-\omega}dz. \tag{X2.2.39}$$

Now, we split this into real and imaginary parts of χ.

$$\chi'(\boldsymbol{k},\omega) = \frac{1}{\pi}P\int_{-\infty}^{\infty} \frac{\chi''(\boldsymbol{k},x)}{x-\omega}dx, \tag{X2.2.40}$$

$$\chi''(\boldsymbol{k},\omega) = -\frac{1}{\pi}P\int_{-\infty}^{\infty} \frac{\chi'(\boldsymbol{k},x)}{x-\omega}dx. \tag{X2.2.41}$$

---

[99] More precisely, we should use the term, *holomorphic.*

[100] This is not a trivial statement. This depends on the definition of the Fourier transformation. It is the lower half plane with our convention.

[101] This was first recognized by H. Takahashi (1942).

Here, P denotes the *Cauchy principal value.* These relations are know as the *Kramers-Kronig relation* (*dispersion relation*) to physicists.[102] As we will see soon, the dissipative part of χ is experimentally easier to access (that is, χ'' is easier to observe), so these relations may be of practical importance to obtain χ experimentally.[103]

X2.2.11 Plemelij formula

The following formula is called the *Plemelij formula*

$$\frac{1}{x \pm i\epsilon} = P\frac{1}{x} \mp \delta(x), \tag{X2.2.42}$$

where ϵ is an infinitesimal positive number. The equality holds inside the integral with respect to x.[104]

There are several ways to demonstrate this. Perhaps the most elementary one is to use the complex function theory. A simple pole on the integral path contributes one half of its residue to the integral, and the remaining integral becomes the Cauchy principal value.

X2.2.12 Spectral representation

Let us reconstruct χ from its imaginary part χ'' with the aid of the Kramers-Kronig relation **X2.2.10**:

$$\chi(\boldsymbol{k}, \omega) = \frac{1}{\pi} P \int_{-\infty}^{\infty} dx\, \frac{\chi''(\boldsymbol{k}, x)}{x - \omega} - i\chi''(\boldsymbol{k}, \omega). \tag{X2.2.43}$$

With the aid of the Plemelij formula (→**X2.2.11**), this can be rewritten as

$$\chi(\boldsymbol{k}, \omega) = \lim_{\epsilon \to +0} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\chi''(\boldsymbol{k}, x)}{x - (\omega - i\epsilon)}. \tag{X2.2.44}$$

Let us define the analytical continuation of χ'' that is analytic on the lower half ω-plane to the upper half ω-plane as

$$\hat{\chi}(\boldsymbol{k}, z) = \frac{1}{\pi} \int_{-\infty}^{\infty} dx \frac{\chi''(\boldsymbol{k}, x)}{x - z}. \tag{X2.2.45}$$

This has no singularity if $Im\, z \neq 0$. That is, the real axis is a branch cut. (X2.2.46) implies that

$$\chi(\boldsymbol{k}, \omega) = \lim_{\epsilon \to +0} = \hat{\chi}(\boldsymbol{k}, \omega - i\epsilon). \tag{X2.2.46}$$

That is, the continuation from the lower half ω-plane gives the physically meaningful value. Therefore, we call the lower half ω-plane as the *physical sheet.* Its analytical continuation beyond the real axis gives the Riemann surface that contains the physically observable values for real z.

The representation of χ in terms of χ'' is called the *spectral representation.*

---

[102] Well known formulas for the Hilbert transformation.

[103] However, to use these formulas the frequency range must be sufficiently widely observed.

[104] so, the limit $\epsilon \to +0$ is the weak limit.

X2.3 Hydrodynamics and Response — Classical case

X2.3.1 What do we know?

If we understand the dynamics of the fluctuation of a density ψ, we can understand the relaxation dynamics of ψ when it is displaced from the equilibrium value (0). The idea is formulated by Onsager as the regression hypothesis (→**4.1.2**). We have seen that the equations of hydrodynamics can be constructed from this point of view (→**5.2.5**).

On the other hand, then, if we know (e.g., from experiments) the hydrodynamic behavior of small deviations, we should be able to know the fluctuation dynamics (with the aid of general kinematics in **X2.2**).

X2.3.2 Simple conservation law and correlation function

Let ψ be a thermodynamic density (say, the particle number density). The hydrodynamic behavior after linearization (for small fluctuations) reads[105]

$$\frac{\partial \psi}{\partial t} = -L\psi \tag{X2.3.1}$$

with a linear operator L. The space-time correlation function $S(\boldsymbol{r},t)$:

$$S(\boldsymbol{r},t) = \langle \psi(\boldsymbol{0},0)\psi(\boldsymbol{r},t)\rangle \tag{X2.3.2}$$

obeys

$$\frac{\partial}{\partial t}S(\boldsymbol{r},t) = -LS(\boldsymbol{r},t). \tag{X2.3.3}$$

This is obtained from (X2.3.1) by multiplying $\psi(\boldsymbol{0},0)$ and averaging over repeated experiments (sample average).[106]

Notice that (X2.3.3) can be used only for $t > 0$, but

$$\langle \psi(\boldsymbol{0},0)\psi(\boldsymbol{r},t)\rangle = \langle \psi(-\boldsymbol{r},-t)\psi(\boldsymbol{0},0)\rangle, \tag{X2.3.4}$$

so

$$S(\boldsymbol{r},t) = S(-\boldsymbol{r},-t) \tag{X2.3.5}$$

classically.[107] To obtain the Fourier transform of the space-time correlation function (if ψ is the particle number density, we are computing the dynamical form factor as we have seen in **3.7.10**), we must study the analytical properties of the Laplace and the Fourier transforms of the correlation function.

X2.3.3 Fourier and Laplace transformations of correlation functions: convention

The space-time Fourier transform of $S(\boldsymbol{r},t)$ has been introduced as

$$S(\boldsymbol{k},\omega) = \int d\boldsymbol{r}\int_{-\infty}^{\infty} dt\, e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}S(\boldsymbol{r},t). \tag{X2.3.6}$$

---

[105] To be faithful to Onsager's program (→**??**), there must be a noise. For the time being we do not need it. We will discuss the fluctuation-dissipation relation of the second kind later (→**X2.3.15**).

[106] There is a fluctuation due to thermal noise, but that does not contribute the averaged result.

[107] As we will see later (→**X2.4.5**), in quantum cases, ψ at different space-time points may not commute, so this symmetry is not generally correct.

The Fourier-Laplace transform of $S(\boldsymbol{r}, t)$ is defined as (spatially it is just the Fourier transform as above, and $s = iz$ in the usual convention of the Laplace transformation)

$$S[\boldsymbol{k}, z] = \int d\boldsymbol{r} \int_0^\infty dt\, e^{-i(zt - \boldsymbol{k}\cdot\boldsymbol{r})} S(\boldsymbol{r}, t). \tag{X2.3.7}$$

X2.3.4 Fourier-Laplace transform of correlation function

From (X2.3.3) we obtain

$$izS[\boldsymbol{k}, z] - S(\boldsymbol{k}, t = 0) = -L_{\boldsymbol{k}} S[\boldsymbol{k}, z], \tag{X2.3.8}$$

or

$$S[\boldsymbol{k}, z] = \frac{S(\boldsymbol{k}, t = 0)}{iz + L_{\boldsymbol{k}}}, \tag{X2.3.9}$$

where $L_{\boldsymbol{k}}$ is the Fourier transform of L, and $S(\boldsymbol{k}, t = 0)$ is

$$S(\boldsymbol{k}, t = 0) = \int d\boldsymbol{r} S(\boldsymbol{r}, 0) e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{X2.3.10}$$

$S(\boldsymbol{k}, t = 0)$ is related to the static susceptibility $\chi(\boldsymbol{k})$ as we see in **X2.3.5**. Therefore, in terms of the $\boldsymbol{k}$-dependent (static) susceptibility, we have

$$S[\boldsymbol{k}, z] = \frac{k_B T \chi(\boldsymbol{k})}{iz + L_{\boldsymbol{k}}}. \tag{X2.3.11}$$

X2.3.5 $\boldsymbol{k}$-dependent static susceptibility

By definition,

$$S(\boldsymbol{k}, t = 0) = \int d\boldsymbol{r}\, \langle \psi(0)\psi(\boldsymbol{r})\rangle e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{X2.3.12}$$

where $\langle\ \rangle$ is the equilibrium average. If we introduce the spatial Fourier transformation of ψ, we have

$$\begin{aligned} \langle \psi_{\boldsymbol{k}} \psi_{-\boldsymbol{k}} \rangle &= \int d\boldsymbol{r} \int d\boldsymbol{r}' \langle \psi(\boldsymbol{r}')\psi(\boldsymbol{r})\rangle e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} && \text{(X2.3.13)} \\ &= V \int d\boldsymbol{r} \langle \psi(0)\psi(\boldsymbol{r})\rangle e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. && \text{(X2.3.14)} \end{aligned}$$

Here, V is the system volume. That is,

$$S(\boldsymbol{k}) \equiv S(\boldsymbol{k}, t = 0) = \frac{1}{V}\langle \psi_{\boldsymbol{k}} \psi_{-\boldsymbol{k}} \rangle. \tag{X2.3.15}$$

If $\boldsymbol{k} = 0$ (uniform perturbation), then we have

$$S(0) = \frac{1}{V}\langle \Psi\Psi \rangle = k_B T \chi, \tag{X2.3.16}$$

where $\Psi = \psi_{\boldsymbol{k}=0}$, which is the space integral of $\psi(\boldsymbol{r})$ = the total amount. The susceptibility χ here is defined per volume. Therefore, it is sensible to define the wave vector dependent susceptibility $\chi(\boldsymbol{k})$ according to[108]

$$S(\boldsymbol{k}) = k_BT\chi(\boldsymbol{k}) = \frac{1}{V}\langle\psi_{\boldsymbol{k}}\psi_{-\boldsymbol{k}}\rangle. \tag{X2.3.17}$$

Note that $\chi(\boldsymbol{k})$ is real positive.

X2.3.6 Fourier transform of space-time correlation function

Let us relate the Fourier transform and the Fourier-Laplace transform.

$$\begin{aligned} S(\boldsymbol{k},\omega) &= \int d\boldsymbol{r}\int_{-\infty}^{\infty} dt e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}S(\boldsymbol{r},t), && \text{(X2.3.18)}\\ &= S[\boldsymbol{k},\omega] + \int d\boldsymbol{r}\int_{-\infty}^{0} dt e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}S(\boldsymbol{r},t), && \text{(X2.3.19)}\\ &= S[\boldsymbol{k},\omega] + \int d\boldsymbol{r}\int_{0}^{\infty} dt e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}S(-\boldsymbol{r},-t). && \text{(X2.3.20)} \end{aligned}$$

If we assume that the system is space-inversion symmetric, then $S(\boldsymbol{r},t)$ is an even function of $\boldsymbol{r}$ and t (→(X2.3.5)), so the above result implies

$$S(\boldsymbol{k},\omega) = S[\boldsymbol{k},\omega] + S[-\boldsymbol{k},-\omega]. \tag{X2.3.21}$$

We know the complex conjugate of $S[\boldsymbol{k},\omega]$ is $S[-\boldsymbol{k},-\omega]$, because $S(\boldsymbol{r},t)$ is real. Therefore,

$$S(\boldsymbol{k},\omega) = 2Re\, S[\boldsymbol{k},\omega]. \tag{X2.3.22}$$

This should be valid for sufficiently small ω and $\boldsymbol{k}$. Theoretically, as we have seen, it is easier to compute $S[\boldsymbol{k},z]$ than $S(\boldsymbol{k},\omega)$, but experimentally, as we have already seen in **3.7.10**, the latter is more direct.

X2.3.7 Macroscopic limit is delicate

However, notice that the macroscopic limit $(\boldsymbol{k},\omega) \to (0,0)$ is a limit to a genuine singularity:

$$\begin{aligned} \lim_{|\boldsymbol{k}|\to 0}\lim_{\omega\to 0} S(\boldsymbol{k},\omega) &= +\infty, && \text{(X2.3.23)}\\ \lim_{\omega\to 0}\lim_{|\boldsymbol{k}|\to 0} S(\boldsymbol{k},\omega) &= 0. && \text{(X2.3.24)} \end{aligned}$$

This implies that really infinitely large nonequilibrium systems are very delicate.

X2.3.8 Adiabatic turning on of perturbation

The relaxation of ψ may be studied by the linearized hydrodynamic equation (X2.3.1). The relaxation process is started by removing the external intensive field $h(\boldsymbol{r})$ at $t = 0$. Therefore, the starting state is the equilibrium state with the external field $h(\boldsymbol{r})$. In our setting of the problem, we cannot obtain this state with the aid of equilibrium statistical mechanics,

---

[108] For small fluctuations, we may assume that the distribution is Gaussian: $\propto \exp[-\beta\sum_{\boldsymbol{k}}\psi_{\boldsymbol{k}}\psi_{-\boldsymbol{k}}/2V\chi(\boldsymbol{k})]$. This agrees with our definition of $\chi(\boldsymbol{k})$. See **X2.3.12**.

because we do not know the system Hamiltonian. However, the hydrodynamic equation must be compatible with the equilibrium statistical mechanics. Therefore, we should be able to make the equilibrium starting state with the aid of the hydrodynamic equation.

The starting point of this relaxation process is prepared by the following "adiabatic switching on" of the conjugate force h:

$$h(\boldsymbol{r}, t) = \begin{cases} 0 & t > 0, \\ h(\boldsymbol{r})e^{\epsilon t} & t < 0. \end{cases} \tag{X2.3.25}$$

Here, $\epsilon > 0$ is an infinitesimal positive number describing the very slow turning-on of the perturbation (*adiabatic turning-on*).

The result of this perturbation may be written with the aid of the response function as

$$\langle \psi(\boldsymbol{r}, t)\rangle_R = \int d\boldsymbol{r}' \int_{-\infty}^{0} dt' \, \phi(\boldsymbol{r} - \boldsymbol{r}', t - t')h(\boldsymbol{r}')e^{\epsilon t'}. \tag{X2.3.26}$$

From now on, let us write the suffix R to indicate that the calculated quantity is for the relaxational experiment. Beyond $t = 0$ this formula describes the relaxation of the system from the initial condition

$$\langle \psi(\boldsymbol{r}, 0)\rangle_R = \int d\boldsymbol{r}' \int_{0}^{\infty} dt' \, h(\boldsymbol{r}')\phi(\boldsymbol{r} - \boldsymbol{r}', t'). \tag{X2.3.27}$$

This value must be identical to the equilibrium response to the static field $h(\boldsymbol{r})$. Therefore,

$$\langle \psi(\boldsymbol{k}, t = 0)\rangle_R = \chi(\boldsymbol{k})h(\boldsymbol{k}). \tag{X2.3.28}$$

Exercise. Check that the above statement is consistent with (X2.3.17). □

X2.3.9 Admittance from hydrodynamics

From (X2.3.26) we obtain

$$\frac{\partial}{\partial t}\langle \psi(\boldsymbol{r}, t)\rangle_R = -\int d\boldsymbol{r}' \, \phi(\boldsymbol{r} - \boldsymbol{r}', t)h(\boldsymbol{r}'). \tag{X2.3.29}$$

From this

$$i\omega\langle \psi[\boldsymbol{k}, \omega]\rangle_R - \langle \psi(\boldsymbol{k}, t = 0)\rangle_R = -\chi(\boldsymbol{k}, \omega)h(\boldsymbol{k}). \tag{X2.3.30}$$

Here, we have used a consequence of causality (→**X2.2.9**): $\chi[\boldsymbol{k}, \omega] = \chi(\boldsymbol{k}, \omega)$. The initial condition is obtained from (X2.3.27) as

$$\langle \psi(\boldsymbol{k}, t = 0)\rangle_R = \chi(\boldsymbol{k})h(\boldsymbol{k}). \tag{X2.3.31}$$

Therefore, (X2.3.30) reads

$$\langle \psi[\boldsymbol{k}, \omega]\rangle_R = \frac{\chi(\boldsymbol{k}) - \chi(\boldsymbol{k}, \omega)}{i\omega}h(\boldsymbol{k}). \tag{X2.3.32}$$

We can also compute the relaxation of ψ with the initial condition (X2.3.27) with the aid of the hydrodynamic equation (X2.3.1) as

$$i\omega\langle \psi[\boldsymbol{k}, \omega]\rangle_R - \chi(\boldsymbol{k})h(\boldsymbol{k}) = -L_{\boldsymbol{k}}\langle \psi[\boldsymbol{k}, \omega]\rangle_R. \tag{X2.3.33}$$

That is,

$$\langle\psi[\boldsymbol{k},\omega]\rangle_R = \frac{\chi(\boldsymbol{k})}{i\omega + L_{\boldsymbol{k}}} h(\boldsymbol{k}). \tag{X2.3.34}$$

Equating (X2.3.32) and (X2.3.34), we obtain

$$\chi(\boldsymbol{k},\omega) = \frac{L_{\boldsymbol{k}}\chi(\boldsymbol{k})}{i\omega + L_{\boldsymbol{k}}}. \tag{X2.3.35}$$

X2.3.10 Relaxation in terms of the loss part of admittance

If we use the Kramers-Kronig relation (→**X2.2.10**), we have

$$\chi(\boldsymbol{k}) = \frac{1}{\pi} P \int \frac{\chi''(\boldsymbol{k},z)}{z} dz, \tag{X2.3.36}$$

$$\chi(\boldsymbol{k},\omega) = \frac{1}{\pi} P \int \frac{\chi''(\boldsymbol{k},z)}{z - (\omega - i\epsilon)} dz. \tag{X2.3.37}$$

Using these relations in (X2.3.32), we obtain

$$\langle\psi[\boldsymbol{k},z]\rangle_R = \frac{1}{i\pi} P \int dz\, \frac{\chi''(\boldsymbol{k},z)}{z(z - (\omega - i\epsilon))} h(\boldsymbol{k}). \tag{X2.3.38}$$

X2.3.11 Fluctuation dissipation relation of the first kind

Comparing (X2.3.35) with the correlation function obtained from hydrodynamics (X2.3.22), and noting that $\chi(\boldsymbol{k})$ is real, we may conclude that

$$\frac{\chi''(\boldsymbol{k},\omega)}{\omega} = \frac{\beta}{2} S(\boldsymbol{k},\omega). \tag{X2.3.39}$$

This is called the *fluctuation-dissipation relation of the first kind.*
Exercise. This is easy to see if $L_{\boldsymbol{k}}$ is real, but this need not be assumed. Demonstrate (X2.3.39). □.

Clearly recognize that this general relation is based on (i) the existence of linear phenomenological law (X2.3.1), (ii) space-time reversal symmetry (X2.3.5), and (iii) causality. The statistical assumption is hidden in deriving the equation for the correlation function from (i). That is, we have used Onsager's regression principle (→**4.1.2**). Then, no microscopic consideration is needed.

X2.3.12 Green-Kubo relation from hydrodynamics

Let us write the hydrodynamic equation with noise

$$\frac{\partial\psi}{\partial t} = -L\psi + w. \tag{X2.3.40}$$

Fourier-transforming this with space, we can regard this as a set of ordinary Langevin equations:

$$\frac{d\psi_{\boldsymbol{k}}}{dt} = -L_{\boldsymbol{k}}\psi_{\boldsymbol{k}} + w_{\boldsymbol{k}}. \tag{X2.3.41}$$

Then, the fluctuation dissipation relation of the first kind (→**X1.1.7**) should give us $L_{\boldsymbol{k}}$. To use this we must rewrite (X2.3.41) in the following form:

$$\frac{d\psi_{\boldsymbol{k}}}{dt} = -\Lambda_{\boldsymbol{k}} \frac{\partial S}{\partial \psi_{-\boldsymbol{k}}} + w_{\boldsymbol{k}}. \tag{X2.3.42}$$

Here, (recall the fluctuation theory, →**2.8.1**, **2.8.5**)

$$S = S_0 - \sum_{\boldsymbol{k}} \frac{\psi_{-\boldsymbol{k}}\psi_{\boldsymbol{k}}}{2Vk_BT\chi(\boldsymbol{k})}, \tag{X2.3.43}$$

and from (X1.1.19)

$$\Lambda_{\boldsymbol{k}} = \int_0^\infty ds \left\langle \frac{d\psi_{-\boldsymbol{k}}(0)}{dt} \frac{d\psi_{\boldsymbol{k}}(s)}{dt} \right\rangle. \tag{X2.3.44}$$

Therefore, $\Lambda_{\boldsymbol{k}}/Vk_BT\chi(\boldsymbol{k}) = L_{\boldsymbol{k}}$:

$$\chi(\boldsymbol{k})L_{\boldsymbol{k}} = \beta \frac{1}{V} \int_0^\infty ds \left\langle \frac{d\psi_{-\boldsymbol{k}}(0)}{dt} \frac{d\psi_{\boldsymbol{k}}(s)}{dt} \right\rangle. \tag{X2.3.45}$$

This can be rewritten in terms of the flux, if ψ is conserved, as we see in **X2.3.13**.

