Math 254A Lecture Notes

Professor: Tim Austin
Scribe: Daniel Raban

Contents

1 Counting Type Classes and Introduction to Shannon Entropy

1.1 Counting type classes

Here is a basic setting we will be working with:

- A is a finite alphabet.
- $P(A) = \{p : A \to \mathbb{R} : p(a) \geq 0, \sum_a p(a) = 1\} \subseteq \mathbb{R}^A$ is the set of probability mass functions on A.
- $\|p - q\| = \sum_a |p(a) - q(a)| = 2\sup_{B\subseteq A} |p(B) - q(B)|$ is the total variation between p and q.
- If $x \in A^n$ (for $n \in \mathbb{N}$), then $N(a \mid x) = |\{i = 1, \ldots, n : x_i = a\}|$ is the number of occurrences of a in x.

Definition 1.1. The **empirical distribution** of x is $p_x(a) = \frac{N(a|x)}{n}$.

Definition 1.2. Given $p \in P(A)$, the **type class** of p is $T_n(p) = \{x \in A^n : p_x = p\}$.

How big is $|T_n(p)|$? Here is a basic answer:

$$|T_n(p)| = \begin{cases} \frac{n!}{(np(a_1))!\cdots(np(a_k))!} & np(a) \in \mathbb{N} \ \forall a \in A \\ 0 & \text{otherwise} \end{cases}, \qquad A = \{a_1, \ldots, a_k\}.$$

We are interested in the exponential asymptotic behavior of $|T_n(p)|$. Stirling's approximation tells us that

$$n! = \frac{n^n}{e^n}\sqrt{2\pi n}e^{o(1)}$$

as $n \to \infty$ (where $e^{o(1)} \to 1$ as $n \to \infty$). We will write this more crudely as

$$n! = \frac{n^n}{e^n}e^{o(n)}.$$

Inserting this into the previous expression gives

$$\begin{aligned} |T_n(p)| &= \frac{(n^n/e^n)e^{o(n)}}{\prod_{i=1}^k((np(a_i))^{np(a_i)}/e^{np(a_i)})e^{o(n)}} \\ &= \frac{n^n}{\prod_i (np(a_i))^{np(a_i)}} \Big/ \frac{e^n}{\prod_i e^{np(a_i)}} \\ &= \frac{e^{n\log n}}{\exp(\sum_i np(a_i)\log np(a_i))} \end{aligned}$$

$$= \exp\left(n \log n - \sum_i np(a_i) \log(np(a_i))\right)$$
$$= \exp\left(n \log n - \sum_i np(a_i) \log n - n \sum_i p(a_i) \log p(a_i)\right).$$

In total, we have

$$\begin{aligned} |T_n(p)| &= e^{-n \sum_i p(a_i) \log p(a_i) + o(n)} \\ &= e^{nH(p)+o(n)}, \end{aligned}$$

where $H(p) = -\sum_a p(a) \log p(a)$. This quantity is called the **Shannon entropy** of $p \in P(A)$.

Later on, high-level real analysis will allow us to make sense of redoing the above computation in more complicated variants of this problem, where we are not just looking at the empirical distribution.

Remark 1.1. We regard H as a function $P(A) \to \mathbb{R}$, with the convention that $0 \log 0 = 0$.

1.2 Basic properties of Shannon entropy

Proposition 1.1. *The Shannon entropy H has the following properties:*

(a) H is continuous.

Proof. $x \log x$ is continuous for $x \in (0, 1]$, and $x \log x \to 0$ as $x \to 0$. □

(b) H is strictly concave; i.e. $H(tp + (1 - t)q) \geq tH(p) + (1 - t)H(q)$ with equality only if either $p = q$ or $t \in \{0, 1\}$.

Proof. The function $x \mapsto x \log x$ is strictly concave on $[0, 1]$ (second derivative is < 0). For strictness, if $p \neq q$ and $0 < t < 1$, then there is some a such that $p(a) \neq q(a)$. Then

$$-(tp(a) + (1 - t)q(a)) \log(tp(a) + (1 - t)q(a)) > -tp(a) \log p(a) - (1 - t)q(a) \log q(a).$$

□

(c) $H(p)$ is symmetric under permutations of A.

(d) $0 \leq H(p) \leq \log |A|$. Equality on the left is achieved iff $p = \delta_a$ for some $a \in A$, and equality of the right is achieved iff $p = (1/|A|, \ldots, 1/|A|)$.

Proof. $-x \log x \geq 0$ and is > 0 unless $x = 0, 1$. So $H(p) \geq 0$, and equals 0 only if $p(a) \in \{0, 1\}$ for all a, i.e. only if $p = \delta_b$ for some b. On the other hand, by concavity and symmetry (properties (b) and (c)), H must be maximized at $p = (1/|A|, \dots, 1/|A|)$, and then $H = \log |A|$. $\square$

Example 1.1. Look at the image of H of the simplex $P(\{1, 2, 3\}) = \{(p_1, p_2, p_3) : p_i \geq 0, \sum_i p_i = 1\}$.

Remark 1.2. Suppose X is a random variable taking values in A, and let $p(a) = \mathbb{P}(X = a)$ for $a \in A$. Then $H(X) := H(p)$ is a canonical way to quantify the "uncertainty" in X.

Next time, we will loosen the counting problem to estimate the size of

$$T_{n,\delta}(p) = \{x \in A^n : \|p_x - p\| < \delta\}.$$

2 Counting Empirical Distributions Close to a Given Distribution

2.1 Easier upper bound for the size of a type class

Recall our setting: A is a finite alphabet, and for $x \in A^n$, $p_x(a) = \frac{|\{i \le n : x_i = a\}|}{n}$ is the empirical distribution. The type class is

$$T_n(p) = \{x \in A^n : p_x = p\}.$$

Last time, we used Stirling's approximation to show that $|T_n(p)| = e^{H(p)n+o(n)}$, where $H(p) = -\sum_a p(a) \log p(a)$.

Today we will focus on a variant of the question: counting how many empirical distributions are close to p. We will prove an alternative proof that $|T_n(p)| \le e^{H(p)n}$, the arguments for which will help us in the later analytic case when there is no exact answer.

Proposition 2.1. $|T_n(p)| \le e^{H(p)n}$.

Proof. Choose $X \in A^n$ at random with iid p coordinates, i.e. the law of x is $p^{\times n}$. Given $x \in T_n(p)$, then

$$\begin{aligned}
\mathbb{P}(X = x) &= \prod_{i=1}^n p(x_i) \\
&= \exp\left(\sum_{i=1}^n \log p(x_i)\right) \\
&= \exp\left(\sum_a p_x(a) \cdot n \cdot \log p(a)\right) \\
&= \exp\left(n \sum_a p(a) \log p(a)\right) \\
&= e^{-H(p)n}.
\end{aligned}$$

So

$$1 \ge \mathbb{P}(x \in T_n(p)) = \sum_{x \in T_n(p)} \mathbb{P}(X = x) = |T_n(p)| e^{-H(p)n}. \qquad \square$$

Remark 2.1. It's also true that $|T_n(p)| \ge e^{H(p)n - o(n)}$ if $p(a) \in \mathbb{N}/n$ for all a.

2.2 Asymptotic analysis of number of empirical distributions close to p

Next, we estimate the size of

$$T_{n,\delta}(p) = \{x \in A^n : \|p_x - p\| < \delta\}.$$

Proposition 2.2. *For any $\varepsilon > 0$ and $p \in P(A)$, there is a $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$, we have*

$$e^{H(p)n-\varepsilon n-o(n)} \le |T_{n,\delta}(p)| \le e^{H(p)n+\varepsilon n+o(n)}.$$

Proof. (Upper bound):

$$T_{n,\delta}(p) = \bigcup_{\substack{\|q-p\|<\delta \\ nq(a)\in\mathbb{N}\,\forall a}} T_n(q),$$

so

$$|T_{n,\delta}(p)| \le \sum_q |T_n(q)| \le \sum_q e^{H(q)n}.$$

H is continuous on $\mathbb{P}(A)$, so there exists a δ_0 such that $\|q-p\| < \delta_0 \implies H(q) < H(p)+\varepsilon$, and then

$$\begin{aligned}|T_{n,\delta}(p)| &\le e^{H(p)n+\varepsilon n}|\{q \in P(A) : \|q-p\| < \delta, nq(a) \in \mathbb{N}\,\forall a\}| \\ &\le (n+1)^{|A|}e^{H(p)n+\varepsilon n} \\ &= e^{H(p)n+\varepsilon n+o(n)}.\end{aligned}$$

(Lower bound): If $X \sim p^{\times n}$, so

$$\begin{aligned}\mathbb{P}(X \in T_{n,\delta}(p)) &= \mathbb{P}\left(\sum_a |p_X(a) - p(a)| < \delta\right) \\ &= \mathbb{P}\left(\sum_a \left|\frac{|\{i : X_i = a\}|}{n} - p(a)\right| < \delta\right) \\ &= \mathbb{P}\left(\sum_a \left|\frac{\sum_{i=1}^n \mathbb{1}_{\{X_i=a\}}}{n} - p(a)\right| < \delta\right).\end{aligned}$$

The $\mathbb{1}_{\{X_i=a\}}$ are iid Bernoulli random variables with mean $p(a)$, so by the Weak Law of Large Numbers, this stays $< \delta/|A|$ with high probability as $n \to \infty$. So this probability equals $1 - o(1)$. So we must have

$$\sum_{x\in T_{n,\delta}(p)} \underbrace{\mathbb{P}(X = x)}_{=e^{-n\sum_a p_x(a)\log p(a)}} = 1 - o(1).$$

Observe that for any $\varepsilon > 0$, there exists a δ such that $\|p_x - p\| < \delta \implies \sum_a p_x(a)\log p(a) \le \sum_a p(a)\log p(a) + \varepsilon$. So for this δ, we get

$$|T_{n,\delta}(p)|e^{-H(p)n+\varepsilon n} \ge \mathbb{P}(X \in T_{n,\delta}(p)) = 1 - o(1),$$

and so $|T_{n,\delta}(p)| \ge e^{H(p)n-\varepsilon n-o(n)}$. $\square$

2.3 Superadditivity and convexity arguments for counting type classes of sets

What we've done is specify a ball in the space of empirical distributions and calculated how many distributions end up in the ball. Here is an approach that does not rely on an exact answer. Given $U \subseteq P(A)$, let $T_n(U) = \{x \in A^n : p_x \in U\}$ and $S_n(U) := \log |T_n(U)|$. Here is a key fact.

Proposition 2.3. *If U is convex, then $S_{n+m}(U) \geq S_n(U) + S_m(U)$ for all n, m; i.e. $S_\cdot(U)$ is* ***superadditive****.*

Proof. Suppose $x \in T_n(U)$ and $y \in T_m(U)$. Then

$$p_{(x,y)}(a) = \frac{n}{n+m} p_x(a) + \frac{m}{n+m} p_y(a),$$

so $p_{(x,y)} \in U$ by convexity of U. So $T_n(U) \times T_m(U) \subseteq T_{n+m}(U)$. This gives $|T_n(U)| \cdot |T_m(U)| \leq |T_{n+m}(U)|$. Now take log. $\square$

Lemma 2.1 (Fekete). *Suppose $a_n \in \mathbb{R}$ for all n is superadditive: $a_{n+m} \geq a_n + a_m$. Then*

$$\lim_n \frac{a_n}{n} = \sup_n \frac{a_n}{n} \in (-\infty, \infty].$$

Proof. By iterating this condition, $a_n \geq n a_1$ for all n. Rearrange this to $a_n/n \geq a_1$ for all n. Now suppose that $c < \sup_n a_n/n$. We will show that $a_n/n > c$ for all sufficiently large n. Choose m such that $a_m/m > c$. Now consider $n \gg m$ such that $n = km + p$, where $k \geq 1$ and $0 \leq p < m$. Then $a_n \geq k a_m + a_p$, so

$$\frac{a_n}{n} \geq \frac{k}{km+p} a_m + \frac{p}{km+p} a_1 = \underbrace{\frac{km}{km+p}}_{\xrightarrow{n\to\infty} 1} \underbrace{\frac{a_m}{m}}_{>c} + \underbrace{\frac{p}{km+p}}_{\xrightarrow{n\to\infty} 0} a_1. \qquad \square$$

Corollary 2.1. *If $U \subseteq P(A)$ is convex, then $S(U) := \lim_n \frac{1}{n} S_n(U)$ exists; i.e. $|T_n(U)| = e^{S(U)n + o(n)}$.*

Next, we will derive properties of S.

Lemma 2.2. *If $U \subseteq V$, then $S(U) \leq S(V)$.*

Here is somewhat of an improvement:

Lemma 2.3. *If $U \subseteq U_1 \cup \cdots \cup U_k$, then $S(U) \leq \max_i S(U_i)$.*

Proof.
$$|T_n(U)| \leq \sum_i |T_n(U_i)| \leq k \cdot \max_i |T_n(U_i)|,$$
so
$$\frac{1}{n} S_n(U) \leq \frac{\log k}{n} + \max_i \frac{1}{n} S_n(U_i).$$
Now let $n \to \infty$. $\square$

How can a function of convex sets U be like this?

Example 2.1. Let $\widetilde{S} : P(A) \to \mathbb{R}$ be continuous, and let $S(U) = \sup\{\widetilde{S}(p) : p \in U\}$. This example will have the property in the above lemma.

Next time, we will give conditions on S for it to have this form. When we come to the analytic case, we will be able to repeat this analysis without needing to know the exact value of S.

3 Turning Set Functions Into Point Functions

3.1 Recap + dealing with the empty set

Last time, we had a finite alphabet A, and given $U \subseteq P(A)$, we looked at $T_n(U) = \{x \in A^n : p_x \in U\}$. We looked at the asymptotic behavior of the size of this set without relying on explicit formulae. We defined $S_n(U) = \log |T_n(U)|$.

What if $T_n(U) = \varnothing$? Here are two answers.

1. If $U \neq \varnothing$ is open, if we pick $p \in U$, let n be very large and pick $X \sim p^{\times n}$. Then $\mathbb{P}(p_X \in U) \to 1$ as $n \to \infty$ by the Weak Law of Large Numbers. So $T_n(U) \neq \varnothing$ for all sufficiently large n.

2. We should let S_n take the value $-\infty$. This will be fine, as long as we're not subtracting negative infinities or multiplying. This is the better answer

Last time, we showed that $S_n(U)$ is superadditive if U is convex:

$$S_{n+m}(U) \geq S_n(U) + S_m(U).$$

By Fekete's lemma, $S(U) = \lim_n \frac{1}{n} S_n(U)$ exists and equals $\sup_n \frac{1}{n} S_n(U)$. This tells us that

$$|T_n(U)| = e^{S(U)n + o(n)}.$$

This produces a set function $S : \{\text{convex open subsets of } P(A)\} \to [-\infty, \infty]$. We would like a point function $S : P(A) \to [-\infty, \infty]$ such that $s(U) = \sup\{S(p) : p \in U\}$.

3.2 General considerations: when do set functions give rise to point functions?

We will step away for a while to a more abstract setting: Let X be a topological space, let $\mathcal{U}$ be a cover of X by open sets, and let $S : \mathcal{U} \to [-\infty, \infty]$. When is there a point function $S : X \to [-\infty, \infty]$ such that $S(U) = \sup\{S(x) : x \in U\}$?

The first necessary condition is

(S1) If $U, U_1, \dots, U_k \in \mathcal{U}$ and $U \subseteq U_1 \cup \cdots \cup U_k$, then $S(U) \leq \max_i S(U_i)$.

Unfortunately, this condition is not sufficient, but we will give a sufficient condition later.

Aside: Call S **locally finite** if for every $x \in X$, there is some $U \in \mathcal{U}$ such that $x \in U$ and $S(U) < \infty$.

Now let's define $S(x) := \inf\{S(U) : U \in \mathcal{U}, U \ni x\}$. Then the following is true.

Lemma 3.1.

$$S(U) \geq \sup\{S(x) : x \in U\}.$$

Lemma 3.2. *The point function S must be upper semicontinuous.*

Proof. If $S(x) < a$, then there exists some $U \in \mathcal{U}$ with $x \in U$ and $S(U) < a$, but then $U \subseteq \{S < a\}$. $\square$

Now suppose that $K \subseteq X$ is compact. We want to define S for these types of sets, rather than just open sets. Define

$$S(K) := \inf\{\max_i S(U_i) : U_1, \dots, U_k \in \mathcal{U}, K \subseteq U_1 \cup \dots \cup U_k\}.$$

Remark 3.1. If S is locally finite, then $S(K) < \infty$ for all compact K.

Remark 3.2. If $K = \{x\}$, then $S(K) = S(x)$.

Lemma 3.3. *If $U \in \mathcal{U}$ and $\overline{U}$ is compact, then $S(U) \leq S(\overline{U})$.*

This is the first moment where we actually use the property (S1).

Proof. If $U_1 \dots, U_k \supseteq \overline{U} \supseteq U$, then by (S1), $S(U) \leq \max_i S(U_i)$. $\square$

Corollary 3.1. *If $U \in \mathcal{U}$ is also compact, then $S(U)$ is unambiguous.*

Proof. The previous lemma gives $S(U) \leq S(\overline{U}) \leq S(U)$. $\square$

Lemma 3.4. *For every compact $K \neq \varnothing$, we have*

$$S(K) = \sup\{S(x) : x \in K\}.$$

Proof. If $\{x\} \subseteq K$, then

$$S(x) = S(\{x\}) \leq S(K).$$

For the other direction, if $\sup_{x \in K} s(x) = \infty$, we are done. So assume that this is $< \infty$ and let $a > \sup_K S(x)$. Then for any $x \in K$, there is some $V_n \in \mathcal{U}$ with $S(V_n) < a$. K is compact, so there exist $x_1, \dots, x_k$ with $K \supseteq V_{x_1} \cup \dots \cup V_{x_k}$, and so

$$S(K) \leq \max_i S(V_{x_i}) < a.$$

Taking the inf over as gives

$$S(K) \leq \sup_K S(x).$$

$\square$

Here is the second necessary condition on the set function S:

(S2) ("Inner regularity") $S(U) = \sup\{S(K) : K \text{ is compact}, K \subseteq U\}$

Lemma 3.5. *If (S1) and (S2) hold, then $S(U) = \sup\{S(x) : x \in U\}$.*

Proof. We already know $\geq$. For the reverse inequality, use (S2): It is enough to show that

$$\sup_K S(x) = S(K) \leq \sup_U S(x).$$

for all compact $K \subseteq U$. This inequality holds by the previous lemma. $\square$

3.3 The settings we will apply this general theory to

The main settings we care about are:

1. Z is some "nice" topological space (usually a compact metric space), $X = M(Z)$, the finite signed Borel measures on Z, and $\mathcal{U}$ is the collection of convex subsets open for the weak topology defined by $C_b(Z)$ (i.e. the weak* topology if Z is a compact metric space).

2. $X = \mathbb{R}^d$ and $\mathcal{U}$ is the collection of convex open sets.

A suitable intermediate generality for us to cover these two cases will be: X is a locally convex topological vector space and $\mathcal{U}$ is the collection of open convex subsets.

Next time, we will

- find conditions making the point function S concave,
- observe a general "sequence counting" situation where those conditions hold.

4 Convexity of Set Functions and Measuring Type Classes

4.1 Recap + addressing superadditivity with $-\infty$

Let's fix a mistake from last time: If a_n are extended reals (i.e. $\in [-\infty, \infty)$ or $(-\infty, \infty]$) and satisfy $a_{n+m} \geq a_n + a_m$ for all n, m, then Fekete's lemma says that $\frac{a_n}{n} \to \sup_m \frac{a_m}{m} \in (-\infty, \infty]$. However, there can be problems if $-\infty$ is allowed among a_ns. For example,

$$a_n = \begin{cases} 0 & n \text{ even} \\ -\infty & n \text{ odd} \end{cases}$$

does not satisfy the conclusion of Fekete's lemma. The fix is that we will need to check separately that $a_n = -\infty$ for all sufficiently large n.

Last time, we discussed in what situations we can turn set functions into compatible point functions. In particular, we had a topological space X, an open cover $\mathcal{U}$, and a map $s : \mathcal{U} \to [-\infty, \infty]$ satisfying:

(S1) If $U \subseteq U_1 \cup \cdots \cup U_k$, then $s(U) \leq \max_i s(U_i)$.

Then

$$s(x) = \inf\{s(U) : U \in \mathcal{U}, U \ni x\},$$

and s is **locally finite** if $s(x) < \infty$ for all x. If we define

$$s(K) = \inf\{\max_i s(U_i) : K \subseteq U_1 \cup \cdots \cup U_k, U_i \in \mathcal{U}\},$$

then we had a lemma that said

$$s(K) = \sup\{s(x) : x \in K\}.$$

If we have the additional property

(S2) $s(U) = \sup\{s(K) : K \subseteq U \text{ is compact}\}$,

then we proved a lemma which says $s(U) = \sup\{s(x) : x \in U\}$.

4.2 Concavity of induced point functions

Now we will specialize to the situation where X is a locally convex topological vector space over $\mathbb{R}$ and $\mathcal{U}$ is the collection of open, convex sets. Another lemma from last time tells us that $s : X \to \mathbb{R}$ is upper semicontinuous, i.e. for all $a \in [-\infty, \infty]$, $\{s < a\}$ is open.

Lemma 4.1. *Suppose a set function s satisfies*

$$s\Bigg(\underbrace{\frac{1}{2}U + \frac{1}{2}V}_{=\{\frac{1}{2}u + \frac{1}{2}v : u \in U, v \in V\}}\Bigg) \geq \frac{1}{2}(s(U) + s(V)) \qquad \forall U, V \in \mathcal{U}$$

and is locally finite. Then the point function s is concave:

$$s(tx + (1-t)y) \geq ts(x) + (1-t)s(y).$$

Proof. Fix x, y, and let $W \in \mathcal{U}$ be a neighborhood of $w := \frac{1}{2}x + \frac{1}{2}y$. Then there exist $U, V \in \mathcal{U}$ such that $U \ni x$, $V \ni y$ and $\frac{1}{2}U + \frac{1}{2}V \subseteq W$. Therefore,

$$s(W) \geq s\left(\frac{1}{2}U + \frac{1}{2}V\right) \geq \frac{1}{2}\left(s(U) + s(V)\right) \geq \frac{1}{2}(s(x) + s(y)).$$

Take the inf over $W \ni w$ to get

$$s\left(\frac{1}{2}x + \frac{1}{2}y\right) \geq \frac{1}{2}(s(x) + s(y)).$$

Now conclude that

$$s(tx + (1-t)y) \geq ts(x) + (1-t)s(y)$$

for all dyadic rational t by induction on the dyadic depth of t. For example,

$$\begin{aligned} s\left(\frac{3}{4}x + \frac{1}{4}y\right) &= s\left(\frac{1}{2}x + \frac{1}{2}\left(\frac{1}{2}x + \frac{1}{2}y\right)\right) \\ &\geq \frac{1}{2}s(x) + \frac{1}{2}s\left(\frac{1}{2}x + \frac{1}{2}y\right) \\ &\geq \frac{1}{2}s(x) + \frac{1}{2}\left(\frac{1}{2}s(x) + \frac{1}{2}s(y)\right) \\ &= \frac{3}{4}s(x) + \frac{1}{4}s(y). \end{aligned}$$

The general dyadic case is similar.

Finally, we get all t by upper semicontinuity: if t_n are dyadic rationals with $t_n \to t$, then

$$s(tx + (1-t)y) \geq \limsup_n s(t_n x + (1-t_n)y).$$

Now apply the previous case. □

4.3 Measuring type classes in this setting

Here is a setting where we can apply these ideas: Let (M, λ) be a σ-finite measure space, let $X, \mathcal{U}$ be as before, and let $\varphi : M \to X$ be a measurable map, where

- "measurable" refers to the Borel σ-algebra of X.
- φ takes values inside a subset $E \subseteq X$ such that the restriction of the topology of X to E is separable and metrizable.

This second condition is a bit technical. Here are some examples:

Example 4.1. $E = X = \mathbb{R}^d$

Example 4.2. Let Z be a compact metric space, and let $X = M(Z)$ be the collection of signed finite measures on Z with the weak* topology, so $\mathcal{U}$ is the collection of weak* open convex sets. Then take $E = P(Z)$, the subset of probability measures, which is a weak*-closed convex subset of $M(Z)$ which is metrizable. In this case, we will usually have $M = Z$, $\lambda \in P(Z)$, and φ sending $z \mapsto \delta_z$.

Example 4.3. Take the same as above, but Z is any complete, separable metric space, and $M(Z)$ has the topology generated by all evaluations $\mu \mapsto \int f \, d\mu$ for $f \in C_b(Z)$. Still restrict φ to take values in $P(Z)$. In this situation, $P(Z)$ still has a complete separable metric, but this is harder; we won't prove this carefully here.

Values of interest: For $U \in \mathcal{U}$, how does

$$\lambda^{\times n}\Bigg(\underbrace{\left\{p \in M^n : \frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\right\}}_{T_n(U)}\Bigg)$$

behave? Previously, we had $M = A$, λ equals counting measure, and $\varphi(p) = \delta_p$, so $\frac{1}{n}\sum_{i=1}^n \varphi(p_i)$ was the empirical distribution of p.

Proposition 4.1. *There exists some* $s : \mathcal{U} \to [-\infty, \infty]$ *such that*

$$\lambda^{\times n}(T_n(U)) = e^{s(U)n + o(n)} \qquad \forall U \in \mathcal{U}.$$

Proof. Observe that if $p \in T_n(U)$ and $q \in T_m(U)$ and $r = pq$ is the concatenation, then

$$\frac{1}{n+m}\sum_{i=1}^{n+m} \varphi(r_i) = \frac{n}{n+m} \cdot \frac{1}{n}\sum_{i=1}^n \varphi(p_i) + \frac{m}{n+m} \cdot \frac{1}{m}\sum_{i=1}^m \varphi(q_i)$$

lies in U if $\frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U$ and $\frac{1}{m}\sum_{i=1}^m \varphi(p_i) \in U$, i.e. $T_{n+m}(U) \supseteq T_n(U) \times T_m(U)$. So

$$\lambda^{\times(n+m)}(T_{n+m}(U)) \geq \lambda^{\times n}(T_n(U)) \cdot \lambda^{\times m}(T_m(U)).$$

Take logs to get superadditivity. This gives

$$\begin{aligned} s(U) &= \lim_n \underbrace{\frac{1}{n}\log \lambda^{\times n}(T_n(U))}_{a_n/n} \\ &= \sup_n \frac{1}{n}\log \lambda^{\times n}(T_n(U)), \end{aligned}$$

provided that either $a_n = -\infty$ for all n or $a_n \neq -\infty$ for all sufficiently large n. We will complete the proof next time. □

5 Eventual Finiteness of $s_n(U)$ and Point Function Conditions

5.1 Recap

From last time, we have a σ-finite measure space (M, λ), a locally convex topological vector space X, and a measurable map $\varphi : M \to X$. We also let $\mathcal{U}$ be the convex open subsets of X. In this case, the equivalent of type classes is $T_n(U) = \{p \in M^n : \frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\}$, and we may let $s_n(U) := \log \lambda^{\times n}(T_n(U))$. We have shown that $T_{n+m}(U) \supseteq T_n(U) \times T_m(U)$, which implies that $s_{n+m}(U) \geq s_n(U) + s_m(U)$ (taking values in $[-\infty, \infty]$), and so, by Fekete,

$$s(U) = \lim_n \frac{s_n(U)}{n} = \sup_n \frac{s_n(U)}{n},$$

provided we show that either $s_n(U) = -\infty$ or $s_n(U) > -\infty$ for all sufficiently large n.

5.2 Eventual finiteness of $s_n(U)$

Lemma 5.1. *Either $s_n(U) = -\infty$ or $s_n(U) > -\infty$ for all sufficiently large n.*

Proof. Suppose $s_m(U) > -\infty$, i.e. $\lambda^{\times m}(T_m(U)) > 0$. Then $T_{km}(U) \supseteq T_m(U)^k$, so $s_{km}(U) > -\infty$. We need to control the indices in between.

Step 1: Reduce to the case where $U \ni 0$.[1] To do this, let $x \in U$ and now consider $\varphi'(m) = \varphi(m) - x$. Then $U' = U - x$ is a neighborhood of 0, and $\{p : \frac{1}{n}\sum_{i=1}^n \varphi'(p_i) \in U'\} = \{p : \frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\}$.

Step 2: Since U is convex and $U \ni 0$, $tU \subseteq U$ for all $t \in [0, 1]$. Also, since U is open, $U = \bigcup_{0 \leq t < 1} t \cdot U = \bigcup_{r \in \mathbb{N}} \frac{r}{r+1} U$; this is because $x \in U$ implies there is some $r \in \mathbb{N}$ such that $\frac{r+1}{r} x \in U$, i.e. $x \in \frac{r}{r+1} U$. The countable union is for measure theory purposes. So $T_n(U) = \bigcup_r T_n(\frac{r}{r+1} U)$, and so

$$\lambda^{\times n}(U) = \lim_{r \to \infty} \lambda^{\times n}\left(T_n\left(\frac{r}{r+1} U\right)\right).$$

So there exists some $r \in \mathbb{N}$ such that

$$\lambda^{\times m}\left(T_m\left(\frac{r}{r+1} U\right)\right) > 0.$$

Step 3: On the other hand, $X = \bigcup_{q \in \mathbb{N}} q \cdot U$, so for all n, we have $\lambda^{\times n}(T_n(q \cdot U)) > 0$ for some q.

[1] This step is not strictly necessary, but it makes our notation easier.

Step 4: Let $n \gg m$ with $n = km + \ell$ with $\ell \in \{0, \dots, m-1\}$. Suppose $p \in M^n$. Then

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) &= \frac{1}{n}\left(\sum_{i=1}^{n}\varphi(p_i) + \sum_{i=m+1}^{2m}\varphi(p_i) + \cdots + \sum_{i=(k-1)m+1}^{km}\varphi(p_i) + \sum_{i=km+1}^{n}\varphi(p_i)\right) \\ &= \frac{m}{n}\left(\frac{1}{m}\sum_{i=1}^{n}\varphi(p_i) + \frac{1}{m}\sum_{i=m+1}^{2m}\varphi(p_i) + \cdots + \frac{1}{m}\sum_{i=(k-1)m+1}^{km}\varphi(p_i)\right) \\ &\quad + \frac{\ell}{n}\cdot\underbrace{\frac{1}{\ell}\sum_{i=km+1}^{n}\varphi(p_i)}_{*}.\end{aligned}$$

For each of these k terms, we have positive measure for the event that $\frac{1}{m}\sum_{i=*}^{*+m}\varphi(p_i) \in \frac{r}{r+1}U$. Hence, we have positive measure that $\frac{1}{k}(\frac{1}{m}\sum_{i=1}^{m}\varphi(p_i)+\cdots+\frac{1}{m}\sum_{i=(k-1)m+1}^{km}\varphi(p_i)) \in \frac{r}{r+1}U$ (and we can even replace this by $\frac{mk}{n}$ times this). By step 3, we have positive measure that $* \in q \cdot U$ for some q independent of n and hence $\frac{\ell}{n} \cdot * \in \frac{q\ell}{n}U$. If all of these positive measure events occur, then

$$\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in \frac{r}{r+1}\cdot U + \frac{q\ell}{n}U.$$

Provided $n \geq q \cdot \ell \cdot (r+1)$, this implies

$$\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in \frac{r}{r+1}U + \frac{1}{r+1}U = U.$$

Hence, $s_n(U) > -\infty$ for this n. ☐

Remark 5.1. It is also possible that $\lambda^{\times n}(T_n(U)) = +\infty$, so $s_n(U) = +\infty$, and we may get $s(U) = +\infty$. Fekete's lemma still works, but the result is not meaningful. You usually want to look for additional reasons of why s is locally finite. The simplest condition is that if $\lambda(M) < \infty$, then $\lambda^{\times n}(T_n(U)) \leq \lambda(M)^n$ for all n.

5.3 Checking conditions to extend s to a point function

Next, we want to switch to point functions $s(x) = \inf\{s(U) : U \in \mathcal{U}, U \ni x\}$.

