THERMODYNAMIC FORMALISM

JESSICA METZGER

Contents

1. Introduction

Everything in thermodynamics follows from the fundamental postulate: *For an isolated system in equilibrium, all accessible microstates are equally likely.* A **microstate** is a configuration of the state specified by micro variables like locations of individual molecules. It follows that the degeneracy, and thus the entropy, of a system's state is maximized:

Definition 1.1 (Entropy, the physics version)**.** The **degeneracy** of a system, $\Omega(c_1, \dots, c_n)$ is the number of microstates accessible to a system specified by macroscopic parameters $c_1, \dots, c_n$.

The **entropy** S is defined to be $S = \log \Omega(c_1, \dots, c_n)$.

If we have a system with N possible states, where the nth has probability $p(n)$, the degeneracy for W copies of the system is given by

$$\Omega = \frac{W!}{\prod_{n=1}^{N} (p(n)W)!} \tag{1.2}$$

by combinatorics. Then we can calculate

$$S = -\sum_{n=1}^{N} p(n) \log p(n). \tag{1.3}$$

We can also think of entropy as the expected amount of information that will be contained in our system, where the "information" of an event with probability p is defined as $-\log p$. This is a decreasing function $(0, 1] \to [0, \infty)$ where unlikely

Date: March 27, 2020.

events store more information (and the trivial event carries zero information), which makes sense. Note that entropy is maximized for a probability distribution where $p(n) = 1/N$ for all n (in which case the entropy is $\log N$).

In this paper, we'll take a mathematically formal exploration of the distribution of thermodynamic systems over states. First (section 2), we consider isolated systems with a fixed amount of energy which seek only the maximization of entropy (following the fundamental postulate above). Then (section 3), we look at systems in thermal contact with their environment, which match temperatures with surroundings but can vary in energy. These systems trade off maximizing entropy and minimizing energy penalty by maximizing a variable called the "pressure." Explicitly, these systems take on the "Gibbs distribution" (section 4).

1.1. Notation and Preliminaries.

T is a measure-preserving transformation of a compact metric space X with respect to measure μ on the Borel σ-algebra $\mathcal{B}$.

For an open cover or partition $\mathcal{A}$ of X,

$$\mathcal{A}_n = \mathcal{A} \vee T^{-1}\mathcal{A} \vee \ldots \vee T^{-(n-1)}\mathcal{A}.$$

A cylinder set is written as $C^{j_1,\ldots,j_n}_{i_1,\ldots,i_n}$ where the indices $i_1, \ldots, i_n$ are fixed to symbols $j_1, \ldots, j_n$.

We frequently use the lemma

Lemma 1.4. *Suppose $p_1, \ldots, p_n$ are positive real numbers satisfying $\sum_{i=1}^n p_i = 1$.*

(1) The expression $-\sum_{i=1}^n p_i \log p_i$ has a maximum of $\log n$ which is attained iff all p_i are equal to $1/n$.

(2) If $a_1, \ldots, a_n$ are fixed real numbers, the expression $\sum_{i=1}^n p_i(a_i - \log p_i)$ has a maximum of $\log \sum_{i=1}^n e^{a_i}$ which is attained iff, for all i,

$$p_i = \frac{e^{a_i}}{\sum_{j=1}^n e^{a_j}}. \tag{1.5}$$

These follow from the strict concavity of $-x \log x$.

2. Variational Principle for Entropy

The fundamental postulate of statistical mechanics dictates that the entropy of a system is maximized. In dynamical systems, the distribution over states is given by the measure. So, we are interested in measures that maximize $h_\mu(T)$. (For a review of metric entropy, see Appendix A.)

Recall that the topological entropy for an open cover is $h_{top}(T, \mathcal{G}) = \lim_n \frac{1}{n} \log |\mathcal{G}_n|$. The theoretical maximum entropy of a partition is $\log |\mathcal{A}|$, attained only when it is equidistributed with respect to μ (lem. 1.4). Therefore the theoretical maximum of $h_\mu(T, \mathcal{A})$ over all μ is $\lim_n \frac{1}{n} \log |\mathcal{A}_n|$, which is basically the topological entropy (except that open covers are not partitions; however, they can approximate each other arbitrarily well in Radon spaces). So the variational principle for entropy makes intuitive sense. (We don't prove it here, as we prove a more general version in section 3.)

Theorem 2.1 (Variational Principle for Entropy)**.** *Let T be a homeomorphism of the compact metric space (X, d). Then the topological entropy $h_{top}(T)$ is the supremum of the metric entropies $h_\mu(T)$ over all $\mu \in \mathcal{M}_T(X)$.*

The variational principle has useful applications for determining which distribution (measure) on a system maximizes its entropy. We'll apply it to a common toy physical system, the Ising model.

The Ising model is a large or infinite lattice of particles with spin either up or down, where all particles are a fixed distance apart and the lattice has a fixed dimension d. Let's start with an infinite 1-dimensional lattice, which is isomorphic to the full shift space on two symbols, $\Sigma_2 = \{-1, +1\}^{\mathbb{N}}$, where the symbols +/-1 represent spin up/down and the shift σ "scans" the lattice one unit at a time.

Example 2.2 (Ising model)**.** Subject to no constraints (i.e. no interactions or external magnetic field), the spins distribute themselves randomly. Formally, the measure

$$\mu_{1/2}(C^{j_1,\ldots,j_k}_{i_1,\ldots,i_k}) = \prod_{l=1}^{k} p(j_l) = \prod_{l=1}^{k} \frac{1}{2} = \frac{1}{2^k} \tag{2.3}$$

maximizes entropy. We can say this with certainty since its entropy $h_{\mu_{1/2}}(\sigma, \mathcal{A})$ for the generating partition $\mathcal{A} = \{C^0_1, C^1_1\}$ (the 1-cylinder sets) is

$$h_{\mu_{1/2}}(\sigma, \mathcal{A}) = -\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{2^n} \frac{1}{2^n} \log\left(\frac{1}{2^n}\right) = \log 2. \tag{2.4}$$

By the Kolmogorov-Sinai theorem, $h_{\mu_{1/2}}(\sigma, \mathcal{A}) = h_{\mu_{1/2}}(\sigma)$ since $\mathcal{A}$ generates the sigma-algebra. As shown in class, the system has topological entropy $\log 2$ since it is topologically isomorphic to the expanding map E_2 which has topological entropy $\log 2$. Therefore, $h_{\mu_{1/2}}(\sigma) = h_{top}(\sigma)$. So by the variational principle, maximum metric entropy is attained by the measure $\mu_{1/2}$. However, we can't yet say if this measure is the unique maximizing distribution. This will follow from the results in section 4.

3. Pressure

In physical situations, there tend to be more complex constraints that require taking into account interaction forces, limited energy, or a fixed temperature. So we introduce potential energy via the pressure.

Let $\psi \in \mathcal{C}(X)$ be a function on X called the "potential;" it accounts for the energy of a state. Let $\mathcal{U}$ be a finite open cover and let

$$\begin{gathered} \alpha(C) = \sup\{S_n\psi(x) \mid x \in C\}, \\ \text{where } S_n\psi(x) = \sum_{k=0}^{m} \psi(T^k x). \end{gathered} \tag{3.1}$$

If we let

$$Z_n(\psi, \mathcal{U}) = \inf_{\Gamma} \left\{ \sum_{C\in\Gamma} \exp\left(\alpha(C)\right) \mid \Gamma \subset \mathcal{U}_n \text{ covers } X \right\} \tag{3.2}$$

then we can define the topological pressure:

Definition 3.3 (Topological Pressure)**.** The topological pressure for an open cover $\mathcal{U}$, with respect to a potential ψ, is

$$P_{top}(\psi, \mathcal{U}, T) = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\psi, \mathcal{U}). \tag{3.4}$$

The topological pressure $P(\psi)$ is defined as

$$P_{top}(\psi, T) = \lim_{\text{diam } \mathcal{U}\to 0} P_{top}(\psi, \mathcal{U}, T), \tag{3.5}$$

if these limits exist (they do for well-enough behaved ψ).