X2.3.13 Generalized diffusion constant from hydrodynamics

From (X2.3.35) we obtain

$$\lim_{\omega\to 0} \lim_{\boldsymbol{k}\to 0} \frac{\omega}{k^2}\chi''(\boldsymbol{k},\omega) = D\chi, \tag{X2.3.46}$$

where $\chi = \chi(0)$ and

$$D \equiv \lim_{\boldsymbol{k}\to 0} L_{\boldsymbol{k}}/k^2. \tag{X2.3.47}$$

With the aid of the fluctuation-dissipation relation (X2.3.39), this implies[109]

$$\begin{aligned} D\chi &= \lim_{\omega\to 0} \lim_{\boldsymbol{k}\to 0} \frac{\beta\omega^2}{2k^2} S(\boldsymbol{k},\omega), & \text{(X2.3.48)} \\ &= \lim_{\omega\to 0} \lim_{\boldsymbol{k}\to 0} \frac{\beta\omega^2}{2k^2} \int d\boldsymbol{r} \int_{-\infty}^\infty dt\, e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} \langle \psi(0,0)\psi(\boldsymbol{r},t)\rangle, & \text{(X2.3.49)} \\ &= \lim_{\omega\to 0} \lim_{\boldsymbol{k}\to 0} \frac{\beta}{2k^2} \int d\boldsymbol{r} \int_{-\infty}^\infty dt\, e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} \left\langle \frac{\partial}{\partial t}\psi(0,0) \frac{\partial}{\partial t}\psi(\boldsymbol{r},t) \right\rangle. & \text{(X2.3.50)} \end{aligned}$$

Now, we assume that the hydrodynamic equation has the following conservation law form:

$$\frac{\partial}{\partial t}\psi(\boldsymbol{r},t) = -div\, \boldsymbol{j}(\boldsymbol{r},t). \tag{X2.3.51}$$

---

[109] Detail:

$$\begin{aligned} \omega^2 \int dt\, e^{-i\omega t}\langle\psi(0)\psi(t)\rangle &= -\int dt \left(\frac{d^2}{dt^2}e^{-i\omega t}\right)\langle\psi(0)\psi(t)\rangle = \int dt \left(\frac{d}{dt}e^{-i\omega t}\right)\langle\psi(0)\frac{d}{dt}\psi(t)\rangle, \\ = &\int dt \left(\frac{d}{dt}e^{-i\omega t}\right)\langle\psi(-t)\frac{d}{dt}\psi(0)\rangle = -\int dt\, e^{-i\omega t}\left\langle\frac{d}{dt}\psi(-t)\frac{d}{dt}\psi(0)\right\rangle = \int dt\, e^{-i\omega t}\left\langle\frac{d}{dt}\psi(0)\frac{d}{dt}\psi(t)\right\rangle. \end{aligned}$$

The last equality is the same as the calculation in **4.1.12**.

Then, (X2.3.50) reads[110]

$$D\chi = \lim_{\omega\to 0}\lim_{\boldsymbol{k}\to 0}\frac{\beta}{2k^2}\int d\boldsymbol{r}\int_{-\infty}^{\infty} dt\, e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\left\langle \nabla\cdot\boldsymbol{j}(0,0)\,\nabla\cdot\boldsymbol{j}(\boldsymbol{r},t)\right\rangle, \tag{X2.3.52}$$

$$= \lim_{\omega\to 0}\lim_{\boldsymbol{k}\to 0}\frac{\beta}{2}\int d\boldsymbol{r}\int_{-\infty}^{\infty} dt\, e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\frac{1}{k^2}\left\langle \boldsymbol{k}\cdot\boldsymbol{j}(0,0)\,\boldsymbol{k}\cdot\boldsymbol{j}(\boldsymbol{r},t)\right\rangle. \tag{X2.3.53}$$

That is,[111]

$$D\chi = \lim_{\omega\to 0}\frac{\beta}{2d}\int d\boldsymbol{r}\int dt\, e^{-i\omega t}\langle\boldsymbol{j}(0,0)\cdot\boldsymbol{j}(\boldsymbol{r},t)\rangle. \tag{X2.3.54}$$

This is the Green-Kubo relation for D. The time integration may be from 0 to ∞ in this formula (but without 2 in the prefactor), so for an isotropic system, we have

$$D\chi = \frac{\beta}{d}\int_0^{\infty} dt\int d\boldsymbol{r}\,\langle\boldsymbol{j}(0,0)\cdot\boldsymbol{j}(\boldsymbol{r},t)\rangle. \tag{X2.3.55}$$

As is noted in **X2.3.12**, again this relation has been derived almost without microscopic consideration, so it should be a reliable result.

X2.3.14 Example: particle flux and diffusion constant

Consider the solute particle number density ρ that obeys the diffusion equation:

$$\frac{\partial\rho}{\partial t} = D\Delta\rho = -div\,\boldsymbol{j}, \tag{X2.3.56}$$

where $\boldsymbol{j} = -D\nabla\rho$. Then, D in this equation is exactly the D we computed in **X2.3.13**, and χ must be $\beta\langle\delta\rho\delta\rho\rangle$. If we may assume that solute particles are all independent, then

$$\langle\boldsymbol{j}\,\boldsymbol{j}(t)\rangle = \sum_i\langle\boldsymbol{v}_i\,\boldsymbol{v}_i(t)\rangle = \rho\langle\boldsymbol{v}\,\boldsymbol{v}(t)\rangle. \tag{X2.3.57}$$

Also, $\beta\langle\delta\rho\delta\rho\rangle$ may be computed by assuming the ideal gas law, and we recover the equation for D obtained in **3.2.9**.

X2.3.15 Fluctuation-dissipation relation of the second kind

We have learned that there are two kinds of fluctuation dissipation relations. Up to this point we have been studying the first kind. The second kind relates the noise w in the equation in (X2.3.1) + noise (cf. **3.2.5**, **X1.2.6**):

$$\frac{\partial}{\partial t}\psi = -L\psi + w \tag{X2.3.58}$$

---

[110] We assume the spatial translational symmetry. We need the same trick as used in (X2.3.50) or see **4.1.14**.

[111] We assume the isotropy of the system:

$$\langle k_i j_i k_l j_l\rangle = k_i k_i\langle\boldsymbol{j}\cdot\boldsymbol{j}\rangle/d$$

with the summation convention.

with L.

Let us study the same time average $\langle\psi_{-\boldsymbol{k}}(t)\psi_{\boldsymbol{k}}(t)\rangle$ for a sufficiently large t. Let us write

$$\frac{1}{V}\langle w_{-\boldsymbol{k}}w_{\boldsymbol{k}}(t)\rangle = 2\sigma(\boldsymbol{k})\delta(t). \tag{X2.3.59}$$

From the solution to (X2.3.58) (we may ignore the initial condition)

$$\langle\psi_{-\boldsymbol{k}}(t)\psi_{\boldsymbol{k}}(t)\rangle = L_{\boldsymbol{k}}^{-1}\sigma(\boldsymbol{k}). \tag{X2.3.60}$$

Here, we can relate $\langle\psi_{-\boldsymbol{k}}(t)\psi_{\boldsymbol{k}}(t)\rangle$ to the susceptibility in (X2.3.17), so

$$\sigma(\boldsymbol{k}) = k_BTL_{\boldsymbol{k}}\chi(\boldsymbol{k}). \tag{X2.3.61}$$

This is an expression of the fluctuation-dissipation relation of the second kind, but it can be rewritten in a more appealing form. Also notice that this is a generalization of the Einstein relation (→**3.1.3**)

X2.3.16 Random current and systematic current

We have rewritten the hydrodynamic equation in the conservation form (X2.3.51). We must also take into account the noise contribution:

$$\frac{\partial\psi}{\partial t} = -div(\boldsymbol{j}+\boldsymbol{j}'), \tag{X2.3.62}$$

where $\boldsymbol{j}$ is the systematic flux that can be written as

$$-div\boldsymbol{j} = -L\psi \tag{X2.3.63}$$

or

$$i\boldsymbol{k}\cdot\boldsymbol{j}_{\boldsymbol{k}} = -L_{\boldsymbol{k}}\psi_{\boldsymbol{k}} = (Dk^2+o[k^2])\psi_{\boldsymbol{k}}. \tag{X2.3.64}$$

For the long wave length limit, if we introduce the diffusion constant D as above, consistent with the D calculated in **X2.3.13**,

$$\boldsymbol{j} = -D\nabla\psi. \tag{X2.3.65}$$

Therefore, the fluctuation dissipation relation of the second kind should relate the random flux $\boldsymbol{j}'$ with D.

X2.3.17 Diffusion constant in terms of random current

Since

$$\frac{d}{dt}\psi_{\boldsymbol{k}} = -Dk^2\psi_{\boldsymbol{k}} + i\boldsymbol{k}\cdot\boldsymbol{j}'_{\boldsymbol{k}}, \tag{X2.3.66}$$

(X2.3.61) may be written in terms of the random current as

$$Dk^2\chi(\boldsymbol{k}) = \frac{\beta}{V}\int_0^\infty\langle\boldsymbol{k}\cdot\boldsymbol{j}'_{\boldsymbol{k}}\,\boldsymbol{k}\cdot\boldsymbol{j}'_{-\boldsymbol{k}}\rangle. \tag{X2.3.67}$$

That is,

$$D\chi = \lim_{\boldsymbol{k}\to 0}\frac{\beta}{dV}\int_0^\infty\langle\boldsymbol{j}'_{\boldsymbol{k}}\cdot\boldsymbol{j}'_{-\boldsymbol{k}}\rangle. \tag{X2.3.68}$$

Thus, we have arrived at a different expression of D, but must be equal to the result (X2.3.55). Note that in our case time decay is fast enough. In such cases, this equality is a general result.

X2.4 Hydrodynamics and Response — Quantum case

The correlation function $S(\boldsymbol{r},t)$ for a classical case is an even function of time (→**X2.3.2**) thanks to the commutativity of observables. However, in the quantum case, this is not generally true, so the Fourier-Laplace transform of the correlation function and its Fourier transform are not related simply as in **X2.3.6**. If we admit the following manifestly quantum mechanical relation (→**X2.4.5**)

$$S(\boldsymbol{k},\omega) = e^{-\beta\hbar\omega}S(-\boldsymbol{k},-\omega), \tag{X2.4.1}$$

where $\hbar$ is the Dirac constant, as a macroscopic relation, then we can almost mimic what we have done for the classical case.[112]

However, this may not be very intuitive. Therefore, our strategy is to derive the above relation from the calculation of the loss part of the admittance χ'' (→**X2.2.3**).

X2.4.1 Quantum transition rate due to external force

Let us apply a monochromatic stimulus (in phasor representation →**X2.2.7**)

$$h(\boldsymbol{r},t) = Re\, h(\boldsymbol{k},\omega)e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}. \tag{X2.4.2}$$

Then, the space-time-averaged power loss reads (→**X2.2.7**)

$$\langle p\rangle = \frac{1}{2}\omega V\chi''(\boldsymbol{k},\omega)|h(\boldsymbol{k},\omega)|^2, \tag{X2.4.3}$$

where V is the system volume.

The power absorbed by the system is lost as heat, so the power loss can be computed, if we can compute the transition probability due to the perturbation (X2.4.2).

If we assume that h to be the (energetic) conjugate variable of the thermodynamic density ψ, the perturbation Hamiltonian (in the Schrödinger picture; the observable ψ is not time-dependent) reads[113]

$$h = -\int d\boldsymbol{r}\, h(\boldsymbol{r},t)\psi(\boldsymbol{r}) = -Re\, h(\boldsymbol{k},\omega)\psi(-\boldsymbol{k})e^{i\omega t}. \tag{X2.4.4}$$

The golden rule tells us that the transition probability rate (probability per unit time) is given by[114]

$$W_{i\to f} = \frac{\pi}{2\hbar^2}|\langle i|h(\boldsymbol{k},\omega)\psi(-\boldsymbol{k})|f\rangle|^2\delta(\omega_f-\omega_i-\omega), \tag{X2.4.5}$$

where $\omega_f = E_{final}/\hbar$ and $\omega_i = E_{initial}/\hbar$.

---

[112] This is essentially the Callen and Welton theory generalizing Nyquist's theorem: H. B. Callen and T. A. Welton, "Irreversibility and generalized noise," Phys. Rev. **83**, 34-40 (1951).

[113] Why is there a minus sign such as $-HM$, even though in the Gibbs relation $dE = \cdots + HdM + \cdots$? Here, H is given, and we do not pay any attention to the energy required to change the field. Therefore, we must consider the energy change of the system under the given H; if due to H M changes by dM, the energy of the magnetic system actually decreases by HdM.

[114] Spontaneous emission is negligible.

X2.4.2 Transition rate in terms of equilibrium average

With the transition rate given by (X2.4.5), the absorption rate $p(\boldsymbol{k},\omega)$ of the energy quantum $\hbar\omega$ from the monochromatic perturbation (X2.4.2) is given by

$$p(\boldsymbol{k},\omega)=\sum_{i,f}\rho_i W_{i\to f}, \tag{X2.4.6}$$

where ρ_i is the initial distribution $\rho_i \propto e^{-\beta E_{initial}} = e^{-\beta\hbar\omega_i}$ ($\rho_i = \langle i|\rho_e|i\rangle$, where $|i\rangle$ is the initial eigenstate; there are many different initial states). Introducing (X2.4.5) into this, we obtain (cf., calculation in **3.7.10**)

$$\begin{aligned}
p(\boldsymbol{k},\omega) &= \sum_{i,f}\rho_i\frac{1}{4\hbar^2}|\langle i|h(\boldsymbol{k},\omega)\psi(-\boldsymbol{k})|f\rangle|^2\int_{-\infty}^{\infty}dt\, e^{it(\omega_f-\omega_i-\omega)}, && \text{(X2.4.7)}\\
&= \frac{1}{4\hbar^2}|h(\boldsymbol{k},\omega)|^2\sum_{if}\int_{-\infty}^{\infty}dt\,\langle i|\rho_e\psi(-\boldsymbol{k})|f\rangle\langle f|e^{iHt/\hbar}\psi(\boldsymbol{k})e^{-iHt/\hbar}|i\rangle e^{-i\omega t}, && \text{(X2.4.8)}\\
&= \frac{1}{4\hbar^2}|h(\boldsymbol{k},\omega)|^2\int_{-\infty}^{\infty}dt\, e^{-i\omega t}Tr(\rho_e\psi(-\boldsymbol{k})\psi(\boldsymbol{k},t)), && \text{(X2.4.9)}\\
&= \frac{1}{4\hbar^2}|h(\boldsymbol{k},\omega)|^2\int_{-\infty}^{\infty}dt\, e^{-i\omega t}\langle\psi(-\boldsymbol{k})\psi(\boldsymbol{k},t)\rangle. && \text{(X2.4.10)}
\end{aligned}$$

Here, $\psi(\boldsymbol{k},t)$ is the Heisenberg representation

$$\psi(\boldsymbol{k},t)=e^{iHt/\hbar}\psi(\boldsymbol{k})e^{-iHt/\hbar}. \tag{X2.4.11}$$

Thus, the final result reads

$$p(\boldsymbol{k},\omega)=\frac{V}{4\hbar^2}S(\boldsymbol{k},\omega)|h(\boldsymbol{k},\omega)|^2. \tag{X2.4.12}$$

Since ψ is real, notice $\psi(\boldsymbol{k})^*=\psi(-\boldsymbol{k})$. If the reader recalls the spectral analysis of noise (→**3.7.4**), she should have seen that the last term in $p(\boldsymbol{k},\omega)$ is related to the space-time correlation function of ψ (→(X2.3.15)).

X2.4.3 Relation between transition rate and correlation function

We have

$$\begin{aligned}
\langle\psi(-\boldsymbol{k})\psi(\boldsymbol{k},t)\rangle &= \int d\boldsymbol{r}_1\int d\boldsymbol{r}_2\,\langle e^{-i\boldsymbol{k}\cdot\boldsymbol{r}_1}\psi(\boldsymbol{r}_1,0)e^{i\boldsymbol{k}\cdot\boldsymbol{r}_2}\psi(\boldsymbol{r}_2,t)\rangle, && \text{(X2.4.13)}\\
&= \int d\rho\int d\boldsymbol{r}_1\, e^{i\boldsymbol{k}\cdot\boldsymbol{\rho}}\langle\psi(\boldsymbol{r}_1,0)\psi(\boldsymbol{r}_1+\boldsymbol{\rho},t)\rangle, && \text{(X2.4.14)}\\
&= VS(\boldsymbol{k},t), && \text{(X2.4.15)}
\end{aligned}$$

which is the spatial Fourier transform of the space-time correlation function

$$\begin{aligned}
C(\boldsymbol{r},t) &= \langle\psi(0,0)\psi(\boldsymbol{r},t)\rangle, && \text{(X2.4.16)}\\
S(\boldsymbol{k},\omega) &= \int d\boldsymbol{r}\int dt\, C(\boldsymbol{r},t)e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}. && \text{(X2.4.17)}
\end{aligned}$$

This is essentially the Wiener-Khinchine theorem (→**3.7.5**).

Combining (X2.4.10) and (X2.4.15), we obtain

$$p(\boldsymbol{k},\omega) = \frac{V}{4\hbar^2}S(\boldsymbol{k},\omega)|h(\boldsymbol{k},\omega)|^2. \tag{X2.4.18}$$

X2.4.4 Power absorption or energy dissipation rate

$p(\boldsymbol{k},\omega)$ given by (X2.4.18) is the transition rate of the system in equilibrium by absorbing the energy quantum $\hbar\omega$. The same perturbation (X2.4.2) also induces the opposite transition: the emission $\hbar\omega$. This rate is given by

$$p(-\boldsymbol{k},-\omega) = \frac{V}{4\hbar^2}S(-\boldsymbol{k},-\omega)|h(\boldsymbol{k},\omega)|^2. \tag{X2.4.19}$$

Therefore, the net power absorption (loss) by the system is given by

$$\langle p\rangle = \hbar\omega[p(\boldsymbol{k},\omega) - p(-\boldsymbol{k},-\omega)]. \tag{X2.4.20}$$

With the aid of (X2.4.18) and (X2.4.3), we arrive at

$$\chi''(\boldsymbol{k},\omega) = \frac{1}{2\hbar}[S(\boldsymbol{k},\omega) - S(-\boldsymbol{k},-\omega)]. \tag{X2.4.21}$$

As seen from $\hbar$ in the denominator, this is purely quantum mechanical. That is why this formula cannot be derived without mechanical considerations.

However, never regard the above derivation of (X2.4.21) a respectable first principle calculation. We have assumed that the system is in equilibrium despite the steady energy input due to the perturbation (X2.4.2). Our system Hamiltonian H is the Hamiltonian of an isolated system, so with this perturbation the system, honestly computed as a quantum mechanical system, must eventually be hot, but the dissipated energy is assumed to be harmlessly discarded.[115]

X2.4.5 Relation between $S(\boldsymbol{k},\omega)$ and $S(-\boldsymbol{k},-\omega)$

To obtain the relation corresponding to (X2.3.39), we need a relation between $S(\boldsymbol{k},\omega)$ and $S(-\boldsymbol{k},-\omega)$. In the classical case this is easy (→**X2.3.6**), and everything could be done without detailed mechanics, so we could totally avoid the serious difficulty inherent in its purely mechanical derivation mentioned at the end of **X2.4.4**.

Notice that

$$\frac{S(\boldsymbol{k},\omega)}{S(-\boldsymbol{k},-\omega)} = \frac{p(\boldsymbol{k},\omega)}{p(-\boldsymbol{k},-\omega)} = \frac{\sum_{i,f}\rho_i W_{i\to f}}{\sum_{i,f}\rho_f W_{f\to i}} \tag{X2.4.22}$$

thanks to (X2.4.18) and (X2.4.6).

