1.3 Elements of statistical physics

1.3.1 *Combinatorial formulation*

A model in statistical physics is defined through a set of allowed configurations $\mathcal{V}$ and an energy function E over this set. We shall now explain these two notions.

1.3.1.1 *Configurations and energy* In Chapter 2, we have exemplified the scope of statistical physics on a gas, or liquid of molecules. In such a system, the degrees of freedom are the N particles locations; these are three-dimensional vectors with real-valued components subjected to some obvious constraints (particles cannot escape from the container).

For the sake of simplicity, we shall consider below the case of degrees of freedom taking values from a finite, discrete set. This is no limitation in most of the combinatorial optimization problems we shall encounter in the next chapters, whose nature is intrisically discrete.

Consider a set of N variables σ_j, $j = 1, \ldots, N$, each valued from a finite set $\mathcal{V}$ of cardinality q. An element $\mathcal{C} = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$ of $\mathcal{V}^N$ is called a configuration of the variables. The simplest non-trivial case is achieved for $q = 2$. Physicists usually consider $\mathcal{V} = \{-1, +1\}$, rather than the more natural $\mathcal{V} = \{0, 1\}$; variables

$$\sigma_i = \pm 1 \qquad (i = 1, \ldots, N) \tag{1.18}$$

are called Ising spins.

Each configuration $\mathcal{C}$ is associated a real-valued number $E(\mathcal{C})$ called energy. An ubiquitous form of energy function for configurations made of Ising spins is the following. Consider a graph G (Fig. 1.10) whose vertices are labelled by integer i, running from 1 to N. We denote by $\langle i, j \rangle$ the edges of G. An ising spin σ_i is attached to each node i of G. The energy function is simply a quadratic form over G defined through

$$E(\sigma_1, \ldots, \sigma_N) = -\sum_{\langle i,j \rangle \in G} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i \qquad . \tag{1.19}$$

Coefficients J_{ij} appearing in the first sum, carried over the edges, are called couplings, while parameters h_i involved in the second sum of (1.19) are usually referred to as magnetic fields, or simply fields. As a particular case, when all couplings and fields are uniform, *i.e.* $J_{ij} = J$, $h_i = h$ for all i, j, energy function E in eqn (1.19) reduces to,

$$E(\sigma_1, \ldots, \sigma_N) = -J \sum_{\langle i,j \rangle \in G} \sigma_i \sigma_j - h \sum_i \sigma_i \qquad , \tag{1.20}$$

the so-called Ising energy function or Ising Hamiltonian.

Physically motivated energy functions e.g. eqn (1.20) are designed to reflect the properties of realistic systems of spins. Models encountered in condensed matter certainly live in our three-dimensional world, or sometimes in some one- or two-dimensional subspaces. As a consequence graphs G considered in physics

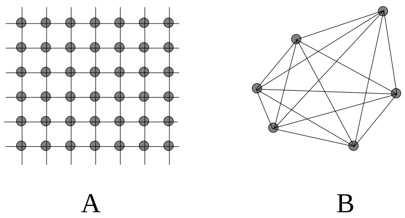

FIG. 1.10. Two examples of underlying graphs G for the Ising Hamiltonian (1.20). **A.** A two-dimensional square lattice; each node (grey dot) is connected to four adjacent vertices. **B.** Complete graph K_6: six vertices are connected two by two.

can always be embedded in a three dimensional space, a constraint that is not necessarily met in situations originating from mathematics or computer science (Fig. 1.10).

1.3.1.2 *Counting energy levels: notion of entropy* Consider the set of energy values, or levels, $\mathcal{E} = \{E(\mathcal{C}), \mathcal{C} \in \mathcal{V}^N\}$. The cardinality $\mathcal{N}$ of $\mathcal{E}$ is bounded from above by q^N. Elements of $\mathcal{E}$ can be sorted in increasing order, from the smallest, E_0, to the largest one, $E_{\mathcal{N}}$. We define the degeneracy, or multiplicity $D(E)$ of energy level $E \in \mathcal{E}$ as the number of configurations with energy equal to E,

$$D(E) = \sum_{\mathcal{C}\in\mathcal{V}} \delta_{E,E(\mathcal{C})} = \sum_{\ell=1}^{\mathcal{N}} \delta_{E,E_\ell} \quad , \tag{1.21}$$

where $\delta_{a,b}$ denotes the Kronecker function: $\delta_{a,b} = 1$ if $a = b$, 0 otherwise. The logarithm of the degeneracy,

$$S(E) = \ln D(E) \quad , \tag{1.22}$$

is referred to as the entropy, or microcanonical entropy, associated to the energy level E.