Proposition 5.1. *Under the same conditions as before, s is concave.*

Proof. $T_{n+m}(U) \supseteq T_n(U) \times T_m(U)$. Similarly, let $x \in T_n(U)$ and $y \in T_m(V)$ (where $U, V \in \mathcal{U}$). Then the concatenation $z = xy$ satisfies

$$\frac{1}{2n}\sum_{i=1}^{2n}\varphi(z_i) = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi(x_i) + \frac{1}{n}\sum_{i=1}^{n}\varphi(y_i)\right) \in \frac{1}{2}U + \frac{1}{2}V.$$

So $T_{2n}(\frac{1}{2}U + \frac{1}{2}V) \supseteq T_n(U) \times T_n(V)$, which tells us that

$$\frac{s_{2n}(\frac{1}{2}U + \frac{1}{2}V)}{2n} \geq \frac{1}{2}\left(\frac{s_n(U)}{n} + \frac{s_n(V)}{n}\right).$$

After letting $n \to \infty$, we get

$$s\left(\frac{1}{2}U + \frac{1}{2}V\right) \geq \frac{1}{2}(s(U) + s(V)).$$

By a previous lemma (the argument with dyadic rationals and applying upper semicontinuity), this gives that the point function $s(x)$ is concave. $\square$

Next, we quickly check that condition (S1) holds: If $U \subseteq U_1 \cup \cdots \cup U_k$, then $T_n(U) \subseteq T_n(U_1) \cup \cdots \cup T_n(U_k)$.. Using subadditivity and taking logs, we get

$$\frac{s_n(U)}{n} \leq \frac{\log K}{n} + \max_i \frac{s_n(U_i)}{n},$$

which gives

$$s(U) \leq \max_i s(U_i).$$

We also need conditions under which we can check (S2): $s(U) = \sup\{s(K) : K \subseteq U, K \text{ compact}\}$, where $s(K) = \inf\{\max_i s(U_i) : K = U_1 \cup \cdots \cup U_k, U_i \in \mathcal{U}\} = \sup_{x \in K} s(x)$ (by a previous lemma). To deduce (S2) in the setting of generalized type-counting, we need to assume:

Every open convex set U can be written as a countable union of compact, convex sets.

Example 5.1. In $\mathbb{R}^d$, by intersecting with balls, we can write every U as a countable union of bounded, open, convex sets, and then we can express each of these as a countable union of compact convex sets by looking at the set of points under a certain distance from the boundary.

Example 5.2. If $X = Y^*$ with the weak*-topology, this property also holds, but we will show this later.

6 Proving the (S2) Condition Via Compact Exhaustion

6.1 Compact exhaustion of convex open sets

Our setting is a σ-finite measure space (M, λ) with a measureable map $\varphi : M \to X$, where $\mathcal{U}$ is the collection of open convex subsets of X. We are trying to measure

$$\lambda^{\times n}\left\{p \in M^n : \frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\right\} = \exp(n \cdot s(U) + o(n))$$

(if it is finite for each each n, otherwise we get $s(U) = \infty$ and RHS $= \infty$). This exceptional case is not an issue if we can guarantee at most exponential growth, e.g. if $\lambda(M) < \infty$.

We can also define a point function

$$s(x) = \inf_{U \ni x} s(U),$$

and this is upper semicontinuous and concave. The next step needs an extra condition:

Each $U \in \mathcal{U}$ is a countable union of compact convex sets.

Here are examples where we can prove this property.

Example 6.1. $X = \mathbb{R}^d$. Let U be convex and open, and let $F_n = \{x \in U : |x| \leq n, \mathrm{dist}(x, U^c) \geq 1/n\}$. This is a closed subset of U (which is bounded and hence compact), and $U = \bigcup_n F_n$. To show that this is convex, we need to make sure the last condition preserves convexity. Observe that this condition holds iff $B_{1/n}(x) \subseteq U$. If this holds at x and y, then

$$B_{1/n}(tx + (1-t)y) = tB_{1/n}(x) + (1-t)B_{1/n}(y) \subseteq U.$$

Example 6.2. $X = Y^*$, where Y is a Banach space and X has the weak*-topology.

To prove the second example, we need the following:

Lemma 6.1. *For $X = Y^*$, if $U \in \mathcal{U}$, then there exist $y_1, \ldots, y_k$ and an open, convex $V \subseteq \mathbb{R}^k$ such that*

$$U = \{x : (\langle x, y_1\rangle, \ldots, \langle x, y_k\rangle) \in V\}$$

i.e. $U = L^{-1}[V]$, where $L : X \to \mathbb{R}^k$ sends $x \mapsto (\langle x, y_1\rangle, \ldots, \langle x, y_k\rangle)$.

Proof. Assume $U \ni 0$, so there exist linearly independent $y_1, \ldots, y_k \in Y$ and a neighborhood W of 0 in $\mathbb{R}^k$ such that $U \supseteq L^{-1}[W]$ (L as above). The main work is showing that $U = L^{-1}[V]$ for some $V \subseteq \mathbb{R}^k$. It is equivalent to show that $U = U + z$ for any $z \in \ker L$.

Suppose $z \in \ker L \subseteq U$ and so $\frac{1}{\varepsilon}z \in U$ for all ε. We have, by convexity, that $U \supseteq (1-\varepsilon)U$ for all $\varepsilon \in [0, 1]$. Similarly, $U \supseteq (1-\varepsilon)U + \varepsilon u$, where $u \in U$. So, in particular,

$$U \supseteq (1-\varepsilon)U + \varepsilon \cdot \frac{1}{\varepsilon}z = (1-\varepsilon)U + z$$

for all ε. Hence,

$$U \supseteq \bigcup_{1>\varepsilon>0} (1-\varepsilon)U + z = U + z.$$

By symmetry, $U = U + z$. $\square$

Proposition 6.1. *$X = Y^*$ has the desired property.*

Proof. Let $U = L^{-1}[V] = \bigcup_n L^{-1}[F_n]$ as above, where $L^{-1}[F_n]$ are weak* closed and convex. By Alaoglu's theorem, $X = \bigcup_n \overline{B_n}$, where $\overline{B_n}$ is compact and convex, and so $U = \bigcup_n (L^{-1}[F_n] \cap \overline{B_n})$. $\square$

6.2 Compact exhaustion implies (S2) condition

Proposition 6.2. *Suppose that X and $\mathcal{U}$ have this property. Then the (S2) condition*

$$s(U) = \sup\{s(K) : K \subseteq U \text{ is compact}\}$$

holds, where

$$s(K) := \inf\{\max_i s(U_i) : U_1, \ldots, U_k \in \mathcal{U}, K \subseteq U_1 \cup \cdots \cup U_k\}.$$

Proof. Recall that if $s(U) > -\infty$, then

$$s(U) = \lim_n \frac{\log \lambda^{\times n}(\{\frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\})}{n} = \sup_n \frac{\log \lambda^{\times n}(\{\frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\})}{n}.$$

Suppose $a < s(U)$. Then by this latter formulation for $s(U)$, there is some m such that $\log(\lambda^{\times m}(\{\cdots \in U\}))/m > a$. Write $U = \bigcup_k F_k$ where the F_k are compact and convex. So $\lambda^{\times m}(\{\cdots \in U\}) =\uparrow \lim_k \lambda^{\times m}(\{\cdots \in F_k\})$. So there is a compact convex $F \supseteq U$ with $\frac{\log \lambda^{\times m}(\cdots \in F\})}{m} > a$. By convexity of F, this gives

$$\frac{\log \lambda^{\times \ell m}(\{\frac{1}{\ell m}\sum_{i=1}^{\ell m} \cdots \in F\})}{\ell m} > a$$

for all ℓ. Now suppose $F \subseteq U_1 \cup \cdots \cup U_k$ with the $U_i \in \mathcal{U}$. Then $\lambda^{\times \ell m}(\{\cdots \in F\}) \leq k \max_i \lambda^{\times \ell m}(\{\cdots \in U_i\})$. So

$$\frac{\log \lambda^{\times \ell m}(\{\cdots \in F\})}{\ell m} \leq o(1) + \max_i \underbrace{\frac{\log \lambda^{\ell m}(\{\cdots \in U_i\})}{\ell m}}_{\to s(U_i)}.$$

The lim sup of this as $\ell \to \infty$ is a lower bound on $\max_i s(U_i)$ whenever $F \subseteq U_1 \cup \cdots \cup U_k$. Hence, $s(F) \geq a$. Since a was arbitrary $< s(U)$, we have (S2). $\square$

6.3 Special cases of our construction

Let's take stock of what we have so far: There exists $s : \mathcal{U} \to [-\infty, \infty]$ satisfying (S1) and (S2) such that

$$\lambda^{\times n}\left(p \in M^n : \frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\right) = \exp(n \cdot s(U) + o(n))$$

as $n \to \infty$ for all $U \in \mathcal{U}$. We also have an upper semicontinuous point function $s : X \to [-\infty, \infty]$ with $s(U) = \sup\{s(x) : x \in U\}$. Also, if $s : \mathcal{U} \to [-\infty, \infty]$ is locally finite, then $s : X \to [-\infty, \infty)$ and is concave.

Here are a few notable special cases:

Example 6.3. Let $M = A$ be a finite alphabet with λ as counting measure. Then $s(U) \leq \log|A|$ for all U, and $\varphi(a) = \delta_a \in P(A)$. Then $\frac{1}{n}\sum_{i=1}^{n}\varphi(a_i)$ is the empirical distribution p_a, and so our conclusion is

$$|T_n(U)| = \exp(n \sup_{p \in U} s(p) + o(n)).$$

Example 6.4. Let $X = \mathbb{R}^d$, and let $\xi_1, \xi_2, \dots$ be iid random variables with values in $\mathbb{R}^d$. So in the background, there is a probability space (M, λ) and measurable $\varphi : M \to \mathbb{R}^d$ such that $(\xi_1, \xi_2, \dots) \overset{d}{=} (\varphi(p_1), \varphi(p_2), \dots)$, where $(p_1, p_2, \dots) \sim \lambda^{\times\infty}$. Then there exists a point function $s : \mathbb{R}^d \to [-\infty, 0]$ such that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i \in U\right) = \exp\left(n \cdot \sup_{x \in U} s(x) + o(n)\right).$$

(Note that $s(x) \leq 0$ for all x because $s(U) \leq \log\lambda(M) = 0$ for all U.) If this event is unlikely (prob $\to 0$ as $n \to \infty$), then the event is called a **large deviation**, and this is the beginning of "Large Deviations Theory."

7 Large Deviations and Affine Approximation of Semicontinuous Functions

7.1 Recap

Here is our main result so far: We have a σ-finite measure space (M, λ) and a locally convex topological vector space, and $\mathcal{U}$ as the collection of open convex sets on X. We assume that every $U \in \mathcal{U}$ is an increasing union of compact, convex sets (e.g. $\mathbb{R}^d, Y^*$). We also have a measurable map $\varphi : M \to X$ which takes values in a metrizable subset. Then

$$\lambda^{\times n}\left(\left\{p \in M^n : \frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\right\}\right) = e^{n \cdot s(U) + o(n)}$$

for $U \in \mathcal{U}$. And if $s : \mathcal{U} \to [-\infty, \infty]$ ($\neq +\infty$ if s is locally finite), then there exists a point function $s : X \to [-\infty, \infty)$ which is upper semicontinuous and concave with $s(U) = \sup\{s(x) : x \in U\}$.

Example 7.1. In our original counting of type classes, we had $M = A$ is a finite alphabet, λ is counting measure, $p(a) = \delta_a$, and $\frac{1}{n}\sum_{i=1}^{n}\varphi(a_i) = p_a$ is the empirical distribution.

Example 7.2. In Large Deviations Theory, (M, λ) is a probability space, and $X = \mathbb{R}$. Then $\xi_1 = \varphi(p_1), \xi_2 = \varphi(p_2), \dots$ are iid random variables. Then the theorem says

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i \in U\right) = \exp\left(n \cdot \sup_U s(x) + o(n)\right).$$

Here, $s \leq 0$ always.

7.2 The large deviations principle

How does this fit into probability theory? Suppose $\mathbb{E}[|\xi_i|] < \infty$ (iff $\varphi \in L^1(\lambda)$). Then the Weak Law of Large Numbers says

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i \in U\right) \to \begin{cases} 1 & \mathbb{E}[\xi_i] \in U \\ 0 & \mathbb{E}[\xi_i] \notin \overline{U}. \end{cases}$$

In the case where $\sup_U s < 0$, this gives an exponential decay, upgrading the result of the Weak Law of Large Numbers. We can see Large Deviations Theory as a refinement of the convergence to zero in the WLLN.

The most "standard" formulation of the large deviations principle says

- $$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i \in U\right) \geq \exp\left(n \cdot \sup_{x \in U} s(x) + o(n)\right)$$

for all open $U \subseteq \mathbb{R}$. [This follows from the observation that LHS $\geq \mathbb{P}(\frac{1}{n}\sum_{i=1}^n \xi_i \in I)$ for all open intervals $I \subseteq U$.]

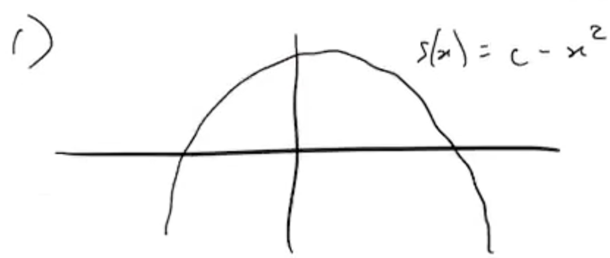
- $$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \xi_i \in C\right) \leq \exp\left(n \cdot \sup_C s(x) + o(n)\right)$$

  for all closed $C \subseteq \mathbb{R}$. [This follows from above if C is compact: if $\sup_C s(x) = \alpha$, then we can cover C with finitely many open intervals $I_1, \ldots, I_k$ such that $\mathbb{P}(\frac{1}{n}\sum_{i=1}^n \xi_i \in I_\ell) \leq e^{n \cdot \sup_{I_\ell} s + o(n)}$ for all $\ell \leq k$. We can extend this to closed sets if $s(x) \to -\infty$ as $x \to \pm\infty$, in which case s is called *good.*[2] If s is good, we can cover a general closed set with far away half infinite intervals on each side and have a compact set in the middle. The apply the previous argument.]

7.3 Approximation of concave, upper semicontinuous functions by affine functions

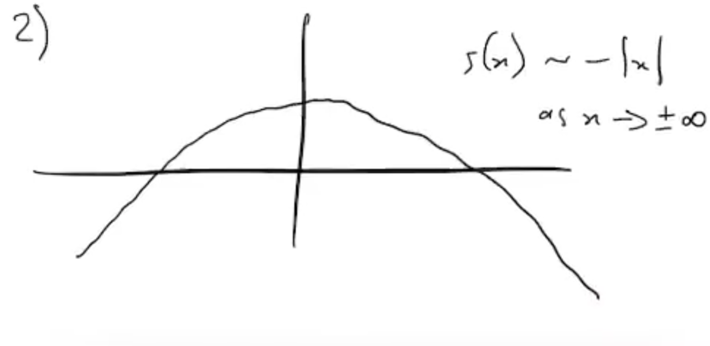
Returning to the general story, assuming local finiteness, $s : X \to [-\infty, \infty)$ is upper semicontinuous and concave. How can we describe these in general? Here are some examples where $X = \mathbb{R}$:

Example 7.3. $s(x) = c - x^2$



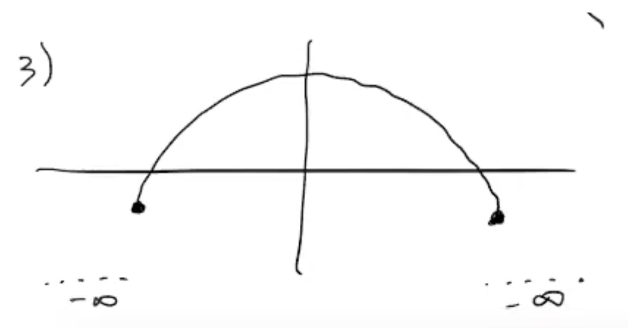
[2] This is also called *proper* in analysis.

Example 7.4. $s(x) \sim -|x|$ as $x \to \pm\infty$.

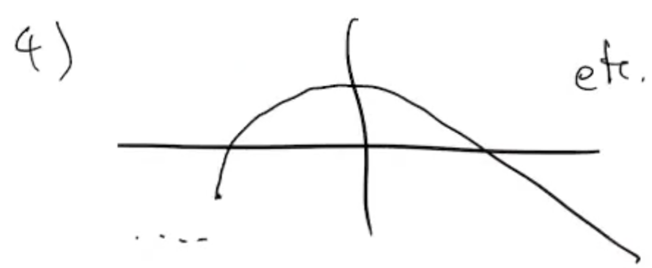


Example 7.5. Upper semicontinuous example with

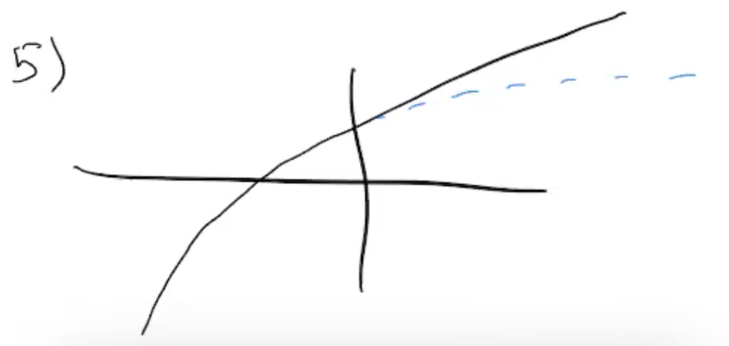
$$s(x) = \begin{cases} c - x^2 & |x| \leq M \\ -\infty & |x| > M. \end{cases}$$



Example 7.6. An example that tends to $-\infty$ on the right:



Example 7.7. An example which tends to $+\infty$ on the right:



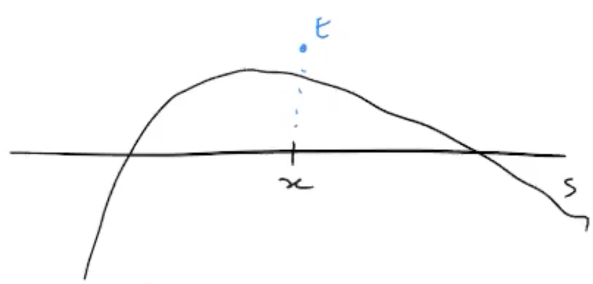
The key to all these cases is whether we can draw a straight tangent line that lies entirely above the graph. In example 3, we run into a bit of trouble at the endpoints, since we cannot draw vertical line (with infinite slope), so we may need an ε bit of wiggle room. What this idea leads to is the fact that any upper semicontinuous function can be written as an infimum of affine functions. Here is a lemma that we need.

Lemma 7.1. *Let X be a locally convex topological vector space, and let $s : X \to [-\infty, \infty)$ be an upper semicontinuous concave function with $x \in X$. If $t > s(x)$, then there exists a $c \in \mathbb{R}$ and a continuous functional y on X such that*

- $c + \langle y, z \rangle \geq s(z)$ *for all* $z \in X$,
- $c + \langle y, x \rangle < t$.

This is the infinite dimensional analogue of whether we can place a line above the graph of $s(x)$ which stays below any point above the graph.

Proof.



We want to think of this as a picture in a larger topological vector space that includes the vertical coordinate. Let $\widetilde{X} = X \times \mathbb{R}$, which is a locally convex topological vector space with the product topology. The point (x, t) lies above the subset $C := \{(x, \theta) : \theta \leq s(z)\}$.

This subset is closed because s is upper semicontinuous and is convex because s is concave. By the Hahn-Banach separation theorem, there exists a $\widetilde{y} \in \widetilde{X}^*$ such that $\widetilde{y}(x,t) > \sup_C \widetilde{y}$. Also, $\widetilde{y}$ can be written as $\widetilde{y}(z,\theta) + \langle y, z\rangle + \alpha\theta$ for some $y \in X^*$ and $\alpha \in \mathbb{R}$. If we let c be the y-intercept of the hyperplane given by Hahn-Banach and rewrite the inequality $\widetilde{y}(x,t) > \sup_C \widetilde{y}$ in terms of c, we get the result. □

Proposition 7.1. *A function $s : X \to [-\infty, \infty)$ is upper semicontinuous and concave if and only if*

$$s(x) = \inf\{c + \langle y, x\rangle : c \in \mathbb{R}, y \in X^*, c + \langle y, z\rangle \geq s(z)\ \forall z \in X\}.$$

7.4 The Fenchel-Legendre transform

How can we give a canonical family in here? For fixed $y \in X^*$, what is the best c to use? We want $c + \langle y, z\rangle \geq s(z)$ for all z. That is, we want $c \geq s(z) - \langle y, z\rangle$, so we want to take

$$c = \sup_{z \in X} s(z) - \langle y, z\rangle =: s^*(y).$$

This is known as the **Fenchel-Legendre** transform of s. Here are some properties of s^*:

Proposition 7.2.

1. *s^* is lower semicontinuous and convex.*

   Proof. s^* is the supremum of lower semicontinuous, convex (affine) functions. □

2. *Provided $s \not\equiv -\infty$, we get $s^* : X^* \to (-\infty, \infty]$ and s^* is not always $+\infty$.*

3. *$s(z) = \inf_y s^*(y) - \langle y, z\rangle$. That is, $s = (s^*)_*$.*

8 Integral Formula for the Fenchel-Legendre Transform

8.1 The Fenchel-Legendre transform and the integral formula

Last time, we defined the **Fenchel-Legendre transform** $s^* = \sup_x s(x) + \langle y, x\rangle$, which is convex, lower semicontinuous, is $s^* : X^* \to (-\infty, \infty]$, and is not always $+\infty$.[3] We also saw that $s = (s^*)_*$, so we can recover s from its Fenchel-Legendre transform.

Let's focus on the $X = Y^*$ case, since this also subsumes the $X = \mathbb{R}^k$ case. Also assume $\lambda \neq 0$.

Theorem 8.1. *In this generalized type counting problem for $X = Y^*$,*

$$s^*(y) = \log \int e^{\langle y,\varphi\rangle}\, d\lambda$$

for $y \in Y$.

Before proving this, observe:

$$\begin{aligned}\exp(s^*(ty + (1-t)w)) &= \int e^{t\langle y,\varphi\rangle + (1-t)\langle w,\varphi\rangle}\, d\lambda \\ &= \int e^{t\langle y,\varphi\rangle} \cdot e^{(1-t)\langle w,\varphi\rangle}\, d\lambda\end{aligned}$$

Using Hölder's inequality,

$$\leq \left(\int e^{\langle y,\varphi\rangle}\, d\lambda\right)^t \left(\int e^{\langle w,\varphi\rangle}\, d\lambda\right)^{1-t},$$

so taking logs gives that this expression is convex. We can also check that this expression is lower semicontinuous.

8.2 Proofs of the upper bound and the lower bound

Proof. ($\leq$): Since $s^*(y) = \sup_x s(x) + \langle y, x\rangle$, we need to show that

$$s(x) + \langle y, x\rangle \leq \log \int e^{\langle y,\varphi\rangle}\, d\lambda$$

for all x. Let $\varepsilon > 0$, and consider $U = \{x' : \langle y, x'\rangle > \langle y, x\rangle - \varepsilon\}$. We know that

$$\begin{aligned}s^{n\cdot(s(x)+\langle y,x\rangle)} &\leq e^{n(s(U)+\langle y,x\rangle)} \\ &= e^{o(n)} e^{n\langle y,x\rangle} \lambda^{\times n}\left(\left\{p : \frac{1}{n}\sum_{i=1}^{n} \varphi(p_i) \in U\right\}\right)\end{aligned}$$

[3] Many authors study $\widetilde{s} = -s$ throughout and then get $s^*(y) = \sup_x \langle y, x\rangle - \widetilde{s}(x)$ and $\widetilde{s}(z) = \sup_y \langle y, z\rangle - s^*(y)$. We use a different convention.

$$= e^{o(n)} e^{n\langle y,x\rangle} \lambda^{\times n}\left(\left\{p : \sum_{i=1}^{n} \langle y, \varphi(p_i)\rangle > n\langle y, x\rangle - n\varepsilon\right\}\right)$$

Exponentiate both sides in the inequality and apply Markov's inequality:[4]

$$\begin{aligned}
&\le e^{o(n)} e^{n\langle y,x\rangle} e^{n\varepsilon - n\langle y,x\rangle} \int e^{\sum_{i=1}^{n}\langle y,\varphi(p_i)\rangle}\, d\lambda^{\times n} \\
&= e^{o(n)+n\varepsilon} \int_{M^n} \prod_{i=1}^{n} e^{\langle y,\varphi(p_i)\rangle}\, d\lambda^n \\
&= e^{o(n)} e^{\varepsilon n} \left(\int e^{\langle y,\varphi\rangle}\, d\lambda\right)^n,
\end{aligned}$$

so

$$n(s(U) + \langle y, x\rangle) \le o(n) + \varepsilon n + n \log \int e^{\langle y,\varphi\rangle}\, d\lambda.$$

Divide by n and send $n \to \infty$ to get

$$s(x) + \langle y, x\rangle \le \; s(U) + \langle y, x\rangle \le \varepsilon + \log \int e^{\langle y,\varphi\rangle}\, d\lambda.$$

Since ε is arbitrary, we get $(\le)$.

To get the lower bound, let's look at the proof of the upper bound and try to make it as tight as possible. The first inequality is close if U is a small enough neighborhood of x. In the Chernoff bound, we want to see when this is close to equality. To prove $(\ge)$, we will look at the Chernoff bound step; here's the idea: Consider

$$e^{n\langle y,x\rangle} \lambda^{\times n}\left(\left\{p : \frac{1}{n}\sum_{i=1}^{n} \langle y, \varphi(p_i)\rangle \in U\right\}\right),$$

where we want to make U small enough around x to force this to be $\approx \langle y, x\rangle$. We then get

$$e^{n\langle y,x\rangle} \lambda^{\times n}\left(\left\{p : \exp \sum_{i=1}^{n} \langle y, \varphi(p_i)\rangle \approx e^{n\langle y,x\rangle}, \frac{1}{n}\sum_{i=1}^{n} \varphi(p_i) \in U\right\}\right).$$

This is

$$\approx e^{\pm\varepsilon n} \int_{\{\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i)\in U\}} e^{\sum_{i=1}^{n}\langle y,\varphi(p_i)\rangle}\, d\lambda^{\times n} \le e^{\varepsilon n} \int_{M^n} e^{\sum_{i=1}^{n}\langle y,\varphi(p_i)\rangle}\, d\lambda^{\times n}.$$

So the question becomes: Can we find an x where most of the mass lies in the set $\{\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\}$?

Now let's prove $(\ge)$ carefully. First assume two conditions:

[4] This is sometimes called a **Chernoff bound**.

1. $Z = \int e^{\langle y,\varphi\rangle}\, d\lambda < \infty$.

2. p takes values in a compact subset K of X.

In this case, we can define a new probability measure on M by

$$d\theta(p) = \frac{1}{Z} e^{\langle y,\varphi(p)\rangle}\, d\lambda(p)$$

(using assumption 1). Now, for any $A \subseteq M^n$,

$$\int_A e^{\sum_{i=1}^n \langle y,\varphi(p_i)\rangle}\, d\lambda^{\times n} = Z^n \theta^{\times n}(A).$$

With $A = \{\frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\}$, we get

$$Z^n \theta^{\times n}\left(\left\{\frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\right\}\right)$$

This suggests we can use the Weak Law of Large Numbers for θ and φ.[5] To do this carefully, we need assumption 2: p takes values in $K \subseteq X$, so it has a barycenter with respect to θ: a unique $x \in K$ such that

$$\int \langle y,\varphi\rangle\, d\theta = \langle y, x\rangle \qquad \forall y \in Y.$$

And now a vector-valued Weak Law of Large Numbers holds: for this x and any weak* neighborhood $U \ni x$, we get

$$\theta^{\times n}\left(\left\{\frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\right\}\right) = 1 - o(1)$$

as $n \to \infty$. As a result, for any weak* neighborhood of this x, we now get

$$\int_{\{\frac{1}{n}\sum_{i=1}^n \varphi(p_i)\in U\}} = Z^n \theta^{\times n}\left(\left\{\frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\right\}\right) \geq Z^n e^{o(n)}.$$

Insert this to reverse the previous upper bound proof to get an x such that $s(x) + \langle y, x\rangle \geq \log Z - \varepsilon$. This gives $s^*(y) = \log Z$.

To remove assumptions 1 and 2, recall that (M,λ) is σ-finite and $X = \bigcup_n K_n$, so for any $a < \int e^{\langle y,\varphi\rangle}\, d\lambda$, there exists a measurable $A \subseteq M$ such that $\infty > \int_A e^{\langle y,\varphi\rangle}\, d\lambda > a$, and $\varphi(A)$ takes values in some K_n. Now run the previous argument with $d\lambda'(p) = \mathbb{1}_A(p)\, d\lambda(p)$ to get that for every ε, there is an x such that $s(x) + \langle y, x\rangle \geq \log a - \varepsilon$. Since $a < \int e^{\langle y,\varphi\rangle}\, d\lambda$ was arbitrary, we get

$$s^*(y) = \log \int e^{\langle y,\varphi\rangle}\, d\lambda,$$

even if this is $+\infty$. $\square$

[5]This is the key idea of the lower bound proof. It is called the **change of measure** idea.

9 Cramér's Theorem and Recovering Entropy as the Exponent

9.1 Cramér's theorem

We have a σ-finite measure space (M, λ), and a measurable map $\varphi : M \to X$, where $X = Y^*$ is a locally convex space with the weak* topology. We found that

$$\lambda^{\times n}\left(\left\{p \in M^n : \frac{1}{n}\sum_{i=1}^{n} \varphi(p_i) \in U\right\}\right) = e^{n \cdot s(U) + o(n)},$$

where $s(U) = \sup_{x \in U} s(x)$ for some point function s which is upper semicontinuous and concave. To study s, we have introduced Fenchel-Legendre duality:

$$s(x) = \inf_{y} s^*(y) - \langle y, x\rangle,$$

where

$$s^*(y) := \sup_{x} s(x) + \langle y, x\rangle$$

is sometimes known as the **convex conjugate** of s. Last time, we proved a formula: if $s(x) < \infty$ for all n, then

$$s^*(y) = \log \int e^{\langle y, \varphi\rangle}\, d\lambda.$$

Remark 9.1. In the proof of this integral formula, to show $(\leq)$, we showed that $s(x) + \langle y, x\rangle \leq \text{RHS}$ for all x, y. For this, given $\varepsilon > 0$, we found $U \ni x$ such that

$$\lambda^{\times n}(\{\cdots \in U\}) \leq e^{\varepsilon n + o(n)} \left(\int e^{\langle y, \varphi\rangle}\, d\lambda\right)^n.$$

This part of the proof does not require that s is finite. In fact, it gives a way to prove $s(U) < \infty$ and hence $s(x) < \infty$. So if there is some $y \in Y$ such that $\int e^{\langle y, \varphi\rangle}\, d\lambda < \infty$, then $s < \infty$ and s^* is as in the theorem. The mantra is that $s < \infty$ everywhere iff $s^* < \infty$ somewhere.

A special case is when (M, λ) is a probability space and $X = \mathbb{R}^d$. In this case, we get the following version of the theorem we proved before:

Theorem 9.1 (Cramér, 1937)**.** *Let $\xi_1, \xi_2, \ldots$ are i.i.d. random vectors in $\mathbb{R}^d$. Then*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \xi \in U\right) = \exp\left(n \cdot \sup_{x \in U} s(x) + o(n)\right),$$

where

$$s(x) = \inf_{y \in \mathbb{R}^d} \Lambda(y) - \langle y, x\rangle,$$

and

$$\Lambda(y) = \log M(y) = \log \mathbb{E}[e^{\langle y, \xi_1 \rangle}]$$

is the ***cumulant generating function****.*

In a number of texts, our s is denoted by $-I$ (so the inf becomes a sup, etc.).

9.2 Connection to the Kullback-Leibler divergence in the case of empirical distributions

Let K be a compact metric space, λ be a finite Borel measure, $X = M(K)$ be the space of measures on K (equal to $C(K)^*$ by Riesz representation), and $\varphi(p) = \delta_p$. In this case, $\frac{1}{n}\sum_{i=1}^n \varphi(p_i)$ is the empirical distribution of $(p_1, \dots, p_n)$.