For all the pairs (i,f) there is a one to one correspondence between the terms $i\to f$ on the denominator and $f\to i$ on the numerator. $W_{i\to f} = W_{f\to i}$ due to the Hermitian nature of

---

[115] If there is no dissipation, then the mechanical derivation can even be regorized. This is the case of the Hall conductance. In this case the electrical current and the driving electric field are orthogonal, so there is no dissipation due to the Hall effect.

the perturbative Hamiltonian.[116] For each pair (i, f) for which $W_{i\to f} \neq 0$ $E_{final} - E_{initial} = \hbar\omega$. Hence,

$$\rho_f = e^{-\beta\hbar\omega}\rho_i. \tag{X2.4.23}$$

Thus we have arrived at the following fundamental relation (an example of the KMS condition **X2.4.11**):

$$S(-\boldsymbol{k}, -\omega) = e^{-\beta\hbar\omega} S(\boldsymbol{k}, \omega). \tag{X2.4.24}$$

X2.4.6 Fluctuation dissipation relation for quantum case

With the aid of (X2.4.24) (X2.4.21) reads

$$\chi''(\boldsymbol{k}, \omega) = \frac{1}{2\hbar}(1 - e^{-\beta\hbar\omega}) S(\boldsymbol{k}, \omega). \tag{X2.4.25}$$

This is the *fluctuation dissipation relation of the first kind.*

Notice that the classical limit $\hbar \to 0$ of the above formula indeed reduces to the classical result (X2.3.39).

X2.4.7 Relation between χ'' and commutator

Let us rewrite (X2.4.21) with the aid of (X2.4.17) as

$$\begin{aligned} \chi''(\boldsymbol{k}, \omega) &= \frac{1}{2\hbar}\int d\boldsymbol{r} \int dt\, \langle \psi(0,0)\psi(\boldsymbol{r}, t)\rangle \left[e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})} - e^{-i(-\omega t + \boldsymbol{k}\cdot\boldsymbol{r})}\right], \qquad \text{(X2.4.26)} \\ &= \frac{1}{2\hbar}\int d\boldsymbol{r} \int dt\, \left[\langle \psi(0,0)\psi(\boldsymbol{r}, t)\rangle - \langle \psi(\boldsymbol{r}, t)\psi(0,0)\rangle\right] e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}. \qquad \text{(X2.4.27)} \end{aligned}$$

Therefore,

$$\chi''(\boldsymbol{k}, \omega) = \frac{1}{2\hbar}\int d\boldsymbol{r} \int dt\, \langle [\psi(0,0), \psi(\boldsymbol{r}, t)]\rangle e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}. \tag{X2.4.28}$$

Here, we have assumed that the system is translationally symmetric in space and time.

X2.4.8 Response function is essentially a commutator

We know we can reconstruct the whole admittance from its loss part with the aid of the Kramers-Kronig relation (the spectral representation →**X2.2.12**).

We make (X2.2.45):

$$\begin{aligned} \tilde{\chi}(\boldsymbol{k}, z) &= \frac{1}{2\hbar\pi}\int d\boldsymbol{r} \int dt \int_{-\infty}^{\infty} d\omega\, \frac{e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}}{\omega - z} \langle [\psi(0,0), \psi(\boldsymbol{r}, t)]\rangle, \qquad \text{(X2.4.29)} \\ &= -\frac{i}{\hbar}\int d\boldsymbol{r} \int dt \langle [\psi(0,0), \psi(\boldsymbol{r}, t)]\rangle \Theta(t) e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}, \qquad \text{(X2.4.30)} \end{aligned}$$

where $\Theta(t)$ is the Heaviside step function. Therefore, (X2.2.46) tells us

$$\chi(\boldsymbol{k}, \omega) = \frac{i}{\hbar}\int dt \int d\boldsymbol{r} \langle [\psi(0,0), \psi(\boldsymbol{r}, t)]\rangle e^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}\Theta(t). \tag{X2.4.31}$$

That is,

$$\phi(\boldsymbol{r}, t) = \frac{i}{\hbar}\langle [\psi(0,0), \psi(\boldsymbol{r}, t)]\rangle \Theta(t). \tag{X2.4.32}$$

---

[116] This is a perturbative version of the detailed balance (→**2.7.6**) of mechanics.

X2.4.9 Green-Kubo relation: quantum version

We have derived the Green-Kubo relation for the classical case (→**X2.3.12**, **X2.3.13**). There, the diffusion constant is related to the correlation function.

Quantum mechanically, let us consider the symmetrized correlation first:

$$\int d\boldsymbol{r}\int dt\,\langle[\psi(0,0),\psi(\boldsymbol{r},t)]_+\rangle e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}=S(\boldsymbol{k},\omega)+S(-\boldsymbol{k},-\omega). \tag{X2.4.33}$$

Here, $[A,B]_+=AB+BA$ is the anticommutator. Thanks to the following relations obtained from (X2.4.25), we have

$$\frac{1}{2\hbar}S(\boldsymbol{k},\omega) = \frac{\chi''(\boldsymbol{k},\omega)}{1-e^{-\beta\hbar\omega}}, \tag{X2.4.34}$$

$$\frac{1}{2\hbar}S(-\boldsymbol{k},-\omega) = \frac{\chi''(\boldsymbol{k},\omega)}{e^{\beta\hbar\omega}-1}. \tag{X2.4.35}$$

(X2.4.33) now reads

$$\frac{1}{2\hbar}\int d\boldsymbol{r}\int dt\,\langle[\psi(0,0),\psi(\boldsymbol{r},t)]_+\rangle e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}=\chi''(\boldsymbol{k},\omega)\coth\frac{\beta\hbar\omega}{2}. \tag{X2.4.36}$$

Since the relation (X2.3.46) comes from the phenomenological equation that defines D, we compute this with the aid of (X2.4.36):

$$D\chi = \lim_{\omega\to0}\lim_{\boldsymbol{k}\to0}\frac{1}{2\hbar}\tanh\frac{\beta\hbar\omega}{2}\frac{\omega}{k^2}\int d\boldsymbol{r}\int dt\,\langle[\psi(0,0),\psi(\boldsymbol{r},t)]_+\rangle e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}, \tag{X2.4.37}$$

$$= \lim_{\omega\to0}\frac{\beta}{4d}\int d\boldsymbol{r}\int dt\,e^{-i\omega t}\langle[\boldsymbol{j}(0,0),\boldsymbol{j}(\boldsymbol{r},t)]_+\rangle. \tag{X2.4.38}$$

This is the quantum version of the Green-Kubo relation for D.

X2.4.10 Relation between $S(\boldsymbol{k},\omega)$ and $S(-\boldsymbol{k},-\omega)$ revisited

In **X2.4.5** we have related $S(\boldsymbol{k},\omega)$ and $S(-\boldsymbol{k},-\omega)$ with the aid of physical reasoning. Here, let us reconsider the relation from the symmetry point of view.

$$S(-\boldsymbol{k},-\omega) = \int d\boldsymbol{r}\int dt\,e^{i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\langle\psi(0,0)\psi(\boldsymbol{r},t)\rangle, \tag{X2.4.39}$$

$$= \int d\boldsymbol{r}\int dt\,e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\langle\psi(0,0)\psi(-\boldsymbol{r},-t)\rangle. \tag{X2.4.40}$$

Here, we have changed the integration variables as $t\to-t$ and $\boldsymbol{r}\to-\boldsymbol{r}$. Explicitly writing the equilibrium average in terms of the canonical density operator, we have

$$S(-\boldsymbol{k},-\omega) = \frac{1}{Z}\int d\boldsymbol{r}\int dt\,e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}Tr\left(e^{-\beta H}\psi(\boldsymbol{r},t)\psi(0,0)\right), \tag{X2.4.41}$$

$$= \frac{1}{Z}\int d\boldsymbol{r}\int dt\,e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}Tr\left(\psi(0,0)e^{-\beta H}\psi(\boldsymbol{r},t)\right). \tag{X2.4.42}$$

Here, in the first line we have translated the coordinates, and then cyclic permutation of operators has been done in the trace.

Now, we wish to explicitly write $\psi(\boldsymbol{r},t)$ as an operator in the Heisenberg picture:

$$\begin{aligned} S(-\boldsymbol{k},-\omega) &= \frac{1}{Z}\int d\boldsymbol{r}\int dt\, e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}Tr\left(\psi(0,0)e^{-\beta H}\psi(\boldsymbol{r},t)e^{\beta H}e^{-\beta H}\right), && \text{(X2.4.43)} \\ &= \frac{1}{Z}\int d\boldsymbol{r}\int dt\, e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}Tr\left(e^{-\beta H}\psi(0,0)e^{(i/\hbar)H(t+i\beta\hbar)}\psi(\boldsymbol{r},0)e^{-(i/\hbar)H(t+i\beta\hbar)}\right). && \text{(X2.4.44)} \end{aligned}$$

That is,

$$S(-\boldsymbol{k},-\omega) = \int d\boldsymbol{r}\int dt\, e^{-i(\omega t-\boldsymbol{k}\cdot\boldsymbol{r})}\left\langle\psi(0,0)e^{(i/\hbar)H(t+i\beta\hbar)}\psi(\boldsymbol{r},0)e^{-(i/\hbar)H(t+i\beta\hbar)}\right\rangle. \qquad \text{(X2.4.45)}$$

Introducing the new variable $\tau = t+i\beta\hbar$, and noting that $A(\tau) = e^{iH\tau/\hbar}Ae^{-iH\tau/\hbar}$, the above relation reads

$$S(-\boldsymbol{k},-\omega) = e^{-\beta\hbar\omega}\int d\boldsymbol{r}\int_{-\infty+i\beta\hbar}^{\infty+i\beta\hbar} d\tau\, e^{-i(\omega\tau-\boldsymbol{k}\cdot\boldsymbol{r})}\langle\psi(0,0)\psi(\boldsymbol{r},\tau)\rangle, \qquad \text{(X2.4.46)}$$

This is close to what we wish to have:

$$S(-\boldsymbol{k},-\omega) = e^{-\beta\hbar\omega}S(\boldsymbol{k},\omega). \qquad \text{(X2.4.47)}$$

We need an extra condition, the KMS condition **X2.4.11**, to reach it.

X2.4.11 Kubo-Martin-Schwinger condition If $S(\boldsymbol{r},t)$ is analytic as a function of t on the strip $Im\, z \in (0,\beta\hbar)$, we may move the integration path for τ in (X2.4.47) to the real axis, so, we obtain

$$S(-\boldsymbol{k},-\omega) = e^{-\beta\hbar\omega}S(\boldsymbol{k},\omega). \qquad \text{(X2.4.48)}$$

This equality is what we have used in the derivation of the fluctuation dissipation relation (→(X2.4.24)).

The analyticity condition on the strip $Im\, t \in (0,\beta\hbar)$ of the correlation function is called the *Kubo-Martin-Schwinger condition* (*KMS condition*).

X2.4.12 Importance of KMS condition

The *KMS condition* may be summarized as follows:

KMS condition. Let T_t be the time evolution operator of a system and A and B be arbitrary observables of the system (e.g., $T_tA(s) = A(s+t)$). The density operator ρ (often called a *state*) is said to satisfy the KMS condition for a positive number β and for T_t, if

$$F_{AB}(t) = Tr\rho AT_tB = \langle AB(t)\rangle \qquad \text{(X2.4.49)}$$

is analytic on the strip $Im\, t \in (0,\beta\hbar)$. □

Notice that

$$F_{BA}(t) = \langle BA(t)\rangle = F_{AB}(t+i\beta\hbar). \qquad \text{(X2.4.50)}$$

The KMS condition characterizes the equilibrium *Gibbs state*. A remarkable thing is that Gibbs state dictates the time evolution operator conversely:

Roughly speaking,

(1) $T_t = e^{iHt/\hbar}$ and ρ satisfies the KMS condition for some β and for T_t, then $\rho \propto e^{-\beta H}$.[117]

(2) $\rho \propto e^{-\beta H}$ satisfies the KMS condition for some β and for T_t, then $T_t = e^{iHt/\hbar}$.[118]□

---

[117] Cf. R B Israel, *Convexity in the Theory of Lattice Gases* (Princeton UP).

[118] The Tomita-Takesaki theory, there is no good introduction.

X2.5 Linear Response — General Theory

X2.5.1 Time evolution of entropy operator in open system[119],

Let H_u be the Hamiltonian of an isolated system that consists of a system under consideration + its surrounding reservoirs

$$H_u = H + H_r + U, \tag{X2.5.1}$$

where H is the system's intrinsic Hamiltonian, H_r the Hamiltonian of the reservoirs, and U the system-reservoir interaction Hamiltonian. The equation of motion for the full density operator ρ_u is

$$i\hbar\frac{d\rho_u}{dt} = [H_u, \rho_u] \tag{X2.5.2}$$

The system density operator ρ is obtained as $\rho = Tr_r\rho_u$, where Tr_r traces out the bath states. Therefore,[120]

$$i\hbar\frac{d\rho}{dt} = [H, \rho] + Tr_r[U, \rho_u]. \tag{X2.5.3}$$

The last term is the effect of the baths. If we could write

$$Tr_r[U, \rho_u] = [h, \rho] \tag{X2.5.4}$$

with the aid of a certain operator h, then the system virtually behaves just as an isolated system. Therefore, such a rewriting should not be possible. Since the equation must be a linear equation, (X2.5.3) should have the following form

$$i\hbar\frac{d\rho}{dt} = [H', \rho] + i\hbar q\rho, \tag{X2.5.5}$$

where the operator q describes the effects of baths that cannot be absorbed in an effective Hamiltonian H' and will be specified later with the aid of phenomenology.

Let us introduce

$$\eta \equiv \log\rho, \tag{X2.5.6}$$

which is called the *entropy operator* of the system. It is easy to check

$$i\hbar\frac{d}{dt}\eta = [H', \eta] + i\hbar q. \tag{X2.5.7}$$

We must express q with the aid of phenomenological considerations. (X2.5.7) suggests that q must be related to the surface term in the entropy balance equation expressed in terms of the fluxes corresponding to thermodynamic densities. Therefore, q is the divergence of the entropy flux.

$$q = -div\left(k_B^{-1}\sum_i \cdot F_i\boldsymbol{j}_i\right). \tag{X2.5.8}$$

---

[119] The exposition here is based on J. A. McLennan, *Introduction to Nno-Equilibrium Statistical Mechanics*, (Prentice Hall, 1989), but there may be sign mistakes.

[120] Only an honest calcualtion is needed: for example, $Tr_r(H\rho_u) = H\rho$ and $Tr_r[H_r, \rho] = 0$, because the system variables not traced out are just mere parameters.

Here, $\boldsymbol{j}_i$ are understood as functions of the dynamical variables of the system. Let the time evolution operator corresponding to H' be $T(t)$. Then, starting from the local equilibrium distribution specified by A_i as $\rho_0 \propto \exp\left(\frac{1}{k_B}\sum_i F_i(\boldsymbol{r})A_i(\boldsymbol{r})\right)$, we obtain

$$\eta(t) = \phi(t) + \frac{1}{k_B}\sum_i \int d\boldsymbol{r}\, F_i(\boldsymbol{r})T(-t)A_i(\boldsymbol{r}) - \frac{1}{k_B}\sum_i \int_0^t ds \int_{\partial S} d\boldsymbol{S}\cdot F_i(\boldsymbol{r},s)T(s-t)\boldsymbol{j}_i(\boldsymbol{r},s), \tag{X2.5.9}$$

where ϕ is determined by the normalization condition, ∂S is the boundary of the system, and $\boldsymbol{S}$ is the surface element.

Now, note the following relations (we have only to compute (X2.5.9) for each i) :

$$\begin{aligned} F(\boldsymbol{r})T(-t)A(\boldsymbol{r}) &= F(\boldsymbol{r})A(\boldsymbol{r},-t) = -\int_0^t ds\, \frac{\partial}{\partial s}\left[F(\boldsymbol{r},s)A(\boldsymbol{r},s-t)\right] + F(\boldsymbol{r},t)A(\boldsymbol{r}), \\ &= F(\boldsymbol{r},t)A(\boldsymbol{r}) - \int_0^t ds\, \frac{\partial F(\boldsymbol{r},s)}{\partial s}A(\boldsymbol{r},s-t) - \int_0^t ds\, F(\boldsymbol{r},s)\frac{\partial A(\boldsymbol{r},s-t)}{\partial s}, \end{aligned} \tag{X2.5.10}$$

and

$$\int_0^t ds \int_{\partial S} d\boldsymbol{S}\cdot F(\boldsymbol{r},s)T(s-t)\boldsymbol{j}(\boldsymbol{r},s) = \int_0^t ds \int d\boldsymbol{r}\, div\left[F(\boldsymbol{r},s)T(s-t)\boldsymbol{j}(\boldsymbol{r},s)\right]. \tag{X2.5.11}$$

Therefore,

$$\begin{aligned} &\int d\boldsymbol{r}\, F(\boldsymbol{r})T(-t)A(\boldsymbol{r}) - \int_0^t \int_{\partial S} d\boldsymbol{S}\cdot F(\boldsymbol{r},s)T(s-t)\boldsymbol{j}(\boldsymbol{r},s) \\ =\ & \int d\boldsymbol{r}\, F(\boldsymbol{r},t)A(\boldsymbol{r}) - \int_0^t ds \int d\boldsymbol{r}\, \frac{\partial F(\boldsymbol{r},s)}{\partial s}A(\boldsymbol{r},s-t) - \int_0^t ds \int d\boldsymbol{r}\, F(\boldsymbol{r},s)\frac{\partial A(\boldsymbol{r},s-t)}{\partial s} \\ &\quad - \int_0^t \int d\boldsymbol{r}\, div\left[F(\boldsymbol{r},s)T(s-t)\boldsymbol{j}(\boldsymbol{r},s)\right] ds, \end{aligned} \tag{X2.5.12}$$

$$= \int d\boldsymbol{r}\, F(\boldsymbol{r},t)A(\boldsymbol{r}) - \int_0^t ds \int d\boldsymbol{r} \left(\frac{\partial F(\boldsymbol{r},s)}{\partial s}A(\boldsymbol{r},s-t) + \nabla F(\boldsymbol{r},s)\cdot\boldsymbol{j}(\boldsymbol{r},s-t)\right) \tag{X2.5.13}$$

$$= \int d\boldsymbol{r}\, F(\boldsymbol{r},t)A(\boldsymbol{r}) - \int_{-t}^0 ds \int d\boldsymbol{r} \left(\frac{\partial F(\boldsymbol{r},t+s)}{\partial s}A(\boldsymbol{r},s) + \nabla F(\boldsymbol{r},t+s)\cdot\boldsymbol{j}(\boldsymbol{r},s)\right). \tag{X2.5.14}$$

Here, we have used the conservation law:

$$\frac{\partial A}{\partial t} = -div\boldsymbol{j}. \tag{X2.5.15}$$

Taking into account the time scales as Mori's original approaches, the above result can be molded into the final form:

$$\eta(t) = \phi(t) + \frac{1}{k_B}\int d\boldsymbol{r}\, F(\boldsymbol{r},t)A(\boldsymbol{r}) - \frac{1}{k_B}\int_{-t}^0 ds \int d\boldsymbol{r} \left(\frac{\partial F(\boldsymbol{r},t+s)}{\partial s}A(\boldsymbol{r},s) + \nabla F(\boldsymbol{r},t+s)\cdot\boldsymbol{j}(\boldsymbol{r},s)\right). \tag{X2.5.16}$$

X2.5.2 Nonequilibrium statistical operator[121]

Now, to obtain the *nonequilibrium statistical operator* = density operator for nonequilibrium states we exponentiate η as $\rho = e^{\eta}$. To use it for the first order calculation, we must expand it in terms of $\delta\eta$. This can be done with the aid of (4.3.9). Thus,

$$\rho(t) = \rho_L(t)\Bigg\{1 - T\int_{-\infty}^{0} du \int d\boldsymbol{r} \sum_j e^{\epsilon u} \int_0^{\beta} d\lambda \, \Big[\nabla F_j(\boldsymbol{r}, t+u)\cdot \boldsymbol{j}_j(\boldsymbol{r}, u+i\lambda\hbar)$$
$$+ \quad \frac{\partial}{\partial u} F_j(\boldsymbol{r}, t+u)\hat{A}_j(\boldsymbol{r}, u+i\lambda\hbar)\Big]\Big]\Bigg\}. \tag{X2.5.17}$$

For the case without spatial dependence this reads

$$\rho(t) = \rho_e \left[1 - T\int_{-\infty}^{0} du \sum_j e^{\epsilon u} F_j(t+u) \int_0^{\beta} d\lambda \, \dot{\hat{A}}_j(u+i\lambda\hbar)\right]. \tag{X2.5.18}$$

X2.5.3 Response function

Let us consider a system under a perturbation that perturb $\hat{A}$ directly. For simplicity, we assume that the system is spatially uniform. Let us assume the perturbative Hamiltonian to be

$$h = -X(t)\hat{A}. \tag{X2.5.19}$$

The effect may be expressed in terms of the nonequilibrium statistical operator (→(X2.5.18)) as follows[122]

$$\rho(t) = \rho_e \left[1 + \int_{-\infty}^{0} du \, X(t+u) \int_0^{\beta} d\lambda \, \dot{\hat{A}}(u - i\lambda\hbar)\right]. \tag{X2.5.20}$$

Here, we set $-TF = X$. The average of another observable B is given by (let us assume, for simplicity, that its average is zero in equilibrium (without the perturbation))

$$\langle B\rangle(t) = \int_{-\infty}^{0} du \, X(t+u) \int_0^{\beta} d\lambda \, \langle \dot{\hat{A}}(u - i\lambda\hbar) B\rangle_e, \tag{X2.5.21}$$

$$= \int_{-\infty}^{t} du \, X(s)\beta\langle \dot{\hat{A}}(s-t); B\rangle_e. \tag{X2.5.22}$$

Therefore, the response function ϕ_{BA} reads

$$\phi_{BA}(t) = \beta\langle \dot{\hat{A}}; B(t)\rangle_e. \tag{X2.5.23}$$

This can be further transformed to our familiar form in terms of the commutator (→**X2.4.8**) with the aid of *Kubo's identity* (→**X2.5.4**).