The generating function Z of energy levels is called partition function in statistical mechanics,

$$Z(x) = \sum_{E\in\mathcal{E}} D(E)\, x^{E} = \sum_{\mathcal{C}\in\mathcal{V}} x^{\,E(\mathcal{C})} \quad . \tag{1.23}$$

where x is some formal variable the physical meaning of which will be unveiled in next section.

1.3.2 *Probabilistic formulation.*

We have so far reduced statistical mechanics to the counting problem of energy levels. If this point of view permits to establish a direct connection with combinatorics which will turn to be very useful for studying optimization problems, statistical physics is founded, at the origin, on a probabilistic description of physical systems.

The implementation of this idea has required the introduction of revolutionary concepts at the end of the ninteenth century by Boltzmann and followers, and in particular, the ideas of ergodicity and thermodynamical equilibrium. We shall not attempt here to provide an exposition of these concepts. The interested reader can consult excellent textbooks (Reif 1965, Huang 1967, Ma 1985).

The main postulate is the existence of a probability measure, called Gibbs measure, over the space $\mathcal{V}^N$ of configurations, built from the energy function introduced in 1.3.1.1. The probability $p_T(\mathcal{C})$ associated to configuration $\mathcal{C}$ depends in addition on some parameter T, called temperature,

$$p_T(\mathcal{C}) = \frac{1}{Z(T)} \exp\left(-\frac{1}{T} E(\mathcal{C})\right) \quad , \tag{1.24}$$

where

$$Z(T) = \sum_{\mathcal{C}\in\mathcal{V}} \exp\left(-\frac{1}{T} E(\mathcal{C})\right) \tag{1.25}$$

ensures the normalization of the probability (1.24).

A configuration is all the more probable than its energy is low. To acquire some intuition on the dependence of the probability upon temperature, consider two limiting cases of (1.24):

- *infinite temperature* $T = \infty$*:* the probability $p_\infty(\mathcal{C})$ becomes independent of $\mathcal{C}$. All configurations are thus equiprobable.
- *zero temperature* $T = 0$*:* the probability $p_0(\mathcal{C})$ is concentrated on the minimum of the energy function E, called the ground state. If there are more than one ground state, that is, more than one configuration realizing the global minimum, all these ground state configurations are equiprobable.

When varying the temperature, the probability p_T interpolates between these two extreme cases.

In this probabilistic formulation, the energy is now a stochastic variable. We use the symbol $\langle\cdot\rangle$ to denote the average over the Gibbs measure p_T, and write the expectation value of the energy,

$$\langle E\rangle = \sum_{\mathcal{C}\in\mathcal{V}} p_T(\mathcal{C})\, E(\mathcal{C}) = T^2 \frac{d}{dT} \ln Z(T) \ , \tag{1.26}$$

from the normalization coefficient $Z(T)$ (1.25). Identity (1.26) can be extended to the higher moments of the energy. For instance, the variance of E can be computed from the second derivative of $Z(T)$,

$$\langle E^2\rangle - \langle E\rangle^2 = \left(T^2 \frac{d}{dT}\right)^2 \ln Z(T) \quad . \tag{1.27}$$

Not only the energy, but any observable, *i.e.* quantity depending upon configurations are stochastic variables whose moments made be computed from knowledge of the Gibbs measure. An example of such an observable is, for Ising spins variables and an energy function of the type (1.19), the magnetization of the configuration,

$$M(\mathcal{C}) = \sum_{i=1}^{N} \sigma_i \qquad . \tag{1.28}$$

The free energy of a model is simply defined through its partition function,

$$F(T) = -T \ln Z(T) \qquad . \tag{1.29}$$

As seen above, the free energy F gives access to all the moments of the energy. Another reason for which F is of central importance is that it contains a lot of information related to the degeneracy of the energy levels $D(E)$ (1.21) as shown in section 1.3.3.