Theorem 9.2. *In this setting, $s(\mu) = -\infty$ unless $\mu \in P(K)$ and $\mu \ll \lambda$, and in that case,*

$$s(\mu) = -\int \frac{d\mu}{d\lambda} \log \frac{d\mu}{d\lambda} \, d\lambda.$$

We will denote the right hand side by $\widetilde{s}(\mu)$ until we have proven the theorem; that way, the proof is to show that $s = \widetilde{s}$.

Remark 9.2. Note that

$$\widetilde{s}(\mu) = \int \eta\left(\frac{d\mu}{d\lambda}\right) d\eta, \qquad \eta(t) = \begin{cases} -t \log t & t > 0 \\ 0 & t = 0. \end{cases}$$

If $|\eta(\frac{d\mu}{d\lambda})| \in L^1(\lambda)$, then $\widetilde{s}(\mu) > -\infty$. Otherwise, we set $s(\mu) := -\infty$.

Remark 9.3. Here is an alternative formula that will be useful:

$$\widetilde{s}(\mu) = -\int \log \frac{d\mu}{d\lambda} \, d\mu.$$

This formula is useful, but it is a little harder to see the natural $-\infty$ convention with this version.

Here are 2 special cases:

Example 9.1. Let K be finite with λ being counting measure. Then $\frac{d\mu}{d\lambda}(a) = \mu(\{a\})$, and so

$$\widetilde{s}(\mu) = -\sum_a \mu(\{a\}) \log \mu(\{a\}) = H(\mu)$$

is the Shannon entropy.

Example 9.2. If $\lambda(K) = 1$, then

$$-\widetilde{s}(\mu) = \begin{cases} \int \frac{d\mu}{d\lambda} \log \frac{d\mu}{d\lambda}\, d\lambda & \\ +\infty & \text{in the cases described above} \end{cases}$$

is called the **Kullback-Leibler divergence**. The standard notation for this is $D(\mu\|\lambda)$.

Lemma 9.1. *If $\lambda(K) = 1$, then $D(\mu\|\lambda) \geq 0$, with equality if $\mu = \lambda$.*

Proof.

$$D(\mu\|\lambda) = \int \frac{d\mu}{d\lambda} \log \frac{d\mu}{d\lambda}$$

$$= \int -\eta\left(\frac{d\mu}{d\lambda}\right) d\lambda$$

$-\eta$ is strictly concave, so using Jensen's inequality gives

$$\begin{aligned} &-\eta\left(\int \frac{d\mu}{d\lambda}\, d\lambda\right) \\ &= -\eta(1) \\ &= 1 \log 1 \\ &= 0. \end{aligned}$$

We get equality iff $\frac{d\mu}{d\lambda}$ is constant for λ-a.e., that is, iff $\mu = \lambda$. □

Let's prove the theorem:

Proof. We want to prove that $s = \widetilde{s}$. Using the expresion for s in terms of the Fenchel-Legendre transform and using the integral formula, we want to show that

$$\inf\left\{\log \int e^{f(p)}\, d\lambda(p) - \langle f, \mu\rangle : f \in C(K)\right\} = \widetilde{s}(\mu).$$

This is known as **Gibbs' variational formula**.

($\geq$): We want

$$\log \int e^f\, d\lambda - \langle f, \mu\rangle \geq -\int \frac{d\mu}{d\lambda} \log \frac{d\mu}{d\lambda}\, d\lambda.$$

The key object is

$$d\mu_f(p) = \frac{e^{f(p)}}{Z(f)}\, d\lambda(p), \qquad Z(f) = \int e^f\, d\lambda,$$

which is sometimes called the **Gibbs measure** of f with respect to λ. Observe that $\lambda \ll \mu_f$ and $\mu_f \ll \lambda$, so if $\mu \ll \lambda$, then $\mu \ll \mu_f$, then $\frac{d\mu}{d\lambda} = \frac{d\mu}{d\mu_f} \frac{d\mu_f}{d\lambda}$, and so

$$\widetilde{s}(\mu) = -\int \log \frac{d\mu}{d\lambda}\, d\mu$$

$$
\begin{aligned}
&= -\int \log \frac{d\mu}{d\mu_f}\, d\mu - \int \log \frac{d\mu_f}{d\lambda}\, d\mu \\
&= -D(\mu\|\mu_f) - \int (f - \log Z)\, d\mu \\
&= -D(\mu\|\mu_f) + \{\log Z - \langle f, \mu\rangle\}\,.
\end{aligned}
$$

Rearrange this to get

$$
\log Z - \langle f, \mu\rangle = \widetilde{s}(\mu) + D(\mu\|\mu_f) \geq \widetilde{s}(\mu),
$$

with equality iff $\mu = \mu_f$.

($\leq$): We already know this if $\mu = \mu_f$ for some $f \in C(K)$. The summary of the rest of the proof is "such measures μ_f are dense as f varies." In more detail:

(a) $\inf\{\log \int e^f\, d\lambda - \langle f, \mu\rangle : f \in C(K)\}$ has the same value if we enlarge $C(K)$ to $B(K)$, the bounded Borel functions. This is because given λ and μ, $C(K)$ is dense in $L^1(\lambda+\mu)$, so for all $g \in B(K)$ (all uniformly bounded), there is some $(f_n)_n$ in $C(K)$ with $f_n \to g$ in $L^1(\lambda)$ and $L^1(\mu)$. Then $\langle f_n, \mu\rangle \to \langle g, \mu\rangle$, and $\int e^{f_n}\, d\lambda \to \int e^g\, d\lambda$.

(b) Now suppose $\mu \not\ll \lambda$. Then there is an A such that $\lambda(A) = 0$ and $\mu(A) > 0$. Let $g = c\mathbb{1}_A \in B(K)$. This gives

$$
\log \int e^g\, d\lambda - \langle g, \mu\rangle = 0 - c\mu(A) \to -\infty
$$

as $c \to +\infty$. So $\inf\{\cdots\} = -\infty$, as required.

(c) Lastly, suppose $d\mu = \rho\, d\lambda$. If $\rho = e^g$ with $g \in B(K)$, we are done by the previoius calculation. Otherwise, choose $(g_n)_n$ in $B(K)$ such that

$$
e^{g_n} \to \rho \begin{cases} \text{from below} & \text{if } \rho > 1 \\ \text{from above} & \text{if } \rho \leq 1. \end{cases}
$$

Now show that:

- $$\log \int_{\{\rho\leq 1\}} e^{g_n}\, d\lambda \to \log \int \rho\, d\lambda = \log 1 = 0,$$
- $$\log \int_{\{\rho>1\}} e^{g_n}\, d\lambda \to \log \int \rho\, d\lambda = \log 1 = 0,$$
- $$\langle g_n, \mu\rangle \to \langle \log\rho, \mu\rangle = \widetilde{s}(\mu).$$ □

10 Introduction to Statistical Physics

10.1 Recap

Last time, we mentioned that in the case $X = M(K)$, λ is finite on K, and $\varphi : K \to X$ sends $p \mapsto \delta_p$ for a compact metric space K, we get

$$s(\mu) = \begin{cases} -\int \frac{d\mu}{d\lambda} \log \frac{d\mu}{d\lambda}\, d\lambda & \mu \in P, \mu \ll \lambda \\ -\infty & \text{else.} \end{cases}$$

For example, if $K = A$ is finite and $\lambda \in P(A)$, we get

$$\lambda^{\times n}(\{a \in A^n : p_a \approx \mu\}) = e^{-nD(\mu\|\lambda)+o(n)},$$

where

$$D(\mu\|\lambda) := \sum_a \frac{\mu(a)}{\lambda(a)} \log \frac{\mu(a)}{\lambda(a)}.$$

10.2 Quick intro to statistical physics

Imagine n point particles located in space with positions $r_i(t)$ for $1 \le i \le n$. The laws of motion give

$$\frac{d^2 r_i(t)}{dt^2} = \frac{1}{m_i} F_i,$$

where the F_i are forces. A typical simple example of forces is

$$F_i = -\nabla V(r_i(t))$$

for some **potential energy** $V : \mathbb{R}^3 \to \mathbb{R}$.

How big is n? Avogadro's number, $\approx 6 \times 10^{23}$, is roughly the number of Carbon atoms in the graphite in a pack of pencils. This tells us that the number of particles is way too large to be solved in the usual way. The founding ansatz in statistical physics is that in a macroscopic physical system, once a few key quantities are fixed, "almost" all possible microscopic states of the system look macroscopically the same.

In an example of classical particles, the total energy is conserved:[6]

$$\Phi(r_i, p_i) = \underbrace{\sum_{i=1}^n V(r_i)}_{\text{potential energy}} + \underbrace{\sum_{i=1}^n \frac{1}{2m_i}|p_i|^2}_{\text{kinetic energy}},$$

where $p_i = m_i \frac{dr_i}{dt}$ is the momentum of a particle, and m_i is the mass.

[6] In practice, you can't directly measure or control the amount of total energy in a system. You have to control factors that would affect the energy in the system, such as temperature and pressure.

The **phase space** is $(\mathbb{R}^3)^n \times (\mathbb{R}^3)^n$, which keeps track of the position and momentum of every particle. The **micro-state** moves around in the level set

$$\Omega(E) = \{(r_1, \dots, r_n, p_1, \dots, p_n) : \Phi(r_i, p_i) = E\}$$

if the system is isolated, i.e. no energy exchange occurs with the surroundings.

We want to restrict to cases where these sets are bounded, which we can do if we impose a restriction on the set of potential energies (for example, if the particles occupy a **potential well**). Any other macroscopic observable quantity will be a function of $(r_1, \dots, r_n, p_1, \dots, p_n)$. We want the functions that arise this way to stay close to some constant value on most of $\Omega(E)$. By "most," we mean most in the sense of the measure on $\Omega(E)$; instead of using $6n-1$ dimensional Hausdorff measure on $\Omega(E)$ (which is a $6n-1$ dimensional manifold), it is easier in practice to thicken the manifold (loosen the restriction on the exact energy E) and use $6n$ dimensional Lebesgue measure.

In the previous formula for the energy, these particles are not interacting. This is the easiest case where we can solve things explicitly. In reality, molecules have pairwise interactions, so we could include a term like $\sum_{i,j} V_{i,j}(r_i, r_j)$. We will first deal with the noninteracting particles case, and once we have developed the tools to talk about these things, we will then deal with systems with interactions.

We call these observables **concentrated** (probability terminology) or **self-averaging** (physics terminology).

11 Mathematical Setup for Statistical Mechanics

11.1 Relationship to type counting

Our goal is to set up some mathematical models of physical systems with large numbers of degrees of freedom and see whether "most microstates look the same" after fixing a few macroscopic parameters.

To begin with, we will focus on n classical[7] non-interacting, identical point particles moving around in a potential. Non-interacting particles can be thought of as particles where the energies of interaction are negligible compared to the total energy of the system.

We will describe the particles via their positions $r_1, \dots, r_n \in \mathbb{R}^3$ and velocities $v_1, \dots, v_n \in \mathbb{R}^3$. Newton's law says

$$m\frac{dv_i}{dt} = m\frac{d^2 r_i}{dt} = F_i = -\nabla V(r_i).$$

We will assume the mass m equals 1, so $v_i = p_i$, the momentum. The total energy is

$$\Phi(r_1, \dots, r_n, p_1, \dots, p_n) = \sum_{i=1}^n \varphi(r_i, p_i), \qquad \varphi(r_i, p_i) = v(r_i) + \frac{1}{2}|p_i|^2.$$

We want to study averages over the set

$$\Omega(n, I) = \{(r_1, \dots, p_n) \in (\mathbb{R}^3)^n \times (\mathbb{R}^3)^n : \frac{1}{n}\Phi(r_1, \dots, p_n) \in I = (E - \varepsilon, E + \varepsilon)\}$$

for some desired total energy E and error tolerance ε.

The first step is to ask: How big is $\Omega(n, I)$ in the sense of Lebesgue measure? This is just an instance of generalized type counting: $M = \mathbb{R}^3 \times \mathbb{R}^3$, $\lambda = m_3 \times m_3$, and $\varphi : M \to [0, \infty)$. Note that we are assuming V is lower bounded, and we are adjusting it by a constant to assume its minimum equals 0. Now the asymptotic behavior of

$$\lambda^{\times n}\left(\left\{(r_1, \dots, p_n) : \frac{1}{n}\Phi \in I\right\}\right) = \exp\left(n \cdot \sup_{E \in I} s(E) + o(n)\right).$$

To go further, we need to know more about s in the present situation.[8]

11.2 Assumptions of the model and properties of the entropy

Here are the salient features of the present situation and a necessary assumption:

- (M, λ) is σ-finite but not finite.

[7] Here, classical means not quantum. You can do this with quantum physics, but it requires making use of the full machinery of Hilbert spaces.

[8] In this situation, s is $\frac{1}{n}$ times the **Boltzman entropy.**

- $\min \varphi = 0 = \text{ess } \min \varphi$, i.e. $\lambda(\{(r,p) : \varphi(r,p) < a\}) > 0$ for all $a > 0$.
- We need V to confine particles strongly enough to bounded regions of space. Mathematically, we will ask that $\int e^{-\beta\varphi}\, d\lambda < \infty$ for all $\beta > 0$. [Note that $\int e^{-\beta\varphi}\, d\lambda = \iint_{\mathbb{R}^3\times\mathbb{R}^3} e^{-\beta V} e^{-(\beta/2)|p|^2}\, dm_3(r)\, dm_3(p)$.][9]

Under these assumptions, we know that $s(E)$ exists, is upper semicontinuous, concave, and is $s : \mathbb{R} \to [-\infty, \infty)$. In fact, we also know that $s \equiv -\infty$ on $[-\infty, 0)$, so we can focus on $s|_{[0,\infty)}$. In this case, we have our variational formula

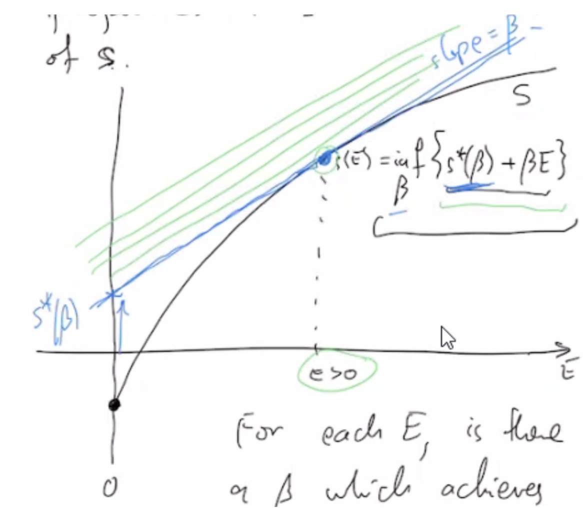
$$s(E) = \inf_{\beta}\{s^*(\beta) + \beta E\},$$

where

$$s^*(\beta) = \log \int e^{-\beta\varphi}\, d\lambda.$$

Note that we have switched y with $-\beta$, as β has a physical interpretation.

We will use the formula for s^* to derive more qualitative features of s. We will set up ways of translating properties of s^* into those of s. Here is a picture of s and s^*:



For each E, is there a β which achieves the equality $s(E) = s^*(\beta) + \beta E$? The answer is yes, if and only if s has finite one-sided derivative on at least one side, and then you can use any $D_+s(E) \le \beta \le D_-s(E)$. In particular, if $s'(E)$ exists, then the unique choice is $\beta = s'(E)$.

[9] Notably, gravity does not satisfy this assumption, but gravity operates on different scales than we are working with, so we will ignore it.

12 Duality: Deriving Properties of s Via Properties of s^*

12.1 Recap

Our setup from last time is a system of n "non-interacting particles." M is the phase space $\mathbb{R}^3 \times \mathbb{R}^3$, $\lambda = m_3 \times m_3$ is a σ-finite but not finite measure, and $\varphi : M \to [0, \infty)$ is $\varphi(r,p) = \varphi_{\text{pot}}(r) + \frac{1}{2}|p|^2$ (potential energy + kinetic energy). We will assume φ is lower bounded and normalize φ so that $\min \varphi = \operatorname{ess\,min} \varphi = 0$. Then, for open interval $I \subseteq \mathbb{R}$,

$$\begin{aligned} \lambda^{\times n} &\left(\left\{ (r_1, \ldots, r_n, p_1, \ldots, p_n) : \frac{1}{n} \Phi_n(r_1, \ldots, p_n) := \frac{1}{n} \sum_{i=1}^n \varphi(r_i, p_i) \in I \right\} \right) \\ &= \exp\left(n \cdot \sup_{E \in I} s(E) + o(n) \right) \end{aligned}$$

The intuition is that

$$\lambda^{\times n}\left(\left\{ \frac{1}{n}\Phi_n \approx E \right\} \right) \approx e^{n \cdot s(E) + o(n)}.$$

We also have that

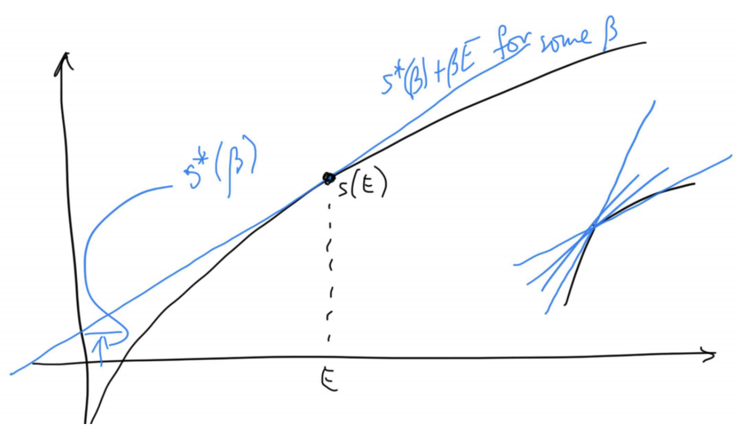
$$s(E) = \inf_{\beta \in \mathbb{R}} \{ s^*(\beta) + \beta E \},$$

$$s^*(\beta) = \sup_{E \geq 0} \{ s(E) - \beta E \} = \log \int e^{-\beta\varphi} \, d\lambda,$$

which is assumed to be $< \infty$ for all $\beta > 0$. Next, we need to understand where these inf and sup are achieved.

12.2 Supporting tangents and conjugacy between β and E

Definition 12.1. A **supporting tangent** to s at E is a line touching the graph of s at E and bounding from above.

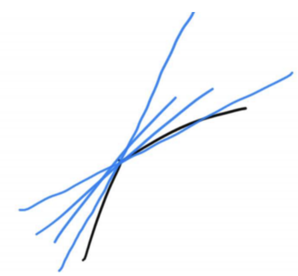


Its slope β must satisfy

$$s(E') \le s(E) + \beta(E' - E) \qquad \forall E'.$$

Equivalently,

$$D_+ s(E) \le \beta \le D_- s(E)$$



or

$$s(E) = s^*(\beta) + \beta E.$$

Up to a sign, this last equation is *symmetric* between "conjugate variables" β and E:

$$s(E) + (-s^*(\beta)) = \beta E.$$

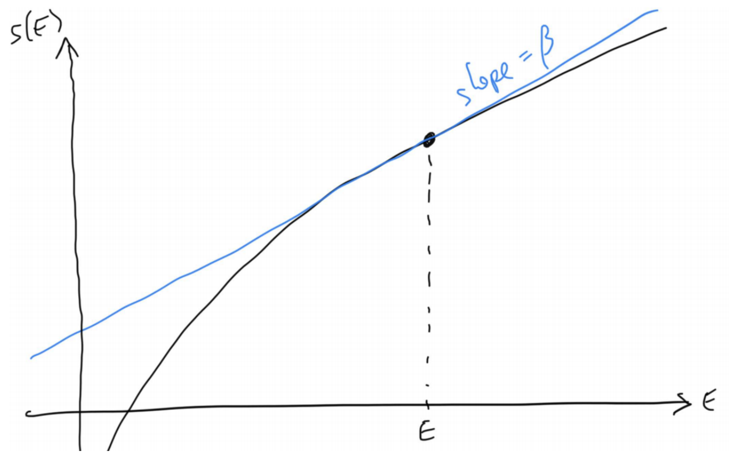
Here, s and $(-s^*)$ are both upper semicontinuous, and they play the same role in this equation. So, by symmetry, β is a slope for a supporting tangent line to s at E iff E is a slope for a supporting tangent line to $-s^*$ at β. That is,

$$D_+ s(E) \le \beta \le D_- s(E) \iff D_- s^*(\beta) \le -E \le D_+ s^*(\beta).$$

This is the key observation for deriving smoothness and differentiability properties of s from those of s^*.

12.3 Leveraging conjugacy to prove differntiability and strict convexity of s

Here is our picture relating s and s^*:



Here are some main features to be proved about this picture:

Proposition 12.1.

$$s(E) \to \begin{cases} \infty & \textit{as } E \to \infty \\ \log \lambda(\{\varphi = 0\}) & \textit{as } E \downarrow 0. \end{cases}$$

The first case implies s is strictly increasing. Also, it could be in this picture (if $\lambda(\{\varphi = 0\}) = 0$) that the graph gets steeper and steeper and never hits the vertical axis.

Proof. First, we have

$$s(E) = \inf_{\beta > 0} \Big\{ \underbrace{\log \int e^{-\beta\varphi}\, d\lambda}_{s^*(\beta)} + \beta E \Big\}.$$

First, here are some properties of s^*:

$$s^*(\beta) \to \begin{cases} \log \lambda(\{\varphi = 0\}) & \text{as } \beta \to \infty \\ \infty & \text{as } \beta \downarrow 0. \end{cases}$$

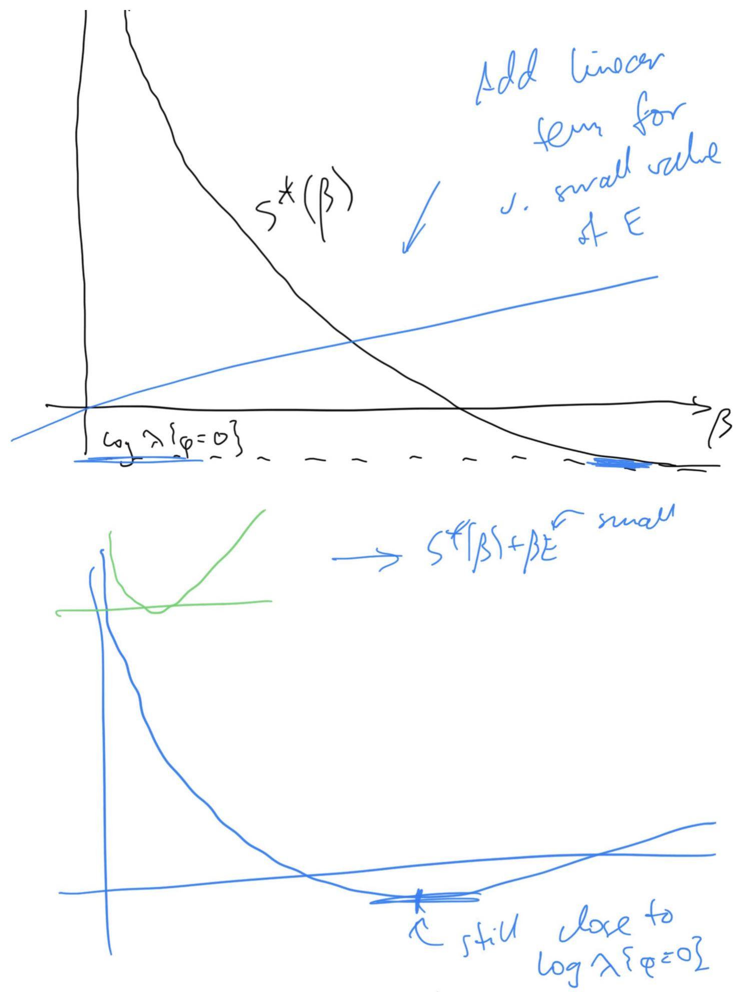
The first of these follows since $\varphi \geq 0$, $\beta_1 > \beta_2 > 0$ implies $e^{-\beta_1\varphi} \leq e^{-\beta_2\varphi}$. As $\beta \to \infty$, $e^{-\beta\varphi} \downarrow \mathbb{1}_{\{\varphi=0\}}$. By the dominated convergence theorem, $s^*(\beta) \to \log \int \mathbb{1}_{\{\varphi=0\}}\, d\lambda = \log \lambda\{\varphi = 0\}$.

Secondly, we have $\lambda(\{\varphi \leq M\}) \to \infty$ as $M \to \infty$, so for all $K > 0$, pick M so that $\lambda(\{\varphi \leq M\}) \geq K$. Now pick β so small that $e^{-\beta M} \geq 1/2$, so now

$$\begin{aligned} s^*(\beta) &= \log \int e^{-\beta\varphi}\, d\lambda \\ &= \log \int_{\{\varphi \leq M\}} e^{-\beta M}\, d\lambda \end{aligned}$$

$$
\begin{aligned}
&\geq \log\left(\frac{1}{2}\lambda(\{\varphi \leq M\})\right) \\
&\geq \log\left(\frac{K}{2}\right) \\
&\xrightarrow{K\to\infty} \infty.
\end{aligned}
$$

For the rest, here are some pictures (which can be justified with some εs and δs):



So $s(E) = \min_{\beta>0}\{s^*(\beta) + \beta E\}$ is close to $\inf_{\beta>0} s^*(\beta) = \log\lambda(\{\varphi = 0\}) = \lim_{E\downarrow 0} s(E)$ if E is small enough. Similarly, if E is very big,

$$s(E) = \min_{\beta>0}\{s^*(\beta) + \beta E\} \to \infty.$$

as $E \to \infty$. $\square$

Lemma 12.1. *s is differentiable on $(0, \infty)$ (i.e. no corners).*

Proof. s is differentiable at E iff $D_+ s(E) = D_- s(E) = s'(E)$. By our previous discussion, this is equivalent to if there is only one slope β for a supporting tangent at E. This is equivalent to if for this E, the solution to $s(E)+(-s^*(\beta)) = \beta E$ in β is unique. Equivalently, this is when $\inf_{\beta>0}\{s^*(\beta) + \beta E\}$ is achieved at exactly one β. This occurs precisely when $s^*(\cdot) + E(\cdot)$ is strictly concave where the minimum is achieved. Quantifying over E this tells us that s is differentiable if and only if s^* is strictly convex.

Now let's show that s^* is strictly convex: Suppose $\alpha > \beta > 0$ and $0 < t < 1$. Then

$$s^*(t\alpha + (1-t)\beta) = \log \int e^{(-t\alpha-(1-t)\beta)\varphi}\, d\lambda$$

Apply Hölder's inequality with exponents $1/t$ and $1/(1-t)$:

$$\leq t \log \int e^{-\alpha\varphi}\, d\lambda + (1-t) \log \int e^{-\beta\varphi}\, d\lambda,$$

with equality iff $e^{-\alpha\varphi}$ is a constant multiple of $e^{-\beta\varphi}$. This is possible only if φ is constant a.e., which is not true. □

Proposition 12.2. *s is strictly concave on $[0, \infty)$.*

Proof. As before, this is equivalent to $s^*(\beta) = \log \int e^{-\beta\varphi}\, d\lambda$ being differentiable. This holds by differentiating under the integral. □

13 Observing Macroscopic Quantities From Microscopic States

13.1 Recap

We have a phase space (M, λ) which is a σ finite but not finite measure space. The energy of one particle is $\varphi : M \to [0, \infty)$, where $\min \varphi = \operatorname{ess\,min} \varphi = 0$. Then we know that

$$\begin{aligned}\lambda^{\times n}\left(\left\{(p_1, \ldots, p_n) \in M^n : \frac{1}{n}\Phi_n(p_1, \ldots, p_n) := \frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in I\right\}\right)\\ = \exp\left(n \cdot \sup_{x \in I} s(x) + o(n)\right),\end{aligned}$$

where

$$s(x) = \inf_{\beta > 0}\{s^*(\beta) + \beta x\}.$$

We also have the Fenchel-Legendre transform

$$s^*(\beta) = \log \int e^{-\beta\varphi}.$$

β achieves equality in the definition of s

$$\begin{aligned}&\iff s \text{ has a tangent of slope } \beta \text{ at } x\\ &\iff D_+ s(x) \le \beta \le D_- s(x)\\ &\iff s^*(\beta + (-s(x)) = -\beta x\\ &\iff D_- s^*(\beta) \le -x \le D_+ s^*(\beta)\\ &\iff s^* \text{ has a tangent of slope } -x \text{ at } \beta.\end{aligned}$$

Using s^*, we can prove:

- $$s^*(\beta) \to \begin{cases}\log \lambda(\{\varphi = 0\}) & \beta \to \infty\\ \infty & \beta \downarrow 0.\end{cases}$$
- s^* is strictly decreasing and strictly convex.
- s^* is differentiable on $(0, \infty)$.
- $$s(x) \to \begin{cases}\log \lambda(\{\varphi = 0\}) & x \downarrow 0\\ \infty & x \to \infty.\end{cases}$$
- s is strictly increasing and strictly concave.
- s is differentiable on $(0, \infty)$.

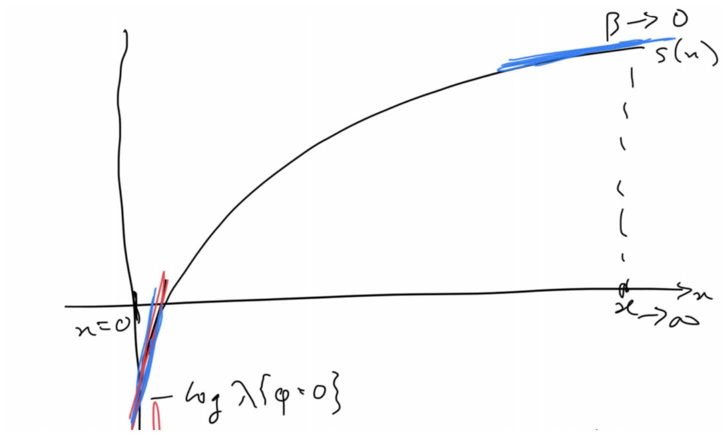
13.2 Behavior of s'

Let's analyze the behavior of s':

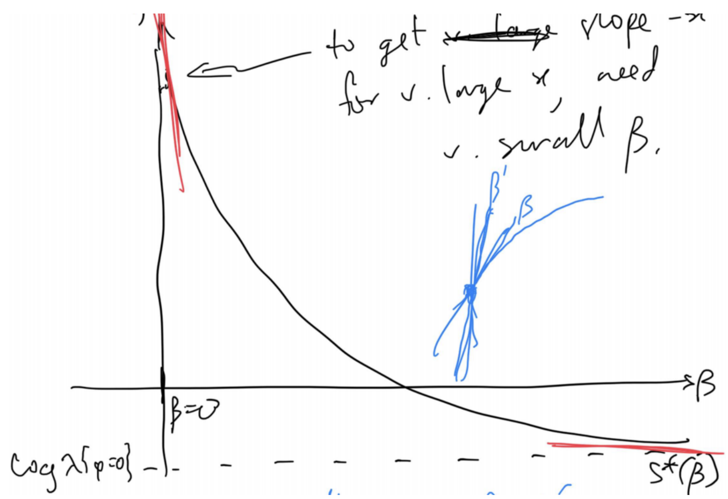
Proposition 13.1.

$$s'(x) \to \begin{cases} 0 & x \to \infty \\ \infty & x \to 0. \end{cases}$$

Instead of a formal proof, here are some pictures. Look at the possible slopes we can get for points on the graph of s and how they correspond to slopes for points on the graph for s^*.



To get slope $-x$ for very large x in the graph of s^*, we need very small β.



13.3 Observing macroscopic quantities froms microscopic states

Now imagine we are looking at some other macroscopic observable quantity of the microscopic state $(p_1, \dots, p_n) \in M^n$. We will study functions for the form

$$\Psi_n(p_1, \dots, p_n) = \sum_{i=1}^n \psi(p_i).$$

If $M = \mathbb{R}^3 \times \mathbb{R}^3$, we could take $\psi(r, p) = \mathbb{1}_D(r)$, which indicates whether a particle is in D or not in D; then Ψ_n would be the total number of particles in D.

