ERGODIC THEORY, ENTROPY AND APPLICATION TO STATISTICAL MECHANICS

TIANYU KONG

Abstract. Ergodic theory originated in statistical mechanics. This paper first introduces the ergodic hypothesis, a fundamental problem in statistical mechanics. In order to come up with a solution, this paper explores some basic ideas in ergodic theory. Next, the paper defines measure-theoretical entropy and shows its connection to physical entropy. Lastly, these results are used to construct Gibbs ensembles, a useful tool in statistical mechanics.

Contents

1. Introduction

Ergodic theory is the study of dynamical systems with an invariant measure, a measure preserved by some function on the measure space. It originated from the proof of the ergodic hypothesis, a fundamental problem in statistical mechanics. A basic example, which illustrates the ergodic hypothesis, is the movement of an ideal gas particle in a box. If the particle satisfies the hypothesis, then over long periods of time, the probability of particle at any position should be the same.

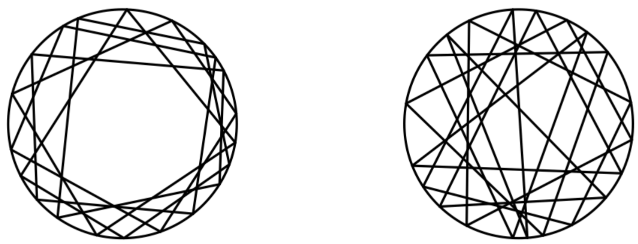

FIGURE 1. Movement of an ideal gas particle. A: Without ergodic hypothesis. B: With ergodic hypothesis

Physicists studying statistical mechanics deal not with physical boxes, but with phase spaces. A phase space is a space in which all possible microscopic states of a system are represented, and each possible state is a point in the phase space. This concept was developed by Boltzmann, Poincaré and Gibbs. They brought up the ergodic hypothesis through their attempts to connect phase space to physical quantities that can be observed and measured, called *observables*.

The remainder of the section provides a rigorous definition of phase spaces, Hamiltonian and Hamiltonian flow. Consider N identical classical particles moving in $\mathbb{R}^d$ with $d \geq 1$, or in a finite subset $\Lambda \subset \mathbb{R}^d$. Suppose that these particles all have mass m and that each is in position q_i and with linear momentum p_i, $1 \leq i \leq N$.

Definition 1.1. The *phase space*, denoted as Γ or Γ_Λ, is the set of all spatial positions and momenta of these particles, so

$$\Gamma = (\mathbb{R}^d \times \mathbb{R}^d)^N, \text{ or } \Gamma_\Lambda = (\Lambda \times \mathbb{R}^d)^N$$

We are interested in the dynamics of the system. Every point $x \in \Gamma$ represents a specific state of the system. The dynamics of the phase space studies how x evolves over time. The notions of potential functions and Hamiltonians are the basic tools needed.

Consider two potential functions $W : \mathbb{R}^d \to \mathbb{R}$ and $V : \mathbb{R}_+ \to \mathbb{R}$. The external potential W comes from outside forces exerted on the system, such as gravity. The pair potential V is the generated by the force one particle exerts on the other, and depends on the distance between these two particles. The potential functions help determine the energy of a system.

Definition 1.2. The *energy of a system of* N *particles* is a function of the positions and momenta of these particles, called the *Hamiltonian*. The Hamiltonian is a function $H : \Gamma \to \mathbb{R}$ given by

$$H(q,p) = \sum_{i=1}^{n} \left(\frac{p_i^2}{2m} + W(q_i) \right) + \sum_{i<j} V(|q_i - q_j|),$$

where $q = (q_1, ..., q_n)$, $p = (p_1, ..., p_n)$, and $n = dN$.

The dynamics in the phase space is governed by the Hamilton's equations of motion.

Definition 1.3. *Hamilton's equations of motion* are

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \text{ and } \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \tag{1.4}$$

Suppose I_n is the $n \times n$ identity matrix, and J denotes the $2n \times 2n$ matrix

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$

Then, (1.4) can be expressed as a vector field $v : \Gamma \to \Gamma$, given by

$$v(x) = J\nabla H(x)$$

where $x = (q, p) \in \mathbb{R}^{2n}$ and

$$\nabla H(x) = \left(\frac{\partial H}{\partial q_1}, ..., \frac{\partial H}{\partial q_N}, \frac{\partial H}{\partial p_1}, ..., \frac{\partial H}{\partial p_N}\right)$$

Let $X(t) \in \Gamma$ be any point in the phase space at given time $t \in \mathbb{R}$. Then, from existence and uniqueness for ODE, for each point $x \in \Gamma$, there exists a unique function $X : \mathbb{R} \to \Gamma$ that satisfies

$$X(0) = x \text{ and } \frac{d}{dt}X(t) = v(X(t)).$$

Definition 1.5. The *Hamiltonian flow* is $\Phi = \{\phi_t | t \in \mathbb{R}\}$, where

$$\phi_t : \Gamma \to \Gamma, \ \phi_t(x) = X(t)$$

for any $t \in \mathbb{R}$.

The Hamiltonian flow can be understood as the evolution of time in phase space. If at $t = 0$, the system has state x in the phase space, then at time $t = T$, the system must have state $\phi_T(x)$. For all $t, s \in \mathbb{R}$, $\phi_t \circ \phi_s = \phi_{t+s}$.

A *microstate* is a point in the phase space, which contains the position and momentum of each particle. A *macrostate*, in comparison to microstate, is determined by measured values called *observables*, including entropy S, internal energy U, temperature T, pressure P, and so on. The observables are bounded continuous functions on the phase space.

Ideally, for an observable f and state $x \in \Gamma$, the measured result is $f(x)$. But, particle movements are small compared to what human can perceive, so making an observation about specific time is not possible. Therefore observations are an average over a relatively long period of time. It is important to know whether taking the average over long time is possible in the phase space and whether the result the same if the initial microstate is different.

In mathematics, the problem becomes: given an observable f, does the limit

$$\tilde{f}(x) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} (f \circ \phi_t)(x)dt$$

exist and is $\tilde{f}(x)$ constant? Section 2 introduces the mathematics needed to answer the question.

2. Ergodic Theory

2.1. **Measure Preserving Maps.** The main focus of ergodic theory is the study of dynamic behavior of measure-preserving maps. Assuming basic knowledge of measure and measure spaces, the paper first explores the probability spaces, and measure-preserving maps in probability spaces.

Definition 2.1. If $(X, \mathscr{B}, \mu)$ is a probability space, $T : (X, \mathscr{B}, \mu) \to (X, \mathscr{B}, \mu)$ from a probability space to itself is *measure-preserving*, or equivalently is an *automorphism*, if $\mu(A) = \mu(T^{-1}(A))$ for every $A \in \mathscr{B}$. μ is called a *T-invariant* measure.

One example of measure-preserving transformation from statistical mechanics is the Hamiltonian flow, defined in Definition 1.5. This fact is the foundation to all discussion of statsitical mechanics in this paper.

Lemma 2.2. *Let Φ be a Hamiltonian flow with Hamiltonian H, then any function $F = f \circ H$ is invariant under ϕ_t for all t. That is $F \circ \phi_t = F$ for all $t \in \mathbb{R}$.*

Proof. $F \circ \phi_t = F$ is equivalent to $H \circ \phi_t = H$. To prove this, we first observe that $H(\phi_0(x)) = H$. Since

$$\frac{d}{dt} H(\phi_t(x)) = H'(\phi_t(x)) \frac{d}{dt}\phi_t(x) = \left(\nabla H(\phi_t(x))\right)^T J \nabla H(\phi_t(X)) = 0,$$

H is ϕ_t-invariant. □

Theorem 2.3. *(Liouville's Theorem) The Jacobian $|\det \phi_t'(x)|$ of a Hamiltonian flow is constant and equals to 1.*

Liouville's Theorem is one of the most important theorems in statistical mechanics. The proof is on page 5 of [3].

Corollary 2.4. *Let μ be a probability measure on $(\Gamma, \mathscr{B}_\Gamma)$. with density $\rho = F \circ H$ with respect to Lebesgue measure, $F : \mathbb{R} \to \mathbb{R}$. Then μ is ϕ_t-invariant for all $t \in \mathbb{R}$.*

Proof. For all $A \in \mathscr{B}_\Gamma$, using change of variable formula,

$$\begin{aligned} \mu(A) &= \int_A \rho(x)dx = \int_\Gamma \chi_A(x)\rho(x)dx = \int_\Gamma \chi_A(\phi_t(x))\rho(\phi_t(x))|\det \phi_t'(x)|dx \\ &= \int_\Gamma \chi_{\phi_t^{-1}(A)}(x)\rho(x)dx = \mu(\phi_t(A)) \end{aligned}$$

□

2.2. **Poincaré's Recurrence Theorem.** In late 1800's, Poincaré proved the first theorem in ergodic theory. He showed that any measure-preserving map has almost everywhere recurrence. This subsection states the theorem and gives one approach to its proof.

Theorem 2.5. *(Poincaré's Recurrence Theorem) Let T be an automorphism of probability space $(X, \mathscr{B}, \mu)$. Given $A \in \mathscr{B}$, let A_0 be the set of points $x \in A$ such that $T^n(x) \in A$ for infinitely many $n \geq 0$, then $A_0 \in \mathscr{B}$ and $\mu(A_0) = \mu(A)$.*

Proof. We first prove $A_0 \in \mathscr{B}$. Let $C_n = \{x \in A | T^j(x) \notin A, \forall j \geq n\}$, and let $A_0 = A \setminus \bigcup_{n=1}^\infty C_n$. Then,

$$C_n = A \setminus \bigcup_{j \geq n} T^{-j}(A) \in \mathscr{B}.$$

Therefore $A_0 \in \mathscr{B}$.