The normalization coefficient of the Gibbs measure (1.25) coincides with the partition function (1.23) when its argument x is related to the temperature T through

$$x = e^{-1/T} \qquad . \tag{1.30}$$

We shall indifferently use the terminology partition function to refer to $Z(T)$ (1.25) or $Z(x)$ (1.23) provided T, x fulfill (1.30). Quantities defined in the probabilistic framework e.g. the average energy $\langle E \rangle$ have therefore combinatorial counterparts e.g. $dZ/d(\ln x)$. Using one or the other description is merely a matter of convenience. We shall sometimes consider in the next chapters that configurations of variables related to various optimization problems are distributed according to a Gibbs measure with a fictitious temperature T, whose value will be eventually sent to zero to reach the optimal assignment.[22]

Notice that within the probabilistic approach, the temperature T ranges from zero to infinity, and the argument x of the partition function from zero to unity from eqn (1.30). No bounds on x are to be imposed from a purely combinatorial point of view[23].

1.3.3 *Large sizes: concentration and deviations.*

Phase transitions, that is, abrupt changes of the average values of observables when a control parameter e.g. temperature is tuned, are not encountered when the set of configurations is finite. A necessary condition for the existence of a

[22] This point of view has proven to be useful in the analytical study of model optimization problems, and fructuous from a practical standpoint too. It is indeed the starting point for the simulated annealing search method exposed in Chapter 11 (Kirkpatrick, Gelatt, Vecchi 1983).

[23] Negative, or complex values of the temperature have indeed been considered by statistical physicists in a series of studies which underline the combinatorial significance of the partition function. Fisher (1965) studied the location of the zeroes of $Z(T)$ for the Ising model in the complex plane of the temperature, following the pioneer work of Lee and Yang (1952), and Yang and Lee (1952). See Itzykson, Drouffe (1989) for a review.

transition in a system is indeed that the size of the latter goes to infinity. The mathematical reason is simple: if the number of terms in the sum (1.23) is finite, the partition function Z, the free-energy F, the average energy, ... are analytic functions of the temperature T and so do not have singularities over the range $0 < T < \infty$. Analytical studies are therefore often devoted to the understanding of the emergence of singularities in the free-energy when the size of the system goes to infinity, the so-called thermodynamic limit.

1.3.3.1 *Concentration of intensive observables.* Most observables we have considered so far e.g. free energy, average energy, average magnetization, ... are said to be extensive: their values depend upon the system size N and, more precisely, are proportional to N for large sizes. It is therefore natural to define their intensive counterparts as the observables divided by the size. It is a widespread convention to use lowercase letters for intensive observables to distinguish them from extensive observable (uppercase letters). For instance, the intensive magnetization function, also called density of magnetization or magnetization per spin, reads

$$m(\mathcal{C}) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \qquad . \tag{1.31}$$

Not surprisingly, the probability distribution of intensive observables are much less spread than the ones of extensives observables. Think for instance of the sum Q of N independently distributed normal random variables w_i, with mean w and variance v. Whereas Q is a random variables which may take any real value with a non zero probability density, $q = Q/N$ converges in law to a highly concentrated distribution in $q = w$, with $O(1/\sqrt{N})$ relative fluctuations.

A similar concentration phenomenon is expected to take place in the thermodynamic limit for intensive observables. For instance, the moments of the intensive energy $e = E/N$ usually scale as

$$\begin{aligned} \langle e \rangle &= O(1) \\ \langle e^2 \rangle - \langle e \rangle^2 &= O(1/N) \qquad , \end{aligned} \tag{1.32}$$

when $N \to \infty$. The extensive energy of a configuration is thus, with high probability, equal to the average value $\langle E \rangle$ up to $O(\sqrt{N})$ fluctuations.