We need some regularity. A simple sufficient condition is that ψ is bounded. A weaker but still sufficient condition is that for every $\beta > 0$, there is an $\varepsilon > 0$ such that $\int e^{-\beta\varphi} e^{-\gamma\psi}\, d\lambda < \infty$ for all $\gamma \in (-\varepsilon, \varepsilon)$.

Let's assume ψ is bounded, and we'll ask about the distribution of Ψ_n on the approximate level set $\{\frac{1}{n}\Phi_n \in I\}$, where I is a small interval. We need to compare $\lambda^{\times n}(\{\frac{1}{n}\Phi_n \in I\})$ and $\lambda^{\times n}(\{\frac{1}{n}\Phi_n \in I, \frac{1}{n}\Psi_n \in J\})$. We use the generalized type-counting machinery with $\mathbb{R}^2$ to get an asymptotic for this:

$$\begin{aligned}\lambda^{\times n}\left(\left\{\frac{1}{n}\Phi_n \in I, \frac{1}{n}\Psi_n \in J\right\}\right) &= \lambda^{\times n}\left(\left\{(p_1, \dots, p_n) \in M^n : \frac{1}{n}\sum_{i=1}^n (\varphi(p_i), \psi(p_i)) \in I \times J\right\}\right) \\ &= \exp\left(n \cdot \sup_{(x,y) \in I \times J} \widetilde{s}(x, y) + o(n)\right),\end{aligned}$$

where $\widetilde{s}(x, y) : \mathbb{R}^2 \to [-\infty, \infty)$ is an upper semicontinuous, concave function with

$$\widetilde{s}(x, y) = \inf_{\beta, \gamma}\{\widetilde{s}^*(\beta, \gamma) + \beta x + \gamma y\}.$$

and Fenchel-Legendre transform

$$\widetilde{s}^*(\beta, \gamma) = \log \int e^{-\beta\varphi} e^{-\gamma\psi}\, d\lambda.$$

Here, we assume ψ is bounded, $|\psi| \le M$, so

$$\widetilde{s}^*(\beta, \gamma) = \begin{cases} \infty & \beta = 0 \\ < \infty & \beta > 0. \end{cases}$$

Here, $\widetilde{s}(x, y) \le s(x)$ for all $y \in \mathbb{R}$. We want to find a y_0 such that $\widetilde{s}(x, y_0) = s(x)$ and $\widetilde{s}(x, y) < s(x)$ for any other y. This will tell us that conditioned on Φ being x, we are likely to have Ψ be y_0 and not likely to have any other y. We have

$$s(x) = \inf_{\beta > 0}\left\{\log \int e^{-\beta\varphi}\, d\lambda + \beta x\right\},$$

which is greater than or equal to

$$\widetilde{s}(x,y) = \inf_{\beta>0,\gamma\in\mathbb{R}} \left\{ \log \int e^{-\beta\varphi} e^{-\gamma\psi}\, d\lambda + \beta x + \gamma y \right\}.$$

Lemma 13.1. *$\widetilde{s}(x, y_0) = s(x)$ and $\widetilde{s}(x,y) < s(x)$ for any other y, where*

$$y_0 = \frac{\int \psi e^{-\beta\varphi}\, d\lambda}{\int e^{-\beta\varphi}\, d\lambda} = \langle \psi, \mu_\beta \rangle$$

and

$$d\mu_\beta(p) = \frac{e^{-\beta\varphi(p)\, d\lambda(p)}}{\int e^{-\beta\varphi}\, d\lambda}$$

*is the **Gibbs measure** obtained from λ, φ, β.*

Proof. First, s is differentiable, so for every $x > 0$, there is a unique $\beta > 0$ such that $s(x) = \log \int e^{-\beta\varphi}\, d\lambda + \beta x$. To achieve $\widetilde{s}(x, y_0) = s(x)$, we must have that the function $\gamma \mapsto \log \int e^{-\beta\varphi} e^{-\gamma\psi}\, d\lambda + \beta x + \gamma y_0$ achieves its minimum uniquely at $\gamma = 0$. This function of γ is convex (by Hölder), strictly convex if ψ is not a.s. constant, and differentiable. Assuming ψ is not a.s. constant, we need y_0 such that

$$\frac{\partial}{\partial\gamma} \left\{ \log \int e^{-\beta\varphi} e^{-\gamma\psi}\, d\lambda + \beta x + \gamma y_0 \right\} = 0$$

at $\gamma = 0$. This is the derivative of the log of the moment generating function. Differentiate under the integral to get

$$\frac{\partial}{\partial\gamma} \log \int e^{-\beta\varphi} e^{-\gamma\psi}\, d\lambda = \left. \frac{\int -\psi e^{-\beta\varphi} e^{-\gamma\psi}\, d\lambda}{\int e^{-\beta\varphi} e^{\gamma\psi}\, d\lambda} \right|_{\gamma=0} = -\langle \psi, \mu_\beta \rangle.$$

So $\frac{\partial}{\partial\gamma}[\cdots]|_{\gamma=0} = -\langle \psi, \mu_\beta \rangle + y_0$, and this equals 0 iff $y_0 = \langle \psi, \mu_\beta \rangle$. □

Corollary 13.1.

$$\lambda^{\times n} \left(\left\{ \left| \frac{1}{n}\Psi_n - \langle \psi, \mu_\beta \rangle \right| > \varepsilon \right\} \;\middle|\; \left\{ \frac{1}{n}\Phi_n \in I \right\} \right) \le e^{-c\cdot n + o(n)},$$

where c is a constant, I is a short enough interval containing x, and we are using conditional probability notation.

Remark 13.1. Given $\frac{1}{n}\Phi_n \approx x$, we found that

$$\begin{aligned} \Psi_n &\approx n(\text{its average over } \{\tfrac{1}{n}\Phi_n \approx n\}^{10}) \\ &\approx n\langle \psi, \mu_\beta \rangle \end{aligned}$$

$$
\begin{aligned}
&= \langle \psi(p_1) + \cdots + \psi(p_n), \mu_\beta^{\times n} \rangle \\
&= \int \Psi_n \, d\mu_{\beta,n},
\end{aligned}
$$

where

$$
d\mu_{\beta,n}(p_1, \ldots, p_n) = \frac{e^{-\beta\Phi_n(p_1,\ldots,p_n)} \, d\lambda^{\times n}(p)}{\int e^{-\beta\Phi_n} \, d\lambda^{\times n}} = \mu_\beta \times \cdots \times \mu_\beta
$$

is called the **canonical ensemble measure**.

[10] This is called the **microcanonical ensemble**.

14 Intro to Interacting Particles and Temperature

14.1 Properties of systems of non-interacting particles

Let's recap what we've proved so far about systems of n non-interacting particles. We have the phase space (M, λ), where $(M^n, \lambda^{\times n})$ describes the total state of all n particles. We have shown that

$$\lambda^{\times n}(\tfrac{1}{n}\Phi_n \in I) = \exp\left(n \cdot \sup_{x \in I} s(x) + o(n)\right),$$

where

$$s(x) = \inf_{\beta > 0}\{s^*(\beta) + \beta x\}$$

can be expressed in terms of its Fenchel-Legendre transform:

$$\begin{aligned} s^*(\beta) &= \underbrace{\log \int e^{-\beta\varphi}\, d\lambda}_{\log Z(\beta)} \\ &= \frac{1}{n} \log \int_{M^n} e^{-\beta\Phi_n}\, d\lambda^{\times n} \\ &= \frac{1}{n} \log Z_n(\beta). \end{aligned}$$

Here $Z_n(\beta)$ is called the **partition function**.

We have also proven some properties about $s : \mathbb{R} \to [-\infty, \infty)$ and s^* using their relationship to each other:

- $s \equiv -\infty$ on $(-\infty, 0)$.
- $$s(x) \to \begin{cases} \infty & x \to \infty \\ \text{const or } -\infty & x \downarrow 0. \end{cases}$$
- s is strictly concave (iff s^* is diferentiable) and differentiable (iff s^* is concave)
- $$s'(x) \to \begin{cases} 0 & x \to \infty \\ \infty & x \downarrow 0. \end{cases}$$

Define the **microcanonical ensemble**[11]

$$d\mu_{n,I}(p_1, \dots, p_n) = \frac{\mathbb{1}_{\{\frac{1}{n}\Phi_n \in I\}}(p_1, \dots, p_n)\, d\lambda(p_1) \cdots d\lambda(p_n)}{\lambda^{\times n}(\{\frac{1}{n}\Phi_n \in I\})}.$$

[11] The term "ensemble" goes back to Gibbs, who used it before measure theory and its terminology were around.

For $\beta > 0$,

$$d\mu_\beta(p) = \frac{1}{Z(\beta)} e^{-\beta\varphi(p)}\, d\lambda(p)$$

is the normalized **Gibbs measure**.

Then

$$\begin{aligned} d\mu_{n,\beta}(p_1, \ldots, p_n) &= d\mu_\beta(p_1) \cdots d\mu_\beta(p_n) \\ &= \frac{1}{Z(\beta)^n} e^{-\beta\varphi(p_1)}\, d\lambda(p_1) \cdots e^{-\beta\varphi(p_n)}\, d\lambda(p_n) \\ &= \frac{e^{-\beta\Phi_n(p_1,\ldots,p_n)}\, d\lambda^{\times n}(p_1, \ldots, p_n)}{Z_n(\beta)} \end{aligned}$$

is the **canonical ensemble**, which applies to all the particles at once.

Last time, we said that

$$\mu_{n,I}(\{\tfrac{1}{n}\Psi_n \approx \langle \psi, \mu_\beta \rangle\}) \approx 1,$$

where $\Psi_n = \psi(p_1) + \cdots + \psi(p_n)$, I is a short interval around E, and β is chosen so that $\langle \varphi, \mu_\beta \rangle = E$. We have that

$$\mu_{n,I}(\{\tfrac{1}{n}\Psi_n \approx \tfrac{1}{n}\langle \Psi_n, \mu_{n,\beta} \rangle\}) \approx 1,$$

so there is an equivalence of the canonical ensemble and the microcanonical in the limit $n \to \infty$.

14.2 Wishlist for extending properties to interacting systems of particles

Suppose we have some sequence of σ-finite but not finite measure spaces (M_n, λ_n) with "total energy" functions $\Phi_n : M_n \to [0, \infty)$. Then we want

$$\lambda_n(\tfrac{1}{n}\Phi_n \in I) = \exp\left(n \cdot \sup_{x \in I} s(x) + o(n) \right),$$

where we can hopefully define s as usual and

$$s^*(\beta) = \lim_{n\to\infty} \frac{1}{n} \log \int_{M^n} e^{-\beta\Phi_n}\, d\lambda = \lim_{n\to\infty} \frac{1}{n} \log Z_n(\beta)$$

We will retain the following properties of s and s^*:

- $s \equiv -\infty$ on $(-\infty, 0)$.
- $$s(x) \to \begin{cases} \infty \text{ (sometimes)} & x \to \infty \\ \text{const or } -\infty & x \downarrow 0. \end{cases}$$

- s will not always be strictly concave but will usually be differentiable.
- $$s'(x) \to \begin{cases} 0 \text{ (not always)} & x \to \infty \\ \infty \text{ (usually)} & x \downarrow 0. \end{cases}$$

We can also define the canonical and microcanonical ensembles and hope for an equivalence of ensembles in the limit, as well.

14.3 Defining temperature

What is temperature? When two bodies of different temperature come into contact for a prolonged period of time, they will eventually both reach some equilibrium temperature. Temperature is a quantity that determines when bodies/systems are in thermal equilibrium. There is a canonical "thermodynamic temperature" (which can be measured, for example, by a mercury thermometer) which we want to be able to define.[12]

To interpret this, consider two systems $(M_n, \lambda_n), \Phi_n : M_n \to [0, \infty)$ and $(\widetilde{M}_n, \widetilde{\lambda}_n), \widetilde{\Phi}_n : \widetilde{M}_n \to [0, \infty)$. There is the combined system is $(M_n \times \widetilde{M}_n, \lambda_n \times \widetilde{\lambda}_n)$ with total energy $\Phi_n(p) + \widetilde{\Phi}_n(\widetilde{p})$. Note that once again we are assuming very weak interaction between the systems in terms of energy. If we condition on

$$\{(p, \widetilde{p}) \in M_n \times \widetilde{M}_n : \tfrac{1}{2n}(\Phi_n(p) + \widetilde{\Phi}_n(\widetilde{p})) \in I\},$$

what is the typical split of total energy between Φ_n and $\widetilde{\Phi}_n$?

Suppose

$$\lambda_n(\{\tfrac{1}{n}\Phi_n \in I\}) = \exp\left(n \cdot \sup_I s + o(n)\right),$$

$$\widetilde{\lambda}_n(\{\tfrac{1}{n}\widetilde{\Phi}_n \in I\}) = \exp\left(n \cdot \sup_I \widetilde{s} + o(n)\right).$$

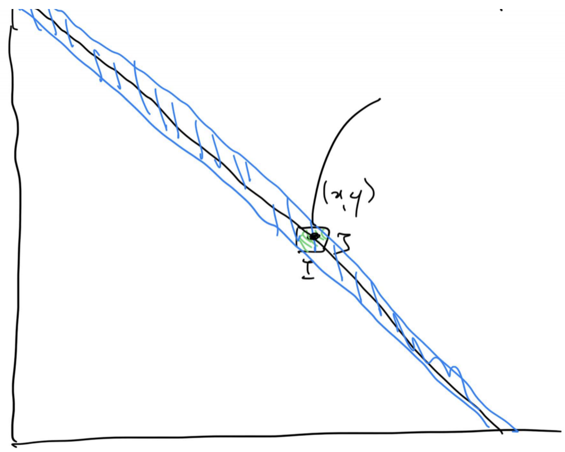
Then consider $(\Phi_n(p), \widetilde{\Phi}_n(\widetilde{p})) : M_n \times \widetilde{M}_n \to [0, \infty)^2$ with

$$\lambda_n \times \widetilde{\lambda}_n(\{(\tfrac{1}{n}\Phi_n, \tfrac{1}{n}\widetilde{\Phi}_n) \in I \times J\}) = \exp\left(n \cdot \sup_{x \in I, y \in J}(s(x) + \widetilde{s}(y)) + o(n)\right)$$

This is the same when $I \times J$ are replaced by general open, convex sets.

In the following picture of the microcanonical ensemble, conditioning on $\frac{1}{2n}(\Phi_n(p) + \widetilde{\Phi}_n(\widetilde{p})) \in \operatorname{int} K$ means conditioning on the blue strip:

[12] Historically, the mysterious quantity "entropy" was discovered first, and temperature was defined relative to it.



The most likely energy split occurs where $s(x)+\widetilde{s}(y)$ is maximized on this strip. Suppose the strip is very thin around $\{x+y=E\}$. We want to maximize $s(x)+\widetilde{s}(E-x)$ as x varies in $[0,E]$. If $s,\widetilde{s}$ are differentiable, this requires

$$\frac{\partial}{\partial x}[s(x)+\widetilde{s}(E-x)]=0,$$

i.e. $s'(x)=\widetilde{s}'(E-x)$. That is, systems are in thermal equilibrium at individual energies x and $y=E-x$ only if $\beta=s'(x)=\widetilde{s}'(y)=\widetilde{\beta}$. This is the unique maximizer, so this is "if and only if" in the case where $s,\widetilde{s}$ are strictly concave.

So we define the **thermodynamic temperature** of the system with entropy function s to be

$$T=\frac{1}{\beta}=\frac{1}{s'(x)}.$$

Here, β is known as the **inverse temperature**.

15 Models With Additional Thermodynamical Parameters

15.1 Recap

Suppose we have two thermodynamical systems with energy functions $\Phi_n : M_n \to [0, \infty)$ and $\widetilde{\Phi}_n : \widetilde{M_n} \to [0.\infty)$. Then

$$\lambda_n \left(\left\{ \frac{1}{n} \Phi_n \in I \right\} \right) = \exp \left(n \cdot \sup_{x \in I} s(x) + o(n) \right),$$

$$\widetilde{\lambda}_n \left(\left\{ \frac{1}{n} \Phi_n \in I \right\} \right) = \exp \left(n \cdot \sup_{x \in I} \widetilde{s}(x) + o(n) \right).$$

We are studying what happens when we put the systems in thermal contact and constrain them so that the total energy is $\approx nE$. In equilibrium, energy split between the systems is decided by maximizing $s(x) + \widetilde{s}(E - x)$. If the exponent functions $s, \widetilde{s}$ are differentiable, then the condition is

$$s'(x) = \widetilde{s}'(E - x).$$

Hence, we denote $\frac{1}{s'(x)}$ as the **thermodynamic temperature** of the first system at energy x. In physics, the dependence of s on energy per particle x and any other parameters in the model is known as the **fundamental relation** of the system.[13]

15.2 Fundamental relation and equivalence of ensembles

In a laboratory, suppose we constrain the temperature of a system to be $T = 1/\beta$ (instead of controlling the total energy). Now we can look for energy per particle as a the root of the equation $s'(x) = \beta$ if you know the fundamental relation of the system.[14]

Alternatively, if we are also expecting equivalence of ensembles, we can make predictions about other thermodynamical quantities based directly on the canonical ensemble/the Gibbs measure

$$d\mu_{n,\beta}(p) = \frac{e^{-\beta \Phi_n(p)} \, d\lambda_n(p)}{Z_n(\beta)}.$$

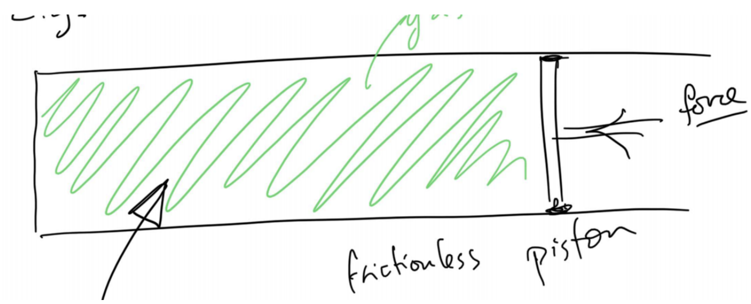
Note that the microcanonical ensemble is determined by the total energy you're constraining around, while the canonical ensemble is determined by the temperature. So in this case, we can determine the behavior of the system using temperature, which is the quantity we can actually control in a lab.

[13] In practice, it may be easier to describe s^* instead.

[14] You can run into trouble here if s has any flat regions.

15.3 Gas in a piston chamber

Example 15.1. Suppose we have gas in a chamber, where at one end, there is a piston. Assume the piston is frictionless.



If the piston can slide back and forth, then it will be not moving when the gas inside is at the same pressure as the atmospheric pressure of the air outside the box. Thus, our system has an additional parameter, which can be modeled using the force on the piston, the pressure of the gas, or the total volume v of the chamber.

We'll begin our discussion with this example as the model, and then we will abstract out what we need to discuss general models with additional parameters. For simplicity, we will assume the area of the piston equals 1. How can we understand the dependence between the pressure and the volume from the fundamental relation?

Let's consider n "non-interacting" classical particles. The total energy of a particle with position r and momentum p is

$$\varphi(r,p) = \varphi_{\text{pot}}(r) + \frac{1}{2}|p|^2,$$

and the total energy of the system is

$$\Phi_n(r_1,\ldots,r_n,p_1,\ldots,p_n) = \sum_{i=1}^{n} \varphi(r_i,p_i).$$

We will now include the volume v as a parameter in all these functions. Assuming the particles bounce off the walls elastically[15], we want to relate the pressure of the gas to the volume of the chamber. Forces are obtained as *gradients* of total potential energy, so the force on the piston is

$$\frac{\partial}{\partial v}\Phi_n(v,r_1,\ldots,r_n,p_1,\ldots,p_n) = \sum_{i=1}^{n} \frac{\partial \varphi_{\text{pot}}(v,r_i)}{\partial v}.$$

[15]In reality, there is a repulsive force that is weak unless the particles are very close together, in which case it becomes very strong.

For most states $(r_1, \dots, r_n, p_1, \dots, p_n)$, this will be accurately predicted by

$$\left\langle \sum_{i=1}^n \frac{\partial \varphi_{\text{pot}}}{\partial v}(v, \cdot), \mu_{n,I} \right\rangle,$$

where $\mu_{n,I}$ is the microcanonical ensemble. Or, if we have equivalence of enesmbles, this is predicted by

$$\left\langle \sum_{i=1}^n \frac{\partial \varphi}{\partial v}(v, \cdot), \mu_{n,\beta} \right\rangle,$$

where $\mu_{n,\beta}$ is the canonical ensemble.

To understand this, look at

$$\begin{aligned}
\frac{\partial}{\partial v} s^*(v, \beta) &= \frac{\partial}{\partial v} \left\{ \frac{1}{n} \log \int e^{-\beta \Phi_n(v,\cdot)} \, d\lambda_n \right\} \\
&= \frac{1}{n} \frac{\int -\beta \frac{\partial \Phi_n}{\partial v} e^{-\beta \Phi_n} \, d\lambda_n}{\int e^{-\beta \Phi_n} \, d\lambda_n} \\
&= -\frac{1}{n} \beta \left\langle \frac{\partial \Phi_n}{\partial v}, \mu_{n,\beta} \right\rangle.
\end{aligned}$$

So

$$\frac{\partial}{\partial v} F(v, \beta) = \left\langle \frac{\partial \Phi_n}{\partial v}, \mu_{n,\beta} \right\rangle,$$

where

$$F(v, \beta) = -\frac{n}{\beta} s^*(v, \beta) = Tns^*(v, \beta).$$

This is known as the **free energy**, the **Helmholtz function**, or the **Helmholtz free energy**. So the pressure is

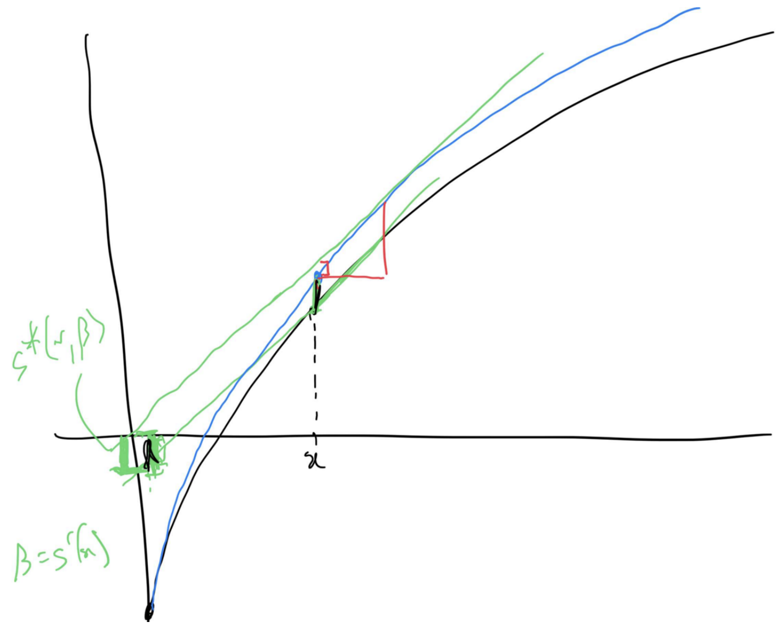
$$P = \frac{\partial}{\partial v} F(v, \beta) = \frac{\partial}{\partial v} [-T \log Z_n(v, \beta)].$$

Here is another way to get this in terms of s itself:

Lemma 15.1. *If $s(v, x)$ is strictly concave in β and C^2 in both parameters, then*

$$\frac{\partial}{\partial v} s^*(v, \beta) = \frac{\partial}{\partial x} s(v, x).$$

Proof. Here is a proof by picture. Draw $s(v, x)$ and $s(v + dv, x)$:



Recall that $s^*(v, \beta)$ is the vertical-axis intercept when we draw the tangent line to s at x. Here, $\beta = s'(x)$. On the tangent line to the new blue curve in this picture, the slope maybe changed a little bit. Instead, find a place where a tangent of the same slope hits the blue curve and consider the difference of those intercepts. The difference between these changes in the intercept end up being a second order difference, so they disappear in the derivative. $\square$

Once we know s, this leads to equations that relate V, P, T or V, P, E, etc. Once you have any two parameters, you can solve for the third. Any such equation is called an **equation of state**.

16 The Ideal Gas Law and Discretization

16.1 Recap

Last time, we set up a model with total energy $\Phi_n : I \times M_n \to [0, \infty)$, where I is an set specifying an extra parameter, such as the volume v of the enclosing system. Let $S_n(v, x) := \log \lambda_n(\{\frac{1}{n}\Phi_n(v, \cdot) \approx x\})$, and we assume that $\frac{1}{n} S_n(v, x) \to s(v, x)$, where s is concave in x, etc. We have the partition function $Z_n(v, \beta) = \int e^{-\beta \Phi_n} \, d\lambda_n$, and we assume that $\frac{1}{n} \log Z_n(v, \beta) \to s^*(v, \beta)$, where $\beta > 0$. These are related by using the Fenchel-Legendre transform:

$$s(v, x) = \inf_{\beta > 0} \{s^*(v, \beta) + \beta x\},$$

where the inf is achieved at $\beta = \frac{\partial}{\partial x} s(v, x)$.

In our piston chamber example, the "pressure" P associated to v was

$$\left\langle \frac{\partial \Phi_n(v, \cdot)}{\partial v}, \mu_{n,\beta} \right\rangle = \frac{\partial}{\partial v} \underbrace{[-T \log Z_n(v, \beta)]}_{=F(v,\beta)},$$

where $F(v, \beta)$ is the **Helmholtz free energy** and $T := 1/\beta$ is the **thermodynamic temperature**.

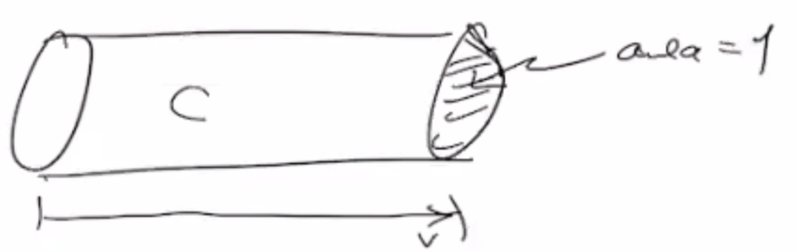
$$\approx \frac{\partial}{\partial v}[-T n s^*(v, \beta)]$$

Here are the assumptions we have been making here:

- For most microscopic states, this quantity stays close to its average with respect to to the microcanonical ensemble.
- We can replace the microcanonical ensemble by the canonical ensemble.
- We are assuming that we can interchange integration and differentiation in the above calculation (this is fortunately not hard to justify using convexity arguments).

16.2 The ideal gas law

Assume a gas of n (mass 1) non-interacting particles is in a region (cylinder C with cross-sectional area 1 and length v).



The state is $(r_1, \dots, r_n, p_1, \dots, p_n) \in (\mathbb{R}^3 \times \mathbb{R}^3)^n$, and the potential energy is

$$\varphi_{\text{pot}}(r) = \begin{cases} 0 & r \in C \\ \infty & r \notin C. \end{cases}$$

The total energy is

$$\Phi_n(r_1, \dots, r_n, p_1, \dots, p_n) = \begin{cases} \sum_{i=1}^n \frac{1}{2}|p_i|^2 & r_i \in C\ \forall i \\ \infty & \text{otherwise.} \end{cases}$$

Our goal is to understand the pressure in terms of temperature and volume. We understand this through

$$\begin{aligned} Z_n(v, \beta) &= \int \cdots \int e^{-\beta \Phi_n}\, dm_3^{\times n}(r_1, \dots, r_n)\, dm_3^{\times n}(p_1, \dots, p_n) \\ &= \int_{C^n} dr_1 \cdots dr_n \cdot \int \cdots \int e^{-\beta \sum_{i=1}^n |p_i|^2/2}\, dp_1 \cdots dp_n \\ &= v^n \cdot \left(\int e^{-\beta |p|^2/2}\, dp \right)^n \\ &= v^n \cdot \left(\left(\frac{2\pi}{\beta} \right)^{3/2} \right)^n. \end{aligned}$$

Then

$$F_n = T \log Z_n = Tn \log v - \frac{3Tn}{2} \log(2\pi T),$$

and so the pressure is

$$P = \frac{\partial F_n}{\partial v} = \frac{Tn}{v}.$$

Thus, we get the **Ideal Gas Law**:[16]

$$PV = nT.$$

In Gay-Lussac's version of this law, he derived a slightly more complicated-looking expression

$$PV = \text{const} \cdot n \cdot (\text{const} + \theta),$$

where θ is the Celsius temperature and the constant next to it is $\approx 273.16^{\text{o}}$ C.

[16]If you are using standard physical units, you need a constant in here to facilitate the conversion of units.

16.3 Discretization in models with interaction

Suppose we have n particles in a region $R_n \subseteq \mathbb{R}^3$ with volume $|R_n|$. Then the position is $(r_1, \ldots, r_n) \in R_n^n$, and

$$\Phi_n(r_1, \ldots, r_n) = \sum_{i=1}^{n} \varphi_{\text{pot}}(r_i) + \sum_{i \neq j} \varphi_{\text{int}}(r_i - r_j) + \sum_{i=1}^{n} \frac{1}{2}|p_i|^2.$$

Here, the measure is $\lambda_n = m_3^{\times n} \times m_3^{\times n}$. The entropy

$$s_n(x) = \log \lambda_n(\{(r_1, \ldots, r_n, p_1, \ldots, p_n) : \tfrac{1}{n}\Phi_n \approx x\}).$$

We have a new kind of limit: The region should depend on n, so that $\frac{|R_n|}{n} \to$ some limit $=$ v. This little v is called the **molar volume**.

Note that

$$\lambda_n(\{(r_1, \ldots, r_n) \in R_n^n\}) = |R_n|^n \sim (nv)^n = n^n v^n.$$

This blows up with n. The solution is to not care about the ordering of the positions of the particles (treating the particles as indistinguishable). Thus, we actually define $\lambda_n = \frac{1}{n!} m_3^{\times n} \times m_3^{\times n}$, and this quantity $\sim (v/e)^n$.

With this choice of λ_n now look at

$$\begin{aligned} Z_n(\beta) &= \int_{R_n^n} e^{-\beta \sum \varphi_{\text{pot}}(r_i) - \beta \sum \varphi_{\text{int}}(r_i - r_j)}\, dr_1 \cdots dr_n \cdot \int_{(\mathbb{R}^3)^n} e^{-(\beta/2)(|p_1|^2 + \cdots + |p_n|^2)}\, dp_1 \cdots dp_n \\ &= \frac{1}{n!} \int_{R_n^n} e^{-\text{potential terms}}\, dr_1 \cdots dr_n \cdot \left(\frac{2\pi}{\beta}\right)^{3n/2}. \end{aligned}$$

So we get

$$\log Z_n(\beta) = \log\left(\frac{1}{n!} \int_{R_n^n} (\cdots)\, dr_1 \cdots dr_n\right) + \frac{3n}{2} \log \frac{2\pi}{\beta}.$$

To "discretize" such a model, focus on the first term, ignore the second term, and then discretize $R_n^n \subseteq (\mathbb{R}^3)^n$ to $(R_n \cap \varepsilon\mathbb{Z}^3)^n \subseteq (\varepsilon\mathbb{Z}^3)^n$. Then we replace m_3 with counting measure times ε^3.

Next time, we will show how these considerations can allow us to derive the ideal gas law again.

