MATH 585 – TOPICS IN MATHEMATICAL PHYSICS – FALL 2006
MATHEMATICS OF MEAN FIELD SPIN GLASSES AND THE REPLICA METHOD
LECTURE 1: INTRODUCTION TO CLASSICAL SPIN SYSTEMS

S. STARR
MATHEMATICS DEPARTMENT, UNIVERSITY OF ROCHESTER

CONTENTS

1. THE ISING MODEL

The Ising model is a basic model of equilibrium statistical mechanics. In this first lecture we will use it primarily as a specific example for describing statistical mechanics in general. One of the main goals is to motivate why so people want to calculate the *free energy* and *pressure* in general. That happens in Sections 1 and 2. In Section 3, we state the Gibbs variational principle directly in infinite volumes, i.e., on $\mathbb{Z}^d$. The reason we do this is that it will set up the prototype we will follow when we use de Finetti's theorem to solve the mean-field models, several lectures from now.

1.1 Set-up.

We will use the word "lattice" to mean a graph. But, abusing notation, we will often also use the same name/symbol to denote just the vertex set of the graph. Let Λ be a finite lattice. Given $x, y \in \Lambda$, if there is an edge between x and y, it will be denoted $\langle x, y\rangle$. In particular, a sum whose index set is $\langle x, y\rangle$ is meant to be summed over edges in the graph. As usual, $\langle x, y\rangle$ and $\langle y, x\rangle$ are the same edge. We will often say that x and y are "nearest neighbors", or write $x \sim y$, to mean that they comprise an edge $\langle x, y\rangle$. We will usually define the edge set in words, instead of introducing a notation for it.

Date: September 13, 2006.

A typical example of a lattice is $\Lambda = \mathbb{Z}_N^d$ for some N and d. Here $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$ is the chain $[1, N] = \{1, 2, \dots, N\}$, with periodic boundary conditions $(N + 1 \equiv 1)$. So $\mathbb{Z}_N^d$ is a discrete d-dimensional torus, with a total of N^d vertices. Then there is an edge $\langle x, y\rangle$ in $\mathbb{Z}_N^d$ if an only if there is an edge $\langle x', y'\rangle$ in $\mathbb{Z}^d$ between some pair of points x' and y' projecting to x and y on the torus.

Another family of lattices are the complete graphs, K_N. Thus, $\Lambda = \{1, 2, \dots, N\}$. In the complete graph, every pair x, y is a nearest-neighbor $\langle x, y\rangle$, except that it is typical to exclude self-edges, i.e., loops: there is no edge $\langle x, x\rangle$, not for any $x \in \Lambda$. This graph has a high degree of symmetry. Most of the models we will discuss after the first lecture will involve K_N rather than $\mathbb{Z}_N^d$.

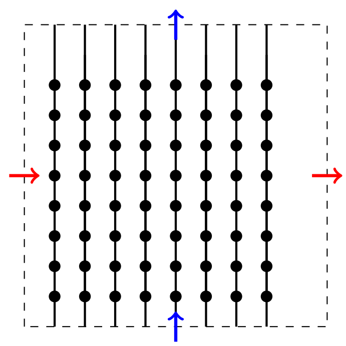

FIGURE 1. A two-dimensional lattice: $\mathbb{Z}_6^2$

The state space (or configuration space) for the Ising model on Λ is: $\Omega_\Lambda = \{+1, -1\}^\Lambda$. This is the set of all spin configurations $(\sigma_x \in \{+1, -1\} \,:\, x \in \Lambda)$. We abbreviate such a configuration, σ. We think of each site x where $\sigma_x = +1$ as having an up-spin $\uparrow$, and each site where $\sigma_x = -1$ as having a down-spin $\downarrow$. This is one of the simplest models for magnetism. For more realistic models, one should consult the literature on quantum spin systems. (See, for example, Daniel Mattis's textbook, *The Theory of Magnetism Made Simple*.) But that will not be the focus of this class.

An equilibrium stat.-mech. model is determined by a Hamiltonian function on the state space: $H_\Lambda : \Omega_\Lambda \to \mathbb{R}$. The Hamiltonian function gives the energies of the spin configurations. For the Ising model, we will write

$$H_\Lambda(\sigma) \;=\; -J \sum_{\langle x,y\rangle} \sigma_x \sigma_y - h \sum_{x\in\Lambda} \sigma_x \,,$$

where J and h are real constants. The first term indicates the preference of two nearby spins to be aligned or misaligned: if $J > 0$ the model is called a ferromagnet; if $J < 0$ it is called an antiferromagnet. The second term is the effect of a uniform external magnetic field, of strength h, in the direction of the up-spins. Of course, h can also be negative, which causes a preference for down-spins instead. We will restrict attention to the ferromagnet, which is slightly simpler.

