

# Part II

## Combinatorial Models

### 5 A Sketch of Basic Statistical Physics

Statistical physics = Thermodynamics (macroscopic)  
+ Statistical mechanics (microscopic)

#### 5.1 Thermodynamics

A *thermodynamic system* is characterised by (macroscopic, observable) variables  $T$  (“temperature”) and  $X_1, \dots, X_n$ . These variables determine “all interesting” properties of the system.

E.g. in the classical ideal gas model a sufficient set of variables is  $T, p, V$  and  $N$ . ( $N \sim$  the number of molecules is here for simplicity thought of as a continuous quantity. This might be easier if  $N$  was replaced by  $n = N/N_0$ , the amount in moles of gas, where  $N_0 = 6.02 \cdot 10^{23}$  is Avogadro’s number.)

The system is in (*thermal*) *equilibrium* if it satisfies a characteristic *state equation*

$$g(T, X_1, \dots, X_n) = 0$$

E.g. ideal gas:  $pV - NkT = 0$ , where  $k = 1.38 \cdot 10^{-23} \text{ J/K}$  is *Boltzmann’s constant*, or  $pV - nRT = 0$ , where  $R = 8.32 \text{ J/Kmol}$  is the *gas constant*.

A *potential* or *energy function* for the system is some sufficiently smooth function

$$F = F(T, X_1, \dots, X_n).$$

In classical thermodynamics, a key role is given to the *total energy* function determined by the *First Law of Thermodynamics*:

$$dU = dQ + dW, \quad (1)$$

where  $dQ$  is the amount of “heat” added to a system and  $dW$  is the amount of “work” performed on it.

Integrating the potential given e.g. the state equation of the ideal gas yields

$$U(T, p, N) = U_0 + \left( \frac{1}{2}fN + N - S_0 \right) (T - T_0) - NT \ln \left( \left( \frac{T}{T_0} \right)^{1+f/2} \frac{p_0}{p} \right),$$

where  $U_0, S_0, T_0$  and  $p_0$  are reference values and  $f/2$  a constant (“specific heat”).<sup>1</sup>

In classical thermodynamics, the system variables are divided into *extensive* and *intensive*, depending on whether their values depend on the “size” of the system or not. E.g.  $T$  and  $p$  are intensive,  $V$  and  $N$  extensive.

Two systems at the same temperature may be “combined”, and if  $F$  is otherwise a function of extensive variables only, then it is linear, i.e.

$$F(T, X_1 + X'_1, \dots, X_n + X'_n) = F(T, X_1, \dots, X_n) + F(T, X'_1, \dots, X'_n).$$

By the total derivative formula:

$$dF = \left( \frac{\partial F}{\partial T} \right) dT + \sum_{i=1}^n \left( \frac{\partial F}{\partial X_i} \right) dX_i. \quad (2)$$

State variables are *conjugate* (with respect to  $F$ ), if

$$X = \frac{\partial F}{\partial Y} \quad \text{or} \quad Y = \frac{\partial F}{\partial X}.$$

In classical thermodynamics conjugates of extensive variables are intensive, and vice versa. The conjugate of  $T$  w.r.t.  $U$ ,

$$S = \frac{\partial U}{\partial T}$$

is called the *entropy* of the system.

---

<sup>1</sup>To be precise, since  $T$  and  $p$  are not “natural” variables of the energy function  $U$  arising from its differential definition (1), this equation refers to a variant of  $U$  expressed in terms of  $T$ ,  $p$  and  $N$ , so called “Gibbs free energy”.

Conjugate variables may be interchanged via the *Legendre transform*, yielding new forms of a given potential function. E.g. in the case of the ideal gas with fixed  $N$ ,  $U = U(S, V)$  and

$$dU = TdS - pdV.$$

Here we may interchange  $S$  for  $T$  by considering instead of  $U$  the *Helmholz free energy*  $F = U - ST$ . This satisfies:<sup>2</sup>

$$dF = dU - SdT - TdS = TdS - pdV - SdT - TdS = -SdT - pdV.$$

For this potential function the “natural” variables are  $T$  and  $V$ , i.e.  $F = F(T, V)$ .