---

[121] This form was obtained by Nakajima (1957) and Zubarev (1961-5). D. N. Zubarev, *Nonequilibrium Statistical Mechanics* (Consultants Bureau, New York, 1974).

[122] The mismatch of the normalization constants due to perturbation does not contribute the final result, so long as we use the operator to compute the expectation values of the obaservables whose equilibrium averages vanish.

X2.5.4 Kubo identity

To show (X2.5.27) we start with

$$\frac{d}{d\beta}[\hat{A}, \rho_e] = -i\hbar\dot{\hat{A}}\rho_e - H[\hat{A}, \rho_e]. \tag{X2.5.24}$$

Now, let us write

$$[\hat{A}, \rho_e] = \rho_e C(\beta). \tag{X2.5.25}$$

Put this into (X2.5.24), and we obtain

$$\dot{C}(\beta) = -i\hbar\dot{\hat{A}}(-i\beta\hbar). \tag{X2.5.26}$$

This is

$$[\rho_e, \hat{A}] = i\hbar \int_0^\beta d\lambda\, \rho_e \dot{\hat{A}}(-i\lambda\hbar) \tag{X2.5.27}$$

This is called *Kubo's identity.*

X2.5.5 Response function in terms of commutator

Let us use Kubo's identity (X2.5.27) to rewrite (X2.5.23) as

$$\phi_{BA}(t) = \frac{1}{i\hbar} Tr\left([\rho_e, \hat{A}(-t)]B\right), \tag{X2.5.28}$$

$$= \frac{1}{i\hbar}\int_0^\beta d\lambda\, Tr(\rho_e[\hat{A}(-t), B]). \tag{X2.5.29}$$

Therefore, we have arrived at the formula we already know:

$$\phi_{B\hat{A}}(t) = \frac{i}{\hbar} Tr \rho_e [B(t), \hat{A}]. \tag{X2.5.30}$$

This is essentially the form we already know (→**X2.4.8**).

X2.5.6 Response function: spatially nonuniform case

It should be almost obvious that

$$\phi_{B\hat{A}}(\boldsymbol{r}, t) = \frac{i}{\hbar} Tr \rho_e [B(\boldsymbol{r}, t), \hat{A}(0, 0)]. \tag{X2.5.31}$$

X2.5.7 Canonical correlation function: summary

We have already introduced the canonical correlation function (→**4.3.5**)[123]

$$\langle X; Y\rangle = k_B T \int_0^\beta d\lambda\, Tr(\rho_e X(-i\lambda\hbar)Y). \tag{X2.5.32}$$

---

[123] Here, for simplicity, we discuss only the cases with X and Y being self-adjoint.

In terms of this the response function may be written, as we know well by now (→**X2.5.3**)

$$\phi_{B\hat{A}}(t) = \beta\langle\hat{\dot{A}}(0); B(t)\rangle = -\beta\langle\hat{A}; \dot{B}(t)\rangle. \tag{X2.5.33}$$

We have already seen how natural the canonical correlation function is in various contexts.[124] Let us summarize its properties.

[A] $\langle X; Y\rangle$ is a scalar product. That is,
- (i) $\overline{\langle X; Y\rangle} = \langle Y; X\rangle$,
- (ii) bilinear,
- (iii) $\langle X; X\rangle \geq 0$ and $X = 0$ iff $\langle X; X\rangle = 0$.

[B] Stationarity: $\langle X(t); Y(t+s)\rangle = \langle X; Y(s)\rangle$.
[C] $\langle X; Y\rangle$ is real.[125]
[D] Symmetry: $\langle X; Y\rangle = \langle Y; X\rangle$.
[E] Let L be the Liouvillian (→**X1.2.4**) of the natural motion. Then

$$\langle LX; Y\rangle = \langle X; LY\rangle. \tag{X2.5.34}$$

[F] Reciprocity:

$$\langle X; Y(t)\rangle_{\boldsymbol{H},\boldsymbol{\omega}} = \epsilon_X\epsilon_Y\langle Y; X(t)\rangle_{-\boldsymbol{H},-\boldsymbol{\omega}}. \tag{X2.5.35}$$

Here, $\boldsymbol{H}$ is the magnetic field and $\boldsymbol{\omega}$ is the rotation angular velocity.

Using [F] and (X2.5.33), we conclude

$$\phi_{BA}(t)_{\boldsymbol{H},\boldsymbol{\omega}} = \epsilon_A\epsilon_B\phi_{AB}(t)_{-\boldsymbol{H},-\boldsymbol{\omega}}. \tag{X2.5.36}$$

It is a good exercise to demonstrate [A] - [F].

X2.5.8 Symmetrized correlation functions

We know that replacing the correlation function with the canonical correlation function, we can generalize the classical results to their quantum counterparts. However, initially it was thought that symmetrization of correlations $\langle AB\rangle \to \langle AB + BA\rangle/2$ was the right procedure for this generalization. Let us define the *symmetrized correlation function*

$$C(X; Y) \equiv \frac{1}{2}Tr\rho_e[XY + YX]. \tag{X2.5.37}$$

Its salient features may be summarized as follows:

[A] $C(X; Y)$ is a scalar product:
- (i) $\overline{C(X; Y)} = C(Y; X)$,
- (ii) bilinear,
- (iii) $C(X; X) \geq 0$ and $X = 0$ iff $C(X; X) = 0$.

[B] Stationarity: $C(X(t); Y(t+s)) = C(X; Y(s))$.
[C] $C(X; Y)$ is real.[126]
[D] Symmetry: $C(X; Y) = C(Y; X)$.
[E] Let L be the Liouvillian (→) of the natural motion. Then

$$C(LX; Y) = C(X; LY). \tag{X2.5.38}$$

---

[124] It is the retarded two time Green's function in the field theoretical approach to many body theory.

[125] if X and Y are self-adjoint, but here we assume this.

[126] if X and Y are self-adjoint, but throughout this section we assume dynamical operators are all self-adjoint.

[F] Reciprocity:

$$C(X;Y(t))_{\boldsymbol{H},\boldsymbol{\omega}} = \epsilon_X\epsilon_Y C(Y;X(t))_{-\boldsymbol{H},-\boldsymbol{\omega}}. \tag{X2.5.39}$$

Here, $\boldsymbol{H}$ is the magnetic field and $\boldsymbol{\omega}$ is the rotation angular velocity.

Thus, there is a very good parallelism between the canonical and symmetrized correlation functions. Their precise relation is the fluctuation dissipation relation originally obtained by Callen and Welton.[127]

X2.5.9 Relation between canonical and symmetrized correlations

The canonical correlation is closely related to responses, and (symmetrized) correlations are more directly related to observable quantities such as scattering functions and power spectra. Therefore, it should be very convenient, if we have a direct relation between them. Actually, this direct relation is the essence of fluctuation dissipation relations.

The relation we seek is

$$\frac{\beta\hbar\omega}{2}\coth\frac{\beta\hbar\omega}{2}\int dt\,\langle X_{-\boldsymbol{k}}(0);Y_{\boldsymbol{k}}(t)\rangle e^{-i\omega t} = \int dt\,C(X_{-\boldsymbol{k}};Y_{\boldsymbol{k}}(t))e^{-i\omega t}. \tag{X2.5.40}$$

Let us start from the canonical correlation: Since $Y\rho_e = \rho_e Y(-i\beta)$ (→(4.3.10))

$$\int dt\,e^{-i\omega t}Tr(\rho_e X_{-\boldsymbol{k}}(0)Y_{\boldsymbol{k}}(t)) = \int dt\,e^{-i\omega t}Tr(X_{-\boldsymbol{k}}(0)\rho_e Y_{\boldsymbol{k}}(t-\beta i\hbar)), \tag{X2.5.41}$$

$$= \int dt\,e^{-i\omega t}Tr(\rho_e Y_{\boldsymbol{k}}(t-i\beta\hbar)X_{-\boldsymbol{k}}(0)). \tag{X2.5.42}$$

Now, we assume the KMS condition (→**X2.4.12**):

$$= \int dt\,e^{-i\omega(t+i\beta\hbar)}Tr[\rho_e Y_{\boldsymbol{k}}(t)X_{-\boldsymbol{k}}(0)), \tag{X2.5.43}$$

$$= e^{\beta\hbar\omega}\int dt\,e^{-i\omega t}\langle Y_{\boldsymbol{k}}(t)X_{-\boldsymbol{k}}(0)\rangle_e. \tag{X2.5.44}$$

Therefore, with the aid of symmetry, from this we obtain

$$\int dt\,e^{-i\omega t}C(X_{-\boldsymbol{k}};Y_{\boldsymbol{k}}(t)) = \frac{1}{2}(1+e^{\beta\hbar\omega})\int dt\,e^{-i\omega t}\langle X_{-\boldsymbol{k}}Y_{\boldsymbol{k}}(t)\rangle_e. \tag{X2.5.45}$$

On the other hand, we have

$$\int dt\,e^{-i\omega t}Tr(\rho_e[X_{-\boldsymbol{k}},Y_{\boldsymbol{k}}(t)]) = (1-e^{\beta\hbar\omega})\int dt\,e^{-i\omega t}\langle X_{-\boldsymbol{k}}Y_{\boldsymbol{k}}(t)\rangle_e. \tag{X2.5.46}$$

Now, we use the two expressions of the response function (→(X2.5.30), (X2.5.33))

$$\frac{i}{\hbar}Tr(\rho_e[X_{-\boldsymbol{k}},Y_{\boldsymbol{k}}(t)]) = -\beta\langle X_{-\boldsymbol{k}};\dot{Y}_{\boldsymbol{k}}(t)\rangle. \tag{X2.5.47}$$

Therefore, (X2.5.46) reads

$$\int dt\,e^{-i\omega t}\langle X_{-\boldsymbol{k}};\dot{Y}_{\boldsymbol{k}}(t)\rangle = \frac{i}{\beta\hbar}(1-e^{\beta\hbar\omega})\int dt\,e^{-i\omega t}\langle X_{-\boldsymbol{k}}Y_{\boldsymbol{k}}(t)\rangle_e. \tag{X2.5.48}$$

Combining this and (X2.5.45), we obtain the desired relation (X2.5.40).

---

[127] H. B. Callen and T. A. Welton, Phys. Rev. **83**, 34 (1951).

X2.6 NMR Spectrum

Here is a highly simplified exposition of the *Kubo-Tomita theory*[128] of magnetic resonance spectrum.

This is the first systematic and general statistical mechanical study of phenomena involving dissipation. However, spins do not flow under the external perturbation, so there is a considerable difference between the spectrum problem and the electrical conductivity discussed in the next section.

X2.6.1 Magnetic resonance spectrum

Magnetic resonance spectrum observes the energy absorption by a sample from the magnetic field. Therefore, the driving force is $\boldsymbol{H}$, and the conjugate displacement is the system magnetization $\boldsymbol{M}$. The response function describing the response of the magnetization due to the perturbative magnetic field is $\phi_{\boldsymbol{MM}}$ (→**X2.5.3**):

$$\phi_{\boldsymbol{MM}}(t) = \beta\langle\dot{\boldsymbol{M}};\boldsymbol{M}(t)\rangle. \tag{X2.6.1}$$

The absorption spectrum is determined by the loss part of the admittance, the Fourier transform of the response function. We know the power absorption is (→**X2.2.8**)

$$p = \frac{1}{2}Re\left(i\omega\chi_{\boldsymbol{MM}}(\omega)\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^{\dagger}\right). \tag{X2.6.2}$$

X2.6.2 Circularly and linearly polarized magnetic fields

The phasor representation of the magnetic field is $\hat{\boldsymbol{H}} = (H, -iH, 0)^T$. Therefore, (→**X2.2.8**)

$$p = \frac{1}{2}(\chi'_{xy} - \chi'_{yx} + \chi''_{xx} + \chi''_{yy})\omega H^2. \tag{X2.6.3}$$

Here, $'$ implies the real part, and $''$ implies the imaginary part.

For linearly polarized magnetic field,

$$p = \frac{1}{2}\omega\chi''_{xx}H^2. \tag{X2.6.4}$$

X2.6.3 Admittance in terms of canonical correlation

Due to causality (→**X2.2.9**), the Fourier transformation is given by

$$\chi_{\boldsymbol{MM}}(\omega) = \int_0^\infty dt\, e^{-i\omega t}\phi_{\boldsymbol{MM}}(t). \tag{X2.6.5}$$

In particular, $\langle M_x; M_x(t)\rangle$ is real from [E] of **X2.5.7**, so we obtain

$$\chi''_{xx}(\omega) = \omega\int_0^\infty dt\,\cos\omega t\,\beta\langle M_x; M_x(t)\rangle. \tag{X2.6.6}$$

---

[128] R. Kubo and K. Tomita, “A general theory of magnetic resonance absorption, ”J. Phys. Soc. Jpn. **9**, 888-919 (1954).

X2.6.4 Calculation of canonical correlation function: strategy

To compute the canonical correlation function, we need $M(t)$. To compute this we need its equation of motion. To this end we write down the Hamiltonian H:

$$H = H_1 + H_2 + H' = H_0 + H', \tag{X2.6.7}$$

where H_1 is the Hamiltonian of the spin system (the system that absorbs the energy from the magnetic field), H_2 is the so-called lattice system (the Hamiltonian governing the degrees of freedom of the system other than spins), and H' is the interaction between these two subsystems. We assume

$$[H_1, H_2] = 0, \;\; [H_2, M] = 0. \tag{X2.6.8}$$

Therefore, the equation of motion for the total magnetization in the Heisenberg picture reads

$$i\hbar \frac{dM}{dt} = [M, H_1 + H']. \tag{X2.6.9}$$

Here, deliberately M lacks any suffix. Any component or linear combination of them obeys this equation.

We solve this equation (at least formally), and then compute the canonical correlation function. Since we do not honestly study the motion of the whole system, the perturbative Hamiltonian H' must be appropriately modeled. Usually, we assume that this interaction fluctuates stochastically.

X2.6.5 Approximate expression of the canonical correlation

Let the time evolution operator for the system be T:

$$i\hbar \frac{dT}{dt} = (H_0 + H')T. \tag{X2.6.10}$$

Let T_0 be the time evolution operator of the unperturbed (non-interacting) system:

$$T_0 = e^{-iH_0 t/\hbar}. \tag{X2.6.11}$$

and define τ as $T = T_0\tau$. Then,

$$T_0 i\hbar \frac{d\tau}{dT}_0 = H'T_0\tau, \tag{X2.6.12}$$

Introducing the Dirac representation of H':

$$\tilde{H}' = T_0^{-1} H' T_0 = e^{iH_0t/\hbar} H' e^{-iH_0t/\hbar}, \tag{X2.6.13}$$

we obtain the equation of motion for τ:

$$i\hbar \frac{d}{dt}\tau = \tilde{H}'(t)\tau. \tag{X2.6.14}$$

Here, the time dependence of the interaction Hamiltonian is explicitly denoted. Its formal solution reads

$$\tau = \mathrm{T}\exp\left[\frac{i}{\hbar}\int_0^t \tilde{H}'(t)dt\right], \tag{X2.6.15}$$

where T is the chronological operator that rearranges everything on its right hand side in the increasing order of time t.[129] The Heisenberg representation of M can be written formally as[130]

$$M(t) = \tau^{-1}\tilde{M}\tau = \mathrm{T}\exp\left[\frac{i}{\hbar}\int_0^t \tilde{H'}^{\times}(t)dt\right]\tilde{M}(0) = \mathrm{T}\exp\left[\frac{i}{\hbar}\int_0^t \tilde{H'}^{\times}(t)dt\right]M(0). \quad \text{(X2.6.16)}$$

Therefore,

$$\begin{aligned}\Phi(t) &\equiv \langle M(0); M(t)\rangle \\ &= \beta^{-1}\int_0^\beta d\lambda\, Tr\left\{\rho_0 e^{\lambda H_0/\hbar} M e^{-\lambda H_0/\hbar} e^{-iH_0 t/\hbar}\mathrm{T}\exp\left[\frac{i}{\hbar}\int_0^t \tilde{H'}^{\times}(t)dt\right] M e^{iH_0 t/\hbar}\right\} \quad \text{(X2.6.17)} \\ &= \beta^{-1}\int_0^\beta d\lambda\, Tr\left\{\rho_0 e^{i(-i\lambda+t)H_0/\hbar} M e^{-i(-i\lambda+t)H_0/\hbar}\mathrm{T}\exp\left[\frac{i}{\hbar}\int_0^t \tilde{H'}^{\times}(t)dt\right] M\right\}. \quad \text{(X2.6.18)}\end{aligned}$$

Here, $\rho_0 \propto e^{-\beta H_0}$. Kubo and Tomita computed this with the high temperature approximation:

$$\Phi(t) = Tr\rho_0 M \exp\left[\frac{i}{\hbar}\int_0^t H'^{\times}(t)dt\right] M. \quad \text{(X2.6.19)}$$

X2.6.6 Stochastic approximation of the canonical correlation

Now we assume that H' is fluctuating stochastically, and apply the second cumulant approximation to Φ:[131]

$$\Phi(t) = \exp\left[-\frac{1}{\hbar^2}\int_0^t dt_1\int_0^{t_1} dt_2\, Tr\left\{\rho_0 M\tilde{H}'^{\times}(t_1)\tilde{H}'^{\times}(t_2)M\right\}\right]. \quad \text{(X2.6.20)}$$

To go further we write the interaction Hamiltonian as[132]

$$\tilde{H}'(t) = \sum_\alpha e^{i\omega_\alpha t}H'_\alpha(t). \quad \text{(X2.6.21)}$$

Notice that $H'^*_\alpha(t) = H'_{-\alpha}(t)$ due to the Hermitian property of the Hamiltonian.

We must compute

$$\begin{aligned} Tr\left(\rho_0 M {H'}^{\times}_{\alpha}(t_1){H'}^{\times}_{\beta}(t_2)M\right) &= Tr\left(M{H'}^{\times}_{\alpha}(t_1){H'}^{\times}_{\beta}(t_2)M\rho_0\right), \quad \text{(X2.6.22)} \\ &= Tr\left([M, H'_\alpha(t_1)]\,[H'_\beta(t_2), M\rho_0]\right), \quad \text{(X2.6.23)} \\ &= Tr\left([M, H'_\alpha(t_1)]\,([H'_\beta(t_2), M]\rho_0] + M[H'_\beta, \rho_0])\right). \quad \text{(X2.6.24)}\end{aligned}$$

---

[129] This is the usual starting point of the Dyson expansion.