To justify eqn (1.32), we rewrite partition function (1.23) and (1.25),

$$Z(T) = \sum_{\ell=1}^{\mathcal{N}} e^{N(s(e_\ell) - e_\ell/T)} = \int_{-\infty}^{\infty} de\ \mu(e)\ e^{N(s(e)-e/T)} \tag{1.33}$$

where the intensive entropy $s(e) = S(N\,e)/N$ is defined from eqn (1.22). The probability density of allowed, distinct energy level reads,

$$\mu(e) = \frac{1}{\mathcal{N}} \sum_{\ell=1}^{\mathcal{N}} \delta(e - e_\ell) \qquad , \tag{1.34}$$

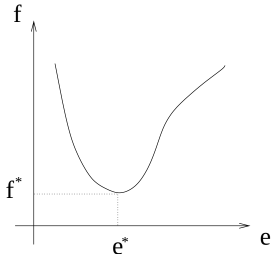

FIG. 1.11. A well-behaved free energy functional $f(e)$. f has a single global minimum with positive curvature. This situation takes place at high temperature.

where δ denotes the Dirac distribution. If $\mu(e)$ converges for large sizes N towards some smooth measure over some well defined support $e_0 < e < e_\infty$, integral (1.33) may be evaluated using the Laplace, or saddle-point method. Let us call

$$f(e) = e - T\, s(e) \qquad , \tag{1.35}$$

the free energy functional associated to the density of entropy s, and f^* its global minimum reached in e^* (Fig. 1.11). Both f^* and e^* depend upon temperature. The intensive free energy f then coincides with f^* from eqn (1.29) and (1.33). As for e^*, it is equal to the most probable value of the energy density. The expansion of the free energy functional f (1.35) around its minimum, $f(e) = f^* + (e-e^*)^2\chi^*/2 + 0\big((e-e^*)^3\big)$ with $\chi^* > 0$, indeed shows that the probability density of the intensive energy to be equal to e scales as, from (1.24),

$$p_T(e) = \mu(e)\,\frac{e^{-Nf(e)/T}}{Z(T)} \simeq e^{-N(e-e^*)^2/(2\chi^*)} \qquad , \tag{1.36}$$

up to non exponential in N contributions. Eqn (1.36) shows that the density of energy cannot deviate with high probability from its most probable value by more than $O(1/\sqrt{N})$. The inverse χ^* of the curvature of f in its minimum is usually called susceptibility.

1.3.3.2 *Phase transitions.* The free energy functional (1.35) has a simple significance. It reflects the competition mechanism between two opposed effects in the Gibbs measure (1.24). One configuration is all the more probable that its energy is low. But the overall probability that the system will be at energy density e is obviously proportional to the number of configurations with this energy. Well-defined physical systems have generally few configurations with low energies and much more numerous configurations with high energies[24]; their entropy $s(e)$

[24]We need only this statement to be correct close, and slightly above the ground state energy. Models considered here, without kinetic energy, may have a decreasing microcanonical entropy e.g. the Ising model on the complete graph studied in this Chapter. The non monotonous character of the entropy would however be seen for arguments of the partition function $x > 1$ only, *i.e.* for negative temperatures.

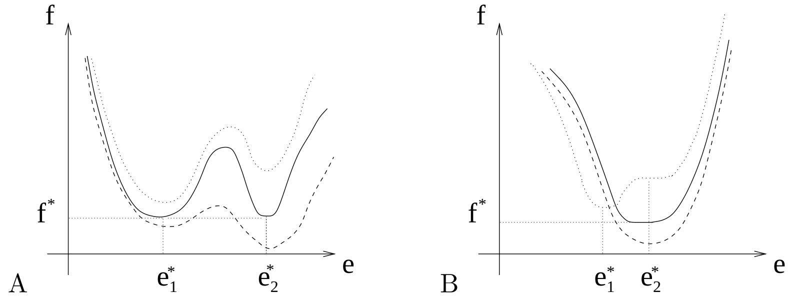

FIG. 1.12. Two examples of free energy functionals $f(e)$ exhibiting phase transitions. Dotted curves: temperature below criticality; full curves: critical temperature; dashed curve: temperature above criticality. **A**. Case of two far-away minima. At the critical temperature, the two minima are at the same heights, and the energy discontinuously jumps from e_2^* to e_1^*. The transition is of the first order. **B**. Case of one global minimum becoming unstable, and giving birth to another global minimum at the critical temperature. The energy is continuous at the transition. The transition is of the second order.

is a monotonously increasing function of the energy, which makes the minimum of the r.h.s. of eqn (1.35) take place at some non trivial value $e^*(T)$. In other words, while the energy term in (1.35) tends to push $e^*(T)$ down to the ground state value, the entropy contribution favorishes larger values. The trade-off is controlled by the value of the temperature.