In the classical setting, it is a law of nature (the *Second Law of Thermodynamics*) that in equilibrium processes (evolutions) entropy never decreases:

$$dS \geq 0.$$

Processes for which entropy stays constant ( $dS = 0$ ) are called *adiabatic*.

## 5.2 Statistical Mechanics

Let us consider a thermodynamic energy function framed in terms of extensive variables:

$$U = U(S, X_1, \dots, X_n),$$

and assume that the value of  $U$  expresses in fact only the *average* of a large number of microscopic potentials:

$$U = \langle H \rangle = \sum_{\omega} p_{\omega} H(\omega).$$

The micropotential function  $H(\omega)$  is also called the *Hamiltonian* of the system. We shall furthermore assume, motivated by the additivity of  $U$ , that the Hamiltonian of a system consisting of two independent subsystems at thermal equilibrium can be decomposed as:

$$H(\langle \omega_1, \omega_2 \rangle) = H(\omega_1) + H(\omega_2).$$

What is now the distribution of the microstates  $p_{\omega}$ , given the constraint that  $\langle H \rangle = U$ ? We assume that all microstates with the same value of the Hamiltonian are equally probable, so that  $p_{\omega}$  has the form  $p_{\omega} = g(H(\omega))$ .

---

<sup>2</sup>There is an unfortunate sign difference here as compared to formula (2). We could have fixed this by defining  $F = ST - U$ , but this would have been against convention.

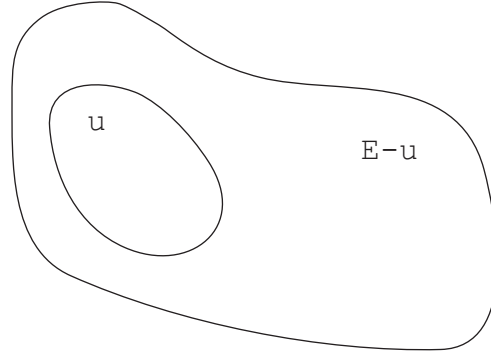


Figure 1: A heat bath.

To further specify the functional form of the distribution, think of our system  $\mathcal{S}$  as being in thermal equilibrium with, but otherwise independent of, a much larger system or “reservoir”  $\mathcal{R}$ . Denote the total system consisting of  $\mathcal{S}$  and  $\mathcal{R}$  by  $\mathcal{T}$  (This is called a “heat bath” arrangement; cf. Figure 1.)

For any given system, denote by  $\Omega(u) = |H^{-1}(u)|$  the number of its microstates at potential  $u$ . (Whether we are referring to  $\mathcal{S}$ ,  $\mathcal{R}$  or  $\mathcal{T}$  should always be clear from the context.) Fix some reference potential level  $E \gg U$  for the total system  $\mathcal{T}$ , and observe that by our assumption, all microstates of  $\mathcal{T}$  with potential  $E$  have the same probability.

Now for every microstate  $\omega$  of  $\mathcal{S}$ , there are exactly  $\Omega(E - H(\omega))$  microstates  $\omega'$  of  $\mathcal{R}$  such that the combined state  $\langle \omega, \omega' \rangle$  of  $\mathcal{T}$  has potential  $E$ . Since all of these are equally probable, it follows that  $p_\omega \propto \Omega(E - H(\omega))$ . Taking logarithms and applying Taylor’s formula yields:

$$\begin{aligned} \ln p_\omega &= \ln \Omega(E - H(\omega)) + \text{const.} \\ &= \ln \Omega(E) - \left( \frac{\partial \ln \Omega(E')}{\partial E'} \right)_{E'=E} H(\omega) + \dots \\ &= \ln \Omega(E) - \beta H(\omega) + \dots, \end{aligned}$$

where  $\beta = \partial \ln \Omega / \partial E$  is a parameter whose value is to be determined later.