[130] Thanks to T, H' and τ commutes, so it is easy to see that for any *fixed s*

$$i\hbar\tau^{-1}\tilde{A}(s)\tau = [\tau^{-1}\tilde{A}(s)\tau, H'],$$

so

$$\tau^{-1}\tilde{A}(s)\tau = \mathrm{T}\exp\left[\frac{i}{\hbar}\int_0^t \tilde{H'}^{\times}(t)dt\right]\tilde{A}(s).$$

Then, set $s = t$.

[131] The average is assumed to be zero; if not, it causes shifts in spectral lines through the first order perturbation effect, but the line shape is not affected. Therefore, we can safely subtract the average, if it is not zero.

[132] That is, into the eigenstates of H_1 the interaction representation is decomposed.

Now, we ignore the last factor:

$$[H'_\beta, \rho_0] \sim \hbar\omega_\beta H'_\beta \rho_0, \tag{X2.6.25}$$

assuming that there is no contribution from the high frequency components. Thus, we have

$$Tr\left(\rho_0 M H'^{\times}_{\alpha}(t_1) H'^{\times}_{\beta}(t_2) M\right) \simeq Tr\left(\rho_0 [M, H'_\alpha(t_1)][H'_\beta(t_2), M]\right). \tag{X2.6.26}$$

Defining

$$f_{\alpha,\beta}(t) = \frac{Tr\left([M, H'_\alpha(t)]\,[H'_\beta, M]\rho_0\right)}{Tr\left([M, H'_\alpha]\,[H'_\beta, M]\rho_0\right)}, \tag{X2.6.27}$$

$$\Delta^2_{\alpha\beta} = \frac{Tr\left([M, H'_\alpha]\,[H'_\beta, M]\rho_0\right)}{\hbar^2 \langle M^2\rangle}, \tag{X2.6.28}$$

we have

$$Tr\left(\rho_0 M H'^{\times}_{\alpha}(t_1) H'^{\times}_{\beta}(t_2) M\right) \simeq \Delta^2_{\alpha\beta} f_{\alpha\beta}(t_1 - t_2)\hbar^2\langle M^2\rangle. \tag{X2.6.29}$$

Thus,

$$\Phi(t) = \langle M^2\rangle \exp\left[-\int_0^t dt_1 \int_0^{t_1} dt_2 \sum_{\alpha,\beta} e^{it\omega_\alpha} e^{it\omega_\beta} \Delta^2_{\alpha\beta} f_{\alpha\beta}(t_1 - t_2)\right]. \tag{X2.6.30}$$

If $f_{\alpha\beta}$ decays quickly, then $\omega_\alpha + \omega_\beta \simeq 0$ would survive, so

$$\Phi(t) = \langle M^2\rangle \exp\left[-\int_0^t d\tau (t-\tau) \sum_\alpha e^{i\tau\omega_\alpha} \Delta^2_\alpha f_\alpha(\tau)\right], \tag{X2.6.31}$$

where $f_\alpha \equiv f_{\alpha-\alpha}$ and $\Delta_\alpha \equiv \Delta_{\alpha-\alpha}$. Here, the term with $\omega_\alpha = 0$ describes the adiabatic effect (*secular effect*), and the rest nonadiabatic.

X2.6.7 Spectral line shape

The spectral line shape is given by the Fourier transform of (normalized) Φ:

$$I(\omega) \propto \frac{1}{2\pi} \int dt\, e^{-i\omega t} \Phi(t). \tag{X2.6.32}$$

To compute this we need f_α in **X2.6.6**, which is a time correlation function of M driven by the spin-lattice interaction; due to this interaction dephasing of the spin motion occurs. Therefore, we wish to study the effect of this dephasing on the spectrum.

Although we need a further detailed modeling of the system to compute f_α, the generic behavior may be captured by

$$f_\alpha(t) = e^{-|t|/\tau_c}, \tag{X2.6.33}$$

which is a very familiar form since the simple model of Brownian motion (→**3.2.7**). This gives (normalized)

$$I(\omega) = \frac{1}{2\pi} \int dt\, e^{-i\omega t} \exp\left[-\sum_\alpha \Delta^2_\alpha (\tau_c + i\tau_c^2\omega_\alpha)t/(1 + \omega_a^2\tau_c^2)\right]. \tag{X2.6.34}$$

Therefore, the line shape is Lorentzian

$$I(\omega) = \frac{1}{\pi} \frac{T_2}{1 + (\omega - \delta)^2 T_2^2} \tag{X2.6.35}$$

with

$$\frac{1}{T_2} = \sum_\alpha \Delta_\alpha^2 \frac{\tau_c}{1 + \omega_\alpha^2 \tau_c^2}, \quad \delta = -\sum_\alpha \Delta_\alpha^2 \frac{\tau_c^2 \omega_\alpha}{1 + \omega_\alpha^2 \tau_c^2}. \tag{X2.6.36}$$

Here, T_2 is called the *transverse relaxation time* representing the time scale of the spin dephasing.[133] Notice that the dephasing causes a Lorentzian line shape (cf. **3.7.6**).

X2.6.8 Motional narrowing

If the dephasing of spins occurs rapidly (i.e., if τ_c is small $\Rightarrow T_2$ large), the line shape becomes narrow as can be seen from (X2.6.35). This is called the *motional narrowing*, because often the dephasing is due to the random motion of nuclei on which spins ride.

The line shape is a very important topic, but truly microscopic studies are often prohibitively hard, so there are many model studies based of Markov chains and processes (→**6.5.10**∼) that are very reasonable and highly useful.

X2.6.9 Infrared spectra and corresponding correlation functions[134]

NMR and EPR spectra study the response of magnetic moments to the perturbing magnetic field, so the perturbation Hamiltonian reads, as we have seen,

$$h = -\boldsymbol{M} \cdot \boldsymbol{H}(t). \tag{X2.6.37}$$

As we know from this we can read off what correlation function we observe: $\langle M; M(t) \rangle$.

A direct analogue is the infrared absorption (or generally VL and UV photon absorption) spectrum in which the electrical dipole moment $\boldsymbol{P}$ is studied. Therefore, the perturbative Hamiltonian is

$$h = -\boldsymbol{P} \cdot \boldsymbol{E}, \tag{X2.6.38}$$

so that we observe the correlation $\langle \boldsymbol{P}; \boldsymbol{P}(t) \rangle$. Its analogy with NMR is straightforward.

X2.7 Electrical Conductivity

Although the Kubo-Tomita theory outlined in **X2.6** was the first systematic study of systems with dissipation, the spin system does not dissipate energy when the magnetic field is time independent.[135] Therefore, this theory was not regarded as a real breakthrough in nonequilibrium statistical mechanics.

---

[133] The *longitudinal relaxation time* T_1 is the relaxation time of the energy of the spin system that is lost to the lattice system.

[134] B. J. Berne, "Time-dependent properties of condensed media," Chapter 9 of *Physical Chemistry, an advanced treatise* Vol VIIIB (Academic Press, New York, 1971) is an excellent intermediate level exposition.

[135] Since there is no monopole, it is absolutely impossible to have a DC current.

Then, stimulated by the formalism of Kubo and Tomita, Nakano gave the general formula for conductivity in 1955.[136] This was the real breakthrough in nonequilibrium statistical mechanics, and many papers followed this including Kubo's supposedly purely mechanical general framework of linear response theory. However, as we have clearly seen up to this point, and as Nakano stresses very explicitly in his review quoted at the beginning of this Chapter, the linear response theory is an eclectic theory.

X2.7.1 Nakano's approach[137]

If the reader wishes to apply a static electric field to the system, the perturbative Hamiltonian must contain infinitely large electric potential. Therefore, Nakano avoided this, and used the vector potential $\boldsymbol{A}$. Thus, the perturbation Hamiltonian reads

$$h = -\hat{\boldsymbol{j}} \cdot \boldsymbol{A}(t), \tag{X2.7.1}$$

where $\hat{\boldsymbol{j}}$ is the (microscopic expression of) electrical current. Then, at least formally, this becomes very similar to the spin system we have discussed in **X2.6**. Beyond this point he followed roughly the calculation outlined in **X2.6** and obtained the perturbative solution to the von Neumann equation:

$$\rho(t) = \rho_e \left[1 - \int_{-\infty}^{t} ds \int_0^\beta d\lambda e^{\lambda H} \dot{\hat{\boldsymbol{j}}}(s-t) \cdot \boldsymbol{A}(s) e^{-\lambda H} + \cdots \right]. \tag{X2.7.2}$$

Integrating this by parts, we obtain

$$\rho(t) = \rho_e \left[1 - \boldsymbol{A}(t) \int_0^\beta d\lambda e^{\lambda H} \hat{\boldsymbol{j}}(0) e^{-\lambda H} + \int_{-\infty}^{t} ds \int_0^\beta d\lambda e^{\lambda H} \hat{\boldsymbol{j}}(s-t) \cdot \frac{\partial}{\partial s} \boldsymbol{A}(s) e^{-\lambda H} + \cdots \right]. \tag{X2.7.3}$$

Now, we assume

$$\boldsymbol{A}(t) = \boldsymbol{E}(t_0) \Theta(t_0 - t) e^{\epsilon t}. \tag{X2.7.4}$$

That is, we assume a pulse-like electric field.[138] Putting this into (X2.7.3), we obtain (→**X2.5.1**)

$$\rho(t, t_0) = \rho_e \left[1 - \int_0^\beta d\lambda e^{\lambda H} \hat{\boldsymbol{j}}(t_0 - t) \cdot \boldsymbol{E}(t_0) e^{-\lambda H} + \cdots \right]. \tag{X2.7.5}$$

The current can be computed as

$$\boldsymbol{j}(t) = Tr \rho(t, t_0) \hat{\boldsymbol{j}} = \int_0^\beta d\lambda \left\langle e^{\lambda H} \hat{\boldsymbol{j}}(0) \cdot \boldsymbol{E}(t_0) e^{-\lambda H} \hat{\boldsymbol{j}}(t - t_0) \right\rangle_e. \tag{X2.7.6}$$

Thus, the 'Green's function' (the pulse response) of the current is given by

$$j_\mu(t) = \beta \sum_\nu \langle j_\nu ; j_\mu(t - t_0) \rangle E_\nu(t_0). \tag{X2.7.7}$$

The corresponding admittance is the (frequency-dependent) conductivity.

---

[136] The first-hand record of this development may be found in: H. Nakano, "Linear response theory: a historical perspective," Int. J. Mod. Phys. B **7**, 2397-2467 (1993). This also contains the way to compute the conductivity using the Boltzmann equation, and a detailed calculation of some of the conductivity formulas (with, e.g., impurity, phonon scattering) from the pedagogical point as well.

[137] If the reader is interested only in the modern approach, begin at **X2.7.2**.

[138] Notice that $\boldsymbol{E} = -\partial \boldsymbol{A}/\partial t$.

X2.7.2 Electrical conductivity from the standard formula

We have already developed a general theoretical framework. The electrical current is the flux of charge, and the conjugate variable is the electrical potential V. Since

$$\boldsymbol{j} = -\sigma\nabla V, \tag{X2.7.8}$$

the general theory tells us that the conductivity tensor reads

$$\sigma = \int_0^\infty \beta\langle \boldsymbol{j}; \boldsymbol{j}(t)\rangle dt. \tag{X2.7.9}$$

This is the most famous Green-Kubo relation called somehow the Kubo formula, but it is derived in the way explained in **X2.7.1** by Nakano, so should be called the *Nakano formula of electrical conductivity.*

X2.7.3 High temperature approximation for conductivity

Under high temperature, we may identify the canonical and ordinary correlation functions, so (X2.7.9) can be symmetrized as

$$\sigma = \frac{1}{k_BT}\int_0^\infty \Gamma(t)dt, \tag{X2.7.10}$$

$$\Gamma(t) = \frac{1}{2}\langle j_\mu(t)j_\mu + j_\mu(-t)j_\mu\rangle. \tag{X2.7.11}$$

To compute this we need $j_\mu(t)$. Let H' be the interaction Hamiltonian, e.g., between the electrons and phonons, and we can almost follow the calculation we have done in the NMR spectrum shape, and we get

$$\Gamma(t) = \langle j_\mu^2\rangle \exp\left[-\int_0^t (t-\tau)F(\tau)d\tau\right], \tag{X2.7.12}$$

where

$$F(t) = \frac{1}{\hbar^2\langle j_\mu^2\rangle}\sum_\omega \langle [H'_\omega, j_\mu]\,[j_\mu, H'_{-\omega}]\rangle \cos\omega t. \tag{X2.7.13}$$

This is essentially the correlation function of microscopic dynamics, so we may assume that it decays very quickly. Therefore, we may approximate this further as

$$\Gamma(t) = \langle j_\mu^2\rangle e^{-\gamma t} \tag{X2.7.14}$$

with

$$\gamma = \int_0^\infty F(\tau)d\tau = \frac{\pi}{\hbar^2\langle j_\mu^2\rangle}\sum_\omega \langle [H'_\omega, j_\mu]\,[j_\mu, H'_{-\omega}]\rangle\delta(\omega). \tag{X2.7.15}$$

Thus, finally we have obtained the following formula

$$\sigma = \frac{\langle j_\mu^2\rangle}{\gamma k_BT} = \frac{\hbar^2\langle j_\mu^2\rangle^2}{\pi k_BT\sum_\omega \langle [H'_\omega, j_\mu]\,[j_\mu, H'_{-\omega}]\rangle\delta(\omega)}. \tag{X2.7.16}$$

X2.7.4 Resistance due to electron-phonon interactions

Let us take the electron-phonon interaction Hamiltonian to be

$$H' = \sum_{\boldsymbol{k},\boldsymbol{q}} C_{\boldsymbol{q}} a^+_{\boldsymbol{k}+\boldsymbol{q}} a_{\boldsymbol{k}} (b_{\boldsymbol{q}} + b^+_{-\boldsymbol{q}}), \tag{X2.7.17}$$

where a^+, a are creation and annihilation operators for conduction electrons, and b^+, b are creation and annihilation operators for phonons. $C_{\boldsymbol{q}}$ is the coupling constant:

$$C_{\boldsymbol{q}} = \alpha q^{1/2}, \ \ \alpha = \left(\frac{2\hbar}{9NMs}\right)^{1/2} C \tag{X2.7.18}$$

with N being the number of lattice ions of mass M per unit volume, C the phonon heat capacity, and s the sound velocity. The current operator reads

$$\hat{\boldsymbol{j}} = \sum_{\boldsymbol{k}} \boldsymbol{j}(\boldsymbol{k}) a^+_{\boldsymbol{k}} a_{\boldsymbol{k}}. \tag{X2.7.19}$$

From this

$$\langle \boldsymbol{j}^2_\mu \rangle = -\left(\frac{e\hbar}{m}\right) k_B T \sum_{\boldsymbol{k}} k^2_\mu \frac{\partial f_0(\epsilon_{\boldsymbol{k}})}{\partial \epsilon_{\boldsymbol{k}}} = \frac{n k_B T e^2}{m}, \tag{X2.7.20}$$

where n is the conduction electron number density, and m is the electron (effective) mass. Here, the summation over $\boldsymbol{k}$ can be computed as

$$\sum_{\boldsymbol{k}} \to \frac{1}{4\pi^3} \int_0^\infty k^2 dk \int_{-1}^{+1} d\cos\theta \int_0^{2\pi} d\phi. \tag{X2.7.21}$$

Next, we compute the commutators[139]

$$[H'_\omega, j_\mu] = \frac{\hbar e \alpha}{m} \sum_{\boldsymbol{k},svq} \sqrt{q} q_\mu (a^+_{\boldsymbol{k}+\boldsymbol{q}} a_{\boldsymbol{k}} b_{\boldsymbol{q}} - a^+_{\boldsymbol{k}} a_{\boldsymbol{k}+\boldsymbol{q}} b^+_{\boldsymbol{q}}), \tag{X2.7.22}$$

Using this, we obtain

$$\begin{aligned} & \sum_\omega \langle [H'_\omega, j_\mu]\,[j_\mu, H'_{-\omega}] \rangle \delta(\omega) \\ = & \left(\frac{\hbar e \alpha}{m}\right)^2 \sum_{\boldsymbol{k},\boldsymbol{q}} q q^2_\mu \left\{ f'_0(1-f_0)(n_0+1) + f_0(1-f'_0) n_0 \right\} \hbar \delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \omega_{\boldsymbol{q}}), \end{aligned} \tag{X2.7.23}$$

$$= 2\left(\frac{\hbar e \alpha}{m}\right)^2 \sum_{\boldsymbol{k},\boldsymbol{q}} q q^2_\mu (f_0 - f'_0) n_0 (n_0+1) \hbar \delta(\epsilon_{\boldsymbol{k}+\boldsymbol{q}} - \epsilon_{\boldsymbol{k}} - \omega_{\boldsymbol{q}}). \tag{X2.7.24}$$

Here, $f_0 = f_0(\epsilon_{\boldsymbol{k}})$ and $f'_0 = f_0(\epsilon_{\boldsymbol{k}+\boldsymbol{q}})$. Next, this sum is computed as integrals as in (X2.7.21). The result reads[140]

$$\sum_\omega \langle [H'_\omega, j_\mu]\,[j_\mu, H'_{-\omega}] \rangle \delta(\omega) = \frac{4\pi C^2 N e^2 T^6}{3M\Theta^6} J_5(T/\Theta), \tag{X2.7.25}$$

---

[139] In the original paper $\sqrt{q} q_\mu$ reads $\sqrt{q} k_\mu$. This is a typo.

[140] Slightly more details can be seen in Nakano's review article already quoted.

where

$$J_5(x) = \int_0^x dy\, \frac{y^5}{(e^y - 1)(1 - e^{-y})} \tag{X2.7.26}$$

is one of the Debye functions, and Θ is the Debye temperature.

Thus we arrive at *Grüneisen's formula*, which was derived with the aid of the Boltzmann equation:

$$\sigma = \frac{n\hbar e^2 k_F^3 M k_B \Theta}{4\pi^3 N C^2 m^2} \left(\frac{\Theta}{T}\right)^5 \frac{1}{J_5(\Theta/T)}. \tag{X2.7.27}$$

Chapter 7

Dynamical Systems Theoretical Approaches

Note:
This chapter should contain many figures, but they have not been included here yet. In the lectures many concepts will be maximally visualized.

The theory of dynamical systems is another theoretical subject every physicists should know its rudiments. Introductory textbooks are

D. Ruelle: *Elements of Differentiable Dynamics and Bifurcation Theory* (Academic Press, 1989),[1]
A. Katok and B. Hasselblatt *Introduction to the Modern Theory of Dynamical Systems* (Cambridge UP, 1995).[2]
This will be quoted as Katok-Hasselblatt.

An authoritative reference series is
Dynamical Systems I- (Encyclopaedia of Mathematical Sciences, Springer)

There is no readable, thin, and serious-error-free introductory textbooks.[3]

---

[1] An Amazon review says, "An undergrad nightmare as a text, a graduate student's dread as a course text but a postgraduate student's delight..." Therefore, this is a good book for serious students.

[2] This is an 800 page textbook containing most important topics.

[3] E. A. Jackson, "Perspectives of Nonlinear Dynamics," two volumes (Cambridge UP 1989) is a readable and enjoyable introductory survey, but very thick (but much more humane than this chapter and with excellent illustrations). It contains many topics such as solitons, but the level is far short of that we need here.

7.1 Dynamical Systems, Introduction

7.1.1 What is dynamical system?

Let T be a set on which addition $+$ is defined. This is the set of time. If there is a one parameter family of maps ϕ_t $(t \in T)$ from a set M (*phase space*) into itself (endomorphism), satisfying

$$\begin{aligned} \phi_0 &= 1 \text{ (identity on } M \text{)}, &(7.1.1)\\ \phi_{s+t} &= \phi_t\phi_s \ (\forall t, s \in T), &(7.1.2)\end{aligned}$$

that is, $\{\phi_t\}_{t\in T}$ is a group or semigroup acting on M, then (M, ϕ_t) is called a *dynamical system.* If the time set T is continuous (discrete), then (M, ϕ_t) is called a continuous (discrete) *dynamical system.*[4]

Let $x \in M$. Then, $\phi_t(x)$ $(t \in T)$ defines an *orbit* in M that passes x at $t = 0$ (that is, x is an initial condition).

The *time one map* ϕ_1 represents the property of the dynamical system (M, ϕ_t) very well. For example, if ϕ_1 is 'chaotic,' then the original dynamical system is 'chaotic.' Therefore, the study of discrete dynamical systems is essentially the study of dynamical systems.