The energies determine the probability to see any given spin configuration, according to the Boltzmann-Gibbs distribution. This also depends on the temperature. Let T be the temperature, and define a quantity β, called the "inverse-temperature", which is $\beta = (k_B T)^{-1}$

where k_B is Boltzmann's constant. (We will always use β instead of T, so we don't need to explicate k_B.) Then the Boltzmann-Gibbs measure is given by

$$\mu_\beta(\sigma) \;=\; Z_\Lambda(\beta)^{-1}\, e^{-\beta H_\Lambda(\sigma)}\,,$$

for each $\sigma \in \Omega_\Lambda$, where

$$Z_\Lambda(\beta) \;=\; \sum_{\sigma\in\Omega} e^{-\beta H_\Lambda(\sigma)}\,.$$

To define the measure more canonically: for any set $S \subseteq \Omega_\Lambda$, let

$$\mu_\beta(S) \;:=\; Z_\Lambda(\beta)^{-1} \sum_{\sigma\in S} e^{-\beta H_\Lambda(\sigma)}\,.$$

1.2 The partition function, pressure and free energy.

The quantity $Z_\Lambda(\beta)$, the normalization for the Boltzmann-Gibbs measure, is important on its own. It is called the "partition function". Let us define some subsidiary functions from it. We will call the function

$$\mathcal{P}_\Lambda(\beta) \;=\; \log\, Z_\Lambda(\beta)$$

the "thermodynamic potential". We will call the function

$$p_\Lambda(\beta) \;=\; \frac{1}{|\Lambda|}\,\log\, Z_\Lambda(\beta)\,,$$

the "pressure". The "pressure" is a term used, for example, in Simon's book, *The Statistical Mechanics of Lattice Gases*, because of its analogy to the real physical concept of pressure for lattice gases. Let us also define the "free energy density[1]",

$$f_\Lambda(\beta) \;=\; -\beta^{-1} p_\Lambda(\beta)\,.$$

This is related to the ground state energy density, defined as

$$e_0(\Lambda) \;:=\; \frac{1}{|\Lambda|}\min_{\sigma\in\Omega_\Lambda} H_\Lambda(\sigma)\,,$$

because $\lim_{\beta\to\infty} f_\Lambda(\beta) = e_0(\Lambda)$. We prefer the pressure because the extra factor β^{-1} in the free energy becomes a nuisance when taking derivatives.

One reason that the pressure is important is that it gives information for certain natural random variables. To see this, first note that $p_\Lambda(\beta)$ is implicitly a function of h and J as well as β. (In fact it is just a function of βJ and βh, but we prefer to think of all three variables as separate.) We write $p_\Lambda(\beta, h, J)$ to make this explicit. A simple calculation will show that

$$\frac{\partial}{\partial h} p_\Lambda(\beta,h,J) \;=\; \beta\,\mathbf{E}^{\mu_\beta}[m_\Lambda(\sigma)] \quad\text{where}\quad m_\Lambda(\sigma) \;=\; |\Lambda|^{-1}\sum_{x\in\Lambda}\sigma_x\,.$$

(We should also write $\mu_{\beta,h,J}$ if we want to emphasize the dependence on h and J, but we will leave it implicit.) We will call $m_\Lambda(\sigma)$ the "magnetization" and we will define

$$\bar{m}_\Lambda(\beta) \;:=\; \mathbf{E}^{\mu_\beta}[m_\Lambda(\sigma)]\,,$$

---

[1]We should call $p_\Lambda(\beta)$ and $f_\Lambda(\beta)$ the "finite volume approximations" to the pressure and the free energy. These are thermodynamic quantities, properly defined only for an infinitely large sample. But it is convenient and harmless to use the shorter names.

to be the average magnetization of the system. Another simple calculation gives

$$\frac{\partial^2}{\partial h^2} p_\Lambda(\beta; h; J) \;=\; \beta^2 |\Lambda| \mathrm{Var}^{\mu_\beta}(m_\Lambda(\sigma))\,.$$

By taking higher derivatives, one obtains higher cumulants for $m_\Lambda(\sigma)$, viewed as a random variable, modulo certain factors such as the $\beta^2|\Lambda|$ term above.

There are similar calculations, taking partial derivatives with respect to J and β instead of h. In particular, one result of these calculations, and the fact that variance is always positive, is the conclusion that $p_\Lambda(\beta, h, J)$ is convex in β, and it is jointly convex in h and J.