17 Deriving The Ideal Gas Law With Nonconstant Volume

17.1 Recap

Last time, we had a model of n interacting particles in a region $R_n \subseteq \mathbb{R}^3$. We will keep the volume per particle constant:

$$\frac{|R_n|}{n} \to v.$$

Then the phase space is $M_n = R_n \times \mathbb{R}^3$ for positions and momenta, and the measure is $\lambda_n = \frac{1}{n!} m_3 \times m_3$, where the $\frac{1}{n!}$ shows that we are treating the particles as indistinguishable. The total energy of our particles is

$$\Phi_n(r_1, \dots, r_n, p_1, \dots, p_n) = \sum_{i=1}^n \varphi_{\text{pot}}(r_i) + \sum_{i,j} \varphi_{\text{int}}(r_i - r_j) + \sum_{i=1}^n \frac{1}{2}|p_i|^2,$$

so we have potential energy, interaction energy, and kinetic energy terms. The partition function is

$$Z_n(\beta) = \int e^{-\beta \Phi_n} \, d\lambda_n = \underbrace{\widetilde{Z}_n(\beta)}_{\text{pot. + int. energy part}} \cdot \underbrace{\left(\frac{2\pi}{\beta}\right)^{3n/2}}_{\text{kinetic energy part}},$$

where

$$\widetilde{Z}_n(\beta) = \frac{1}{n!} \int \cdots \int_{R_n^n} e^{-\beta \sum_{i=1}^n \varphi_{\text{pot}}(r_i) - \beta \sum_{i,j} \varphi_{\text{int}}(r_i - r_j)} \, dm_3^{\times n}.$$

17.2 Derivation of the ideal gas law with nonconstant volume

Now discretize[17] space: pick $\varepsilon > 0$ and let $B_n = R_n \cap \varepsilon \mathbb{Z}^3$. We will analyze the resulting approximation to $\widetilde{Z}_n(\beta)$ and then let $\varepsilon \downarrow 0$ to derive an equation of state. The approximation to $\widetilde{Z}_n(\beta)$ is to

$$\text{replace } \int \cdots \int_{R_n^n} \text{ with } \frac{1}{n!} \sum_{\substack{r_1, \dots, r_n \in B_n \\ r_1, \dots, r_n \text{ distinct}}}.$$

or equivalently with the sum

$$\sum_{\omega \subseteq B_n, |\omega| = n}.$$

We will actually use the indexing

$$\sum_{\substack{\omega \in \{0,1\}^{B_n}, \\ |\omega| = n}}, \qquad \text{where } |\omega| = \sum_{i \in B_n} \omega_i.$$

[17] Discretization is not actually necessary, but without it, our arguments will take too long for the purposes of this course.

So we need to analyze

$$\widehat{Z}_n(\beta) = \sum_{\substack{\omega \in \{0,1\}^{B_n} \\ |\omega| = n}} e^{-\beta \sum_{i \in B_n} \varphi_{\text{pot}}(i) \cdot \omega(i)} e^{-\beta \sum_{i,j \in B_n} \varphi_{\text{int}}(i-j) \omega_i \omega_j}.$$

Now let's rederive the ideal gas law again: $\varphi_{\text{pot}} = \varphi_{\text{int}} = 0$, leaving

$$\begin{aligned} \widehat{Z}_n(\beta) &= |\{\omega \in \{0,1\}^{B_n} : |\omega| = n\}| = \binom{|B_n|}{n} \\ &= \binom{|B_n|}{\frac{n}{|B_n|}|B_n|} \end{aligned}$$

where $n/|B_n| \to \varepsilon^3/v$.

$$= \exp\left(|B_n| \cdot H\left(\frac{\varepsilon^3}{v}, 1 - \frac{\varepsilon^3}{v}\right) + o(|B_n|)\right)$$

As $n \to \infty$, this looks like

$$= \exp\left(n\frac{v}{\varepsilon^3} H\left(\frac{\varepsilon^3}{v}, 1 - \frac{\varepsilon^3}{v}\right) + o(n)\right).$$

So $\frac{1}{n} \log \widehat{Z}_n(\beta) \to \frac{v}{\varepsilon^3} H(\frac{\varepsilon^3}{v}, 1 - \frac{\varepsilon^3}{v})$ as $n \to \infty$. So in this model.

$$\begin{aligned} P &= \frac{\partial}{\partial v}[-T \log \widehat{Z}_n(\beta)] \\ &= nT \frac{\partial}{\partial v}\left[\frac{v}{\varepsilon^3} H\left(\frac{\varepsilon^3}{v}, 1 - \frac{\varepsilon^3}{v}\right)\right] + o(n) \end{aligned}$$

Now we calculate

$$\begin{aligned} &\frac{\partial}{\partial v}\left[\frac{v}{\varepsilon^3}\left(-\frac{\varepsilon^3}{v} \log \frac{\varepsilon^3}{v} - \left(1 - \frac{\varepsilon^3}{v}\right) \log \left(1 - \frac{\varepsilon^3}{v}\right)\right)\right] \\ &= \frac{\partial}{\partial v}\left[\log \frac{v}{\varepsilon^3} - \left(\frac{v}{\varepsilon^3} - 1\right) \log \left(1 - \frac{\varepsilon^3}{v}\right)\right] \\ &= \frac{1}{v} - \frac{\partial}{\partial v}\left[\left(\frac{v}{\varepsilon^3} - 1\right) \log \left(1 - \frac{\varepsilon^3}{v}\right)\right], \end{aligned}$$

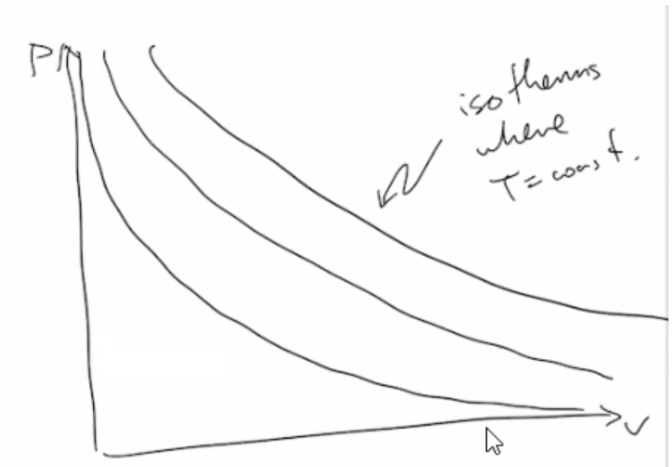
where the right term becomes negligible as $\varepsilon \to 0$ (there may be some calculation errors).

So

$$P = \frac{nT}{v} + o_{\varepsilon \downarrow 0}(n),$$

and we get the ideal gas equation of state, $PV = nT$, after letting $\varepsilon \to 0$. The value of this method is that we can vary the volume as we increase the number of particles. If we increased the number of particles without increasing the volume, the pairwise interactions of the particles may blow up. Note that most quantities here admit direct control or observation in the laboratory.

Here is how we may plot experimental data following the ideal gas law. The lines are isotherms, curves where T is held constant.

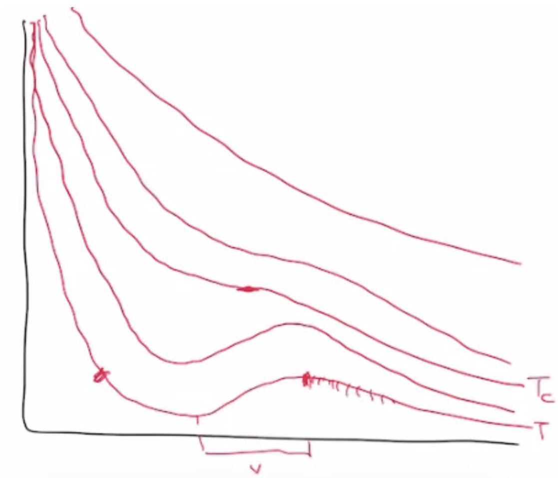


17.3 The van der Waals equation of state

In 1872, van der Waals wrote the more accurate following equation of state

$$\underbrace{\left(P + \frac{a}{v^2}\right)}_{\substack{\text{effective}\\ \text{pressure}}} \cdot \underbrace{(v - b)}_{\substack{\text{effective}\\ \text{compressible}\\ \text{volume}}} = nT,$$

where a, b are constants. This equation tells you what the effective addition to pressure a/v^2 is when the volume changes, accounting for the approximation errors in our model. This equation of state predicts real world behavior in a larger range of contexts than the ideal gas law. Here is a picture of what this equation of state predicts:



However, this predicts that at low temperatures, gasses can “catastrophically collapse.” Maxwell later adjusted the model by assuming that we adjust the curve so the total area under the curve is the total work. This is known as “Maxwell’s equal area correction.” We will discuss this story next time.

18 Deriving van der Waal's Equation

18.1 Recap

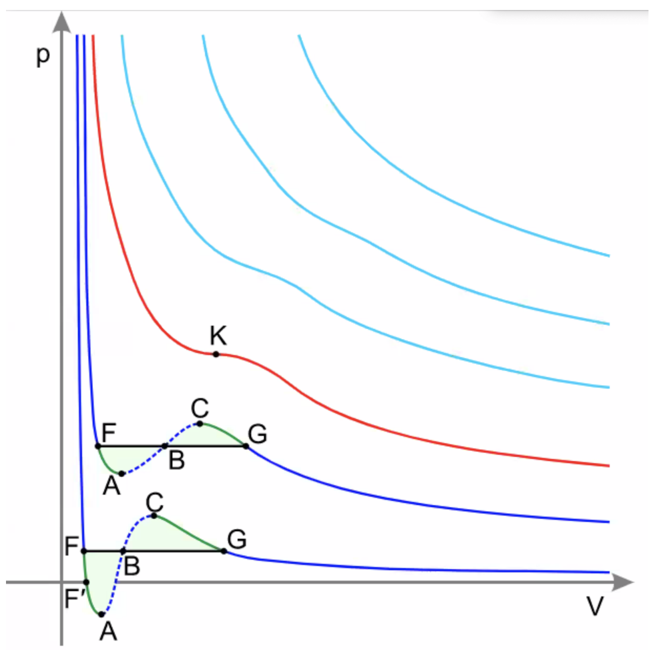
Last time, we derived the ideal gas law

$$Pv = nT,$$

where the volume is not held constant as the number n of particles increase. We mentioned van der Waal's equation, which is a better description of real gasses:

$$\left(P + \frac{a}{v^2}\right)(V - b) = nT,$$

where a and b are constants. Here is what the equation predicts:



The flat regions in this picture are the Maxwell correction to van der Waal's equation. Our next goal is to derive this equation from some simple model.

18.2 Overview of van der Waal's equation

The continuous model (with the kinetic part removed) has the partition function

$$\widetilde{Z}_n(\beta) = \frac{1}{n!} \int \cdots \int_{R_n^n} \exp\left(-\beta \sum_{i=1}^{n} \varphi_{\text{pot}}(r_i) - \beta \sum_{i,j} \varphi_{\text{int}}(r_i - r_j)\right) dm_3^{\times n}.$$

The discrete analogue is

$$\widetilde{Z}_n(\beta) = \sum_{\substack{\omega\in\{0,1\}^{B_n} \\ |\omega|=n}} \exp\left(-\beta \sum_{i\in B_n} \varphi(i)\omega_i - \beta \sum_{i,j\in B_n} \varphi_{\text{int}}(i-j)\omega_i\omega_j\right),$$

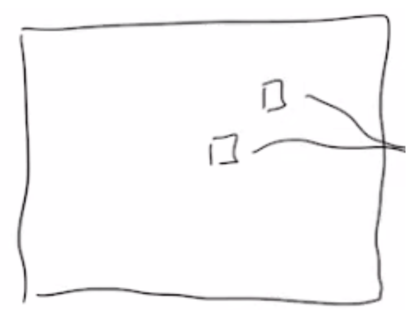
where $B_n = R_n \cap \varepsilon\mathbb{Z}^3$.

Under what conditions can we derive van der Waal's equation? We will do this for the case $b = 0$, i.e.

$$\left(P + \frac{a}{v^2}\right) v = nT.$$

(Getting the case $b > 0$ is similar but more intricate, so we will not do it for the sake of time.)

Imagine two tiny regions in a gas.



The mass in each tiny region is proportional to the density, $1/v$. So the force between the two regions is dependent on $1/v$. This gives intuition for why there should be a $1/v$ in the equation. This idea of interactions between molecules will lead to the equation.[18]

18.3 Setup and notation

To incorporate a limit of "long range forces," fix an attractive potential energy of interaction $\varphi : \mathbb{R}^3 \to [0, \infty)$ (since φ is attractive, to avoid having a negative potential, we assume φ is positive and just change the sign in the partition function equation). We will also assume

- $\varphi \in C^1$,
- $\varphi(x) = 0$ for $|x| \geq 1$.
- $\varphi(x) = \varphi(-x)$.

[18] These forces between molecules are now known as **van der Waals forces**.

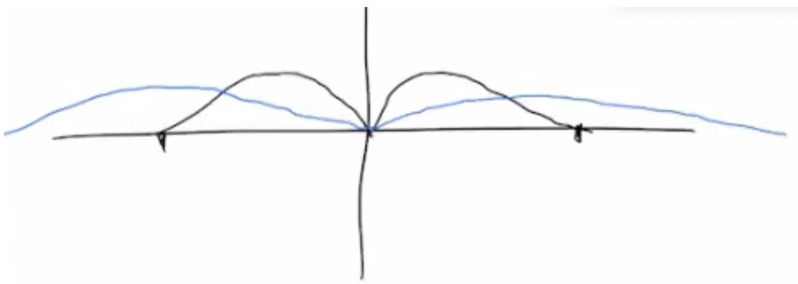
Let

$$\varphi^r(x) = \frac{1}{r^3}\varphi\left(\frac{x}{r}\right), \qquad \text{for } r > 0,$$

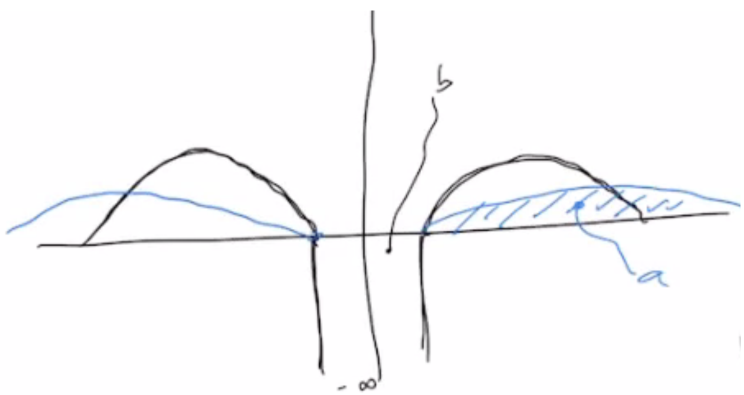
so

$$\int \varphi^r \, dm_3 = \int \varphi \, dm_3 = a$$

for all $r > 0$. We will be taking $n \to \infty$ and then taking $r \to \infty$. Here's what this looks like:



To get $a > 0$ and $b > 0$, we would need to treat a picture like this:



We want to estimate

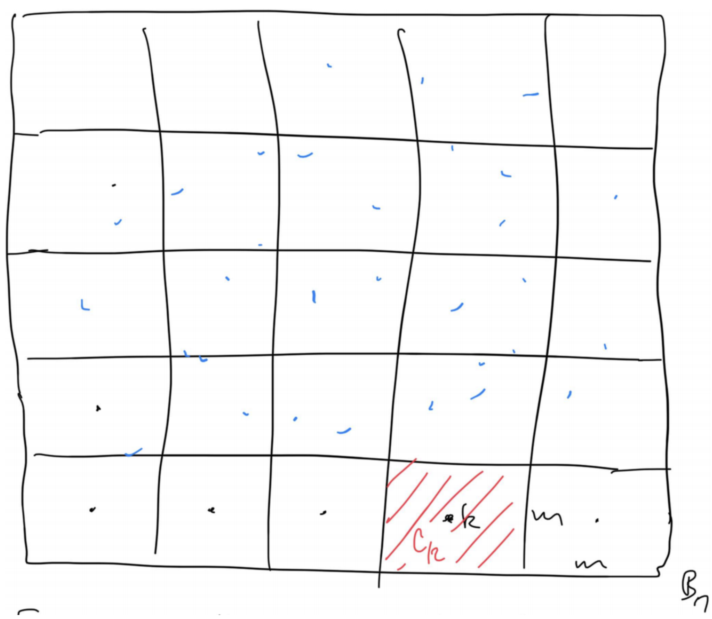
$$\widetilde{Z}_n(\beta) = \sum_{\substack{\omega \in \Omega_n \\ |\omega| = N_n}} \exp\left(-\beta \Phi_n^r(\omega)\right),$$

where $\Omega_n := \{0,1\}^{B_n}$, N_n is the number of particles in B_n, and $B_n = \{1, \ldots, n\}^3$ (note we are changing notation to make n be some geometric parameter rather than the number of particles). We also have

$$\Phi_n^r(\omega) = \sum_{i,j \in B_n} \varphi^r(\varepsilon(i-j))\omega_i \omega_j.$$

18.4 Splitting space into boxes with mass pooled around the centers

We want to estimate $\frac{1}{n}\log \widehat{Z}_n(\beta)$, and we will let $n, r \to \infty$, $\varepsilon \to 0$, and differentiate with respect to v. Here is a picture of how we will do it:



Fix another $m \in \mathbb{N}$, and divide B_n into $m \times m$ boxes. Let $\mathcal{C}_n$ be the set of centers of these boxes. If $k \in \mathcal{C}_n$, then C_k will be the box with center k. (We will assume that $m \mid n$ for simplicity.). Picture $1 \ll m \ll r$, and let $\omega \in \Omega_n$ with $|\omega| = N_n$. We will define a map $D : \Omega_n \to \widetilde{\Omega}_n = \{0, 1/m^3, 2/m^3, \dots, 1\}^{\mathcal{C}_n}$ by

$$D(\omega)_k = \frac{1}{m^3}\sum_{i \in C_k} \omega_i.$$

The idea is that $\Phi_n^r(\omega)$ is approximately a function only of $D(\omega)$, provided $r \gg m$.

Define the **effective energy** of a configuration $\rho \in \widetilde{\Omega}_n$:

$$\widetilde{\Phi}_n^r(\rho) = m^6 \sum_{k,\ell \in \mathcal{C}_n} \rho_k \rho_\ell \varphi^r(\varepsilon(k-\ell))$$

Lemma 18.1. *If $D(\omega) = \rho$, then*

$$\Phi_n^r(\omega) = \widetilde{\Phi}_n^r(\rho) + O\left(n \cdot \frac{1}{mr}\right).$$

Proof. Suppose $i \in C_k$ and $j \in C_\ell$. Then

$$\varphi^r(\varepsilon(i-j)) - \varphi^r(\varepsilon(k-\ell)) \le \|\nabla \varphi^r\|(|\varepsilon(i-k)| + |\varepsilon(j-\ell)|)$$

$$= \frac{1}{r^4}\|\nabla\varphi\| \cdot O\left(\frac{m\varepsilon}{r^4}\right)$$
$$= O\left(\frac{m\varepsilon}{r^4}\right).$$

Therefore,

$$\begin{aligned}
&\left|\sum_{i\in C_k, j\in C_\ell} \varphi^r(\varepsilon(i-j))\omega_i\omega_j - m^6\varphi^r(\varepsilon(k-\ell))\rho_k\rho_\ell\right| \\
&\qquad = \left|\sum_{i\in C_k, j\in C_\ell} [\varphi^r(\varepsilon(i-j)) - \varphi^r(\varepsilon(k-\ell))]\omega_i\omega_j\right| \\
&\qquad \le m^6 O\left(\frac{m\varepsilon}{r^4}\right) \\
&\qquad = O\left(\frac{m^7\varepsilon}{r^4}\right).
\end{aligned}$$

All that remains will be to sum over all pairs of boxes, which we will do next time. □

19 Deriving van der Waal's Equation (Cont.)

19.1 Recap+partitioning space into boxes lemma

In our current setting we have a box $B_n = \frac{1}{\varepsilon}(R_n \cap \varepsilon\mathbb{Z}^3) = \{1, \ldots, n\}^n$. We have N_n particles in B_n, where $\frac{|B_n|}{N_n} \to \frac{v}{\varepsilon^3}$. The particles are located at $\omega \in \Omega_n = \{0,1\}^{B_n}$, where $|\omega| = N_n$. We have a "local density map" $D : \Omega_n \to \widetilde{\Omega}_n = \{0, 1/m^3, \ldots, 1\}^{\mathcal{C}_n}$ with

$$D(\omega)_k = \frac{1}{m^3} \sum_{i \in C_k} \omega_i,$$

where $m \mid n$ and $\{C_k : k \in \mathcal{C}_n\}$ is a partition of B_n into $(m \times m \times m)$-boxes and $\mathcal{C}_n$ is the set of centers of boxes.

The original energy of $\omega \in \Omega_n$ is

$$\Phi_n^r(\omega) = -\sum_{i,j \in B_n} \varphi^r(\varepsilon(i-j))\omega_i\omega_j,$$

where $\varphi : \mathbb{R}^3 \to [0, \infty)$ is C^1, symmetric, has support $\subseteq \overline{B_1(0)}$, and $\varphi^r(x) = r^{-3}\varphi(x/r)$ is a dilation for $r > 0$.

The **effective energy** of $\omega \in \widetilde{\Omega}_n$ is

$$\widetilde{\Phi}_n^r(\rho) = -m^6 \sum_{k,\ell \in \mathcal{C}_n} \varphi^r(\varepsilon(k-\ell))\rho_k\rho_\ell.$$

We also had the following lemma:

Lemma 19.1. *If $D(\omega) = \rho$, then*

$$\Phi_n^r(\omega) = \widetilde{\Phi}_n^r(\rho) + O\left(\frac{n^3 m}{\varepsilon^2 r}\right).$$

Proof. Last time, we showed that

$$\left| \sum_{i \in C_k, j \in C_\ell} \varphi^r(\varepsilon(i-j))\omega_i\omega_j \right| \leq m^6 O\left(\frac{m^7 \varepsilon}{r^4}\right).$$

Finally, we sum over $k, \ell \in \mathcal{C}_n$:

$$|\Phi_n^r(\omega) - \widetilde{\Phi}_n^r(\rho)| = \left| \sum_{k,\ell \in \mathcal{C}_n} \left[\sum_{i \in C_k, j \in C_\ell} \varphi^r(\varepsilon(i-j))\omega_i\omega_j - m^6 \varphi^r(\varepsilon(k-\ell))\rho_k\rho_\ell \right] \right|.$$

Observe that if $\mathrm{dist}(C_k, C_\ell) > r/\varepsilon$, then the expression in the square braces equals 0. How many pairs (k, ℓ) are left? The number of k we can choose first is $(n/m)^3$. Then the number of ℓs that "hit" k equals $O(r^3/(\varepsilon m^3))$. The total number of nonzero terms is $O\left(\frac{n^3 r^3}{\varepsilon^3 m^6}\right)$. Now multiply by the previous bound on those terms to get

$$\leq O\left(\frac{n^3 m}{\varepsilon^2 r}\right). \qquad \square$$

Note that we will let $n, r, m \to \infty$ (with $m = \sqrt{r}$), and then finally let $\varepsilon \to 0$.

19.2 Estimating the size of the partition

Now use this lemma to approximate the partition

$$\begin{aligned}
Z_n^r &= \sum_{\substack{\omega \in \Omega_n \\ |\omega| = N_n}} \exp(-\beta \Phi_n^r(\omega)) \\
&= \sum_{\substack{\rho \in \widetilde{\Omega}_n \\ |\rho| = N_n/m^3}} \sum_{D(\omega)=\rho} \exp(-\beta \Phi_n^r(\omega)) \\
&= \sum_{\substack{\rho \in \widetilde{\Omega}_n \\ |\rho| = N_n/m^3}} \sum_{D(\omega)=\rho} \exp(-\beta \widetilde{\Phi}_n^r(\omega)) \cdot \exp\left(O\left(\frac{n^3 m}{\varepsilon^2 r}\right)\right) \\
&= \underbrace{\sum_{\substack{\rho \in \widetilde{\Omega}_n \\ |\rho| = N_n/m^3}} |D^{-1}(\{\rho\})| \exp(-\beta \widetilde{\Phi}_n^r(\omega))}_{\widetilde{Z}_n^r} \cdot \exp\left(O\left(\frac{n^3 m}{\varepsilon^2 r}\right)\right).
\end{aligned}$$

Next, estimate $|D^{-1}(\rho)|$. This equals

$$\begin{aligned}
\prod_{k \in \mathcal{C}_n} (\#\text{ ways to put } m^3\rho_k \text{ particles into } m^3 \text{ holes}) &= \prod_{k \in \mathcal{C}_n} \binom{m^3}{m^3 \rho_k} \\
&= \prod_k e^{m^3 H(\rho_k, 1-\rho_k) + o(m^3)} \\
&= e^{n^3 [W(\rho) + o(1)]},
\end{aligned}$$

where

$$W(\rho) := \frac{1}{n^3} \sum_{k \in \mathcal{C}_n} m^3 H(\rho_k, 1 - \rho_k) = \frac{1}{(n/m)^3} \sum_{k \in \mathcal{C}_n} H(\rho_k, 1 - \rho_k).$$

Now insert this new approximation to get

$$\widetilde{Z}_n^r = e^{o(n^3)} \underbrace{\sum_{\substack{\rho \in \widetilde{\Omega}_n \\ |\rho| = N_n/m^3}} \exp(n^3 W(\rho) - \beta \widetilde{\Phi}_n^r(\rho))}_{\widehat{Z}_n^r}.$$

The key observation is that the number of terms here is $\le (m^3+1)^{n^3/m^3} = \exp O(n^3 \cdot \frac{\log m}{m^3})$. We will use this via the following:

Lemma 19.2. *Let $a_i \ge 0$ for all $i \in I$ (with $|I| < \infty$). Then*

$$\max_i a_i \le \sum_{i \in I} a_i \le |I| \max_i a_i.$$

Corollary 19.1.

$$\begin{aligned} n^3 \max &\left\{ W(\rho) - \frac{\beta}{n^3} \widetilde{\Phi}_n^r(\rho) : \rho \in \widetilde{\Omega}_n, |\rho| = \frac{N_n}{m^3} \right\} \\ &\le \log \widetilde{Z}_n^r \\ &\qquad \le n^3 \max \left\{ W(\rho) - \frac{\beta}{n^3} \widetilde{\Phi}_n^r(\rho) : \rho \in \widetilde{\Omega}_n, |\rho| = \frac{N_n}{m^3} \right\} + O\left(n^3 \cdot \frac{\log m}{m^3} \right). \end{aligned}$$

Our main remaining task is to understand the maximum of

$$W(\rho) - \frac{\beta}{n^3} \widetilde{\Phi}_n^r(\rho)$$

for $\rho \in \widetilde{\Omega}_n$ such that $|\rho| = N_n/m^3$. Let's unpack this:

$$\begin{aligned} W(\rho) - \frac{\beta}{n^3} \widetilde{\Phi}_n^r(\rho) &= \frac{1}{(n/m)^3} \sum_k H(\rho_k, 1-\rho_k) + \frac{\beta m^6}{n^3} \sum_{k,\ell} \varphi^r(\varepsilon(k-\ell)) \rho_k \rho_\ell \\ &= \frac{1}{(n/m)^3} \left[\sum_k H(\rho_k, 1-\rho_k) + \beta m^3 \sum_{k,\ell} \varphi^r(\varepsilon(k-\ell)) \rho_k \rho_\ell \right] \end{aligned}$$

The key idea is to bound the right term above by something with no cross terms. Observe that $\rho_k \rho_\ell \le \frac{1}{2}(\rho_k^2 + \rho_\ell^2)$ using the AM-GM inequality. Insert this into the second term above:

$$\begin{aligned} m \sum_{k,\ell} \varphi^r(\varepsilon(k-\ell)) \rho_k \rho_\ell &\le m \sum_{k,\ell} \varphi^r(\varepsilon(k-\ell)) \left[\frac{\rho_k^2 + \rho_\ell^2}{2} \right] \\ &= m^3 \sum_{k,\ell} \rho_k^2 \varphi^r(\varepsilon(k-\ell)) \end{aligned}$$

$$= \frac{1}{\varepsilon^3} \sum_{k \in \mathcal{C}_n} \rho_k^2 \underbrace{\left[(\varepsilon m)^3 \sum_{\ell \in \mathcal{C}_n} \varphi^r(\varepsilon(k-\ell)) \right]}_{=:\alpha(n,m,r,\varepsilon,k)}.$$

So we will try to maximize

$$\begin{aligned} &\frac{1}{(n/m)^3} \left[\sum_k H(\rho_k, 1-\rho_k) + \beta m^3 \sum_k \rho_k^2 \cdot \alpha(n,m,r,\varepsilon,k) \right] \\ &\qquad = \frac{1}{(n/m)^3} \sum_{k \in \mathcal{C}_n} \left[H(\rho_k, 1-\rho_k) + \beta m^3 \cdot \alpha(n,m,r,\varepsilon,k) \rho_k^2 \right] \end{aligned}$$

Now consider

$$\alpha(n,m,r,\varepsilon,k) = (\varepsilon m)^3 \sum_{\ell \in \mathcal{C}_n} \varphi(\varepsilon(k-\ell))$$

Ignoring that some k can be on the boundary of the box,

$$\begin{aligned} &\leq (\varepsilon m)^3 \sum_{v \in \varepsilon m \mathbb{Z}^3} \varphi^r(v) \\ &= (\varepsilon m)^3 \sum_{v \in \varepsilon m \mathbb{Z}^3} \frac{1}{r^3} \varphi(v/r) \\ &= \frac{\varepsilon^3 m^3}{r^3} \sum_{v \in (\varepsilon m/r)\mathbb{Z}^3} \frac{1}{r^3} \varphi(v). \end{aligned}$$

As $r \to \infty$, this will give a Riemann sum for $\int \varphi$. We will plug this back into the previous expression next time.

20 Deriving van der Waal's Equation (Part 3)

20.1 Bound on α

Last time, we had a quantity α which depended on various factors. We bounded it by a term $\alpha(m, r, \varepsilon)$ which only depended on these 3 quantities. We will compare this to the integral $\int \varphi = \int \varphi^r$.

$$\left|\alpha(m, r, \varepsilon) - \int \varphi^r\right| = \left|(\varepsilon m)^3 \sum_{v \in \varepsilon m \mathbb{Z}^3} \varphi^r(v) - \int \varphi^r\right|$$

Let Q_v be the cube with side length εm and center v.

$$\begin{aligned}
&= \left|\sum_{v \in \varepsilon m \mathbb{Z}^3} \left((\varepsilon m)^3 \varphi^r(v) - \int_{Q_v} \varphi^r\right)\right| \\
&= \left|\sum_v \left(|Q_v| \varphi^r(v) - \int_{Q_v} \varphi^r\right)\right| \\
&\leq \sum_v \int_{Q_v} |\varphi^r(v) - \varphi^r(x)|\, dx \\
&\leq \sum_{v : Q_v \cap \overline{B_r} \neq \varnothing} (\varepsilon m)^3 \cancel{\frac{\sqrt{3}}{2}} \varepsilon m \cdot \frac{\cancelto{O(1)}{\|\nabla \varphi\|}}{r^4} \\
&= O\left(\frac{r^3}{(\varepsilon m)^3} \cdot \frac{(\varepsilon m)^4}{r^4}\right) \\
&= O\left(\frac{\varepsilon m}{r}\right).
\end{aligned}$$

So

$$\alpha(m, r, \varepsilon) = \alpha + O\left(\frac{\varepsilon m}{v}\right)$$

for some constant α.

20.2 Maximizing the entropy term

Now consider maximizing

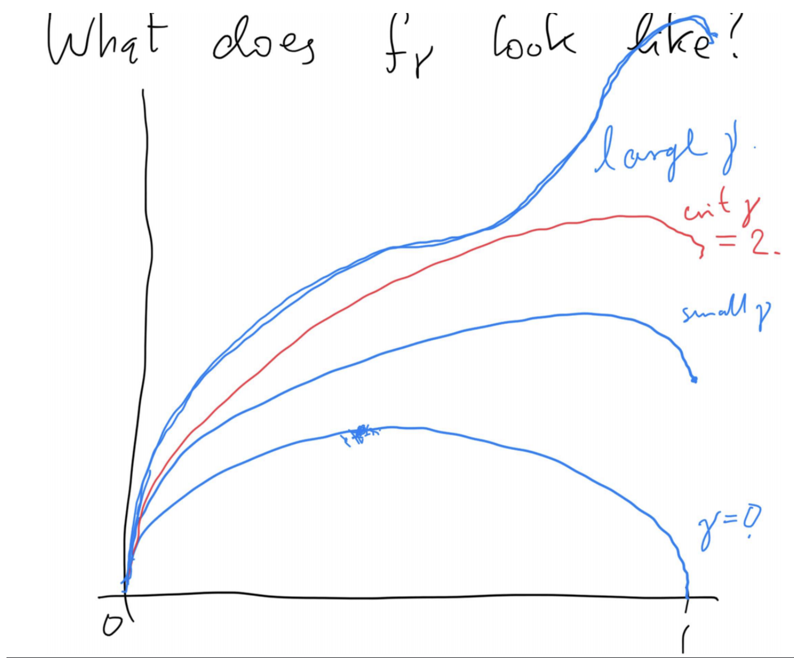
$$\frac{1}{(n/m^3)} \sum_{k \in \mathcal{C}_n} [H(\rho_k, 1 - \rho_k) + \gamma \rho_k^2],$$

where $\gamma = \beta \alpha(m, \varepsilon, r)/\varepsilon^3 = (\beta/\varepsilon^3)(\alpha + O(\varepsilon m/r))$. Now, we want to try to maximize this over $\rho \in \widetilde{\Omega}_n$ with $|\rho| = N_n/m^3$ (recall $N_n/n^3 \to \varepsilon^3/v$ as $n \to \infty$). We can also write this expression as

$$\frac{1}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n} f_\gamma(\rho_k),$$

where $f_\gamma(x) = H(x, 1-x) + \gamma x^2$ for $0 \le x \le 1$.