Note: this is the same as if we had defined it using spanning/separated sets, just as in topological entropy, i.e.

$$\begin{aligned} Z_n'(\psi, \varepsilon) &= \inf \left\{ \sum_{x\in E} \exp\left(S_n\psi(x)\right) \mid E \text{ is } (n,\varepsilon)\text{-spanning} \right\} \\ P_{top}'(\psi, T) &= \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log Z_n'(\psi, \varepsilon). \end{aligned} \tag{3.6}$$

Then we will find (by a relatively unenlightening computation) that $P'_{top}(\psi, T) = P_{top}(\psi, T)$. We could likewise define $Z''_n(\psi, \varepsilon)$ as the supremum of the same set over all (n, ε)-separated sets, and get the same result. We will use the equivalence of these definitions without proof.

Just like entropy, pressure has both topological and metric forms. Define the metric entropy:

Definition 3.7. For a measure $\mu \in \mathcal{M}_T(X)$ and a potential function ψ, let the metric entropy $P_\mu(\psi, T)$ be

$$P_\mu(\psi, T) = h_\mu(T) + \int_X \psi \, d\mu. \tag{3.8}$$

What is the physical justification for these definitions? Like the metric entropy, the metric pressure is maximized to equal the topological pressure at equilibrium. This actually follows from the fundamental postulate of maximizing entropy, but takes on a different form because our system can exchange energy with its surroundings via ψ. See Appendix B for physical justification.

3.1. **Variational Principle for Pressure.** We'll now prove this variational principle in the language of measure spaces. Assume the topological entropy of the system, $h_{top}(T)$, is finite. Then we have

Theorem 3.9 (Variational principle for pressure)**.** *The topological pressure $P_{top}(\psi, T)$ is the supremum of $h_\mu(T) + \int \psi \, d\mu$ over all $\mu \in \mathcal{M}_T(X)$.*

Proof. First we will show that, for any $\mu \in \mathcal{M}_T(X)$, $h_\mu(T) + \int \psi \, d\mu \le P(\psi)$. Let $\mathcal{P}$ be an arbitrary measurable partition of X with cardinality N. Let $\eta > 0$ be arbitrary and let $\varepsilon > 0$ satisfy $\varepsilon N \log N < \eta$.

Approximate $\mathcal{P} = \{P_0, \dots, P_{N-1}\}$ by compact subsets of each set: for each $P_i \in \mathcal{P}$, let $Q_i \subset P_i$ be such that $\mu(P_i \setminus Q_i) < \varepsilon$. Let $Q_N = X \setminus \cup_i Q_i$, and denote this new, nicer partition by $\mathcal{Q}$. Then since $\mu(Q_N) < N\varepsilon$, $\frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} = \delta_{ij}$ (if $j < N$), and $\sum_i \frac{\mu(P_i \cap Q_N)}{\mu(Q_N)} = 1$, we have

$$\begin{aligned} H_\mu(\mathcal{P}|\mathcal{Q}) &= \sum_j -\mu(Q_j) \sum_i \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} \log \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} \\ &= \sum_i -\mu(Q_N) \frac{\mu(P_i \cap Q_N)}{\mu(Q_N)} \log \frac{\mu(P_i \cap Q_N)}{\mu(Q_N)} \\ &\leq \mu(Q_N) \log N \leq \varepsilon N \log N < \eta. \end{aligned} \tag{3.10}$$

This implies, invoking lemma A.7,

$$h(T, \mathcal{P}) \leq h(T, \mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{Q}) < h(T, \mathcal{Q}) + \eta. \tag{3.11}$$

Basically, $\mathcal{P}$ and $\mathcal{Q}$ are close enough for the entropy of $\mathcal{P}$ to be not much more than $\mathcal{Q}$, even though $\mathcal{Q}$ is basically a refinement. We will use this observation later.

Now, fix n and construct an open cover $\mathcal{G}$ which approximates $\mathcal{Q}$. Choose a δ such that

$$\delta < \frac{1}{2} \min d(Q_i, Q_j) \tag{3.12}$$

and define $\mathcal{G}$ as

$$\mathcal{G} = \{Q_N\} \cup \{G_i = \cup_{x \in Q_i} B_\delta(x) \mid \forall Q_i \in \mathcal{Q},\ i < N\}. \tag{3.13}$$

Then this is very close to $\mathcal{Q}$, except it is an open cover. If we note that

$$H_\mu(\mathcal{Q}_n) + \int_X S_n \psi \, d\mu \leq \sum_{C \in \mathcal{Q}_n} \mu(C)(\alpha(C) - \log \mu(C)) \leq \log \sum_{C \in \mathcal{Q}_n} e^{\alpha(C)} \tag{3.14}$$

then this looks very similar to $\log Z_n(\psi, \mathcal{G})$, except with $\mathcal{Q}$.
Since $C' \subset C$ for any $C' \in \mathcal{Q}_n$ and $C \in \mathcal{G}_n$, we have

$$\sum_{C' \in \mathcal{Q}_n} e^{\alpha(C')} \leq \sum_{C \in \mathcal{G}_n} e^{\alpha(C)}. \tag{3.15}$$

We then can guess that

$$\sum_{C \in \mathcal{G}_n} e^{\alpha(C)} \leq h_{top}(T)^n \inf_\Gamma \sum_{C \in \Gamma} e^{\alpha(C)} = h_{top}(T)^n Z_n(\psi, \mathcal{G}) \tag{3.16}$$

(where the inf is over all Γ that cover X), since the exponential "proliferation rate" of duplicate elements of the refined partitions is given by at most the topological entropy (which we assume to be finite). Putting this all together, we find

$$
\begin{aligned}
\frac{1}{n}H_\mu(\mathcal{Q}_n) + \int_X \psi\,d\mu &= \frac{1}{n}\left(H_\mu(\mathcal{Q}_n) + \int_X S_n\psi\,d\mu\right) \leq \frac{1}{n}\log\sum_{C'\in\mathcal{Q}_n} e^{\alpha(C')} \\
&\leq \frac{1}{n}\log\left[h_{top}(T)^n Z_n(\psi,\mathcal{G})\right] = h_{top}(T) + \frac{1}{n}Z_n(\psi,\mathcal{G}) \\
\implies h_\mu(T,\mathcal{Q}) + \int_X \psi\,d\mu &\leq h_{top}(T) + P_{top}(\psi,\mathcal{G},T).
\end{aligned} \tag{3.17}
$$

Now turn this into our original arbitrary partition. By our previous observation (eq. 3.11),

$$
h_\mu(T,\mathcal{P}) + \int_X \psi\,d\mu \leq \eta + h_\mu(T,\mathcal{Q}) + \int_X \psi\,d\mu \leq \eta + h_{top}(T) + P_{top}(\psi,\mathcal{G},T). \tag{3.18}
$$

Taking the limit as the maximum diameter of $\mathcal{P}$ approaches zero, our choice of δ requires that diam $\mathcal{G} \to 0$; therefore $P_{top}(\psi,\mathcal{G},T) \to P_{top}(\psi,T)$. So, we have

$$
h_\mu(T) + \int_X \psi\,d\mu \leq \eta + h_{top}(T) + P_{top}(\psi,T). \tag{3.19}
$$

However, we could've done all the above calculations using T^M and $S_M\psi$ for any integer M. This gives us

$$
\begin{aligned}
M\left(h_\mu(T) + \int_X \psi\,d\mu\right) &= h_\mu(T^M) + \int_X S_M\psi\,d\mu \\
&\leq \eta + h_{top}(T) + P_{top}(S_M\psi, T^M) \\
&= \eta + h_{top}(T) + MP_{top}(\psi,T).
\end{aligned} \tag{3.20}
$$

Dividing by M and taking $M \to \infty$, we find the desired inequality.