1.3 The Gibbs variational principle.

One of the most important tools in statistical mechanics is the Gibbs variational principle. To begin with, note that there is an obvious variational principle for the ground state energy, namely the definition:

$$e_0(\Lambda) \;=\; \min_{\sigma \in \Omega_\Lambda} |\Lambda|^{-1}\, H_\Lambda(\sigma)\,,$$

We will state a generalization of this, applicable for $\beta < \infty$.

Let $M_1(\Omega_\Lambda)$ be the set of all probability measure on Ω_Λ. The relative entropy of $\nu \in M_1(\Omega_\Lambda)$, relative to $\mu \in M_1(\Omega_\Lambda)$, is defined as

$$S_\Lambda(\nu|\mu) \;:=\; \int_{\Omega_\Lambda} u\left(\frac{d\nu}{d\mu}\right)\, d\mu \;=\; \sum_{\sigma \in \Omega_\Lambda} u\left(\frac{\nu(\{\sigma\})}{\mu(\{\sigma\})}\right)\, \mu(\{\sigma\})\,,$$

if ν is absolutely continuous with respect to μ, where

$$u(x) \;=\; \begin{cases} -x\log(x) & x \in (0, \infty]\,; \\ 0 & x = 0\,. \end{cases}$$

If ν has a singular component with respect to μ, then $S_\Lambda(\nu|\mu) := +\infty$. Let μ_0 be the uniform probability measure on Ω_Λ. (This matches the definition of the Boltzmann-Gibbs measure when $\beta = 0$.) For each probability measure ν on Ω_Λ, define

$$\phi_\Lambda(\beta;\nu) \;:=\; |\Lambda|^{-1}\Big(\mathbf{E}^\nu[H_\Lambda(\sigma)] - \beta^{-1} S(\nu|\mu_0) - \beta^{-1}\log(|\Omega_\Lambda|)\Big)\,.$$

Theorem 1.1 (Gibbs Variational Principle)

$$f_\Lambda(\beta) \;=\; \min_{\nu \in M_1(\Omega_\Lambda)} \phi_\Lambda(\beta;\nu)\,,$$

and, moreover, $\nu = \mu_\beta$ is the unique minimizer.

The proof of the Gibbs variational principle is trivial once one knows the right properties of the relative entropy.

Lemma 1.2 *For any $\mu \in M_1(\Omega_\Lambda)$,*

$$\max_{\nu \in M_1(\Omega_\Lambda)} S_\Lambda(\nu|\mu) \;=\; 0\,,$$

and the unique maximizer is $\nu = \mu$.

Proof of Lemma 1.2. Note that u is strictly concave. Therefore, $S_\Lambda(\cdot|\mu)$ is strictly concave on $M_1(\Omega_\Lambda)$. Therefore if any local maximizer exists, it is the unique global maximizer. By Jensen's inequality,

$$S_\Lambda(\nu|\mu) \,=\, \mathbf{E}^\mu\Big[u\Big(\frac{d\nu}{d\mu}(\sigma)\Big)\Big] \,\leq\, u\Big(\mathbf{E}^\mu\Big[\frac{d\nu}{d\mu}(\sigma)\Big]\Big) \,=\, 0\,,$$

because $\mathbf{E}^\mu[d\nu/d\mu] = \nu(\Omega_\Lambda) = 1$. On the other hand, by a direct calculation, $S_\Lambda(\mu|\mu) = 0$. Therefore, μ is the unique maximizer of $S_\Lambda(\cdot|\mu)$. □

Proof of Theorem 1.1. By a direct calculation

$$\phi_\Lambda(\beta;\nu) \,=\, f_\Lambda(\beta) - \beta^{-1} S_\Lambda(\nu|\mu_\beta)\,.$$

So the theorem follows from Lemma 1.2. □

The Gibbs variational principle can equally well be stated for the pressure as for the free energy.

Corollary 1.3 *Define the functional*

$$\psi_\Lambda(\beta;\nu) \,:=\, |\Lambda|^{-1}\Big(S_\Lambda(\nu|\mu_0) - \mathbf{E}^\nu[\beta H_\Lambda(\sigma)] + \log(|\Omega_\Lambda|)\Big)\,.$$

Then

$$\max_{\nu\in M_1(\Omega_\Lambda)} \psi_\Lambda(\beta;\nu) \,=\, p_\Lambda(\beta)\,,$$

and $\nu = \mu_\beta$ is the unique maximizer.

2. Phase Transitions

The most important facet of statistical mechanical systems is the possibility to have phase transitions. For spin systems, phase transitions can only occur in the thermodynamic limit. In order to be precise, let us now restrict attention just to the set of lattices $\Lambda_N = \mathbb{Z}_N^d$, for some fixed dimension d. Then the thermodynamic limit is $N \to \infty$.