Taking exponentials again, we obtain the so called *Gibbs* (or *Boltzmann*) *distribution*

$$p_\omega \propto e^{-\beta H(\omega)} \tag{3}$$

with normalisation constant (actually, function)

$$Z = Z_\beta = \sum_{\omega} e^{-\beta H(\omega)}, \tag{4}$$

known as the *partition function*.<sup>3</sup> Now the value of  $\beta$  is in principle determined implicitly by the condition

$$\langle H \rangle = \frac{1}{Z} \sum_{\omega} e^{-\beta H(\omega)} H(\omega) = U,$$

but we shall obtain a more transparent representation for it below.

The (logarithm of the) partition function (4) can be used to compute several macroscopic quantities:

First:

$$\begin{aligned} \frac{\partial \ln Z}{\partial \beta} &= \frac{1}{Z} \frac{\partial Z}{\partial \beta} \\ &= \frac{1}{Z} \frac{\partial}{\partial \beta} \sum_{\omega} e^{-\beta H(\omega)} \\ &= \frac{1}{Z} \sum_{\omega} e^{-\beta H(\omega)} (-H(\omega)) \\ &= - \sum_{\omega} p_{\omega} H(\omega) \\ &= -U. \end{aligned}$$

Second: Consider an extensive variable  $X_i$  and its conjugate  $\mu_i = \partial U / \partial X_i$ .

$$\begin{aligned} \frac{\partial \ln Z}{\partial X_i} &= \frac{1}{Z} \sum_{\omega} \frac{\partial}{\partial X_i} e^{-\beta H(\omega; X_i)} \\ &= \frac{1}{Z} \sum_{\omega} e^{-\beta H(\omega; X_i)} \left( -\beta \frac{\partial H(\omega; X_i)}{\partial X_i} \right) \\ &= -\beta \sum_{\omega} p_{\omega} \frac{\partial H(\omega; X_i)}{\partial X_i} \\ &= -\beta \left\langle \frac{\partial H(\omega; X_i)}{\partial X_i} \right\rangle \\ &= -\beta \mu_i. \end{aligned}$$

---

<sup>3</sup>In fact,  $Z = Z(\beta, X_1, \dots, X_n)$ . Note also that  $Z$  is a kind of a *generating function* for the sequence of values  $\Omega(u)$ , since  $Z(\beta) = \sum_u \Omega(u) \cdot (e^{-\beta})^u$ .

Third:

$$\begin{aligned}
 d \ln Z &= \frac{\partial \ln Z}{\partial \beta} d\beta + \sum_{i=1}^n \frac{\partial \ln Z}{\partial X_i} dX_i \\
 &= -U d\beta - \beta \sum_{i=1}^n \mu_i dX_i \\
 &= -d(\beta U) + \underbrace{\beta dU - \beta \sum_{i=1}^n \mu_i dX_i}_{\beta T dS}.
 \end{aligned}$$

$$\therefore T dS = \frac{1}{\beta} d(\ln Z + \beta U)$$

$$\therefore \frac{1}{\beta} = kT, \quad dS = kd(\ln Z + \beta U), \quad k = \frac{1}{\beta T} = \text{constant}$$

$$\therefore \frac{1}{\beta} = kT, \quad S = k \ln Z + \frac{U}{T} + \text{const.} \sim k \ln Z + \frac{U}{T}$$

$$\therefore \beta = \frac{1}{kT}, \quad -kT \ln Z \sim U - TS = F \quad (\text{Helmholz free energy})$$

Conversely, let us expand the entropy variable as a microscopic average:

$$\left. \begin{aligned}
 S &= k \ln Z + k\beta U \\
 &= k \ln Z + k \sum_{\omega} p_{\omega} \beta H(\omega) \\
 &= k \left( \ln Z - \sum_{\omega} p_{\omega} (\ln Z + \ln p_{\omega}) \right) \\
 &= -k \sum_{\omega} p_{\omega} \ln p_{\omega}.
 \end{aligned} \right| \begin{aligned}
 p_{\omega} &= \frac{1}{Z} e^{-\beta H(\omega)} \\
 \Rightarrow \beta H(\omega) &= -\ln(Z p_{\omega}) \\
 \sum_{\omega} p_{\omega} &= 1
 \end{aligned}$$