Next, we prove $\mu(A_0) = \mu(A)$. This statement is equivalent to $\mu(C_n) = 0, \forall n \geq 0$. To prove this, notice

$$A \subset \bigcup_{j\geq 0} T^{-j}(A),\ C_n \subset \bigcup_{j\geq 0} T^{-j}(A) \setminus \bigcup_{j\geq n} T^{-j}(A)$$

This implies

$$\mu(C_n) \leq \mu\left(\bigcup_{j\geq 0} T^{-j}(A) \setminus \bigcup_{j\geq n} T^{-j}(A)\right) = \mu\left(\bigcup_{j\geq 0} T^{-j}(A)\right) - \mu\left(\bigcup_{j\geq n} T^{-j}(A)\right)$$

So, since

$$\bigcup_{j\geq n} T^{-j}(A) = T^{-n}\left(\bigcup_{j\geq 0} T^{-j}(A)\right)$$

and since T is measure preserving,

$$\mu\left(\bigcup_{j\geq 0} T^{-j}(A)\right) = \mu\left(\bigcup_{j\geq n} T^{-j}(A)\right)$$

Therefore, $\mu(C_n) = 0$. □

This theorem shows that after sufficiently long time, a certain system will return to a state very close to its initial state. However, Poincaré's theorem is not able to provide important information such as the rate of convergence.

2.3. **Birkhoff's Theorem.** Birkhoff's Theorem originated from the works of statistical physicists Boltzmann and Gibbs introduced in Section 1. The discrete version of the problem can be stated as follows: given a measure-preserving map, T, of a probability space, $(X, \mathscr{B}, \mu)$, and an integrable function $f : X \to \mathbb{R}$, under what conditions does the limit

$$\lim_{n\to\infty} \frac{1}{n} \left(f(x) + f(T(x)) + ... + f(T^n(x))\right) \tag{2.6}$$

exist and is constant almost everywhere. In 1931 Birkhoff proved that the limit above exists almost everywhere for all T and f. This subsection states the Birkhoff's theorem, as well as some of its important implications.

Theorem 2.7. *(Birkhoff's Ergodic Theorem) Let $(X, \mathscr{B}, \mu)$ be a probability space, and let $T : X \to X$ be a measure preserving map. If $f \in L^1(X)$, the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$$

exists for almost every point $x \in X$.

The proof is on page 92 of [1]. Birkhoff's Ergodic Theorem has many useful implications. One of the most important is the following corollary, which builds theoretical background for microscopic dynamics.

Corollary 2.8. *Let $(X, \mathscr{B}, \mu)$ be a probability space, and let $T : X \to X$ be a measure preserving map. If $f \in L^p(X)$, $1 \le p < \infty$, function $\tilde{f}$ defined by*

$$\tilde{f} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \tag{2.9}$$

is in $L^p(X)$, and satisfies

$$\lim_{n\to\infty} \left\| \tilde{f} - \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \right\|_p = 0. \tag{2.10}$$

For almost every $x \in X$,

$$\tilde{f}(T(x)) = \tilde{f}(x). \tag{2.11}$$

Proof. First, we show $\tilde{f} \in L^p(X)$: From (2.9), we know

$$|\tilde{f}(x)| \le \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))| \tag{2.12}$$

This along with the fact that $\tilde{f}$ is non-negative, for any p such that $1 \le p < \infty$ implies

$$|\tilde{f}(x)|^p \le \lim_{n\to\infty} \left(\frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))| \right)^p \tag{2.13}$$

Since the sequence from the right side of the inequality is strictly increasing,

$$\lim_{n\to\infty} \left(\frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))| \right)^p = \liminf_{n\to\infty} \left(\frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))| \right)^p$$

The functions in the sequence are measurable,

$$\liminf_{n\to\infty} \int_X \left(\frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))| \right)^p d\mu < \infty$$

Then, from Fatou's Lemma we get

$$\begin{aligned} \int_X |\tilde{f}(x)|^p d\mu &\le \int_X \liminf_{n\to\infty} \left(\frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))| \right)^p d\mu \\ &\le \liminf_{n\to\infty} \int_X \left(\frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))| \right)^p d\mu \end{aligned}$$

To show $\tilde{f} \in L^p(X)$:

$$\int_X \left(\frac{1}{n} \sum_{i=0}^{n-1} |f(T^i(x))| \right)^p = \left\| \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i \right\|_p^p \le \left(\frac{1}{n} \sum_{i=0}^{n-1} \|f \circ T^i\|_p \right)^p$$

Since T is measure-preserving, $\|f\|_p = \|f \circ T\|_p$,

$$\left(\frac{1}{n} \sum_{i=0}^{n-1} \|f \circ T^i\|_p \right)^p = \left(\frac{1}{n} \sum_{i=0}^{n-1} \|f\|_p \right)^p = \|f\|_p^p < \infty$$

So from (2.13), $\|\tilde{f}\|_p^p \le \|f\|_p^p$, therefore $\tilde{f} \in L^p(X)$.

Next, we show (2.10) holds. Suppose $f \in L^\infty(X)$, that is f is almost everywhere bounded. This is enough to use the dominated convergence theorem. Note $\|f\|_\infty = \inf\{k|$ for almost every $x \in X,\ |f(x)| \leq k\}$. From Theorem 2.7, the sequences,

$$\left|\tilde{f} - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right| \to 0, \quad \left|\tilde{f} - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right|^p \to 0$$

almost everywhere. Then from (2.12) we get,

$$|\tilde{f}(x)| \leq \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |f(T^i(x))| \leq \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \|f\|_\infty = \|f\|_\infty$$

Therefore

$$\left|\tilde{f} - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right| \leq \left|\|f\|_\infty + \frac{1}{n}\sum_{i=0}^{n-1} \|f \circ T^i\|_\infty\right| \leq 2\|f\|_\infty$$

is bounded by a constant. By the dominated convergence theorem:

$$\int_X \left|\tilde{f} - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right|^p d\mu \leq \int_X 0 d\mu = 0$$

So (2.10) holds when $f \in L^\infty$.

When $f \notin L^\infty$, we will approximatie f by a L^∞ function. Since L^∞ is dense in L^p, for any ϵ, choose $g \in L^\infty$ such that $\exists N \in \mathbb{N}$,

$$\|f - g\|_p < \frac{\epsilon}{3}, \quad \left\|\tilde{g} - \frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i\right\|_p < \frac{\epsilon}{3}$$

for all $n \geq N$. By the triangular inequality,

(2.14)
$$\left\|\tilde{f} - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right\|_p \leq \|\tilde{f} - \tilde{g}\|_p + \left\|\tilde{g} - \frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i\right\|_p + \left\|\frac{1}{n}\sum_{i=0}^{n-1} (f-g) \circ T^i\right\|_p$$

From $\|\tilde{f}\|_p \leq \|f\|_p$,

$$\|\tilde{f} - \tilde{g}\|_p = \|(\widetilde{f - g})\|_p \leq \|(f-g)\|_p < \frac{\epsilon}{3}$$

and

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1} (f-g) \circ T^i\right\|_p \leq \frac{1}{n}\sum_{i=0}^{n-1} \|(f-g) \circ T^i\|_p = \frac{1}{n}\sum_{i=0}^{n-1} \|(f-g)\|_p = \|(f-g)\|_p < \frac{\epsilon}{3}$$

This combined with (2.14) shows that

$$\left\|\tilde{f} - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right\|_p < \epsilon$$

Lastly, we need to show $\tilde{f}(T(x)) = \tilde{f}(x)$. To do this, we observe

$$\begin{aligned}\tilde{f}(T(x)) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(T^{i+1}(x)) = \lim_{n\to\infty}\frac{1}{n}\left(\sum_{i=0}^{n} f(T^i(x)) - f(x)\right)\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n} f(T^i(x)) - \lim_{n\to\infty}\frac{f(x)}{n} = \tilde{f}(x)\end{aligned}$$

□

Corollary 2.15. *Let $(X, \mathscr{B}, \mu)$ be a probability space, and let $T : X \to X$ be a measure preserving map. For every $f \in L^p$,*

$$\int_X \tilde{f} d\mu = \int_X f d\mu.$$

Proof. From Corollary 2.8,

$$\int_X \tilde{f} d\mu = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\int_X f\circ T^i d\mu = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\int_X f d\mu = \int_X f d\mu$$

□

Definition 2.16. The function $\tilde{f}$ from Corollary 2.8 is called the *orbital average* of f. When f is the characteristic function χ_A of a set $A \in \mathscr{B}$, $\tilde{\chi}_A(x)$ is called the *average time* x spends in set A and is written as $\tau_A(x)$.

Observe that

$$\tau_A(x) = \lim_{n\to\infty}\frac{1}{n}\left|\{0 \le j \le n-1 | T^j(x) \in A\}\right|$$

$$\int_X \tau_A d\mu = \int_X \chi_A d\mu = \mu(A).$$

2.4. **Ergodicity.** Ergodic transformations are an important subset of measure-preserving transformations. This subsection gives the definition of ergodicity and then introduces some properties of ergodic maps.

Definition 2.17. A set $A \in \mathscr{A}$ is T*-invariant* if $T^{-1}(A) = (A)$. T is *ergodic* if every T-invariant set has measure 0 or 1.

Ergodic functions have some nice properties. For example, if T is ergodic, any T-invariant function is almost everywhere constant. This, along with some other properties, gives a criterion to determine whether a map is ergodic or not.

Proposition 2.18. *The following statements are equivalent:*

(1) T is ergodic.
(2) If for $1 \le p < \infty$, $f \in L^p(X)$ is T-invariant, then f is constant almost everywhere.
(3) For every $A, B \in \mathscr{B}$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mu(T^{-i}(A)\cap B) = \mu(A)\mu(B)$$

(4) For every $f \in L^1(X)$, $\tilde{f} = \int_X f d\mu$ almost everywhere.