Now show the opposite inequality, $\sup\left\{h_\mu(T,\mathcal{P}) + \int_X \psi\,d\mu\right\} \geq P_{top}(T,\psi)$. Work with the separating set definition of the topological pressure. Fix some $\varepsilon > 0$. We want to exhibit a measure μ such that $h_\mu(T) + \int_X \psi\,d\mu \geq P''_{top}(\psi,\varepsilon,T) = \liminf_{n\to\infty}\frac{1}{n}\log Z''_n(\psi,\varepsilon)$, where

$$
Z''_n(\psi,\varepsilon) = \sup\left\{\sum_{x\in E}\exp\left(S_n\psi(x)\right) \mid E \text{ is } (n,\varepsilon)\text{-separated}\right\}. \tag{3.21}
$$

Since for each n, $Z''_n(\psi,\varepsilon)$ is the sup of $\sum_{x\in E} e^{S_n\psi(x)}$ over (n,ε)-separated sets, we can choose some (n,ε)-separated set $E_n(\varepsilon)$ such that

$$
\log\sum_{x\in E_n(\varepsilon)} e^{S_n\psi(x)} \geq \log Z''_n(\psi,\varepsilon) - 1. \tag{3.22}
$$

Define the following quantities:

$$(3.23)\quad \begin{aligned} \mathcal{Z}_n &= \sum_{x\in E_n(\varepsilon)} e^{S_n\psi(x)} \quad \text{(the partition function for this } n,\varepsilon) \\ \Delta_n &= \frac{1}{\mathcal{Z}_n}\sum_{y\in E_n(\varepsilon)} e^{S_n\psi(y)}\delta_y \in \mathcal{M}(X) \\ \mu_n &= \frac{1}{n}\sum_{k=0}^{n-1} T_*^k\Delta_n \in \mathcal{M}(X) \\ \mu &= \lim_{j\to\infty}\mu_{n_j} \in \mathcal{M}_T(X) \text{ is the weak-* limit of some subsequence } (n_j). \end{aligned}$$

We will show this μ satisfies $h_\mu(T) + \int_X \psi\, d\mu \ge P''_{top}(\psi, \varepsilon, T)$.

Let $\mathcal{P}$ be a reasonable partition with diam $(P) < \varepsilon$ for all $P \in \mathcal{P}$. We can then write

$$(3.24)\qquad H_{\Delta_n}(\mathcal{P}_n) + \int_X S_n\psi\, d\Delta_n = \sum_{y\in E_n(\varepsilon)} \Delta_n(\{y\})(S_n\psi(y) - \log\Delta_n(\{y\})$$

since there can be at most one element of the (n, ε)-separated $E_n(\varepsilon)$ in each member of $\mathcal{P}_n$. Since $\sum_{y\in E_N(\varepsilon)} \Delta_n(\{y\}) = 1$, this expression is maximized by $\Delta_n(\{y\}) = e^{S_n\psi(y)}/\mathcal{Z}_n$ (by the properties of log), which is exactly how we have constructed Δ_n. The maximum, which it attains, is then $\log \sum_{y\in E_n(\varepsilon)} e^{S_n\psi(y)} = \log \mathcal{Z}_n$. So we have

$$(3.25)\qquad H_{\Delta_n}(\mathcal{P}_n) + \int_X S_n\psi\, d\Delta_n = \log\mathcal{Z}_n \ge \log Z''_n(\psi, \varepsilon) - 1$$

because of how we chose $E_n(\varepsilon)$, see eq. 3.22. We now have to manipulate this expression to be in terms of μ_n (the average of the first n push-forwards of the Δ_n). First, note that

$$(3.26)\quad \begin{aligned} \int_X \psi\, d\mu_n &= \frac{1}{n}\sum_{k=0}^{n-1}\int_X \psi\circ T^k\, d\Delta_n = \frac{1}{n}\sum_{k=0}^{n-1}\sum_{y\in E_n(\varepsilon)} \psi\circ T^k(y)\frac{e^{S_n\psi(y)}}{\mathcal{Z}_n} \\ &= \frac{1}{n}\sum_{y\in E_n(\varepsilon)} S_n\psi(y)\frac{e^{S_n\psi(y)}}{\mathcal{Z}_n} = \frac{1}{n}\int_X S_n\psi\, d\Delta_n. \end{aligned}$$

Thus we have

$$(3.27)\qquad H_{\Delta_n}(\mathcal{P}_n) + n\int_X \psi\, d\mu_n \ge \log Z''_n(\psi, \varepsilon) - 1.$$

So we only need to trade $H_{\Delta_n}(\mathcal{P}_n)$ for $H_{\mu_n}(\mathcal{P}_n)$ and show this still holds. For the following steps, follow [2].

Fix $0 < q < n$ and $0 \le j < q$. Define the set U_j to be

$$(3.28)\quad \begin{aligned} U_j = \{j, j+1, \ldots, a_jq + j - 1\}, \quad a_j = \left\lfloor \frac{n-j}{q} \right\rfloor \\ \implies \{0, \ldots, n-1\} = U_j \cup \{0, \ldots, j-1\} \cup \{a_jq + j, \ldots, n-1\} = U_j \cup V_j \end{aligned}$$

where $|V_j| \le 2q$. Then note that we can rewrite $\mathcal{P}_n$ as

$$\mathcal{P}_n = \left(\bigvee_{r=0}^{a_j-1}\bigvee_{i=0}^{q-1} T^{-(rq+j+1)}\mathcal{P}\right) \vee \bigvee_{l\in V_j} T^{-l}\mathcal{P} = \left(\bigvee_{r=0}^{a_j-1} T^{-(rq+j+1)} \bigvee_{i=0}^{q-1}\mathcal{P}\right) \vee \bigvee_{l\in V_j} T^{-l}\mathcal{P} \tag{3.29}$$

and thus, using $H(\alpha\vee\beta) \le H(\alpha) + H(\beta)$ and $H(\mathcal{P}) \le \log|\mathcal{P}| = \log N$, rewrite $H_{\Delta_n}(\mathcal{P}_n)$ as

$$\begin{aligned} H_{\Delta_n}(\mathcal{P}_n) &\le \sum_{r=0}^{q-1} H_{\Delta_n}(T^{-(rq+j)}\bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + H_{\Delta_n}(\bigvee_{l\in V_j} T^{-l}\mathcal{P}) \\ &\le \sum_{r=0}^{a_j-1} H_{T_*^{rq+j}\Delta_n}(\bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + 2q\log N \\ &= \sum_{r=0}^{a_j-1} H_{T_*^{rq+j}\Delta_n}(\mathcal{P}_q) + 2q\log N \end{aligned} \tag{3.30}$$

where we have used $|V_j| \le 2q$. Summing over $j = 0, \dots, q-1$, we find

$$\begin{aligned} qH_{\Delta_n}(\mathcal{P}_n) &\le \sum_{j=0}^{q-1}\sum_{r=0}^{a_j-1} H_{T_*^{rq+j}\Delta_n}(\mathcal{P}_q) + 2q^2\log N \\ &\le n\sum_{k=0}^{n-1}\frac{1}{n} H_{T_*^k\Delta_n}(\mathcal{P}_q) + 2q^2\log N \\ &\le nH_{\mu_n}(\mathcal{P}_q) + 2q^2\log N \end{aligned} \tag{3.31}$$

where we have used the fact that

$$\sum_{k=0}^{n-1}\frac{1}{n} H_{T_*^k\Delta_n}(\mathcal{A}) \le H_{\sum_{k=0}^{n-1}\frac{1}{n}T_*^k\Delta_n}(\mathcal{A}) = H_{\mu_n}(\mathcal{A}) \tag{3.32}$$

(the "superadditivity" of entropy), which follows from the properties of log. Taking the whole inequality, we now have

$$\frac{1}{n}\log Z_n''(\psi,\varepsilon) - \frac{1}{n} \le \frac{1}{n}H_{\Delta_n}(\mathcal{P}_n) + \int_X \psi\, d\mu_n \le \frac{1}{q}H_{\mu_n}(\mathcal{P}_q) + 2\frac{q}{n}\log N + \int_X \psi\, d\mu_n. \tag{3.33}$$

