Contents

Chapter 1

Introduction

1.1 Historical Background

Statistical Physics deals with the collective behaviour of many interacting particles. The systems which have been more widely studied are *homogeneous systems*: systems where the particles are identical. However there are a lot of systems where particles are different one from the other, for example because they interact differently with their neighboring particles. In order to understand this kind of systems, a lot of efforts have been concentrated on the study of disordered systems like spin glasses, structural glasses and polymer networks. The interest towards these models started to grow in the seventies.

The physics of these systems resulted to be very rich and, moreover, spin glasses showed to be very useful to describe a huge variety of systems (both from physics and other disciplines). Among the systems that can be described and analyzed with the language of disordered systems, there are the problems of combinatorial optimization.

It has been observed that some algorithms used to solve problems of combinatorial optimization display an easy-hard transition. The presence of such a transition suggested to study it with the typical tools of phase transitions in statistical mechanics.

There has been an upsurge of interest in studying combinatorial optimization problems from a statistical physics point of view, both for theoretical analysis and for algorithmic developments. Conversely, some very clever optimization algorithms are adopted by physicists in their studies of ground states of disordered systems.

1.2 Motivations

Several optimization problems can be stated as disordered systems problems. This fact encouraged a fruitful exchange of knowledge and technical tools from one field to the other. For example the informations obtained by investigations about the structure of the phases space of disordered systems gave a deep insight in the nature of combinatorial optimization problems. This insight made possible to design new algorithms based on some physical ideas.

In both the disciplines there are several open problems. Probably the most known questions in the study of algorithm complexity is about the comprehension of the intrinsic complexity of problems: in the Non Deterministic Polynomial set of problems there is a subset of problems that look to be intrinsicly more difficult to solve than the rest. The root of the complexity of this set of problems is not well understood.

One of the problem pepoles would like to approach with the physics tools is the study of approximated algorithms: algorithms that respond to the every day life need to find good solutions, not necessarily the best one for optimization problems.

Cavity fields have been introduced in the context of disordered systems physics. The cavity fields are some effective fields that give informations about the variables of a given system, they can be easily generalized so to give informations about the variables of optimization problems.

In this thesis some approximated algorithms which use cavity fields are analyzed.

1.3 Outline of the Thesis

The first part of this thesis is devoted to an introduction to the physics of disordered systems, with a specific attention to its relation with combinatorial optimization. In chapter 2 we introduce only the basic concepts of statistical mechanics that will be used in the rest; then in chapter 3 we show how disordered systems have been studied thanks to the introduction of the replica methods and how they are described by a breaking of replica symmetry. In chapter 4 we briefly introduce combinatorial optimization. In chapter 5 we describe the main physical tool we use in the thesis: cavity fields.

The second part of the thesis is devoted to the description of some

new results. In chapter 6 we describe how it is possible, for a fixed instance of an optimization problem (in our case the *Max-Cut problem*), to find some bounds for the energy of its optimal state using the cavity fields. This general method here is applied for the specific case of Ising Spin Glasses on Random Graphs. In chapter 7 we analyze the behaviour of an algorithm based on the passing of messages (Belief Propagation) on the *Assignment Problem*. This algorithm surprisingly results to be exact for the Assignment problem. We give a proof of the exactness of this algorithm and describe its behaviour. In chapter 8 we discuss some variants of the Assignment problem and introduce then a variant: the one-in-two problem, suggested us as a simpler but no easier version of the Traveling Salesman Problem. For this problem here we show the NP-completeness.

1.4 Contributions of this Thesis

In this thesis we give contributions on some points mainly concerning approximated algorithms.

More precisely we introduce a new algorithmic method for estimating the energy of the ground state of the Spin Glasses Model (the basic disordered systems model). Altough this method could be introduced also without the introduction of cavity fields, their use is crucial to make the algorithm well performing. An upper-bound of the energy is found just by performing an iterated decimation of the spins using cavity fields. A lower-bound is found using iteratively an exact inequality on the ground-state energy.

A study of the performance of belief-propagation on the assignment problem has been done. It has been showed that in a finite time (instance dependent) this algorithm gives the solution of the problem and a certificate of the fact that the feasible solution obtained is the optimal one. For each instance we find an upper bound on the time needed by the belief propagation algorithm we describe. There is numerical evidence of the fact that this limit is not improvable substantially. In fact, given an instance with a high expected solution time, a high solution time is effectively needed to find the solution [9].

We introduce the one-in-two problem, a problem with some remarkable properties: it can be described on a two dimensional modular graph, it is possible to encode other known problems with a small degree polynomial, its interactions are local. For this problem we prove

NP-completeness. In doing that we obtain some reductions from the Satisfiability problem to the 3-dimensional assignment problem.

Part I

Statistical Physics and Combinatorial Optimization

Chapter 2

Equilibrium Statistical Mechanics of Ordered Systems

This chapter is a brief introduction to the concepts of equilibrium statistical mechanics. It is written for the reader who lacks of familiarity with this concepts in order to build a background needed for reading the remaining part of this work. This chapter is not exhaustive of the subject and the reader already familiar with statistical mechanics can skip it.

In the context of classical mechanics, a system composed of some interacting particles could be studied in principle by solving the fundamental laws for the whole set of particles. For a system of thermodynamic size in which the number of particles is huge such a program is not practicable and not even useful.

What we can do in this case is to determine the macroscopic properties (not mechanical) that are a manifestation of the underlying microscopic structure.

This program has been approached with two different methods: statistical mechanics and thermodynamics.

Thermodynamics was born during the industrial revolution when the structure of matter was not well known and was mainly based on empirical observations. Thermodynamics determines relations among macroscopic properties (pressure, volume, temperature) without any need of mechanical laws or any reference to the microscopic structure of the system. To do this, anyway, thermodynamics needs the introduction

of new principles.

Statistical mechanics instead tries to deduces the macroscopic properties starting from the laws of mechanics. However statistical mechanics needs to make some hypotesis on the microscopic structure, on the interactions and needs some approximations. Usually statistical mechanics needs a model: a simplified system enclosing the essential feature of the system we want to study. The importance of statistical mechanics lies in the fact that it does not need some more principles and it gives an understanding of the relationship between microscopic structure and macroscopic behaviour.

It is not possible to make a list of all the fields of application of thermodynamics. In fact every system with many particles, from a stone to the man, can be studied with thermodynamics and its tools.

One important chapter in statistical mechanics is the study of phase transitions. An example of phase transition is the passage from ice to water. A little decrease of the temperature of water can induce some macroscopic changes in the structure of the water that becomes solid.

In many cases it has been shown that the thermodynamic behaviour of a sample does not depend on the microscopic details of the particles and their interactions. This fact justifies the use of simple models and approximations to get the macroscopic behaviour of systems at phase transition.

2.1 Equilibrium Statistical Mechanics: basics

The most general result of equilibrium statistical mechanics is concerns the occupation probability for the microscopic states. Here we do not discuss how it can be shown, then we just assume as a principle the fact that the probability that a given system in contact with a thermal bath at inverse temperature $\beta = 1/T$ with Hamiltonian H be on the state x is the following function only of x's energy:

$$\mathrm{Prob(x)} = \frac{1}{\mathrm{Z}}\mathrm{e}^{-\beta \mathrm{H(x)}} \tag{2.1}$$

where Z is the normalization (such that $Prob(x)$ is a probability measure). Then Z is a function of the temperature:

$$Z(\beta) = \sum_x e^{-\beta H(x)}$$

this function is called partition function and has a special role in statistical mechanics, in fact the thermodynamical observables can be deduced from the partition function, here we give three examples: the *internal energy*, the *free energy*, and the *entropy*:

$$\begin{aligned} U(\beta) &= -\frac{\partial}{\partial\beta}\ln(Z(\beta)) \\ F(\beta) &= -\frac{1}{\beta}\ln(Z(\beta)) \\ S(\beta) &= \beta\left[E(\beta) - F(\beta)\right] \end{aligned}$$

The energy of the system is the average value of the energy of the microscopic system, averaged with the Gibbs measure:

$$U(\beta) := \sum_x H(x)e^{-\beta H(x)}$$

The entropy is the meaures of the information contained in the Gibbs distribution, in Shannon sense and with an offset given by the fact that the uniform distribution (corresponding to $\beta = 0$ infinite temperature) there is no information at all and, if the probability distribution is concentrated on a single state, then there is complete information about the state of the system. The entropy is then a measure of the disordered of the system. Its definition is the following:

$$S(\beta) = -\sum_x Prob(x)\ln(Prob(x))$$

An important remark should be made about the entropy: among all the distributions probability with a given value of internal energy, the equilibrium one (the Gibbs measure) is the one that maximizes the entropy.

A system at thermodynamic equilibrium can be viewed as being in a state x and evolving according to an ergodic Markov chain such that in infinite time the fraction of time spent in each state will be proportional to $\exp(-\beta H[x])$

If the system is isolated then its energy can not change. In this case the system (constrained to have an energy fixed by the initial conditions) will be with equal probability in any allowed configuration. Remark that the uniform distribution is the distribution maximizing the entropy within the fixed-energy constraint. Such modelization of a macroscopic

system is called microcanonical and we will refer to the corresponding probability distribution as *microcanonical ensemble*.

Anyway, also if it is never possible to isolate a system, microcanonical ensemble is meaningful because, for large systems, the Gibbs distribution often concentrate on a value of energy.

2.2 Phase Transitions

We talk of phase transition when a continuous change of the temperature of a system induces a discontinuos variation of a thermodynamical (macroscopic) observables.

More precisely we say that there is a phase transition of the n-th order when a (C^∞) change of the temperature induces a discontinuos variation of the n-th derivative of the partition function.

Remark first of all that a given system with a finite number of states has always an analytical partition function, then the observables obtained as sums or products of a finite number of derivatives of Z with respect to β are always analytical: there are no discontinuities in the observables or in their derivatives.

It can be observed that in correspondence of a phase transition often the distribution probability of many systems undergoes a break of the simmetry of the Hamiltonian.

Formally both these phenomena (phase transitions and breaks of the simmetry) should be possible only for infinite systems: in any finite system the asymptotic distribution is always the Gibbs one and thus no break of the Hamiltonian's simmetry is possible. Anyway also for finite system the time needed to observe the asymptotic Gibbs distribution can be larger than observation time.

The ergodicity breaking of the phase space happens when, in the local dynamics, the passage from a phase to another requires the passage from an high free energy region (w.r.t. β) of the phase space between the valleys (for the free energy landscape).

Often the phases are charectarized by some quantities as for example the magnetization for the transition from the paramagnetic state to the ferromagnetic. For the Ising model (we will discuss it in the next section) used to describe this transition, while the high-temperature phase is a zero-magnetization one, there are two low-temperature equivalent phases characterized by the (non zero) value of the magnetization.

As we are often interested both in phase transitions and in very

large systems, then we have to consider infinite systems or to consider the *thermodynamic limit*: the physical observables (like the energy) can be defined in such a way to have a meaningful limit as N, the size of the system goes to infinity. We can for example define the energy per variable $\varepsilon := E/N$, so that the limit for $N \to \infty$, if it exists, can be interpreted as the energy per site in the infinite system.

2.3 Ising System and Mean Field

Ising model is an important archetypal model for the description of phase transitions in ordered systems. Consider N spins $\{\sigma_i\}_{i=1,\dots n}$ complexively denoted by the letter σ, and each spin can be in one of two states: 1 and -1 also referred as *up* and *down*. Consider the following Hamiltonian:

$$H(\sigma) = J \sum_{\langle i,j \rangle} \sigma_i \sigma_j + h \sum_i \sigma_i \tag{2.2}$$

where the first summation is performed only over the nearest neighbours. J is the strenght of the interaction between the sites and h is an external magnetic field.

This model can be easily solved in a one-dimensional lattice as described in section 3 and it has been solved in two dimensions by Onsager [42]. In dimensions higher than one a phase transition of the second order (involving discontinuities in the second derivative of partition function) occurs at a finite value of the temperature.

Here we will see how the *mean field* approximation works in the Ising Model. The Ising Hamiltonian can be rewritten as

$$H[\sigma] = \sum_i \sigma_i \left(h + \frac{Jq}{2q} \sum_{j \in V(i)} \sigma_j \right)$$

where q is the number of neighbours of the i-th site, and the summation is on the neighbours $V(i)$ of i. If we are not able to solve this equation we can do the following substitution: $\frac{1}{q} \sum_{j \in V(i)} \sigma_j \to \frac{1}{N-1} \sum_{j \neq i} \sigma_j$. So that the new Hamiltonian reads

$$\tilde{H}[\sigma] = H \sum_i + \frac{Jq}{N-1} \sum_{i,j} \sigma_i \sigma_j$$

Where the information about the coordination in our original model is in q. If we define the magnetization $m = M/N$ as the average value of

the spin variables: $M[\sigma] = \sum_i \sigma_i$ then $M^2[\sigma] = N + \sum_{i,j} \sigma_i \sigma_j$ and the mean field Hamiltonian is

$$\tilde{H}[\sigma] = H \cdot M[\sigma] + \frac{J\,q}{N-1} \frac{M^2[\sigma] - N}{2}$$

Let r be the number of down spins then $M[\sigma] = N - 2r$ and $Z(\beta) = \sum_{r=0}^{N} c_r$ with $c_r = \binom{N}{r} \exp\{\beta H(N-2r) + \frac{J\beta q}{N-1}\frac{1}{2}[(N-2r)^2 - \frac{N}{2}]\}$. As we said $m = \frac{1}{N}\langle M[\sigma]\rangle = \frac{1}{Z}\sum_r c_r(1 - 2\frac{r}{N})$. An easy calculation shows that c_r has is maximum value for $M = \frac{1-e^{-2\beta(qJM+H)}}{1+e^{2\beta(qJM+H)}}$, moreover the function c_r is concentrated on this value: it has a width of order $N^{1/2}$. For $\beta q\, J < 1$ the solution to the equation for the magnetization is $M = 0$. For $\beta q\, J > 1$ there are three solutions, and we should look for the one minimizing the free energy.

In other words there are a lot of configurations with zero magnetization, but these configurations do not minimize the energy, at high temperature the maximization of entropy is more important than the minimization of energy so the configurations on which the probability measure concentrates are the zero-magnetization configurations. At low temperatures the minimization of energy is more important then entropy maximization so the probability distribution will be dominated by configurations of non zero magnetization.

The phase transition then consists in an increase of the correlation between the spin variables in the sample. If the dynamic (as often happens in many physical systems) is such that the positive magnetization states are not enough easy to be reached starting from the negative ones, then the system will fall in one of the two phases

Remark that our mean-field approach is equivalent to the following approximation: $H = \sum_i \sigma_i h_{\text{eff}}$ where the effective external field has to be chosen consistently with the magnetization of the system: $h_{\text{eff}} = (h + \frac{Jq}{N-1}\sum_i \sigma_i) = h + Jq\langle\sigma_i\rangle$. In other words the mean field approximation is equivalent to the assumption of independence of different spins, in fact the only interaction between spins emerge by the consistence condition.

2.4 Cluster Property

It is also possible to define pure phases in a more formal way. This approach starts from the assumption that a given ground-state of the system can be perturbed only locally under local perturbations.

This property must be true only for a given ground state: when the system allows for different ground states if we perform averages over a probability distribution (concentrated on more then one phase) then we get systematic correlations.

We can say that a probability distribution μ describes a pure phase if and only if for each pair of local physical observables $A(x)$ and $B(x)$, the connected correlation function vanishes in the large distance limit:

$$\lim_{|x-x_0|\to\infty} (\langle A(x_0)B(x)\rangle_\mu - \langle A(x_0)\rangle_\mu \langle B(x)\rangle_\mu) = 0$$

For Ising model the observable that allows to distinguish the two low temperature phases is the magnetization (in the language above $A = B = \overline{\sigma}$ is already a valid choiche). The two low temperature pure phases can be selected by adding an infinitesimal external magnetic field (if we are performing the thermodynamic limit then the external magnetic field times the size of the system should be kept not smaller than $1/\beta$, any finite external field works).

The distributions minimizing the free energy at the leading order are all the measures that can be obtained as convex combination of the two extremal measures satisfying the cluster property.

Chapter 3

Disordered Systems

In this chapter we will shortly describe some of the tools and results in the field of disordered systems.

The basics ingredients of disordered systems are frustration and disorder. Given a spin system with space of configurations σ and a Hamiltonian $H[\sigma]$ given by the summation of several contribution each one involving a subset of the spin variables, then we say that there is frustration if there is not a configuration such that all the contributions to the Hamiltonian are minimal. An example of frustrated system is the antiferromagnetic Ising model on a triangular lattice, indeed it is easy to check that for each plaquette, at least one of the three boundary edge on the plaquette gives a positive contribution to the Hamiltonian. Frustration, as happens in this example, is often associated to the presence of several ground-states, or low energy states with energy almost equal to the energy of the ground-state, more precisely an exponential number on the large N limit. For comparison, in ordinary ferromagnet, as the Ising model at zero external field, there are 2 ground states and a number of states with k excitations only polynomial in N (of degree k).

The other fundamental ingredient of disordered systems is disorder. As we saw in chapter 2, statistical mechanics of ordered systems concerns systems where all the particles are equivalent. This is not true for all the physical systems. For example there are several systems that have two different scales of times of thermalization, a well known example is the one of glasses, where the mutual position of molecules changes slightly, in the years as the result of little arrangements of the molecules, while small energy excitations diffuse over the glass at faster time-scales.

Suppose that we have two set of variables: $\phi_1\,\phi_2$, in the context of statistical mechanics of ordered systems we would be involved in the study of the properties of the measure obtained by extracting the possible configurations with weights given by the Gibbs distribution:

$$e^{-\beta H[\phi_1,\phi_2]}$$

When we have two thermalization times, we have to extract a configuration ϕ_2 with an opportune weight given by an effective Hamiltonian of only the slow variables: $\tilde{H}[\phi_2]$ and then we should study the fast degree of freedom with fixed ϕ_2, so with the measure given by $H_{\phi_2}[\phi_1]$. In other words we are interested in studying the partition function

$$Z(\beta) = \sum_{\phi_2} Z_{\phi_2}(\beta)\mathrm{Prob}(\phi_2)$$

As we are concerned with thermodynamic properties, we want to calculate the free energy that is $\overline{\ln Z_{\phi_2}(\beta)}$, the average over the slow variables of the free energy of the system of fast variables with fixed ϕ_2.

Consider the average values of $\overline{Z^n_{\phi_2}[\beta]}$, these quantities physically represent a setting with n independent systems (n replicas of the same system, i.e. whith ϕ_2 fixed). The free energy of n replicas is $F_n := -\frac{1}{\beta n}\overline{Z^n_{\phi_2}[\beta]}$. It has a well-defined meaning for n integer. The replica trick consists in letting n free to have real values so to calculate the logarithm of Z using the limit: $\ln Z = \lim_{n\to 0}\frac{Z^n-1}{n}$, and obtain the average free energy as $\overline{F} = \lim_{n\to 0} F_n$.

The replica trick involves an analytical continuation that is not clear under which assumptions it holds. We will discus some of these issues through one of the most simple examples: the random energy problem (REM) introduced by Deridda [11], which is an easy example of glassy phase transition.

3.1 REM

Given a set of N colour variables (each one can assume values $1,\dots q$), an instance is defined by the cost of each configuration: $E[\sigma]_{\sigma\in\{1,\dots q^N\}}$, where the costs are drawn independently with the same distribution. We draw the energies from a Gaussian distribution:

$$\rho_N(E) = \frac{e^{-E^2/(NJ^2)}}{\sqrt{\pi N J^2}}$$

defined in such a way to have a non trivial thermodynamic limit.

Remark that the optimization problem associated to this model is not in the NP class. In fact a check of the fact that a given feasible solution is the one which minimizes the cost needs to read the whole list of energies.

Given a temperature $1/\beta$ the probability distribution for the states is given by Gibbs distribution (2.1). The partition function $Z_E(\beta)$ depends on the temperature and on the instance.

Since the energies associated to the configurations are i.i.d. random variables, the number of configurations with energy in a "window" $[E, E+\epsilon]$ is a binomial random variable. So, for values of E out of some interval $[-\epsilon^*, \epsilon^*]$ the typical density of energy levels is exponentially small in N.

With $q = 2$ the density of levels of given energy E is

$$n(E) = 2^N P(E) \simeq e^{N[\log 2 - \frac{E^2}{N^2 J^2}]} \tag{3.1}$$

We now work in the *microcanonical ensemble*: we suppose the system to be constrained to have an energy in a narrow window, and we give to every allowed configuration the same probability of being realized.

The entropy $S(E) = \log n(E)$ is then $N[\log 2 - \frac{E^2}{N^2 J^2}]$ for $|E| < E_0$ with $E_0 = NJ(\log 2)^{1/2}$, and $S(E) = 0$ for $|E| < E_0$. The temperature is

$$\frac{1}{T} = \frac{\mathrm{d}S}{\mathrm{d}E} = -\frac{2E}{NJ^2}$$

we concentrate on the region $-E_0 < E < 0$ that means $T > T_0 > 0$ with $T_0 = \frac{NJ^2}{2E_0}$. The free energy is $F = E - TS = -\frac{nJ^2}{4T} - NT\log 2$. Remark that $F(T_0) = -E_0$. Let's study the REM in the replica formalism to check the solution against the one just given.

The replica partition function is

$$Z^n = \sum_{\sigma_1 \dots \sigma_n = 1}^{2^N} \prod_{\sigma^* = 1}^{2^N} e^{-\beta E \dot{\sum}_{i=1}^{n} \sigma \chi(\sigma_i = \sigma^*)}$$

Given the distribution (3.1) we can perform the sum over the all the possible energies the level can assume, a Gaussian integral gives that the average over the external configurations of Z^n, the replica partition

function is

$$\overline{Z^n} = \sum_{\sigma_1 \dots \sigma_n = 1}^{2^N} e^{\frac{\beta^2 N}{4} \sum_{\sigma^*=1}^{2^N} \sum_{i=1,\dots,n} \chi(\sigma_i = \sigma^*)} = \sum_{\sigma_1 \dots \sigma_n = 1}^{2^N} e^{\frac{\beta^2 N}{4} \sum_{i,j=1}^{n} \chi(\sigma_i = \sigma_j)} \tag{3.2}$$

Now this expression appears like a partition function for the system of n replicas, but is no longer dependent on an external disorder (that has been integrated out). However the replicas are now no longer independent, this is not strange, in fact it is just the effect of the integration over the disorder. In fact, especially at low temperatures, all the replicas tend to stay in the low-energy states, so that there is a sort of "attractive" interaction among the replicas.