What does f_γ look like? Here is a picture:



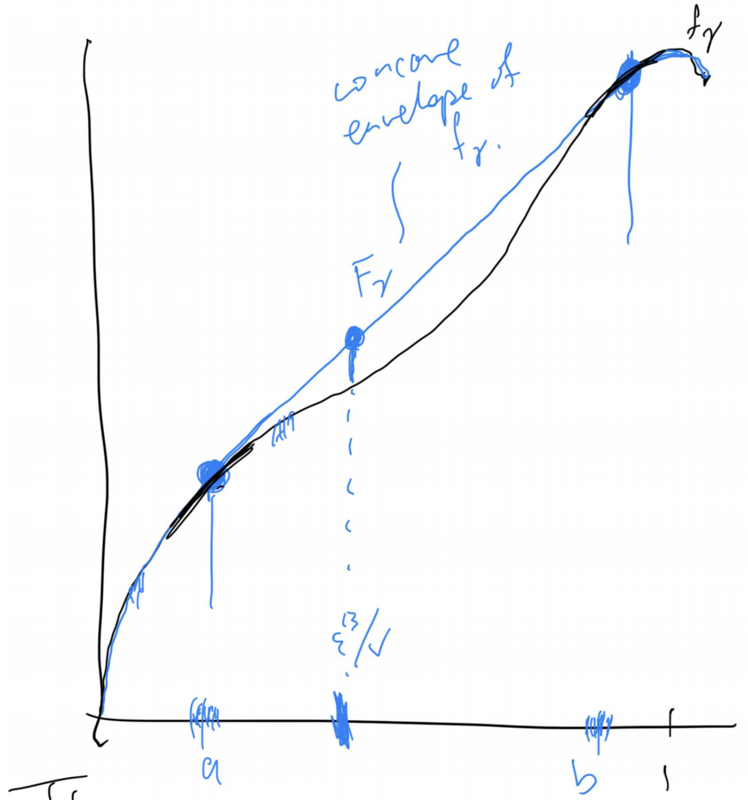
There is a critical value of γ which is the largest γ for which this is still concave. Check using calculus that the critical γ equals 2.

If $\gamma \le 2$, f_γ is concave, so Jensen's inequality gives

$$\begin{aligned}
\frac{1}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n} f_\gamma(\rho_k) &\le f_\gamma\left(\frac{1}{|C_n|}\sum_k \rho_k\right) \\
&= f_\gamma\left(\frac{N_n}{m^3}\right) \\
&= f_\gamma\left(\frac{N_n}{m^3}\right) \\
&\xrightarrow{n\to\infty} f_\gamma\left(\frac{\varepsilon^3}{v}\right).
\end{aligned}$$

We can make this close to tight by taking $\rho_k \approx \varepsilon^3/v$ for all k.

What if $\gamma > 2$? Use the *concave envelope* F_γ of f_γ:



Then

$$\begin{aligned}
\frac{1}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n} f_\gamma(\rho_k) &\le \frac{1}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n} F_\gamma(\rho_k) \\
&\le F_\gamma\left(\frac{1}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n}\rho_k\right) \\
&\approx F_\gamma\left(\frac{\varepsilon^3}{v}\right)
\end{aligned}$$

as $n\to\infty$. This also can be brought as close as we like once $n\to\infty$ and m is large.

- If $\varepsilon^3/v \notin (a,b)$, then $F_\gamma(\varepsilon^3/v) = f_\gamma(\varepsilon^3/v)$. Then just take $\rho_k \approx \varepsilon^3/v$ for all k.
- If $a < \varepsilon^3/v < b$, then $\rho_k \approx \varepsilon^3/v$ for all k will give you

$$\frac{1}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n} f_\gamma(\rho_k) = f_\gamma\left(\frac{\varepsilon^3}{v}\right) < F_\gamma\left(\frac{\varepsilon^3}{v}\right).$$

Instead, express $\varepsilon^3/v = ta + (1-t)b$. Now choose the values ρ_k so that $\rho_k \approx a$ for $\approx t|\mathcal{C}_n|$ many ks and $\rho_k \approx b$ for $\approx (1-t)|\mathcal{C}_n|$ many ks. Then

$$\frac{1}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n} f_\gamma(\rho_k) \approx t f_\gamma(a) + (1-t) f_\gamma(b) = F_\gamma\left(\frac{\varepsilon^3}{v}\right)$$

The conclusion is that

$$\frac{1}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n} f_\gamma(\rho_k) = F_\gamma\left(\frac{\varepsilon^3}{v}\right) + O\left(\frac{1}{m}\right)$$

as $n \to \infty$.

20.3 Maximizing the effective partition function

Is the maximization problem for $\log \widehat{Z}_n^r$ close to the same value? Yes!

Go back to

$$\max_{|\rho|=N_n/m^3}\left\{W(\rho) - \frac{\beta}{n^3}\widetilde{\Phi}_n^r(\rho)\right\}.$$

Can we get this close to the same value? Yes. Since f_γ is strictly convex near a and b, we must have roughly a t fraction of ρ_ks close to a and roughly a $(1-t)$ fraction of ρ_ks close to b

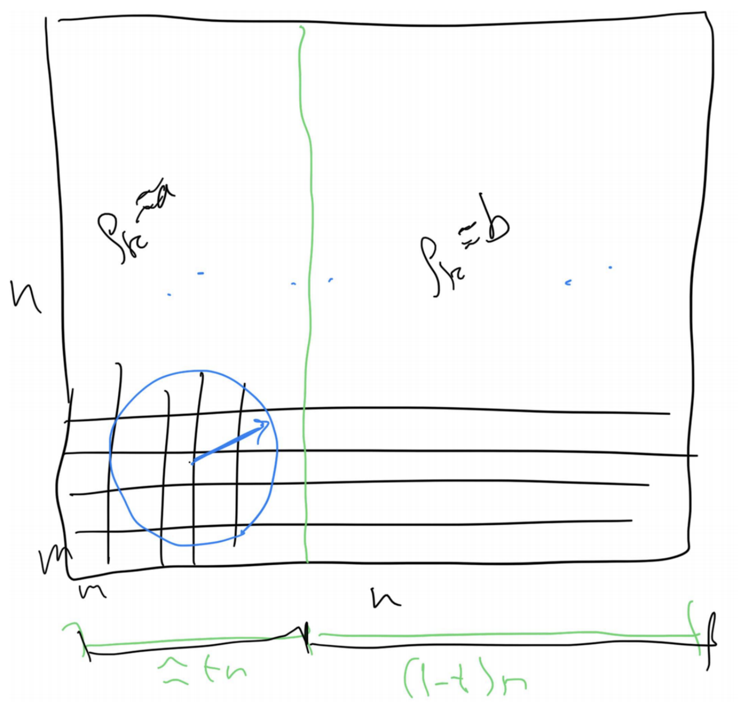
When is the above maximum close to the average of f_γ? We had the bound via AM-GM:

$$\sum_k H(\rho_k, 1-\rho_k) + \beta m^3 \sum_{k,\ell}\varphi^r(\varepsilon(k-\ell))\rho_k\rho_\ell \le \sum_k H(\rho_k, 1-\rho_k) + \beta m^3 \sum_{k,\ell}\varphi^r(\varepsilon(k-\ell))\frac{\rho_k^2+\rho_\ell^2}{2}.$$

The difference equals

$$\beta m^3 \sum_{k,\ell}\varphi^r(\varepsilon(k-\ell))\frac{1}{2}(\rho_k-\rho_\ell)^2.$$

This is small if $\rho_k \approx \rho_\ell$ for most pairs (k,ℓ) where $\varphi^r(\varepsilon(k-\ell))$ is not negligible. What kinds of ρ achieve all these requirements? Choose it according to this picture:



Then we do get

$$W(\rho) - \frac{\beta}{n^3}\widetilde{\Phi}_n^r(\rho) \approx \frac{1}{|\mathcal{C}_n|}\sum_k f_\gamma(\rho_k) \approx F_\gamma\left(\frac{\varepsilon^3}{v}\right).$$

We will finish off this story next time.

21 Deriving van der Waal's Equation: The Final Chapter

21.1 Combining accumulated approximations for the partition function

So far, we had our original partition function

$$Z_n^r = \frac{1}{N_n!} \int \cdots \int_{R_n^{N_n}} e^{-\beta \sum_{i,j} \varphi^r(q_i - q_j)}\, dm_3^{\times N_n}(q_1, \ldots, q_{N_n}).$$

We replaced it with the discretized partition function

$$\widetilde{Z}_n^r = \sum_{\substack{\omega \in \{0,1\}^{B_n} \\ |\omega| = N_n}} e^{-\beta \Phi_n^r(\omega)}, \qquad B_n = R_n \cap \varepsilon \mathbb{Z}^3,$$

which is what we will prove results about. We also introduced the **effective partition function**

$$\widehat{Z}_n^r = \sum_{\substack{\rho \in \widetilde{\Omega}_n \\ |\rho| = N_n/m^3}} e^{n^3 W(\rho) - \beta \widetilde{\Phi}_n^r(\rho)},$$

which approximates $\widetilde{Z}_n^r$.

So far, our approximations have been:

- (Unproved heuristic):
$$Z_n^r \approx \widetilde{Z}_n^r e^{[\text{small}] \cdot n} \qquad \text{as } n \to \infty.$$

-
$$\begin{aligned}
\log \widetilde{Z}_n^r &= \log \widehat{Z}_n^r + O\left(\frac{n^3 m}{\varepsilon^2 r}\right) + o(n^3) \\
&= n^3 \max_\rho \left\{ W(p) - \frac{\beta}{n^3} \widetilde{\Phi}_n^r(\rho) \right\} + O\left(n^3 \cdot \frac{\log m}{m^3}\right) + O\left(\frac{n^3 m}{\varepsilon^2 r}\right) + o(n^3) \\
&= n^3 \left[F_{\beta\alpha(m,\varepsilon,r)}\left(\frac{\varepsilon^3}{v}\right) + \cancel{o_{n\to\infty}(1)} + \cancel{O\left(\frac{1}{m^3}\right)} \right] + O\left(n^3 \cdot \frac{\log m}{m^3}\right) \\
&\qquad + O\left(\frac{n^3 m}{\varepsilon^2 r}\right) + o(n^3)
\end{aligned}$$

Recall that $f_\gamma(x) = H(x, 1-x) + \gamma x^2$ and $F_\gamma(x)$ is the concave envelope of $f_\gamma(x)$ for $0 \le x \le 1$.

$$\begin{aligned}
&= n^3 \left[F_{\beta\alpha/\varepsilon^3}\left(\frac{\varepsilon^3}{v}\right) + \cancel{O\left(\frac{m}{\varepsilon^2 r}\right)} \right] + (\text{other error terms}) \\
&= n^3 F_{\alpha\beta/\varepsilon^3}\left(\frac{\varepsilon^3}{v}\right) + O\left(n^3 \cdot \frac{\log m}{m^3}\right) + O\left(\frac{n^3 m}{\varepsilon^2 r}\right) + o(n^3).
\end{aligned}$$

We want everything in terms of N_n, rather than in terms of the volume of the box. So let's write

$$\begin{aligned}
\frac{1}{N_n}\log \widehat{Z}_n^r &= \frac{1}{n^3}\frac{n^3}{N_n}[\text{above stuff}] \\
&= \left(\frac{v}{\varepsilon^3} + o_{n\to\infty}(1)\right)[\text{above stuff}] \\
&= \left(\frac{v}{\varepsilon^3} + o(1)\right)\left[F_{\alpha\beta/\varepsilon^3}\left(\frac{\varepsilon^3}{v}\right) + O\left(\frac{\log m}{m^3}\right) + O\left(\frac{m}{\varepsilon^2 r}\right) + o(1)\right].
\end{aligned}$$

21.2 Taking limits to find the asymptotic behavior of the partition function

Let $n \to \infty$. Then $r \to \infty$ and $m \to \infty$ (with $m = \sqrt{r}$). Then, we will let $\varepsilon \to 0$. We get

$$\lim_{r\to\infty}\lim_{n\to\infty}\frac{1}{N_n}\log \widetilde{Z}_n^r = \frac{v}{\varepsilon^3}F_{\alpha\beta/\varepsilon^3}\left(\frac{\varepsilon^3}{v}\right).$$

What happens here as $\varepsilon \to 0$?

We need the following lemma (proven in Homework 3):

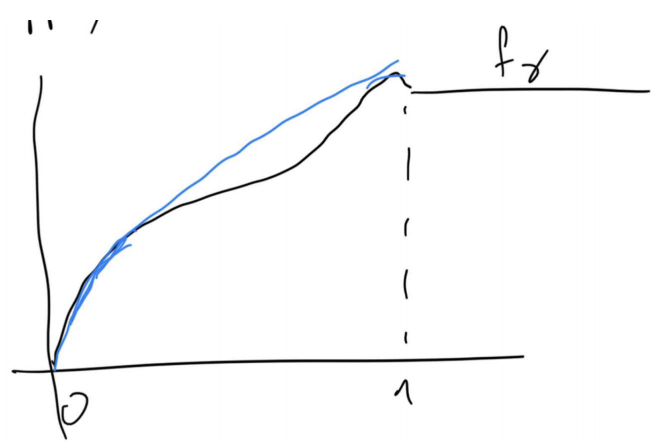
Lemma 21.1. *Suppose $f : [0,\infty) \to \mathbb{R}$ is continuous with concave envelope $F : [0,\infty) \to \mathbb{R}$. Assume $f(x)/x \to 0$ as $x \to \infty$. Then $g : [0,\infty) \to \mathbb{R}$, defined by*

$$g(v) = \begin{cases} v \cdot f(1/v) & v > 0 \\ 0 & v = 0, \end{cases}$$

has concave envelope equal to

$$\begin{cases} v \cdot F(1/v) & v > 0 \\ 0 & v = 0. \end{cases}$$

We will apply this to f_γ (making it flat to the right of the unit interval):



So the remaining expression above is the concave envelope of $\frac{v}{\varepsilon^3} f_{\alpha\beta/\varepsilon^3}(\frac{\varepsilon^3}{v})$, which is explicit. This is

$$\begin{aligned}
&= \frac{v}{\varepsilon^3}\left[-\frac{\varepsilon^3}{v}\log\frac{\varepsilon^3}{v} - \left(1-\frac{\varepsilon^3}{v}\right)\log\left(1-\frac{\varepsilon^3}{v}\right) + \frac{\alpha\beta}{\varepsilon^3}\left(\frac{\varepsilon^3}{v}\right)^2\right] \\
&= \log\frac{v}{\varepsilon^3} - \left(\frac{v}{\varepsilon^3}-1\right)\log\left(1-\frac{\varepsilon^3}{v}\right) + \frac{\alpha\beta}{v} \\
&= \log v - \log\varepsilon^3 + \left(\frac{v}{\varepsilon^3}-1\right)\left(\frac{\varepsilon^3}{v} + O\left(\frac{\varepsilon^6}{v^2}\right)\right) + \frac{\alpha\beta}{v} \\
&= \log v - \log\varepsilon^3 + 1 + O\left(\frac{\varepsilon^3}{v}\right) + \frac{\alpha\beta}{v}.
\end{aligned}$$

In our original formulas for $\widehat{Z}_n^r$ and $\widetilde{Z}_n^r$, we should have had a factor of $(\varepsilon^3)^{N_n}$ to account for the number of particles per box. Putting that in (and carrying it throughout the whole calculation), we are left with

$$\frac{v}{\varepsilon^3} f_{\alpha\beta/\varepsilon^3}\left(\frac{\varepsilon^3}{v}\right) = \log v + 1 + \frac{\alpha\beta}{v} + O\left(\frac{\varepsilon^3}{v}\right).$$

This is a uniform limit as $\varepsilon \downarrow 0$ for v bounded away from 0. Check that we also get convergence of the derivatives in v and that we get the same convergence for the concave envelopes. So

$$\lim_{r\to\infty}\lim_{n\to\infty}\frac{1}{N_n}\log\widetilde{Z}_n^r = \text{conc. env. of } \underbrace{\left(\log v + 1 + \frac{\alpha\beta}{v}\right)}_{g(v)}.$$

21.3 Recovering van der Waal's equation and Maxwell's equal area correction

What does this have to do for the van der Waal's equation? Maxwell's equal area correction is precisely what you get when you replace $\log v + 1 + \frac{\alpha\beta}{v}$ by its concave envelope. Explicitly, we get:

$$P = \frac{\partial}{\partial v}[T \log \text{partition function}].$$

We have

$$\frac{\partial}{\partial v}\left[\frac{1}{\beta}\log v + \frac{1}{\beta} + \frac{1}{v}\right] = \frac{1}{\beta v} - \frac{\alpha}{v^2},$$

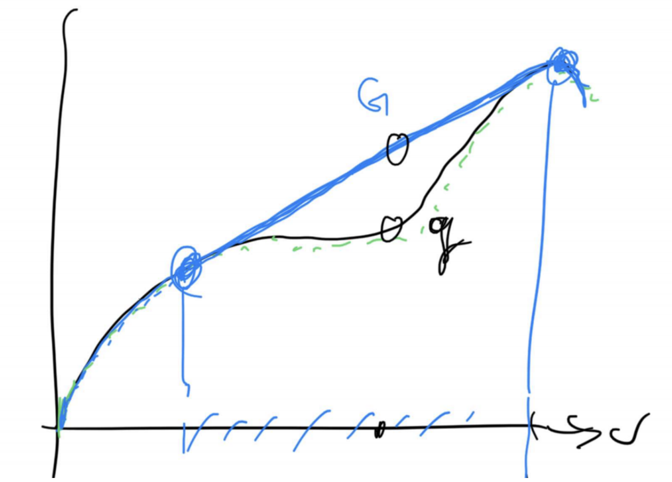
so

$$v\left(P + \frac{\alpha}{v^2}\right) = \frac{1}{\beta} = T.$$

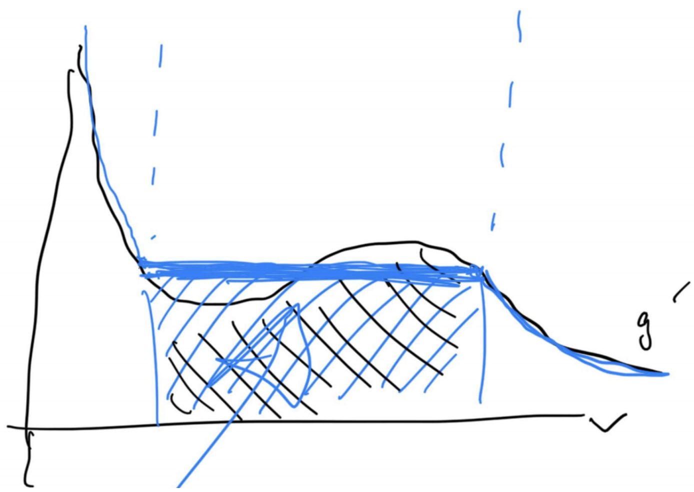
That is, we get van der Waal's equation,

$$v(P + \frac{\alpha}{v^2}) = NT.$$

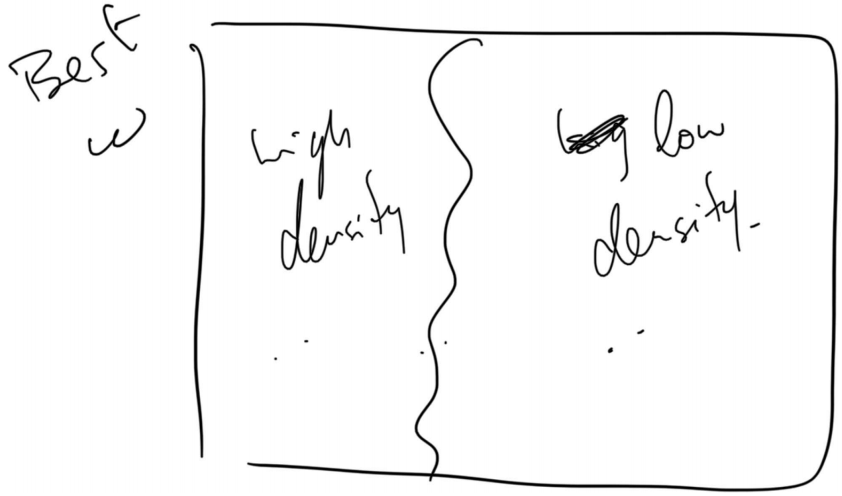
In the case of non-concavity, how do we fix it?



$P = \frac{\partial}{\partial v}[T \cdot G(v)]$, so let's graph g' and G', the derivative of the concave envelope.



Inside this shaded region above, recall which ρs carried most of the mass in $\widehat{Z}_n^r = \sum_\rho \exp(n^3 W(\rho) - \beta\widehat{\Phi}_n^r(\rho))$. Our analysis told us this, i.e. which micro configurations carry most of the mass in the canonical ensemble. This means that the best ω have regions of high density and regions of low density:



That is, the substance separates into a high density region (liquid) and a low density region (gas). For example, some of the water in a glass of water will evaporate into water vapor. Thus, van der Waal's equation correctly predicts the existence of a phase transition between gas and liquids.

22 Basics of Lattice Models

22.1 Lattices

In the derivation of van der Waal's equation, we used a discretization and let $\varepsilon \to 0$. Now we will begin looking at lattice models, where there is a fixed lattice; this precludes any notion of particles getting too close to each other. Consider boxes $B_n \subseteq \mathbb{Z}^d$, which are cuboids with all sides $\to \infty$. At each site $i \in B_n$, the set of possible "local states" is a finite set A (the **alphabet**). So the microscopic states of the system are elements $\omega \in \Omega_n = A^{B_n}$.

Example 22.1 (Lattice gases)**.** If $A = \{0, 1\}$, we can interpret "$\omega_i = 1$" as "there is a particle at i" and interpret "$\omega_i = 0$" as 'there is no particle at i."

Example 22.2 (Magnetizable solid)**.** If $A = \{-1, 1\}$, we can interpret ω_i as the "direction" of a magnetic spin located at i inside a magnetizable solid. (More realistic magnet models allow A to be a sphere in $\mathbb{R}^3$.)

Example 22.3. More generally, we could have $A = \{0, a_1, a_2, \dots, a_k\}$. Here, 0 represents the absence of a particle, and the a_i represent possible internal states of particles.

22.2 Interactions

The total energy of $\omega \in \Omega_n$ will be given by an *interaction*.

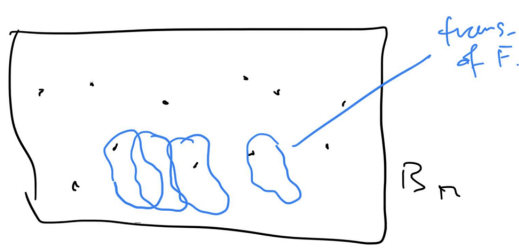
Definition 22.1. An **interaction** is a family $(\varphi_F : F \subseteq \mathbb{Z}^d,\ F \text{ finite})$, where

1. $\varphi_F : A^F \to \mathbb{R}$
2. translation invariant:

$$\varphi_F(\underbrace{(a_v)_{v \in F}}_{\in A^F}) = \varphi_{F+u}(\underbrace{(a_{v+u})_{v \in F}}_{\in A^{u+F}}).$$

Then for $\omega \in \Omega_n$, its **total (potential) energy** is

$$\Phi_{B_n}(\omega) = \sum_{F \subseteq B_n} \varphi_F(\omega_F).$$



Example 22.4. Most simply, a **pair interaction** has $\varphi_F = 0$ unless $|F| = 1$ or $|F| = 2$.

For example, as in our study of van der Waal's equation, we could take $A = \{0, 1\}$ and

$$\varphi_F(\omega) = \begin{cases} -\varphi^r(i-j)\omega_i\omega_j & F = \{i, j\} \\ 0 & \text{otherwise.} \end{cases}$$

22.3 Interaction decay

We want to understand asymptotic behavior as $B_n \to \mathbb{Z}^d$. This is known as the **thermodynamic limit**. To get a meaningful limit, we need enough decay in interaction strength with distance. Possible additional assumptions are:

1. **Finite range**: There exists some $R < \infty$ such that $\varphi_F = 0$ if $\operatorname{diam} F \geq R$.
2. A bit more general: φ is in the **big space** of interactions if $\sum_{F \ni 0} \frac{\|\varphi_F\|_\infty}{|F|} < \infty$. This guarantees "finite energy per particle."
3. φ is in the **small space** of interactions if $\sum_{F \ni 0} \|\varphi_F\|_\infty < \infty$. This is more restrictive than the big space.

Note that the big space and small space of interactions are Banach spaces, and these quantities are norms. We will tend to prove results assuming finite range, with the understanding that a bit more careful reasoning will work for the more general assumptions.

So now we need to look at sets of the form

$$\Omega_{B_n}(\varphi, I) = \left\{ \omega \in A^{B_n} : \frac{\Phi_{B_n}(\omega)}{|B_n|} \in I \right\},$$

for $I \subseteq \mathbb{R}$. Here, we are keeping track of energy per unit volume.

22.4 Observables

Next, we need a notion of macroscopic observables. We will study these as "averages over B_n."

Definition 22.2. An **observable** is a function $\psi : A^W \to \mathbb{R}$ with $W \subseteq \mathbb{Z}^d$, and

$$\Psi_{B_n}(\omega) = \sum_{i + W \subseteq B_n} \psi(\omega_{i+W}).$$

Example 22.5. If $A = \{0, 1\}$, $W = \{0\}$, and $\psi(a) = a$, then

$$\Psi_{B_n}(\omega) = \sum_{i \in B_n} \omega_i = \#\text{ particles in } B_n.$$

22.5 Entropy

We want to study the growth of the cardinality of

$$\Omega_{B_n}(\varphi, I; \psi, J) = \left\{ \omega \in A^{B_n} : \frac{\Phi_{B_n}(\omega)}{|B_n|} \in I, \frac{\Psi_{B_n}(\omega)}{|B_n|} \in J \right\},$$

where I is an open interval $\subseteq \mathbb{R}$, and J is an open convex subset of $\mathbb{R}^n$.

Theorem 22.1. *Let $s_n(I, J) = \log |\Omega_{B_n}(\varphi, I; \psi, J)|$. Then there exists a concave and upper semicontinuous function $s : \mathbb{R} \times \mathbb{R}^r \to [-\infty, \infty)$ such that*

$$s_n(I, J) = |B_n| \cdot \sup_{(x,y) \in I \times J} s(x, y) + o(|B_n|).$$

We will prove this assuming φ has finite range. In this case, we will simplify our work by studying

$$\frac{\sum_{i+F \subseteq B_n} \varphi_F(\omega_{i+F})}{|B_n|} \in I_F$$

for every $\operatorname{diam}(F) < R$, rather than

$$\sum_{\substack{\text{equiv. classes of} \\ \operatorname{diam} F < R \text{ up} \\ \text{to translation}}} \frac{\sum_{i+F \subseteq B_n} \varphi_F(\omega_{i+F})}{|B_n|} \in I.$$

This lets us write

$$\Omega_{B_n}(\psi, J) = \left\{ \frac{\Psi_n(\omega)}{|B_n|} \in J \right\}$$

for a single observable $\psi : A^W \to \mathbb{R}^{r'}$ with r' bigger than r. Let's restate the theorem:

Theorem 22.2. *In the setting above,*

$$s_n(\psi, J) = |B_n| \cdot \sup_{x \in J} s(x) + o(|B_n|),$$

where $s : \mathbb{R}^r \to [-\infty, \infty)$ is concave and upper semicontinuous.

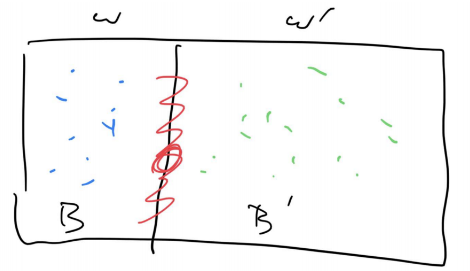
We would like to show that $S_n(\psi, J)$ is superadditive. In fact, previously, we had

$$s_n(\psi, J) + s_m(\psi, J') \leq s_{n+m}\left(\psi, \frac{n}{n+m} J + \frac{m}{n+m} J'\right)$$

for non-interacting systems. First, we need a version for cuboids, something like

$$s_B(\psi, J) + s_{B'}(\psi, J') \leq s_{B \cup B'}\left(\psi, \frac{|B|}{|B| + |B'|} J + \frac{|B'|}{|B| + |B'|} J'\right),$$

when



But

$$\Psi_{B\cup B'}(\omega,\omega') = \Psi_B(\omega) + \Psi_{B'}(\omega') + \text{boundary terms}.$$

We will have to take care of these boundary terms of make this argument work in this case.

23 Existence of the Thermodynamic Limit for Lattice Models

23.1 Recap

Let B be a big finite box in $\mathbb{Z}^d$ (all sides "long enough," which may be specified later). We have a finite set A of single-site states telling us what is happening at a site (such as whether a particle is present at that site). We will look at microscopic states $\omega \in A^B$ and macroscopic observables such as

$$\Psi_B(\omega) = \sum_{i+W \subseteq B} \psi(\omega_{i+W}),$$

where $W \subseteq \mathbb{Z}^d$ is a finite "window" and $\psi : A^W \to \mathbb{R}^n$ and W is fixed. Given $U \subseteq \mathbb{R}^n$, let

$$\Omega_B(\psi, U) = \{\omega \in A^B : \tfrac{1}{|B|}\Psi_B(\omega) \in U\}.$$

Theorem 23.1. *There exists a concave, upper semicontinuous function $s : \mathbb{R}^n \to [-\infty, \infty)$ such that*

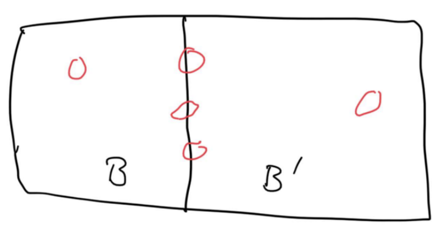
(a) $\max_x s(x) = \log |A|$.

(b) If $U \subseteq \mathbb{R}^n$ is a convex open set such that either $U \cap \{s > -\infty\} = \varnothing$ or $U \cap \operatorname{int}\{s > -\infty\} \neq \varnothing$, then

$$|\Omega_B(\psi, U)| = \exp(|B| \cdot \sup_U s + o(|B|))$$

as $B \uparrow \mathbb{Z}^d$ (i.e. for any sequence $\langle B_n \rangle$ with side lengths $\to \infty$).