For each $N \in \mathbb{N}_+$, $p_N(\beta, h, J)$ is a smooth function of β, J and h. We define the thermodynamic pressure to be

$$p(\beta,h,J) \,=\, \lim_{N\to\infty} p_N(\beta,h,J)\,,$$

if the limit exists. This is a pointwise limit of convex functions. Therefore, it is convex, wherever it is defined. In particular, this means that it is continuous where it is defined, and even has one-sided directional derivatives.

If the pressure is not smooth at some point, then that point is said to correspond to a phase transition. One can be more specific, however. If, for example, $\partial p/\partial h$ has a jump discontinuity at (β, J, h), then one says there is a first-order phase transition there, with respect to the variable h. If, on the other hand, $\frac{\partial^2}{\partial h^2} p(\beta; h; J)$ is undefined, then one says there is a second-order phase transition with respect to h. Of course, it is possible to have phase transitions with respect to the other variables, β and J, as well.

2.1 Pictures for the Ising model.

Let us state the known results for the Ising model in dimensions $d \geq 2$. We will state some things which are not mathematically proved, or at least not easily proved.

Ackowledgement. *We have copied the pictures here (not entirely faithfully) from Minlos's nice, short book, "Introduction to Mathematical Statistical Mechanics".*

Let us restrict attention to the ferromagnet. Then we fix $J = 1$, which is no loss of generality because of the redundancy in (β, h, J). If β is small enough, or if $|h|$ is large enough, then $p(\beta, h) := p(\beta, h, J = 1)$ can be proved to be analytic. But, for large β one can prove that there is a first-order phase transition with respect to the variable h, at $h = 0$. In other words, defining

$$\bar{m}(\beta, h) \; := \; \lim_{N\to\infty} \bar{m}_{\Lambda_N}(\beta, h, J = 1)\,,$$

there is a jump discontinuity in $\bar{m}(\beta, h)$, across $h = 0$, for each $\beta > \beta_0$.

What is widely understood is that $p(\beta, h)$ is smooth everywhere except on the ray $\{(\beta, h) : \beta > \beta_c, h = 0\}$, and there is a jump discontinuity in $\bar{m}(\beta, h)$, with respect to h, across this ray. (Of course, $\beta_0 > \beta_c$: one cannot prove more than is true.)

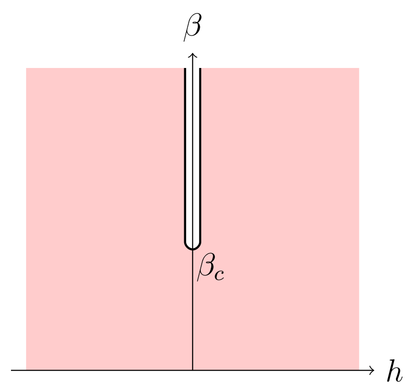

FIGURE 2. The region where $p(\beta, h)$ is smooth.

Here $(\beta, h) = (\beta_c, 0)$ is known as the critical point and it is characterized by its own special properties. As mentioned before, there is a jump discontinuity in $\bar{m}(\beta, h)$ at $h = 0$ for $\beta > \beta_c$. Let us denote this jump as $[\bar{m}(\beta, 0)]_h := \bar{m}_+(\beta, 0) - \bar{m}_-(\beta, 0)$, where

$$\bar{m}_\pm(\beta, 0) \; = \; \lim_{h\to 0^\pm} \bar{m}(\beta, h)\,.$$

But $[\bar{m}(\beta, 0)]_h$ goes to 0, continuously, as β approaches β_c from the right.

The behavior is supposed to be power-law: $[\bar{m}(\beta, 0)]_h = C(\beta - \beta_c)^{\mathrm{\beta}}(1 + o(1))$ for $\beta > \beta_c$. Here $\mathrm{\beta}$ is a number not to be confused with the inverse temperature β, $o(1)$ symbolizes a quantity which vanishes as $\beta \to \beta_c^+$, and C is a non-"universal" constant, which is therefore of less interest. The constant $\mathrm{\beta}$ is called a "critical exponent". It is one of the numbers that physicists care about when they consider critical phenomena. It is dependent on d. But supposedly it is independent of some of the microscopic structure defining the model, although it is dependent on the symmetry group. The idea is that at criticality there is a divergent