The minimization procedure of the free energy functional can be unambiguously carried out as long as $f(e)$ is a convex function of the type displayed in Fig. 1.11. This is generally the case at sufficiently high temperatures from the above discussion, and $e^*(T)$ is generally an analytic function of the temperature T. This needs not to be true at low temperatures, where the minimum of f may exhibit discontinuities or singularities with T, a signal of the occurence of phase transitions.

An example of a discontinuous transition is proposed in Fig. 1.12A. The graph of f shows to local minima located at $e_1^*(T)$ and $e_2^*(T)$, with respective heights $f_1^*(T)$ and $f_2^*(T)$. At relatively high temperature, $f_2^*(T) < f_1^*(T)$, and the system has density energy equal to $e_2^*(T)$ with high probability. At some temperature T_c, called critical temperature, both free energies are equal: $f_1^*(T) = f_2^*(T) \equiv f^*$. When decreasing the temperature below T_c, the global minimum of f is not located in e_2^* any longer, but in e_1^*. When crossing T_c from above, the intensive energy therefore jumps down from e_2^* to e_1^*. The energy is a discontinuous function of the temperature. So is the associated entropy s^* from eqn (1.35) to keep the free energy continuous at the transition,

$$f^* = e_1^* - T_c\, s_1^* = e_2^* - T_c\, s_2^* \qquad (T = T_c)\ . \tag{1.37}$$

Such a transition, called first order phase transition, takes often place in Nature, a celebrated example being the freezing of liquid water into ice (Reif 1965).

Another possibility for a transition to arise is through an unstability of the minimum, that is, the vanishing of its curvature. This is shown in Fig. 1.12B. At high temperature, the free energy functional f exhibits a minimum at e_2^*, which becomes a saddle-point at some critical temperature T_c. As T decreases from T_c, the true global minimum, reached in e_1^*, gets away from e_2^* and migrates to the left. Both the energy and entropy (and, obviously, the free energy) are continuous functions of the temperature at criticality, though their derivatives may be not. Such a transition is called second order phase transition. It describes for instance the paramagnetic to ferromagnetic ordering in real magnets (Kittel 1985).

1.3.3.3 *Large deviations.* The concept of large deviation is intrisically connected to the one of concentration. As explained above, we expect physical quantities e.g. the energy density to be concentrated around their average values when the size N goes to infinity. At finite but huge sizes, fluctuations can be observed. Small fluctuations of the order of $N^{-1/2}$ happen with finite probability[25], but much larger fluctuations of the order of unity can be oberved too, though with very small probability. How small is this probability is expressed by the rate function.

What a rate function is can be understood from the following and simple example. Suppose you are given an unbiased coin, toss it N times, and measure the fraction h of these draws that have resulted into head. Being the coin unbiased, head and tail have the same probability 1/2 to be obtained after a draw. Then, we guess, and the central limit theorem tells us that, after a large number N of draws, the fraction h is very likely to be close to $\bar{h} = 50\%$. Actually, fluctuations around h_0 are of the order of $1/\sqrt{N}$, and vanish in the large N limit. Yet, this result is true on average only, and does not tell us anything about the probability that the measured fraction will be, say, $h = 78\%$, except that this probability with be vanishingly small as N increases. But how small? Each sequence of N draws {head,tail, head,head,..., tail} has probability $1/2^N$ to occur. The probability that the fraction of heads equals h after N draws is therefore,

$$\text{Prob}[h, N] = \frac{1}{2^N}\binom{N}{N\,h} \quad , \tag{1.38}$$

by definition of the above combinatorial coefficient. Use of the Stirling formula allows us to obtain the following asymptotic expression for p,

$$\text{Prob}[h, N] = 2^{N\,\omega(h)+o(N)} \quad , \tag{1.39}$$

where

$$\omega(h) = -1 - h\,\log_2 h - (1-h)\,\log_2(1-h) \quad . \tag{1.40}$$

[25] The order of magnitude of fluctuations can be larger at transition where the susceptibility χ^* in (1.36) diverges, see Section 3.1.3.