First taking n_j instead of n and then taking the weak limit as $j \to \infty$, we find

$$\begin{aligned} P''_{top}(\psi,\varepsilon) &\le \lim_{j\to\infty}\frac{1}{n_j} Z''_{n_j}(\psi,\varepsilon) \\ &\le \lim_{j\to\infty}\left(\frac{1}{q}H_{\mu_{n_j}}(\mathcal{P}_q) + 2\frac{q}{n_j}\log N + \int_X \psi\, d\mu_{n_j}\right) = \frac{1}{q}H_\mu(\mathcal{P}_q) + \int_X \psi\, d\mu. \end{aligned} \tag{3.34}$$

Then, taking the limit as $q \to \infty$, we find

$$P''_{top}(\psi,\varepsilon) \le h_\mu(T) + \int_X \psi \, d\mu. \tag{3.35}$$

Since such a measure exists for each $\varepsilon > 0$, the topological pressure $P_{top}(\psi, T) = P''_{top}(\psi, T)$ is at most the supremum of $h_\mu(\mathcal{P}_n) + \int_X \psi \, d\mu$ over all $\mu \in \mathcal{M}_\mu(X)$. Thus the variational principle for pressure is true. $\square$

Now if we choose $\psi = 0$, this reduces to the variational principle for entropy. This is the situation where there are no physical constraints on the system, and the probability distribution is determined by the maximization of entropy alone. We can also fix ψ and vary β, which determines the relative sensitivity of the system to the energy constraints. With very low β, the system effectively just maximizes entropy; whereas when $\beta \to \infty$, the system "freezes" into whatever configuration is preferred by the potential function. For these reasons, β can be thought of as the inverse temperature. Actually, that's what it is: $\beta = 1/k_B T$.

Example 3.36 (Ising model with a potential)**.** Consider again the 1d Ising model with entries $x_j = \pm 1$, but this time with a potential $\psi : \Omega \to \mathbb{R}$, defined as

$$\psi(x) = -\beta(-x_0) = \beta x_0. \tag{3.37}$$

So if $x_0 = -1$, $\psi(x) = -\beta$, while if $x_0 = 1$, $\psi(x) = \beta$. This accounts for the local effects of an external magnetic field which prefers spin up, and has no consideration of interactions. Our goal is the find the measure that strikes the best balance between maximizing entropy, and obeying the potential ψ, at a fixed temperature β. We do this using the variational principle for pressure.

First, calculate the topological pressure $P_{top}(\psi, \sigma)$. Use the separated set definition. For any $\varepsilon > 0$ and $n \in \mathbb{N}$, if we let M be such that $2^{-(M+1)} \le \varepsilon < 2^{-M}$ then a set $E_n(\varepsilon)$ containing one point from each $M + n$-cylinder is (n, ε)-separated. In fact, these sets account for all maximal-cardinality (n, ε)-separated sets, and can be used to calculate the nth partition functions Z''_n:

(3.38)

$$\begin{aligned}
Z''_n(\psi,\varepsilon) &= \sup \left\{ \sum_{x \in E_n(\varepsilon)} \exp\left(\sum_{k=0}^{n-1} \psi(\sigma^k x) \right) \mid E_n(\varepsilon) \text{ has one point from each } n+M\text{-cylinder} \right\} \\
&= \sum_{(i_0,\ldots,i_{n+M-1})} \exp\left(\sum_{k=0}^{n-1} \psi(\sigma^k [i_0,\ldots,i_{n+M-1}]) \right) \\
&= \sum_{(i_0,\ldots,i_{n+M-1})} \exp\left(\sum_{k=0}^{n-1} \beta i_k \right) \\
&= 2^M \sum_{(i_0,\ldots,i_{n-1})} \exp\left(\sum_{k=0}^{n-1} \beta i_k \right) \\
&= 2^M \sum_{l=0}^{n} \binom{n}{l} \exp(\beta l - \beta(n-l)) = 2^M \sum_{l=0}^{n} \binom{n}{l} \exp(2l\beta - n\beta)
\end{aligned}$$

where the variable l represents the number of particles within the first n lattice positions with spin up. We can immediately calculate the topological pressure, replacing the limit as $\varepsilon \to 0$ with the limit as $M \to \infty$:

$$\begin{aligned}
P_{top}(\psi,\sigma) &= \lim_{M\to\infty}\lim_{n\to\infty}\frac{1}{n}\log\left[2^M\exp(-n\beta)\sum_{l=0}^{n}\binom{n}{l}\exp(2\beta l)\right] \\
&= \lim_{n\to\infty}\frac{1}{n}\log\left[\sum_{l=0}^{n}\binom{n}{l}\exp(2\beta l)\right]-\beta \\
&= \lim_{n\to\infty}\frac{1}{n}n\log\left[1+e^{2\beta}\right]-\beta \\
&= \log\left(1+e^{2\beta}\right)-\beta.
\end{aligned} \tag{3.39}$$

Now determine the metric pressure $P_\mu(\psi,\sigma)$ for a measure μ, and find one which attains $P_{top}(\psi,\sigma)$. We only calculate it for Bernoulli measures μ_p (with the probability of $x_j=-1$ being p and the probability of $x_j=1$ being $1-p$) and maximize with respect to p (it will turn out that one of these is the optimal measure).

First, calculate metric entropy. By Kolmogorov-Sinai, we only need to do this for the generating partition of the cylinder sets $\mathcal{A}=\{C_0^0,C_0^1\}$. We have the metric entropy for this partition is

$$\begin{aligned}
h_{\mu_p}(T) &= \lim_{n\to\infty}\frac{1}{n}H_{\mu_p}(T,\mathcal{A}) \\
&= -\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{2^n}p(\omega_i(0))\cdot\ldots\cdot p(\omega_i(n))\log\left[p(\omega_i(0))\cdot\ldots\cdot p(\omega_i(n))\right] \\
&= -\lim_{n\to\infty}\frac{1}{n}\left[\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\left[k\log p+(n-k)\log(1-p)\right]\right] \\
&= -\lim_{n\to\infty}\frac{1}{n}\left[\sum_{k=0}^{n}\frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k}\log p+\sum_{k=0}^{n}\frac{n!}{k!(n-k-1)!}p^k(1-p)^{n-k}\log(1-p)\right] \\
&= -\lim_{n\to\infty}\frac{1}{n}\left[np\log p+n(1-p)\log(1-p)\right] \\
&= -\left(p\log p+(1-p)\log(1-p)\right)
\end{aligned} \tag{3.40}$$

Now, calculate the potential energy:

$$\int_\Omega \psi(x)\,d\mu_p=\beta\int_\Omega x_0\,d\mu_p=\beta(-p+(1-p))=\beta(1-2p). \tag{3.41}$$

So, the pressure for a measure μ_p is

$$P_{\mu_p}(\psi,\sigma)=-\left(p\log p+(1-p)\log(1-p)\right)+\beta(1-2p) \tag{3.42}$$

which we can maximize by taking the derivative with respect to p and setting it equal to zero:

$$\begin{aligned}
&\frac{\partial}{\partial p} P(\mu_p) = 0 \\
&\implies 2\beta = \log \frac{1-p}{p} \\
&\implies p = \frac{1}{1+e^{2\beta}}, \quad 1-p = \frac{e^{2\beta}}{1+e^{2\beta}}.
\end{aligned} \tag{3.43}$$

We can confirm that this is indeed the optimal distribution by calculating $P_{\mu_p}(\psi, \sigma)$:

(3.44)

$$\begin{aligned}
P_{\mu_p}(\psi,\sigma) = -&\left[\frac{1}{1+e^{2\beta}} \log\left(\frac{1}{1+e^{2\beta}}\right) + \frac{e^{2\beta}}{1+e^{2\beta}} \log\left(\frac{e^{2\beta}}{1+e^{2\beta}}\right)\right] - \beta\left(1 - \frac{2}{1+e^{2\beta}}\right) \\
&= \log(1+e^{-2\beta}) - \beta = P_{top}(\psi,\sigma) \quad \text{(eq. 3.39)}
\end{aligned}$$

which is the theoretical maximum of P_μ, by the variational principle.