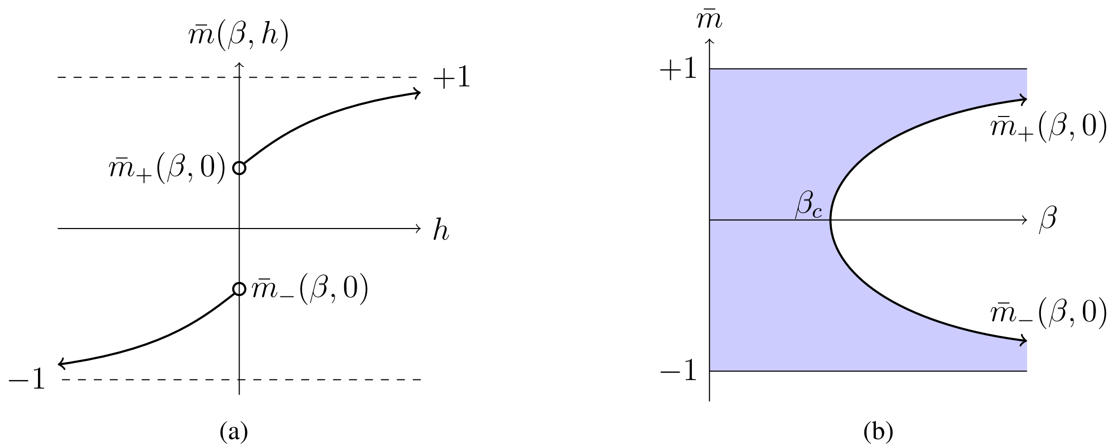

FIGURE 3. (a) $\bar{m}(\beta, h)$ as a function of h, for a fixed $\beta > \beta_c$; (b) the region of accessible values for $\bar{m}(\beta, h)$: the height of the excluded region at $\beta > \beta_c$ is $[\bar{m}(\beta, 0)]_h$.

"correlation length" and therefore one obtains continuum equations replacing the lattice variables. This is connected with the ideas of "scaling" and the "renormalization group". On the other hand, at the level of rigorous mathematics, universality is not very well understood. In two dimensions universality is related to the [stochastic, Schramm-] Loewner evolution. Therefore, much recent progress has been made in understanding universality there.

In two lectures we will consider the mean-field Ising model, also known as the Curie-Weiss model. One sometimes thinks of this as an analogue of the Ising model on $\mathbb{Z}_N^d$, except that d goes to ∞. For the mean-field model, we can calculate $p(\beta, h, J)$, fairly explicitly. Then we will see that there are phase transitions, for that model, and we will be able to verify all these pictures in that context. We will also be able to calculate the critical exponent $\beta = 1/2$ for the mean-field model. So, if we write $\beta(d)$ for the critical exponent on $\mathbb{Z}^d$, as a function of d, one fully expects $\beta(d) \to 1/2$, as $d \to \infty$. (For certain models one has mathematical proofs of such facts, and in fact for some models, there is even a critical dimension d_0 such that for $d \geq d_0$, the critical exponents equal to the mean-field values. But this is a little beyond our present scope.)

3. THE GIBBS VARIATIONAL PRINCIPLE ON INFINITE LATTICES

There is a version of the Gibbs variational principle that works directly on the infinite lattice $\mathbb{Z}^d$. Let us attempt to explain this now, in a very brief way.

The configuration space $\Omega_{\mathbb{Z}^d} = \{+1, -1\}^{\mathbb{Z}^d}$ has the product topology, and is therefore compact. This topology can be understood very concretely. Let $\mathscr{P}_f$ be the set of all finite subsets $X \subset \mathbb{Z}^d$. Given any $X \in \mathscr{P}_f$, let $\mathscr{F}_X$ be the set of all cylinder sets in $\Omega_{\mathbb{Z}^d}$ which depend only on the spins in X. Let $\mathscr{C}_X$ be all the functions which are measurable with respect to $\mathscr{F}_X$. Then $\mathscr{C}$, the set of all continuous functions on $\Omega_{\mathbb{Z}^d}$, is

$$\mathscr{C} = \overline{\bigcup_{X \in \mathscr{P}_f} \mathscr{C}_X}\,,$$

where the closure is with respect to the sup-norm.

Let $\mathscr{F}$ be the Borel σ-algebra on $\Omega_{\mathbb{Z}^d}$. Let $M_1(\mathscr{F})$ denote the set of Borel probability measures on $\Omega_{\mathbb{Z}^d}$. There are some interesting subsets of $M_1(\mathscr{F})$. First of all, there are those measures which are "trivial at infinity". Specifically, let the "algebra at infinity" be

$$\mathscr{F}_\infty = \bigcap_{X \in \mathscr{P}_f} \mathscr{F}_{X^{\complement}}\,,$$

where $\mathscr{F}_{X^{\complement}}$ is the algebra of all events in $\mathscr{F}$ which only depend on spins *outside* of X. A measure $\mu \in M_1(\mathscr{F})$ is *trivial at infinity* if and only if for every event $E \in \mathscr{F}_\infty$ one has $\mu(E) \in \{0,1\}$. It satisfies a type of 0-1 law.