A given configuration (set of $\sigma_1 \dots \sigma_n$) enters in the effective Hamiltonian of the partition function (3.2) through the overlaps of the replicas. We introduce a matrix $Q_{i,j} := \chi(\sigma_i = \sigma_j)$ with elements in $\{0, 1\}$. This matrix is symmetric, the entries on the diagonal are 1. Its most general form (up to permutations) is a set of diagonal square blocks filled with 1-s and 0 elsewhere

Remark that the system described by the replicas is invariant under the group of permutations of the replicas. This simmetry, called *replica simmetry*, follows by the definition itself of the replica system and is a general feature when using the replica trick. This simmetry of the replica Hamiltonian suggests us that the system itself is symmetric under permutations of the replicas, when it happens one says that the system is replica symmetric.

So we suppose that the system is invariant under permutations and then $Q_{i,i} = 1$ and $Q_{i,j} = q \ \forall i \neq j$, so there are two possible choice for Q: $Q_{i,j} = 1 \ \forall i, j$ this correspond to having all the replicas in the same configurations. $Q_{i,j} = \delta_{i,j}$ this one corresponds to the configurations with all the replicas in different configurations. Clearly we expect the first situation to describe low temperatures behaviour, as if all the replicas were constrained to remain in the same energy level, and we expect the second picture to dominate the summation in the partition function. (3.3). Clearly this is an approximation at fixed N size of the system, but as we are interested in the thermodynamic limit it could give us the exact result.

We now let n to be real. We can write the partition function in the following way:

$$\overline{Z^n} = \sum_{Q} \exp(N g(Q)) \tag{3.3}$$

as a sum over all the possible Q matrices of an opportune function given by a term that correspond to the energy of Q and another that keep account of the number of different ways to realize that given Q.

As we are interested in the limit $N \to \infty$, we search the minima of $g(Q)$. We observe (for $n > 1$) that the function $g(Q)$ has a phase transition at temperature $\beta_c = \sqrt{4 \log 2/n}$, in other words the replica symmetric configuration minimizing the value of $g(Q)$ is the $q = 0$ one for high temperature and the $q = 1$ for low temperatures. for $n < 1$ there is a phase transition at $\beta_c = \sqrt{4 \log 2/n}$ but surprisingly it goes in the other direction: the high temperature phase correspond to the $q = 1$ configuration for Q. This is a general problem when considering $n < 1$. The solution of this problem is linked to the fact that the eigenvalues undergo a change of sign (because of the prefactor $n(n-1)$) when n is decreased below 1, so that the minimum is actually realized by the maxima for $n > 1$.

We are using an estimation of the free energy valid in the large N limit and then we are performing the $n \to 0$ limit. Remark that the correct calculation consists in doing the limit on the number of replicas before. In doing so we are implicitly assuming that the two limits commute.

We obtain that for any fixed temperature the sum in the partition function (3.3) is dominated by the phase with $q = 0$ meaning that all the replicas are in different states, the expression that we get for the free energy is

$$-\beta F := \lim_{N \to \infty} \frac{1}{N} \overline{Z^N} = \frac{\beta^2}{4} + \log 2$$

this is the same result obtained in the *microcanonical ensemble*, but we should investigate the stability of this solution after breaking the replica simmetry, we have to investigate if there is a low temperature regime when $\beta < \beta_c = 2\sqrt{\ln 2/n}$.

The problem is to find a larger set of matrices Q such to include the ones minimizing $g(Q)$ (or maximizing for $n < 1$). In the RSB scheme first proposed by Parisi, a recursive procedure is defined, it consists in the breaking of the group of permutations.

Suppose that n is a multiple of m, then we can divide the n replicas in n/m groups of m replicas, we will restrict, at the first step of the breaking of the simmetry to the matrices such that $Q_{i,i} = 1$ (for definition), $Q_{i,j} = q_1$ if the replicas are in the same group and $Q_{i,j} = q_2$ if the two replicas are in different groups.

Obviously $q_1, q_2 = 0$ correspond to the configuration with all the replicas on different levels, $q_1, q_2 = 1$ correspond to the configuration with all the replicas in the same level, so we are interested in the new set of matrices obtained with $q_1 = 1, q_2 = 0$ where all the m replicas in the same group are in the same configuration and replicas which belong to different groups are in different states.

Calculating the function $g(Q)$ we have a contribution from the energy that is $\frac{\beta^2}{4}nm$ and a contribution from the entropy (the logarithm of the number of ways to realize the configuration) that is $\frac{n}{m}\log 2$. This expression is minimized by $m(\beta) = \frac{2\sqrt{\log 2}}{\beta}$. Remark that the inequality $1 < m < n$ valid for integer values become $n = 0 < m < 1$ in the $n \to 0$ limit, so the expression for m makes sense only if $\beta < 2\sqrt{\log 2} = \beta_c$. We then obtain that for this solutions the energy is $-\beta F = \beta\sqrt{log2/n}$.

In the 1RSB scheme, an arbitrarly chosen replica has a probability m to be in the same state of another replica and a probability $1 - m$ to be in the same state. This situation is represented by a growing function $q : [0, 1] \to [0, 1]$ that in the RS solution takes only one value and in the 1RSB takes only two values.

3.2 Spin Glasses

A typical example of disordered system are spin glasses. Consider a spin system on a lattice (some particles on a lattice with some possible spin configurations) and an interaction between neighbouring spins as in equation (2.2), with the remarkable difference that the strenght and nature of the interaction is not constant:

$$H(\sigma) = -J_{i,j}\sum_{\langle i,j\rangle}\sigma_i\sigma_j - h\sum_i \sigma_i \tag{3.4}$$

Here the interactions J are the quenched variables, remark that if they are positive then (at low temperatures) the spins tend to take the same value, if they are negative the spins will tend to have different values.

Remark that it is not possible to satisfy all the tendency to be aligned or disaligned on a *loop* $(i_1, i_2, \ldots, i_l)$ of spins if on that loop the product of interactions $J_{i_1,i_2}J_{i_2,i_3} \cdot J_{i_l i_1}$ is negative.

As we have done for the case of ordered system, we begin by discussing the mean-field case: the case where every spin interacts with every other spin.

3.2.1 The Sherrington Kirkpatrick model

In the Sherrington Kirkpatrick model of Hamiltonian (3.4) and null external field the interactions $J_{i,j}$ between the spins are i.i.d. drawn from a probability distribution $\mu_0(J)$.

$$\mu_0(J) = \sqrt{\frac{N}{2\pi}} e^{-\frac{N}{2}J^2}. \tag{3.5}$$

This model has an infinite range of interactions, then it does not correspond exactly to a physical system defined in a finite dimensional space: there is not a notion of neighbouring or a geometric distance between pairs of spins. The dependence on N in the distribution (3.5)is such that in the large N limit the complete interaction on every site has a non trivial limit.

The partition function of the replicas is:

$$\overline{Z_N^n} = \sum_{\sigma} \int \prod_{i<j} \left(\mathrm{d}J_{i,j} \sqrt{\frac{N}{2\pi}} e^{-\frac{N}{2}J_{i,j}^2 + \sum_{a=1}^{n} \beta J_{i,j} \sigma_i^a \sigma_j^a} \right)$$

where the summation on σ is a summation over the 2^{nN} configurations on the replicas. We integrate over J analogously as we have done for the REM, we get:

$$\overline{Z_N^n} = \sum_{\sigma} e^{\frac{\beta^2}{2N} \sum_{i<j} (\sum_{a=1}^{n} \sigma_i^a \sigma_j^a)^2} = e^{\frac{(N-n)n\beta^2}{4}} \sum_{\sigma} e^{\frac{\beta^2}{2N} \sum_{a<b} (\sum_{i=1}^{N} \sigma_i^a \sigma_j^a)^2}$$

After Gaussian integration we get

$$\overline{Z_N^n} = e^{\frac{(N-n)n\beta^2}{4}} \int \prod_{a<b} \frac{\mathrm{d}Q_{a,b}}{\sqrt{2\pi/(N\eta^2)}} e^{-NS[Q]}$$

where $Q_{a,b}$ is a $n \times n$ matrix, zero on the diagonal, which describes the couplings of the n-sites system. Remark that the dependence on N has been highlighted in this expression, so to evaluate the partition function we will use (as for the REM) a saddle-point estimate. $S[Q] = -\ln M[Q] - \ln Z[Q]$ where the first term is the entropic part and the second is the energetic part. As in the REM there is an interaction between the independent replicas due to the fact that they share the same quenched interactions.

$$M[Q] = e^{-\frac{\beta^2}{2} \sum_{a<b} Q_{a,b}^2}$$

$$Z[Q] = \sum_{\sigma \in \{\pm 1\}^n} e^{-\beta^2 \sum_{a<b} Q_{ab}\sigma_a\sigma_b}$$

As for the REM model the action is invariant under the action of the group of the permutations of n elements. In the replica system, with integer n, the matrix Q has the physical meaning of representing the *overlap* between the replicas, in fact the solution of the saddle-point equation:

$$\left.\frac{\partial S[Q]}{\partial Q_{a,b}}\right| = 0$$

gives us

$$Q^* = \langle \sigma_a \sigma_b \rangle = \frac{1}{N} \sum_i \overline{\langle \sigma_a^i \sigma_b^i \rangle} \tag{3.6}$$

With "overlap" between two configurations we mean the number of spins with the same value minus the number of spins having different value: $q(\sigma^a, \sigma^b) = \frac{1}{N}\sum_i \sigma_i^a \sigma_i^b$, we define a distance between configuration of spins: $d(\sigma_i^a, \sigma_i^b) = 1 - q(\sigma^a, \sigma^b) = \frac{1}{2N}\sum_i (\sigma_i^a \sigma_i^b)^2$. At fixed instance the distribution

$$p(q) = \langle \delta(q(\sigma^a, \sigma^b) - q) \rangle$$

gives a measure of the correlation and the distance between configurations in different replicas. This quantity is not dependent on the number of replicas because they are mutually independent. Remark that $p(Q) = lim_{n\to 0} \frac{2!}{n(n-1)} \sum_{a<b} \delta(Q_{a,b}^* - Q)$ and the equation (3.6) implies that $p(Q) = \overline{p(q)}$

3.2.2 Replica Symmetric Ansatz

We seek for a solution of the saddle point equations. It is natural to propose an ansatz symmetric under permutations of the replicas, we thus try a Replica Symmetric one:
For all $a \neq b$, $Q_{ab} = q$.

Plugging this result in the equation (large N limit) and then performing the $n \to 0$ limit, one find that there is a critical temperature such that for $T > T_c$ the solution is $q = 0$ and for $T < T_c$ there is a not null solution. Unfortunately, it turns out that at low temperatures, not only the free energy density does not agree with numerical simulations, but also, the entropy is negative. This is a clear sign that the RS Ansatz is not stable at all the temperatures.

We are thus forced to propose an Ansatz breaking the simmetry of the action.

It is possible to classify all the possible matrices Q with the group of simmetry under permutation of its elements. Each group G_Q can be represented as a rooted tree with n leaves. At each node a subset (with cardinality c) of the arcs going to a lower level is grouped if and only if all the $c!$ permutations which exchange rigidly the sets of leaves on the branches are in G_Q.

We will restrict to a smaller class: the class of the fully simmetric trees: the trees in which each node has a unique grouped set of itd brnches (downgoing arcs). This implies that the tree is graded (it exists an height function such that all the leaves have the same height), it will have a certain height $k+1$, the nodes at distance k' from the root have all the same number of leaves $m_{k'}$ below of them. Of course $n = m_0 > m_1 > \cdots > m_k > m_{k+1} = 1$ and each m_h is a divisor of m'_h if $h' < h$. If we define the distance between two leaves as the height of the lowest ancestor (the height is bigger as we approach the root) then the simmetry of Q implies that in the class of full simmetric trees the value of $Q_{a,b}$ depends only on the distance between the leaves.

Just to get confidence: the simplest full simmetric tree (matrix) is the two level tree with all the leaves connected to the root with n grouped arcs. This correspond to a matrix where all the leaves are equivalent and it is invariant for permutations, this matrix is a RS one.

In the case of the REM: 1 step of replica simmetry breaking the tree representing the simmetry group of Q is a tree with two levels, n leaves linked to m/n sites at the first level (each grouping m leaves) and finally a root grouping m/n branches. This means that the replicas are interchangeable in every subset of m replicas and whole sets are interchangeable.

Remark that, as we told for the REM, the inequality on m at different levels makes no more sense for $n < 1$ and should be replaced with $m_i \in [0,1]$ and $m_i \geq m_j$ if $i > j$. Furthermore we have that $q_1 \geq q_2$ meaning that q is bigger for more distant (on the tree metric) replicas.

The choice to search solutions for the saddle point equation gives a better result than the Replica Symmetric one, but it gives negative entropy at zero temperature. We then can consider deeper and deeper trees, at every step we have a better approximation of the solution.

The limit process of taking in account a bigger and bigger level of simmetry breaking is called full Replica Simmetry Breaking (∞RSB)

Define at each k level of breaking of the simmetry the quantity:

$$x(q) = \int_0^q \mathrm{d}Qp(Q)$$

the set of allowed x(q) is the set of monotonic k-step functions mapping $\mathcal{R}$ into $[0, 1]$, the p(q) function is a sum of delta functions. In the limit $k \to \infty$ we get that $x(q)$ is a monotonic function. And the probability distribution $p(q)$ is not more constrained to be a sum of Dirac's delta. Solving the saddle point equation we find that $p(q)$ has only two singular contributions:

$$p(q) = x_{min}\delta(q - q_{min}) + x_{max}\delta(q - q_{max}) + \tilde{p}(q)$$

where $\tilde{p}(q)$ is a smooth convex function with support on $[q_{min,q_{max}}]$: the probability density to have $Q_{a,b}$ smaller than q_{min} or bigger than q_{max} is zero.

The replicas with the metrics induced by the distance as the overlap is an ultrametric space. We will introduce ultrametric spaces in next section.

This surprising form is not a peculiarty of the *SK model*. The whole replica machinery can be applied to different disorderd systems. For some of them, the RS ansatz turns out to be the correct one for all the temperatures. To this group belongs the *Assignment problem* and also the *Traveling Salesman Problem*. For other systems, like the *K-SAT* problem or the *Viana-Bray* model the correct solution is given by the 1-RSB ansatz in a low temperature region, and by the RS Ansatz in a high temperature region. There exists systems, like the *p-spin Model*, where the structure of the phase space is described by an 1-RSB Ansatz at a intermediate range of temperature, and by full RSB Ansatz in a low temperature region.

3.2.3 Ultrametric Spaces

A metric on a space is a function $d(x, y) \to \mathcal{R}^+$ such that

$$d(x, y) = 0 \iff x = y$$

$$d(x, y) = d(y, x)$$

$$d(x, y) \leq d(x, z) + d(y, z)$$

In vectorial spaces it is possible to define a metric starting from a norm. A space is ultrametric if the triangular disequality is replaced by the stronger condition:

$$d(x, y) \leq \max[d(x, z), d(z, y)] \; \forall x, y, z$$

As a consequence one get

$$d(x, y) < d(y, z) \Rightarrow d(x, z) = d(y, z)$$

meaning that all the triangles in the space are wheter equilateral wheter isosceles (with the two equal size edges longer than the smaller edge). Remark also that every point in a ball is its center.

Remark also that the remarkable property holds: $|x+y|_p \leq \max(|x|_p, |y|_p)$, so a triangle is wheter equilateral wheter isosceles with the two equal size edges longer than the other one. Remark that in an ultrametric space two balls are wheter non intersecating wheter one is inside the other. Remark also that for the balls in an ultrametric space the diameter is equal to the radius. The space is not an Archimedean because there are some x such that $|x + x|_p$ is not bigger then $|x|_p$.

A way to represent an ultrametric space is a rooted tree. The nodes of the tree are an ultrametric space if we use as distance the height of the lowest common ancestor.

3.2.4 Ground States in Spin Glasses

The problem of finding the ground-state of spin glasses is a typical optimization problem. We will introduce it in greater detail in chapter 6, we will introduce it also in a slightly different (but equivalent) form in chapter 4.1 as *Max Cut Problem*. In fact, given a spin glass on a graph where the nodes represents the spins, every spin configuration induce a natural partition of the set of sites in two subset (the up-sites and the down-sites). Let γ be the set of edges with an endpoint on an up-site and the other endpoint on a down-site. The energy of the configuration can then be written as:

$$E(\sigma) = -H_0 + 2 \sum_{\gamma} J_{i,j}$$

Where H_0 is a constant given by the sum over all the edges and does not depend on the partition. Thus finding the ground-state for the spin glass is equivalent to find the partition of the sites in two set such

that $2\sum_\gamma J_{i,j}$ is maximized. This problem belongs to the class of *NP-complete* problems. The set of edges γ is a cut and to every cut one associate the weight $2\sum_\gamma J_{i,j}$. Anyway it has been showed that on some graphs the problem is "easily" solvable, for example it is polinomially solvable on planar graphs.

Chapter 4

Combinatorial Optimization

The basic ingredients of a combinatorial optimization problem are a finite space of feasible solutions, and a cost function over this space. We will consider two classes of problems: optimization and decision. The optimization problem is solved when a feasible solution which minimizes the cost is found. The decision problem is solved when a feasible solution that costs less than a given threshold is found.

We are interested in the solution of problems via algorithms (a sequence of elementary steps). Algorithmic Complexity in particular studies the scaling of he required time with problem size. An algorithm is, loosely speaking a procedure for solving a problem. More precisely it is a set of instructions understandable by an appropriate automatic machine, such that, given some input data, in a finite number of steps leads to some output.

It is easy to see that the solution of an optimization problem contains the solution of the related decision problem, it is less obvious the fact that an algorithm for the decision problem can be adapted for the related optimization one. Given a procedure (algorithm) able to solve the decision problem it is often possible to find the optimal solution by bisecating the range of possible costs. It is for example the case for the Traveling Salesman Problem with integer distances: since the cost function is the sum of n terms, the interval between a trivial lower-bound and a trivial upper-bound is at most of order nL with L the difference between the largest distance and the shortest one. Bisecating over that range involves $\log_2(nL)$ iterations in the worst case.

The just proposed one is a first example of reduction of a problem to another: we showed that if we are able to "solve" one problem (the decision one) then we are able to "solve" another one (the optimization one) in comparable time. We will examine in more details how reduction permits to construct complexity classes for optimization problems.

4.1 Some Optimization Problems

Optimization Problems can often be given in mathematical terms with the use of graph theory.

Here we give a list of optimization problems defined on a connected graph G of vertices V and weighted edges E.

- *Min-Cut Problem*
  It is desired to find the cut of minimum cost for the graph G. A subset of arcs is called a cut if it is such that when these arcs are removed from G, the graph becomes disconnected.

- *Minimum Spanning Tree Problem*
  It is desired to find the minimum cost *Spanning Tree* subset of G. A subgraph S of G is spanning if the set of vertices $V(S)$ of the subgraph is equal to $V(G)$; a graph is a tree if it does not contain loops and consists of a unique connected component.

- *Chinese Postman Problem*
  It is desired to find the tour (closed path) of minimum lenght that passes through every edge at least once.

- *Assignment Problem*
  Given a bipartite weighted graph G this problem consists in finding the optimal dimer covering of the graph. A dimer covering is a spanning subgraph that contains N edges, $2N$ vertices and N disjoint connected components (the dimers).

- *Eulerian Circuit*
  This decision problem consists, given a graph, in finding if there is a circuit that visits all the edges exactly once and returns to the starting point. Such a circuit is called Eulerian because the problem was first solved by Euler in 1736, when he proved that in a connected graph there exists an Eulerian circuit if and only if every vertex has even degree.

For all these problems there are algorithms able to solve them. What we mean by saying to "solve"? Rigorously speaking any optimization problem is solvable: the naive algorithm that lists all the feasible solutions (they are finite) and pick up the best in fact solves the problem.

This kind of solution of the problem is not satisfactory for someone actually needing a solution forhis problem, in fact, given the exponential asymptotic behaviour of the number of feasible solutions, the running time needed to solve problems too large also for not-so-small instances. Suppose that the naive algorithm runs in 1 second on a given machine to solve a problem with $N!$ feasible solutions when $N = 20$. Then to solve a problem with $N = 40$ it will need $40!/20! \gg 20^{20}$ seconds. Say that a problem is not solved if we can not live enough to see the answer.

We need a more formal criterium to decide whether an algorithm solves a problem or not. A well-defined criterium of goodness is that an algorithm is good if the required number of elementary steps is bounded by a polynomial in the size of the problem. An algorithm is super-polynomial if there is not a polynomial bounding the number of operations (polynomial in the size of the problem). The size of the problem can be assumed to be the number of bits required to encode it. One can define the class P of polynomial problems as the class of problems having a polynomial algorithm.

Remark that, with the given definitions, a polynomial algorithm could be slower than a super-polynomial one for small instances. The threshold between small and large instances could be also greater than the typical one in every-day usage.

There are some problems that have not been "solved" in this sense, here we list some examples:

- *Max-Cut Problem*
  It is desired to find the cut with maximal cost.

- *Steiner Network Problem*
  for a given set of points (for example in the plane) the Steiner network is the shortest set of lines which connects all of them.

- *Traveling Salesman Problem*
  This is probably the most famous optimization problem. Given a weighted graph (where the sites are towns in the pictorial description and the cost is the lenght of the road), it is desired to find if there exists a tour that passes through each vertex at least once and costs less than a given threshold.

- *Three-Dimensional Assignment Problem*
  Given a tripartite weighted hypergraph (with hyperedges of coordination exactly three) the problem consists of finding a trimer covering (a subset of disjoint hyperedges such that the union of their vertices is the set of vertices of the hypergraph) with a cost (sum of the cost of the hyperedges) smaller than a given threshold.

- *Hamiltonian Cycle*
  it consists, similarly to the Eucledian circuit problem in finding if, given a graph, there exists an Hamiltonian cycle (a tour that visits every edge exactly once).

- *K-Satisfiability Problem*
  Given a set of N boolean variables and M clauses, each of them involving exactly K literals, the problem consists in finding (if it exists) a configuration of the variables such that every clause is satisfied.

No one proved that the problems above are not solvable by polynomially bounded algorithms. Howewer it has been shown that either all of this problems can be solved or none of them can be.