So at $\beta = 0$, the optimal distribution is $p = 1 - p = 1/2$. As $\beta \to \infty$, $p \to 0$ and $1 - p \to 1$. Since β is the inverse temperature, this makes sense: if $\beta = 0$, the temperature is infinite and the distribution only maximizes entropy without regard for potential (this is effectively the unconstrained system we saw in example 2.2). If $\beta = \infty$, the temperature is zero and the distribution freezes into the configuration preferred by the potential function, which prefers spin up (which has probability $1 - p$). Note also that this behavior is very simple: there are no points of non-analyticity of p at any β; i.e. no phase transitions.

4. Gibbs Measures

Usually in thermodynamics we are interested in systems with fixed temperatures, in thermal contact with a large reservoir, since in practice it is difficult to isolate systems from the rest of the world (and thermal contact matches temperatures). In this section, we will give a more precise description of the distribution taken on by a system subject to a potential function ψ, at a fixed inverse temperature β. We will restrict ourselves to subshifts of finite type defined by a periodic, irreducible transition matrix A and develop the theory of **Gibbs measures**.

Definition 4.1 (Gibbs measures)**.** A measure $\mu_\psi \in \mathcal{M}_\sigma(\Sigma_A)$ is a **Gibbs measure** for some potential function $\psi : \Sigma_A \to \mathbb{R}$ if for all n-cylinders C and all $x \in C$,

$$\mu_\psi(C) \asymp \exp\left[S_n\psi(x) - nP\right], \tag{4.2}$$

i.e. there are constants $a_1, a_2 > 0$ such that

$$a_1 \leq \frac{\mu_\psi(C)}{\exp\left[S_n\psi(x) - nP\right]} \leq a_2. \tag{4.3}$$

This is known as the **Gibbs condition**.

We will show that, given a sufficiently well-behaved potential function, this measure both exists and is the unique equilibrium state that maximizes the metric pressure. First, we will show it is an equillibrium state.

Proposition 4.4. *For a Hölder potential function ψ, the corresponding Gibbs measure μ_ψ is an equilibrium state.*

i.e., it achieves the supremum of the pressure over all possible distributions.

Proof. Let $\mathcal{A}$ denote the partition of Σ_A into 1-cylinders, and let $\mathcal{A}_n$ be as in the preceding sections. Define the nth partition function $\mathcal{Z}_n$ to be

$$\begin{aligned} \mathcal{Z}_n &= \sum_{A \in \mathcal{A}_n} \exp\left(\alpha(A)\right), \\ \alpha(A) &= \sup\left\{S_n\psi(x) \mid x \in A\right\}. \end{aligned} \tag{4.5}$$

Suppose $\mu = \mu_\psi$ is a Gibbs measure for ψ. Then there are constants $a_1, a_2 > 0$ such that

$$a_1 \frac{\mathcal{Z}_n}{e^{Pn}} \leq \sum_{A \in \mathcal{A}_n} \mu(A) \leq a_2 \frac{\mathcal{Z}_n}{e^{Pn}}. \tag{4.6}$$

But since $\sum_{A \in \mathcal{A}_n} \mu(A) = 1$, this gives

$$\begin{aligned} &\log a_1 + \log \mathcal{Z}_n \leq Pn \leq \log a_2 + \log \mathcal{Z}_n \\ \implies &\frac{\log a_1 + \log \mathcal{Z}_n}{n} \leq Pn \leq \frac{\log a_2 + \log \mathcal{Z}_n}{n} \\ \implies &P = \lim_{n\to\infty} \frac{1}{n} \log \mathcal{Z}_n. \end{aligned} \tag{4.7}$$

We will show this constant P is equal to (1) $h_{\mu_\psi}(\sigma) + \int_X \psi \, d\mu_\psi$ and (2) $P_{top}(\sigma, \psi)$.

Recall that from lem. 1.4 that $\sum_i p_i(a_i - \log p_i) \leq \log \sum_i e^{a_i}$, which is attained when the p_i follow the Gibbs distribution. If we replace the p_i by the measures of the elements of the partition $\mathcal{A}_n$ and replace the potential $\sum_i p_i a_i$ with $\sum_{A \in \mathcal{A}_n} \mu(A)\alpha(A)$ where $\alpha(A)$ is defined at the beginning of the proof, then we find

$$\begin{aligned} H_{\mu_\psi}(\mathcal{A}_n) + \int_X S_n\psi(x)\, d\mu_\psi &\leq -\sum_{A \in \mathcal{A}_n} \mu(A) \log \mu(A) + \sum_{A \in \mathcal{A}_n} \mu(A)\alpha(A) \\ &\leq \sum_{A \in \mathcal{A}_n} e^{\alpha(A)} = \mathcal{Z}_n. \end{aligned} \tag{4.8}$$

This gives us

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \left[H_{\mu_\psi}(\mathcal{A}_n) + \int_X S_n\psi(x)\, d\mu_\psi \right] &\leq \lim_{n\to\infty} \frac{1}{n} \mathcal{Z}_n \\ \implies h_{\mu_\psi}(\sigma) + \int_X \psi \, d\mu_\psi &\leq P. \end{aligned} \tag{4.9}$$

To show the opposite inequality, we use the fact that ψ is Hölder. Let the corresponding constant and exponent be C, α. Then if we let $\alpha(A)$ be as above and $\omega(A)$ be similarly defined as the inf of $S_n\psi(x)$ over A, we can use this property to find:

$$\alpha(A) - \omega(A) \leq \sum_{k=0}^{n-1} C\alpha^k < \frac{C}{1-\alpha}. \tag{4.10}$$

It follows that for any $x \in A$,

$$S_n\psi(x) \geq \omega(A) \geq \alpha(A) - \frac{C}{1-\alpha}, \tag{4.11}$$

from which it follows that, for any $x \in A$,

(4.12)

$$\begin{aligned} -\mu(A)\log\mu(A) + \int_A S_n\psi(x)\,d\mu &\geq -\mu(A)\log\mu(A) + \mu(A)\left(S_n\psi(x) - \frac{C}{1-\alpha}\right) \\ &\geq -\mu(A)\left[\log\left(a_2\exp\left[S_n\psi(x) - nP\right]\right) - S_n\psi(x) + \frac{C}{1-\alpha}\right] \\ &= \mu(A)\left[-\log a_2 + nP - \frac{C}{1-\alpha}\right]. \end{aligned}$$

Summing over all $A \in \mathcal{A}_n$, we find

$$\begin{aligned} & H_\mu(\mathcal{A}_n) + \int_X S_n\psi(x) \geq -\log a_2 + nP - \frac{C}{1-\alpha} \\ \Longrightarrow & \lim_{n\to\infty}\frac{1}{n}\left[H_\mu(\mathcal{A}_n) + \int_X S_n\psi(x)\right] \geq \lim_{n\to\infty}\frac{1}{n}\left[-\log a_2 + nP - \frac{C}{1-\alpha}\right] \\ \Longrightarrow & h_\mu(\sigma) + \int_X \psi\,d\mu \geq P. \end{aligned} \tag{4.13}$$

So, we have $P = h_\mu(\sigma) + \int_X \psi\,d\mu$.