Proof.

(1) $\Rightarrow$ (2) Suppose $f \in L^1(X)$ is T-invariant. Then the set $A_n = \{x|f(x) \leq n\}$ is invariant for all $n \in \mathbb{R}$. Since T is ergodic, $\mu(A_n) = 0$ or 1. If f is not constant almost everywhere, then there exists k such that $0 < \mu(A_k) < 1$, which is a contradiction.

(2) $\Rightarrow$ (4) From Corollary 2.8, $\tilde{f} \in L^p(X)$ and is T-invariant. Therefore, $\tilde{f}$ is constant almost everywhere. Then,

$$\int_X f d\mu = \int_X \tilde{f} d\mu = \tilde{f}.$$

(4) $\Rightarrow$ (3) Consider $f = \chi_A$ the characteristic function of A. From Theorem 2.7, for almost every x,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^j(x)) = \tilde{\chi}_A(x) = \int_X \chi_A d\mu = \mu(A)$$

Therefore,

$$\begin{aligned}
\mu(A)\mu(B) &= \mu(A) \int_X \chi_B d\mu = \int_X \tilde{\chi}_A \chi_B d\mu \\
&= \int_X \left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^j(x)) \right) \chi_B d\mu \\
&= \lim_{n\to\infty} \frac{1}{n} \int_X \sum_{i=0}^{n-1} \chi_A(T^j(x)) \chi_B d\mu \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-j}(A) \cap B)
\end{aligned}$$

(3) $\Rightarrow$ (1) Let $A \in \mathscr{B}$ be a T-invariant set. Therefore by (3),

$$\begin{aligned}
\mu(A)\mu(X \setminus A) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap X \setminus A) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(A \cap X \setminus A) = 0
\end{aligned}$$

So $\mu(A) = 0$ or 1, and T is ergodic. □

The proposition above shows if T is ergodic, then the limit from Birkhoff's Ergodic Theorem is constant almost everywhere. This solves the final part of the problem described in Section 2.3. The limit (2.6) is constant almost everywhere if and only if T is ergodic.

T is ergodic if and only if for every $A \in \mathscr{B}$, $\tau_A(x) = \mu(A)$ almost everywhere. It is a direct corollary of Proposition 2.18 by substituting f with χ_A. Intuitively, the time average of an ergodic function is equal to the spacial average of the whole set.

2.5. **Application to Statistical Mechanics.** The results from last two sections are enough to show the limit

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} (f \circ \phi_t)(x) d\mu$$

exists. With the construction of classical dynamical systems, the limit above is the continuous version of

$$\lim_{T\to\infty} \frac{1}{2T} \sum_{t=-T}^{T} (f \circ \phi_t)(x)$$

from Birkhoff's Ergodic Theorem (Theorem 2.7) and Proposition 2.18

Definition 2.19. A *classical dynamical system* $(\Gamma, \mathscr{B}, \mu; \Phi)$ consists of a probability space $(\Gamma, \mathscr{B}, \mu)$ and a group Φ of actions $\phi : \mathbb{R} \times \Gamma \to \Gamma$ defined by $\phi(t, x) = \phi_t(x)$ such that the following statements hold:

(a) $f_T : \mathbb{R} \times \Gamma \to \mathbb{R}$, $(x, t) \mapsto f(\phi_t(x))$ is measurable for any measurable $f : \Gamma \to \mathbb{R}$.
(b) $\phi_t \circ \phi_s = \phi_{t+s}$ for all $t, s \in \mathbb{R}$.
(c) $\mu(\phi_t(A)) = \mu(A)$ for all $t \in \mathbb{R}$ and $A \in \mathscr{B}$.

If a particle moves inside a box with finite volume $\Lambda \subset \mathbb{R}^d$, then the equation of motion (1.4) does not hold on its boundaries. Therefore, it is often necessary to add boundary conditions on $\partial\Lambda$. One of the most common boundary condition is the elastic reflection condition, in which the angle of reflection is equal to the angle of incidence.

From the definitions above and the Birkhoff's Ergodic theorem, we can deduce the following theorem.

Theorem 2.20. *Let $(\Gamma, \mathscr{B}, \mu; \Phi)$ be a classical dynamical system. For every $f \in L^1(\Gamma, \mathscr{B}, \mu)$, consider*

$$\tilde{f}_T(x) = \frac{1}{2T} \int_{-T}^{T} (f \circ \phi_t)(x) dt$$

There exists $A \in \mathscr{B}$ with $\mu(A) = 1$ such that

(a) The limit

$$\tilde{f}(x) = \lim_{T\to\infty} \tilde{f}_T(x)$$

exists for all $x \in A$.

(b) $\tilde{f}(x) = \tilde{f}(\phi_t(x))$ for all $t \in \mathbb{R}$ and $x \in A$.

(c)

$$\int_\Gamma \tilde{f}(x) d\mu = \int_\Gamma f(x) d\mu$$

Suppose that $\Gamma_1 \subset \Gamma$ is invariant under the flow Φ, $\phi_t(\Gamma_1) \subset \Gamma_1$, that $f : \Gamma \to \mathbb{R}$ is an observable, and that the system is always contained in Γ_1. If the measurement of f takes place over a sufficient long period of time, then the observed value $\tilde{f}$ of f is

$$\tilde{f}(x) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} (f \circ \phi_t)(x) dt$$

Another assumption we often make for observables is that the observed value should be independent of the position at $t = 0$, $\tilde{f}(x) = \tilde{f}$ for a constant $\tilde{f}$. This assumption is closely tied to ergodicity introduced in Section 2.3.

Definition 2.21. Let Φ be a flow on Γ_1, and let μ be a probability measure on Γ_1 that is invariant with respect to Φ. Φ is *ergodic* if for every measurable set $F \subset \Gamma_1$ such that $\phi_t(F) = F$ for all $t \in \mathbb{R}$, then $\mu(f) = 0$ or 1.

Theorem 2.22. *Φ is ergodic if and only if all functions $f \in L^2(\Gamma_1, \mu)$ that satisfy $f \circ \phi_t = f$ are constant functions.*

In light of this theorem, if we suppose Φ to be ergodic, then the observable can be expressed as

$$\tilde{f} = \int_{\Gamma_1} \tilde{f}(x) d\mu = \int_{\Gamma_1} f(x) d\mu$$

This is often referred to by saying the time average of observables equals the average over entire the phase space. In practice, proving the ergodicity of Hamiltonian flows turns out to be the most difficult part of applying this theory.

3. Measure-Theoretic Entropy

The concept of entropy was first introduced to solve a fundamental problem in ergodic theory, deciding whether two automorphisms T_1, T_2 of probability spaces $(X_1, \mathscr{B}_1, \mu_1)$ and $(X_2, \mathscr{B}_2, \mu_2)$ respectively are equivalent. The entropy $h(T)$ is a non-negative number that is the same for equivalent automorphisms.

Shannon and Kolmogorov took different paths in defining the entropy of automorphisms. This paper follows Kolmogorov's notion of entropy, which is defined in three stages: first, the entropy of a finite sub-σ-algebra of $\mathscr{B}$, then the entropy of T relative to a finite sub-σ-algebra, and lastly the entropy of T. The structure of this section is based on Chapter 4 of [2].

3.1. **Partitions and Subalgebras.** This subsection defines finite sub-σ-algebras and partitions.

Definition 3.1. A *partition* of $(X, \mathscr{B}, \mu)$ is a disjoint collection of elements of $\mathscr{B}$ whose union is X. Suppose ξ and η are two finite partitions, η is a *refinement* of ξ, or $\xi \leq \eta$, if every element of ξ is a union of elements in η.

Finite partitions are denoted in Greek letters, such as $\xi = \{A_1, A_2, ...A_k\}$. If ξ is a finite partition of $(X, \mathscr{B}, \mu)$, the collection of all elements of $\mathscr{B}$ that are unions of elements of ξ is a finite sub-σ-algebra of $\mathscr{B}$, denoted as $\mathscr{A}(\xi)$. If $\mathscr{C}$ is a finite sub-σ-algebra, $\mathscr{C} = \{C_i | i = 1, 2, ..., n\}$. Then the non-empty sets of form $B_1 \cap B_2 \cap ... \cap B_n$, where $B_i = C_i$ or $B_i = X \setminus C_i$, form a finite partition, denoted by $\xi(\mathscr{C})$.

It is easy to see $\mathscr{A}(\xi(\mathscr{C})) = \mathscr{C}$ and that $\xi(\mathscr{A}(\eta)) = \eta$. This constructs a one-to-one correspondence between partitions and sub-algebras. Also, ξ is a refinement of η if and only if $\mathscr{A}(\xi) \subseteq \mathscr{A}(\eta)$.

Definition 3.2. Let $\xi = \{A_1, A_2, ...A_n\}$, and $\eta = \{C_1, C_2, ..., C_k\}$ be two finite partitions of $(X, \mathscr{B}, \mu)$. Their *join* $\xi \vee \eta$ is the partition

$$\xi \vee \eta = \{A_i \cap C_j | 1 \leq i \leq n, 1 \leq j \leq k\}.$$

For sub-algebras, if $\mathscr{A}$ and $\mathscr{C}$ are finite sub-σ-algebras of $\mathscr{B}$, then $\mathscr{A} \vee \mathscr{C}$ is the smallest sub-σ-algebra of $\mathscr{B}$ containing both $\mathscr{A}$ and $\mathscr{C}$. $\mathscr{A} \vee \mathscr{C}$ consists of all sets which are unions of sets $A \cap C$, $A \in \mathscr{A}$ and $C \in \mathscr{C}$.