Also note that $\mathbb{Z}^d$ is a *lattice* in the group-theoretical sense. Let $M_1^I(\mathscr{F})$ denote the set of translation-invariant measures. For each $x \in \mathbb{Z}^d$, let $\tau_x : \Omega_{\mathbb{Z}^d} \to \Omega_{\mathbb{Z}^d}$ be the action of the translation by x on the spin configurations. Given $\mu \in M_1^{\mathcal{I}}(\mathscr{F})$, let $\mathscr{F}_0^\mu$ be the set of events $E \in \mathscr{F}$ such that $\mu\big(E \triangle \tau_x(E)\big) = 0$ for all $x \in \mathbb{Z}^d$ (where for any events E, F, $E \triangle F = (E \setminus F) \cup (F \setminus E)$ is the usual symmetric difference). These are the events E, which are μ-a.s. invariant, with respect to each τ_x.

Definition 3.1 *A measure $\mu \in M_1^{\mathcal{I}}(\mathscr{F})$ is called ergodic if and only if, for every event $E \in \mathscr{F}_0^\mu$, one actually has $\mu(E) \in \{0,1\}$.*

The set $M_1^{\mathcal{I}}(\mathscr{F})$ is obviously convex : the mixture of two translation invariant measures is translation invariant. (Note if $\mu, \nu \in M_1^{\mathcal{I}}(\mathscr{F})$, the mixture is simple. Given any $\theta \in [0,1]$, the measure $\theta\mu + (1-\theta)\nu$ describes the law for a random variable: toss a biased coin with probability θ of being "HEADS". If it comes up HEADS, use a random variable X, distributed by μ. If it comes up TAILS, use a random variable Y, distributed by ν.) It turns out that the extreme points of $M_1^{\mathcal{I}}(\mathscr{F})$ are the ergodic measures. We will not prove this fact now, but in a few lectures we will prove a completely analogous fact called de Finetti's theorem. We may state the *mutatis mutandis* necessary to get this theorem then.

The *extreme points*, which we will denote $M_1^{\mathcal{I},\mathcal{E}}(\mathscr{F})$, are the measures which cannot be written as any nontrivial mixture of any other measures in $M_1^{\mathcal{I}}(\mathscr{F})$. I.e., $\mu \in M_1^{\mathcal{I},\mathcal{E}}(\mathscr{F})$ if and only if: first, $\mu \in M_1^{\mathcal{I}}(\mathscr{F})$; and, second, for every $\theta \in (0,1)$, the only pair of measures $\nu_1, \nu 2 \in M_1^{\mathcal{I}}(\mathscr{F})$ for which $\mu = \theta\nu_1 + (1-\theta)\nu_2$ are $\nu_1 = \nu_2 = \mu$.

Even more compelling is the fact that every measure in $M_1^{\mathcal{I}}(\mathscr{F})$ is a unique mixture of ergodic measures. In other words $M_1^{\mathcal{I}}(\mathscr{F})$ is a "Choquet simplex".

Definition 3.2 (Specialized definition of Choquet simplex) *Suppose $K \subset M_1^{\mathcal{I}}(\mathscr{F})$ is a convex subset, which is closed relative to the weak (or equivalently vague) topology. Also suppose that the set of extreme points $K^{\mathcal{E}}$ is Borel measurable relative to that topology on $M_1^{\mathcal{I}}(\mathscr{F})$. Then K is a Choquet simplex if any only if, for each $\mu \in K$, there is a unique Borel probability measure ρ on $M_1^{\mathcal{I}}(\mathscr{F})$ (Borel with respect to the weak topology) which is concentrated on $K^{\mathcal{E}}$, $\rho(K^{\mathcal{E}}) = 1$, and such that for every continuous function f on $\Omega_{\mathbb{Z}^d}$, one* has

$$\int_{\Omega(\mathbb{Z}^d)} f\,d\mu \;=\; \int_{M_1^{\mathcal{I}}(\mathscr{F})} \left(\int_{\Omega(\mathbb{Z}^d)} f\,d\nu\right)\,d\rho(\nu)\,.$$

Note that the map $\nu \mapsto \int f\,d\nu$ is continuous in the weak topology, for any continuous $f : \Omega_{\mathbb{Z}^d} \to \mathbb{R}$, therefore it is certainly measurable. Thus the integral on the right hand side of the display makes sense. Also note, if ρ were not required to be concentrated on $K^{\mathcal{E}}$ it would be easy to satisfy the equation by taking ρ to be a point-mass at μ. If $\mu \in K^{\mathcal{E}}$ that is what ρ is, anyway. But if μ is not in $K^{\mathcal{E}}$ that is not the decomposition: good thing, because the decomposition has to be unique.