One more, simplified expression for entropy: partition the range of possible potential values into narrow bands (of width  $\Delta U$ , say), and denote the number of microstates falling in band  $r$  as

$$\Omega(U_r) = \left| \{ \omega : U_r \leq H(\omega) < U_r + \Delta U \} \right|$$

Then the partition function is approximately

$$Z \approx \sum_r \Omega(U_r) e^{-\beta U_r}$$

In fact, since the number of microstates in a typical system is huge, the microstate potentials are highly concentrated around the average  $U = \langle H \rangle$ , and so in fact

$$Z \approx \Omega(U)e^{-\beta U},$$

whence

$$S = \frac{1}{T}(-F + U) = k \ln Z + \frac{U}{T} \approx k \ln \Omega(U) \underbrace{-\beta k U + \frac{U}{T}}_{=0} \approx k \ln \Omega(U).$$

## 6 The Ising Model, Spin Glasses and Neural Networks

### 6.1 The Ising Model

The following model was introduced by Ernst Ising in 1925 to explain magnetism in materials.

At a microscopic level, Ising's model system consists of  $N$  sites arranged in a lattice, either 1-D, 2-D ( $N = L^2$ ), or maybe even 3-D. At each site  $i = 1, \dots, N$  is located a magnetic ion or *spin* pointing either *up* or *down* ( $S_i = \pm 1$ ). Neighbouring sites  $\langle ij \rangle$  are related by an *interaction coefficient*  $J_{ij}$ , which in Ising's model is uniformly either a positive  $J > 0$  ("ferromagnetic case") or a nonpositive  $J \leq 0$  ("antiferromagnetic case"). A system whose internal interactions are all weak ( $J_{ij} \approx 0$ ) is "paramagnetic". In addition, there may be an *external field*  $h$  influencing the orientation of each of the spins. (More generally, one could have separate fields  $h_i$  for each spin  $S_i$ .)

The Hamiltonian of spin state  $\sigma = \langle S_1, \dots, S_N \rangle$  is

$$H(\sigma) = -J \sum_{\langle ij \rangle} S_i S_j - h \sum S_i,$$

where the sum is taken over *nearest neighbour pairs*  $\langle ij \rangle$  and periodic boundary conditions are assumed for simplicity.

States  $\sigma$  yielding the global minimum value of  $H(\sigma)$  are called *ground states* of the system. For a ferromagnetic system, the ground state has either all  $S_i = +1$  if  $h > 0$ , or all  $S_i = -1$  if  $h < 0$ . If  $h = 0$ , these two states are both equally good.

As a very simple example, let us compute the partition function for a trivial Ising paramagnet with  $N$  spins and  $J = 0$ . Denote  $\Omega = \{+1, -1\}^N$ . Then:

$$\begin{aligned}
 Z_\beta &= \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)} \\
 &= \sum_{\sigma \in \Omega} \exp(\beta h \sum_i S_i) \\
 &= \sum_{S_1=\pm 1} \sum_{S_2=\pm 1} \cdots \sum_{S_N=\pm 1} e^{\beta h S_1} e^{\beta h S_2} \cdots e^{\beta h S_N} \\
 &= \left( \sum_{S=\pm 1} e^{\beta h S} \right)^N \\
 &= (2 \cosh(\beta h))^N
 \end{aligned}
 \quad \left| \quad \cosh x = \frac{e^x + e^{-x}}{2} \right.$$