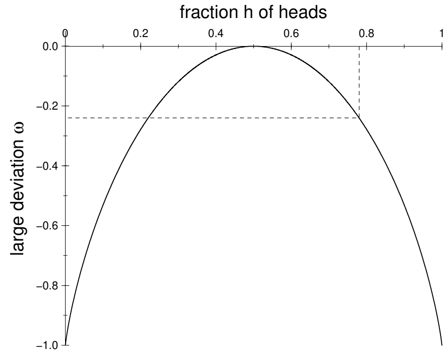

FIG. 1.13. Rate function $\omega(h)$ for the coin problem, giving the probability $2^{N\omega(h)+o(N)}$ that a fraction h of N draws are heads. The maximum, with height $\omega = 0$, corresponds to the most likely value, $h = 1/2$. The dashed line indicates how this curve can be used to read the probability that a fraction $h = 78\%$ of N draws are heads; this probability scales as $2^{-0.240\times N+o(N)}$ for large N.

Function $\omega(h)$ is plotted in Fig. 1.13. Being the logarithm of a probability (divided by N), ω is negative. It vanishes in $h = \bar{h}(= 1/2)$, the only value of h with a non exponentially–in–N small probability. At the boundaries of h–range, $h = 0$ and $h = 1$, ω reaches the value -1 as expected: there is a unique sequence of draws realising $h = 0$ (or $h = 1$), and this sequence has probability $1/2^N$.

Actually the notion rate function was already implicit in the discussion of the concentration phenomenon in the thermodynamic limit. The Gibbs probability for the energy density e reads

$$p_T(e) = e^{N\,\omega(e)+o(N)} \tag{1.41}$$

where the rate function is, see (1.36),

$$\omega(e) = -\frac{1}{T}\big(f(e) - f(e^*)\big) \quad . \tag{1.42}$$

This formula gives a probabilistic interpretation for the free energy $f(e)$ for all values of its argument e and not only the minimum e^*.

1.3.4 *Dynamical formulation.*

The third and alternative formulation of statistical mechanics is of dynamical nature. After all, physical systems are certainly not quenched forever and evolve with time. It is therefore sensible to assume that an adequate description of their properties should take into account this dynamical evolution. A possible dynamics on the configuration set commonly considered is the following. Let $\mathcal{C}$ and $\mathcal{C}'$ be two different configurations. We define

$W[\mathcal{C}', \mathcal{C}]$ = Probability of transition from $\mathcal{C}$ to $\mathcal{C}'$ per unit of time.[26]

Beware of the matrix convention for W: the final and initial configurations are, respectively, rank (left) and column (right) indices. The evolution is then modeled in terms of a dynamical change of the probabilities over configurations, $P(\mathcal{C}, t)$, which explicitely depend on time t. These probabilities obey the first-order differential equation

$$\frac{d}{dt}\, P(\mathcal{C}', t) = \sum_{\mathcal{C}} W[\mathcal{C}', \mathcal{C}]\ P(\mathcal{C}, t)\ , \quad \forall\, \mathcal{C}' \tag{1.43}$$

called master equation. Notice that the above sum over configurations includes the diagonal term $W[\mathcal{C}, \mathcal{C}]$, the value of which has not been specified yet and should be chosen to ensure the normalisation of probability,

$$\sum_{\mathcal{C}} P(\mathcal{C}, t) = 1 \tag{1.44}$$

holding at all times. Summing master equation (1.43) over configurations $\mathcal{C}'$, we see that normalisation (1.44) is fulfilled with the choice

$$W[\mathcal{C}, \mathcal{C}] = -\sum_{\mathcal{C}'(\neq \mathcal{C})} W[\mathcal{C}', \mathcal{C}] \quad . \tag{1.45}$$

The above definition for the diagonal rates simply expresses that $W[\mathcal{C}, \mathcal{C}]$ is minus the probability of escaping from configuration $\mathcal{C}$ per unit of time.

What are sensible choices for the transition rates W? First, each off-diagonal entry of W expresses a probability per unit of time, and thus should be a positive (or null) real-valued number. Secondly, and most importantly, the dynamics should allow us to find back the probabilistic statistical mechanics description of Section 1.3.2 whatever initial condition we consider for the probability over configurations. In other words, the solution of master equation (1.43) should converge to the Gibbs measure $p_T(\mathcal{C})$ (1.24) at long times.