4.2 Non-deterministic Polynomial Problems

We will restrict our attention to the class of *non-deterministic polynomial* problems (NP problems).

This is the class of problems which can be solved in polynomial time by a non-deterministic algorithm. A non deterministic-algorithm is, roughly speaking, an algorithm that can run in parallel on an arbitrarly large number of processors.

This is the class of search problems such that, given a feasible solution x_s, the subproblem of checking whether x_s is a solution or not is a polynomial problem. All the problems given above belong to the NP class.

We say that a problem is in NP class if, given a feasible solution, the check that it is a solution can be done in polynomial time. In other words, the time for this algorithms should be calculated following the formal rule that the time of a forked process is the maximum of the two forked times of the forked subprocesses (as if the forked processes were running on different processors), instead of the sum of the two times (as if the forked processes were running sequencially on a unique processor).

Consider two problems A and B. We say thta A is *polinomially reducible* to B in the following case: say that $\{\mathcal{I}_A\}$ are the possible instances and $\{\sigma\}$ are the feasible configurations for A, while $\{\mathcal{I}_B\}$ and $\{\tau\}$ are the analogue for B. Call $\sigma^*(\mathcal{I}_A)$ a solution for instance $\mathcal{I}_A$, and $\tau^*(\mathcal{I}_B)$ a solution for instance $\mathcal{I}_B$. Then a map ϕ exists, the *encoding*, which associates an instance of B to an instance of A, $\mathcal{I}_B = \phi(\mathcal{I}_A)$ such that the size of $\mathcal{I}_B$ is bounded by a polynomial in the size of $\mathcal{I}_A$, and a map ψ exists, the *decoding*, which, provided a solution $\tau^*\big(\phi(\mathcal{I}_A)\big)$ of the new problem, gives a solution of the original problem, $\psi\big(\tau^*\big(\phi(\mathcal{I}_A)\big)\big) \in \{\sigma^*(\mathcal{I}_A)\}$.

The maps ϕ and ψ must be evaluable in polynomial time, and it is commonly understood that this is the case, and indeed the complexities of the encoding and decoding operations are negligible w.r.t. the complexity of solving problem B.

So the algorithm consisting of this encoding, followed by problem-B algorithm, and finally the decoding, is a polynomial-time algorithm for problem A, with polynomial $P_A(N) = P_B(Q(N))$. This idea leads to the concept of *polynomial reduction* of problems.

Remark that all the polynomial problems are NP problems, (i.e. P $\subseteq$ NP). A main question is whether every NP problem is polynomial (i.e. P = NP), in fact in that case almost the whole of the optimization and decision problems of pratical relevance would be polynomially solvable!

Remark that the idea of polynomial reduction holds also for the class NP. Polynomial reduction is a powerful tool in the study of algorithmic complexity, especially in the more "elusive" class of NP problems, indeed it can be used to construct inequalities among the intrinsic complexity of different problems. In the example above we showed that the time needed to solve the worst case of the optimization of the *Traveling Salesman Problem* is at most $\log(nL)$ times then the time needed to solve the worst decision problem (worst over the threshold a).

As we are interested mainly in a distinction among polynomial and super-polynomial, we are principally interested in polynomial reduction. Given a set of problems S, if there exists a problem such that if we are able to solve it, then we are able by polynomial reduction to solve any problem in the set, we say that this problem is S-complete. Remark that the possibility to reduce one problem to another one gives us the possibility to construct classes of problems such that whenever we are able to solve one problem in this class in polynomial time then we are able to solve every problem of this class in polynomial time. To enlarge

this class it is enough to find loops of reductions involving at least one complete problem.

The *K-Satisfiability* problem, discussed above, is of great importance also for a reason trascending its pratical applications: it is *NP-complete*. This fact has been proved by Cook [10], starting from the formal definition of problems in NP, and of non-deterministic Turing Machines. The proof of Cook's Theorem estabilishes that K-SAT is NP-complete by showing reduction of the general SAT to K-SAT, and then of the formal class of NP-complete problems to SAT.

Although the question wheter P = NP or not is open, it is widely believed that NP problems are not polynomial. For example the first set of problems given in section 4.1 is polynomial, the second set of problems (page 29) is NP-complete, instead of the fact that they can been stated on a very similar way. One of the more fascinating tasks in the study of algorithm complexity is to understand the origin of this intrinsic hardness. Heuristic arguments give a hint that their algorithmic hardness is due to some intrinsic structure of the problem.

Another example of NP-complete problem is in chapter 8, where we show that if one is able to solve an instance of the *one-in-two problem* then in a time polynomial it is possible to encode any instance of *K-SAT* problem so that, using Cook's theorem, the NP-completeness of *one in two problem* is proven.

There are some problems that do not belongs to the class of NP problems but are difficult at least as an NP-complete problem (and then at least as any NP problem). They are the problems which do not allow for a short certificate of optimality. For example the optimization version of the Traveling Salesman Problem (TSP). If I'm able to find the optimal solution of the TSP problem then automaticlly I can say if there exists a solution that cost less then the threshold.

4.3 Average case analysis

Here we presented the worst-case complexity of problems. Anyway not all the efforts are concentrated on the worst case: pratical considerations may also be directed to a form of *average-case* analysis, which indeed is the principal case of many real life applications. In fact in every day life we are often interested in the possibility to solve in a reasonable time an instance extracted from a given distribution. We will refer to the scaling of the solution time with respect to the size of the problem

as the average complexity. Remark that the average complexity of an optimization problem is well defined when a measure on the instances is given. Remark that the worst-case complexity does not depend on the probability measure we use to draw instances of the problem.

A lot of efforts are also concentrated on the study of approximated algorithms: some problems (also some NP-complete) admits some polynomial algorithms able to find an approximated solution (a feasible solution with cost almost optimal). For pratical purpose it is in fact often important to find just a "good" solution. Also for some polynomial problems (problems for which a polynomial algorithm yet exists) it could be useful to find a faster approximated algorithm.

An optimization problem can be stated as a physical problem of statistical mechanics. Indeed, the set of possible solutions can be interpreted as a configuration space, and its cost can be chosen as the Hamiltonian of the system: cost minimization corresponds then to the search of the ground state of the physical system, when it is frozen at zero temperature.

In general the hardness of the optimization problems is related to the fact that the corresponding disordered system is frustated, as the ground state configuration is not the one which simply minimizes all local interactions, it is possible that a local algorithm finding the ground state does not exists.

Many conjectures have been formulated about a relation between phase transitions and algorithmic complexity. The NP-completeness is not directly linked to the presence of a glass phase transition, in fact the NP-completeness is a property of the worst instances while the statistical analysis is about the typical instances. There are some polynomial problems that have a phases transition so we should exclude also the possibility that a phase transition implies a sort of NP-completeness on average. There are both problems with continuous (2-coloring) and with discontinuous (3-core problem) phase transitions admitting a polynomial time algorithm.

Probably it is not possible to find a simple correspondence between statistical properties of the phase transitions and the NP-completeness in a given set of instances because statistical mechanics, describing the structure of the phase space cannot take into account the possibility, for some problems, of a specific smart algorithm able to visit in a global way configurations far apart with respect to a natural distance.

The existence of an algorithm solving the problem is often linked

to the existence of symmetries (as for the XOR-SAT problem and for the Assignment problem). In this sense it is instructive to analyze the XOR-SAT problem (see [35], [38] and [7]). This problem admits a smart algorithm (Gauss algorithm) solving it, but has a transition of the same kind as 3-COL and 3-SAT.

Statistical Mechanics approach can be fruitful in order to design good heuristic algorithms based on some physically meaningful approximations. In this field two major applications have been achieved. One is the use of the so called *simulated annealing algorithms*. Since most of the "hard" problems we deal with in this chapter are frustrated, algorithms trying to minimize locally the energy often get trapped in a local minimum of the cost function, and the resulting configuration can have a cost much higher than the optimal one. A solution to this issue comes from the physical formulation: if we introduce a fictious temperature, allowing for thermal fluctuations, it is possible to avoid being stuck in local minima, exploring a larger part of the configuration space. When temperature is decreased, first the system get trapped in the region of the space of configurations containing the minimum, then, inside a valley, further cooling allows to find the ground state or a low-energy state. In case this procedure fails, a reheating can be performed, in order to select a new valley.

A second class of algorithms based on statistical mechanics approach is more recent and is inspired by the cavity method. This method allows to analyze on a statistical average but also for single instances of a given disordered system, and leads to self-consistency relations for the set of cavity fields, when the phase space is described by a RS-pattern, or for the surveys of the cavity fields, when the phase space is described by a 1-RSB pattern. A heuristic algorithm which uses cavity equations at zero temperature to find the solution of the associated satisfiability problem can be implemented.

Programmers have been using for long time heuristic message-passing algorithm, called Belief Propagation [48], which have been recognized to corespond to cavity equations in the RS approximation. Belief Propagation is highly performing on certain problems but not o some other. The physical interpetation of this fact is that Belief Propagation fails when the space of configurations contains many phases, and the Survey Propagation Algorithm is the proper generalization to 1-RSB landscapes. The first application of Survey Propagation has been to the 3-SAT problem [6].

4.3.1 The Random K-SAT

In an average-case analysis of the K-SAT problem a natural control parameter is the clause-to-variable ratio $\alpha = M/N$.

The *Random K-SAT* (as the *K-SAT*) problem results to be deeply different for $K = 1$, 2, 3.

The $K = 1$ case is almost surely UNSAT for every $\alpha > 0$, in fact the number of interactions per variable is a Poissonian variable with mean α, then for every finite value of α there is a finite fraction of variables with at least two interactions. The probability for such variables to be unsatisfiable is greater than 1/2. All the satisfiable variables with at least one interaction are then constrained to assume a given value. The $Poiss_\alpha(0)$ variables without interactions are free to be 0 or 1. Then the entropy of the zero temperature has an important contribution ($N\ln 2$) by the fully unconstrained variables. The other contribution comes from the unsatisfied variables.

The critical α for $K > 1$ is different from 0. Also the 2- and 3-SAT have a non-zero entropy for $\alpha = 0$. The number of solutions is exponential (in N) for any value of $\alpha < \alpha_c$.

The space of solutions is RS for 2-SAT: two solutions belong almost surely to the same cluster of Hamming diameter dN.

The space of solutions of 3-SAT is RS below a critical value $\alpha_{RSB} < \alpha_c$. When replica symmetry is broken the space of solutions breakes in an exponential number of clusters. In the thermodynamic limit the distance between configurations belonging to the same cluster decreases, while the distance between different clusters remains approximatively the same.

While passing through the critical value for α there is a phase transition behaviour in the number of *backbones variables*: fully constrained variables among all solutions of the instance. This transition is continuous for 2-SAT while it is discontinuous for 3-SAT. For this problem a relation among the nature of this transition and the behaviour of algorithms able to solve this problem has been shown.

If α is sufficently low, the problem is satisfied with probability going to 1 in the thermodynamic limit, while if it is sufficently high, the problem is unsatisfied with probability going to 1. It turns out that, in the thermodynamic limit, the satisfiability probability undergoes a sharp SAT-UNSAT phase transition: the probability $p_N(\alpha)$ that a random 3-SAT instance of size N and ratio α is satisfied, in the large limit is a step function: $\lim_{N\to\infty} p_N(\alpha) = \theta(\alpha_c - \alpha)$. with $\alpha_c = 4.267\ldots$

The SAT region presents a second transition point: for values of α below $\alpha_d = 3.921...$ the system contains an exponentially large number of solutions, and the phase space is described by a RS-pattern, while for values of α between α_d and α_c the exponential number of solutions is "clustered" into an exponential number of pure phases, this number vanishing however in the limit $\alpha \to \alpha_c$. The pure-phase clusterization pattern is well-described by a 1-RSB ansatz.

This statistical mechanics analysis has a practical counterpart. Heuristic algorithms which only use local informations undergo a dynamical transition at value $\alpha_d^{(\mathrm{alg})}$: beyond this value, they fail to retrive a solution. When there are many pure phases, these algorithms always get stuck on metastable valleys, so α_d is an upper bound for all the $\alpha_d^{(\mathrm{alg})}$, for algorithms in this class. On the other hand, when using Survey-Propagation inspired algorithms, the non local information propagated by the survey over pure phases, and by the parameter y (see section 5.4) which accounts for the reshuffling of the free energies, allows to retrieve a solution also for large sizes, and values for α near to α_c.

Chapter 5

Theory of Cavity Method

5.1 A trivial example: One Dimensional Spin Chains

Consider the following physical system: let σ_i be N spins on a chain with value in $\{-1, 1\}$. Given the Hamiltonian

$$H[\vec{\sigma}] = -\sum_{i=1,N-1} \sigma_i \sigma_{i+1} J_{i,i+1}$$

Given an instance (the set of $N-1$ values of the interactions) it is easy to find the ground-state, in fact it is possible to satisfy all the interactions (i.e. exists a configuration $\tilde{\sigma}$ of cost $H[\tilde{\sigma}] = \sum_{i=1,N-1} |J_{i,i+1}|$).

It is also easy to find the partition function for this system. Suppose to know the two partition functions obtained summing all the configurations for the spins from 1 to n with σ_n constrained to be ± 1 $H[\vec{\sigma}, \sigma_n = \pm 1] = \sum_{i=1,n-1} \sigma_i \sigma_{i+1} J_{i,i+1}$. Let the two partition functions be $\binom{f^n_{+1}(\beta)}{f^n_{-1}(\beta)}$. We can easily deduce the analogous partition functions for the system with $n+1$ spins, in fact the following equality holds:

$$\begin{pmatrix} f^{n+1}_{+1}(\beta) \\ f^{n+1}_{-1}(\beta) \end{pmatrix} = \begin{pmatrix} e^{-\beta J_{n,n+1}} & e^{\beta J_{n,n+1}} \\ e^{\beta J_{n,n+1}} & e^{-\beta J_{n,n+1}} \end{pmatrix} \cdot \begin{pmatrix} f^{n}_{+1}(\beta) \\ f^{n}_{-1}(\beta) \end{pmatrix}$$

The matrix used to pass from the n sites system to the $n+1$ sites system is called *transfer matrix* and were introduced in the context of ordered systems. For ordered systems the matrix is the same for any value of n then the partition function for the whole system is given by $\mathrm{Tr}(M_{b.c.} \cdot T^{n-1})$ the trace of the product of $n-1$ transfer matrices and

a matrix $M_{b.c.}$ keeping account of the specific boundary conditions. An effective field h_{eff} acting on the spin n-th in absence of the interaction with the $n+1$-th one can be defined. So that

$$\begin{pmatrix} f_{+1}^{n+1}(\beta) \\ f_{-1}^{n+1}(\beta) \end{pmatrix} \propto \begin{pmatrix} e^{-\beta h_{\text{eff}}} \\ e^{\beta h_{\text{eff}}} \end{pmatrix}$$

It is easy to check that the same approach is valid also for a Hamiltonian with external fields acting on the spin variables:

$$H[\vec{\sigma}] = -\sum_{i=1,N-1} \sigma_i \sigma_{i+1} J_{i,i+1} - \sum_{i=1,N} h_i \sigma_i$$

In this case the transfer matrix reads:

$$\begin{pmatrix} e^{-\beta(J_{n,n+1}+h_{n+1})} & e^{-(\beta-J_{n,n+1}-h_{n+1})} \\ e^{-\beta(-J_{n,n+1}+h_{n+1})} & e^{-(\beta J_{n,n+1}-h_{n+1})} \end{pmatrix}$$

It is easy to check that the effect of transfer matrix on the effective fields is equivalent to the application of an equation linking the effective fields (the *cavity fields*) on the n-th site and the cavity field on the $n+1$-th one. For the Ising model the so called cavity equation is:

$$h_{\text{eff}}^{n+1} = \sum \frac{1}{\beta} \text{atanh}[\tanh(\beta J_{n,n+1}) \tanh(\beta h_{\text{eff}}^{n})]$$

and its zero temperature version is:

$$h_{\text{eff}}^{n+1} = \sum \min(|J_{n,n+1}|, |h_{\text{eff}}^{n}|) \cdot \text{sign}(J_{n,n+1} \cdot h_{\text{eff}}^{n})$$

We call *cavity fields*, the effective external fields here introduced because they represent the external field experienced by the site n in the system where the interaction with the spin variable $n+1$ has been removed.

The system described in this section result to be so easy to study because of the fact that there are no loops, it is a general fact that on loopless graphs the transfer matrix approach here described for a one dimensional chain gives an exact (and fast) algorithm to find the partition function, this is due to the fact that every edge in the graph is a *bridge* and then, if removed the system factorizes.

5.2 Factor Graphs: a general framework

We want to deal with a class of statistical mechanics models sufficently large to include problems like the *Sherrington Kirkpatrick* spin glass and the *Random K-SAT* as prototype cases.

The variables are defined on a certain state space S, for example $S = \pm 1$ for spin systems, or $S = \{true, false\}$ for boolean problems. A configuration $\vec{\sigma}$ (that we often indicate shortly σ) is then an element of S^N.

The cost function can be represented as a global function of the whole configuration. Often the cost function can be expressed as a sum of factorized interactions: interactions involving only a subset of spins.

Each interaction a, involving k_a variables, corresponds to a real function $E_a : S^{k_a} \to \mathbb{R}$. There are also one-body terms, corresponding to real functions $W_i : S \to \mathbb{R}$ acting on a single variable and (for convenience) not labeled by interaction indices. One instance consists in the parameters defining these functions, clearly they have to be considered quenched. So the generic cost function can be expressed in the following way:

$$H(\sigma) = \sum_{a=1}^{M} E_a(\sigma_{i_1}, \dots, \sigma_{i_{k_a}}) + \sum_{i=1}^{N} W_i(\sigma_i)$$

For a statistical model on a lattice, the typical picture we have in mind is a regular lattice, in which variables are sitting on the vertices, and the pattern of interactions is quite simple: usually each interaction involves a fixed number of spins, and relate spins neighbouring on the lattice.

Dealing with disordered systems often need of more generic graphs and interactions that involve more than two bodies.

The graphical representation often used are the factor graphs: given a system of N variables and M interactions, we can represent it via a bipartite graph G, i.e. a graph in which we have $V(G)$ of two species $V_1(G)$ and $V_2(G)$ (say, circles and squares), and edges $E(G)$ only between vertices of different species

$$V(G) = V_1(G) \cup V_2(G) \qquad E(G) \subseteq V_1(G) \times V_2(G) \tag{5.1}$$

Associate to each variable a circle-vertex i (variable node) and to each interaction a square-vertex a (function node). Draw an edge between a circle and a square if the variable associated with the circle is involved in the interaction associated with the square. Often in factor graph's

graphical representations there are no square vertices corresponding to one-body interactions $W_i(\sigma_i)$, because they are implicitly represented as circle vertices.

Call $\delta(i)$ the neighbours of a variable node i and $\delta(a)$ the neighbours of a function node a. We denote $c(i)$, $c(a)$ their cardinalities. A factor graph inherits the natural notion of distance over graphs, in fact given two sites i and j in G we can define the distance $d(i,j)$ as the length of the shortest path on the graph going from i to j.

5.2.1 The Cavity Approach

The basic idea of Cavity Method is that we can understand the properties of a large system of the kind defined in appendix B by comparison with systems which only locally differ from the original one (cavity systems). These systems are chosen in order to almost decorrelate certain highly correlated variables close to the deformation, while minimally perturbing other observables, with support far away from the deformation (Cavity assumption).

We denote the set of neighbours of a given interaction a, but one given neighbour i, with $\delta(a)_i$ or $\mathcal{V}(a) \setminus i$ and similarly, the set of neighbours of a given variable i, but one given neighbour a, with $\delta(i)_a$ $(\mathcal{V}(i) \setminus a)$.

Given an edge a, i connecting a function a to the variable i we define a new Hamiltonian in a slightly different system. We disconnect i to a and introduce a new variable i_a to connect to a (so that $c(a)$ is not changed).

$$H_{i \nleftrightarrow a}(\sigma \cup \sigma_i) = \sum_i W_i(\sigma_i) + \sum_{a' \neq a} E'_a(\sigma|_{\delta(a')}) + E_a(\sigma_{i_a}, \sigma|_{\delta(a)_i})$$

This is the Hamiltonian for the cavity system where the effect of interaction a on variable i has been removed.

Remark that through iterated applications of this cavity transformation we can disconnect the graph, for example by disconneting a function from all its variables we obtain a factorized Hamiltonian (the disconnected function will be totally independent): its contribution to the partition function is trivially factorized.

Remark that the one-body interaction in the modified Hamiltonian is present only on the original variable i and do not need to be duplicated on the auxiliar variable i_a (this follows automatically dealing with one-body interactions as for the other interactions).

Given an operator O we denote with $\langle O \rangle_{a \nleftrightarrow i}$ the averages over cavity systems; with $\langle O \rangle_{/i}$ averages over systems where the i variable has been removed and with $\langle O \rangle_{/a}$ averages over systems where the a function has been removed. In particular F is the Free Energy of the cavity system, then:

$$F_{a \nleftrightarrow i} = -\frac{1}{\beta} \ln Z_{a \nleftrightarrow i} \qquad Z_{a \nleftrightarrow i} = \int \mathrm{d}\sigma e^{-\beta H_{a \nleftrightarrow i}}.$$

We can parametrize the distribution of probability over S for every single variable with an element h of an opportune space. For example for spin variable we can use the usual parametrization with $h \in [-\infty, \infty]$:

$$f_h(\sigma) = \frac{e^{-\beta h \sigma}}{2 \cosh(\beta h)}$$

so that a given value of h corresponds to a probability distribution.

We call magnetic field the element h_i such that the marginal probability distribution over site i in the original systems is reproduced.

We call cavity field the element $h_{i \to a}$ such to reproduce the marginal probability distribution over site i in the cavity system with the interaction i, a has been removed.

We call cavity bias the element $h_{a \to i}$ such that parametrizes the marginal probability distribution over site i_a in the cavity system where the interaction i, a has been removed.

As we said in section 2.4 we expect (inside a pure phase) the connected expectation values to decrease for points far away. In the following we will assume that for our system, correlation functions of operators far away inside a pure phase can be neglected in the large N limit.