A measure preserving $T : X \to X$ also preserves partitions and sub-σ-algebras. Let $\xi = \{A_1, A_2, ...A_n\}$ be a finite partition, then $T^{-n}(\xi)$ denotes the partition $\{T^{-n}(A_1), T^{-n}(A_2), ...T^{-n}(A_n)\}$. Suppose $\mathscr{A}$ is a sub-σ-algebra of $\mathscr{B}$, then $T^{-n}(\mathscr{A})$ denotes sub-σ-algebra $\{T^{-n}(A) | A \in \mathscr{A}\}$.

3.2. **Entropy of Partitions.** In the following sections, the expression $0 \log 0$ is assumed to be 0.

From a probabilistic viewpoint, a partition, $\xi = \{A_1, A_2, ...A_k\}$, of probability space, $(X, \mathscr{B}, \mu)$, is a list of possible outcomes of a random variable. The probability of event A_i happening is $\mu(A_i)$. The entropy of partitions $H(\xi)$ can be used to measure the amount of information required to describe this random variable.

Definition 3.3. Let $\mathscr{A}$ be a finite sub-algebra of $\mathscr{B}$, $\xi(\mathscr{A}) = \{A_1, A_2, ...A_k\}$. The *entropy* of $\mathscr{A}$, or of $\xi(\mathscr{A})$, is

$$H(\mathscr{A}) = H(\xi(\mathscr{A})) = -\sum_{i=1}^{k} \mu(A_i) \log \mu(A_i)$$

$H(\mathscr{A})$ is always positive. Suppose $\mathscr{A}$ is the trivial σ-algebra $\{\emptyset, X\}$, then $H(\mathscr{A})$ $= 0$. If a random variable has one definitive outcome, then no new information is required to describe that variable. In comparison, if a partition has k elements, every element has measure $\frac{1}{k}$, it has the maximum possible entropy of $\log k$.

Similarly, the conditional entropy can be understood as the amount of information needed to describe one random variable given the knowledge of another.

Definition 3.4. Let $\mathscr{A}$ and $\mathscr{C}$ be finite sub-σ-algebras of $\mathscr{B}$, and let $\xi(\mathscr{A}) = \{A_1, A_2, ...A_k\}$, $\xi(\mathscr{C}) = \{C_1, C_2, ...C_p\}$. Then, the *conditional entropy* of $\mathscr{A}$ given $\mathscr{C}$ is

$$\begin{aligned} H(\xi(\mathscr{A})|\xi(\mathscr{C})) = H(\mathscr{A}|\mathscr{C}) &= -\sum_{j=1}^{p} \mu(C_j) \sum_{i=1}^{k} \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \\ &= -\sum_{i,j} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \end{aligned}$$

omitting terms where $\mu(C_j) = 0$.

The convexity of the following auxiliary function is useful in determining properties of conditional entropy.

Lemma 3.5. *The function* $\phi : [0, \infty) \to \mathbb{R}$ *defined by*

$$\phi(x) = \begin{cases} 0, & x = 0 \\ x \log x, & x \neq 0 \end{cases}$$

is strictly convex.

Therefore $\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y)$ for all $x, y \geq 0$, and $t \in (0, 1)$. The equality holds only when $x = y$.

Theorem 3.6. *Let* $(X, \mathscr{B}, \mu)$ *be probability space. Suppose* $\mathscr{A}, \mathscr{C}$*, and* $\mathscr{D}$ *are finite sub-algebras of* $\mathscr{B}$*. Then,*

(a) $H(\mathscr{A} \vee \mathscr{C}|\mathscr{D}) = H(\mathscr{A}|\mathscr{D}) + H(\mathscr{C}|\mathscr{A} \vee \mathscr{D})$

(b) $\mathscr{A} \subseteq \mathscr{C} \Rightarrow H(\mathscr{A}|\mathscr{D}) \leq H(\mathscr{C}|\mathscr{D})$

(c) $\mathscr{C} \subseteq \mathscr{D} \Rightarrow H(\mathscr{A}|\mathscr{C}) \geq H(\mathscr{A}|\mathscr{D})$

Proof. Let $\xi(\mathscr{A}) = \{A_1, ..., A_i\}$, and $\xi(\mathscr{C}) = \{C_1, ..., C_j\}$, $\xi(\mathscr{D}) = \{D_1, ..., D_k\}$. From Definition 3.4, we can assume all elements in these partitions have strictly positive measure.

(a) By definition,

$$H(\mathscr{A} \vee \mathscr{C}|\mathscr{D}) = -\sum_{p,q,r} \mu(A_p \cap C_q \cap D_r) \log \frac{\mu(A_p \cap C_q \cap D_r)}{\mu(D_r)}$$

If $\mu(A_p \cap D_r) = 0$, the left side will be 0, and omitted from total sum. For $\mu(A_p \cap D_r) \neq 0$,

$$\frac{\mu(A_p \cap C_q \cap D_r)}{\mu(D_r)} = \frac{\mu(A_p \cap C_q \cap D_r)}{\mu(A_p \cap D_r)} \frac{\mu(A_p \cap D_r)}{\mu(D_r)}$$

Therefore:

$$\begin{aligned}
&H(\mathscr{A} \vee \mathscr{C}|\mathscr{D}) \\
=& -\sum_{p,q,r} \mu(A_p \cap C_q \cap D_r) \left(\log \frac{\mu(A_p \cap C_q \cap D_r)}{\mu(A_p \cap D_r)} + \log \frac{\mu(A_p \cap C_q \cap D_r)}{\mu(D_r)} \right) \\
=& -\sum_{p,q,r} \mu(A_p \cap C_q \cap D_r) \log \frac{\mu(A_p \cap C_q \cap D_r)}{\mu(D_r)} + H(\mathscr{C}|\mathscr{A} \vee \mathscr{D}) \\
=& -\sum_{p,r} \mu(A_p \cap D_r) \log \frac{\mu(A_p \cap D_r)}{\mu(D_r)} + H(\mathscr{C}|\mathscr{A} \vee \mathscr{D}) \\
=& H(\mathscr{A}|\mathscr{D}) + H(\mathscr{C}|\mathscr{A} \vee \mathscr{D})
\end{aligned}$$

(b) $\mathscr{A} \vee \mathscr{C} = \mathscr{C}$, $H(\mathscr{C}|\mathscr{D}) = H(\mathscr{A} \vee \mathscr{C}|\mathscr{D})$. From (a),

$$H(\mathscr{A} \vee \mathscr{C}|\mathscr{D}) = H(\mathscr{A}|\mathscr{D}) + H(\mathscr{C}|\mathscr{A} \vee \mathscr{D}) \geq H(\mathscr{A}|\mathscr{D}).$$

(c) For fixed p, q, Let

$$t_r = \frac{\mu(D_r \cap C_q)}{\mu(C_q)},\ x_r = \frac{\mu(D_r \cap A_p)}{\mu(D_r)}$$

Notice $\sum_{r=1}^{k} t_r = 1$. Consider the function ϕ defined in Theorem 3.5,

$$\phi(\sum_{r=1}^{k} t_r x_r) \leq \sum_{r=1}^{k} t_r \phi(x_r)$$

that is

$$\phi\left(\sum_{r=1}^{k} \frac{\mu(D_r \cap C_q)}{\mu(C_q)} \frac{\mu(D_r \cap A_p)}{\mu(D_r)}\right) \leq \sum_{r=1}^{k} \frac{\mu(D_r \cap C_q)}{\mu(C_q)} \phi\left(\frac{\mu(D_r \cap A_p)}{\mu(D_r)}\right)$$

Because $\mathscr{C} \subseteq \mathscr{D}$, $C_q = D_{r'}$ for some $r' \in [1, k] \cap \mathbb{N}$. Therefore

$$\sum_{r=1}^{k} \frac{\mu(D_r \cap C_q)}{\mu(C_q)} \frac{\mu(D_r \cap A_p)}{\mu(D_r)} = \frac{\mu(C_q \cap A_p)}{\mu(C_q)}$$

and

$$\frac{\mu(C_q \cap A_p)}{\mu(C_q)} \log \frac{\mu(C_q \cap A_p)}{\mu(C_q)} \leq \sum_{r=1}^{k} \frac{\mu(D_r \cap C_q)}{\mu(C_q)} \phi\left(\frac{\mu(D_r \cap A_p)}{\mu(D_r)}\right)$$

Multipling both sides by $\mu(C_q)$ and summing over p and q gives

$$\begin{aligned}-H(\mathscr{A}|\mathscr{C}) &= \sum_{p,q} \mu(C_q \cap A_p) \log \frac{\mu(C_q \cap A_p)}{\mu(C_q)} \\ &\leq \sum_{p,q,r} \mu(D_r \cap C_q) \frac{\mu(D_r \cap A_p)}{\mu(D_r)} \log \frac{\mu(D_r \cap A_p)}{\mu(D_r)} \\ &= \sum_{p,r} \mu(D_r) \frac{\mu(D_r \cap A_p)}{\mu(D_r)} \log \frac{\mu(D_r \cap A_p)}{\mu(D_r)} \\ &= -H(\mathscr{A}|\mathscr{D})\end{aligned}$$

Therefore, $H(\mathscr{A}|\mathscr{C}) \geq H(\mathscr{A}|\mathscr{D})$.

□

These properties all have intuitive probabilistic meanings. $\mathscr{A} \subseteq \mathscr{C}$ means $\mathscr{A}$ has less entropy than $\mathscr{C}$, and so requires less information to describe. The join $\mathscr{A} \vee \mathscr{C}$ is the combined outcomes of both random variables. It requires the knowledge of both $\mathscr{A}$ and $\mathscr{C}$ given $\mathscr{A}$ to properly describe the entropy of the join.