A practical way to achieve this objective is to impose what is called 'detailed balance' to W. Asking that the Gibbs measure is a stationary distribution over the configurations gives, from eqn (1.43),

$$0 = \sum_{\mathcal{C}} W[\mathcal{C}', \mathcal{C}]\ p_T(\mathcal{C}) = \sum_{\mathcal{C}(\neq \mathcal{C}')} \Big\{ W[\mathcal{C}', \mathcal{C}]\ p_T(\mathcal{C}) - W[\mathcal{C}, \mathcal{C}']\ p_T(\mathcal{C}') \Big\} \tag{1.46}$$

which must hold for any $\mathcal{C}'$. This stationarity condition expresses a global balance between the in- and out-flows of probability (per unit of time and within the Gibbs measure) for configuration $\mathcal{C}'$. Detailed balance is a stronger condition

[26] In other words, the probability that the configuration $\mathcal{C}$ is changed into $\mathcal{C}'$ during the time interval δt equals $W[\mathcal{C}', \mathcal{C}] \times \delta t$ in the limit of very small intervals $\delta t \to 0$.

requiring that, for any pair of configurations $\mathcal{C}, \mathcal{C}'$, the probabilities (per unit of time and within the Gibbs measure) of going from $\mathcal{C} \to \mathcal{C}'$ and from $\mathcal{C}' \to \mathcal{C}$ are equal; it is mathematically expressed as

$$W[\mathcal{C}', \mathcal{C}]\, p_T(\mathcal{C}) = W[\mathcal{C}, \mathcal{C}']\, p_T(\mathcal{C}') \quad . \tag{1.47}$$

Detailed balance is therefore a sufficient (but restrictive) condition for stationarity amounting to impose that the term in the brackets of eqn (1.46) vanishes for any $\mathcal{C}$, Furthermore, as announced above, identity (1.47) ensures that the Gibbs measure is not only stationary but also the long time limit of the solution of the master equation[27]. We will come back to what 'long' means in practice in Section 3.1.

A further consequence of detailed balance is that the ratio of the transition rates from $\mathcal{C} \to \mathcal{C}'$ and back equals[28]

$$\frac{W[\mathcal{C}', \mathcal{C}]}{W[\mathcal{C}, \mathcal{C}']} = \frac{p_T(\mathcal{C}')}{p_T(\mathcal{C})} = e^{-(E(\mathcal{C}')-E(\mathcal{C}))/T} \quad , \tag{1.52}$$

and is thus a simple function of the difference of the energies of the two configurations. Notice that only the ratio of the forward and backward rates is imposed

[27]The proof can be sketched as follows. Define the modified transition matrix

$$\hat{W}[\mathcal{C}', \mathcal{C}] = \big(p_T(\mathcal{C}')\big)^{-\frac{1}{2}}\, W[\mathcal{C}', \mathcal{C}]\, \big(p_T(\mathcal{C})\big)^{\frac{1}{2}} \tag{1.48}$$

Detailed balance implies that $\hat{W}$ is a symmetric, and thus diagonalisable matrix. In addition, using eqn (1.45), we see that the quadratic form

$$Q \equiv \sum_{\mathcal{C}, \mathcal{C}'} \hat{W}[\mathcal{C}', \mathcal{C}]\, \hat{X}[\mathcal{C}']\, \hat{X}[\mathcal{C}] = -\frac{1}{2} \sum_{\mathcal{C} \neq \mathcal{C}'} W[\mathcal{C}', \mathcal{C}] \left[\hat{X}[\mathcal{C}'] - \left(\frac{p_T(\mathcal{C}')}{p_T(\mathcal{C})} \right)^{\frac{1}{2}} \hat{X}[\mathcal{C}] \right]^2 \tag{1.49}$$

is negative since all off-diagonal entries of W are positive (or null). The eigenvalues of $\hat{W}$ are negative; the zero eigenvalue is associated to the eigenvector $\hat{X}_0[\mathcal{C}] \propto \big(p_T(\mathcal{C})\big)^{\frac{1}{2}}$ and is not degenerate (provided the set of configurations is connected through non zero transition rates). From eqn (1.48) and (1.43), the modified time-dependent probability

$$\hat{P}(\mathcal{C}, t) = \big(p_T(\mathcal{C})\big)^{-\frac{1}{2}}\, P(\mathcal{C}, t) \tag{1.50}$$