Define the (*total*) magnetisation of state  $\sigma$  as

$$M(\sigma) = \sum_{i=1}^N S_i.$$

The corresponding thermodynamic average at given  $\beta$  is

$$\begin{aligned}
 \langle M \rangle &= \frac{1}{Z} \sum_{\sigma \in \Omega} M(\sigma) \exp(-\beta H(\sigma)) \\
 &= \frac{1}{Z} \sum_{\sigma \in \Omega} \underbrace{\left( \sum_i S_i \right) \exp(-\beta H(\sigma))}_{(\star)}.
 \end{aligned}$$

However now in fact  $(\star) = \frac{\partial Z}{\partial(\beta h)}$ , so fortuitously:

$$\begin{aligned}
 \langle M \rangle &= \frac{1}{Z} \frac{\partial Z}{\partial(\beta h)} = \frac{\partial \ln Z}{\partial(\beta h)} \\
 &= N \frac{\partial \ln(2 \cosh(\beta h))}{\partial(\beta h)} \\
 &= N \frac{2(\partial \cosh(\beta h) / \partial(\beta h))}{2 \cosh(\beta h)} \\
 &= N \frac{2 \sinh(\beta h)}{2 \cosh(\beta h)} \\
 &= N \tanh(\beta h).
 \end{aligned}$$



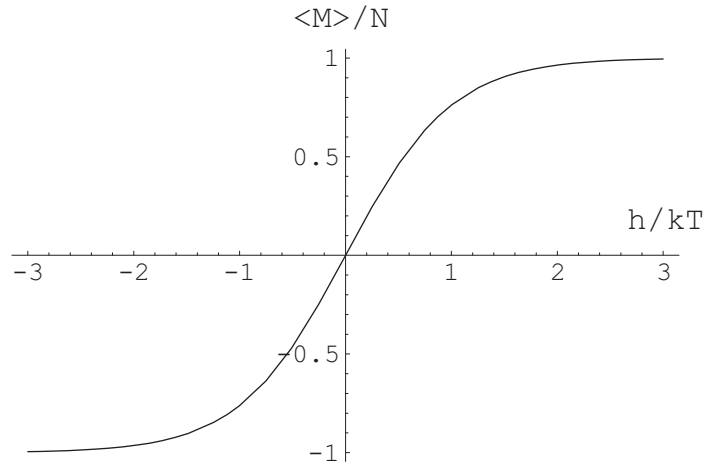


Figure 2: Magnetisation of an Ising paramagnet.

Thus the average magnetisation per site or “magnetisation density” of a totally decoupled Ising paramagnet at external field  $h$  and temperature  $T = 1/k\beta$  equals

$$\langle M \rangle = \tanh\left(\frac{h}{kT}\right).$$

A plot of this function is presented in Figure 2.

The ferromagnetic 1-D Ising model is also explicitly solvable with somewhat more work. The 2-D ferromagnetic case with  $h = 0$  was solved by L. Onsager in 1944, and in a simpler way by Kasteleyn & Fisher in 1961. The 2-D case with  $h \neq 0$  and higher dimensions are still open.

## 6.2 Spin Glasses

*Spin glasses* generalise the Ising model with more general interactions. Also the spins may be nonbinary, in which case such models are called *Potts glasses*.

The general form of the (binary-state) spin glass Hamiltonian is

$$H(\sigma) = - \sum_{\langle ij \rangle} J_{ij} S_i S_j - \sum_i h_i S_i,$$

where  $J_{ij}, h_i \in \mathbb{R}$ . Also the neighbourhood relation may correspond to an arbitrary *graph*, not necessary a lattice.

Several varieties of spin glass models have been introduced, e.g.:

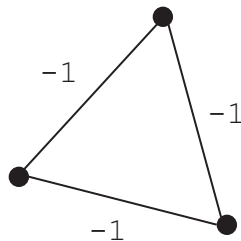


Figure 3: Frustrated spin glass configuration.