From the definition, it is also clear that if T is measure-preserving, $H(T^{-1}(\mathscr{A})) = H(\mathscr{A})$ and $H(T^{-1}(\mathscr{A})|T^{-1}(\mathscr{C})) = H(\mathscr{A}|\mathscr{C})$. So measure-preserving maps also preserve entropy and conditional entropy.

Sometimes calculating conditional entropy can be simplified.

Theorem 3.7. *Let $(X, \mathscr{B}, \mu)$ be probability space, and suppose $\mathscr{A}$ and $\mathscr{C}$ are finite sub-algebras of $\mathscr{B}$, then:*

(1) $H(\mathscr{A}|\mathscr{C}) = 0$ if and only if for all $A \in \mathscr{A}$, there exists $C \in \mathscr{C}$ such that $\mu(A \triangle C) = 0$.

(2) $H(\mathscr{A}|\mathscr{C}) = H(\mathscr{A})$ if and only if $\mathscr{A}$ and $\mathscr{C}$ are independent, that is for all $A \in \mathscr{A}$, $C \in \mathscr{C}$, $\mu(A \cap C) = \mu(A)\mu(C)$.

Proof. Let $\xi(\mathscr{A}) = \{A_1, ..., A_i\}$ and $\xi(\mathscr{C}) = \{C_1, ..., C_j\}$. Also, assume all elements of these partitions have strictly positive measure.

(a) Suppose for all $A \in \mathscr{A}$, there exists $C \in \mathscr{C}$ such that $\mu(A \triangle C) = 0$. Then, for any p, q, either $\mu(A_p \cap C_q) = 0$ or $\mu(A_p \cap C_q) = \mu(C_q)$. Therefore by definition $H(\mathscr{A}|\mathscr{C}) = 0$. If $H(\mathscr{A}|\mathscr{C}) = 0$,

$$-\sum_{p,q} \mu(A_p \cap C_q) \log \frac{\mu(A_p \cap C_q)}{\mu(C_q)} = 0$$

because

$$-\mu(A_p \cap C_q) \log \frac{\mu(A_p \cap C_q)}{\mu(C_q)} \geq 0$$

Therefore, $\forall p, q$ such that $1 \leq p \leq i$, $1 \leq q \leq j$,

$$-\mu(A_p \cap C_q) \log \frac{\mu(A_p \cap C_q)}{\mu(C_q)} = 0$$

So, either $\mu(A_p \cap C_q) = 0$ or $\mu(A_p \cap C_q) = \mu(C_q)$.

(b) If $\mathscr{A}$ and $\mathscr{C}$ are independent, from definition $H(\mathscr{A}|\mathscr{C}) = H(\mathscr{A})$. If $H(\mathscr{A}|\mathscr{C}) = H(\mathscr{A})$,

$$-\sum_{p,q} \mu(A_p \cap C_q) \log \frac{\mu(A_p \cap C_q)}{\mu(C_q)} = -\sum_{p} \mu(A_p) \log \mu(A_p)$$

Fix p and consider

$$t_q = \mu(C_q),\ x_q = \frac{\mu(A_p \cap C_q)}{\mu(C_q)}$$

Notice $\sum_{q=1}^{j} t_q = 1$. Let the function ϕ be defined as Theorem 3.5. Then,

$$\sum_{q=1}^{j} \mu(A_p \cap C_q) \log \frac{\mu(A_p \cap C_q)}{\mu(C_q)} = \sum_{q=1}^{j} t_q \phi(x_q) \geq \phi(\sum_{q=1}^{j} t_q x_q) = \mu(A_q) \log \mu(A_q)$$

The equality

$$\sum_{q=1}^{j} t_q \phi(x_q) = \phi(\sum_{q=1}^{j} t_q x_q)$$

holds only when x_q is constant for all q. This means $\mu(A_p \cap C_q) = K_p \mu(C_q)$. Summing over all q gives $\mu(A_p) = K_p$. Therefore, $\mu(A_p \cap C_q) = \mu(A_p)\mu(C_q)$, $\mathscr{A}$ and $\mathscr{C}$ are independent.

□

These statements also have intuitive meaning from a probabilistic viewpoint. If $\mathscr{A}$ and $\mathscr{C}$ are independent, knowing one random variable does not affect the knowledge of the other.

3.3. **Entropy of Measure Preserving Maps.** The second stage of Kolmogorov's definition of entropy is to define the entropy of measure preserving map relative to a finite sub-algebra.

Definition 3.8. Suppose $T : X \to X$ is a measure preserving map of a probability space, $(X, \mathscr{B}, \mu)$, and that $\mathscr{A}$ is a finite sub-σ-algebra of $\mathscr{B}$. Then, the *entropy of* T *given* $\mathscr{A}$ is

$$h(T, \xi(\mathscr{A})) = h(T, \mathscr{A}) = \lim_{n\to\infty} \frac{1}{n} H\left((\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}) \right) \tag{3.9}$$

The *entropy of* T is $h(T) = \sup h(T, \mathscr{A})$, where the supremum is taken over all finite sub-algebras $\mathscr{A}$ of $\mathscr{B}$.

The definition does not imply the existence of limit (3.9). The following theorems and corollary provide a proof based on sequences of real numbers.

Theorem 3.10. *If $\{a_n\}$ is a sequence of real numbers such that for every $m, n \in \mathbb{R}$, $a_{m+n} \leq a_m + a_n$, then*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_n \frac{a_n}{n}$$

Proof. Fix $p > 0$. For all $n > 0$, n can be expressed as $n = kp + i$, $k \in \mathbb{N}$, $0 \leq i < p$. Therefore

$$\frac{a_n}{n} = \frac{a_{kp+i}}{kp+i} \leq \frac{a_i + a_{kp}}{kp} \leq \frac{a_i}{kp} + \frac{ka_p}{kp} = \frac{a_i}{kp} + \frac{a_p}{p}$$

When $n \to \infty$ and $k \to \infty$,

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_p}{p}$$

Taking the infimum over all p gives

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \inf_p \frac{a_p}{p}$$

But,

$$\inf_p \frac{a_p}{p} \leq \liminf_{n\to\infty} \frac{a_n}{n}$$

Therefore,

$$\limsup_{n\to\infty} \frac{a_n}{n} = \liminf_{n\to\infty} \frac{a_n}{n} = \lim_{n\to\infty} \frac{a_n}{n} = \inf_n \frac{a_n}{n}$$

□

Corollary 3.11. *If $T : X \to X$ is a measure preserving map of a probability space, $(X, \mathscr{B}, \mu)$, and $\mathscr{A}$ is a finite sub-σ-algebra of $\mathscr{B}$, then,*

$$\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}))$$

exists.

Proof. Let

$$a_n = H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)$$

Therefore,

$$\begin{aligned}
a_{n+p} &= H\left(\bigvee_{i=0}^{n+p-1} T^{-i}(\mathscr{A})\right) \\
&= H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) + H\left(\bigvee_{i=n}^{n+p-1} T^{-i}(\mathscr{A}) \middle| \bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) \\
&\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) + H\left(\bigvee_{i=n}^{n+p-1} T^{-i}(\mathscr{A})\right) \\
&= a_n + H\left(\bigvee_{i=0}^{p-1} T^{-i}(\mathscr{A})\right) = a_n + a_p
\end{aligned}$$

Then apply Theorem 3.10 to show the corollary is true. □

There are some basic properties of $h(T, \mathscr{A})$.

Theorem 3.12. *Suppose $\mathscr{A}$ and $\mathscr{C}$ are finite sub-algebras of $\mathscr{B}$, and T is a measure preserving map of the probability space $(X, \mathscr{B}, \mu)$, then:*

(a) $h(T, \mathscr{A}) \leq H(\mathscr{A})$
(b) $h(T, \mathscr{A} \vee \mathscr{C}) \leq h(T, \mathscr{A}) + h(T, \mathscr{C})$
(c) $\mathscr{A} \subseteq \mathscr{C} \Rightarrow h(T, \mathscr{A}) \leq h(T, \mathscr{C})$.
(d) $h(T, \mathscr{A}) \leq h(T, \mathscr{C}) + H(\mathscr{A}|\mathscr{C})$
(e) If $k \geq 1$,

$$h(T, \mathscr{A}) = h\left(T, \bigvee_{i=0}^{k-1} T^{-i}(\mathscr{A})\right)$$

Proof. Most of these statements are immediate corollaries of Theorem 3.6, so only a sketch of proof is provided. The complete proof is on Page 89 of [2].

(a)
$$H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) = H\left(\mathscr{A} \vee T^{-1}(\mathscr{A}) \vee ... \vee T^{-(n-1)}(\mathscr{A})\right) \leq \sum_{i=0}^{n-1} H(T^{-i}(\mathscr{A}))$$
(b)
$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A} \vee \mathscr{C})\right) &= H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}) \vee \bigvee_{i=0}^{n-1} T^{-i}(\mathscr{C})\right) \\ &\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) + H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{C})\right) \end{aligned}$$
(c) Because $\mathscr{A} \subseteq \mathscr{C}$, $T^{-i}(\mathscr{A}) \subseteq T^{-i}(\mathscr{C})$. So,
$$H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) \leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{C})\right)$$
(d)
$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) &\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}) \vee \bigvee_{i=0}^{n-1} T^{-i}(\mathscr{C})\right) \\ &= H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{C})\right) + H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}) \middle| \bigvee_{i=0}^{n-1} T^{-i}(\mathscr{C})\right) \end{aligned}$$
$$H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}) \middle| \bigvee_{i=0}^{n-1} T^{-i}(\mathscr{C})\right) \leq \sum_{i=0}^{n-1} H\left(T^{-i}(\mathscr{A}) \middle| T^{-i}(\mathscr{C})\right) = nH\left(\mathscr{A} \middle| \mathscr{C}\right)$$
(e)
$$\begin{aligned} h\left(T, \bigvee_{i=0}^{k-1} T^{-i}(\mathscr{A})\right) &= \lim_{n\to\infty} H\left(\bigvee_{j=0}^{n-1} T^{-j}\left(\bigvee_{i=0}^{k-1} T^{-i}(\mathscr{A})\right)\right) \\ &= \lim_{n\to\infty} H\left(\bigvee_{j=0}^{n+k-1} T^{-j}(\mathscr{A})\right) \\ &= h(T, \mathscr{A}) \end{aligned}$$
□

Theorem 3.13. *Let T be a measure-preserving map of a probability space, $(X, \mathscr{B}, \mu)$.*
(a) For $k > 0$, $h(T^k) = kh(T)$.
(b) If T is invertible, then $h(T^k) = |k|h(T)$, $k \in \mathbb{Z}$.