Cavity ansatz: In the thermodynamic limit, inside a pure thermodynamic state, for the cavity systems in which we removed site i we have as a consequence of the cavity ansatz

$$p(\sigma, h_{i \to a}) \propto e^{-\beta W_i(\sigma)} \prod_{b \in \delta(i) \setminus a} p(\sigma, h_{b \to i}) \tag{5.2}$$

$$p(\sigma, u_{a \to i}) \propto e^{-\beta E_a(\sigma_i = \sigma, \sigma_j)} \prod_{j \in \delta(a) \setminus i} (\mathrm{d}\sigma_j \, p(\sigma_j, h_{j \to a})) \tag{5.3}$$

Remark that in the Cavity Approximation the cavity fields can be easily related to the difference between the Free energy in the original

and in the cavity system:

$$F - F_{a\leftrightarrow i} \simeq -\frac{1}{\beta}\ln\int \mathrm{d}\sigma p(\sigma, h_{i\to a})p(\sigma, u_{a\to i}) \tag{5.4}$$

$$F - F_{/i} \simeq -\frac{1}{\beta}\ln\int \mathrm{d}\sigma e^{-\beta W_i(\sigma)}\prod_{b\in\delta(i)} p(\sigma, u_{b\to i}) \tag{5.5}$$

$$F - F_{/a} \simeq -\frac{1}{\beta}\ln\int \prod_{j\in\delta(a)} (\mathrm{d}\sigma_j p(\sigma_j, h_{j\to a}))e^{-\beta E_a(\sigma_j)} \tag{5.6}$$

The equations 5.2 are expressed in full generality, for the two-state variables with the language of standard magnetic fields we have

$$h_{i\to a} = w_i + \sum_b u_{b\to i} \tag{5.7}$$

$$u_{a\to i} = -\frac{1}{2\beta}\ln\frac{\sum_{\sigma_j}\exp(-\beta E_a(+1,\sigma_j) - \beta\sum_j h_{j\to a}\sigma_j)}{\sum_{\sigma_j}\exp(-\beta E_a(-1,\sigma_j) - \beta\sum_j h_{j\to a}\sigma_j} \tag{5.8}$$

5.2.2 Ergodicity Breaking

The ergodicity breaking and the presence of many pure phases consists in a non-trivial decomposition of Gibbs measure into a sum of measures, each one being almost coincident with the Gibbs measure restricted on a subset of the phase space.

The subdivision of the phase space needs a notion of distance in this space and is temperature dependent, in fact every subset should be separated by the other valleys by not valicable walls (both entropically and energetically).

Say we have $\mathcal{N}$ pure phases labeled with an index $\alpha = 1, \ldots \mathcal{N}$. The pure-phase measures are defined via a partition of the phase space $\chi^{(\alpha)}(\sigma)$. So we define:

$$\mu_\alpha(\sigma) = \frac{1}{z_\alpha}\chi^\alpha(\sigma)e^{-\beta H(\sigma)} \qquad Z_\alpha = \int \mathrm{d}\sigma\chi^{(\alpha)}(\sigma)e^{-\beta H(\sigma)}$$

For each pure phase the free energy is $F_\alpha = -(1/\beta)\ln Z_\alpha$ and, for each observable O, the expectation in the α-th pure phase is

$$\langle O\rangle^{(\alpha)} = \int \mathrm{d}\mu_\alpha(\sigma)O(\sigma)$$

The expectation value of the operator in the full Gibbs measure (remember that it has a physical meaning only if the measure time is of an order of magnitude such that the system can visit different phases) is

$$\langle O \rangle = \sum_{\alpha} e^{-\beta(F-F^{\alpha})} \langle O \rangle^{(\alpha)}$$

We will work in the hypotesis that the Cluster Property holds in each pure phase and that, to any pure phase in the original system corresponds a phase in the cavity system.

We saw how in the Ising system a change in temperature induce the appearence of two phases. In disordered frustrated systems it can happens to have a number of pure phases exponentially large (in N the size of the system). Then it is possible to approximate the spectrum of free energies with a continuos distribution, analogously we introduce a complexity function $\Sigma(F)$ such that

$$e^{\Sigma(F)} \mathrm{d}F = \#\{\text{pure phases } \alpha : F^{\alpha} \in [F, F + \mathrm{d}F]\}$$

5.3 Some Examples

To make more clear the idea of cavity fields and how they can be used both to deduce analitycally new properties of some problems and both to design new algorithms we present here two examples.

5.3.1 Assignment

In the Assignment Problem, we choose a Factor Graph representation of the cost function written in variables n_{ij}. So, the variables are N^2, each for any pair (ij), the factors $\exp(-\beta\epsilon_{ij}n_{ij})$ are the one-variable factors $F^{(0)}$ we described above, while the $2N$ constraints (7.3) required in order to guarantee that n is a valid assignment are the clauses F_a. More precisely, if $\delta(k)$ is the "Kronecker" delta, such that for $k \in \mathbb{Z}$ gives 1 if $k = 0$ and 0 otherwise, we have

$$\exp\left(-\beta\mathcal{H}_{\epsilon}(n)\right) = \prod_{i,j=1}^{N} \exp(-\beta\epsilon_{ij}n_{ij}) \prod_{i=1}^{N} \delta\Big(1 - \sum_{j} n_{ij}\Big) \prod_{j=1}^{N} \delta\Big(1 - \sum_{i} n_{ij}\Big) . \tag{5.9}$$

So, the factor graph corresponding to the Assignment Problem is a "decoration" of the $\mathcal{K}_{N,N}$ we already said to represent the problem:

there is a square on each node of $\mathcal{K}_{N,N}$, while each edge has a circle "in the middle", thus of degree 2 (each n_{ij} enters a "row" and a "column" constraint). The constraints, conversely, have all degree N.

The Assignment Problem can be studied as a problem of Statistical Mechanics of Disordered Systems, in which one seeks for the zero-temperature limit of the Gibbs measure, which identifies the ground states. Given a set of variables as in the formulation of equation (7.2), and the Gibbs measure at finite temperature $T = 1/\beta$, i.e. $\mu_{\epsilon;\beta}(n) \propto \exp(-\beta\mathcal{H}_\epsilon(n))$, for an arbitrary function $\mathcal{O}(n)$ (called *observable*), one defines the *expectation value* of $\mathcal{O}$ just as

$$\langle\mathcal{O}\rangle_{\epsilon;\beta} = \frac{\sum_n \mathcal{O}(n)\exp(-\beta\mathcal{H}_\epsilon(n))}{\sum_n \exp(-\beta\mathcal{H}_\epsilon(n))}\,. \tag{5.10}$$

In particular, it is of interest to study the most elementary of these observables, i.e., for a pair (ij), the operator n_{ij}. Clearly one has that

$$\langle n_{ij}\rangle = \frac{\mathrm{prob}(n_{ij}=1)}{\mathrm{prob}(n_{ij}=0)+\mathrm{prob}(n_{ij}=1)} \in [0,1]\,,$$

so we can equivalently parametrize this value with a real parameter, called *local magnetic field*

$$\mathrm{prob}(n_{ij}) = \frac{e^{-\beta h_{ij} n_{ij}}}{1+e^{-\beta h_{ij}}}\,, \qquad \langle n_{ij}\rangle = \frac{e^{-\beta h_{ij}}}{1+e^{-\beta h_{ij}}}\,, \tag{5.11}$$

so that $h \ll -1/\beta$ means $\langle n\rangle \sim 1$ and $h \gg +1/\beta$ means $\langle n\rangle \sim 0$. If the instance is non-degenerate, there are some values of β sufficiently large such that the measure concentrates on the single groud state of the system, and thus we enter the "extreme" regime above, but the scaling in N of the value of β at which this occurs is quite strong.

In order to determine these parameters $\{h_{ij}\}_{i,j\in[N]}$, one can think of using what is called a *Replica-symmetric Cavity Approximation*.

The cavity fields are the fields which parametrize the marginal distribution of the variables n_{ij} in systems in which a certain little portion of the factor graph nearby node (ij) has been removed (the cavity).

It is customary to define two sets of fields, called *cavity fields* and *biases*. In the special case in which the interactions involve pairs of variables, as e.g. in an Ising Spin Glass, it is easy to solve one set w.r.t. the other. The same happens here, although for the "dual" reason, that variables enters only in pairs of interactions.

So, we indicate with $g_{i\to j}$ the bias which parametrizes the probability distribution for n_{ij} in the modified system where only the row-interaction on the i-th row is acting on this variable. The cavity field on site (ij) and "direct towards" column j is then given by $g_{i\to j}+\epsilon_{ij}$. Analogously $h_{j\to i}$ is the bias on n_{ij} in the modified system where only the interaction on the j-th column is acting on this variable. So, in the cavity approximation, one can hope that the combination $\epsilon_{ij}+g_{i\to j}+h_{j\to i}$ is a good approximation of the seeked value h_{ij}, where, if there is a single pure phase in the system, the corrections to this relation are expected to decrease with system size analogously to some observables on the "topology of the loops" on the graph.

The Cavity Equations are thus a set of self-consistent equations for the biases:

$$g_{i\to j}=\frac{1}{\beta}\log\sum_{j'\neq j}e^{-\beta(\epsilon_{ij'}+h_{j'\to i})}\,;\quad h_{j\to i}=\frac{1}{\beta}\log\sum_{i'\neq i}e^{-\beta(\epsilon_{i'j}+g_{i'\to j})}\,; \tag{5.12}$$

and, as we are interested in the ground states of the system, we can hope that it suffices to restrict to the zero-temperature version of these equations:

$$g_{i\to j}=\max_{j'\neq j}(-\epsilon_{ij'}-h_{j'\to i})\,;\qquad h_{j\to i}=\max_{i'\neq i}(-\epsilon_{i'j}-g_{i'\to j})\,. \tag{5.13}$$

These equations raise in general many issues: first, in many cases we do not know how to state rigorously that there is a single thermodynamic phase; then, even if this is the case, and one expects that the cavity fields do satisfy the cavity equations, we don't know if there are other "spurious" solutions of these equations; finally, even if this does not happen, we should also device a reasonable technique for finding the solution of our system of non-linear equations in many variables and with no symmetries. A last point is that, even if everything works fine up to here, if one is interested in the zero-temperature limit, maybe some prescription is required in the order in which the $\beta\to\infty$ limit and the "thermodynamic" $N\to\infty$ limit are performed.

For our problem we get (see chapter 7) strong answers to many of these questions, by the analysis of the "Cavity algorithm" induced by the equations. Indeed, often a solution of this kind of equations is obtained by noticing that they correspond to the fixed-point condition of a certain recursive map. Then, the method will work if one determines that the

solution is an attractive point w.r.t. this map, and finds a suitable initial condition, in the bassin of attraction of this point.

Our equations have special characteristics, coming from the mix of linearity with "max" operations, that we discuss here. We consider in all our chapter the so-called *parallel updating*, i.e. we have a dynamics on a discrete time, and thus every field, e.g. $g_{i\to j}$, is raised to a function $g^{(t)}_{i\to j}$, which is updated through

$$g^{(t)}_{i\to j} = \max_{j'\neq j}(-\epsilon_{ij'} - h^{(t)}_{j'\to i})\,; \quad h^{(t+1)}_{j\to i} = \max_{i'\neq i}(-\epsilon_{i'j} - g^{(t)}_{i'\to j})\,. \tag{5.14}$$

Under parallel updating, the fields in the sets $\{g^{(t)}_{i\to j}\}_{j\in[N]}$, $\{h^{(t)}_{j\to i}\}_{i\in[N]}$ are all equal, except for at most a single smaller value, because of the action of the max. Indeed, for example, for a given value i, the set of values $\{-\epsilon_{ij} - h^{(t)}_{j\to i}\}_{j\in[N]}$ will have either two or more indices realizing the maximum (in this case, call j^* and j^{**} two of them), or one site realizing the maximum (call it j^*), and one or more realizing the second-maximum (call j^{**} one of them). In both cases, the new fields are

$$g^{(t)}_{i\to j} = \begin{cases} -\epsilon_{ij^{**}} - h^{(t)}_{j^{**}\to i} & j = j^* \\ -\epsilon_{ij^*} - h^{(t)}_{j^*\to i} & j \neq j^* \end{cases} \tag{5.15}$$

(For this reason, we will adopt the notation $\mathrm{2nd}_i(x_i)$ and $\text{arg-2nd}_i(x_i)$, besides $\max_i(x_i)$ and $\text{arg-max}_i(x_i)$.) A high cavity field on a site means that the site probably has occupancy 0 ($n_{ij} = 0$) in the optimal matching, on the contrary a low cavity field on a site means that the site should have occupancy 1. So, in a non-degenerate instance, near to an "extreme" zero-temperature limit, when the measure concentrates on a single configuration, we expect a single negative field per row and per column, identifying the matching, this being qualitatively consistent with the update above (we will be more precise later on).

Note however that such a behaviour is *not* what results from an analytic approach to the cavity equations, or, in other words, an analysis of the update equations which assumes time stationarity (and symmetry under exchange of rows with columns). Furthermore, in this approach it is also used the genuinely cavity idea that the distribution of a cavity field cannot be sensitive to the parameters of the instance *downstream*, while this is not the case in our "extreme zero limit" argument above, where the strongly negative field is aware of the fact that the ϵ_{ij} entry immediately downstream is so good that it is viable for the optimal

matching. Conversely, the cavity idea would imply that, as $N-1$ fields out of N are achieved from the "max" function over the whole set of N indices j, and only one takes the "second max", because of the $j' \neq j$ constraint, and as this fraction of entries is not correlated with the choice of the relevant entries at the following time step, we can neglect the infinitesimal fraction and write the distributional equations

$$x \stackrel{d}{=} \max_i(-\xi_i - x_i) \qquad\qquad x' \stackrel{d}{=} \operatorname{2nd}_i(-\xi_i - x_i) \tag{5.16}$$

where x_i are i.i.d. variables distributed according to the unknown distribution $f(x)$, ξ_i is the Poisson Process that corresponds to a measure probability on the entries of the cost matrix with i.i.d. variables distributed with a probability distribution continue in 0 with value 1. It is easy to find that

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{4(\cosh(x/2))^2} \tag{5.17}$$

and it is possible to find a general formula for the k^{th} highest value in each row (or column) for the process $(-\xi_i - x_i)$, where ξ_i are a P.P.P. and x_i are reshuffling distributed with the Logistic distribution above. This formula is

$$f_k(x) = f(x)\frac{\left(\ln(1+e^{-x})\right)^{k-1}}{(k-1)!} \tag{5.18}$$

and so x' is distributed according to this formula with $k=2$.

$$f_2(x) = f(x)\ln(1+e^{-x})\,. \tag{5.19}$$

Further analysis of the distributional equation for this random ensemble allow to determine the expectation over the instances of the minimum cost. Indeed the entries of the optimal matching are distributed, up to the scaling of N, as the entries ξ_i, with i realizing the maximum in the distributional equation 5.16, so, using the distribution for the x_i's and some properties of the Poisson Process, one easily gets that

$$\mathrm{d}\mu(\xi) = \theta(\xi)\frac{e^{-\xi}(e^{-\xi}-1+\xi)}{(1-e^{-\xi})^2}$$

which is indeed normalized. The first moment gives directly the average cost of the optimal matching for random instances, as the scaling factor N simplifies with the fact that we have N summands, and gives:

$$\langle \min_n(H_\epsilon(n))\rangle_\epsilon = \int \mathrm{d}\mu(\xi)\cdot\xi = \frac{\pi^2}{6}$$

This is the result conjectured by Parisi and later proved in [41]

5.3.2 K-SAT

Many computational problems are known to be *NP-complete* through a *polynomial reduction* onto the *K-Satisfiability* (SAT) problem which is the first problem shown to be *NP-Complete*.

We are mainly interested in the *Random K-SAT*. Consider N variables x_i with $i = 1, \ldots, N$ (each boolean variable could be represented by a spin σ_i with values in $\{\pm 1\}$). In the random version of the SAT problem there are M clauses randomly chosen: a clause is the logical OR of K randomly chosen variables, each of them randomly negated or not (with equal probabilities). The problem consists of the logical AND of the M clauses. The specific instance is said to be solvable if there exists a logical assignment of the N boolean variables. It is unsatisfiable otherwise.

Let $M = \alpha \cdot N$. Numerical experiments showed that (for $K \geq 2$) the probability for a random instance to be satisfiable tend to be (for large N) a step function (it is one when $\alpha < \alpha_c^{(K)}$ and it is zero when $\alpha < \alpha_c^{(K)}$). It has been observed that hard radom instances are generated for values of α close to the critical value.

The cavity equations for the K-SAT are:

$$h_{j \to a} = \sum_{b \in V(j) \setminus a} J_{bj} u_{b \to j}$$

where $h_{j \to a}$ is the cavity field on the variable j in absence of the interaction with the clause a; J_{bj} is -1 if j is negated in the clause b; $u_{b \to j}$ is the cavity bias of the interaction b on the variable j. If there is at least one variable $j \in V(a) \setminus i$ such that $h_{j \to a} J_j^a \leq 0$ then $u_{a_i} = 0$ otherwise $u_{a \to i} = 1$. Then, more formally:

$$u_{a \to i} = \prod_{j \in V(a) \setminus i} \theta \left(h_{j \to a} J_j^a \right)$$

It is possible to interpret these equations in the following way: suppose that we want to determine the effect of a clause (interaction) b on one of its variables. If there is a variable j satisfying the interaction b then the interaction will not ask to the variable i to satisfy it, otherwise ($u \neq 0$) the interaction b will ask to be satysfied by biasing the variable i. The

effect of all the clauses on the variable i is:

$$h_i = \sum_{b \in V(i)} J_{bj} u_{b \to j}$$

The difficult task to solve the cavity equations can be approached in several different ways. (Remark that there is always a trivial solution with all fields and biases equal 0). One way is to use the equations as recursive equations: one can use some arbitrary (not all null) starting fields and biases, then one have to perform some iterated updates of the fields hoping to reach a self-consistent solution.

Remark that this algorithm not always converge. If one is interested in finding a solution of the specific problem one strategy is to fix the variables with strong fields and try to solve the so decimated problem.

One important remark that should be clear is that this algorithm always converge for tree-like problems (problems such that the structure of the factor graph is a tree).

The belief propagation is a message passing algorithm consisting in the propagation of distributions of the fields: let $Q_{a \to i}(u) = (1 - \eta_{a \to i})\delta(u) + \eta_{a \to i}\delta(u-1)$. Where $\eta_{a \to i}$ is the probability for $u_{a \to i}$ to be 1. For these probability the following equalities hold:

$$\eta_{a \to i} = \sum_u Q(u) \prod_{j \in V(a) \setminus i} \theta \left(\sum_{b \in V_a} u_{b \to j} \right)$$

This way to search for solutions of the cavity equations (as distribution of probability) is suitable to keep account of the intra-cluster entropy.

When a large number of cluster of solutions exists, a message passing algorithm can find incongruences, in fact in two regions far apart messages inherent to different clusters often diffuses and this lead to contradictions when the messages "meet".

To keep correctly account of the possible existence of several clusters we need to do a step forward: we need to implement a message passing of distributions of distributions of fields.

A simple equivalent way to keep account of the existence of several clusters is to introduce a new state for the variables. This new state has as effect to enlarge the space of warnings and to allow to transmit messages belonging to different clusters.

5.4 Survey of Cavity Equations over many states

All the equations introduced in 5.2 are valid in each pure phase. In particular if cluster property holds then we can write some equations (the so called cavity equations) that relates the cavity fields. So that any pure phase is characterized by a set of numbers: the Free energy and $|E|$ cavity fields and $|E|$ cavity bias ($|E|$ is the number of edges in the factor graph, remark that for the Ising spin glass they are $2 \cdot |V_1|$: there are two fields per each interaction). Every pure phases works as if there were not the other phase. Pay attention to the following fact: the free energy $F^{\alpha}_{a\to i}$ in the cavity system is a function of the free energy in the original system F^{α}, of cavity bias on the variable i and of cavity fields on a, thus the set of free energies F^{α}, if ordered for magnitude, will be reshuffled respect to the initial order after an application of the cavity equations. In other words, the phase with smaller free energy $F^{\alpha}_{i\to a}$ in a given cavity system is not sure to be the smaller also for free energy in the original system.

To keep control of the possible reshuffling of energies we now makes some hypotesis:

- The complexity function for the free energy is convex and such that there is a parameter y selecting a narrow interval of free energies relevant for the statistical description of the system (F^*), such that
$$\frac{\partial}{\partial F}\sum_{/a}(F)|_{F^*} \simeq \frac{\partial}{\partial F}\sum_{/i}(F)|_{F^*} \simeq y$$

- We neglect the dependence of the distributions of cavity fields and bias on F

- We consider the distributions of the cavity field (bias) on a function (variable) as if they were almost factorized. Then

$$Prob(h_{i\to a}) \propto \int \prod_{b\in\delta(i)_a} Prob_{b\to i}\mathrm{d}u_{b\to i}) \; e^{-y\Delta F(\{u_{b\to i}\})\delta(h_{i\to a},h_{i\to a}(u_{b\to i}))} \quad (5.20)$$

$$Prob(u_{a\to i}) \propto \int \prod_{j\in\delta(a)_i} Prob_{j\to a}\mathrm{d}h_{j\to a}) \; e^{-y\Delta F(\{h_{j\to a}\})\delta(u_{a\to i},u_{a\to i}(h_{j\to a}))} \quad (5.21)$$

So, as we are not able to deal with all the infinite number of pure phases, we renounce to calculate the infinite $h^{\alpha}_{i\to a}$ associated to each link i, a and work on their probability distribution. It can be calculated (in our hypotesis) starting from the set of neighbouring cavity bias $\{u_{b\to i}\}$, we then calculate (by integrating) the probability that the cavity bias are such that $h_{i,a}$ as a function of the bias has values $h_{i,a}$. We need to keep account of the reshuffling of the energy occurring when we apply the cavity equations. It is possible that our system with a given cavity field gain free energy when reinserting the removed edge i, a, so that, forgetting the reshuffling, we would underestimate its probability.

Part II

Three Optimization Problems

Chapter 6

Bounds on the Ground State Energy of Ising Spin Glasses

Ising Spin Glasses on random graphs have been often used as a playground for the study of cavity methods both for optimization problems and statistical mechanics. See for example [33] [32].

Spin glasses present an interesting behaviour also at the mean field level (as seen in section 3.2). One of the main defects of the mean field model is the lack of notion of neighbourhood: in the mean field model all the spins are neighbouring. Random graphs (see appendix B) have the advantage to have a concept of locality, and locally a tree-like structure.

The problem of finding the ground state of a spin glass is an NP-hard problem, so if we are interested to know its energy for large instances we should discard the idea to find it by an exact algorithm. For this reason approximated algorithms solving this problem have been introduced and here we present a way to calculate (in polynomial time) a lower-bound in the case of an Ising Spin Glass.