This is a funny definition if you have not thought much about measures-on-measures before. But there are two things to keep in mind. First, this is in some sense a stronger version of the conlcusion of the Krein-Milman theorem for closed convex sets. (Note, we have not *proved*, and will not prove in this lecture, that $M_1^{\mathcal{I}}(\mathscr{F})$ is a Choquet simplex.) But now one does not need to take limits – every measure $\mu \in K$ has a fixed decomposition, but in place of a finite "convex combination" of extreme points, there is a probability measure on the presumably infinite set $K^{\mathcal{E}}$ – and moreover that decomposition is *unique*. The second thing to keep in mind is that sometimes one does consider measures-on-measures, without realizing it. For example, random point processes, such as the Poisson point process, are measures-on-measures. A point configuration, on $\mathbb{R}$ for instance, is actually a locally-finite integer-valued measure (which forces it to be a countable sum of point-masses). But since one considers a random point process, the distributional law is a measure on the set of these. For most purposes it is important that if x_n is a sequence of points converging to x, then we think of the point-masses at x_n as converging to the point-mass at x. This is the reason we consider the weak topology. Because the underlying space $\Omega_{\mathbb{Z}^d}$ is compact, the topological space $M_1(\mathscr{F})$, with the weak topology is compact. So, in some senses, it is really a small set.

Let us also mention that there are finite-dimensional Choquet simplices, as well as infinite ones. In finite dimensions, these are exactly the usual simplices. I.e., the n-dimensional simples is an invertible affine transformation of the standard simplex:

$$\Delta^n = \{(a_1,\dots,a_{n+1})\ :\ a_1,\dots,a_{n+1} \geq 0\ \text{ and }\ a_1+\dots+a_{n+1}=1\}\,.$$

In order to state the Gibbs variational principle on $M_1^{\mathcal{I}}(\mathscr{F})$, we need to say how to take the relative entropy density of a measure $\mu \in M_1^{\mathcal{I}}(\mathscr{F})$. Given any measure $\mu \in M_1(\mathscr{F})$, and given any $X \in \mathscr{P}_f$, we can restrict μ to the σ-algebra $\mathscr{F}_X$. This is equivalent to a measure on the finite set Ω_X. Let us call that measure $\mu \restriction X$. The following theorem is an important result, which we won't prove.

Theorem 3.3 *For each n, let $\Lambda_n = [-n,n]^d \in \mathbb{Z}^d$. Given any $\mu \in M_1^I(\mathscr{F})$, the following limit exists:*

$$s(\mu) \;:=\; \lim_{n\to\infty} |\Lambda_n|^{-1} S_{\Lambda_n}(\mu \restriction \Lambda_n)\,,$$

although it may take the value $-\infty$. The function s is affine (the analogue of linearity for convex sets) and upper semicontinuous. It is also bounded above by 0, and it equals 0 if and only if μ is the product measure which assigns equal probability to $\{\sigma_x = +1\}$ and $\{\sigma_x = -1\}$ for all $x \in \mathbb{Z}^d$.

Excellent references for this theorem, as well as the Gibbs principle, to follow, are the monographs by Israel or Simon.

It remains to state what is the Gibbs variational principle for measures μ in $M_1^{\mathcal{I}}(\mathscr{F})$. The function s gives us a way to calculate the entropy density for any such measure. We need a way to calculate the energy density. Define

$$\mathrm{H}_{h,J}(\sigma) \,=\, -\frac{J}{2} \sum_{\substack{x \in \mathbb{Z}^d \\ \|x\|=1}} \sigma_x \sigma_0 - h\sigma_0\,.$$

Because of translation invariance, the energy density of the measure μ really is $\mathbf{E}^{\mu}[\mathrm{H}_{h,J}(\sigma)]$.

Also, we should have mentioned before now that $p(\beta, h, J)$ always does exist, although we previously left open the possibility thatit might fail to exist at some point (β, h, J). The proof of this important fact can be found in any of the monographs of Ruelle, Israel, or Simon (or many others). It relies only on approximate subadditivity of $p_{\Lambda_n}(\beta, h, J)$. (When we consider mean-field spin glasses, this will be one of the first tricky points to overcome, proving that the limiting pressure exists.)

Now, the Gibbs variational principle on $M_1^{\mathcal{I}}(\mathscr{F})$ states that

$$p(\beta, h, J) \,=\, \log(2) + \sup_{\mu \in M_1^{\mathcal{I}}(\mathscr{F})} \Big(s(\mu) - \mathbf{E}^{\mu}[\beta\, \mathrm{H}_{h,J}(\sigma)] \Big)\,.$$

Moreover, the supremum is attained. Compare this to Corollary 1.3.