Proof. This is also direct corollary of Theorem 3.12. A sketch of proof is provided, and the complete proof is on Page 90 of [2]

(a) For any finite sub-σ-algebra $\mathscr{A}$:
$$\begin{aligned} h(T^k, \mathscr{A}) = h\left(T^k, \bigvee_{i=0}^{k-1} T^{-i}(\mathscr{A})\right) &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-kj}\left(\bigvee_{i=0}^{k-1} T^{-i}(\mathscr{A})\right)\right) \\ = k \lim_{n\to\infty} \frac{1}{nk} H\left(\bigvee_{j=0}^{nk-1} T^{-j}(\mathscr{A})\right) &= kh(T, \mathscr{A}) \end{aligned}$$

(b) From (a), the statement is equivalent to $h(T^{-1}) = h(T)$. Since T is measure-preserving, for every finite sub-σ-algebra $\mathscr{A}$,

$$H\left(\bigvee_{i=0}^{n-1} T^i(\mathscr{A})\right) = H\left(T^{n-1}\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)\right) = H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)$$

□

The theorem above provides a convenient way to compute the entropy of some special transformations. Let $T : S^1 \to S^1$ be the rotation of angle $\alpha = \frac{2}{q}\pi$ on a circle, $q \in \mathbb{Z}$. Then, T^q is the identity map, and

$$h(Id) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} \mathscr{A}\right) = 0$$

Therefore a rotation of S^1 also has 0 entropy.

3.4. **Kolmogorov-Sinai Theorem.** At the end of last subsection, we gave a very simple example of an entropy calculation. Yet these calculations can sometimes be difficult. In most cases, it is hard to take the supremum of every finite sub-algebra. The Kolmogorov-Sinai theorem states if the sub-algebra $\mathscr{A}$ satisfies certain conditions, then $h(T) = h(T, \mathscr{A})$. This will make calculation of entropy much easier.

Theorem 3.14. *(Kolmogorov-Sinai Theorem) Let T be an invertible measure-preserving map of a probability space $(X, \mathscr{B}, \mu)$, and let $\mathscr{A}$ be a finite sub-algebra of $\mathscr{B}$ such that $\bigvee_{i=-\infty}^{\infty} T^i(\mathscr{A}) \mathrel{\mathring{=}} \mathscr{B}$. Then, $h(T) = h(T, \mathscr{A})$.*

Some lemmas are required before we can prove this theorem. The lemma below states that two very similar partitions have very small conditional entropy. The same is also true for two similar sub-algebras.

Lemma 3.15. *Fix $r \geq 1$. For every $\epsilon > 0$, there exists $\delta > 0$ such that if $\xi = \{A_1, A_2, ..., A_r\}$ and $\eta = \{C_1, C_2, ..., C_r\}$ are any two partitions of $(X, \mathscr{B}, \mu)$ with $\sum_{i=1}^{r} \mu(A_i \triangle C_i) < \delta$, then $H(\xi|\eta) + H(\eta|\xi) < \epsilon$.*

Theorem 3.16. *Let $(X, \mathscr{B}, \mu)$ be a probability space. Let $\mathscr{B}_0$ be an algebra such that the σ-algebra generated by $\mathscr{B}_0$, $\mathscr{B}(\mathscr{B}_0)$, satisfies $\mathscr{B}(\mathscr{B}_0) \mathrel{\mathring{=}} \mathscr{B}$. Let $\mathscr{C}$ be a finite sub-algebra of $\mathscr{B}$. Then for every $\epsilon > 0$, there exists a finite algebra $\mathscr{D} \subseteq \mathscr{B}_0$ such that $H(\mathscr{C}|\mathscr{D}) + H(\mathscr{D}|\mathscr{C}) < \epsilon$.*

The proofs of Lemma 3.15 and Theorem 3.16 are on page 94 by of [2]. The main idea is to let $\delta < \frac{1}{4}$ and $-r(r-1)\delta \log \delta - (1-\delta)\log(1-\delta) < \frac{\epsilon}{2}$ so that the entropy is sufficiently small.

Finally, given an increasing sequence of finite sub-σ-algebras, Theorem 3.16 can be used to show convergence in conditional entropy. Some notations about sequences of sub-σ-algebras are necessary in the following theorems. Suppose $\{\mathscr{A}_n\}$ is a sequence of sub-σ-algebras of $\mathscr{B}$, then $\bigvee_n \mathscr{A}_n$ denotes the sub-σ-algebra generated by $\{\mathscr{A}_n\}$, that is the intersection of all sub-σ-algebras that contain every $\mathscr{A}_n$. $\bigcup_n \mathscr{A}_n$ is the collection of sets that belong to some $\mathscr{A}_n$. If $\{\mathscr{A}_n\}$ an increasing sequence of sub-σ-algebras, $\bigcup_n \mathscr{A}_n$ is an algebra since it is closed under complements, finite unions and intersections. However, it is not necessarily a σ-algebra, and counter-examples are easy to construct.

Corollary 3.17. *If $\{\mathscr{A}_n\}$ is an increasing sequence of finite sub-algebras of $\mathscr{B}$, and $\mathscr{C}$ is a finite sub-algebra with $\mathscr{C} \subset \bigvee_n \mathscr{A}_n$, then $H(\mathscr{C}|\mathscr{A}_n) \to 0$ as $n \to \infty$.*

Proof. Let $\mathscr{B}_0 = \bigcup_{i=1}^{\infty} \mathscr{A}_i$. Then $\mathscr{B}_0$ is an algebra, and $\mathscr{C} \subset \mathscr{B}(\mathscr{B}_0)$. Let $\epsilon > 0$. By Theorem 3.16, there exists a finite sub-algebra $\mathscr{D}_\epsilon$ of $\mathscr{B}_0$ such that $H(\mathscr{C}|\mathscr{D}_\epsilon) < \epsilon$. Since $\mathscr{D}_\epsilon$ is finite, there exists N such that $\mathscr{D}_\epsilon \subseteq \mathscr{A}_N$. For all $n \geq N$, from Theorem 3.12.c

$$H(\mathscr{C}|\mathscr{A}_n) \leq H(\mathscr{C}|\mathscr{A}_N) \leq H(\mathscr{C}|\mathscr{D}_\epsilon) < \epsilon$$

Since ϵ is arbitrary, $H(\mathscr{C}|\mathscr{A}_n) \to 0$. □

The corollary is the final piece in the proof of Theorem 3.14.

Proof. (Kolmogorov-Sinai) The theorem is equivalent to the statement that for every finite $\mathscr{C} \subseteq \mathscr{B}$, $h(T, \mathscr{C}) \leq h(T, \mathscr{A})$. By Theorem 3.12,

$$\begin{aligned} h(T,\mathscr{C}) &\leq h\left(T, \bigvee_{i=-n}^{n} T^i(\mathscr{A})\right) + H\left(\mathscr{C} \middle| \bigvee_{i=-n}^{n} T^i(\mathscr{A})\right) \\ &= h(T,\mathscr{A}) + H\left(\mathscr{C} \middle| \bigvee_{i=-n}^{n} T^i(\mathscr{A})\right) \end{aligned}$$

Let $\mathscr{A}_n = \bigvee_{i=-n}^{n} T^i(\mathscr{A})$. By Corollary 3.17, $H(\mathscr{C}|\mathscr{A}_n) \to 0$ as $n \to \infty$. Then $h(T, \mathscr{C}) \leq h(T, \mathscr{A})$.

□

For T not necessarily invertible, a similar result to Kolmogorov-Sinai still holds.

Theorem 3.18. *If T is a measure-preserving map of a probability space, $(X, \mathscr{B}, \mu)$, and $\mathscr{A}$ is a finite sub-algebra of $\mathscr{B}$ such that $\bigvee_{i=0}^{\infty} T^{-i}(\mathscr{A}) \stackrel{\circ}{=} \mathscr{B}$, then $h(T) = h(T, \mathscr{A})$.*

3.5. **Example: Entropy of Shifts.** The Bernoulli shift is an important example of a measure-preserving map. Calculation of its entropy is a direct application of the Kolmogorov-Sinai theorem. Understanding the entropy of shifts can be useful in understanding entropy in general. The Bernoulli shift is a discrete stochastic process, each random variable may have k different states ($k \geq 2$), and each state i occurs with probability p_i, $\sum_{i=1}^{k} p_i = 1$.

Definition 3.19. Let $Y = \{1, 2, ..., k\}$ be the state space, and put a measure on Y that gives each point i measure p_i. Let $(Y, 2^Y, \nu)$ denote the measure space. Let $(X, \mathscr{B}, \mu) = \prod_{-\infty}^{\infty} (Y, 2^Y, \nu)$. The σ-algebra $\mathscr{B}$ is the product σ-algebra. The *two-sided shift* is $T : X \to X$, $T(\{x_n\}) = \{y_n\}$ where $y_n = x_{n+1}$ for all $n \in \mathbb{Z}$. This is also denoted as the two-sided $(p_1, ...p_k)$ shift.