- The Sherrington-Kirkpatrick model: Hamiltonian as above, complete inter-connection graph, coefficients  $J_{ij}$  according to a specific probability distribution.
- The Edwards-Anderson model: Hamiltonian

$$H(\sigma) = - \sum_{\langle ij \rangle} J_{ij} S_i S_j,$$

regular lattice topology (e.g. cubic),  $J_{ij}$  independent Gaussian variables.

A phenomenon that makes spin glass models even less tractable than the Ising model is *frustration*. E.g. in the spin glass neighbourhood in Figure 3 there is no completely “consistent” choice of spin values.

Frustration means that the “landscape” determined by the Hamiltonian can have a very complicated structure, with large numbers of local minima, and no obvious location for the globally minimal ground state.

In fact, the problem of determining the ground state of a given SK-spin glass instance  $\langle \bar{J}, \bar{h} \rangle$  is *NP-complete*, even with  $\bar{h} = 0$ .

This can be seen by reduction from the well-known NP-complete MAX CUT problem: Given a graph  $G = (V, E)$ , determine the partition  $V = V_1 \cup V_2$  that maximizes  $w(V_1, V_2) = \left| \{(i, j) \in E : i \in V_1 \wedge j \in V_2\} \right|$ .

The reduction is as follows:

Given a graph  $G = (V, E)$ , let  $\bar{J}$  be an SK system with sites corresponding to  $V$ , and  $J_{ij}$  determined by

$$J_{ij} = \begin{cases} -1, & \text{if } \langle i, j \rangle \in E, \\ 0, & \text{otherwise.} \end{cases}$$

Let then  $C = (V_1, V_2)$  be a cut in  $G$ , and divide the edges in  $G$  corresponding as

$$\begin{aligned} E_1 &= \{\langle i, j \rangle \in E : i, j \in V_1\}, \\ E_2 &= \{\langle i, j \rangle \in E : i, j \in V_2\}, \\ E_C &= \{\langle i, j \rangle \in E : i \in V_1 \wedge j \in V_2\}. \end{aligned}$$

Consider the spin glass state  $\sigma$  determined as

$$S_i = \begin{cases} +1, & \text{if } i \in V_1, \\ -1, & \text{if } i \in V_2. \end{cases}$$

For this,

$$\begin{aligned} H(\sigma) &= - \sum_{\langle ij \rangle} J_{ij} S_i S_j = \sum_{\langle ij \rangle \in E} S_i S_j \\ &= \sum_{\langle ij \rangle \in E_1} S_i S_j + \sum_{\langle ij \rangle \in E_2} S_i S_j + \sum_{\langle ij \rangle \in E_C} S_i S_j \\ &= |E_1| + |E_2| - |E_C| \\ &= |E| - 2|E_C| \\ &= |E| - 2w(C). \end{aligned}$$

Conversely, given any spin glass state  $\sigma$ , one obtains a cut  $C$  satisfying  $w(C) = \frac{1}{2}|E| - \frac{1}{2}H(\sigma)$ .

Thus, graph cuts and spin glass states correspond one-to-one, with  $w(C) \propto -H(\sigma)$ , and minimizing one is equivalent to maximising the other.

The result means that the SK spin glass ground state problem is in a sense “universal” difficult problem, i.e. it contains as special cases all the  $\sim 2000$  other known NP-complete problems.

For  $J_{ij} > 0$  and arbitrary  $\bar{h}$  the problem reduces to network flow, and can be solved in polynomial time. For planar  $G$  and  $\bar{h} = 0$  the problem also has a polynomial time algorithm (Fisher 1966 (2-D lattices), Barahona 1982). However, for planar  $G$  with  $\bar{h} \neq 0$ , and for 3-D lattices the problem is NP-complete (Barahona 1982). It is also NP-complete for every other nonplanar crystal lattice graph (Istrail 2000). Thus, the dimensionality of the system is not crucial to the complexity of the ground state problem; the key is rather the planarity of the interconnection graph.

### 6.3 Neural Networks

John Hopfield proposed, in an influential paper in 1982, to use the SK model as a basis for “neural associative memories”. The idea is to create an  $N$ -site SK