Lastly, we have to show $P = P_{top}(\sigma, \psi)$. This is fairly apparent, from the spanning set definition of the topological pressure (eq. 3.6). Letting $\varepsilon > 0$ be arbitrary, if we choose $M \in \mathbb{N}$ such that $2^{-(M+1)} \leq \varepsilon < 2^{-M}$, a set consisting of one point from each $M + n$ cylinder is (n, ε)-separated for any $n \in \mathbb{N}$. Thus we can also write $\mathcal{Z}_n$ as

$$\mathcal{Z}_{n+M} = \sup\left\{\sum_{x\in E}\exp\left(S_n\psi(x)\right) \mid E \text{ is } (n,\varepsilon)\text{-separated}\right\}. \tag{4.14}$$

But this is just the $Z'_n(\psi, \varepsilon)$ we introduced in the separating definition of topological entropy, so we have for any ε there is some M such that $\mathcal{Z}_{n+M} = Z'_n(\psi, \varepsilon)$. Taking the limit as $n \to \infty$ we find that the two pressures are equal:

$$P = \lim_{n\to\infty}\frac{1}{n}\log\mathcal{Z}_n = \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log Z'_n(\psi, \varepsilon) = P_{top}(\psi). \tag{4.15}$$

□

Thus the metric pressure of the Gibbs distribution μ_ψ is maximal.

To show the existence, σ-invariance, and ergodicity of the Gibbs measure, we use the functional analysis version of the Perron-Frobenius theorem, applied to the Ruelle-Perron-Frobenius operator.

Definition 4.16 (RPF operator)**.** The **Ruelle-Perron-Frobenius (RPF) operator** $\mathcal{L}_\psi$ for a potential function ψ is defined as

$$\mathcal{L}_\psi f(x) = \sum_{y\in\sigma^{-1}x} e^{\psi(y)} f(y). \tag{4.17}$$

The dual operator $\mathcal{L}_\psi^*$ acts on measures ν by

$$\int f\, d(\mathcal{L}_\psi^* \nu) = \int \mathcal{L}_\psi f\, d\nu. \tag{4.18}$$

We present without proof the operator analog of the Perron-Frobenius theorem:

Theorem 4.19. *If Σ_A has a periodic and irreducible A, then there is a continuous positive $h : \Sigma_A \to \mathbb{R}$, a unique $\lambda > 0$, and a $\nu \in \mathcal{M}_\sigma(\Sigma_A)$ such that*

$$\mathcal{L}_\psi h = \lambda h, \quad \mathcal{L}_\psi^* \nu = \lambda \nu. \tag{4.20}$$

This makes intuitive sense by the positivity of $\mathcal{L}_\psi$ (if $f \geq 0$, then $\mathcal{L}_\psi f \geq 0$). The RPF operator is useful, because its eigenfunction and measure provide a Gibbs measure for ψ, defined by $d\mu = hd\nu$. In the next few propositions, we will show this μ is σ-invariant, satisfies the Gibbs property, is ergodic, and is unique.

Proposition 4.21. *The measure μ defined as $d\mu = hd\nu$ is σ-invariant.*

Proof. Note that for any functions f, g, we have

$$g(x) \cdot \mathcal{L}_\psi f(x) = \sum_{y\in\sigma^{-1}(x)} e^{\psi(y)} f(y) g(\sigma(y)) = \mathcal{L}_\psi (f \cdot g \circ \sigma). \tag{4.22}$$

So we have

$$\begin{aligned} \int f\, d\mu &= \int f \cdot h\, d\nu = \frac{1}{\lambda} \int f \cdot \mathcal{L}_\psi h\, d\nu \\ &= \frac{1}{\lambda} \int \mathcal{L}_\psi (h \cdot f \circ \sigma)\, d\nu = \frac{1}{\lambda} \int h \cdot f \circ \sigma\, d(\mathcal{L}_\psi^* \nu) \\ &= \int f \circ \sigma h\, d\nu = \int f \circ \sigma\, d\mu. \end{aligned} \tag{4.23}$$

So μ is σ-invariant. □

Note that we can scale h so that μ is a probability measure.

Proposition 4.24. *The measure μ satisfies the Gibbs property with $P = \log \lambda$.*

Proof. Note that for any f, $\mathcal{L}_\psi^n f = \sum_{y\in\sigma^{-1}x} e^{S_n\psi(y)} f(y)$. So, we have for any $z \in \Sigma_A$ and any n-cylinder C,

$$\mathcal{L}_\psi^n (h \cdot \mathbf{1}_C) = \sum_{y\in\sigma^{-n}x} e^{S_n\psi(y)} h(y) \mathbf{1}_C(y) \leq e^{\frac{C}{1-\alpha}} (\sup h) e^{S_n\psi(x)} = a_2 e^{S_n\psi(x)} \tag{4.25}$$

since there is at most one $y \in C$ such that $\sigma^n y = z$. So

$$\mu(C) = \int_{\Sigma_A} h \cdot \mathbf{1}_C \, d\nu = \frac{1}{\lambda^n} \int_{\Sigma_A} \mathcal{L}_\psi^n (h \cdot \mathbf{1}_C) \, d\nu \leq \frac{a_2}{\lambda^n} e^{S_n \psi(x)}. \tag{4.26}$$

And by the irreducibility of A, there is some $M \in \mathbb{N}$ and some $y \in C$ such that $\sigma^{n+M}(y) = z$ (the same M for all $z \in \Sigma_A$). So,

(4.27)
$$\mathcal{L}_\psi^n (h \cdot \mathbf{1}_C) = \sum_{y \in \sigma^{-n} x} e^{S_n \psi(y)} h(y) \mathbf{1}_C(y) \geq e^{S'_{n+M} \psi(y)} h(y) \geq e^{-M \sup \psi} e^{\frac{C}{1-\alpha}} (\inf h) e^{S_n \psi(x)} = a_1 e^{S_n \psi(x)}.$$

Thus,

$$\mu(C) = \frac{1}{\lambda^n} \int_{\Sigma_A} \mathcal{L}_\psi^n (h \cdot \mathbf{1}_C) \, d\nu \geq \frac{a_1}{\lambda^n} e^{S_n \psi}. \tag{4.28}$$

So, we find

$$a_1 \leq \frac{\mu(C)}{\lambda^{-n} e^{S_n \psi(x)}} \leq a_2 \tag{4.29}$$

for all $x \in C$, and for all n-cylinders C. This is the Gibbs property for $P = \log \lambda$. □

Proposition 4.30. *σ is mixing with respect to μ.*

Proof. By induction on previous results, for any functions f, g on Σ_A we have

$$g(x) \cdot \mathcal{L}_\psi^n f(x) = \mathcal{L}_\psi^n (f \cdot g \circ \sigma^n). \tag{4.31}$$

Therefore we have for any cylinder sets E, F

$$\begin{aligned} \mu(E \cap \sigma^{-n} F) &= \int \mathbf{1}_E \cdot \mathbf{1}_{\sigma^{-n} F} \, d\mu = \int \mathbf{1}_E \cdot (\mathbf{1}_F) \circ \sigma^n \, d\mu \\ &= \int \mathbf{1}_E \cdot h \cdot (\mathbf{1}_F) \circ \sigma^n \, d\nu = \frac{1}{\lambda^n} \int \mathcal{L}_\psi^n (\mathbf{1}_E \cdot h \cdot (\mathbf{1}_F) \circ \sigma^n) \, d\nu \\ &= \frac{1}{\lambda^n} \int \mathbf{1}_F \cdot \mathcal{L}_\psi^n (\mathbf{1}_E \cdot h) \, d\nu \end{aligned} \tag{4.32}$$

which satisfies

$$\begin{aligned} \left| \mu(E \cap \sigma^{-n} F) - \mu(E)\mu(F) \right| &= \left| \int \left(\mathbf{1}_F \cdot \frac{\mathcal{L}_\psi^n (\mathbf{1}_E \cdot h)}{\lambda^n} - h \mathbf{1}_E \cdot h \mathbf{1}_F \right) d\nu \right| \\ &\leq \nu(F) \left| \int \left(\frac{\mathcal{L}_\psi^n (\mathbf{1}_E \cdot h)}{\lambda^n} - h \mathbf{1}_E \cdot h \right) d\nu \right| \\ &\to 0 \end{aligned} \tag{4.33}$$

by a manipulation of a result from the operator version of the Perron-frobenius theorem. So σ is mixing with respect to μ. □