Theorem 3.20. *The two-sided $(p_1, ...p_k)$ shifts are measure-preserving and have entropy $-\sum_{i=1}^{k} p_i \log p_i$.*

Proof. It is easy to show shifts are measure preserving. Let $A_i = \{\{x_k\}|x_0 = i\}$, $1 \leq i \leq k$. Then $\xi = \{A_1, ...A_k\}$ is a partition of X. Let $\mathscr{A} = \mathscr{A}(\xi)$ be the sub-σ-algebra generated by ξ. By the definition of $\mathscr{B}$, every element $B \in \mathscr{B}$ can be represented as

$$B = \bigcap_{i=-\infty}^{\infty} T^i(A_{n(i)})$$

where $A_{n(i)} \in \xi$. Therefore,

$$\bigvee_{i=-\infty}^{\infty} T^i(\mathscr{A}) = \mathscr{B}.$$

By Theorem 3.14,

$$h(T) = h(T,\mathscr{A}) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)$$

In the partition $\xi(\mathscr{A} \vee T^{-1}(\mathscr{A}) \vee ... \vee T^{-(n-1)}(\mathscr{A}))$, every element can be represented as

$$A_{i_0} \cap T^{-1}(A_{i_1}) \cap ... \cap T^{-(n-1)} A_{i_n} = \{\{x_n\} | x_0 = i_0, ..., x_{n-1} = i_{n-1}.\}$$

which has measure $p_{i_0} p_{i_1} ... p_{i_{n-1}}$. Therefore

$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) &= - \sum_{i_0,i_1,...,i_{n-1}=1}^{k} (p_{i_0} p_{i_1} ... p_{i_{n-1}}) \log(p_{i_0} p_{i_1} ... p_{i_{n-1}}) \\ &= - \sum_{i_0,i_1,...,i_{n-1}=1}^{k} (p_{i_0} p_{i_1} ... p_{i_{n-1}}) [\log p_{i_0} + \log p_{i_1} + ... + \log p_{i_{n-1}}] \\ &= -n \sum_{i=1}^{k} p_i \log p_i \end{aligned}$$

So,

$$h(T) = h(T,\mathscr{A}) = -\sum_{i=1}^{k} p_i \log p_i$$

□

Consider the two sided shifts $T_1 = (\frac{1}{2}, \frac{1}{2})$-shift and $T_2 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$-shift. By Theorem 3.36, $h(T_1) = \log 2$, and $h(T_2) = \log 3$. Therefore, T_1 and T_2 are not equivalent to each other.

3.6. **Boltzmann's Entropy.** In physics, entropy measures the level of disorder and chaos in a system. Specifically in statistical mechanics, entropy measures the uncertainty that remains in the system after the observable macroscopic properties are observed. Boltzmann came up with the formula for the entropy of a physical system.

$$S = k_B \log W \tag{3.21}$$

where $k_B = 1.38065 \times 10^{-23} Joules/Kelvin$ is the Boltzmann constant, and W is number of microstates corresponding to the macrostate. Therefore, an increase in possible microstates implies increase in uncertainty about the system.

The definition of entropy in statistical mechanics is connected to Kolmogorov's definition of entropy. Recall from Definition 3.3, given a finite partition $\xi = \{A_1, ..., A_N\}$ with probability measure $\mu(A_i) = p_i$, the entropy of the partition is

$$H(\xi) = -\sum_{i=1}^{N} p_i \log p_i$$

Then for a system with a discrete set of microstates, if E_i is the energy of microstate i, and p_i is the probability that it occurs , then the entropy of the system is

$$S = -k_B \sum_{i=1}^{W} p_i \log p_i \tag{3.22}$$

This entropy is called the Gibbs entropy formula. This formula only differs from the entropy of a partition by a constant. It can also be concluded that Kolmogorov's entropy generates an upper bound for the entropy of an arbitrary thermodynamical system.

A fundamental postulate in statistical mechanics states that for an isolated system with an exact macrostate, every microstate that is consistent with the macrostate should be found with equal probability. Therefore if $p_i = W^{-1}$ for all i, then (3.22) becomes

$$S = -k_B \sum_{i=1}^{W} \frac{1}{W} \log \frac{1}{W} = k_B \log W$$

which is exactly (3.21) by Boltzmann.

4. Gibbs Ensembles

Gibbs proposed three Gibbs ensembles in 1902, the microcanonical, the canonical and grandcanonical ensemble. These ensembles can be viewed as probability measures on a subset phase space with a specified density with respect to Lebesgue measure. From Lemma 2.2, all Hamiltonian flows are measure-preserving maps with respect to the ensembles.

The reason to study Gibbs ensembles rather than general phase spaces is that some thermodynamics quantities are conserved for each ensemble. Phase spaces are general and introduce too much complexity in the study the system. The restrictions on conserved quantity can provide some geometric properties that are useful in calculations. The example in Section 4.4 is a perfect example of this.

A *micro-canonical ensemble* is the statistical ensemble that is used to represent the possible states of the system with an exactly specified total energy. The system is isolated from the outer environment, and the energy remains the same. It is also called a NVE ensemble, since the number of particles, N, the volume, V, and the energy, E, are conserved.

A *canonical ensemble* is the statistical ensemble that represents the possible states at a fixed temperature. The system can exchange energy with the outer environment, so the states of the system will differ in total energy, but the temperature stays the same. It is also called a NVT ensemble, since the number of particle, N, the volume, V and the temperature, T are conserved.

A *grand canonical ensemble* is the statistical ensemble that is used to represent the possible states that are open in the sense that the system can exchange energy and particles with the outer environment. This ensemble is therefore sometimes called the μVT ensemble, since the chemical potential, μ, the temperature, T, and volume, V, are conserved.

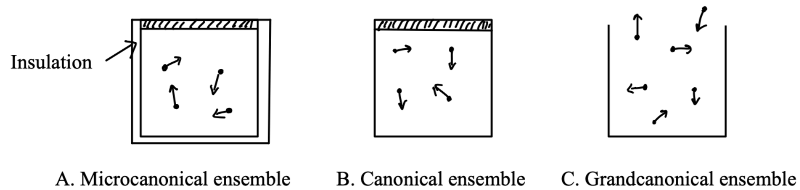

FIGURE 2. Visual representation of three Gibbs ensembles

4.1. **Microcanonical Ensemble.** The energy in the microcanonical ensemble is fixed. For a system with Hamiltonian H and energy fixed at E,

Definition 4.1. For any $E \in \mathbb{R}_+$, the *energy surface* Σ_E for a given Hamiltonian H is defined as

$$\Sigma_E = \{(q,p) \in \Gamma \mid H(q,p) = E\}$$

Σ_E is a hypersurface in the phase space. If H is continuous, then $\Sigma_E = H^{-1}(\{E\})$ is closed. Also, since $H \circ \phi_t = H$, $\phi_t(\Sigma_E) = \Sigma_E$.

Theorem 4.2. *(Riesz-Markov Representation Theorem) For any positive linear functional l, there is a unique Borel measure μ on Γ such that*

$$l(f) = \int_\Gamma f(x) d\mu$$

Let $l_E(f)$ be defined as

$$l_E(f) = \lim_{\delta \to 0} \int_{\Sigma_{[E.E+\delta]}} f(x) dx$$

From the Riesz-Markov theorem, there exists a unique μ'_E such that

$$l_E(f) = \int_\Gamma f(x) d\mu'_E$$

Definition 4.3. If $\omega(E) = \mu'_E(\Gamma) < \infty$, then μ'_E can be normalized as

$$\mu_E = \frac{\mu'_E}{\omega(E)}$$

which is a probability measure on $(\Gamma, \mathscr{B}_\Gamma)$. The probability measure μ_E is called the *microcanonical measure*, or *microcanonical ensemble*. The function $\omega(E)$ is called the *microcanonical partition function*.

The microcanonical measure can also be defined explicitly using curvilinear coordinates. Let $d\sigma_E$ be the surface area element of Σ_E, then

$$d\mu'_E = \frac{d\sigma_E}{\|\nabla H\|}, \text{ and } \omega(E) = \int_{\Sigma_E} \frac{d\sigma_E}{\|\nabla H\|} \tag{4.4}$$

This fact is useful for calculations in Section 4.5. From the definition above, $\omega(E)$ is the number of microstates on the energy surface, so $W = \omega(E)$. From Boltzmann's formula (3.21), the microcanonical entropy is

$$S = k_B \log \omega(E)$$

4.2. **Canonical Ensemble.** In the canonical ensemble, the temperature T is fixed, and so is the inverse temperature $\beta = \frac{1}{k_B T}$.

Definition 4.5. For fixed β, the *canonical Gibbs ensemble* is the probability measure $\gamma^{\beta}_{\Lambda,N}$ with density

$$\rho^{\beta}_{\Lambda,N}(x) = \frac{e^{-\beta H^{(N)}_{\Lambda}(x)}}{Z_{\Lambda}(\beta, N)}, x \in \Gamma_{\Lambda}$$

with respect to the Lebesgue measure. The *partition function* $Z_{\Lambda}(\beta, N)$, or *normalisation*, is defined as

$$Z_{\Lambda}(\beta, N) = \int_{\Gamma_{\Lambda}} e^{-\beta H^{(N)}_{\Lambda}(x)} dx$$

Given fixed values of thermodynamic quantities, the Gibbs ensembles always maximize the Gibbs entropy. This is called the *maximum principle for entropy.* The following theorem is the canonical version of the principle.