6.1 The Model

The Ising Spin Glass, introduced in section 3.2 has Hamiltonian:

$$H(\vec{\sigma}) = -\sum_{i \neq j} J_{i,j}\sigma_i\sigma_j - h_{\text{ext}} \sum_{i=1}^{N} \sigma_i \tag{6.1}$$

Let α be the average coordination, let the interactions $J_{i,j}$ be i.i.d. random variables distributed according to:

$$\text{prob}(J_{i,j}) = \left(1 - \frac{\alpha}{N}\right) \cdot \delta(J_{i,j}) + \frac{\alpha}{N}\frac{1}{2}\chi_{[-1,1]}(J_{i,j}) \tag{6.2}$$

the spin variables σ_i are in $\{-1, 1\}$, the first summation in (6.1) runs over all the pairs of different spins (remark that per each spin a non zero contribution comes only from an average number α of neighbours). The interaction variables $J_{i,j}$ as given by (6.2) can be obtained also by drawing, for every edge in a random graph of average coordination α some $J_{i,j}$ i.i.d. from a random distribution: $\chi_{[-1,1]}(J)$, where for $b > a$ $\chi_{[a,b]}(x) = \theta(b - x) \cdot \theta(x - a)$.

Since we are interested in the ground state (GS) we consider only the zero temperature properties of the system. Since we are interested in the thermodynamic limit we consider the average energy per spin: $H(\vec{\sigma}_{GS})/N$. At zero temperature the system lies on the ground state. In the thermodynamic limit we expect its ground state energy per site to converge to a constant.

Given an instance (a graph and a set J of interactions) the energy of the ground state is bigger than

$$E_T = -\sum_{\langle i,j \rangle} |J_{i,j}| - h_{\text{ext}} \cdot N \tag{6.3}$$

then, averaging over all the possible instances, using the probability distribution in (6.2) for the interactions J, we obtain: $\overline{E_T} = -\frac{\alpha}{2} \cdot \overline{|J|} - h = -\frac{\alpha}{4} - h$.

Similarly we can deduce an easy upper bound on the average energy of the ground-state: consider the configuration with all the spins up: $E_U = -\sum_{\langle i,j \rangle} J_{i,j} - h$ and then its average $\overline{E_U} = -h$.

The problem we will try to solve is, for a fixed instance, to give better bounds on the energy of the ground state of the system. We will do it by introducing two cavity-based algorithms based on cavity method.

The strategy we will use for the upper-bounds is similar to the bound given above: the energy of a wathever given spin configuration is always an upper-bound of the energy. Then our efforts are concentrated on finding a "good" configuration.

To find a "good" configuration (with energy close to the ground-state one) we implemented an algorithm that uses the informations that can be deduced by the cavity fields. Other strategies are possible, see for example Extremal Optimization algorithms [5].

Improving the lower-bound in (6.3) is harder. Lower bounds for disordered systems with symmetric distribution functions of the random disorder were introduced in [25], remark that for the system in (6.1) the lower-bound in (6.3) is recovered. An algorithmic way to calculate the lower-bound is by semidefinite programming, see [19]. Here we introduce a way that uses cavity methods as a basic ingredient.

6.2 Cavity Equations

The cavity equations can be easily found:

$$h_{i\to j} = \sum_{j'\in V(i)\setminus j} u_{j'i\to i} \tag{6.4}$$

$$u_{ij\to j} = \frac{1}{\beta}\text{arctanh}\left[\tanh(\beta J_{ij})\cdot\tanh(\beta h_{j\to ij})\right] \tag{6.5}$$

We will use these equations both on specific instances and on populations of cavity fields to deduce thermodynamic observables.

6.2.1 Zero-Temperature Cavity Equations

First of all remark that the zero temperature cavity field on a given spin i is one half the difference between the ground state's energy when the spin i is constrained to be up and down in the system where the interaction $J_{i,j}=0$:

$$h_{i\to j} = \frac{E_{GS}(\sigma_i=+1, J_{i,j}=0) - E_{GS}(\sigma_i=-1, J_{i,j}=0)}{2}$$

By keeping in mind the definition above of zero temperature cavity field it is easy to see on a tree graph that the consistence rule is:

$$h_{i\to j} = \sum_{j'\in V(i)\setminus j} \text{sign}(J_{j',i}, h_{j'\to i})\cdot\min(J_{j',i}, h_{j'\to i}) \tag{6.6}$$

The cavity bias is:

$$u_{i\to j} = \text{sign}(J_{j',i}, h_{j'\to i})\cdot\min(J_{j',i}, h_{j'\to i}) \tag{6.7}$$

It's easy to check that the zero temperature cavity equations above are obtained as the $T\to 0$ limit of 6.4

Cavity fields are the messages that variables "send" to interactions, so no cavity field is associated to a spin variable with connectivity zero: the only field on a zero connectivity spin is the external field. Cavity fields "live" on the edges of the graph and on each edge there are two cavity fields.

If site i has connectivity one then the cavity field $h_{i\to j} = h_{\text{ext}}$ is equal to the external field on i. The cavity bias on j is given by equation (6.7). The simplicity of interactions allows us to write the consistence equation for the cavity equations without need to calculate biases.

6.2.2 Site Addition

The energy of the system for large sizes can be calculated as the difference between the energy of systems of slightly different sizes:

$$E := \lim_{N\to\infty} \frac{H_N(\{\sigma\}_{GS})}{N} = \lim_{N\to\infty} H_{N+1} - H_N$$

We can estimate this energy by calculating the energy in a system before and after the addition of a spin. The extra spin has to be chosen with coordination distributed according to the Poissonian distribution (as we are dealing with Erdős Rényi random graphs), however if we perform this operation on a random graph several times we introduce too many edges. The contribution to the energy site addition due to the extra edges addition has to be subtracted:

$$E := \Delta E_s - \frac{\alpha}{2}\Delta E_l$$

The increment in the energy due to the addition of a spin σ_0 connected to k spins in the hypothesis of independence between its neighbours is given by:

$$\Delta E_s = -|u_0| - \sum_{i=1}^{k} \max(|h_{i\to 0}|, |J_{i,0}|) + \sum_{i=1}^{k} |h_{i\to 0}|$$

where u_0 is the external field on the spin 0. The formula above is exact if the k spins are mutually disconnected before the addition or if the energies of the 2^k ground states obtained constraining the values on the spins $\sigma_{i=1,\dots,k}$ are given by $Const + \sum_{i=1}^{k} \sigma_i \cdot h_i$.

Remark that when we want to calculate average quantities we need to add spins according to the appropriate distribution: both the old

system and the new one should be extracted with the weight we are interested in. For example in our case the cavity fields come from sites of coordination $k+1$ with k being a Poissonian random variable of average α. For example Cayley graph (where all the sites have the same connectivity k) the site has to be added taking care of keeping the right connectivity on all the spins.

The energy corresponding to a link addition is the difference between the ground state energy of the whole system with or without the edge. In the hypothesis of independence of the spin variables i and j the link contribution on the global ground state energy with spins i, j constrained is

$$\Delta E_l = |J_l| - \theta(-J_l \cdot \sigma_i \sigma_j) \cdot \min(J_l, \sigma_i, \sigma_j)$$

Given an edge $\{i, j\}$, the probability distribution that we need to use to draw the coordination α of the two endpoints is $k \cdot Prob(k)$. In the random graph case, thanks to the fact that $k \cdot Poiss_\alpha(k) = \alpha \cdot Poiss_\alpha(k-1)$ we can extract two vertices each with i.i.d. poissonian coordination and then connect them with the edge i, j.

6.3 Population dynamics

A useful tool to find the probability distribution for the cavity fields and to perform calculations are population dynamics algorithms.

If some cavity fields on the edges of a given system solve the cavity equations then this set of fields is stable under application of the cavity equations. If moreover the incoming fields on each site are independent (as happens on tree graphs) then these fields are distributed according to a law stable under iteration of the cavity equations. A discrete approximation of the distribution of probability over the cavity fields consists in considering a finite set of fields, each one allowed to assume a real value.

The *population* is a set of cavity fields, we ask to the population to be stable under random iteration of the cavity equations. Let $P(h)$ be their probability distribution, then the autoconsistency condition consists of the fact that by applying the cavity equations (6.6) to a random subset of fields of our population the new cavity field obtained is distributed according to $P(h)$.

This autoconsistency test for the probability distribution $P(h)$ can be turned in a method to find it: the cavity equations can be used to

implement a dynamic on the population.

For example as dynamic we randomly choose a subset of fields and use the cavity equation to find a new cavity field, as if we picked up randomly one edge of an actual system and calculate one of its two fields. The new field is used to update the distribution: we choose randomly one of the elements of the population and we substitute its old field with the new one. In the scheme of RS theory the population consists of a population of fields, anyway we could also consider other kind of populations, for example populations of distributions as in 1RSB-analysis.

The autoconsistency equation for the cavity fields of a spin glass on random graphs is:

$$P(h) = \sum_{k=0}^{\infty} e^{-c} \frac{c^k}{k!} \delta(x - h_{\text{ext}}) * (P'(x))^{*k} \tag{6.8}$$

where $*$ denotes the convolution and $f(x)^{*n}$ means that the function $f(x)$ has to be convoluted $n-1$ times with itself. P' is not the first derivative of P but the following function:

$$P'(x) := \frac{P(x) + P(-x)}{2} \cdot |x| + \chi_{[-1,1]}(x) \int_{|x|}^{\infty} \mathrm{d}x' P(x')$$

In the RS assumption the system is described by a distribution $P(h)$ that satisfies equation (6.8). So, a randomly chosen edge has two independent fields, each one distributed according to (6.8). The RS assumption consists in keeping in account of the existence of only one pure phase.

We compared the distribution which satisfies equation (6.8) (unique unless $h_{\text{ext}} = 0$) with the distributions actually obtained when running a message-passing algorithm on a given large-size instance. The distributions result to be the same.

Given the distribution of probabilities of the cavity fields we calculated the energy as a result of site and link addition as described in section 6.2.2.

The energy found in this way is plotted in the figures of pages 66 and 67 as "Population RS".

The results (for graphs with coordination up to 4) are in good agreement with the energy of the configuration found by decimation with a message-passing algorithm.

6.3.1 Stability of the RS solution

We know that the RS solution could be not able to describe completely the system for low values of the external field: we can test its goodness by testing the stability of the RS solution as a solution of the 1RSB equations.

In the 1RSB scheme on each site there is not simply a field, but a distribution of probabilities of fields and then the whole system is described no longer by a distribution but by a distribution of distributions. In this scheme the RS solution consists of a distribution of delta distributions.

So a set $h_{i\to j}$ of fields corresponds to a set $f_{i\to j}(x) = \delta(x - h_{i\to j})$ of distributions. In the context of 1RSB the equations (6.6) (6.7) (6.4) for the cavity fields have to be interpreted as equations for the distributions of cavity fields.

We tested the stability of the delta-like distributions in different numerical ways: the simpler way is by assuming the existence of some perturbations small enough to admit the linearization of the propagation equation for the cavity field. This method does not require the knowledge of the 1RSB distribution and gives no information about it.

We assume to have a perturbed RS solution consisting of a set of distributions: $f_{i\to j}(x) = (1-\epsilon)\delta(x - h_{i\to j}) + \epsilon_{i\to j}\tilde{f}_{i\to j}(x)$ where $\tilde{f}_i$ is a probability distribution. Then the equations linearized in ϵ_i are:

$$\begin{aligned} h_{i\to j} &= (1 - \sum_{j'\neq j} \epsilon_{j'\to i}) \cdot \delta(h_{i\to j} - \sum_{j'\neq j} u_{j'i}) + \\ &\quad + \sum_{j'\neq j} \epsilon_{j'\to i} \left[\delta(h_{i\to j} - \sum_{j''\neq j,j'} u_{j''\to i}) * f_{j'\to i}(u_{j'\to i}) \right] \\ u_{i\to j} &= \operatorname{sign}(J_{ij} \cdot h_{j\to i}) \cdot \min(J_{ij}, h_{j\to i}) \end{aligned} \tag{6.9}$$

When the outgoing perturbation is on average bigger than the ingoing one, the RS solution is not stable.

We tested numerically the correctness of the results so found also using the finite temperature cavity equations and then performing the zero temperature limit.

The curve obtained in this way has been plotted in figure 6.2.
We also analyzed the behaviour of two different sets of random fields following the same evolution laws: we compared the ovelap of the two distributions after the iterated application of the cavity fields on the

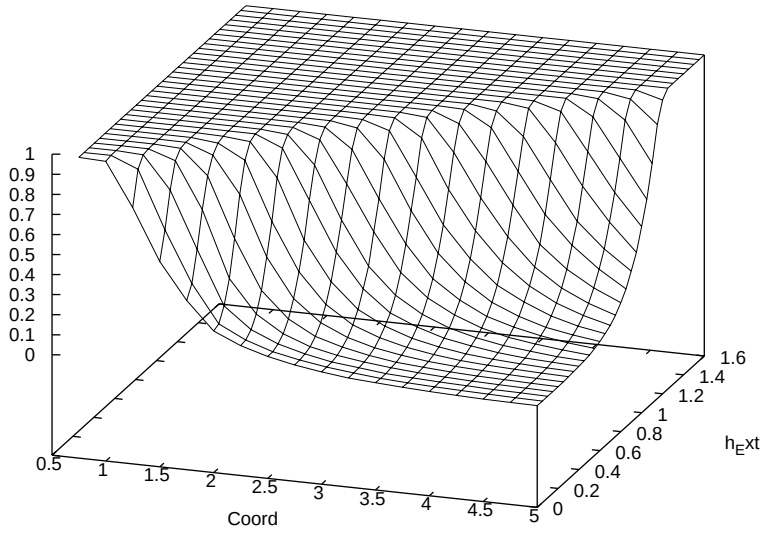


Figure 6.1: Overlap between two populations

same elements of the two different populations. The overlap is plotted in figure 6.1. The corresponding deduced instability curve has been plotted in figure 6.2.

6.3.2 Replica Symmetry Breaking

It is possible also to keep account numerically of the presence of several phases also by using many fields for each site. The different fields associated to each site represents different phases.

This method introduced in general in chapter 5 consists in the application of the cavity equations to the cavity fields coming from some different phases. Then each cavity field so calculated has to be represented in the new element of the population with a weight related to the free energy induced by the addition of the site: $\exp(-\mu\Delta F)$. The parameter μ is the so called *Parisi parameter* that has to be chosen so to select the relevant region of the space of pure phases.

With such a dynamics (for a more detailed description see [33] and [32]) a population of sets of cavity fields is obtained. These cavity fields

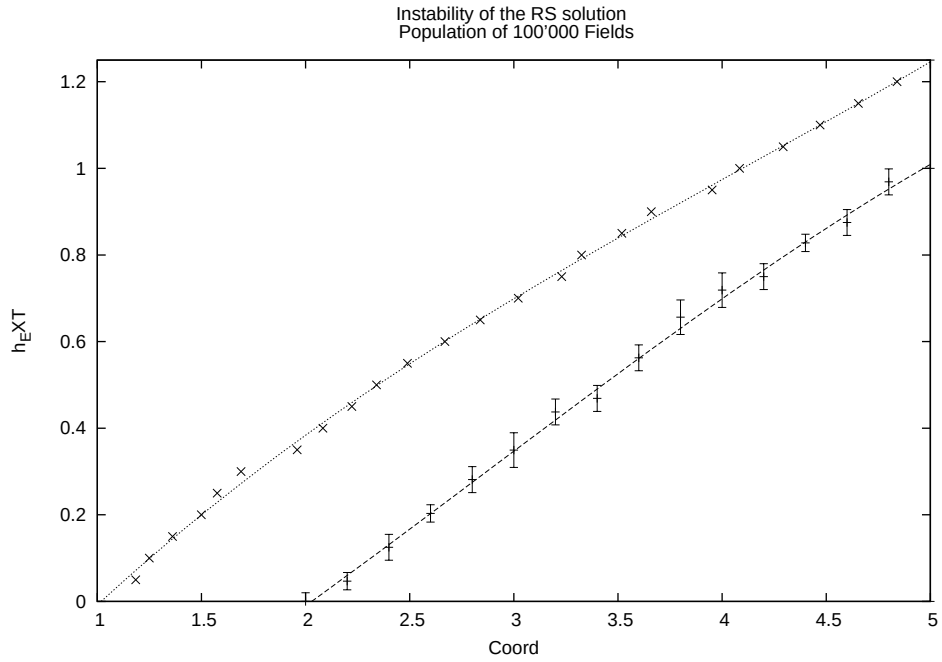


Figure 6.2: Local and global instability lines

can be used to calculate physical observables.

6.4 Upper bound by BP-Decimation

To find a good solution to the *Max-Cut* problem (that is a low cost configuration for the Ising Spin Glass, see 3.2.4) we used a message passing algorithm starting with some randomly drawn cavity fields. Then we updated sequencially all the fields several times. When working at low values of the external field usually the fields do not converge to a solution of the cavity equations: not all the fields of the system are stable under iteration of the equations.

Then we fix the more biased sites among the ones with stable values for the cavity field. So to obtain a smaller system. The energy for the configuration found in this way is an upper-bound for the energy of the ground-state of the system.

In general, for the Ising Spin Glass the energy of a configuration found by belief propagation is a good approximation of the actual ground

state energy. For this reason we did not try to improve this calculation.

In the figures of pages 66 and 67 we plotted the upper bounds given by this algorithm.

6.5 Lower bound

The algorithm we introduced to find a lower bound of the ground-state energy works by reducing the Ising Spin model to a smaller one. A basic step consists in the deletion of an edge. The basic step is iteratively applied so to reduce the problem to another consisting only of isolated spin variables with some auxiliar fields. In a sense this could be seen as an algorithmic version of the approach in [17] to obtain lower bounds to the free energy.

The energy of the ground state is the minimum over all possible configurations $\{\vec{\sigma}\}$ of $H(\vec{\sigma})$. Using the fact that

$$\min_x(f(x)) = \min_x(f(x)+g(x)-g(x)) \geq \min_x(f(x)+g(x)) + \min_x(-g(x)) \tag{6.10}$$

Fix an edge i, j, then consider the following inequality, obtained from the one above.

$$\begin{aligned} E_{GS}(H(\vec{\sigma})) \quad \geq \quad & E_{GS}(H(\vec{\sigma}) - h_1\sigma_i - h_2\sigma_j - J_{ij}\sigma_i\sigma_j) + \\ & + \min(h_1\sigma_i + h_2\sigma_j + J_{ij}\sigma_i\sigma_j) \end{aligned} \tag{6.11}$$

The system is reduced to a "cavity" system without the edge i, j. We keep account of the energy of the deleted edge with the second summand. Remark that the inequality (6.10) is tight if the function $f(x) + g(x)$ is constant. Then in (6.11) the inequality is tight only if the auxiliar (and arbitrary) fields h_1, h_2 are such that the ground state energy of the new system is constant (as function of σ_1 and σ_2). This is not always possible, anyway the fields that make the minimizing expression as small as possible are the cavity fields ($h_{i\to j}$,$h_{j\to i}$).

In general a function of two spin variables can be always written by using four constants. The energy of the ground state of a spin system at given fixed values for two spins i, j in general is a function:

$$E_{GS}(\sigma_i, \sigma_j) = E_0 + h_1\sigma_i + h_2\sigma_j + \Delta_{ij}\sigma_i\sigma_j \tag{6.12}$$

the inequality in (6.11) is surely tight only if $\Delta_{ij} = 0$ and we are using the appropriate values for the auxiliar fields.

Our algorithm to calculate the lower bound of the Ising Spin Glass iteratively searches a solution in terms of cavity fields for the cavity equations, then delete one edge by using equation (6.11). At each step the algorithm returns a slightly modified system and a real value, representing the energy contributions of the deleted edge.

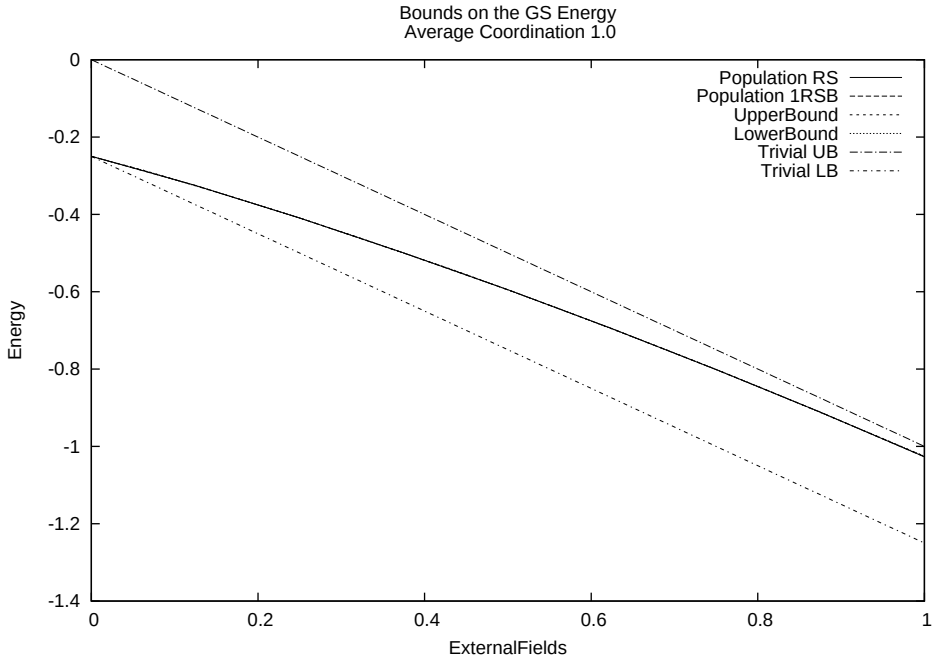
The main limits on the efficence of this algorithm lie in the fact that it is not always possible to find a solution to the cavity equations and in the fact that often the quantity Δ_{ij} in equation (6.12) is non-zero.