Proposition 4.34. *The Gibbs measure μ of a potential function ψ is unique.*

Proof. Suppose for some ψ we have two Gibbs measures μ, μ' with pressures P, P' and sets of constants a_1, a_2, a_1', a_2'. Note that eq. 4.15 can be rewritten as

$$P = \lim_{n\to\infty} \frac{1}{n} \log \sum_{A\in\mathcal{A}_n} e^{\alpha(A)} \tag{4.35}$$

where $\alpha(A)$ is the sup of $S_n\psi(x)$ over A. This doesn't depend on μ, so we have $P = P'$. So we can divide the Gibbs conditions of μ, μ' by each other to get

$$\frac{a_1}{a_1'} \leq \frac{\mu(C)}{\mu'(C)} \leq \frac{a_2}{a_2'} \tag{4.36}$$

for any cylinder set C. Or, $b_1\mu'(C) \leq \mu(C) \leq b_2\mu'(C)$. Therefore the two measures are equivalent, and have the same generic and null sets. So applying the ergodicity of the Gibbs measure (prop. 4.30), there is a full-measure set of points G for which, for any $x \in G$ and any continuous function $f : \Sigma_A \to \mathbb{R}$,

$$\int f\,d\mu = \lim_{n\to\infty} \frac{1}{n} S_n f(x) = \int f\,d\mu'. \tag{4.37}$$

So $\mu = \mu'$, and the RPF operator gives us the unique Gibbs measure of the system. □

Example 4.38 (Ising model with a full potential)**.** Consider again the Ising model. However, this time our domain will be finite. We impose cyclic boundary conditions; that is, we identify lattice points N and 0 for some N. We will only be interested in large N. Define the potential function $\psi : \Omega \to \mathbb{R}$ as an extension of the external magnetic field potential (ex. 3.36)

$$\psi(x) = -\beta\left[-Jx_0x_1 - x_0\right] \tag{4.39}$$

where J represents the interaction forces. If $J > 0$ spins prefer to be aligned (ferromagnetism), and if $J < 0$ spins prefer to be mis-aligned. (See [4] for more details.) The optimal measure for this system would be very difficult to calculate using the variational principle alone. However, we can easily verify that a Markov-like Measure satisfies the Gibbs property for this system, and is therefore the unique stationary distribution.

Following [3, 4], define the transfer matrix P to be

$$P = \begin{pmatrix} e^{-\beta(-J+1)} & e^{-\beta(J+1)} \\ e^{-\beta(J-1)} & e^{-\beta(-J-1)} \end{pmatrix} = \begin{pmatrix} p_{-1,-1} & p_{-1,1} \\ p_{1,-1} & p_{1,1} \end{pmatrix} \tag{4.40}$$

where we have let the ijth cell represent the "probability" of a spin i particle being to the left of a spin j particle, just as in a transfer matrix. (We index our matrices and vectors using the symbols ± 1.) Note that this isn't exactly a transfer matrix since the sum of the rows aren't one; for this reason we will have to worry about normalization.

Suppose the lattice is periodic with period N. Find the measure of the state ($\sim$ cylinder set) $C_{0,\ldots,N-1}^{x_0,\ldots,x_{N-1}}$. Conjecture that this is

$$\mu(C_{0,\ldots,N-1}^{x_0,\ldots,x_{N-1}}) = \frac{p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{N-1}x_0}}{\sum_{x_0=\pm1}\cdots\sum_{x_{N-1}=\pm1}\prod_{j=0}^{N-2}p_{x_jx_{j+1}}} = \frac{p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{N-1}x_0}}{Tr(P^N)} \tag{4.41}$$

where Tr is the trace. We need to include the denominator as a normalization so that μ is a probability measure; the reader can verify that with this definition it is true.

Now show μ satisfies the Gibbs condition. We need a constant P such that for all n, all n-cylinders, and all x in such n-cylinders,

$$\begin{aligned}
\mu(C_{0,\ldots,n-1}^{x_0,\ldots,x_{n-1}}) &= \sum_{x_n=\pm1}\cdots\sum_{x_{N-1}=\pm1}\frac{p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{N-1}x_0}}{Tr(P^N)}\\
&= \frac{p_{x_0x_1}\cdots p_{x_{n-2}x_{n-1}}}{Tr(P^N)}\sum_{x_n=\pm1}\cdots\sum_{x_{N-1}=\pm1}p_{x_{n-1}x_n}\cdots p_{x_{N-1}x_0}\\
&\asymp \exp\left(\beta\sum_{k=0}^{n-1}(Jx_kx_{k+1}+x_k)-nP\right) = \frac{p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-1}y_n}}{e^{nP}}, \quad \forall y_n\\
\Longrightarrow \frac{p_{x_{n-1}y_n}}{e^{nP}} &\asymp \frac{1}{Tr(P^N)}\sum_{x_n=\pm1}\cdots\sum_{x_{N-1}=\pm1}p_{x_{n-1}x_n}\cdots p_{x_{N-1}x_0}, \quad \forall y_n
\end{aligned} \tag{4.42}$$

So if we take $P = \frac{1}{N}\log Tr(P^N) = \frac{1}{N}\log(\lambda_+^N + \lambda_-^N)$ where $\lambda_\pm$ are the bigger/smaller eigenvalues, the requirement becomes

$$\begin{aligned}
1 &\asymp \sum_{x_n=\pm1}\cdots\sum_{x_{N-1}=\pm1}\frac{p_{x_{n-1}x_n}\cdots p_{x_{N-1}x_0}}{[Tr(P^N)]^{\frac{N-n}{N}}\,p_{x_{n-1}y_n}}\\
&= \sum_{x_n=\pm1}\cdots\sum_{x_{N-1}=\pm1}\frac{p_{x_{n-1}x_n}\cdots p_{x_{N-1}x_0}}{\lambda_+^{N-n}\left[1+(\lambda_-/\lambda_+)^N\right]^{\frac{N-n}{N}}\,p_{x_{n-1}y_n}}\\
&\approx \sum_{x_n=\pm1}\cdots\sum_{x_{N-1}=\pm1}\frac{p_{x_{n-1}x_n}\cdots p_{x_{N-1}x_0}}{\lambda_+^{N-n}p_{x_{n-1}y_n}}.
\end{aligned} \tag{4.43}$$

The $\left[1+(\lambda_-/\lambda_+)^N\right]^{\frac{N-n}{N}}$ term is negligible because we are only interested in very large N (technically it should be taken into account, but it makes the next part less clear). So we need constants a_1, a_2 such that

$$a_1 \le \sum_{x_n=\pm1}\cdots\sum_{x_{N-1}=\pm1}\frac{p_{x_{n-1}x_n}\cdots p_{x_{N-1}x_0}}{\lambda_+^{N-n}p_{x_{n-1}y_n}} \le a_2 \tag{4.44}$$

for all n and all $x_{n-1},\ldots,x_{N-1},x_0,y_n$. But since we fix N to be finite, there are finite possibilities for the middle quantity, so we can let a_1 be the minimum and a_2 be the maximum. (In fact, since each p_{ij} is of order λ_+, the constants will each be something of order 1 to the Nth power.) So μ satisfies the Gibbs condition, and the topological pressure is $P_\mu(\psi,\sigma) = \frac{1}{N}\log Tr(P^N) \approx \log\lambda_+$. We can explicitly calculate this:

$$P_{top}(\psi, \sigma) \approx \log \lambda_+ = \log \left[e^{\beta J} \cosh \beta + \sqrt{e^{2\beta J} \cosh^2 \beta - 2 \sinh(2\beta J)} \right] \tag{4.45}$$

and can likewise explicitly calculate μ. We calculate the probabilities of different configurations for different values of J in the interactive graph at at `https://www.desmos.com/calculator/8whmo4ydju`. It displays very intuitive behavior; once again, at $\beta = 0$ the configuration is completely random (just maximizing entropy), while as $\beta \to \infty$ the configuration freezes into the all spin-up configuration. The behavior is faster with higher J and with $J = 0$ it reduces to the result from ex. 3.39. Once again, the measures and pressure are all smooth with respect to β, at all values of β and for all N; i.e. there is no phase transition.