7.1.2 Vector field

If (M, ϕ_t) is a continuous time dynamical system, and if it is differentiable with respect to time, then

$$\frac{d}{dt}\phi_t(x) = X(x) \tag{7.1.3}$$

is called the *vector field* (on M). If a vector field on a manifold M is given, then constructing ϕ_t is a problem of ordinary differential equation. The theory of dynamical systems originated from the qualitative theory of ordinary differential equations, so sound knowledge of ordinary differential equations is advantageous.[5]

For discrete dynamical systems, the time one map $\phi = \phi_1$ generates all ϕ_t as $\phi_t = \phi^t$.

7.1.3 Limit sets

The ultimate destination of the orbit stating from x is called the *ω-limit set* for x. To make this concept precise, we need the concept of *ω-limit point*: y is an ω-limit point for x, if there is an infinite sequence $\{\phi_{t_i}(x)\}$ $(t_1 < t_2 < \cdots \to \infty)$ converging to y. The totality of the ω-limit points for x is called the ω-limit set for x, and is denoted by $\omega(x)$. The *ω-limit set* Ω of a dynamical system (M, ϕ_t) is $\Omega = \cup_{x\in M}\omega(x)$.

Similarly, we can introduce the *α-limit set* $\alpha(x)$ for x. $\alpha(x)$ is the totality of *α-limit points*; y is an α-limit point for x, if there is an infinite sequence $\{\phi_{t_i}(x)\}$ $(t_1 > t_2 > \cdots \to -\infty)$ converging to y.

If the trajectory starting from a point x eventually goes away from this point, x is called a *wandering point.* That is, x is a wandering point, if there is a neighborhood U of x such that for some $\tau > 0$

$$(\cup_{t>\tau}\phi_t(U)) \cap U = \emptyset. \tag{7.1.4}$$

---

[4] A C^r-class invertible dynamical system is a C^r-class Lie group isomorphic to $\boldsymbol{R}$ or $\boldsymbol{Z}$ acting on M.

[5] M. Hirsch and S. Smale, *Differential equations, dynamical systems, and linear algebra* (Academic Press, 1974) is an excellent introduction.

if x is not a wandering point, it is called a *non-wandering point.* That is, x is a non-wandering point, if for any neighborhood U of x and for any $\tau > 0$ there is $t > \tau$ such that

$$\phi_t(U) \cap U \neq \emptyset. \tag{7.1.5}$$

A set consisting of nonwandering points is called a *nonwandering set.*

$N \subset M$ is called an *invariant set*, if

$$\phi_t(N) = N \tag{7.1.6}$$

for all $t \in T$.
(1) ω-limit and α-limit sets are nonwandering closed sets.
(2) The totality of nonwandering points is an invariant set.
(3) Let A be a non-empty compact invariant set. Then, A contains an ω-limit set.

7.1.4 Attracting set and attractor

A closed invariant set A is called an *attracting set* of a dynamical system (M, ϕ_t), if within its any neighborhood is a set B (called a trapping region) such that

$$\cap_{t>0}\phi_t(B) = A. \tag{7.1.7}$$

Notice that this set need not be connected.

The concept "*attractor*" is more restricted. There are different definitions, but in essence we wish to call an attracting set an attractor if the restriction of the original dynamics on it defines a nice dynamics that can capture typical long time asymptotic behaviors of the original system. Therefore, inseparability as a set and lack of any genuinely smaller attracting set in it are often required. Thus, one possibility is to say that an invariant set is an *attractor* of (M, ϕ_t) , if it is an attracting set containing a dense orbit.

A sink and a stable limit cycle are typical attractors. There are other types of attractors, and they are vaguely called *strange attractors.*

7.1.5 Fixed point, periodic point

If $\phi_t(x) = x$ for some $t \in T$, x is called a *periodic point.* In this case, if $\phi_s(x) \neq x$ for $0 < s < t$, then t is called the *period* of x. If $\phi_t(x) = x$ for all $t \in T$, x is called a *fixed point.* A fixed point is called a *hyperbolic fixed point*, if the linearization of the map ϕ_t at x is hyperbolic for any t, that is, the eigenvalues of $d\phi_t(x)/dx$ are not on the unit circle.

The hyperbolicity of dynamics near a fixed point x implies that the dynamical system is essentially described by the linearized approximation near x. This is guaranteed by the following theorem:

Theorem [Hartman].[6] Near a hyperbolic fixed point, the structure of ϕ_1[7] is determined by its linear approximation. More precisely, let $L = D_x\phi_1$ be the linear approximation of ϕ_1 around the fixed point x. Then, there is a homeomorphism[8] h such that $h(x) = x$ and

$$\phi_1 = hLh^{-1}. \tag{7.1.8}$$

□

Between two maps ϕ and ψ, if we have a homeomorphisms h such that $h\phi = \psi h$, we

---

[6] For a proof, see, e.g., Katok-Hasselbratt p260. There, it is called the Hartman-Grobman theorem.

[7] It should be clear that for any $t \in T$ ϕ_t is also determined.

[8] One to one map that is continuous in both ways.

say ϕ and ψ are *conjugate*. Thus, if x is a hyperbolic fixed point of ϕ, then the dynamical systems defined by ϕ and $D_x\phi$ are conjugate in an appropriate neighborhood of x.

Remark 1. This is the theorem justifying the linear stability analysis around a fixed point of an ordinary differential equation. This implies that the reliability of linear stability analysis of a fixed point of an ordinary differential equation is guaranteed only when the fixed point is hyperbolic.□

Remark 2. h is not unique, because the theorem is only local.□

Remark 3. h may not be differentiable (that is, it cannot generally be a diffeomorphism[9]). □

7.1.6 Poincaré map

Let x be a periodic point of a continuous dynamical system (M, ϕ_t). Furthermore, let us assume that $d\phi_t(x)/dt$ (the velocity along the orbit going through x) is not zero. Then, we can imagine a hyperplane[10] perpendicular to this orbit passing through x. All the nearby orbits (the orbits going through the points in a sufficiently small neighborhood of x) must also be nearly orthogonal to this plane.[11] If the neighborhood is taken small enough, after almost one period all the trajectories return to this hyperplane. Therefore, we can define a map in a neighborhood of x from a transversal hyperplane into itself. This is called the *Poincaré map* or the *first-return map*. The periodic point x becomes a fixed point of this map.

If x is a hyperbolic fixed point of the Poincaré map, the periodic orbit is called a *hyperbolic periodic orbit*, and x is called a *hyperbolic periodic point*.

In a certain sense, for a generic differentiable dynamical system all the periodic orbits are hyperbolic.[12]

7.1.7 How can we tell two dynamical systems are similar or not?

There are many occasions where we wish to judge how close two dynamical systems (M, ϕ) and (M, ψ) are. They may be the original system and its perturbed counterpart. In this case, we wish to know how stable the original dynamical system is against *structural perturbations*.

To answer this question, we must introduce a metric in the set of all the dynamical systems defined on M. Or, if we wish to be more specific, we must study the totality $C^r(M, M)$ of C^r endomorphisms of M with its natural metric given by the *Whitney norm*.

We assume M is a compact manifold.[13] On each chart, we can introduce the usual

---

[9] A homeomorphism that is differentiable both ways is called a *diffeomorphism*.

[10] = codimension one manifold

[11] more precisely, *transversal*, that is not tangential.

[12] **Theorem [Kupka-Smale]**. For diffeomorphisms from M into itself, it is generic that all the periodic orbits are hyperbolic. □

(I. Kupka, Contrib. Diff. Eqs. **2**, 457-84 (1963); **3**, 411-420 (1964); S. Smale, Ann. Scuola Normale Superiore Pisa **17**, 97-116 (1963); M. M. Peixoto, J. Diff. Eq., **3**, 214-227 (1966).)

Here, *generic* is a Baire categorical terminology. A property holds *generically* for (or the property is a *generic* property of) a set A, if the subset on which the property does not hold can be written as at most countably many join of totally sparse sets; a set is totally sparse, if its closure does not have any internal point. Genericity is a much stronger concept than being dense, but much weaker than being open dense.

[13] For such differential topological concepts, I. M. Singer and J. A. Thorpe, "Lecture notes on elementary topology and geometry," (Scott, Foresman and Co., 1967) is strongly recommended. M. Spivak, "A comprehensive introduction to differential geometry," Vol. 1 (Publish or Perish 1979) is also readable.

C^s-norm.[14] Then, take the supremum of C^s-norms for all the charts in the atlas for M. This supremum value is called the Whitney norm W^s.

The metric is, as usual, defined by the norm. If M is a compact manifold, then $C^r(M, M)$ becomes a complete metric space with the Whitney metric.

We say two C^r-dynamical systems (M, ϕ_t) and (M, ψ_t) are close if $\|\phi_t - \psi_t\|_r$ is small for all t.

7.1.8 Dynamical systems can be approximated by smoother counterparts

Suppose M is a compact manifold. A C^s-dynamical system may be approximated as much as one wishes with a smoother dynamical system (i.e., with a C^r-dynamical system with $r > s$).

More precisely, $C^r(M, M)$ is dense in $C^s(M, M)$ with the Whitney norm $\|\cdot\|_s$ for $s < r$.

7.2 Stable and Unstable Manifolds

As was already noted in **7.1.1**, a dynamical system (M, ϕ_t) may be studied through a map $\phi = \phi_1$ defined on M irrespective of the nature of the time set T. Therefore, in this section, we mainly concentrate our attention to a discrete dynamical system (M, ϕ).

7.2.1 Local stable manifold and stable manifold

Let x be a fixed point: $\phi(x) = x$ and B_ϵ be the ball of radius ϵ centered at x.

$$W^s_\epsilon(x) = \{y \in B_\epsilon \,|\, \phi^n(y) \in B_\epsilon, \forall n \in \boldsymbol{N}\} \tag{7.2.1}$$

is called the *local stable manifold* of size ϵ of x. If ϕ is replaced by ϕ^{-1}, we obtain the *local unstable manifold*:

$$W^u_\epsilon(x) = \{y \in B_\epsilon \,|\, \phi^{-n}(y) \in B_\epsilon, \forall n \in \boldsymbol{N}\} \tag{7.2.2}$$

We can extend the local manifold into *global* stable and unstable manifolds:

$$W^s(x) = \cup_{n=0}^{\infty} \phi^{-n}(W^s_\epsilon(x)) \tag{7.2.3}$$

is called the *stable manifold* of x. Analogously, the unstable manifold can be defined as

$$W^u(x) = \cup_{n=0}^{\infty} \phi^{n}(W^u_\epsilon(x)). \tag{7.2.4}$$

That is, (for diffeomorphic systems) the stable manifold of ϕ^{-1} is the unstable manifold of ϕ.

Remark 1. Unstable and stable manifolds need not be submanifolds of M (look, e.g., at the horseshoe system in **7.3.5**). However, if there are only a finite number of fixed points, then these manifolds are submanifolds. □

---

[14] For example, we can write $\|f\|_s = \sum_{k=0}^{s} \|f^{(k)}\|$, where $\|f\|$ is the usual sup norm: $\|f\| = \sup_{x \in D} |f(x)|$ with D being the domain.

7.2.2 Stable manifold theorem[15]

Let ϕ be a r-times continuously differentiable diffeomorphism on M (i.e., D^r-dynamical system), and x be a hyperbolic fixed point (→**7.1.5**) of ϕ. Let E^s be the stable subspace of $D\phi$ at x.[16] Then,

(1) $W^s(x)$ is a C^r-manifold injectively immersed in M.
(2) E^s is the tangent space of $W^s(x)$ at x.
(3) Let J be the immersion map[17] of $W^s(x)$ into M. Then, there is a contraction map $g : W^s(x) \to W^s(x)$ such that

$$Jg = fJ. \tag{7.2.5}$$

7.2.3 Homoclinic and heteroclinic points

Stable manifolds and unstable manifolds need not be disjoint. A point y belonging to $W^s(x)$ and $W^u(x)$ (i.e., $y \in W^s(x) \cap W^u(x)$) is called a *homoclinic point.*

If y belongs simultaneously to an unstable manifold of a fixed point x and to a stable manifold of another fixed point x' (i.e., $y \in W^s(x) \cap W^u(x')$), then y is called a *heteroclinic point.*

The orbit going through a homoclinic point consists of homoclinic points only.

There are infinitely many periodic points in any neighborhood of a homoclinic point.

Roughly speaking, for a generic diffeomorphism stable and unstable manifolds cross transversally.[18,19]

7.3 Structural Stability

7.3.1 Various stabilities

If a dynamical system is given, there are many ways to study its stability = robustness in a certain sense against certain perturbations.[20] We can perturb an orbit and then ask what happens to the orbit. This is a standard method to analyze the stability of a solution.

There are other kinds of perturbations we can impose to the system. Now, the system itself (the dynamical law, map, or the vector field) can be modified, and we can ask what happens to the general behavior of its orbits. In this case the structure of the dynamical system is modified, so such a perturbation is called a *structural perturbation.* If the behavior of the system does not change by this perturbation, we call the system to be *structurally stable.*

---

[15] See, e.g., Katok-Hasselblatt, Section 6.2.; S. Smale, Bull. Amer. Math. Soc., **73**, 747-817 (1967). The latter is a famous article worth reading.

[16] That is, the subspace of T_xM corresponding to the eigenvalues inside the unit disk.

[17] If D_xJ is injective, we say J is an *immersion.*

[18] That two manifolds A and B cross *transversally* at x means that $T_xA \cap T_xB = \{0\}$.

[19] **Theorem [Kupka-Smale]**. For a generic diffeomorphism, for any fixed points x and y $W^s(x)$ and $W^u(y)$ cross transversally, if they cross at all. □
(The reference for this theorem is the same as the previous Kupka-Smale theorem.)

[20] Incidentally, the 'stability' question of an ecosystem includes an interesting conceptual problems.

7.3.2 Structural stability, Ω-stability

Let ϕ be a diffeomorphism. We say ϕ is structurally stable, if there is a neighborhood U of ϕ (with respect to the Whitney norm →**7.1.7**) such that $\forall\psi \in U$ is homeomorphic to ϕ (that is, there is a homeomorphism h such that $h\psi = \phi h$), we say ϕ is *structurally stable.*

If ϕ is structurally stable when restricted to the ω-limit set Ω, ϕ is said to be Ω-stable.

7.3.3 Peixoto's theorem[21]

The following Peixoto's theorem was one of the break throughs that initiated the modern study of dynamical systems.

A vector field X is structurally stable if it has a neighborhood whose elements are all homeomorphic to X.

Theorem [Peixoto]. A necessary and sufficient condition for a vector field X on a compact 2-manifold to be structurally stable is that X satisfies the following four conditions:
(1) There are only finitely many singular points[22] and they are all hyperbolic.
(2) All the ω- and α-limit sets are singular points or periodic orbits.
(3) There is no orbit connecting saddles.
(4) There are only finitely many periodic orbits and they are all hyperbolic. □

Theorem [Peixoto]. Structural stable vector fields on a compact 2-manifold are open-dense with respect to the W^1-norm.[23] □

Can we generalize this to higher dimensional manifolds? This was a major driving force to study dynamical systems. The answer is, unfortunately, no.

7.3.4 Morse-Smale dynamical system

Abstracting the Peixoto conditions (1)-(4) in **7.3.3** Smale introduced the *Morse-Smale system* defined as a diffeomorphism satisfying the following three conditions:
(MS-1) The Ω-limit set (the totality of the ω-limit sets) consists of periodic points only.
(MS-2) All the periodic orbits are hyperbolic.
(MS-3) For any $x, y \in \Omega$, $W^s(x)$ and $W^u(y)$ are transversal.

Theorem [Palis-Smale].[24] Let M be a compact C^∞- manifold without boundary. Then, the totality of Morse-Smale systems on M is structurally stable in $C^\infty(M.M)$. □

However, not all the structurally stable systems are Morse-Smale. Actually, it is known that the Morse-Smale systems are not even dense if the dimension of M is not smaller than 3.

Theorem [Shub].[25] In the set $D^r(M)$ of C^r-diffeomorphisms structurally stable dynamical systems are C^0-dense. □

Even if a set A is open dense in B, it does not mean that most elements of B are in A.[26] Thus, the reader should clearly recognize that Shub's theorem is a very week result.

7.3.5 Smale's horseshoe dynamical system

Smale constructed an apparently simple dynamical system to understand a complicated

---

[21] M. M. Peixoto, Topology **1**, 101-120 (1962); however, a large part is in M. C. Peixoto and M. M. Peixoto, Ann. Acad. Brasil. Sci., **31**, 135-160 (1959) as far as I remember.

[22] the points where X vanishes

[23] This is the Whitney norm made of local C^1-norm.

[24] S. Smale, Bull. Amer. Math. Soc., **66**, 43 (1960); J. Palis, Topology **8**, 385 (1969). It is also explained in Smale's review article that introduced the horseshoe system.

[25] M. Shub, Am. J. Math., **91**, 175-199 (1969).

[26] For example, there is an open dense set whose total length is indefinitely close to zero in $[0, 1]$. See, e.g., B. R. Gelbaum and J. M. H. Olmsted, *Counterexamples in analysis* (Dover version, 2003). Incidentally, this is a book recommended to every careful thinker.

dynamics a periodically perturbed nonlinear oscillator exhibits.[27,28] It is constructed by superposing a square on itself after stretching and bending as shown in the figure.

Notice that a complicated kneading occurs in the horseshoe as seen in the following figure. We can embed this structure into a diffeomorphism on S^2.The resultant diffeomorphism is called the *Smale horseshoe dynamical system.*

Theorem [Smale] The horseshoe dynamical system is structurally stable. □
Theorem [Smale] The ω-limit set of the horseshoe system consists of a source, a sink, and a Cantor set[29] Ω_0 . If the system is restricted to Ω_0, it is isomorphic to the full shift on $\{0,1\}^{\boldsymbol{Z}}$.[30] Therefore, there are infinitely many periodic orbits with arbitrary periods. □

However, Ω_0 is not an attracting set, so we can see complicated dynamical behavior only transiently.

---

[27] N. Levinson, "A second order differential equation with singular solutions," Ann. Math., **50**, 127-153 (1949) [See also, M. Levi, "Quantitative analysis of the periodically forced relaxation oscillators," Mem. Amer. Math. Soc., **244** (1981)]. This paper wanted to understand what was found by Cartwright and Littlewood.

[28] S. Smale, "Differentiable dynamical systems," Bull. Amer. Math. Soc. **73**, 747-817 (1967). This is a readable review article every serious student interested in dynamical systems should read.

[29] = a nowhere dense perfect set; perfect set = set without any isolated point

[30] For a *shift dynamical system* see **7.4.7**.

7.3.6 Anosov system[31]

The Morse-Smale system has only finitely many periodic orbits, but we already know that homoclinicity implies infinitely many periodic orbits (see **7.2.3**). Furthermore, as we have seen in **7.3.5**, there are definitely non-Morse-Smale systems that are structurally stable. Anosov introduced the following dynamical systems motivated by the example below:

A diffeomorphism on a manifold M such that the closure of its Ω (the ω-limit set) is M (i.e., Ω is dense in M) and hyperbolic.

Such a system is called an *Anosov system.*

Example 1 [Thom diffeomorphism or Arnol'd cat map]. Let $f : T^2 \to T^2$ such that it is described by the application of the following linear map

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \tag{7.3.1}$$

on the universal covering space of T^2 (i.e., $\boldsymbol{R}^2$). $W^s(0)$ and $W^u(0)$ transversally cross each other, and the homoclinic points are dense in T^2. Consequently, periodic orbits are dense on T^2. That is, this is an Anosov system. □

Theorem [Anosov]. Anosov systems are structurally stable. If there is a Riemann volume defined on M, then it is an invariant measure and the system is ergodic.[32] □

7.3.7 Axiom A dynamical system[33]

All the systems we have so far discussed satisfy the following two conditions called *Axiom A*:[34]

(a) The Ω-limit set is hyperbolic.

(b) Periodic orbits are dense in Ω.

Theorem [Spectral decomposition theorem]. Let (M, ϕ) be an axiom A system. Then, its Ω-limit set can be decomposed into disjoint sets Ω_i such that Ω_i is invariant, closed, and ϕ_{Ω_i} is topologically transitive.[35] □

It is known that Axiom A systems are not dense among diffeomorphisms.[36]

7.4 Measure Theoretical Dynamical Systems

An excellent introductory textbook for this section is:

P. Walters, *An Introduction to Ergodic Theory* (Springer, 1982).

---

[31] An excellent reference is R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lect. Notes Math. **490** (Springer, 1975).