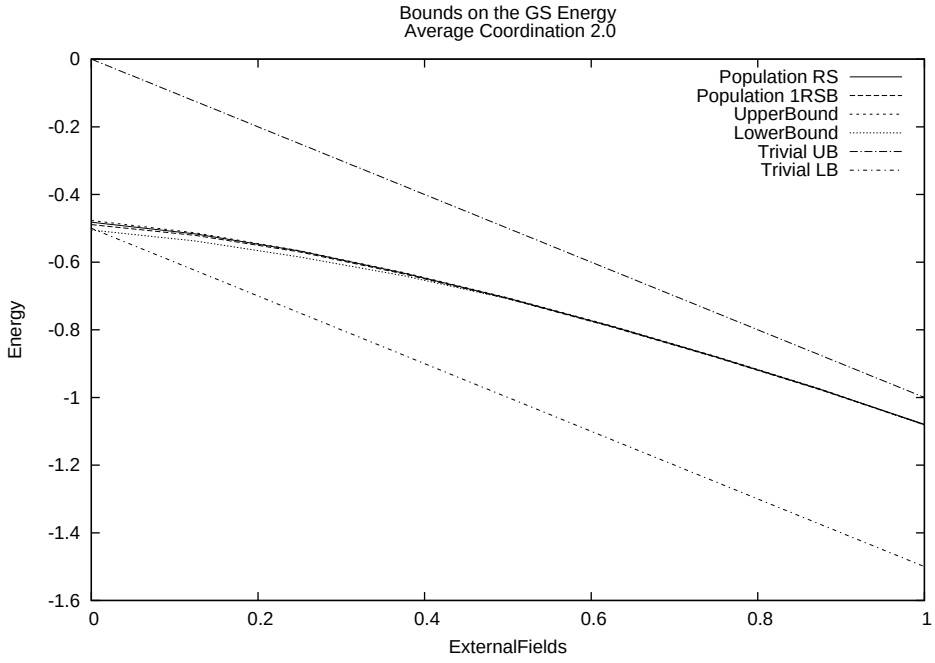
We expect Δ_{ij} to be small if the system is tree-like (near the edge we are going to remove). It is easier to find a solution for the cavity equations if there is a strong external field biasing a large fraction of the spin variables. In the case of strong external field we observe that the lower and the upper bounds are close, while they start to separate when the external field is lowered in correspondence of a value bigger then the one corresponding to the instability of the replica symmetric solution.

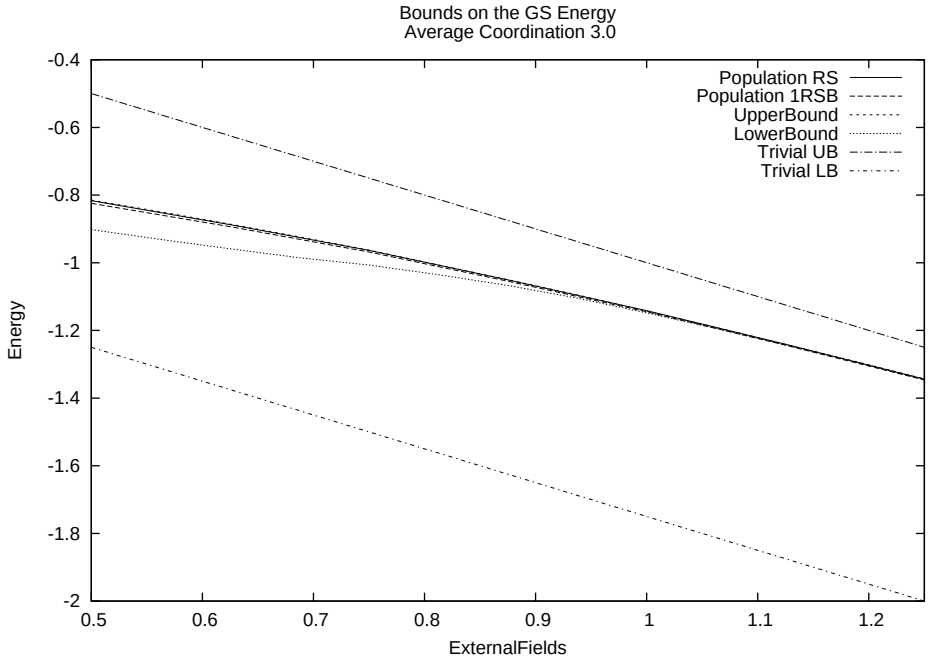
In figures 6.3, 6.4, 6.5, 6.6, 6.7 it is possible to see that there are relatively strong finite sizes corrections: as the size $\to \infty$ the difference between the upper and the lower bound seems to go to a given value. The limit curve is compatible with the constant zero only on random graphs of average coordination around 1.0, in accord with the natural prediction.

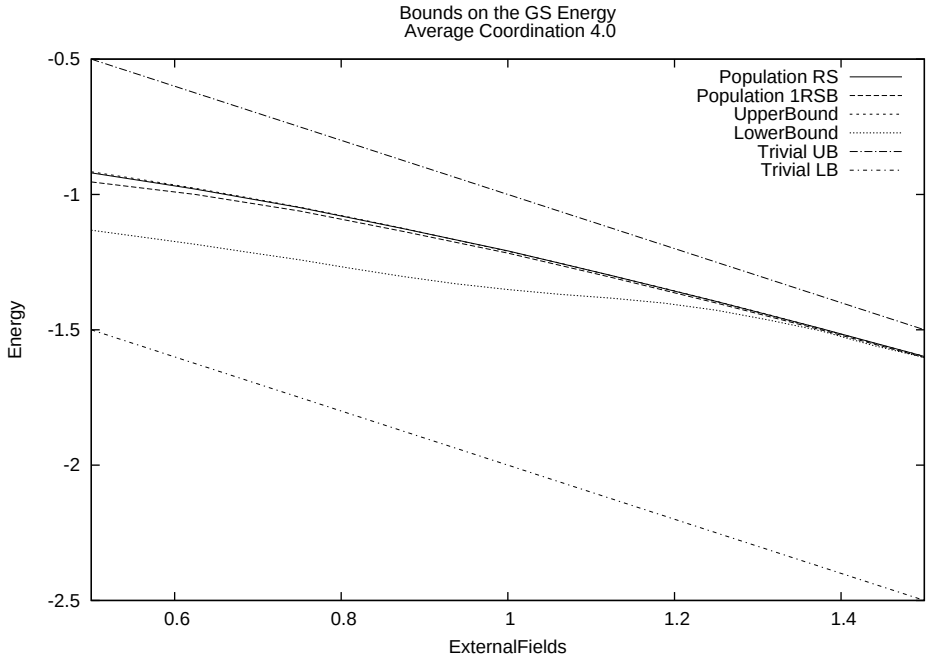
For random graphs of coordination $\alpha = 1$ the limit difference of upper and lower bound is null. This fact is linked to the fact that for these graphs the RS solution is stable. Discrepancies between the bounds arise when we consider finite size systems where there is not full independence between two sites in the cavity system.

The number of edges removed altough the cavity equations did not find convergenceon their value is plotted in the upperside of figures 6.3 ... 6.7. It is possible to see that for $\alpha > 1$ there is a threshold value of the external field such that the fraction of not-converged edges removed is one. This threshold value (to be determined via a finite size scaling analysis) is probably in some relation with the instability of the replica symmetric solution.


Bounds on the GS Energy
Average Coordination 1.0
Population RS
Population 1RSB
UpperBound
LowerBound
Trivial UB
Trivial LB
Energy
ExternalFields
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
0
0.2
0.4
0.6
0.8
1


Bounds on the GS Energy
Average Coordination 2.0
Population RS
Population 1RSB
UpperBound
LowerBound
Trivial UB
Trivial LB
Energy
ExternalFields
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
0
0.2
0.4
0.6
0.8
1


Bounds on the GS Energy
Average Coordination 3.0
Population RS
Population 1RSB
UpperBound
LowerBound
Trivial UB
Trivial LB
Energy
ExternalFields
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
-1.8
-2
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2


Bounds on the GS Energy
Average Coordination 4.0
Population RS
Population 1RSB
UpperBound
LowerBound
Trivial UB
Trivial LB
Energy
ExternalFields
-0.5
-1
-1.5
-2
-2.5
0.6
0.8
1
1.2
1.4

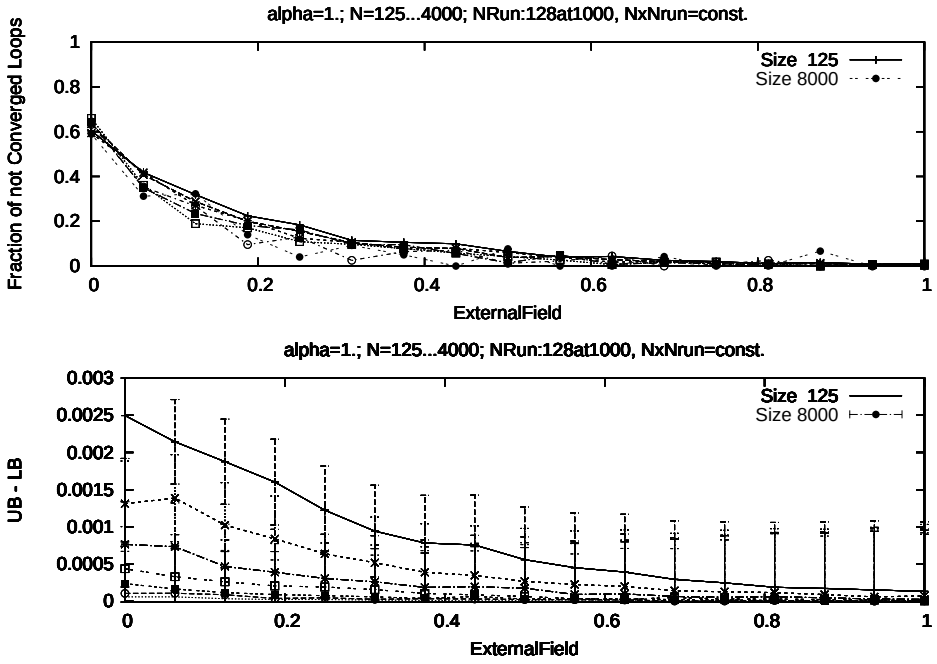


Figure 6.3: Coordination 1.0 (critical $h_{EXT} = 0$)

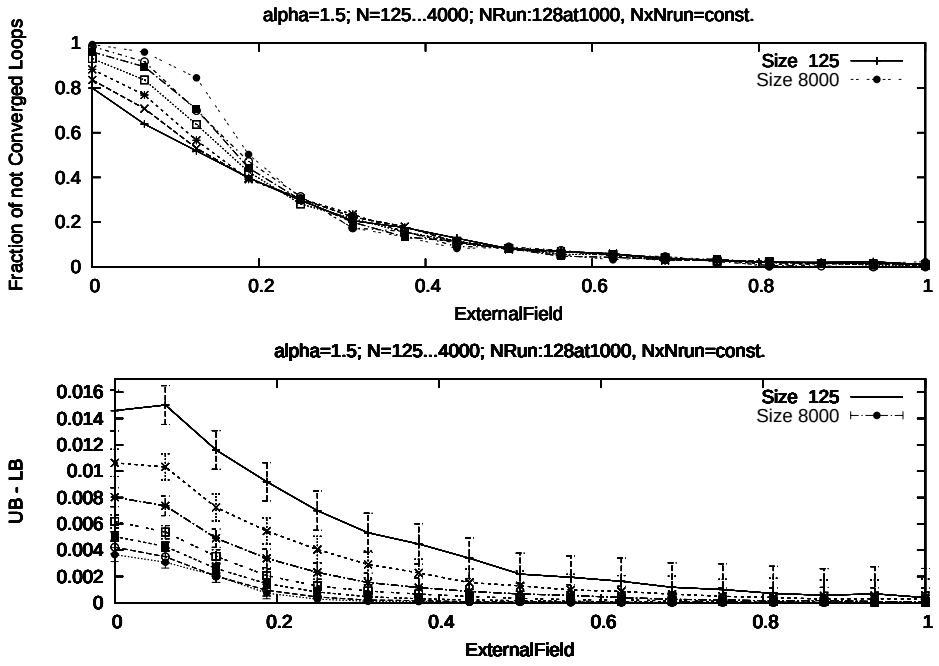


Figure 6.4: Average Coordination 1.5

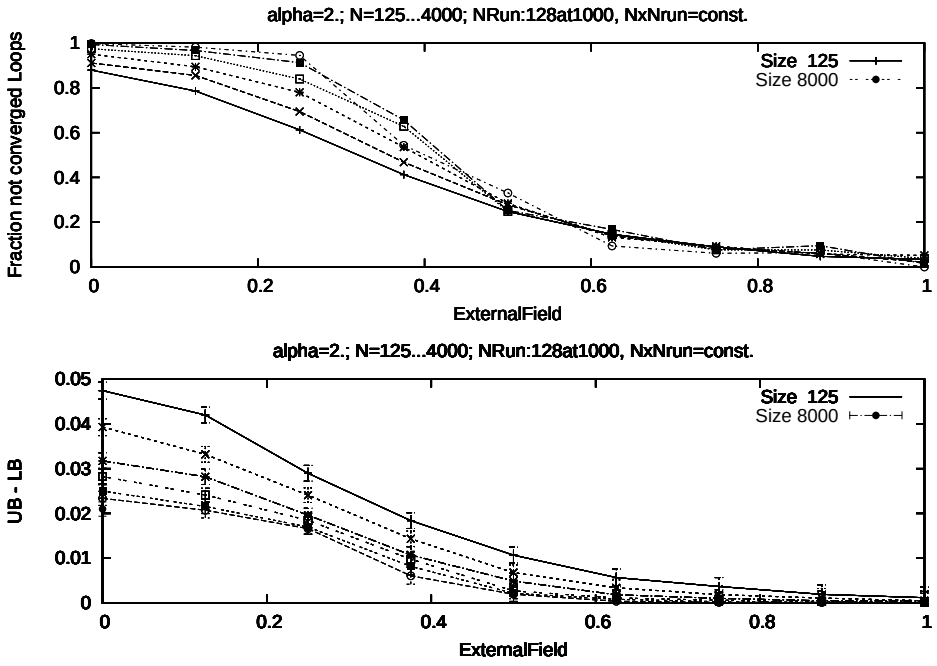


Figure 6.5: Average Coordination 2.0

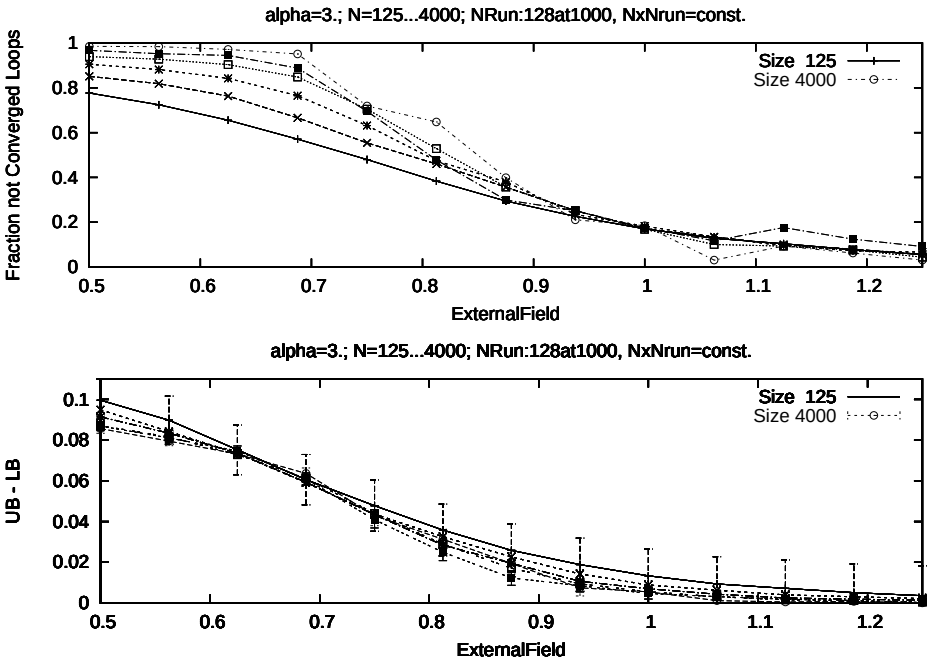


Figure 6.6: Average Coordination 3.0

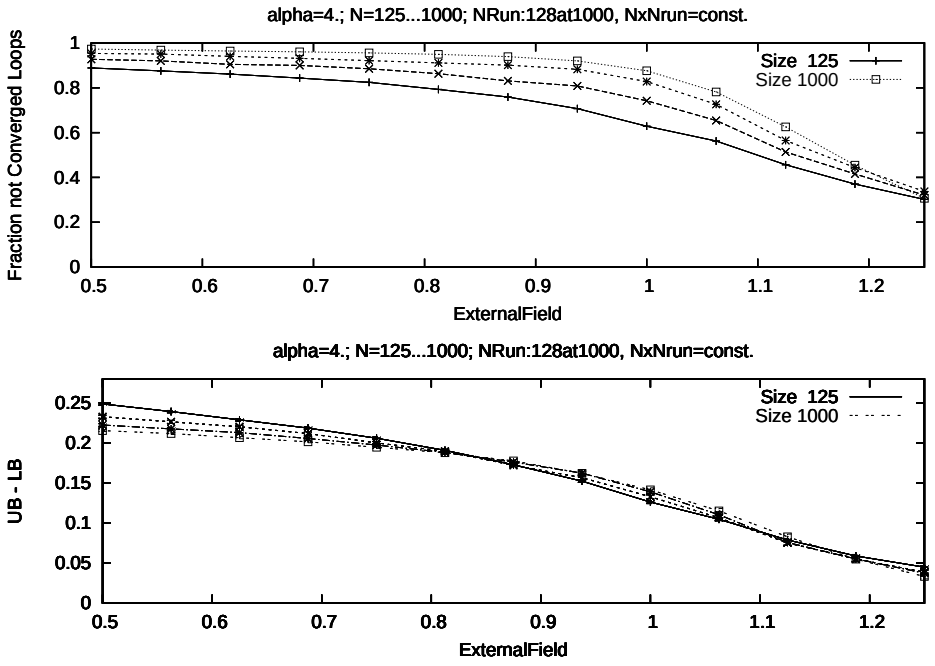


Figure 6.7: Average Coordination 4.0

Chapter 7

The Assignment problem

There are many "pictorial" representations of the Assignment Problem. For example, one can consider the problem of assigning N jobs to N machines, one per each, given a set of costs $\{\epsilon_{ij}\}$ for executing the i-th job on the j-th machine, where the goal is to minimize the sum of all the N costs.

So, a valid *assignment* consists in a one-to-one mapping of jobs onto machines, i.e. a permutation π in the symmetric group $\mathcal{S}_N$, and the cost of π is thus encoded in the cost function (or *Hamiltonian*)

$$H_\epsilon(\pi) = \sum_i \epsilon_{i\,\pi(i)}\,. \tag{7.1}$$

One can give a graphical representation of this problem. Given $\mathcal{K}_{N,N}$, the complete bipartite graph of order N, one can identify the two sets of N vertices, V_r and V_c, as the "jobs" and the "machines", and naturally assign weights ϵ_{ij} to the edges (ij) with $i \in V_r$ and $j \in V_c$. Then, a valid assignment consists of a *matching* M on the graph, i.e. a subset of the edge set $E \equiv V_r \times V_c$ such that each vertex has degree one. The weight of the matching is the sum of the weights on the occupied edges.

Another useful representation in terms of permutations is more "algebraic". Let $[N] \equiv \{1, \ldots, N\}$. One can encode a permutation π through a $N \times N$ matrix n_{ij} valued on $\{0, 1\}$, such that $n_{i\,\pi(i)} = 1$ and $n_{ij} = 0$ for $j \neq \pi(i)$. In this reformulation, the cost function above is written as

$$H_\epsilon(n) = \sum_{i,j} \epsilon_{ij}\, n_{ij} = \mathrm{tr}(\epsilon^T n)\,, \tag{7.2}$$

where the constraint that n corresponds to a permutation is encoded in

the $2N$ constraints

$$\sum_{i=1}^{N} n_{ij} = 1 \quad \forall\, j \in [N]\,; \qquad \sum_{j=1}^{N} n_{ij} = 1 \quad \forall\, i \in [N]\,. \tag{7.3}$$

We have chosen to use sub- and superscripts r and c for "row" and "column" quantities, referred to this formulation (and, pictorially, a row is a job and a column is a machine). Indeed, we index both the jobs and the machines with integers in $[N]$, as natural in the algebraic formulation, instead of what could have been natural in the graph formulation, e.g. jobs labeled from 1 to N and machines from $N+1$ to $2N$.

If the instance ϵ has a single configuration π realizing the minimum, we say that it is *non-degenerate*, and conversely, if there are two or more permutations with optimal cost, we say that the instance is degenerate.

Remark that the difference of cost among two configurations π and π' is a non-trivial linear combination, with coefficients in $\{0, \pm 1\}$, of the costs ϵ_{ij}. So, in any random i.i.d. ensemble of real positive costs drawn from a non-singular distribution, an instance is non-degenerate with probability 1 in Lesbesgue measure.

7.1 Gauge Invariance

7.1.1 Alternated paths and cycles

Given the complete graph $\mathcal{K}_{N,N}$, with edge costs $\epsilon_{i,j}$, and a matching M on it, we say that a path γ is *alternating* on M if one every two edges along their "path" ordering is in M. An analoug definition can be given for a even cycle γ.

We define the alternate cost of the path (cycle) γ as:

$$E^{\pm}(\gamma; M) := -\sum_{(ij)\in\gamma\cap M} \epsilon_{i,j} + \sum_{(ij)\in\gamma\setminus M} \epsilon_{i,j} \tag{7.4}$$

if γ is an alternating cycle $E^{\pm}(\gamma; M)$ is the extra cost of the matching $M' := M\delta\gamma$, in fact $H_\epsilon(M') = H_\epsilon(M) + E^{\pm}(\gamma; M)$.

Conversely for any two matchings M, M' their symmetric difference $M\Delta M'$ is a collection of cycles$(\cup_{\alpha=1}^{k}\, \gamma_\alpha)$, moreover:

$$H_\epsilon(M') = H_\epsilon(M) + \sum_{\alpha=1}^{k} E^{\pm}(\gamma_\alpha; M) \tag{7.5}$$

These facts have many consequences, remark for example that the E optimality gives that $E^{\pm}(\gamma; E) \geq 0$. Here and in the following we consider non degenerate instances, so the inequality above is strict.

For any cicle *alternating* on the optimal matching E let

$$\Delta(\gamma_i) := \frac{E^{\pm}(\gamma_i; E)}{||\gamma_i||} \tag{7.6}$$

consider now the matching $M' \neq E$ that minimizes the following quantity

$$\frac{H_\epsilon(M') - H_\epsilon(E)}{||E \Delta M'||} \tag{7.7}$$

it is not hard to prove that it differs from E by a single *alternating* cycle γ^*, which minimizes the quantity in 7.6.