Theorem 4.6. *(Maximum Principle for Entropy) Let $\beta > 0$, $\Lambda \subset \mathbb{R}^d$ and $N \in \mathbb{N}$. The canonical Gibbs ensemble $\gamma^{\beta}_{\Lambda,N}$ maximizes the entropy*

$$S(\gamma) = -k_B \int_{\Gamma_{\Lambda}} \rho(x) \log \rho(x) dx$$

for any γ having a Lebesgue density ρ, subject to the constraint

$$U = \int_{\Gamma_{\Lambda}} \rho(x) H^{(N)}_{\Lambda}(x) dx$$

Moreover, the temperature, T, and the partition function are determined by

$$U = -\frac{\partial}{\partial \beta} \log Z_{\Lambda}(\beta, N), \;\; \beta - \frac{1}{k_B T}$$

The proof is on page 22 of[3]. It uses the fact that for $a, b \in (0, \infty)$,

$$a \log a - b \log b \leq (a - b)(1 + \log a)$$

Let $a = \rho^{\beta}_{\Gamma,N}$, the density of $\gamma^{\beta}_{\Gamma,N}$ with respect to Lebesgue measure and let $b = \rho$.

From the theorem above, the entropy of the canonical ensemble can be written as

$$S_{\Lambda}(\beta, N) = k_B \log Z_{\Lambda}(\beta, N) + \frac{U}{T}$$

where the internal energy U is the expectation of H_{Λ},

$$U = \int_{\Gamma_1} H^{(N)}_{\Lambda} \rho^{\beta}_{\Lambda,N}(x) dx$$

4.3. **Grandcanonical Ensemble.** In the grandcanonical ensemble, the number of particles is no longer fixed. From Definition 1.1, the phase space for exactly N particles in box $\Lambda \subset \mathbb{R}^d$ can be written as

$$\Gamma_{\Lambda,N} = \{\omega \subset (\Lambda \times \mathbb{R}^d) | \; \omega = \{(q, p_q) | q \in \hat{\omega}\}, |\hat{\omega}| = N\}$$

where $\hat{\omega}$ is the set of positions occupied by the particles, which is a finite subset of Λ. The momentum of the particle at position q is denoted as p_q.

Definition 4.7. If the number of particles N is not fixed, the *phase space* is

$$\Gamma_\Lambda = \bigcup_{N=1}^{\infty} \Gamma_{\Lambda,N} = \{\omega \subset (\Lambda \times \mathbb{R}^d)|\ \omega = \{(q,p_q)|q \in \hat{\omega}\}, |\hat{\omega}| \text{ finite}\}$$

Definition 4.8. Let $\Lambda \subset \mathbb{R}^d$, $\beta > 0$ and $\mu \in \mathbb{R}$. The *grandcanonical ensemble* for fixed inverse temperature β and chemical potential μ is the probability measure $\gamma_\Lambda^{\beta,\mu}$ of Γ_Λ that satisfies $\gamma_\Lambda^{\beta,\mu}|_{\Gamma_{\Lambda,N}}$ and has density

$$\rho_{\beta,\mu}^{(N)}(x) = \frac{e^{-\beta(H_\Lambda^{(N)}(x)-\mu N)}}{Z_\Lambda(\beta,\mu)}$$

The *partition function* is

$$Z_\Lambda(\beta,\mu) = \sum_{N=0}^{\infty} \int_{\Gamma_{\Lambda,N}} e^{-\beta(H_\Lambda^{(N)}(x)-\mu N)} dx.$$

The observables for the grandcanonical ensemble are a sequence of functions $f = (f_0, f_1, ...)$ where $f_N : \Gamma_{\Lambda,N} \to \mathbb{R}$ are functions on the N-particle phase space, and $f_0 \in \mathbb{R}$. So the expectation in grandcanonical ensemble is

$$E_{\gamma_\Lambda^{\beta,\mu}}(f) = \frac{\sum_{n=0}^{\infty} e^{\beta\mu N} Z_\Lambda(\beta,\mu)}{Z_\Lambda(\beta,\mu)} \int_{\Gamma_{\Lambda,N}} f_N(x)\rho_{\beta,\mu}^{(N)}(dx)$$

Suppose $\mathscr{N}$ is the particle number observable, the expected number of particles in the system is

$$E_{\gamma_\Lambda^{\beta,\mu}}(\mathscr{N}) = \frac{1}{\beta}\frac{\partial}{\partial\mu} \log Z_\Lambda(\beta,\mu)$$

The principle of maximum entropy introduced in canonical ensemble is still valid in grandcanonical ensemble.

Theorem 4.9. *(Principle of Maximum Entropy) Let P be a probability measure on Γ_Λ such that its restriction $P_N = P|_{\Gamma_{\Lambda,N}}$ is absolutely continuous with respect to the Lebesgue measure. That is for any $A \in \mathscr{B}_\Lambda^{(N)}$:*

$$P_N(A) = \int_A \rho_N(x)dx$$

Define the entropy of probability measure P as

$$S(P) = -k_B\rho_0 \log\rho_0 - k_B \sum_{N=1}^{\infty} \int_{\Gamma_{\Lambda,N}} \rho_N(x) \log(N!\rho_N(x))dx$$

Then the grandcanonical measure $\gamma_\Lambda^{\beta,\mu}$, where β and μ are determined by $E_{\gamma_\Lambda^{\beta,\mu}}(H_\Lambda) = E$ and $E_{\gamma_\Lambda^{\beta,\mu}}(\mathscr{N}) = N_0$, maximizes the entropy.

The proof is on page 29 of [3].

Let $U = E_{\gamma_\Lambda^{\beta,\mu}}(H_\Lambda)$ be the internal energy, and let $N_0 = E_{\gamma_\Lambda^{\beta,\mu}}(\mathscr{N})$ be the expected number of particles in the system. The entropy of grandcanonical ensemble is

$$S = k \log Z_\Lambda(\beta,\mu) + \frac{1}{T}(U - \mu N_0)$$

4.4. **Example: Ideal Gas in the Microcanonical Ensemble.** One of the most basic examples in thermodynamics is the ideal gas, an abstraction of gas particles in an insulated box. Using the microcanonical ensemble, it is possible to construct the thermodynamic functions that are consistent with the empirical observations, namely the ideal gas law:

$$PV = Nk_BT \tag{4.10}$$

Consider a gas of N identical particles of mass m in d-dimensions contained in box Λ. The volume of the box is $|\Lambda| = V$. In an ideal setting, the particles are not affected by external forces and do not interact with one another. So, the Hamiltonian from Definition 1.4 can be reduced to

$$H_\Lambda(x) = \sum_{i=1}^{n} \frac{p_i^2}{2m}$$

and the gradient of Hamiltonian for this system is

$$\nabla H_\Lambda(x) = \frac{1}{m}(0, ..., 0, p_1, ..., p_n)$$

Observe that

$$|\nabla H_\Lambda(x)|^2 = \frac{1}{m^2}\sum_{i=1}^{n} p_i^2 = \frac{2}{m}H(x)$$

For all $x \in \Sigma_E$, $H(x) = E$ and

$$|\nabla H_\Lambda| = \sqrt{\frac{2E}{m}}$$

Notice that

$$|(p_1, ..., p_n)| = m|\nabla H_\Lambda(x)| = \sqrt{2mE}$$

Since the norm of p is constant, the energy surface Σ_E can be expressed as

$$\Sigma_E = \Lambda^N \times S_n(\sqrt{2mE})$$

where $S_d(r)$ is the hyper sphere of radius r in dimension d. The surface area of a hypersphere with dimension d is

$$A = c_d r^{d-1}$$

where c_d is constant. Therefore from (4.4),

$$\begin{aligned}\omega(E) &= \int_{\Sigma_E} \frac{d\sigma_E}{|\nabla H_\Lambda|} = \sqrt{\frac{m}{2E}} \int_{\Sigma_E} d\sigma_E = \sqrt{\frac{m}{2E}} V^N c_n(\sqrt{2mE})^{n-1} \\ &= mV^N c_{Nd}(2mE)^{\frac{1}{2}Nd-1}\end{aligned}$$

From Boltzman's entropy formula (3.21),

$$S = k_B \log \omega(E), \ \frac{S}{k_B} = \log\left(mV^N c_{Nd}(2mE)^{\frac{1}{2}Nd-1}\right)$$

Since all energy is conserved in the box, the internal energy is exactly the energy E of the system

$$U = E = \frac{1}{2m}\left(\frac{e^{S/k_B}}{mV^N c_{Nd}}\right)^{2/(Nd-2)}$$

From the thermodynamics identity,

$$dU = TdS - PdV$$

Suppose N is sufficiently large. The temperature of the ideal gas is

$$T = \left(\frac{\partial U}{\partial S}\right)_V = \frac{1}{k_B}\frac{2}{Nd-2}U \approx \frac{2U}{k_B Nd}$$

and the pressure is

$$P = -\left(\frac{\partial U}{\partial V}\right)_S = \frac{2N}{Nd-2}\frac{U}{V} \approx \frac{2U}{dV} \approx \frac{Nk_BT}{V}$$

This relation is exactly the empirical ideal gas law (4.10). This shows that the Gibbs ensembles are useful and accurate for describing the microscopic behavior of particles in a way that is consistent with the macroscopic properties.

Acknowledgments

I would like to thank my mentor Meg Doucette for her help throughout the program. She helped me understand a lot of concepts in a relative short period of time, provided many helpful resources. I am also grateful for her comments and advice on this paper. I would also like to thank Peter May for organizing the UChicago Maths REU. It has been a wonderful experience and a great opportunity to explore my interest in mathematics.

References

[1] Ricado Mañé. Ergodic Theory and Differentiable Dynamics. Springer. 1987.
[2] Peter Walters. An Introduction to Ergodic Theory. Springer. 2000.
[3] Stefan Adams. Lectures on Mathematical Statistical Mechanics. https://warwick.ac.uk/fac/sci/maths/people/staff/stefan_adams/lecturenotestvi/cdias-adams-30.pdf