obeys master equation

$$\frac{d}{dt}\, \hat{P}(\mathcal{C}', t) = \sum_{\mathcal{C}} \hat{W}[\mathcal{C}', \mathcal{C}]\, \hat{P}(\mathcal{C}, t) \tag{1.51}$$

and thus approaches $\hat{X}_0(\mathcal{C})$ as time increases. Going back to $P(\mathcal{C}, t)$ using (1.50), we conclude that the probability converges to $\hat{X}_0(\mathcal{C}) \times \big(p_T(\mathcal{C})\big)^{\frac{1}{2}} \propto p_T(\mathcal{C})$. The normalisation of probability (holding at any time t) ensures that the proportionality constant is unity.

[28]It is assumed here that the configuration $\mathcal{C}$ has a positive Gibbs probability and can be reached from the other configuration $\mathcal{C}'$. For configurations $\mathcal{C}, \mathcal{C}'$ that cannot be reached from one another, eqn (1.43) is trivially true with both left and right hand sides equal to zero. An illustration is given by the transition matrix (1.53) for the Ising model with configurations differing by more than one spin.

by detailed balance; the rates themselves can be chosen with some freedom depending on the physical situation to be modeled.

To illustrate the above concepts consider the Ising model with energy function (1.20). We denote by $\mathcal{C}^j$ the configuration of spins obtained from $\mathcal{C} = \{\sigma_i, i = 1, \ldots, N\}$ by flipping spin j, $\mathcal{C}^j = \{\sigma_1, \sigma_2, \ldots, -\sigma_j, \ldots, \sigma_N\}$. The dynamics can be described at follows. Let δt be a small time interval; we consider discrete-valued times $t = n\,\delta t$ with integer n. Initially ($n = 0$), a configuration of the spins $\mathcal{C} = \{\sigma_i, i = 1, \ldots, N\}$ is randomly chosen. At each step $n(\geq 1)$, a spin is chosen, uniformly and at random, say, j, and the difference $\Delta E = E(\mathcal{C}^j) - E(\mathcal{C})$ of energies between the current configuration and the one with spin j flipped is calculated. Then, spin j is flipped with probability $w(\Delta E) \times \delta t$; n is increased by one and the process resumes; possible function w will be proposed below. In the limit of vanishingly small time intervals $\delta t \to 0$, the dynamics is described by eqn (1.43), with transition rates

$$W[\mathcal{C}', \mathcal{C}] = \begin{cases} w(E(\mathcal{C}^j) - E(\mathcal{C})) & \text{if} \quad \mathcal{C}' = \mathcal{C}^j \\ -\sum_{j=1}^{N} w(E(\mathcal{C}^j) - E(\mathcal{C})) & \text{if} \quad \mathcal{C}' = \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{1.53}$$

Detailed balance is fulfilled provided the rate w satisfies (1.47), that is,

$$\frac{w(\Delta E)}{w(-\Delta E)} = e^{-\Delta E/T} \quad . \tag{1.54}$$

Among the (infinite set of) possible rate functions, three popular choices are

$$\text{Metropolis rate :} \qquad w(\Delta E) = \min\left(1, e^{-\Delta E/T}\right) \tag{1.55}$$

$$\text{Glauber rate :} \qquad w(\Delta E) = \frac{1}{1 + e^{\Delta E/T}} \tag{1.56}$$

$$\text{Exponential rate :} \qquad w(\Delta E) = e^{-\Delta E/(2T)} \tag{1.57}$$

These transition rates have in common to be decreasing functions of the energy change, which essentially vanish for large (with respect to T) positive arguments. The case of $N = 2$ spins is illustrated in Exercise ...

1.4 Exercises

1. Consider the following algorithm for solving 1-SAT instances involving N variables and M clauses:
   (a) Define two Boolean arrays, $Presence(i)$ and $Value(i)$, where i is an index running from 1 to N. All the entries of $Presence$ are initialized to F (the elements of $Value$ need not be specified). Set clause counter c to 1.
   (b) Loop over the clause index c. Call y the literal in clause number c, i the index of the associated variable, v a Boolean variable equal to T if $y = x_i$, to F if $y = \overline{x_i}$;