It turns out, the 1D Ising model is the only one without a phase transition; in all higher dimensions, there is one. Imagine there is no external magnetic field and the dimension is at least 2. If you start from a very low temperature where the distribution is dominated by low-energy states (e.g. all spin-up) and then slowly raise the temperature, the distribution will increasingly add more weight to states with "droplets" of different spins (e.g. small pockets of spin-down particles). Past some critical temperature, it is impossible to distinguish between a typical state (typical here means full-measure in the $N \to \infty$ limit) being "down with up pockets" or "up with down pockets," i.e. the spins are just random. The calculation is very long but an intuitive explanation is given in [4].

The reason this happens only in higher dimensions is that the energy penalty for a "droplet" scales with its perimeter (in the higher-dimension version of eq. 4.39). It doesn't happen in 1D because the perimeter is the same no matter the size of the droplet; there is a smooth transition in 1D from being "up with down pockets" to being "down with up pockets" because the cost to add a new pocket is always incremental.

Appendix A. Metric Entropy

We define the entropy of a partition of a probability measure space to assume the same form, where "events" are elements of the partition.

Definition A.1 (Entropy of a partition)**.** Let $(X, \mathcal{B}, \mu)$ be a Borel probability measure space, and let $\mathcal{A} = \{A_1, \ldots, A_N\}$ be a finite partition of X. The **entropy** $H_\mu(\mathcal{A})$ of $\mathcal{A}$ is

$$H_\mu(\mathcal{A}) = -\sum_{i=1}^{N} \mu(A_i) \log \mu(A_i). \tag{A.2}$$

Once again, if we define the "information" stored in a set A_i as $-\log \mu(A_i)$, then the entropy represents the expected information contained in an "event" from this partition, as seen by μ.

Consider the entropy $H_\mu(\mathcal{A}_n)$. Adding the extra information of (potentially very complicated) orbits increases the expected amount of information contained in our system. So $H_\mu(\mathcal{A}_n)$ is a way to look at how much information is added to $\mathcal{A}$ by the first n iterations of T.

Definition A.3 (Entropy of a dynamical system)**.** Let $\mathcal{A}$ be a partition of X. Let the metric entropy of T with respect to $\mathcal{A}$, $h_\mu(T,\mathcal{A})$, be

$$h_\mu(T,\mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{A}_n). \tag{A.4}$$

Let the **metric entropy** $h_\mu(T)$ be the supremum of $h_\mu(T,\mathcal{A})$ over all finite partitions $\mathcal{A}$.

By our above interpretation, $h_\mu(T,\mathcal{A})$ is the information added per iterate of T. And, $h_\mu(T)$ is the maximal "rate of information production" T can have.

We can also define conditional entropy of one partition with respect to another:

Definition A.5 (Conditional entropy of a partition)**.** Let $\mathcal{A} = \{A_i\}$, $\mathcal{B} = \{B_j\}$ be partitions of X. The conditional entropy of $\mathcal{A}$ with respect to $\mathcal{B}$ is defined as

$$H_\mu(\mathcal{A}|\mathcal{B}) = -\sum_j \mu(B_j) \sum_i \mu(A_i|B_j) \log \mu(A_i|B_j). \tag{A.6}$$

This lemma will be useful:

Lemma A.7. *Let ξ and η be finite partitions. Then*

$$h_\mu(T,\xi) \le h_\mu(T,\eta) + H_\mu(\xi|\eta). \tag{A.8}$$

We proved this in class, and a proof can be found on p. 214 of [1].

A partition generates the σ-algebra (under iteration by T) if all measurable sets can be expressed arbitrarily well by finite words in the partition and its iteration by T.

Definition A.9 (Generator partition)**.** Let the partition $\mathcal{G}$ be a **generator** of the σ-algebra $\mathcal{B}$ if for all $\varepsilon > 0$ and all $B \in \mathcal{B}$, there are finite $G_1, \ldots, G_k \in \mathcal{G}_n$ such that $\mu\left(B \,\Delta\, (\cup_i G_i)\right) < \varepsilon$.

It turns out, being a generator of the σ-algebra is sufficient to guarantee that a partition causes T to take on the maximal "rate of information production," as seen by μ:

Theorem A.10 (Kolmogorov-Sinai)**.** *Suppose $\mathcal{G}$ is a generator of the σ-algebra. Then $h_\mu(T) = h_\mu(T,\mathcal{G})$.*

We'll skip the proof because we did it in class.

Appendix B. The Free Energy Principle

Here we'll answer the question: what is the physical justification for the maximization of pressure?

Think of our system as depending on the independent variables, like the details of the distribution μ and the inverse temperature β. We can define quantities $f(\mu,\beta,x_1,x_2,\ldots)$ and can fix or free the independent variables as needed. We are interested in determining the behavior of the quantities $f(\ldots)$ under different constraints (usually constant temperature, constant entropy, etc. but it can be anything smooth). One example of a quantity we look at is the internal energy $U(\beta,\mu)$:

$$U(\beta, \mu) = -\frac{1}{\beta} \int_X \psi \, d\mu. \tag{B.1}$$

If likewise the metric entropy is $S(\mu) = h_\mu(T)$, the metric pressure becomes $P(\mu, \beta) = S(\mu) - \beta U(\mu, \beta)$. Thermodynamics dictates that this $P(\mu, \beta)$ is maximized at equilibrium for a system at constant β (constant temperature). We outline the justification for this here.

The maximization of $P(\mu, \beta)$ is evidently equivalent to the minimization of $A(\mu, \beta) = U(\mu, \beta) - S(\mu)/\beta$, the **Helmholtz free energy** of the system. Thermodynamics dictates that, when the temperature is fixed, this quantity is minimized at thermodynamic equilibrium. The justification, like everything else in thermodynamics, comes from our fundamental postulate ("all microstates are equally likely, thus entropy is maximized"). By the fundamental postulate, for any unconstrained parameter x of the system, $\frac{\partial S}{\partial x} = 0$ and $\frac{\partial^2 S}{\partial x^2} < 0$. (Think of x as a variable that perturbs μ near equilibrium.) By a straightforward calculation which we omit, the behavior of the quantity $A(\beta, x) = U(\mu(x), \beta) - S(\mu(x))/\beta$ or $= U(x, \beta) - S(x)/\beta$ at equilibrium is given by

$$\begin{aligned} \left.\frac{\partial A}{\partial x}\right|_\beta &= \ldots = -\frac{\left.\frac{\partial S}{\partial x}\right|_A}{\left.\frac{\partial S}{\partial A}\right|_x} = 0 \\ \left.\frac{\partial^2 A}{\partial x^2}\right|_\beta &= \ldots = -\frac{\left.\frac{\partial^2 S}{\partial x^2}\right|_A}{\left.\frac{\partial S}{\partial A}\right|_x} > 0, \end{aligned} \tag{B.2}$$

so A is minimized. Showing the second expression is > 0 requires knowing the sign of $\partial S/\partial A$, which you can find to be positive by inverting $A(\beta, x)$ and using the real definition of temperature: $\beta = \frac{\partial(\log \Omega)}{\partial U}$ where U is the internal energy of the system and Ω is the number of microstates (this is consistent with all our other uses of β).omit these calculations. So we can conclude that the fixed temperature system finds a distribution μ which maximizes $P(\mu, \beta) = S(\mu) - \beta U(\beta, \mu)$, given the assumptions of thermodynamics.

References

[1] "Introduction to Dynamical Systems" by Brin and Stuck
[2] "Notes on Thermodynamic Formalism" by H. Bruin
[3] "Notes on Thermodynamic Formalism" by Danny Calegari
[4] "Statistical Physics" notes by David Tong