[32] Actually, it is much more 'chaotic' than mixing.

[33] Bowen quoted already is excellent.

[34] Axiom B is not popular in these days. It requires, roughly speaking, if a stable and unstable manifolds of periodic orbits cross, there must be a transversal crossing point.

[35] A dynamical system (M, ϕ) is *topologically transitive* if there is an orbit that can go through in any neighborhood of any point of M,

[36] However, Pugh demonstrated that (b) is generic.

Up to this point in this Chapter we have discussed dynamical systems without measure. Since the main topic of statistical mechanics is the distribution = measure, we must discuss measure theoretical dynamical systems.

7.4.1 Measure theoretical dynamical systems

If a dynamical system (M, ϕ_t) has an invariant measure μ (→**2.7.4**)

$$\phi_t^{-1}\mu = \mu \tag{7.4.1}$$

we can construct a *measure theoretical dynamical system* (M, ϕ_t, μ).[37]

For a given dynamical system (M, ϕ) there is at least one invariant measure (Theorem due to Krylov and Bogoliubov). Generally speaking, for a given dynamical system there are many (often uncountably many) distinct invariant measures. Consequently, many distinct measure theoretical dynamical systems can be constructed for a given dynamical system. For physicists, thus, a major question is which is natural (which can describe the phenomenon at hand). As we have seen, a lot of difficulty arises from this question. A dynamical system has no intrinsic preference of a particular invariant measure. Thus, the choice of a measure theoretical dynamical system is always extrinsic from the dynamical system's point of view. That is why we need an extra input to do statistical physics.

7.4.2 Ergodicity and transitivity

Let (M, ϕ_t, μ) be a measure-theoretical dynamics. It is called *ergodic* if for any integrable function

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T ds\, f(\phi_s(x)) = \int_M d\mu(x)\, f(x). \tag{7.4.2}$$

The following theorem was mentioned before, but is reproduced here:
Theorem [Birkhoff].[38] For any ϕ_t-invariant measure μ and for any μ-integrable function$f$
(1) There is a time average function $\overline{f}$ such that

$$\overline{f}(x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T ds\, f(\phi_s(x)) \tag{7.4.3}$$

for μ-almost all $x \in M$. This average is invariant under ϕ_t.
(2) For any invariant set A

$$\int_A d\mu(x)\, \overline{f}(x) = \int_A d\mu(x)\, f(x). \tag{7.4.4}$$

(3) If $\phi_t^{-1} = \phi_{-t}$, then we can define a backward time average $\underline{f}(x)$

$$\underline{f}(x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T ds\, f(\phi_{-s}(x)) \tag{7.4.5}$$

and for μ-almost all x $\overline{f} = \underline{f}$.

If any invariant set A satisfies $\mu(A)(1-\mu(A)) = 0$, then the dynamical system is said to be *metrically transitive*. Metrical transitivity and ergodicity are equivalent for measure theoretical dynamical systems.

---

[37] (7.4.1) implies for 'any' set A $\mu(A) = \mu(\phi_t^{-1}(A))$. Precisely speaking, we must set up a measurable space with a σ-field $\mathcal{F}$ of measurable sets, and for any $A \in \mathcal{F}$ (7.4.1) holds. As usual, we ignore such details just as we ignored it when we discussed probability.

[38] See Walters for a proof.

7.4.3 Mixing and decay of correlation

Let A and B be arbitrary sets in M. Then, if

$$\lim_{n\to\infty} \mu(A\cap \phi_t^{-1}(B)) = \mu(A)\mu(B), \tag{7.4.6}$$

we say the dynamical system is (strongly) *mixing*.

It is known that the time correlation function decays to zero for mixing systems, so we definitely have irreversible behaviors (or relaxational behaviors). We cannot, however, say anything about the decay rate.

7.4.4 Partition and entropy of partition

Let (M,μ) be a measurable space. A family $\xi = \{X_\lambda \,|\, \lambda \in \Lambda\}$ of subsets of M is called a *measurable partition* of M, if $X_\lambda \cap X_{\lambda'} = \emptyset$ for $\lambda \neq \lambda'$, and $M = \cup X_\lambda$.[39]

If we have two measurable partitions of M $\xi = \{X_i\}_{i\in\Lambda}$ and $\eta = \{Y_j\}_{j\in K}$, their *join* $\xi \vee \eta$ (the smallest partition compatible with both) is defined as

$$\xi \vee \eta = \{X_i \cap Y_j \,|\, i \in \Lambda, j \in K\}. \tag{7.4.7}$$

The *entropy* $H(\xi)$ of a partition ξ is defined as

$$H(\xi) = -\sum_{i\in\Lambda} \mu(X_i)\log\mu(X_i). \tag{7.4.8}$$

A motivation to introduce a partition of a phase space M is to introduce a coarse-grained observable. That is, a partition $\mathcal{A} = \{A_i\}$ corresponds to a macro observable, and each element A_i corresponds to a particular outcome of a macroscopic observation of $\mathcal{A}$. Observing $\mathcal{A}\vee\mathcal{B}$ corresponds to a simultaneous observation of two observables corresponding to $\mathcal{A}$ and $\mathcal{B}$.

7.4.5 Kolmogorov-Sinai entropy

Let ϕ be the map (unit-time map in the case of the continuous dynamical system) from M onto itself and has an invariant measure μ. Let $\mathcal{A}$ be a measurable partition of M. Then,

$$h(\mu,\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{A} \vee \phi^{-1}\mathcal{A} \vee \cdots \vee \phi^{-n+1}(\mathcal{A})) \tag{7.4.9}$$

is well defined.

The supremum with respect to the partition is called the *Kolmogorov-Sinai entropy*:

$$h(\mu) = \sup_{\mathcal{A}} h(\mu,\mathcal{A}). \tag{7.4.10}$$

The Kolmogorov-Sinai entropy for ϕ_t is proportional to t, so we can compute it from h above.

If there is a partition $\mathcal{G}$ such that $h(\mu,\mathcal{G}) = h(\mu)$, $\mathcal{G}$ is called a *generator*.

The most important property of the Kolmogorov-Sinai entropy is its invariance under isomorphisms. That is, h is *isomorphism invariant*. Two dynamical systems (M,ϕ,μ), (M',ϕ',μ') are isomorphic (we also say two dynamical systems are *conjugate*), if there is a

---

[39] Of course these subsets must be μ-measurable, but such conditions are not explicitly stated in these notes. Each element of a partition can be any so long as it is measurable, It need not be connected nor singly connected,

one-to-one map $q : M \to N$ such that $\mu'(q(A)) = \mu(A)$ and $\phi' \circ q = q \circ \phi$. The Kolmogorov-Sinai entropy is invariant under q.

Let $x \in M$ and $\mathcal{G}$ be a generator. Let $g_n(x)$ be the member of $\mathcal{G} \vee \cdots \vee \phi^{-n+1}(\mathcal{G})$. Then,
Theorem [Shannon-McMillan-Breiman].

$$-\lim_{n\to\infty} \frac{1}{n} \log \mu(g_n(x)) = h_\mu \tag{7.4.11}$$

for μ-almost all $x \in M$. □

7.4.6 Small tube and Brin-Katok's theorem.[40]

The Kolmogorov-Sinai entropy tells us how complicated (on the average) the dynamics is. This complication is due to the sensitivity of each trajectory to small change of the initial condition. If we can measure the 'volume' (probability) of the initial conditions that give orbits staying close to the orbit stating from x, then we could quantitatively study the extent of sensitivity of the orbit to the initial condition.

For $x \in M$ consider

$$B_T(x, \epsilon) = \{y \in M \mid d(\phi^t x, \phi^t y) \le \epsilon, t \in [0, T)\}. \tag{7.4.12}$$

This is the volume of the initial conditions that give orbits that are within distance ϵ of the orbit starting at x for time span of T.

Theorem [Brin-Katok]. Suppose (M, ϕ, μ) is an ergodic dynamical system. Then, for μ-almost all $x \in M$

$$h(\mu) = -\lim_{\epsilon\to 0}\lim_{T\to\infty} \frac{1}{T} \log \mu(B_T(x, \epsilon)). \tag{7.4.13}$$

□

The theorem is sometimes called the local entropy theorem. This should not be a surprise, if one knows the Shannon-McMillan-Breiman theorem (→**7.4.5**).

7.4.7 Bernoulli shift, Bernoulli system and Ornstein's theorem[41]

Let M be the totality of both-side infinite sequences $\{q_n\}_{n=-\infty}^{\infty}$, where $q_n \in \{a_1, \cdots, a_p\}$. The *shift* σ is an operation to shift the sequence as

$$\sigma\{x_i\} = \{y_i\}, \tag{7.4.14}$$

where $y_i = x_{i-1}$. If we introduce an invariant measure through the following probability of cylinder sets,

$$\mu\left(\{x_i\}_{i=\alpha}^{\beta}\right) = \prod_{i=\alpha}^{\beta} p(x_i), \tag{7.4.15}$$

where $p(x_i) = p_s$ if $x_i = a_s$ with $\sum_{s=1}^{p} p_s = 1$ and $p_s > 0$. The dynamical system (M, σ, μ) is called the *Bernoulli shift* $B(p_1, \cdots, p_p)$.

A dynamical system conjugate (i,.e., isomorphic) to a Bernoulli shift is called a *Bernoulli dynamical system* (*Bernoulli automorphism*).

Theorem [Ornstein]. The Kolmogorov-Sinai entropy is a perfect invariant of Bernoulli automorphisms. □

---

[40] M. Brin and A. Katok, *On local entropy* Lect. Notes Math.,

[41] D. Ornstein, *Ergodic theory, randomness, and dynamical systems*, (Tale UP, 1974).

That is, if the Kolmogorov-Sinai entropy is identical, then two Bernoulli automorphisms are isomorphic. Actually, the Kolmogorov-Sinai entropy was initially introduced to classify Bernoulli systems.

Example 1 [Baker's transformation]. Let $M = [0,1)^2$ and μ be the Lebesgue measure μ. Let ϕ be defined as

$$\phi(x,y)) = \begin{cases} (2x, y/2) & x \in [0,1/2) \\ (2x-1, y/2+1/2) & x \in [1/2,1). \end{cases} \tag{7.4.16}$$

The dynamical system (M, μ, ϕ) is called the *baker's transformation*. This is a Bernoulli system conjugate to $B(1/2, 1/2)$.□

7.4.8 Chaos

Let (M, ϕ, μ) be a measure-theoretical dynamical system. If there is an invariant set A with a positive measure of ϕ^n for some $n > 0$ such that (A, ϕ^n, μ) is isomorphic to a Bernoulli system,[42] we say the dynamical system exhibits *chaos*.[43]

Let $\phi : I \to I$ be a continuous map from an interval I into itself. This is called an *interval C^0-endomorphism*.

Theorem [Oono-Osikawa].[44] The following conditions are necessary and sufficient conditions for a C^0-endomorphism of an interval to exhibit chaos:

(1) There is an orbit with period $\neq 2^n$.

(2) The Kolmogorov-Sinai entropy is positive.

(3) There is a positive integer n such that ϕ^n has a mixing invariant measure.

These can be proved from the equivalence to

(4) There are two intervals $A, B \in I$ sharing at most one point such that for some positive integers n and m, $\phi^n(A) \cap \phi^m(B) \supset A \cup B$. □

This implies that positivity of the Kolmogorov-Sinai entropy is a good characteristics of a chaotic system. Later, Ornstein simply defined that the system with a positive Kolmogorov-Sinai entropy as a chaotic system.[45]

Li and Yorke introduced the following definition of chaos and wrote the famous paper, "Period 3 implies chaos."[46] They use the *scrambled set*: S is a scramble set if

(1) $\forall x, y \in S$

$$\limsup_{n\to\infty} d(\phi^n(x), \phi^n(y)) > 0, \quad \liminf_{n\to\infty} d(\phi^n(x), \phi^n(y)) = 0, \tag{7.4.17}$$

(2) For any periodic point $p \in M$ and for any $x \in S$

$$\limsup_{n\to\infty} d(\phi^n(x), \phi^n(p)) > 0, \quad \liminf_{n\to\infty} d(\phi^n(x), \phi^n(p)) = 0, \tag{7.4.18}$$

---

[42] Or to a Bernoulli endomorphism, when the system is not invertible.

[43] due to Y. Oono "Period $\neq 2^n$ implies chaos," Prog. Theor. Phys. **59**, 1029-30 (1978). The equivalence of this definition and the currently popular Devaney's definition proposed much later is shown in S.-H Li, "ω-chaos and topological entropy," Trans. Amer. Math. Soc. **339**, 243-249 (1993).

[44] Y. Oono and M. Osikawa, "Chaos in nonlinear difference equations I qualitative study of (formal) chaos," Pror. Theor. Phys., **64**, 54-67 (1980); M. Osikawa and Y. Oono, "Chaos in C^0-endomorphisms of intervals," Publ. RIMS **17**, 165-77 (1981).

[45] D. S. Ornstein, "In what sense can a deterministic system be random?" Chaos, Solitons & Fractals **5**, 139 (1995),

[46] T.-Y. Li and J. A. Yorke, "Period three implies chaos," Am. Math. Month. **82**, 985-992 (1975). Although this is very similar to "Period $\neq 2^n$ implies chaos," the definition of chaos in these statements are distinct. Li and Yorke's chaos does not imply the positivity of the Kolmogorov-Sinai entropy.

Theorem [Li-Yorke]. If the system has a periodic orbit of period 3, then there is a scrambled set. □

Li-Yorke's chaos implies chaos, but the converse is not true. Thus, the Li-Yorke chaos lacks nice equivalent characterizations. Furthermore, for typical chaos S is, if measurable, it is measure zero:

Theorem [Baba-Kubo-Takahashi].[47] If ϕ is piecewise C^1 with an absolutely continuous invariant measure with a positive Kolmogorov-Sinai entropy, S is, if measurable, measure zero. □

7.4.9 Shift and random number

The definition of chaos given in **7.4.8** and random sequences are intimately connected, because we may interpret Bernoulli dynamical systems as a model of coin-tossing or some other gambling games. It is not an easy task to characterize randomness, but intuitively chaos is something like a deterministic mechanism to 'produce' randomness. Thus, the definition **7.4.8** achieves to capture this intuition without explicitly characterizing randomness.

To be more direct, we must mathematically clarify the concept of randomness to characterize chaos. Perhaps currently the most popular definition of random sequences is due to Solomonov, Kolmogorov and Chaitin.[48] The idea is roughly as follows:

Let $x = \{x_n\}$ be an infinite sequence, and q_n is the shortest program required by a universal Turing machine[49] to print out the first n symbol sequence $[x]_n$ of x. If the length of the program ℓ_n of p_n grows linearly with n (i.e., $\ell_n/n > 0$ for all n), we say x is an *algorithmically random* sequence. That is, if we cannot compress the sequence$x$ efficiently (because we cannot find any regularity in x), we say x is a random sequence.

There is a theorem (Brudno's theorem[50]) asserting that ℓ_n/n converges to the Kolmogorov-Sinai entropy of the dynamical system for μ-almost all initial condition x

7.4.10 Lyapunov exponent and Oseledec's theorem

Let $x \in M$ and $v \in T_xM$ (that is, v is a vector in the tangent space of M at x). The *Lyapunov exponent* of v at x is defined by

$$\lambda(x, v) = \limsup_{n\to\infty} \frac{1}{n} \log \|D\phi^n(x)v\|. \tag{7.4.19}$$

For each x, $\lambda(x, \cdot)$ can take only finitely many distinct values:

$$\lambda^{(1)}(x) > \cdots > \lambda^{(q)}(x), \tag{7.4.20}$$

where q depends on x (but no more than the dimension of M). If the dynamical system is ergodic, there is no x-dependence for μ-almost all x.

Oseledec's theorem asserts not only the existence of these exponents for μ-almost all x,

---

[47] Y. Baba, I. Kubo and Y. Takahashi, "Li-Yorke's scrambled sets have measure 0" Nonlinear Analysis **26**, 1611-3 (1996).

[48] Perhaps the best introductory textbook is: M. Li and P. Vitányi, *An introduction to Kolmogorov complexity and its applications* (Springer 1993). An outline may be found in my lecture notes for Applicable Analysis (the last section).

[49] Roughly, it is an ordinary digital computer with infinite memory.

[50] A. A. Brudno, "Entropy and the complexity of trajectories of a dynamical system," Trans. Mosc. Math. Soc., **44**, 127-151 (1983); a readable paper.

but also asserts that $T_x \simeq \boldsymbol{R}^n$ ($n = dim M$) is decomposed into the direct sum of eigenspaces:
Theorem [Oseledec].[51] At each $x \in M$ T_xM ($\simeq \boldsymbol{R}^n$) is decomposed as

$$T_xM = \bigoplus_{i=1}^{q(x)} H_i(x) \tag{7.4.21}$$

and if $v \in H_j(x)$

$$\limsup_{n\to\infty} \frac{1}{n} \log \|D\phi^n(x)v\| = \lambda_j(x). \tag{7.4.22}$$

□

If the dynamical system (M, ϕ_t, μ) is continuous, and the support of μ is not a single point, then there must be at least one exponent that vanishes (whose eigendirection is tangent to the orbits).

The Lyapunov exponents of the time reversed system are the same with opposite sign. Since Hamiltonian systems are time reversal symmetric, if λ is a Lyapunov exponent, so is $-\lambda$. That the total sum of all the Lyapunov exponents vanishes is nothing but Liouville's theorem.

The directions of the eigenspaces corresponding to the positive exponents may be understood as sensitive directions to perturbation. Therefore, if a dynamical system has invariants (e.g., conserved quantities), then some exponents must be zero. This implies that a closed Hamiltonian system must have at least two vanishing Lyapunov exponents.

7.4.11 Ruelle inequality

Let

$$\lambda_+(x) = \sum_{i:\lambda_i(x)>0} \lambda_i(x). \tag{7.4.23}$$

Here, the summation is calculated with the multiplicity of these eigenvalues taken into account. Then,

$$h_\mu(\phi) \le \int \lambda_+(x) d\mu. \tag{7.4.24}$$

This is called the *Ruelle inequality.*[52] If a system is ergodic, then, the Kolmogorov-Sinai entropy is bounded from above by the sum of positive Lyapunov exponents.

Notice that the exponential growth rate of the volume element $dx_1 \wedge \cdots \wedge dx_k$ in the local unstable manifold $W^u(x)$ is given by the exponent $\lambda_+(x)$.

For 'typical chaotic systems' (e.g., axiom A systems) the above inequality becomes equality (*Pesin's equality*). More precisely, if a system is chaotic, then there is an invariant measure that is observable (→**7.5.2**) experimentally that satisfies Pesin's equality as we will see later (→**7.5.3**).

Just as the relation between the tube volume around the orbit and the Kolmogorov-Sinai entropy (→**7.4.6**), we can have the following theorem as a version of Oseledec's theorem:
Theorem. $B_T(x, \epsilon)$ is defined as in **7.4.6**. Then,

$$\lambda_+ = -\lim_{\epsilon\to 0} \lim_{T\to\infty} \frac{1}{T} \log vol(B_T(x, \epsilon)) \tag{7.4.25}$$

for μ-almost all x. □

---

[51] For a proof, see Katok-Hasselblatt p665. The original Oseledec's theorem is about the general *cocycle* of a dynamical system, and is called Oseledec's *multiplicative ergodic theorem*, but here we confine ourselves to $D\phi$.

[52] D. Ruelle, "An inequality for the entropy of differentiable maps," Bol. Soc. Brasil. Mat., **9**, 83-87 (1978).

7.5 Thermodynamic Formalism

Standard references are
R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lect. Notes Math. **490** (Springer, 1975).
This is a real classic.
D. Ruelle, *Thermodynamic Formalism* (Addison-Wesley, 1978).

There are (usually) uncountably many invariant measures for a given dynamical system (M, ϕ), so a serious question arises about the 'natural' measure-theoretical dynamical system (M, ϕ, μ). What is the best choice of μ? The question is answered by the large deviation theoretical principle. The large-deviation theoretical framework for dynamical systems is called the *thermodynamic formalism* for dynamical systems.