We will refer to γ^* as the optimal *alternating* cycle; we also define γ^{**} as the cycle alternating on E and disjoint from γ that minimizes 7.6

7.1.2 Gauge Invariance

It is easy to check that there is a $2N$-parameter family of transformations Φ for the set of costs ϵ which leaves the problem unchanged. We call such a transformation a *gauge transformation*. Given two real-valued vectors $\{\lambda_i\}_{i\in[N]}$ and $\{\mu_j\}_{j\in[N]}$, the new costs $\epsilon' = \Phi_{\vec{\lambda},\vec{\mu}}(\epsilon)$ are defined as

$$\epsilon'_{ij} = \epsilon_{ij} - \lambda_i - \mu_j \tag{7.8}$$

and are such that

$$H_{\epsilon'}(\pi) = H_\epsilon(\pi) + h_0\,; \qquad h_0 = \sum_i \lambda_i + \sum_j \mu_j\,; \tag{7.9}$$

i.e. every feasible solution π has the same cost in the old and in the new problem, up to a shift of h_0, independent from π and thus irrelevant at the aim of finding the optimal assignment.

Remark that one global parameter has a trivial effect, as do not even affect the single entries of ϵ': $\Phi_{\vec{\lambda}+a,\vec{\mu}-a} = \Phi_{\vec{\lambda},\vec{\mu}}$. One can fix this trivial mode, for example, by choosing $\mu_1 = 0$, or, in a more abstract way, one can think to the corresponding $(2N-1)$-dimensional quotient space.

Also remark that a number of combinations are gauge-independent, among which the alternated cost of a cycle as defined in equation 7.4, as each gauge parameter either does not enter in this linear combination, or it enters with coefficient $+1$ on a single summand, and -1 on a single

other summand. Of course, the gauge-independence of $E^{\pm}(\gamma; M)$ is also a consequence of the relation $E^{\pm}(\gamma; M) = H_\epsilon(M\delta\gamma) - H_\epsilon(M)$, and of the fact that the new costs are unchanged up to a shift overall.

The space of gauged matrices ϵ' (or, more shortly, of *gauges*), accessible from a given ϵ is in bijection with the pairs of vectors (λ, μ) with the quotient above, and inherits from this the topology and the metric of $\mathbb{R}^{2N-1}$, so it makes sense to say that a subset of gauges is *connected*, or *compact*, or *convex*, or *a polytope*.

We call a gauge *proper* if every element ϵ'_{ij} is non-negative. Given some proper gauge ϵ', call $Z \subseteq \mathcal{K}_{N,N}$ the spanning subgraph of the complete bipartite graph whose edges (ij) are the ones such that $\epsilon'_{ij} = 0$.

On a proper gauge, h_0 is a trivial lower bound to the cost of the optimal assignment, and, if one could find an assigment π such that $\epsilon'_{i\,\pi(i)} = 0$ for all i (i.e. totally contained in Z), this would also certificate that π is an optimal assignment for the original instance, with cost exactly h_0. We call a proper gauge *Hungarian* if the corresponding Z contains a matching of cardinality N.

All the Hungarian gauges of a single instance ϵ have a set Z which contains a certain $Z_{\min}$, union of all the matchings M which realize the optimal cost (if more than one).

We call a gauge *proper and non-trivial* if every row and every column of ϵ' contain at least one zero (i.e. if no vertex is isolated in Z). Clearly, this is a necessary but not sufficient condition for being Hungarian, and a proper non-trivial gauge is easily found for every instance, e.g. by composing first the gauge with $\mu_j = 0$ and $\lambda_i = \min_j(\epsilon_{ij})$, and then applying to the resulting gauge ϵ' the analogous transformation with $\lambda_i = 0$ and $\mu_j = \min_i(\epsilon'_{ij})$. We call Φ^{trivial} the map above, which acts as a projector from the space of proper gauges to the subspace of proper non-trivial gauges, and increases h_0 by the amount $\sum_j \min_j(\epsilon_{ij}) + \sum_i \min_i(\epsilon'_{ij}) \geq 0$.

A few remarks are in order. First of all, while gauge transformations Φ are clearly a group (isomorphic to $\mathbb{R}^{2N}$ with vector sum), the restriction to proper gauges breaks this structure, and leaves with a convex polytope of $\mathbb{R}^{2N}$. Furthermore, the set of Hungarian gauges is a connected convex polytope, subset of the boundary of this set. It is of dimension at most N, and exactly N if the optimal matching is unique. If one considers the quotient w.r.t. the global gauge transformation (e.g. by keeping $\mu_1 = 0$), both the set of non-trivial proper gauges and of Hungarian gauges are compact.

So, the space of Hungarian gauges is relatively large. Contrarily to

what could have been argued, there does not exist a single Hungarian gauge, and not even a "canonical" one. Both the Hungarian Algorithm reviewed in the next section and our algorithm discussed in this chapter are deterministic, and halt through a Hungarian gauge, but their output gauges are in general different from each other, and different if one interchanges the role of rows and columns in the algorithms. An interesting fact, that we prove in 7.6, is that for every instance ϵ and every index $j \in [N]$ there exists a Hungarian gauge, deterministically described in terms of ϵ, such that any other column has at least two zeroes (exactly two, and exactly one zero on column j, if the instance is non-degenerate).

Finally, a flavour of the hardness of the problem and of the strength of the Hungarian Algorithm is given by the fact that, even if one knows in advance the optimal solution π, it is not easy to find a gauge transformation on ϵ producing a ϵ' Hungarian, or any other certificate of optimality for π (without using the algorithm itself).

7.1.3 Technical details on statements about gauge structure

Here we prove the statements that have been given above without proof.

Convexity of the set of proper gauges is easily checked. Given $(\vec{\lambda}, \vec{\mu})$ and $(\vec{\lambda}', \vec{\mu}')$ proper,

$$\epsilon_{ij} - \lambda_i - \mu_j \geq 0 \qquad \forall\ i, j\,, \qquad\qquad \epsilon_{ij} - \lambda'_i - \mu'_j \geq 0 \qquad \forall\ i, j\,.$$

If we take a combination $a(\vec{\lambda}, \vec{\mu}) + (1-a)(\vec{\lambda}', \vec{\mu}')$, with $a \in [0, 1]$, by combining the two equations above we get

$$\epsilon_{ij} - (a\lambda_i + (1-a)\lambda'_i) - (a\mu_j + (1-a)\mu'_j) \geq 0 \qquad \forall\ i, j\,,$$

as was to be proven. Then, the set is a polytope because the boundary is given by a set of linear inequalities.

The proof for the set of Hungarian gauges is identical, just with inequalities replaced by equalities for pairs $(ij) \in Z_{\min}$. It is important, at this aim, the fact that $Z_{\min}$ is intrinsic to ϵ.

Now we prove the compactness of the set of Hungarian gauges. We can assume without loss of generality that the optimal assignment is the identity permutation $\pi(i) = i$, so we must have

$$\lambda_i + \mu_j \leq \epsilon_{ij} \qquad \forall\ i, j\,; \qquad\qquad \lambda_i + \mu_i = \epsilon_{ii} \qquad \forall\ i\,.$$

So, using the fixing $\mu_1 = 0$, on one side we get $\lambda_1 = \epsilon_{11}$, and thus $\mu_j \leq \epsilon_{1j} - \epsilon_{11}$, on the other side we get $\lambda_i \leq \epsilon_{i1}$, and thus $\mu_i = \epsilon_{ii} - \lambda_i \geq \epsilon_{ii} - \epsilon_{i1}$. As each μ is bounded on both sides (and, through $\lambda_i + \mu_i = \epsilon_{ii}$, also each λ), we have that (the quotient of) the set of Hungarian gauges is defined by a set of linear inequalities with $\geq$, and is contained in an interval, so it is compact. If non-empty, its dimension is $2N - |Z_{\min}|$, i.e. N in the non-degenerate case ($N - 1$ in the quotient).

Compactness for the set of non-trivial proper gauges through analogous reasonings is proven, although a bit more involved. Of course,

$$\lambda_i \leq \epsilon_{i1}$$

from the gauge fixing $\mu_1 = 0$. Then, for each column j there exists at least one index $i^*(j)$ such that $\epsilon'_{i^*(j)j} = 0$. Take, say, the lowest one. In particular, for $j = 1$ this gives that $\lambda_{i^*(1)} = \epsilon_{i^*(1)1}$. So, for each $j > 1$ we get

$$\mu_j \leq \epsilon_{i^*(1)j} - \lambda_{i^*(1)} = \epsilon_{i^*(1)j} - \epsilon_{i^*(1)1} \leq \max_i (\epsilon_{ij} - \epsilon_{i1}) \,,$$

even if we do not know who is $i^*(1)$. On the other side we have

$$\mu_j = \epsilon_{i^*(j)j} - \lambda_{i^*(j)} \geq \min_i (\epsilon_{ij} - \lambda_i) \geq \min_i (\epsilon_{ij} - \epsilon_{i1}) \,. \tag{7.10}$$

Also for each row i there exists at least an index $j^*(i)$ such that $\epsilon'_{ij^*(i)} = 0$, so

$$\lambda_i = \epsilon_{ij^*(i)} - \mu_{j^*(i)} \geq \epsilon_{ij^*(i)} - \max_{i'} (\epsilon_{i'j^*(i)} - \epsilon_{i'1}) \geq \min_{i',j} (\epsilon_{ij} - \epsilon_{i'j} + \epsilon_{i'1}) \,.$$

where we used 7.10 for $j = j^*(i)$ in the first inequality, while the second one has been obtained by relaxing the restriction $j = j^*(i)$. So both the λ's and the μ's are bounded on both sides. As again the inequalities are not strict, then compactness follows.

Chains of ϵ's with alternated signs, and consecutive items having in common alternatively the row- and the column-index, as in the expression $\epsilon_{ij} - \epsilon_{i'j} + \epsilon_{i'1}$ above, will appear many times in our "cavity" arguments, and will be a *leitmotif* of our proofs. In this case, remarkably, finite-length chains suffice to prove our statement in generality, so this proof is also a kind of introduction to our methods.

7.2 The Hungarian Algorithm

A classical algorithm for the Assignment Problem which finds an optimal matching in worst-case polynomial time [1] is due to H. Kuhn [28], who called it "Hungarian Algorithm" as a tribute to the mathematicians authors of the two main lemmas on which is based, Kőnig and Egerváry.

Most of the material in this section can be found in [29]. We state here only the main facts, and without proof, except for the points where, at our advice, a restatement of the proof in the language of gauge transformation (see section 7.1) allows to better understand the connections between Hungarian Algorithm and Cavity Method techniques.

7.2.1 Kőnig and Egerváry theorems

As we said previously, the basic idea is to perform a series of gauge transformations, up to get a gauge with a graph of zeroes Z which has a matching M on it of cardinality N. Kőnig theorem provides a useful equivalent condition for the existence of such a matching.

Consider a graph G edge-subgraph of the complete bipartite graph $\mathcal{K}_{n,n}$, for $X \subseteq V_r$, let $\mathcal{V}(X)$ denote the subset of V_c of vertices having a neighbour in X in the graph Z. The difference of the two cardinalities, $d(X) := |X| - |\mathcal{V}(X)|$, is called the *deficit number* of X. Then the theorem states that:

Theorem 1 (Kőnig 1916) *In a bipartite graph $G = (V_r, V_c; E)$ the minimum number of unmatched elements in V_r (or V_c) over all the possible matchings is equal to the maximum over all the subsets X of the deficit number $d(X)$. In particular, a perfect matching is possible if and only if $|\mathcal{V}(X)| \geq |X|$ for all $X \subset V_r$.*

The Egerváry theorem states, in the language of section 7.1, as follows:

Theorem 2 (Egerváry 1931) *The cost of the optimal assignment is equal to the maximum value of $h_0 = \sum_i \lambda_i + \sum_j \mu_j$ a proper gauge can have. So, a proper gauge that realizes the maximum h_0 in the whole set of proper gauges has the property that its graph of zeroes Z admits a matching. Such a gauge always exists.*

[1] $\mathcal{O}(n^3)$ after the work of Munkres [40] for speeding up the recovering procedure, as reported in Knuth [26].

The first part of the theorem is proven, e.g. in [29], constructively through Kőnig theorem, while existence follows then easily from compactness of the set of non-trivial gauges. Of course, Egerváry theorem is also both a corollary of the stronger structure theorem that we prove in 7.6, or of the analysis of the Hungarian Algorithm as in [29].

We already know that the set of non-trivial proper gauges is non-empty, as $\Phi^{\text{trivial}}(\epsilon)$ always exists, and that, if $\epsilon'' = \Phi^{\text{trivial}}(\epsilon') \neq \epsilon'$, one has $h_0(\epsilon'') > h_0(\epsilon')$. Then, given any non-trivial proper gauge ϵ' such that $Z(\epsilon')$ does not admit a matching, by Kőnig theorem there exists $X \subseteq V_r$ with positive deficit. A further gauge $\Phi_{\vec{\lambda},\vec{\mu}}$ with parameters

$$\lambda_i = \begin{cases} \delta & i \in X \\ 0 & i \notin X \end{cases} \quad \mu_j = \begin{cases} -\delta & j \in \mathcal{V}(X) \\ 0 & j \notin \mathcal{V}(X) \end{cases} \quad \delta = \min_{\substack{i \in X; \\ j \in V_c \setminus \mathcal{V}(X);}} (\epsilon'_{ij}) \quad (7.11)$$

applied to ϵ' gives a new proper gauge. But $\delta > 0$ by definitions of $\mathcal{V}(X)$, and the variation of h_0 is exactly $\delta \cdot d(X)$, so it is positive (notice that any intermediate value $0 < \delta' \leq \delta$ would suffice at this purpose).

Then, the gauge obtained so far can be projected back to the space of non-trivial gauges by Φ^{trivial}, and also to the gauge-fixing $\mu_1 = 0$. The first step can only further increase h_0, and the second one does not change its value, so h_0 has increased of a positive amount. This proves that non-trivial proper gauges which are not Hungarian cannot be local maxima of h_0, and from compactness and the fact that h_0 is a continuous function of parameters $(\vec{\lambda}, \vec{\mu})$ we conclude that the set of Hungarian gauges is non-empty.

This mechanism sheds some light on the structure of the problem. "Hard" problems in computational complexity are expected to show the emergence of a pseudo-glassy structure in the phase space, such that a blind local search gets trapped in local minima. Conversely, if all the local minima are also global, and all other points have finite gradient, one could hope to reach a minimum by local search. Assignment problem is not hard, as it is polynomial, but it is not either a trivial problem. In a sense, gauge operations are the proper tool to disentangle the landscape induced by the Assignment cost function into a multidimensional parabola-shaped profile.

7.2.2 Kuhn's contribution

Egerváry theorem seems to be the solution of the problem: just project ϵ through Φ^{trivial}, device some method to find a positive-deficit set X in

polynomial time, and apply a sequence of gauge transformations up to saturate the upper bound.

Actually, this is not enough. Indeed, the compactness argument only proves an existence statement, and that a sequence of gauges would induce a monotonically increasing sequence of h_0's, but as we do not have a positive lower bound on the gain $\delta \cdot d(x)$ attained at each step (unless we work with integer costs, and the bound is exponentially small if we work with N-bit integers), the naive application of the theorem would not provide a polynomial algorithm (more precisely, in the case of N-bit integers, the algorithm would be only *quasi-polynomial* instead of *strongly-polynomial*).

The situation is analogous to the well-known case of maximum flow: for integer or rational capacities, the max-flow algorithm of Ford-Fulkerson [15] is finite, but not strongly-polynomial, while with real capacities examples exists for which the algorithm is not even finite. Nonetheless Dinic [12], and independently Edmonds-Karp [13] proved that with a specific recipe (*Breadth First Search* for Edmonds-Karp) for selecting the augmenting flows, the Ford-Fulkerson algorithm can be transformed in a strongly-polynomial one.

Also in our case, the algorithm can be proven to be strongly polynomial if in each gauge transformation the set X in Egerváry theorem is chosen with the appropriate prescription, namely to choose the X which has the smaller size, among the ones with maximum deficit. More precisely, first one easily proves that such a set is unique. Then, if one calls $(d, s) \in \{0, \dots, N-1\}^2$ respectively the maximum deficit and the size of X as above, one has that at each step with $d > 0$, either $d^{\text{new}} < d$ or $d^{\text{new}} = d$ and $s^{\text{new}} > s$, and the algorithm halts if $d = 0$. As both d and s are polynomially bounded integers, the rules above force the number of gauge transformations to be $\lesssim N^2$. Finally, a procedure implicit in the proof of Kőnig theorem allows at every gauge step to determine the appropriate set X in polynomial time, thus completing the proof of polynomiality of the algorithm.

7.2.3 How do the Cavity Equation Works

The cavity equations for the Assignment Problem are the equations (5.12),(5.13).

Since we want to solve the optimization problem we will concentrate on the zero temperature cavity equations. We search for solutions of this equations as fixed points of a certain discrete-time map on the cavity

fields. This method works if the solution of the equations is a fixed point and if the initial conditions are in the bassin of attraction of this point.

We will consider for the assignment problem the so-called *parallel updating* i.e. we have a discrete-time dynamic and we update every field at each step. The bipartite nature is such that a parallel update of all the variables results in the evolution of two initial condition, for this reason the update steps are divided in even (for the g-fields) and odd (for the h-fields).

It is legitimate to ask what is the actual distribution of the cavity fields in a numerical simulation with parallel updating. The distribution of fields in equations 5.17 and 5.19 are reproduced quite closely in a very short time (of order $\mathcal{O}(1)$ or $\mathcal{O}(\ln N)$) but for a translation overall for each of the four sets of the relevant fields (in each row $N-1$ g-fields and in each column $N-1$ h-fields have the same value).

Let $prob_1(g_i^t = x)$ be the probability that the max in the i row has value x; $prob_2(g_i^t = x)$ be the probability that the second higher value of g_{ij} in row i has value x, and define analoug quantities for h. Then

$$\begin{aligned} prob_1(g_i^t = x) &\sim f(x - G_1(t)) \qquad prob_2(g_i^t = x) \sim f(x - G_2(t)) \\ prob_1(h_i^t = x) &\sim f(x - H_1(t)) \qquad prob_2(h_i^t = x) \sim f(x - H_2(t)) \end{aligned} \tag{7.12}$$

These four functions of the number t of performed steps are linked by the fact that both: $G_1(t) + H_2(t)$ and $G_2(t) + H_2(t)$ converge to the cost of the optimal matching (for the disorder distribution we use $\sim \pi^2/6$). The functions $G_1, -G_2, H_1, -H_2$ asymptotically behave like $-t \cdot \delta$ where δ is a positive "drift" parameter.

In the next sections, with theorem 3 we will be able to explain how these behaviours follow from the dynamic and how to interpret it in terms of gauge transformations.

Strictly speaking the stationarety of the analytic cavity predictions does not hold. In reality it holds with good approximation, as $\langle \Delta \rangle \to 0$ for $N \to \infty$, and holds for a set of parameters which excludes the mean value of the distribution.

7.2.4 How to construct a gauge from the fields

In order to have an algorithm we need that our belief propagation procedure be stopped when the cavity fields are such that we can infer the solution.

This happens for example if we are able to show a gauge, see equation (7.8), such that its zeroes-graph contains a matching in $\mathcal{K}_{N,N}$.

We found a recipe to construct such a gauge using the cavity fields. The gauge parameters are the ones showed in lemma (6). They are appropriate sums of g-fields and h-fields so to delete the effect of the drift and of the oscillation during a period.

In the quasi-periodic regime the quantities $\{g_{i\to j}^{(sec)}, h_{j\to i}^{(max)}\}$ defined as

$$\begin{aligned} g_i^{(sec)}(t) &= \frac{1}{T}\sum_{t'=t}^{t+T-1} 2\underset{j}{\mathrm{nd}}\, g_{i\to j}(t'); \\ h_j^{(max)}(t) &= \frac{1}{T}\sum_{t'=t}^{t+T-1} max_i h_{j\to i}(t'); \end{aligned} \tag{7.13}$$

are a proper gauge. Namely they provide a certificate that the feasible solution found is the optimal one.

The Algorithm so designed find with probability one the solution of the Assignment problem, but (on the set of instances we consider) as appear by numerical study the mean time needed to solve a problem is not finite, this is due to the presence of slow instances where the mechanisms described in the proof of the theorem are slow, this is due to a quasi-degeneracy in the value of the drifts. In 7.4 we will explain how works an average-case finite-time algorithm.

7.3 The Main Theorem

One consequence of Egerváry's theorem (2) is that, to solve the Assignment Problem it is sufficient to find a Hungarian Gauge.

In this section we show that the Belief Propagation Algorithm with parallel updating is always able to find a Hungarian Gauge in a finite time.

The Belief Propagation Algorithm we consider consists in the iterate application of the cavity equations (see equations 5.14) on some arbitrarily chosen starting fields.

This procedure is usually used to decimate the problem: the cavity fields give informations about the biases on the variables and so allow to fix the more biased ones, reducing the problem to a smaller one.

In this section we show that it is possible to write a belief propagation algorithm able to find exactly the solution. It consists in performing the belief propagation updating steps till a Hungarian gauge (a proper gauge such that the zero-subgraph it induces contains a matching) is obtained.

The proof uses some ideas of the Bayati Shah and Sharma's work [3] but enhances their conclusion with a better bound and a really more precise set of statements about the stationary phase. These statements permit to write an Algorithm that find the solution in a finite time.

We first present some definitions and useful notations.

7.3.1 Some Definitions (The Matching Problem on the Unwrapped Graph)

We are given the Complete Bipartite Graph $\mathcal{K}_{N,N}$ with bipartition $\{r, c\}$ where $|r| = |c| = N$ and each vertex of r (*row-vertex*) is adjacent to each vertex in c (*col-vertex*). We refer to the edge which links i and j as $\{i, j\}$.