We want to use a superadditivity argument with the following type of configuration:



The problem is that when you write down $\Psi_{B\cup B'}(\omega, \omega')$, the translates of W may lie on the boundary of B and B'. So there will be boundary terms we need to deal with:

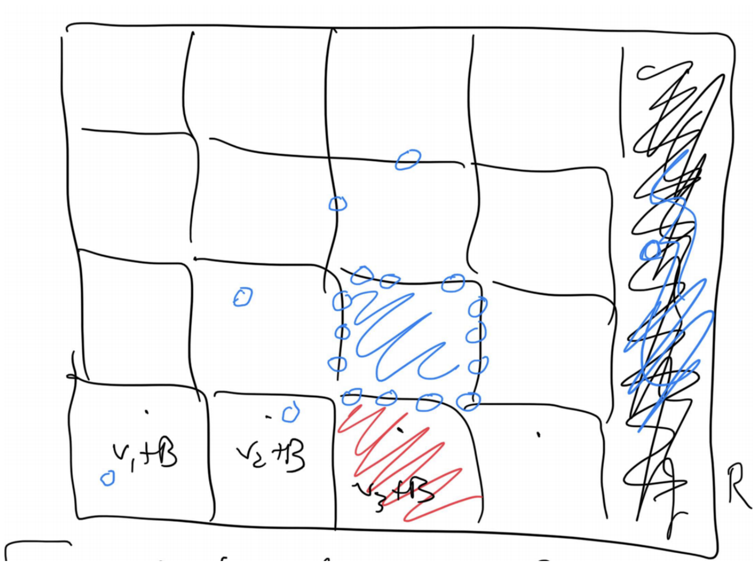
$$\Psi_{B\cup B'}(\omega, \omega') = \Psi_B(\omega) + \Psi_{B'}(\omega') + (\text{boundary terms}).$$

23.2 Proving superadditivity with extra boundary terms

Proposition 23.1. *Fix W, ψ. For every $\varepsilon > 0$, there is an M such that if B has all side-lengths $\geq M$ and R is larger, "big enough" box in terms of B, the the following holds: If $v_1 + B, \dots, v_m + B$ is a maximum-sized collection of disjoint B-translates in R and $U_1, \dots, U_m \subseteq \mathbb{R}^n$ are convex and open, then*

$$|\Omega_R(\psi, U)| \geq \prod_{i=1}^{m} |\Omega_B(\psi(B_i)_\varepsilon)|,$$

where $U = \frac{1}{m}U_1 + \cdots + \frac{1}{m}U_m$ and $V_\varepsilon = \{x : \overline{B_\varepsilon(x)} \subseteq V\}$.



In fact, if $\omega \in A^R$ and $\omega|_{v_i+B} \in \Omega_{v_i+B}(\psi, (U_i)_\varepsilon)$ for all i, then $\omega \in \Omega_B(\psi, U)$.

Proof. We are assuming that $\frac{1}{|B|}\Psi_{v_i+B}(\omega_{v_i+B}) \in (U_i)_\varepsilon$ for all i. Consider

$$\begin{aligned}
\psi_R(\omega) &= \sum_{v+W\subseteq R} \psi(\omega_{v+W}) \\
&= \sum_i \sum_{v+W\subseteq v_i+B} \psi(\omega_{v+W}) + \underbrace{\sum_{\substack{v+W\not\subseteq v_i+B \\ \text{for any } i}} \psi(\omega_{v+W})}_{X} \\
&\in |B|\cdot(U_1)_\varepsilon + \cdots + |B|\cdot(U_m)_\varepsilon + X \\
&= |R|\frac{|B|}{|R|}\cdot((U_1)_\varepsilon + \cdots + (U_m)_\varepsilon) + X
\end{aligned}$$

$$= |R|\left(\frac{1}{m} + o_{R\uparrow\mathbb{Z}^d}(1)\right)((U_1)_\varepsilon + \cdots + (U_m)_\varepsilon) + X$$

For big enough R,

$$\subseteq |R|\left(\frac{U_1}{m} + \cdots + \frac{U_m}{m}\right)_{\varepsilon/2} + X$$

$$= |B|U_{\varepsilon/2} + X$$

Now estimate

$$\begin{aligned}
|X| &= \left|\sum_{\substack{v+W\not\subseteq v_i+B \\ \text{for any } i}} \psi(\omega_{v+W})\right| \\
&\leq \|\psi\|_\infty \cdot \operatorname{diam}(W) \cdot \left(\sum_{v_i+B} |\partial(v_i+B)| + \left|R \setminus \bigcup_i (v_i+B)\right|\right) \\
&= O(1) \cdot (\underbrace{m \cdot |\partial B|} + o_{R\uparrow\mathbb{Z}^d}(|R|)),
\end{aligned}$$

where this bracketed part will be small relative to $m|B| \leq |R|$ if B is big enough. So as $R \uparrow \mathbb{Z}^d$ and then $B \uparrow \mathbb{Z}^d$, we have

$$X = O(1)(o_{R\to\infty}(|R|) + o_{B\to\infty}(|R|)) = o_{R\to\infty, B\to\infty}(|R|).$$

So if B is big enough given ε and then R is big enough given B, then

$$\Psi_R(\omega) \in |R|U_{\varepsilon/2} + |X| \subseteq |R| \cdot U.$$

That is, $\omega \in \Omega_R(\psi, U)$. □

Remark 23.1. ε, B, R did not depend on $U_1, \ldots, U_m$.

We can therefore restate the proposition as follows:

Corollary 23.1. *There exists a function* $\{boxes\} \to (0, \infty)$ *sending* $B \mapsto \varepsilon(B)$ *such that if*

- $\varepsilon(B) \downarrow 0$ *as* $B \uparrow \mathbb{Z}^d$,
- *For all* B*, if* R *is big enough in terms of* B *and* $U_1, \ldots, U_m$ *and* $v_1, \ldots, v_m$ *as before, then*

$$|\Omega_R(\psi, U)| \geq \prod_{i=1}^m |\Omega_B(\psi, (U_i)_{\varepsilon(B)})|.$$

Corollary 23.2. *There exists a function $s : \mathcal{U} = \{\text{open convex subsets of } \mathbb{R}^n\} \to [-\infty, \infty)$ such that for all $B_n \uparrow \mathbb{Z}^d$ and all $U \in \mathcal{U}$, we have*

$$\frac{1}{|B_n|} \log |\Omega_{B_n}(\psi, U_{2\varepsilon(B_n)})| = s(U) + o(1),$$

where

$$s(U) = \lim_n \frac{1}{|B_n|} \log |\Omega_{B_n}(\psi, U_{2\varepsilon(B_n)})|.$$

Proof. Let

$$s(U) := \sup_{\text{boxes } B} \frac{1}{|B|} \log |\Omega_B(\psi, U_{2\varepsilon(B)})|.$$

We will show that this agrees with the limit. The reason is that $\limsup_n \frac{1}{|B_n|}(\cdots) \leq \sup_{\text{boxes}} = s(U)$, so it is enough to show that $\liminf \geq s(U)$.

Let $B_n \uparrow \mathbb{Z}^d$ and fix a box B. Once B_n is big enough in terms of B, we can use the previous corollary to get

$$|\Omega_{B_n}(\psi, V)| \geq \prod_{i=1}^{m} |\Omega_B(\psi, (V_i)_{\varepsilon(B)})|$$

for all $V \in \mathcal{U}$, where m is the cardinality of a maximal packing of B-translates into B_n. Hence,

$$\frac{1}{|B_n|} \log |\Omega_{B_n}(\psi, V)| \geq \underbrace{\frac{m}{|B_n|}}_{1/|B|+o(1)} \log |\Omega_B(\psi, V_{\varepsilon(B)})|.$$

Apply this with $V = U_{2\varepsilon(B_n)}$. We get

$$\begin{aligned} \frac{1}{|B_n|} \log |\Omega_{B_n}(\psi, U_{2\varepsilon(B_n)})| &\geq \left(\frac{1}{|B|} + o(1)\right) \log |\Omega_B(\psi, U_{2\varepsilon(B)+\varepsilon(B)})| \\ &\geq \left(\frac{1}{|B|} + o(1)\right) \log |\Omega_B(\psi, U_{2\varepsilon(B)})| \end{aligned}$$

if n is big enough. Let $n \to \infty$ to get

$$\liminf_n \frac{1}{|B_n|} \log |\Omega_{B_n}(\psi, U_{2\varepsilon(B_n)})| \geq \frac{1}{|B|} \log |\Omega_B(\psi, U_{2\varepsilon(B)})|.$$

Take the sup over B and get $\lim_n = s(U)$. $\square$

24 Thermodynamic Limits for Counting Empirical Measures

24.1 Recap + rest of proof of the thermodynamic limit

In our lattice models, we have an alphabet $|A| < \infty$ of local states. If $W \subseteq \mathbb{Z}^d$ is finite, then an observable is a function $\psi : A^W \to \mathbb{R}^r$. For a box B and $\omega \in A^B$,

$$\Psi_B(\omega) = \sum_{v+W \subseteq B} \psi(\omega_{v+W}).$$

We wanted to measure the size of

$$\Omega_B(\psi, U) = \{\omega \in A^B : \tfrac{1}{|B|}\Psi_B(\omega) \in U\}.$$

We were trying to prove the existence of the thermodynamic limit in this situation:

Theorem 24.1. *There exists a concave, upper semicontinuous function $s : \mathbb{R}^r \to [-\infty, \infty)$ such that*

(a) $\max_x s(x) = \log |A|$.

(b) If either $U \cap \{s > -\infty\} \neq \varnothing$ or $\overline{U} \cap \overline{\{s > -\infty\}} = \varnothing$, then

$$|\Omega_B(\psi, U)| = \exp\left(|B| \cdot \sup_{x \in U} s(x) + o(|B|)\right).$$

Last time, we showed that there is a function boxes $\to (0, \infty)$ sending $B \mapsto \varepsilon(B)$ such that $\varepsilon(B) \to 0$ as $B \uparrow \mathbb{Z}^d$ and

$$|\Omega_B(\psi, U_{2\varepsilon(B)})| = \exp(|B| \cdot s(U) + o(|B|))$$

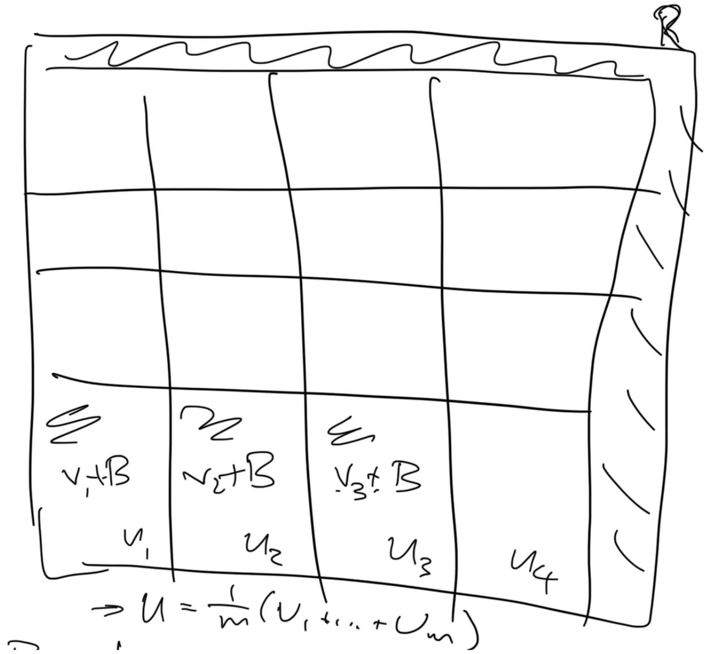
for some $s(U) \in [-\infty, \infty)$, where $U_{2\varepsilon(B)} := \{x : \overline{B_{2\varepsilon(B)}}(x) \subseteq U\}$. We can define

$$s(x) = \inf\{s(U) : U \ni x \text{ is open, convex}\}.$$

This s is automatically upper semicontinuous.

Last time, we showed the estimate that if R is big enough, then

$$|\Omega_R(\psi, U_{2\varepsilon(R)})| \geq \prod_{i=1}^{m} |\Omega_B(\psi, (U_i)_{2\varepsilon(B)})|.$$



Here is the rest of the proof of the theorem:

Proof. If $1/2+O(1/m)$ of the U_is are U and $1/2+O(1/m)$ of them are U', then this inequality gives

$$|\Omega_R(\psi, (\tfrac{1}{2}U+\tfrac{1}{2}U')_{2\varepsilon(R)+O(1/n)})| \geq |\Omega_B(\psi, U_{2\varepsilon(B)})|^{m/2+o(1)} \cdot |\Omega_B(\psi, U'_{2\varepsilon(B)})|^{m/2+o(1)}.$$

Let $R \uparrow \mathbb{Z}^d$ and then $B \uparrow \mathbb{Z}^d$, so we get

$$s\left(\frac{1}{2}U+\frac{1}{2}U'\right) \geq \frac{1}{2}(s(U)+s(U')).$$

Next we show that $s(U)=\sup_{x\in U} s(x)$. As before, this follows if $s(U)=\sup\{s(K) : K \subseteq U \text{ compact, convex}\}$. This works the same as in the non-interacting case because

$$\lim_{B\uparrow\mathbb{Z}^d} \frac{1}{|B|}\log|\Omega_B(\psi, U_{2\varepsilon(B)})| = \sup_B \frac{1}{|B|}\log|\Omega_B(\psi, U_{2\varepsilon(B)})|.$$

So if c is $<$ this, then there is a box B such that $\frac{1}{|B|}\log|\Omega_B(\psi, U_{2\varepsilon(B)})| \geq c$. There exists a compact set K such that $\Omega_B(\psi, U_{2\varepsilon(B)}) = \Omega_B(\psi, K)$. Take $\frac{1}{|B|}\log|\cdot|$, let $B \uparrow \mathbb{Z}^d$ and use superadditivity to get $s(K) \geq c$.

So $s(U) = \sup s(K)$, and so $s(U) = \sup_{x \in U} s(x)$. Now we have a concave upper semicontinuous function such that

$$|\Omega_B(\psi, U_{2\varepsilon(B)})| = \exp\left(|B| \cdot \sup_{x \in U} s(x) + o(|B|)\right).$$

for all open convex U. If we remove the ε, certainly

$$|\Omega_B(\psi, U)| \geq \exp\left(|B| \cdot \sup_{x \in U} s(x) + o(|B|)\right).$$

But if $U \cap \{s > -\infty\} \neq \varnothing$ or $\overline{U} \cap \overline{\{s > -\infty\}} = \varnothing$, then for every $\varepsilon > 0$, there is a $\delta > 0$ such that

$$\sup_{x \in B_\delta(U)} s(x) < \sup_{x \in U} s(x) + \varepsilon, \qquad \text{where } B_\delta(U) = \bigcup_{y \in U} B_\delta(y).$$

But then $U \subseteq (B_\delta(U))_{2\varepsilon(B)}$ for all large enough boxes B, and we have

$$\begin{aligned}
|\Omega_B(\psi, U)| &\leq |\Omega_B(\psi, (B_\delta V)_{2\varepsilon(B)}| \\
&= \exp\left(|B| \cdot \sup_{x \in B_\delta(U)} s(x) + o(|B|)\right) \\
&\leq \exp\left(|B| \cdot (\sup_{x \in U} s(x) + \varepsilon) + o(|B|)\right).
\end{aligned}$$

Therefore,

$$\limsup_{B \uparrow \mathbb{Z}^d} \frac{1}{|B|} \log |\Omega_B(\psi, U)| \leq \sup_{x \in U} s(x) + \varepsilon.$$

Here, ε is arbitrary, so in fact $\lim_{B \uparrow \mathbb{Z}^d} = \sup_U s(x)$.

Here is the last detail: Take $U = \mathbb{R}^r$ to get

$$|A|^{|B|} = |\Omega_B(\psi, U)| = e^{|B| \cdot \sup_x s(x) + o(|B|)}.$$

This gives

$$\sup_{\mathbb{R}^r} s = \log |A|. \qquad \square$$

24.2 The exponent function for measure-valued observables

In the non-interacting case, we described s:

(a) in general via s^*,

(b) explicitly in case ψ is measure-valued.

We will aim to do the same in this setting.

Let's try to approach (b). To set this up, fix again a finite window $W \subseteq \mathbb{Z}^d$ and define $\psi : A^W \to M(A^W) = \mathbb{R}^{A^W}$ sending $a \mapsto \delta_a$. Then we have

$$\Psi_B(\omega)(\{a\}) = \sum_{v+W \subseteq B} \psi(\omega_{v+W})(\{a\}) = |\{v : v + W \subseteq B, \omega_{v+W} = a\}|.$$

We now look at $\frac{1}{|B|}\Psi_B(\omega)$, but it would be cleaner to look at $\frac{1}{|\{v:v+W\subseteq B\}|}\sum_{v+W\subseteq B}\psi(\omega_{v+W})$ so this can be an average. Fortunately, these are asymptotically equivalent, as

$$|\{v : v + W \subseteq B\}| = |B| + o(|B|),$$

so both averages behave the same asymptotically.

Definition 24.1.

$$P_\omega^W = \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \delta_{\omega_{v+W}} \in P(A^W)$$

is called the W**-empirical measure** of $\omega \in A^B$.

What are the possible limits of empirical measures, and what is the exponent function s for those? We will answer this as $W \uparrow \mathbb{Z}^d$ (after everything else). Here is the first observation: Suppose $W \subseteq W'$ and $\pi : A^{W'} \to A^W$ is the projection. Consider $\omega \in A^B$ and

$$\begin{aligned}
\pi_* P_\omega^{W'} &= \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W' \subseteq B} \pi_* \delta_{\omega_{v+W'}} \\
&= \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W' \subseteq B} \delta_{\omega_{v+W}} \\
&= P_\omega^W + O\left(\frac{|W'|}{\text{min side length}(B)}\right),
\end{aligned}$$

where the big O term is a bound on the total variation $\|\pi_* P_\omega^{W'} - P_\omega^W\|$.

This is an "approximate compatibility" of empirical measures. This means that we can look at $\mu \in P(A^{\mathbb{Z}^d})$ and a weak*-neighborhood of the form $U = \{\nu : \|(\pi_W)_*\nu - (\pi_W)_*\mu\| < \varepsilon\}$ for some $\varepsilon > 0$ and finite $W \subseteq \mathbb{Z}^d$. Then consider

$$s(U) = \lim_{B\uparrow\mathbb{Z}^d} \frac{1}{|B|} \log |\{\omega \in A^B : \|P_\omega^W - (\pi_W)_*\mu\| < \varepsilon\}|.$$

This lets us define $s(U)$ for any weak* open set U of this form for some W. These are a base for the weak* topology on $P(A^{\mathbb{Z}^d})$. This will let us find a concave, upper semicontinuous exponent function $s : P(A^{\mathbb{Z}^d}) \to [-\infty, \infty)$.

25 The Entropy Rate of Shift-Invariant Measures

25.1 Recap

Our alphabet is $A^{\mathbb{Z}^d}$ as before, and we have been moving around finite windows $W \subseteq \mathbb{Z}^d$ and looking at what patterns appear. The empirical distribution of x in W is

$$P_x^W = \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \delta_{v+W} \qquad (x \in A^B).$$

Last time, we saw that if U is an open, convex subset of $P(A^W)$ (or $\mathbb{R}^{A^W}$), then

$$\underbrace{|\{x \in A^B : P_x^W \in U\}|}_{=:\Omega_B(W,U)} = e^{|B| \cdot s(U) + o(|B|)},$$

if $U \cap \{s > -\infty\} \neq \varnothing$ or $\overline{U} \cap \overline{\{s > -\infty\}} = \varnothing$. Here, $s(U) = \sup\{s(x) : x \in U\}$. We have not yet verified that if $U \subseteq U_1 \cup \cdots \cup U_k$, then $s(U) \leq \max_i s(U_i)$, but this is a quick check.

25.2 Counting microscopic configurations by their empirical measures — consistency of the entropy rate

If $W \subseteq W'$, B is large, and $\pi : A^{W'} \to A^W$ is the projection, then

$$\pi_* P_x^{W'} = P_x^W + O\left(\frac{|W|}{\text{min-side-length}(B)}\right)$$

As a result, inside $P(A^{\mathbb{Z}^d})$, consider weak* open sets of the form $\widehat{U} := \{\mu \in P(A^{\mathbb{Z}^d}) : \mu_W \in U\}$ for some finite $W \subseteq \mathbb{Z}^d$ and open convex $U \subseteq P(A^W)$, where $\mu \mapsto \mu_W$ is the projection of μ to A^W. These sets form a base $\mathcal{U}$ for the weak* topology on $P(A^{\mathbb{Z}^d})$.

We would like to try to define

$$s(\widehat{U}) := s(U),$$

where the right hand side is defined using the particular window W. We must show that this is consistent with respect to the choice of W: We want $s^{(W)}(U) = s^{(W')}(U')$ whenever $U \subseteq P(A^W)$ is open and convex and $U' = \{\nu \in P(A^{W'}) : \nu_W \in U\}$. This holds because of the result proven last time:

If U and U' are as above, assume $U \cap \{s^{(W)} > -\infty\} \cap \varnothing$ or $\overline{U} \cap \overline{\{s^{(W)} > -\infty\}} = \varnothing$. This condition implies that

$$\inf_{\delta > 0} s^{(W)}(B_\delta(U)) = s^{(W)}(U) = \sup_{\delta > 0} s^{(W)}(U_\delta).$$

Now observe from the aforementioned result that for any $\delta > 0$, if B is large enough,

$$P_x^W \in U \implies (P_x^{W'})_W = P_x^W + O\left(\frac{|W|}{\text{min-side-length}(B)}\right).$$

Hence,

$$|\Omega_B(W,U)| \le |\Omega_B(W',U')|,$$

and similarly,

$$|\Omega_B(W,U)| \ge |\Omega_B(W',U')|.$$

Now let $B \uparrow \mathbb{Z}^d$ and then $\delta \downarrow 0$. Then set $s^{(W')}(U') = s^{(W)}(U)$. We then obtain

$$|\Omega_B(\widehat{U})| = \exp\left(|B| \cdot \sup_{\mu \in \widehat{U}} s(\mu) + o(|B|)\right),$$

as $B \uparrow \mathbb{Z}^d$. Interpret $\Omega_B(\widehat{U})$ as $\Omega_B(W,U)$ for any suitable W and U. Note that the left hand side is not precisely well-defined, but it is asymptotically well-defined by these considerations, so this statement still makes sense. This exponent function s is a concave, upper semicontinuous function on $M(A^{\mathbb{Z}^d})$.

25.3 The entropy rate of shift-invariant measures

Proposition 25.1. *Consider the collection of measures*

$$\{\mu \in M(A^{\mathbb{Z}^d}) : s(\mu) > -\infty\} = \{\mu : \forall B_n \uparrow \mathbb{Z}^d,\ \exists x_n \in A^{B_n} \text{ s.t. } P_{x_n}^W \to \mu_W\ \forall W\}.$$

This is contained in

$$P^T(A^{\mathbb{Z}^d}) = \{\mu \in P(A^{\mathbb{Z}^d}) : \text{shift-invariant, i.e. } T_*^v\mu = \mu\ \forall v \in \mathbb{Z}^d\},$$

where $T^v : A^{\mathbb{Z}^d} \to A^{\mathbb{Z}^d}$ *sends* $\langle a_n\rangle_n \mapsto \langle a_{n-v}\rangle_n$ *and* $(T_*^v\mu)(B) = \mu(T^{-v}(B))$ *for all Borel* $B \subseteq A^{\mathbb{Z}^d}$.

Proof. Here is the proof of shift invariance: Suppose $B_n \uparrow \mathbb{Z}^d$ and $x_n \in A^{B_n}$ are such that $P_{x_n}^W \to \mu_W$ for all finite $W \subseteq \mathbb{Z}^d$. Pick a window V and $a \in A^V$. We will show that $\mu_V(A) = \mu_{V-u}(a)$ for all $u \in \mathbb{Z}^d$.

Pick $W \supseteq V \cup (V-u)$, and let $\psi_1, \psi_2 : A^W \to \{0,1\}$ be defined by

$$\psi_1(b) = \mathbb{1}_{\{b_V = a\}}, \qquad \psi_2(b) = \mathbb{1}_{\{b_{V-u}=a\}}.$$

We know $\mu_W = \lim_n P_{x_n}^W$, and so

$$\mu_V(a) = (\mu_W)_V(a) = \lim_n (P_{x_n}^W)_V(a)$$

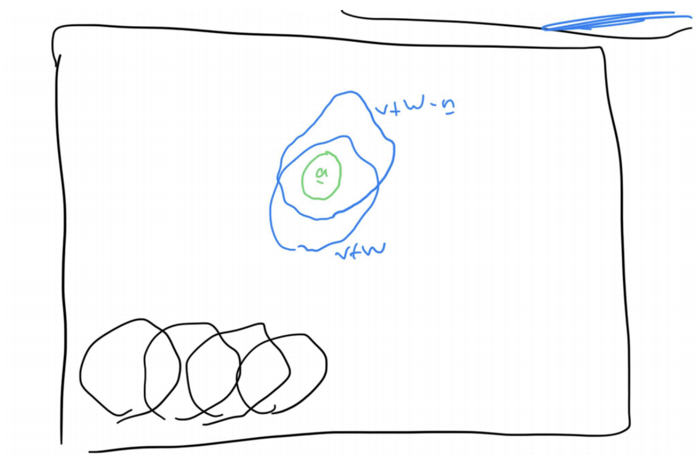
and

$$\mu_{V-u}(a) = \lim_n (P_{x_n}^W)_{V-u}(a).$$

These respectively equal:

$$= \frac{1}{|\{v : v + W \subseteq B_n\}|} |\{v : v + W \subseteq B_n, (x_n)_{v+V} = a\}|,$$

$$= \frac{1}{|\{v : v + W \subseteq B_n\}|} |\{v : v + W \subseteq B_n, (x_n)_{v+V-u} = a\}|,$$



These will agree except for points on the boundary. So the difference is

$$\mu_V(a) - \mu_{V-u}(a) = O\left(\frac{(|v| + |u|)|\text{boundary of } B_n|}{|B_n|}\right) \xrightarrow{n\to\infty} 0.$$

So $T_*^V \mu = \mu$. $\square$

So $\{s > -\infty\} \subseteq P^T(A^{\mathbb{Z}^d})$. We want to generalize the formula "$s(p) = H(p)$ for $p \in P(A)$" from the non-interacting case. To do this we need a digression into the properties of Shannon entropy.

From before, we had that if $p \in P(A)$, then

$$H(p) = -\sum_{a\in A} p(a) \log p(a).$$

Here is some notation: If α is an A-valued random variable and if the distribution of α is p: $\mathbb{P}(\alpha = a) = p(a)$, then $H(\alpha) = H(p)$. We interpret this as a measure of the "uncertainty" in α.

Recall that $0 \leq H(\alpha) \leq \log|A|$, where equality is achieved on the left iff α is deterministic (i.e. $p = \delta_a$ for some letter a) and equality on the right is achieved iff $\alpha \sim \text{Unif}(A)$. Next time, we will discuss some more properties of Shannon entropy and return to $s(\mu)$ for $\mu \in P(A^{\mathbb{Z}^d})$.

26 Basics of Shannon Entropy and Connection to Entropy Rate

26.1 Basic inequalities for Shannon entropy

Definition 26.1. Let A be a finite set with $p \in P(A)$, and let $\alpha \sim p$ be an A-valued random variable. Then

$$H(\alpha) := -\sum_{\alpha \in A} \mathbb{P}(\alpha = a) \log \underbrace{\mathbb{P}(\alpha = a)}_{p(a)} = H(p)$$

is the **Shannon entropy** of α (or of p).

The Shannon entropy quantifies how "uncertain" α is. We have seen that $H(p) \geq 0$ and is $\leq \log |A|$, with equalities achieved with a point mass and with the uniform distribution on $|A|$, respectively.

Next consider random variables α, β with values in A, B. Regard (α, β) as a random variable with values in $A \times B$. The joint distribution is $p_{\alpha,\beta} \in P(A \times B)$. Then

$$\begin{aligned}
H(\alpha, \beta) &= -\sum_{a,b} p_{\alpha,\beta}(a,b) \log p_{\alpha,\beta}(a,b) \\
&= -\sum_{a,b} p_{\alpha,\beta}(a,b) \log \left(p_\alpha(a) \underbrace{p_{\beta|\alpha}(b \mid a)}_{\mathbb{P}(\beta = b | \alpha = a)} \right) \\
&= -\sum_{a,b} p_{\alpha,\beta}(a,b) \log p_\alpha(a) - \sum_{a,b} p_\alpha(a) p_{\beta|\alpha}(b \mid a) \log p_{\beta|\alpha}(b \mid a) \\
&= -\sum_a p_\alpha(a) \log p_\alpha(a) + \sum_a p_\alpha(a) \cdot H(p_{\beta|\alpha}(\cdot \mid a)) \\
&= H(\alpha) + H(\beta \mid \alpha),
\end{aligned}$$

where $H(\beta \mid \alpha) := \sum_a p_\alpha(a) \cdot H(p_{\beta|\alpha}(\cdot \mid a))$.

Here is the generalization of this fact:

Theorem 26.1 (Chain rule)**.**

$$H(\alpha_1, \dots, \alpha_m) = H(\alpha_1) + H(\alpha_2 \mid \alpha_1) + H(\alpha_3 \mid \alpha_1, \alpha_2) + \cdots + H(\alpha_m \mid \alpha_1, \dots, \alpha_{m-1}).$$

We also have the following property.

Lemma 26.1.

$$H(\beta \mid \alpha) \leq H(\beta),$$

and equality holds iff α, β are independent, in which case

$$H(\alpha, \beta) \leq H(\alpha) + H(\beta)$$

Proof.
$$H(\beta \mid \alpha) = \sum_a p_\alpha(a) H(p_{\beta|\alpha}(\cdot \mid a)).$$
By the Law of Total Probability, for all $b \in B$,
$$p_\beta(b) = \sum_\alpha p_\alpha(a) p_{\beta|\alpha}(b \mid a).$$
Since H is strictly concave, Jensen's inequality gives that
$$H(\beta) = H(p_\beta) \geq \sum_\alpha p_\alpha(a) H(p_{\beta|\alpha}(\cdot \mid a)) = H(\beta \mid \alpha).$$
Equality holds in Jensen's inequality iff $p_{\beta|\alpha}(\cdot \mid a) = p_\beta$ whenever $p_\alpha(a) > 0$, i.e. α, β are independent. $\square$

Corollary 26.1.
$$H(\gamma \mid \alpha, \beta) \leq H(\gamma \mid \beta)$$
and similarly with more random variables. Equality holds iff α, γ are conditionally independent given β.

Here is a corollary of the chain rule:

Corollary 26.2. *Let A be a finite set, $p \in P(A)$, and $0 \leq \varepsilon < 1/2$. Suppose $A = B \sqcup C$ with $|B| \leq |C|$ and $p(C) \leq \varepsilon$. Then*
$$H(p) \leq H(\varepsilon, 1-\varepsilon) + (1-\varepsilon)\log|B| + \varepsilon \log|C|.$$

Proof. Let $\alpha \sim p$, and let
$$\beta = \mathbb{1}_B(\alpha) = \begin{cases} 1 & \alpha \in B \\ 0 & \alpha \in C. \end{cases}$$
So $H(\alpha) = H(\alpha) + H(\beta \mid \alpha) = H(\alpha, \beta)$. Expanding via β first instead, we get
$$\begin{aligned} H(\alpha) &= H(\alpha, \beta) \\ &= H(\beta) + H(\alpha \mid \beta) \\ &= H(\beta) + \mathbb{P}(\beta = 1) H(p(\cdot \mid B)) + \mathbb{P}(\beta = 0) H(p(\cdot \mid C)) \\ &\leq H(\varepsilon, 1-\varepsilon) + p(B) \cdot \log|B| + p(C) \cdot \log|C| \\ &\leq H(\varepsilon, 1-\varepsilon) + (1-\varepsilon)\log|B| + \varepsilon \log|C|. \end{aligned}$$
$\square$

Here is the last information-theoretic inequality we need.