In this section, the thermodynamical formalism is explained from the large deviation point of view in the style of theoretical physics as much as possible (i.e., demonstrations are through formal manipulation of formulas as much as possible).[53]

7.5.1 Probability of observing an invariant measure

Let us estimate the probability to observe an invariant measure μ of a diffeomorphism ϕ. If this is positive, then it is fair to say that such a statistical behavior of a given dynamical system can be actually observed experimentally or computationally. Needless to say, we must choose an appropriated sampling measure m of the initial condition. We choose this to be the ordinary Lebesgue measure.[54]

We need essentially the level-2 large deviation problem to estimate:[55]

$$P\left(\frac{1}{T}\sum_{k=0}^{T-1}\delta_{\phi^k(x)} \sim \mu\right) \tag{7.5.1}$$

However, for a diffeomorphism, it is better to use a time-symmetric quantity:

$$P\left(\frac{1}{2T+1}\sum_{k=-T}^{T}\delta_{\phi^k(x)} \sim \mu\right) \tag{7.5.2}$$

We use the generating function (partition function) to compute this:[56]

$$Z_T(U) = \int dm(x) \exp\left(\sum_{k=-T}^{T} U(\phi^k(x))\right). \tag{7.5.3}$$

---

[53] The exposition freely uses unpublished lecture notes of and private communications with Y. Takahashi (RIMS, Kyoto University).

[54] the reason why we choose this transcends physics as we have already discussed in conjunction to the principle of equal probability.

[55] Introducing symbolic dynamics and discuss this problem from the level-3 point of view may be more transparent, but here we minimize our tools. See, e.g., Y Oono, "Large deviation and Statistical Mechanics," Prog. Theor. Phys. Supple **99**, 165 (1989)

[56] The reader might wonder what happens if we introduce a temperature parameter β. This parameter controls the Hausdorff dimension of the support of the invariant measure.

Here, m is the sampling measure of the initial condition (the Lebesgue measure in this section).

If we introduce the 'free energy density' (or it is often called in the present context as *topological pressure*):[57]

$$p(U) = \lim_{T\to\infty} \frac{1}{2T+1} \log Z_T(U), \tag{7.5.4}$$

then, the probability we wish can be written (formally) as the following functional integral (our exposition is informal = theoretical physics level, so $T \to \infty$ is not always explicitly written; if T is sufficiently large, these formulas should be accurate):

$$P\left(\frac{1}{2T+1}\sum_{k=-T}^{T} \delta_{\phi^k(x)} \sim \mu\right) = \int \mathcal{D}(U) \exp\Big((2T+1)\int U(x)d\mu(x) - (2T+1)p(U)\Big). \tag{7.5.5}$$

The calculation above is quite parallel to the formal calculation explained in the demonstration of Sanov's theorem (→**1.6.8**). Thus, if we introduce the rate function $I_m(\mu)$ as[58]

$$P\left(\frac{1}{2T+1}\sum_{k=-T}^{T} \delta_{\phi^k(x)} \sim \mu\right) \sim e^{-(2T+1)I(\mu)}, \tag{7.5.6}$$

we obtain

$$I(\mu) = \sup_U \left[\int d\mu(x)U(x) - p(U)\right]. \tag{7.5.7}$$

7.5.2 Observable measure

Following the formal calculation we did long ago (→**1.6.8**), the functional derivative of the RHS of (7.5.7) wrt U gives

$$\mu(\delta_x) = \lim_{T\to\infty} \frac{1}{2T+1}\int dm(y) \left(\sum_j \delta(x-\phi^j(y))e^{\sum_{k=-T}^{T} U(\phi^k y)}\right)\Bigg/ m\left(e^{\sum_{k=-T}^{T} U(\phi^k y)}\right), \tag{7.5.8}$$

where δ_x is an atomic measure concentrated at a, so formally $\mu(\delta_x)dx = d\mu(x)$. This equation determines U that gives $I(\mu)$.

Consider

$$\tilde{\mu}(x) = \lim_{T\to\infty} \int dm(y)\delta(x-y)e^{\sum_k U(\phi^k y)}/m\left(e^{\sum_k U(\phi^k y)}\right). \tag{7.5.9}$$

Then,

$$\tilde{\mu}(\phi(x)) = \lim_{T\to\infty} \int dm(y)\delta(\phi(x)-y)e^{\sum_k U(\phi^k y)}/m\left(e^{\sum_k U(\phi^k y)}\right), \tag{7.5.10}$$

---

[57] The existence of the thermodynamic limit is assumed. The correspondence between the dynamical systems and the 1D equilibrium lattice system was first realized by Sinai. The Gibbs measure in equilibrium statistical mechanics corresponds to the invariant measure with potential U given in (7.5.13). A good exposition is Bowen quoted at the beginning of this section; this also explains a lot about axiom A systems.

[58] Mathematically speaking, we must show that the rate function exists. However, the existence is essentially the 'extensivity condition.' We assume this or equivalently we assume the large deviation principle for μ just as physicists always assume that the thermodynamic limit exists.

Now, let $y = \phi(z)$. Then,

$$\begin{aligned}\tilde{\mu}(\phi(x)) &= \lim_{T\to\infty} \int dm(z)\, |\phi'(z)|\, \delta(\phi(x) - \phi(z)) e^{\sum_k U(\phi^{k+1}z)}/m\left(e^{\sum_k U(\phi^k y)}\right), && (7.5.11)\\ &= \lim_{T\to\infty} \int dm(z) \delta(x - z) e^{\sum_k U(\phi^{k+1}z)}/m\left(e^{\sum_k U(\phi^k y)}\right). && (7.5.12)\end{aligned}$$

If the convergence in the $T \to \infty$ is assumed, then this must be the same as $\tilde{\mu}(x)$. Therefore, $\mu = \tilde{\mu}$.

This implies

$$\frac{d\mu}{dm}(x) = \lim_{T\to\infty} e^{\sum_k U(\phi^k x)}/m\left(e^{\sum_k U(\phi^k x)}\right), \tag{7.5.13}$$

Notice that this has the form of the canonical distribution function. If we know U, we can have an 'explicit' expression of an observable invariant measure.

Using this formula, we obtain a counterpart of Sanov's theorem as follows. Taking the logarithm of (7.5.13), we obtain

$$\frac{1}{2T+1} \log \frac{d\mu}{dm}(x) = \frac{1}{2T+1} \sum_k U(\phi^k x) - p(U). \tag{7.5.14}$$

Integrating this with respect to μ as

$$\frac{1}{2T+1} \int d\mu(x) \log \frac{d\mu}{dm}(x) = \int d\mu(x) U(x) - p(U), \tag{7.5.15}$$

we obtain the rate function:

$$I(\mu) = \frac{1}{2T+1} \int d\mu(x) \log \frac{d\mu}{dm}(x). \tag{7.5.16}$$

This formula is a strange formula, because the integrand does not seem to have T-dependence. However, we must carefully look at the approximate formula for $d\mu/dm$. The problem is in $\mu(\delta_x)$ already, We need a slightly more careful formulation. For finite T, we should also coarse-grain the space appropriately so that the volume Δ_x containing x satisfies the condition that $\phi^i(\Delta_x)$ for each $i \in \{-T, \cdots, T\}$ is in a ball of radius ϵ (a small positive number; this automatically goes to zero as $T \to \infty$). Now, (7.5.13) reads

$$\frac{\mu(\Delta_x)}{m(\Delta_x)} = e^{\sum_k U(\phi^k \Delta_x x)}/m\left(e^{\sum_k U(\phi^k \Delta_x)}\right). \tag{7.5.17}$$

Thus, (7.5.16) reads, more precisely,

$$I(\mu) = \lim_{T\to\infty} \frac{1}{2T+1} \int d\mu(x) \log \frac{\mu(\Delta_x)}{m(\Delta_x)}. \tag{7.5.18}$$

Remember that Δ_x shrinks to zero as $T \to \infty$.

(7.5.13) tells us that it is convenient to have an explicit formula for U that realizes the infimum. To this end, we must study the structure of $I(\mu)$.

7.5.3 Expression of rate function for ergodic systems

The Radon-Nikodym derivative in (7.5.16) can be computed as

$$\frac{d\mu}{dm}(x) = \lim_{T\to\infty} \frac{\mu(\Delta_x)}{m(\Delta_x)}. \tag{7.5.19}$$

$$\log \frac{\mu(\Delta_x)}{m(\Delta_x)} = \sum \left(\log \frac{\mu(\phi^{-j}\Delta_x)}{\mu(\phi^{-j+1}\Delta_x)} - \log \frac{m(\phi^{-j}\Delta_x)}{m(\phi^{-j+1}\Delta_x)} \right). \tag{7.5.20}$$

so (recall that m is the Lebesgue measure)

$$I(\mu) = \lim_{\epsilon\to 0}\lim_{T\to\infty} \frac{1}{T} \log \frac{\mu(B_T(x,\epsilon))}{m(B_T(x,\epsilon))} = \lambda_+ - h(\mu). \tag{7.5.21}$$

The above formula can be proved for compact M.

Since the rate function cannot be negative, (7.5.21) implies Ruelle's inequality. For μ to be observable, $P > 0$, so $I(\mu)$ must vanish. This implies Pesin's equality.

7.5.4 The relation between equilibrium lattice system and dynamical system

If we could identify $U = U_\mu$ that gives the equality (7.5.21) from

$$I(\mu) = \int U_\mu d\mu - p(U_\mu), \tag{7.5.22}$$

we can identify the observable measure. However, an oversimplified exposition up to this point[59] seems to prevent an easy identification.

Therefore, a new ingredient, actually the original form of the thermodynamic formalism due to Sinai is briefly outlined here.[60]

The basic idea is that a discrete dynamical system may be mapped to a shift dynamical system by a suitable coding (→**7.4.7**).[61] If an orbit containing $x \in M$ is coded as $\{a_n\}$, then we can interpret this as a spin configuration on a 1D lattice. A bunch of orbits sharing n symbols $a_1, \cdots, a_n$ is called a *cylinder set* of length n and is denoted as $[a_n \cdots, a_n]$.

The partition function for the lattice system reads[62]

$$\hat{Z}(U) \simeq \hat{Z}_n(U) = \sum_{[a_1,\cdots,a_n]} \exp\left(\sum_{k=0}^{n-1} U(\phi^k x) \right), \tag{7.5.23}$$

where the sum is take over all the cylinder sets of length n, $x \in [a_n \cdots, a_n]$ is a representative point in the cylinder set, and U is, as before, a function (potential function). Here, $\simeq$ implies that we must take the thermodynamic limit ($n \to \infty$) appropriately.

The entropy per spin in this correspondence just becomes the Kolmogorov-Sinai entropy of the dynamical system.[63] Therefore, if we introduce 'free energy' (this is called *topological*

---

[59] In essence, the oversimplification corresponds to a level-3 large deviation theory told as if it were a level-2 theory (→**3.2.1** for levels); since the system is deterministic, a single point (i.e., the level-2 specification) on an orbit completely determines the orbit (= the elementary object of the level-3 theory). Perhaps I should have started with this entry and reorganized this chapter more statistical mechanically, but for now the revision is kept to a minimum.

[60] For details, see Bowen Chapters 1 and 2.

[61] To be concrete a lot of devices such as Markov partition etc., are needed.

[62] This is with a hat, because this partition function is distinct from Z considered up to this point.

[63] This may be intuitively obvious, but more formally, we can show this relation by going back to the definitions of the both quantities.

pressure, because mathematicians regard the sum as the grand canonical partition function) $P(U)$ as[64]

$$P(U) = \lim \frac{1}{n} \log \hat{Z}_n(U) \tag{7.5.24}$$

Then, the Gibbs variational principle reads

$$P(U) = \sup_{\mu} \left\{ h(\mu) + \int U d\mu \right\}. \tag{7.5.25}$$

The equilibrium distribution function (that is, μ maximizing (7.5.25)) is given by (7.5.13) or, more precisely, by (7.5.17). It is just the canonical distribution.

7.5.5 Characterization of observable measure

We must choose an appropriate U to make the canonical distribution (7.5.13) as the observable measure, which is characterized by $I(\mu) = \lambda_+ - h(\mu) = 0$ according to **7.5.3**.

Let $U = -\lambda_+(x)$ (→**7.4.11**). Then, $P(-\lambda_+) \leq 0$: in this case the partition function can be written as

$$\hat{Z}_n = \sum e^{-\sum \lambda_+(\phi^k x)} \leq m(\cup_x B_n(\epsilon, x)), \tag{7.5.26}$$

where $B_n(\epsilon, x)$ is defined in (7.4.12). We have used the relation between λ_+ and the volume of the small tube according to Oseledec (→**7.4.10**). Thus, $\hat{Z}(-\lambda_+) \leq 1$, so $P \leq 0$.

We already know (→**7.5.3**) that for the observable measure Pesin's equality holds, so $P(-\lambda_+) = 0$. That is, if we choose $U = -\lambda_+(x)$ the resultant canonical distribution is the observable measure.

7.5.6 Sinai-Ruelle-Bowen measure

If we apply the thermodynamic formalism to Axiom A systems, the observable measure is called the *SRB measure* (Sinai-Ruelle-Bowen measure).

Axiom A systems are characterized by the hyperbolic Ω-set in which periodic orbits are dense. Consider the set of points starting from a small ball. The ball will be pushed toward W^u and then be spread along W^u. λ_+ is the spreading rate of the volume along the unstable manifold. As is already clear from this explanation, the support of the SRB measure is the unstable manifold in the Ω-limit set.[65] $U = -\lambda_+$ implies that the weight is inversely proportional to the spreading rate along the unstable manifold. This is a natural weight, because the weight is proportional to the sojourn time of the point along the unstable manifold.

The support of a SRB measure is generally locally isomorphic to the direct product of a Cantor set and an interval. Notice that the actual volume of the support of the SRB measure is usually zero. That is, the ball discussed above spreads along the unstable manifold, but also it is compressed toward W_u and loses its volume.

---

[64] Throughout this chapter if the limit does not exist, replace it with lim sup,

[65] For an axiom A system, its Ω-limit sets are decomposed into disjoint sets Ω_i (the *spectral decomposition theorem* for axiom A system →**7.3.7**). Therefore, invariant measures must be considered for individual spectral components.

7.6 Chaotic Hypothesis and Fluctuation Theorem

7.6.1 Chaotic hypothesis

The hypothesis reads, according to its proposer[66]

> A many-particle system in a stationary state can be regarded, for the purpose of computing macroscopic properties, as a smooth dynamical system with a transitive axiom A global attractor. In the reversible case it can be regarded, for the same purpose, as a smooth transitive Anosov system.

No one knows whether this can be a reliable principle.

There are at least two motivations for this hypothesis. One is that any open system cannot be modeled by a Hamiltonian system (cf., **X2.5.1**), and often (computationally) modeled by a dissipative (i.e., phase volume non-conserving) dynamical system.[67] The other is that for highly chaotic system, as we will see, the fluctuation theorem (→**4.1.3**) becomes a theorem.

7.6.2 Phase contraction and entropy production

Consider a dynamical system whose equation of motion is given by

$$\frac{d}{dt}x = f(x). \tag{7.6.1}$$

The density distribution function ρ obeys the following conservation law (→**3.5.1**):

$$\frac{\partial}{\partial t}\rho = -div(f\rho). \tag{7.6.2}$$

Now, let us compute the time derivative of the Gibbs-Shannon entropy:

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{d}{dt}\int dx\,\rho\log\rho = -\int dx\,(1+\log\rho)\frac{\partial\rho}{\partial t}, && (7.6.3)\\
&= \int dx\,(1+\log\rho)div(f\rho) = -\int dx\, f\cdot grad\, f, && (7.6.4)\\
&= \int dx\,\rho\, div f && (7.6.5)
\end{aligned}$$

Thus, in the steady state

$$\frac{dS}{dt} = \langle div f\rangle. \tag{7.6.6}$$

Notice that $div f$ is the growing rate of the phase volume.

If the system is a finite system, then S must be bounded from above. Therefore, $\langle div f\rangle$ cannot be positive. Recall that this is indeed negative for axiom A systems.[68] If we

---

[66] G. Gallavotti, "Chaotic hypothesis: Onsager reciprocity and fluctuation-dissipation theorem," J. Stat. Phys., **84**, 899 (1996).

[67] A typical example is the Nosé-Hoover algorithm. See, for example, D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids* (Academic Press, 1990).

[68] A rigorous proof of this is found in D. Ruelle, "Positivity of entropy production in nonequilibrium statistical mechanics," J. Stat. Phys., **85**, 1 (1996).

may interpret S as entropy,[69] then this decrease should be due to the escape of dissipated energy into the bath. Therefore, we interpret

$$\langle\sigma\rangle = -\langle div f\rangle \tag{7.6.7}$$

as the entropy production rate.

Remark 1. In reality, there must be an interaction with the heat bath of the system that is not explicitly described in (7.6.1). With such a term added the effect of the phase contraction is compensated by the interaction. Thus, the situation is just as the ordinal Langevin equation. □

Remark 2. For a Hamiltonian system with a stochastic term (+ dissipation), the absolute continuity of the stationary measure with respect to the phase volume is believed to be preserved. Thus, the steady distribution purported by Chaotic Hypothesis is unrealistic. There must be a macro observable that denies this hypothesis.

7.6.3 Entropy production and fluctuation theorem

Let us consider the level-1 large deviation theory for entropy production rate:

$$P\left(\frac{1}{T}\int \sigma dt \sim s\right) \sim e^{-T\zeta(s)}. \tag{7.6.8}$$

Here, ζ is the rate function. The needed probability for an axiom A system may be written as

$$P\left(\frac{1}{T}\int \sigma dt \sim s\right) \propto \sum_{\overline{\sigma}=s} \prod_{k=-T/2}^{T/2} \exp\left(-\lambda_+(\phi^k x)\right). \tag{7.6.9}$$

Here, we have adopted the observable measure (the SRB measure) constructed in the previous section, and the summation is over all x satisfying $\overline{\sigma} = s$ with $\overline{\sigma} = (1/T)\int_{-T/2}^{T/2} \sigma dt$. On the other hand

$$P\left(\frac{1}{T}\int \sigma dt \sim -s\right) \propto \sum_{\overline{\sigma}=-s} \prod_{k=-T/2}^{T/2} \exp\left(-\lambda_+(\phi^k x)\right). \tag{7.6.10}$$

We wish to study its time-reversed version. Let $\mathcal{I}$ be the time reversal operator. $\mathcal{I}W_x^u = W_{\mathcal{I}x}^s$, and $\mathcal{I}\overline{\sigma}(x) = -\overline{\sigma}(\mathcal{I}x)$. Therefore,

$$P\left(\frac{1}{T}\int \sigma dt \sim -s\right) \propto \sum_{\overline{\sigma}=s} \prod_{k=-T/2}^{T/2} \exp\left(\lambda_-(\phi^k x)\right). \tag{7.6.11}$$

Here, λ_- is the sum of negative Lyapunov exponents. From this we can compute the ratio

$$\frac{P\left(\frac{1}{T}\int \sigma dt \sim s\right)}{P\left(\frac{1}{T}\int \sigma dt \sim -s\right)} = \frac{\sum_{\overline{\sigma}=s}\prod_{k=-T/2}^{T/2} \exp\left(-\lambda_+(\phi^k x)\right)}{\sum_{\overline{\sigma}=s}\prod_{k=-T/2}^{T/2} \exp\left(\lambda_-(\phi^k x)\right)}. \tag{7.6.12}$$

---

[69] This is actually a big if. We do not know how to measure nonequilibrium entropy, so no operational definition of S is available. That is, there is no physically respectable characterization of nonequilibrium entropy.

Notice that the terms in the numerator and those in the denominator are one to one correspondent. Consider the following ratio:

$$\frac{\prod_{k=-T/2}^{T/2}\exp\left(-\lambda_+(\phi^k x)\right)}{\prod_{k=-T/2}^{T/2}\exp\left(\lambda_-(\phi^k x)\right)} = \prod_{k=-T/2}^{T/2}\exp\left(-\lambda_+(\phi^k x) - \lambda_-(\phi^k x)\right). \tag{7.6.13}$$

Because

$$\exp\left(-\lambda_+(\phi^k x) - \lambda_-(\phi^k x)\right) \tag{7.6.14}$$

is the reciprocal of the volume ratio at time $T/2$ and $-T/2$, (7.6.13) is the volume contraction ratio in time T. Therefore (→(7.6.7)),

$$\exp(-T[\zeta(s) - \zeta(-s)]) = \exp(-Ts) \tag{7.6.15}$$

or

$$\zeta(s) = \zeta(-s) + s \tag{7.6.16}$$

This is the *fluctuation theorem.*

Index