Any measure attaining the supremum is called a translation-invariant equilibrium measure. The set of translation-invariant equilibrium measures is, itself, another Choquet-simplex. There is always at least one equilibrium measures, a result one can prove relatively easily in the present context, using compactness and the semicontinuity of the right hand side above. If there is more than a single translation-invariant equilibrium measure, then one says that there is a phase transition. More generally, there are other types of equilibrium measures than translation invariant ones, but we do not have to worry about that here. What we would like to say is that this notion of phase transition matches the previous one: i.e., if there are multiple equilibrium measures, then one can prove that some derivative of the pressure is discontinuous. (The mathematically fancy way of saying this is that there are multiple tangent functionals, since one knows that the pressure is always convex, and therefore supports such tangent planes/functionals.)

Let us make a final point. When looking for the maximum value, one can restrict to the set of extreme points. It seems reasonable to guess this result (but not entirely trivial to prove) because one is trying to optimize an affine function. At least in finite dimensions, affine functions on convex sets are always optimized at the extreme points. In infinite dimensions it is also true, but only because of the semicontinuity. So one can restrict to $\mu \in M_1^{\mathcal{I},\mathcal{E}}(\mathscr{F})$. One might, initially, hope that restricting thus simplifies the optimization problem. Unfortunately that is not true. For one thing, $M_1^{\mathcal{I},\mathcal{E}}(\mathscr{F})$ is actually dense in $M_1^{\mathcal{I}}(\mathscr{F})$. Refer to Israel or Simon for this truly bizarre fact. This is sometimes put as saying that $M_1^{\mathcal{I}}(\mathscr{F})$ is a Poulsen simplex. (Up to isomorphisms there is only one Poulsen simplex. E.g., there is obviously no finite Poulsen simplex.)

On the other hand, there are cases where that type of ambition can be realized. When we consider mean-field models, in a few lectures, the analogue of translation invariance will be full permutation invariance. For measures, this is called "exchangeability". Then the extreme points, instead of being ergodic, will actually be i.i.d., product measures. It is easy to imagine that for such simple measures calculating things like the relative entropy density does simplify drastically. As we will see that is the case, and that is why classical mean-field models are actually solvable.

4. Homework Exercises

1. For any finite lattice Λ, calculate $p_\Lambda(\beta; h; 0)$.

2. For $\Lambda = K_N$, derive a PDE satisfied by $p_\Lambda(\beta; h; J)$ involving $\frac{\partial}{\partial J}$, $\frac{\partial}{\partial h}$ and $\frac{\partial^2}{\partial h^2}$.

3. Check that $\psi_\Lambda(\beta; \mu_\beta) = p_\Lambda(\beta)$.

4. For any finite lattice Λ, calculate $\min_{\nu \in M_1(\Omega_\Lambda)} |\Lambda|^{-1} S_\Lambda(\nu|\mu_0)$, where μ_0 is the uniform probability measure on Ω_Λ.

5. Suppose that ν_1 and ν_2 in $M_1(\Omega_\Lambda)$ are mutually singular. (They are concentrated on disjoint sets.) Prove that, for any $\theta \in [0, 1]$,

$$S_\Lambda\big(\theta\nu_1 + (1-\theta)\nu_2 \big| \mu_0\big) \;=\; \theta S_\Lambda(\nu_1|\mu_0) + (1-\theta) S_\Lambda(\nu_2|\mu_0) - \theta\log(\theta) - (1-\theta)\log(1-\theta)\,.$$

6. *(a)* Give an example of an event in the tail algebra, $\mathscr{F}_\infty$. *(b)* Give an example of a translation-invariant measure $\mu \in M_1^I(\mathscr{F})$ which is ergodic, but not trivial at infinity. *(c)* Give an example of a measure which is translation invariant, but not ergodic.

References

[1] R.B. Israel, *Convexity in the Theory of Lattice Gases*, Princeton Series in Physics. With an introduction by Arthur S. Wightman. Princeton University Press, Princeton, N.J., 1979.

[2] R.A. Minlos, *Introduction to Mathematical Statistical Physics*, University Lecture Series, Volume 19. American Mathematical Society, Providence, R.I., 1999.

[3] D. Ruelle, *Statistical Mechanics: Rigorous Results*, Addison-Wesley Publishing Co., Inc., 1969. Reprinted by Imperial College Press and World Scientific Publishing Co. Pte. Ltd., London and Singapore, 1999.

[4] B. Simon, *The Statistical Mechanics of Lattice Gases*, Vol. I., Princeton Series in Physics, Princeton University Press, Princeton, N.J., 1993.

Mathematics Department, University of Rochester, Rochester, NY 14627

E-mail address: sstarr@math.rochester.edu