Theorem 26.2 (Shearer's inequality). *Let $\alpha_1, \ldots, \alpha_m$ be valued in $A_1, \ldots, A_m$, let $\mathcal{S} \subseteq \mathscr{P}(\{1, \ldots, m\})$, and let $k \geq 1$. Assume that every $i \in \{1, \ldots, m\}$ is contained in $\geq k$ members of $\mathcal{S}$. Then*

$$H(\alpha_1, \ldots, \alpha_m) \leq \frac{1}{k} \sum_{S \in \mathcal{S}} H(\alpha_i : i \in S).$$

Proof. Here is the proof in the case $m = 3$ and $\mathcal{S} = \{\{1,2\}, \{1,3\}, \{2,3\}\}$ ($k = 2$); the argument generalizes well.

$$\begin{aligned}
H(\alpha_1, \alpha_2) &= H(\alpha_1) + H(\alpha_2 \mid \alpha_1) \\
H(\alpha_1, \alpha_3) &= H(\alpha_1) + \qquad\qquad\quad + H(\alpha_3 \mid \alpha_1) \\
H(\alpha_2, \alpha_3) &= \qquad\qquad\; H(\alpha_2) \qquad + H(\alpha_3 \mid \alpha_2)
\end{aligned}$$

Adding together the columns, the first column is $H(\alpha_1)$, the second column is $\geq 2H(\alpha_2 \mid \alpha_1)$, and the third column is $\geq 2H(\alpha_3 \mid \alpha_1, \alpha_2)$. So we get

$$\begin{aligned}
H(\alpha_1, \alpha_2) + H(\alpha_1, \alpha_3) + H(\alpha_2, \alpha_3) &= 2[H(\alpha_1) + H(\alpha_2 \mid \alpha_1) + H(\alpha_3 \mid \alpha_1, \alpha_2)] \\
&= 2H(\alpha_1, \alpha_2, \alpha_3).
\end{aligned}$$

□

26.2 Applying Shearer's inequality to lattice models

Here is a corollary of Shearer's inequality.

Corollary 26.3. *Let $W, B \subseteq \mathbb{Z}^d$ be finite with $0 < |A| < \infty$ and $\mu \in P(A^B)$. Then*

$$H(\mu) \leq \frac{1}{|W|} \sum_{v + W \subseteq B} H(\mu_{v+W}) + O\left(\frac{\log |A| \cdot |B| \cdot \operatorname{diam}(W)}{\text{min-side-length}(B)} \right).$$

Proof. Let $\mathcal{S}_0 = \{v + W : v + W \subseteq B\}$, and define $\mathcal{S}_1 = \{(v + W) \cap B : (v + W) \cap B \neq \varnothing\}$. Then $\mathcal{S}_0 \subseteq \mathcal{S}_1$, and $\mathcal{S}_1$ covers every element of B exactly $|W|$-many times. Apply Shearer's inequality to get

$$H(\mu) \leq \frac{1}{|W|} \sum_{(v+W) \cap B \in \mathcal{S}_1} H(\mu_{(v+W) \cap B}) = \frac{1}{|W|} \sum_{\mathcal{S}_0} H(\mu_{v+W}) + \text{error}.$$

The number of terms put into the error is $|\mathcal{S}_1 \setminus \mathcal{S}_0| = O(\frac{\operatorname{diam}(W) \cdot |B|}{\text{min-side-length}(B)})$. Each of these terms is $\leq \log |A^W| = |W| \cdot \log |A|$. □

Now return to shift-invariant measures $\mu \in P^T(A^{\mathbb{Z}^d})$.

Lemma 26.2. *The limit $\lim_{B \uparrow \mathbb{Z}^d} \frac{1}{|B|} H(\mu_B)$ exists, and*

$$\lim_{B \uparrow \mathbb{Z}^d} \frac{1}{|B|} H(\mu_B) = \inf_B \frac{1}{|B|} H(\mu_\beta).$$

Here is a proof using Shearer's inequality:

Proof. Apply the previous corollary to a shift-invariant measure μ, and observe $\mu_{v+W} = \mu_W$ (up to fixing indexing). Then

$$\begin{aligned}\frac{1}{|B|}H(\mu_B) &\leq \frac{1}{|B|}\sum_{v+W\subseteq B}\frac{1}{|W|}H(\mu_W) + o(1)\\ &= \frac{|\{v : v+W \subseteq B\}|}{|B|}\cdot\frac{1}{|W|}H(\mu_W) + o(1)\\ &\leq \frac{1}{|W|}H(\mu_W) + o(1).\end{aligned}$$

So in fact,

$$\lim_{B\uparrow\mathbb{Z}^d}\frac{1}{|B|}H(\mu_\beta) = \inf_{|W|<\infty}\frac{1}{|W|}H(\mu_W).$$

□

Definition 26.2. The quantity

$$h(\mu) = \lim_{B\uparrow\mathbb{Z}^d}\frac{1}{|B|}H(\mu_B) \qquad (\mu \in P^T(A^{\mathbb{Z}^d}))$$

is called the **entropy rate** of μ.

The entropy rate satisfies

$$0 \leq h(\mu) \leq H(\mu_{\{0\}}).$$

Theorem 26.3. *$s = h$ on $P^T(A^{\mathbb{Z}^d})$, and so $\{s > -\infty\} = \{s \geq 0\} = P^T(A^{\mathbb{Z}^d})$.*

27 Equality of Entropy Rate and the Exponent Function

27.1 Proving that the entropy rate equals the exponent function for lattice models

In our current setting, we have a shift-invariant measure $\mu \in P^T(A^{\mathbb{Z}^d})$, and

$$s(\mu) = \inf_{W,U \ni \mu_W} \lim_{B \uparrow \mathbb{Z}^d} \frac{1}{|B|} \log |\Omega_B(U)|,$$

where $\Omega_B(U) = \{x \in A^B : P_x^W \in U\}$.

The Shannon entropy is

$$H(\mu_F) = -\sum_{y \in A^F} \mu_F(y) \log \mu_F(y),$$

and the entropy rate is

$$h(\mu) = \lim_B \frac{1}{|B|} H(\mu_B) = \inf_W \frac{1}{|W|} H(\mu_W)$$

Theorem 27.1.

$$s(\mu) = h(\mu).$$

To prove this, we will use two tools from last lecture:

Lemma 27.1. *If $A = B \sqcup C$, $p \in P(A)$, and $p(C) \leq \varepsilon \leq 1/2$, then*

$$H(p) \leq H(\varepsilon, 1-\varepsilon) + (1-\varepsilon) \log |B| + \varepsilon \log |A|.$$

Last time, we assumed $|B| \leq |C|$ in the above and got $\log |C|$ instead of $\log |A|$; this version is more useful. We also have the following corollary of Shearer's inequality:

Lemma 27.2. *If $W, B \subseteq \mathbb{Z}^d$ are finite and $\mu \in P(A^B)$, then*

$$H(\mu) \leq \frac{1}{|W|} \sum_{v : v+W \subseteq B} H(\mu_{v+W}) + O\left(\frac{\log |A| \cdot |B| \cdot \operatorname{diam}(W)}{\text{min-side-length}(B)} \right).$$

Now let's prove the theorem:

Proof. We will prove the inequalities $\geq$ and $\leq$ separately.

($\geq$): Denote $h = h(\mu)$. We want to show for any W, μ_W, we have

$$\frac{1}{|B|} \log |\{x \in A^B : P_x^W \in U\}| \geq h - o(1)$$

as $B \uparrow \mathbb{Z}^d$. Suppose we knew that

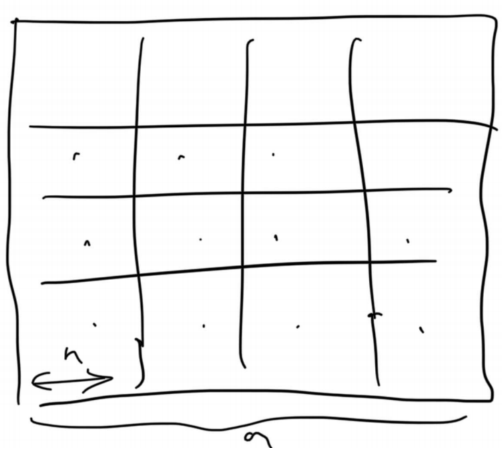
$$\mu_B(\{x \in A^B : P_x^W \in U\}) = 1 - o(1)$$

as $B \uparrow \mathbb{Z}^d$. Then, by the first lemma, we get

$$\begin{aligned}
\frac{1}{|B|} H(\mu_B) &\leq \frac{1}{|B|} H(\varepsilon(B), 1 - \varepsilon(B)) + \frac{1 - \varepsilon(B)}{|B|} \log |\Omega_B(W, U)| + \frac{\varepsilon(B)}{|B|} \log |A^B| \\
&\leq \frac{\log 2}{|B|} + \frac{1}{|B|} \log |\Omega_B(W, U)| + \varepsilon(B) \log |A| \\
&= \frac{1}{|B|} \log |\Omega_B(W, U)| + o(1)
\end{aligned}$$

as $B \uparrow \mathbb{Z}^d$. So $s(\mu) \geq h(\mu)$ if we have this property.

In general, this property does not hold, so we need a replacement for it. To do this, we may restrict attention to $B_n = \{0, \dots, n^2 - 1\}^d$. Let $\mathcal{Q}_n$ be the natural partition of B_n into boxed of side length n.



Let $\nu_n = \bigtimes_{Q \in \mathcal{Q}_n} \mu_Q$. Observe that $H(\nu_n) = \sum_W H(\mu_Q)$, so

$$\frac{1}{n^{2d}} H(\nu_n) = \frac{1}{n^d} H(\mu_{\{0, \dots, n-1\}^d}) \to h(\mu)$$

as $n \to \infty$. Also, if $x \in A^{B_n}$,

$$\begin{aligned}
P_x^W &= \frac{1}{|\{v : v + W \subseteq B_n\}|} \sum_{v + W \subseteq B_n} \delta_{x_{v+W}} \\
&= \frac{1}{|\{v : v + W \subseteq B_n\}|} \sum_{Q \in \mathcal{Q}_n} \sum_{v + W \subseteq Q} \delta_{x_{v+W}} + \text{boundary terms}
\end{aligned}$$

If we do the same analysis as we did before with this type of partition, we get

$$= \frac{1}{n^d} \sum_{Q \in \mathcal{Q}_n} P^W_{x_Q} + o(1).$$

The $P^W_{x_Q}$ are independent if $x \sim \nu_n$. The average of $P^W_{x_Q}(a)$ (with $a \in A^W$) over $x_Q \sim \mu_W$ is $\mu_W(a)$. So by the weak law of large numbers, $P^W_x \in U$ with high probability if $x \sim \nu_n$ and n is large enough. So the property we assumed works if we replace μ_B by ν_n. Now complete the argument as before.

($\le$): We want to show that if $\varepsilon > 0$, W is large enough, and $U \ni \mu_W$ is small enough, then

$$\frac{1}{|B|} \log |\Omega_B(W,U)| \le h + \varepsilon + o(1)$$

as $B \uparrow \mathbb{Z}^d$. To estimate the left hand side, let ν_B be the uniform probability distribution on $\Omega_B(W,U)$, so the left hand side equals $\frac{1}{|B|}$. So the second lemma (the corollary of Shearer's inequality) tells us that

$$\frac{1}{|B|} H(\nu_B) \le \frac{1}{|B|} \sum_{v+v+W \subseteq B} H(\nu_{v+W}) + \underbrace{O\left(\frac{\log |A| \cdot \operatorname{diam}(W)}{\text{min-side-length}(B)} \right)}_{=o_B(1)}.$$

What can we say about the family ν_{v+W}, where $v + W \subseteq B$? Observe that

$$\begin{aligned}
\frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \nu_{v+W} &= \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \int \delta_{x_{v+W}} \, d\nu_B(x) \\
&= \int \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \delta_{x_{v+W}} \, d\nu_B(x) \\
&= \int P^W_x \, d\nu_B(x) \\
&=: \widehat{\mu} \in U.
\end{aligned}$$

Since Shannon entropy is concave and continuous, we get

$$\frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} H(\nu_{v+W}) \le H(\widehat{\mu}) \le H(\mu_W) + \varepsilon$$

if we choose U small enough.

If we put it all together, we get

$$\frac{1}{|B|} \log |\Omega_B(W,U)| = \frac{1}{|B|} H(\nu_B) \le (1 + o_B(1)) \frac{1}{|W|} (H(\mu_W) + \varepsilon) + o_B(1).$$

So for every ε and W, there is a U such that

$$\lim_B \frac{1}{|B|} \log |\Omega_B(W,U)| \leq \frac{1}{|W|} H(\mu_W) + \varepsilon.$$

So, if we choose W large enough depending on ε, we get that the left hand side is $\leq h + 2\varepsilon$, Since ε is arbitrary, we get $s(\mu) \leq h(\mu)$. $\square$

27.2 A digression concerning ergodic measures

If $\mu \in P^T(A^{\mathbb{Z}^d})$, B is a large box, and $x \in A^{\mathbb{Z}^d}$, then

$$\begin{aligned} P_x^W &= \frac{1+o(1)}{|B|} \sum_{v+W \subseteq B} \delta_{x_{v+W}} \\ &= (1+o(1)) \cdot \frac{1}{|B|} \sum_{v \in B} \delta_{x_{v+W}}. \end{aligned}$$

When do we have $P_{x_B}^W \to \mu_W$ in weak* as $B \uparrow \mathbb{Z}^d$ when $x \sim \mu$? Equivalently, we test against $\psi : A^{\mathbb{Z}^d} \to \mathbb{R}$ dependent only on coordinates in W: When do we have

$$\mu\left(\left\{x \in A^{\mathbb{Z}^d} : \left|\frac{1}{|B|} \sum_{v \in B} \psi(x_{v+W}) - \int \psi \, d\mu\right| < \varepsilon\right\}\right) \to 1$$

as $B \uparrow \mathbb{Z}^d$? Write $\frac{1}{|B|} \sum_{v \in B} \psi(x_{v+W}) = \frac{1}{|B|} \sum_{v \in B} \psi(T^v x)$. Then we really want

$$\frac{1}{|B|} \sum_{v \in B} \psi \circ T^v \to \int \psi \, d\mu$$

in probability for all ψ.

Theorem 27.2 (Mean Ergodic Theorem)**.** *Let (X, μ) be a probability space (e.g. above $X = A^{\mathbb{Z}^d}$). Let $(T^n)_{n \in \mathbb{Z}^d}$ be an action on X that preserves μ (e.g. above this equals translation). The following are equivalent:*

1. *For all $\psi \in L^1(\mu)$, we have*

$$\frac{1}{|B|} \sum_{v \in B} \psi \circ T^v \to \int \psi \, d\mu$$

   in L^1 as $B \uparrow \mathbb{Z}^d$.

2. *The system (X, μ, T) is **ergodic**: there is no measurable partition $X = Y \sqcup Z$ such that $T^v(Y) = Y$ and $T^v(Z) = Z$ for all v and $\mu(Y), \mu(Z) > 0$.*

28 Variational Principles for the Entropy Rate

28.1 Recap

Last time, we showed that

$$\begin{aligned} s(\mu) &:= \inf_{W,U\ni\mu} \lim_{B\uparrow\mathbb{Z}^d} \frac{1}{|B|}\log|\{x\in A^B : P_x^W \in U\}| \\ &= \begin{cases} h(\mu) := \lim_B \frac{1}{|B|}H(\mu_B) & \text{if } \mu\in P^T \\ -\infty & \text{otherwise.}\end{cases}\end{aligned}$$

Here, we extend h by $h(\mu)=-\infty$ if $\mu\notin P^T$. Then $h: M(A^{\mathbb{Z}^d})\to[-\infty,\log|A|]$ is concave and upper semicontinuous, and the set $\{h>-\infty\}=\{h\geq 0\}=P^T$. The upper bound $\log|A|$ is achieved when $\mu=\mathrm{Unif}_A^{\times\mathbb{Z}^d}$.

Now, we will see two variational principles.

28.2 The first variational principle

Theorem 28.1. *Let $\psi : A^{\mathbb{Z}^d}\to\mathbb{R}^r$ depend only on coordinates in a finite $W\subseteq\mathbb{Z}^d$. For $x\in\mathbb{R}^r$, let*

$$s(\psi,y) = \inf_{V\ni x}\lim_{B\uparrow\mathbb{Z}^d}\frac{1}{|B|}\log|\{x\in A^B : \tfrac{1}{|B|}\Psi_B(x)\in V\}|,$$

where the inf is over open, convex neighborhoods of x in $\mathbb{R}^r$. Then

$$\begin{aligned} s(\psi,y) &= \sup\{h(\mu):\mu\in P^T, \langle\psi,\mu\rangle = y\} \\ &= \sup\{h(\mu):\mu\in P, \langle\psi,\mu\rangle = y\}.\end{aligned}$$

with the convention that $\sup\varnothing=-\infty$.

Proof.

$$\begin{aligned}\frac{1}{|B|}\Psi_B(x) &= \frac{1}{|B|}\sum_{v+W\subseteq B}\psi(T^v x) \\ &= \frac{1}{|\{v : v+W\subseteq B\}|}\sum_{v+W\subseteq B}\psi(T^v x) + o(|B|) \\ &= \langle\psi, P_x^W\rangle + o(|B|),\end{aligned}$$

This gives ($\geq$): For any $V\subseteq\mathbb{R}^r$ and finite $W\subseteq\mathbb{Z}^d$, we have

$$\frac{1}{|B|}\log|\{x\in A^B : \langle\psi,P_x^W\rangle\in V\}|.$$

The condition $\langle\psi, P_x^W\rangle \in V$ defines any convex neighborhood of any μ such that $\langle\psi,\mu\rangle = y$. So taking $\lim_B$ of the above, we get that it is $\geq h(\mu)$ for any such μ.

Now consider $(\leq)$. Let

$$h = \sup\{h(\mu) : \mu \in P, \langle\psi,\mu\rangle = y\}.$$

The set $\{\mu \in P : \langle\psi,\mu\rangle = y\}$ is compact, so there exists a window W and open convex sets $U_1,\ldots,U_r$ in $P(A^W)$ such that $\{\mu \in P : \langle\psi,\mu\rangle = y\} \subseteq \bigcup_i \{\mu \in P : \mu_W \in U_i\}$ and

$$\frac{1}{|B|}\log|\{x : P_x^W \in U_i\}| \leq (h+\varepsilon) + o(1)$$

for all i. Finally, by compactness again, if $V \subseteq \mathbb{R}^r$ is a small enough neighborhood of y, then

$$\bigcup_i \{\mu \in P : \mu_W \in U_i\} \supseteq \{\mu : \langle\psi,\mu\rangle \in V\}$$

So

$$\frac{1}{|B|}\log|\{x : \langle\psi, P_x^W\rangle \in V\}| \leq \max_i \frac{1}{|B|}\log|\{x : P_x^W \in U_i\}| + \frac{\log s}{|B|} \leq h + \varepsilon$$

as $B \uparrow \mathbb{Z}^d$. Since $\varepsilon > 0$ is arbitrary, we get $s(\psi, y) = h$, as desired. $\square$

Corollary 28.1. *For any convex, open $V \subseteq \mathbb{R}^r$,*

$$\begin{aligned} s(\psi, V) &= \lim_B \frac{1}{|B|}\log|\{x : \langle\psi, P_x^W\rangle \in V\}| \\ &= \sup_{y\in V} s(\psi, y) \\ &= \sup\{h(\mu) : \mu \in P^T, \langle\psi,\mu\rangle \in V\} \\ &= \sup\{h(\mu) : \mu \in P, \langle\psi,\mu\rangle \in V\}. \end{aligned}$$

From this, we can return to interactions giving the total potential energy $\varphi = (\varphi_F)_F$, assumed (for simplicity) to be a finite range interaction. Look at

$$|\{x \in A^B : \tfrac{1}{|B|}\Phi(x) \in I\},$$

where I is a small open interval, and $\Phi_B(x) = \sum_{F\subseteq B}\varphi_F(x_F) = \sum_{F'} |B|\langle\varphi_{F'}, P_x^W\rangle + o(|B|)$. Here, W is a big enough window to see all nonzero translates, and F' runs over one copy of each finite set $\subseteq W$ up to translation. So this set is

$$\left|\left\{x \in A^B : \sum_{F'}\langle\varphi_{F'}, P_x^W\rangle \in I\right\}\right|.$$

$\sum_{F'}\langle\varphi_{F'}, P_x^W\rangle \in I$ is an open, convex condition in $\mathbb{R}^r$, so

$$\frac{1}{|B|}\log\left|\left\{x \in A^B : \sum_{F'}\langle\varphi_{F'}, P_x^W\rangle \in I\right\}\right| \xrightarrow{B\uparrow\mathbb{Z}^d} \sup\left\{h(\mu) : \mu \in P^T, \sum_{F'}\langle\varphi_{F'}, \mu\rangle \in I \approx y\right\}.$$

We can use this result to predict the most likely values of any other observable if there is a unique measure μ that maximizes $h(\mu)$ subject to the constraint $\sum_{F'}\langle\varphi_{F'}, \mu\rangle = y$.

Remark 28.1. There always exists a μ achieving the supremum if the set $\{\mu : \sum_{F'}\langle\varphi_{F'}, \mu\rangle = y\} \neq \varnothing$ by upper semicontinuity of h on the above weak* compact set.

So the key question is when we get uniqueness of that maximizer. We will discuss this next time.

28.3 A variational principle for the Fenchel-Legendre transform of h

To understand the second variational principle, we need to extend the first version from $\varphi : A^{\mathbb{Z}^d} \to \mathbb{R}^r$ to any $\psi \in C(A^{\mathbb{Z}^d})$. To apply ψ "inside a box," given $x \in A^B$, let $\widehat{x}$ be any element of $A^{\mathbb{Z}^d}$ such that $\widehat{x}_B = x$. Given B and $\psi \in C(A^{\mathbb{Z}^d})$, let

$$s_B\psi(x) = \sum_{v\in B}\psi(T^v\widehat{x}).$$

Lemma 28.1. *If $\widehat{x}, \check{x}$ are two choices of extension, then*

$$\left|\sum_{v\in B}\psi(T^v\widehat{x}) - \sum_{v\in B}\psi(T^v\check{x})\right| = o(|B|).$$

Now a fiddly extension of the first variational principle gives

$$\frac{1}{|B|}\log|\{x \in A^B : \tfrac{1}{|B|}s_B\psi(x) \in V\}| = \sup\{h(\mu) : \mu \in P^T, \langle\psi, \mu\rangle \in V\}.$$

This version is good because we can now handle the whole Banach space $C(A^{\mathbb{Z}^d})$, which is the dual of $M(A^{\mathbb{Z}^d})$, equipped with the weak* topology. This leads to a description of the Fenchel-Lengendre transform of h:

Theorem 28.2 (2nd variational principle)**.** *On $C(A^{\mathbb{Z}^d})$,*

$$\begin{aligned} h^*(f) &:= \sup\{h(\mu) - \langle f, \mu\rangle : \mu \in M(A^{\mathbb{Z}^d})\} \\ &= \lim_{B\uparrow\mathbb{Z}^d}\frac{1}{|B|}\log\sum_{x\in A^B} e^{-s_B f(x)}. \end{aligned}$$

The $e^{-s_B f(x)}$ are the Gibbs weights that define the canonical distribution on A^B. In ergodic theory and much of mathematical physics, this limit is called the *pressure* of f (denoted $p(f)$). Caution: this is not always the physical pressure of the system.

29 Equilibrium Measures, the D-L-R Equations, and Uniqueness vs Non-uniqueness

29.1 Second variational principle and equilibrium measures

Here is the main variational principle we proved last time: For $f \in C(A^{\mathbb{Z}^d})$, let

$$p(f) := \lim_{B \uparrow \mathbb{Z}^d} \frac{1}{|B|} \log \sum_{x \in A^B} \exp(-s_B f(x)),$$

where

$$s_B f(x) = \sum_{v \in B} f(T^v \widehat{x}),$$

and $\widehat{x}$ is an extension of x to an element of $A^{\mathbb{Z}^d}$. Then

$$p(f) = \sup\{h(\mu) - \langle f, \mu \rangle : \mu \in P^T(A^{\mathbb{Z}^d})\} = h^*(f).$$

Here, p is called the **pressure** of f.[19] (Recall that $h : M(A^{\mathbb{Z}^d}) \to [-\infty, \log |A|]$ is concave and upper semicontinuous.

Proof. Here is a sketch: Note that

$$\begin{aligned} s_B f(x) &= \sum_{v \in B} f(T^v \widehat{x}) \\ &= |B| \frac{1}{|B|} \sum_{v \in B} f(T^v \widehat{x}) \\ &= |B| \langle f, \underbrace{\frac{1}{|B|} \sum_{v \in B} \delta_{T^v \widehat{x}}}_{=:P_x^B} \rangle, \end{aligned}$$

. where $P_x^B = P_x^W + o_{B \uparrow \mathbb{Z}^d}(1)$. So in the limit, it is enough to prove the formula for

$$\sum_{x \in A^B} \exp(-|B| \langle f, P_x^B \rangle)$$

Proof of ($\geq$): Fix $\mu \in P^T$, fix $\varepsilon > 0$, and let $U = \{\nu \in P^T : \langle f, \nu \rangle < \langle f, \mu \rangle + \varepsilon\}$. This is a weak* open, concave neighborhood of μ. Now

$$\sum_{x \in A^B} \exp(\cdots) \geq \sum_{x : P_x^B \in U} \exp(-|B| \langle f, P_x^B \rangle)$$

[19] This is not always the same thing as pressure in physics. Sometimes, physicists will call this the **Massieu function**.

$$
\begin{aligned}
&\geq \sum_{x:P_x^B\in U} \exp(-|B|(\langle f,\mu\rangle+\varepsilon)) \\
&\geq \underbrace{|\Omega_B|}_{\geq \exp(|B|h(\mu)+o(|B|))} \exp(-|B|\langle f,\mu\rangle - \varepsilon|B|).
\end{aligned}
$$

So we get

$$
\frac{1}{|B|}\log(\cdots) \geq h^*(\mu) - \langle f,\mu\rangle - \varepsilon - o(1).
$$

Since ε is arbitrary, we get $(\geq)$.

Proof of $(\leq)$: Let $\varepsilon > 0$, and pick a finite cover $P \subseteq U_1 \cup \cdots \cup U_r$ such that

$$
\sup_{\mu\in U_i} h(\mu) + \sup_{\mu\in U_i}(-\langle f,\mu\rangle) \leq h + \varepsilon,
$$

where h is the desired right hand side. Now

$$
\begin{aligned}
\sum_x \exp(-s_B f(x)) &\leq \sum_{i=1}^n \sum_{x:P_x^B\in U_i} \exp(-s_B f(x)) \\
&\leq \sum_{i=1}^r \exp(s(U_i)\cdot|B| + o(|B|) + \sup_{\mu\in U_i}(-\langle f,\mu_i\rangle) + \varepsilon \\
&\leq \sum_{i=1}^r \exp\left(\sup_{\mu\in U_i}(h(\mu)-\langle f,\mu\rangle) + 2\varepsilon|B|\right) \\
&\leq r\cdot\max_i(\cdots).
\end{aligned}
$$

Apply this to $\frac{1}{|B|}\log(\cdots)$ to get that this is $\leq RHS + 2\varepsilon + o(1)$. $\square$

We have a max, rather than a sup:

$$
p(f) = \max\{h(\mu) - \langle f,\mu\rangle : \mu\in P^T\}.
$$

Definition 29.1. $\mu \in P^T$ is an **equilibrium measure** for f if $h(\mu) - \langle f,\mu\rangle = p(f)$.

Observe that

$$
\sum_{x\in A^B} \exp(-|B|\cdot\langle f, P_x^B\rangle) = \exp(p(f)\cdot|B| + o(|B|)).
$$

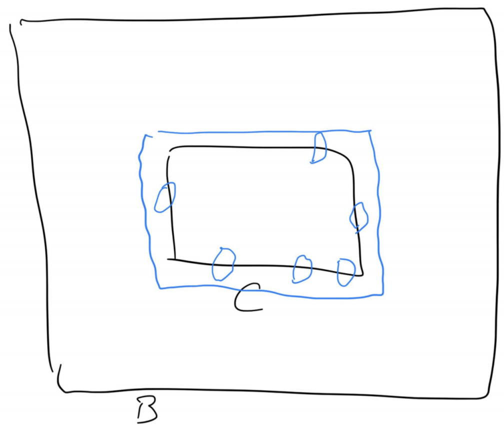
Given $\mu \in P^T$ and a small enough neighborhood U, we have

$$
\sum_{x:P_x^B\in U} \exp(-|B|\langle f,P_x^B\rangle)\exp\left((h(\mu)-\langle f,\mu\rangle\pm\varepsilon)|B| + o(|B|)\right)
$$

Microscopic states will look very similar to equilibrium measures. Equilibrium measures measures always exist. If an equilibrium measure is unique, then we can integrate against it to predict values of any other observable. If there are multiple equilibrium measures, they describe a system with several possible phases of a given temperature.

29.2 The D-L-R equations and uniqueness vs non-uniqueness

The next stage in the story is how to characterize equilibrium measures. For simplicity, we will study a finite-range interaction $\varphi = \varphi_F$ for some fixed finite $F \subseteq \mathbb{Z}^d$. Consider $C \subseteq B$:



Look at

$$\Phi_B(x) = \sum_{v \in F \subseteq B} \varphi(x_{v+F})$$

$$= \sum_{v+F \subseteq C} \varphi(x_{v+F}) + \sum_{v+F \subseteq B \setminus C} \varphi(x_{v+F}) + \sum_{\substack{v+F \subseteq B \\ (v+F) \cap C \neq \varnothing \\ (v+F) \cap (B \setminus C) \neq \varnothing}} \varphi(x_{v+F})$$

$$= \Phi_C(x) + \Phi_{B \setminus C}(x) + \Phi^{\text{int}}_{B,C}(x).$$

The canonical distribution on A^B is

$$\begin{aligned} d\gamma(x) &= \frac{1}{Z} \exp(-\Phi_B(x)) \\ &= \frac{1}{Z} \exp(-\Phi_C - \Phi_{B \setminus C} - \Phi^{\text{int}}_{B,C}). \end{aligned}$$

Use this to write

$$\gamma(x_C = y \mid x_{B \setminus C} = z) = \frac{\exp(-\Phi_C(y) - \cancel{\Phi_{B \setminus C}(x)} - \Phi^{\text{int}}(y,z))}{\sum_{y' \in A^C} \exp(-\Phi_c(y') - \cancel{\Phi_{B \setminus C}(z)} - \Phi^{\text{int}}(y',z))}$$

This depends on z on through $z_{\partial_F C}$, where $\partial_F C = [\bigcup_{(v+F) \cap C \neq \varnothing} (v+F)] \setminus C$. So we can define a family of conditional measures for $y \in A^c$ and $z \in A^{\mathbb{Z}^d \setminus C}$:

$$\gamma_{C,z}(dy) = \frac{\exp(-\Phi_C(y) - \Phi^{\text{int}}(y, z_{\partial_F C}))}{\sum_{y' \in A^C} \exp(-\Phi_C(y') - \Phi^{\text{int}}(y', z_{\partial_F C}))}$$

Definition 29.2. The family $\gamma_{C,z}$ for finite $C \subseteq \mathbb{Z}^d$ and $z \in A^{\mathbb{Z}^d \setminus C}$ is called the **specification** associated to φ.

The specification describes the conditional behavior inside C under a canonical distribution. A $\mu \in P(A^{\mathbb{Z}^d})$ satisfies the D-L-R equations with respect to φ if for all finite $C \subseteq \mathbb{Z}^d$, we have

$$\mu(x_C = y \mid x_{\mathbb{Z}^d \setminus C} = z) = \gamma_{C,z}(y)$$

for all C, z. Equivalently, for all $f \in C(A^{\mathbb{Z}^d})$, we have

$$\int f\, d\mu = \iint f(y, x_{\mathbb{Z}^d \setminus C}) \gamma_{C, x_{\mathbb{Z}^d \setminus C}}(dz)\, d\mu(x).$$

Theorem 29.1. *μ is an equilibrium measure for φ if and only if $\mu \in P^T$ and satisfies all D-L-R equations.*

This is much easier to analyze. This is the gateway to theorems such as the following:

Theorem 29.2. *For any local interaction φ, $\beta\varphi$ has a unique equilibrium state for all sufficiently small β. That is, there is a critical $\beta > 0$ such that for all $\beta < \beta_c$, $T > T_c$.*

The above is a corollary of Dobrushin's uniqueness theorem. In general, things are easier at high temperatures because of techniques like that theorem.

On the other hand, if $A = \{-1, 1\}$, $F = \{0, c_1, c_2, \ldots, c_d\}$, then

$$\varphi(x_0, x_{e_1}, \ldots, x_{e_d}) = -\sum_{i=1}^{d} x_0 x_{e_i}.$$

This is the basis for what is called the **Ising model**.

Theorem 29.3 (Peierls)**.** *If β is high enough and $d \geq 2$, then $\beta\varphi$ has multiple equilibrium states.*