Associate to $\mathcal{K}_{N,N}$ its Unwrapped Graph, a loopless cover graph of $\mathcal{K}_{N,N}$. The Unwrapped Graph is a tree such that locally (except on the leaves) looks like the original graph $\mathcal{K}_{N,N}$. Every vertex of the Unwrapped Graph is a copy of a vertex of $\mathcal{K}_{N,N}$. More precisely, given an integer t, an ordered pair (i, j), and a "binary choice" among "row"and "col". Let $\mathcal{T}(t, i, j, col)$ be the *Unwrapped Graph* of the original Complete Bipartite Graph $\mathcal{K}_{N,N}$, namely the rooted tree in which the root is a copy of the *col-vertex* i in $\mathcal{K}_{N,N}$. The root of the tree is only connected with one vertex, that is a copy of the j vertex of r-kind in $\mathcal{K}_{n,n}$. The j vertex (the only one vertex at distance one from the root) is connected to N vertices (like in the Complete Bipartite Graph), each one of them is a representative for a vertex of the Complete Bipartite Graph. For $1 < d < 2t$ there are $(N-1)^{d-1}$ vertices at distance d from the root and they are a copy of a *col-vertex* if d is even or a copy of a *row-vertex* if d is odd. Each vertex has N neighbouring except for the root and the vertices at distance $2t$ from the root that are the leaves of the tree $\mathcal{T}(t, i, j, col)$

Given the Weight matrix ϵ_{ij} on $\mathcal{K}_{N,N}$, in a natural way they are induced some weights on the edges of $\mathcal{T}(t, i, j, col)$ so to obtain the *Weighted Unwrapped Graph* $\mathcal{T}(t, i, j, col)$. The Weight on an edge of $\mathcal{T}(t, i, j, col)$ that links two vertex representative of two vertex i', j' in $\mathcal{K}_{N,N}$ is $\epsilon_{i',j'}$.

Consider the matching problem on the Unwrapped Graph as the search for a subset of the set of the edges in $\mathcal{T}(t, i, j, col)$ such that:

- each internal vertex is covered exactly once,
- the leaves are free to be covered or not

- the sum of the weights on covered edges is minimal.

We refer to this problem as the Matching Problem on the Unwrapped Graph. It is solved exactly by the iteration of the cavity equations $4t$ times because the factor graph corresponding to this problem is a tree.

The cavity equations for the matching problem on the internal vertex of the Unwrapped Graph are the same as the ones for the matching problem on $\mathcal{K}_{N,N}$,

$$\hat{g}^t_{i\to j} = \max_{j'\neq j}(-\epsilon_{ij'} - \hat{h}^{t-1}_{j'\to i}) \tag{7.14}$$

$$\hat{h}^t_{j\to i} = \max_{i'\neq i}(-\epsilon_{i',j} - \hat{g}^t_{i'\to j}) \tag{7.15}$$

they are different only on the root and on the leaves because here there are not the interactions (they are free to be matched or not). In order to find if the $\{i, j\}$ edge on the root of the graph is covered or not we do not need to perform the equations $4t$ times, but only $2t$ times (the deepness of the tree). In fact after this time the cavity field on the top will not change under successive iterations of cavity equations. The field on the edge incident on the root is positive if the edge does not belongs to the optimal matching, it will be negative otherwise. We call $\hat{g}^t_{i\to j}$ the cavity field on the top edge It is easy to check that $\hat{g}^t_{i\to j}$ coincides with the cavity field $g_{i\to j}$ on $\mathcal{K}_{N,N}$ after $2t$ iterations of the cavity equations.

One remark should be done about non-null initial fields: starting with arbitrary initial fields is equivalent to fix an external field on the leaves (equal to the initial conditions used on the first run of the cavity equations).

Let E be the set of edges in $\mathcal{K}_{N,N}$ belonging to the optimal matching. Let π be the *Optimal Permutation* that describes the optimal matching

Let M_E be the set of vertices on the unwrapped graph that are copies of the vertices m belonging to the Optimal matching.

Let the *speaking edge* of the edge $\{ij\}$ be the edge $j'i$ in the Unwrapped Graph such that the argmax appearing in equation (7.14) is j'. So that the value of $\hat{g}^t_{i\to j}$ is equal to $(-\epsilon_{ij'} - \hat{h}^{t-1}_{j'\to i})$

For each vertex on the Unwrapped Graph there is and is unique a speaking-edge that is, among the edges that connect it to the lower level, the one which realizes the maximum in the cavity equation. The subgraph of Speaking Edges is such that, given a vertex there's one and only one path of speaking-edges passing through it, we call this path *speaking-path*. Every speaking-path has exactly one end on a leaf (the root node is generally not considered a leaf node).

7.3.2 Scheme Of The Proof

Let the *drift* $\Delta(E')$ of a matching be

$$\Delta(E') = \frac{H(E') - H(E)}{N - |E \cap E'|}$$

Consider the set E^* of all the matchings E_{γ_i} that differs from M for only one loop (the symmetric difference of the matching E_{γ_i} and the optimal matching $M \triangle E_{\gamma_i}$ consists of only one loop γ_i). Let E_1 be the matching in E^* that minimize the drift and E_2 the matching that realizes the second minimum for the drift. Define $\Delta := \Delta(E_1)$, $\Delta := \Delta(E_2) - \Delta(E_1)$.

First we show that the cavity fields on the Assignment problem always identify the solution (lemma 1). At this stage the algorithm find a feasible solution but has not a certificate of optimality. Then we identify the asymptotic (in the time) behaviour of the fields (lemma 4) and give a recipe (6) to give a Hungarian gauge that is a certificate of optimality.

Lemma (1) is proved using the connection between the cavity fields of the Assignment problem and the cavity fields of the Matching problem on the Weighted Unwrapped Graph (where they furnish exactly the solution). If the Unwrapped Graph is enough deep to reproduce well the features of the Bipartite Complete Graph then the edge on the top of the Unwrapped Graph belongs to the Matching if and only if its corresponding on the original graph belongs to the Matching on the Bipartite Graph.

Then, using the same ideas as in (1) we identify the asymptoticly (in the time) behaviour of the fields (lemma 4). The fields result to be a quasi-periodic function of the algorithmic steps, with a period T common to all the fields. The period T is the length of the optimal cycle) and a drift of strength Δ positive for the fields on edge not in E and negative for edges in E.

The quasi-periodic behaviour of the fields allows to find (lemma 6) a Hungarian gauge that is a certificate of optimality.

We call $\overline{E'}$ the maximum over $k \in \{1, \ldots, k\}$, of the maximum over pairs $(i, j) \in M$ and among $(i', j') \in \gamma^* \cap M$ of the minimum among paths γ connecting (i', j') to (i, j), of $-E^{\pm}(\gamma, E)$, and having length equal k modulo $T = ||\gamma^*||$. Remark that $\overline{E'}$ is always smaller than the maximum entry.

Theorem 3 *Given a non-degenerate cost matrix (a matrix such that* $\Delta, \Delta_{12} \neq 0$*), defining,*

$$\overline{E} = \min(\max_{i,j}(\epsilon_{i,j}), min_M(H_\epsilon(M))) \tag{7.16}$$

after a number

$$t^{**} = 2N + \frac{\overline{E}}{2\Delta} + \frac{\overline{E'} + N\Delta}{2\Delta_{12}} \tag{7.17}$$

of iterations of the cavity equations on some arbitrary fields (for example null-fields) the two quantities $g_i^{sec}(t^{**}), h_j^{max}(t^{**})$ *as defined in lemma (6) give a Hungarian Gauge.* t^{**} *is finite and is a function of the instance and of the starting fields.*

7.4 The Untypical Slow Instances

Two remarks are needed. The main theorem give some upper-bounds of the solution time. This time is instance dependent so, there are instances surely easy (if Δ and $\Delta_{1,2}$ are not $\ll 1$), and there are instances that could be slow (the ones with $\Delta \ll 1$ or $\Delta_{1,2} \ll 1$).

Numerical Simulations show on a large set of samples that the smallness of Δ and $\Delta_{1,2}$ is the reason for the slowness of the algorithm: meaning that the time to find the solution is big if $\Delta \ll 1$ or $\Delta_{1,2} \ll 1$).

This is a sign of the fact that it is not possible to ameliorate the inequality of theorem (3). An algorithm able to solve the slow instances in a faster way need some new mechanism. With the aim of writing a new algorithm belief-propagation based it is useful to know deeply the cavity fields dynamic (as it is explained in the proof of the theorem).

We tested a fork algorithm that reduce a slow instance to the solution of two typical problems. Sadly this fork works on average-case instances (in the worst case the two new problems are hard and not forkable).

7.5 Proof of the Main Theorem

As it is proved in lemma (5) after a time t^{**} the cavity fields enters in a quasi-periodic regime (described in lemma (5)). As showed in lemma (6) this is sufficient to construct a gauge with the parameters of the thesis.

□

Lemma 1 *If*

$$t > \frac{\overline{E}}{2\Delta} + N - 1$$

the sign of $\hat{g}^t_{i\to j}$ *is negative or positive respectively if* $\pi(i) = j$ *or not.*

First observe that the sign of $\hat{g}^t_{i\to j}$ is negative or positive respectively if the edge on the top of the Unwrapped Graph belongs to the Optimal Matching on $\mathcal{T}(t, i, j, col)$. If an edge $\{a, b\}$ belongs to the Optimal Matching on the Unwrapped Graph M_U and it does not belong to the Optimal Matching on the Original Graph E then it exists (unique) a path that contains the edge $\{a, b\}$, and that contains alternatively edges in M_U and edges copies of E. This path either ends on two leaves of $\mathcal{T}(t, i, j, col)$ or end on the root and on one leaf. So, if the edge on the root of $\mathcal{T}(t, i, j, col)$ belongs to one among E, M_U but not to the other, there is unique a path alternating in the edges belonging to the two matchings with one end on the root and the other on one leaf.

The proof works through a constructive absurd. If the thesis were false it would be possible to construct a new matching (M'_U) on the Unwrapped Graph such that its cost is smaller than the one of (M_U). If the thesis were false there would be an alternating path on the Unwrapped Graph connecting the root to a leaf (the path is long $2t$ edges), let M_p be the set of edge belonging to this path, ${M_U}^{(1)} = M_U \cap M_p$ and ${M_U}^{(2)} = M_p \setminus M_U$.

Consider the set ${M_U}^* = ((M_U \cup {M_U}^{(2)}) \setminus {M_U}^{(1)})$ namely the set obtained by inverting the variables on the path.

The image of this alternating path on $\mathcal{K}_{N,N}$ can be decomposed in a set of cycles and a single self-avoiding path (not longer than $2N - 2$).

The cost of the matching ${M_U}^*$ minus the cost of the matching M_U is not bigger than $(2t - (2N - 2)) \times (-\Delta) + \min(\text{MaxEntry}, CostM)$. This upper bound of the difference of costs is surely negative when $t > \min(\text{MaxEntry}, CostM)/(2\Delta) + N - 1$ so the thesis is proved.

□

The case with not-null starting fields needs only little modifications. Let $CostM$ be the Cost of the Optimal Matching on $\mathcal{K}_{N,N}$; let $Const_1$ be the minimum starting field and $Const_2$ the maximum one. If the starting fields are not-null Cost(${M_U}^*$) - Cost(M_U) should be not bigger than: $(2t-(2N-2))\cdot(-\Delta)+\min(\text{MaxEntry}, CostM)+Const_2-Const_1$ and so $t^* = [Const2 - Const1 + \min(\text{MaxEntry}, CostM)]/(2\Delta) + N - 1$. The time t^* is bigger but the thesis is still true.

So, after t^* iterations the fields on the edges identify a matching and this matching is the optimal one. A working algorithm needs a halting condition: it must return a certificate that the found matching is the optimal one.

One should also remark that the value Δ is both arbitrarily near to zero, and impossible to deduce from the instance, unless using some procedure which is essentially equivalent to solve the problem otherwise. So we both do not have a bound on the running time, and not even a certified fixed-instance threshold time such that, stopping the algorithm after that time, we can safely conclude the reconstructed matching is optimal.

We now make a statement about the trees with more than $2t^*$ levels. We need in fact that something more happens in order to have the possibility to construct a certificate of optimality for our feasible solution.

Lemma 2 *For each $t \geq t^*$, for any vertex not farther than $t - t^*$ from the root (all the first $t - t^*$ levels of the tree) all the edges $\in M_E$ are speaking-edges.*

This is a corollary of lemma (1) It follows by the fact that in order to have that $\hat{g}^t_{i\to j} = (-\epsilon_{ij'} - \hat{h}^{t-1}_{j'\to i})$ be positive, the field h must be negative, but this happens only if the edge $\{ij'\}$ is in E.

□

If $\hat{g}^t_{i\to j}$ on the root of the Unwrapped Graph $\mathcal{T}(t, i, j, col)$ is positive for every $j \neq \pi(i)$ and is negative only for $j = \pi(i)$ this implies that among the entering messages $(-h_{ij} - \epsilon_{ij})$ with $j = 1, \ldots, N$ there is only one positive and it is the one such that $\pi(i) = j$.

Remark 1 *In the region of distance smaller than $t - t^*$ from the root, the Speaking-path is alternating in the matching.*

On the speaking-path starting fronm the root of the Unwrapped Graph, alternatively one edge of the speaking path is in M_e and one not. This happens because each vertex has exactly one edge that belongs to M_e; a path constraint to pass through a given vertex either it pass in the incoming edge through an edge belonging to M_e or it pass in the outgoing one through an edge belonging to M_e.

Lemma 3 *The Speaking-path on a vertex i linked upward to $\pi(i)$ is the one which maximize a global cost function on the path starting on i and*

ending at the level t^*. The Speaking-path on a vertex i linked upward to $j \neq \pi(i)$ is the one which minimize an analog quantity.

The fact that the speaking-path is alternating in the matching E, permits to write the equation

$$\hat{g}^t_{i,j} = \max_{j' \neq j}(-\epsilon_{i,j'} + \min_{i' \neq i}(+\epsilon_{i',j'} + \hat{g}^{t-1}_{i',j'})$$

in the form of equation (7.18). Maximizing this quantity at each level of the tree is equivalent to maximizing it on the whole path where it consists in a sum of ϵ with alternate signs with $2(t - t^*) + 1$ addends.

If $t > t^*$ the Lemma (2) can be used to write the equations for two steps:

$$\hat{g}^t_{i \to j} = \max_{j' \neq j}(-\epsilon_{ij'} + \epsilon_{\pi^{-1}(j') \to j'} + \hat{g}^{t-1}_{\pi^{-1}(j') \to j'}) \tag{7.18}$$

$$\hat{h}^t_{j \to i} = \max_{i' \neq i}(-\epsilon_{i',j} + \epsilon_{i',\pi(i')} + \hat{h}^{t-1}_{i' \to \pi(i')})$$

in both the cases we have to maximize a quantity each two steps on paths running on a tree from the root to the level t^* where (7.18) ceases to hold. This is equivalent to a global maximization. Analog relations hold if $j \neq \pi(i)$ So, the thesis is proved.

□

The quantity $\overline{E}$ defined in 7.16 is related to the cost of a simple open path alternating on the optimal matching, namely, while a closed path γ has always a positive cost, an open path can also have a gain, but it is bounded by $-\overline{E}$.

Lemma 4 *The speaking path on the unwrapped graph $\mathcal{T}(t, i, j, col)$, if wrapped on $\mathcal{K}_{N,N}$ must intersect γ^* if*

$$t \geq t^{**} := t^* + N + \frac{\overline{E} + 2\overline{E'} + N\Delta}{\Delta_{12}} \tag{7.19}$$

Assume first $j = \pi(i)$ then the speaking path is alternating on the optimal matching, start from the cavity field g_0 at the bottom leaf. The path, when wrapped on $\mathcal{K}_{N,N}$ is decomposable into an open path of length at most $N - 1$ alternating on E (thus costing less than $\overline{E}$), and a set of cycles. If the path does not touch the feedback cycle γ^*, this happens in particular for all the cycles in the decomposition, and thus,

by definition of $\Delta_{1,2}$ these cycles provide at least a rate $-(\Delta+\Delta_{12})$ per step, for at least $t-t^*-N$ steps. So, the maximum alternated-cost of the speaking path is bounded by:

$$g_0+\overline{E}-(t-t^*-N)(\Delta+\Delta_{1,2})$$

The alternated-cost of a path which goes through the optimal cycle many times is easy to bound using the quantity $\overline{E'}$. The complete cycles provide an alternate-cost larger than $-(t-t^*-2k-l)\Delta$. The remaining open path has an alternate-cost boundable from below so that:

$$-E^{\pm}(\gamma,E)=g_0-2\overline{E'}-l\Delta-(t-t^*-l)\Delta$$

The cost difference of the so constructed cycle and the cycle not intersecting γ^* is at least $\overline{E}+2\overline{E'}-(t-t^*-N)\Delta_{12}+N\Delta$ which for $t\geq t^{**}$ with t^{**} defined in 7.19 implies that the speaking path must touch the feedback cycle.

□

As a consequence we have that asymptotically in t the fields $g_{i\to j}\sim Const_{i,j}+t\cdot\Delta$ with $j\neq\pi(i)$ and $g_{i\to j}\sim Const_{i,j}-t\cdot\Delta$ if $j=\pi(i)$. Where $Const_{i,j}$ is a constant depending on the history of the cavity fields in the first t^* steps.

Lemma 5 *After $t^*+N'+T$ steps the cavity fields are quasi-periodic functions of the number of iterations. Meaning that*

$$g^t_{i\to j}=g^{t-T}_{i\to j}-\Delta\cdot T\ \forall i,j:j\neq\pi(i)$$

$$g^t_{i\to j}=g^{t-T}_{i\to j}+\Delta\cdot T\ \forall i,j:j=\pi(i)$$

As we know for every $t>t^*+N'+T+1$ the speaking-edge run at least one loop on the optimal cycle. Suppose that $g^t_{i\to j}<g^{t-T}_{i\to j}+\Delta\cdot T$. Let γ_{T-t} be the speaking path for $g^{t-T}_{i\to j}$ and γ_T be the speaking path for $g^t_{i\to j}$. The speaking path are path maximizing (see lemma 3) the value on $g^t_{i\to j}$. The $g^t_{i\to j}$ constructed via a modified γ_{T-t} cycle that run one more time on the optimal path is equal to $g^{t-T}_{i\to j}-\Delta\cdot T$. This fact is in contradiction with the optimality of the path γ_T.

Suppose now $g^t_{i\to j}>g^{t-T}_{i\to j}+\Delta\cdot T$. Then the $g^{t-T}_{i\to j}$ constructed via a modified γ_T path that run one less time on the optimal cycle is $=g^T_{i\to j}+\Delta\cdot T$. But it is in contradiction with the optimality of the path γ_{T-t}.

Similarly can be proved the quasi-periodicity for the fields $g_{i\to\pi(i)}$.

□

Lemma 6 *In a quasi-periodic regime (see lemma (5)), and $t > t^{**}$ (where t^{**} has been introduced in lemma (4)) the quantities $\{g_i^{(sec)}(t), h_j^{(max)}(t)\}$ defined in equation 7.13 are a good set of dual variables, meaning that the gauge with $\lambda_i = g_i{}^{(sec)}(t)$ and $\mu_i = h_j{}^{(max)}(t)$ provides a Hungarian Gauge: a proper gauge whose zeroes subgraph admits a perfect matching.*

Call ϵ'_{ij} the weights shifted by the variables, i.e., $\epsilon_{ij} + \lambda_i + \mu_j$. We have two statement to prove: $\epsilon'_{i\,\pi(i)} = 0 \,\forall\, i$ and $\epsilon'_{ij} \geq 0 \,\forall\, i, j$. In the first case we have

$$\begin{aligned}
\epsilon'_{i\,\pi(i)} &= \epsilon_{i\,\pi(i)} + \frac{1}{T}\sum_{t'=t}^{t+T-1}(g_i^{(sec)}(t) + h_{\pi(i)}^{(max)}(t)) = \\
&= \epsilon_{i\,\pi(i)} + \frac{1}{T}\sum_{t'=t}^{t+T-1}(g_i^{(sec)}(t) + (-g_i^{(sec)}(t) - \epsilon_{i\,\pi(i)})) = \\
&= 0 \qquad (7.20)
\end{aligned}$$

while for the second case, we have that if $\pi \neq \pi(i)$

$$\begin{aligned}
\epsilon'_{ij} &= \frac{1}{T}(h_j^{(max)}(t+T+1) + g_i^{(sec)}(t) + \qquad (7.21)\\
&\quad + \sum_{t'=t+1}^{t+T-1}(h_j^{(max)}(t+1) + g_i^{(sec)}(t))) \qquad (7.22)\\
&= \Delta + \frac{1}{T}\sum_{t'=t}^{t+T-1}(g_i^{(sec)}(t) + (\epsilon_{ij} + h_j^{(max)}(t-1))) \qquad (7.23)
\end{aligned}$$

but, because of the cavity-equations features for times after $t^\star$, each summand is at sight positive (and zero if j is the 'arg-second' index at all the timesof the period).

□

7.6 Special Gauges

As said in section 7.1 for every instance there are many *Hungarian gauges*. We know that to find a solution we only need a Hungarian gauge with only N zeroes, anyway the typical output gauge of the Hungarian Algorithm tipically contains many zeroes. Moreover we know that, if the

instance is not degenerate, can not contain more than $2N-1$ zeroes. In this section we give and prove the following theorem stating that there is always the possibility to find a special Hungarian gauge.

Theorem 4 *Given an instance ϵ of Assignment Problem, and a column j, it is always possible to perform a Hungarian gauge transformation such that in the gauged matrix ϵ' all the colums $j' \neq j$ have at least two zeroes. Moreover, if the instance is non-degenerate, there is exactly one zero in column j and two zeroes in all the other columns, and the graph of zeroes Z is a spanning tree on $\mathcal{K}_{N,N}$.*

Up to a relabeling of the columns, we can assume that $j = 1$. Choose an optimal configuration π. Again, up to a relabeling of the rows, we can assume that π is the identity permutation, and in particular that $\pi(1) = 1$. Call M_1 the corresponding matching on $\mathcal{K}_{N,N}$, with permutation π, and H_1 the cost.

For each $i \geq 2$, analogously define M_i and H_i the optimal matching and cost in the subensembles in which $\pi(i) = 1$ is forced. If M_i is not unique, consider a whatever optimum M_i'. The symmetric difference $M_1 \triangle M_i'$ is composed of a set of self-avoiding loops, one of which contains the edges (11) and $(i1)$ by construction (the "special" loop).

By the optimality hypothesis, on each loop the alternated-sign sum of entries ϵ_{ij} on edges (ij) is zero, except for the special one, for which it makes $H_i - H_1$. This implies that the matching M_i such that its symmetric difference with M_1 contains only the special loop has the same cost H_i as M_i'.

Consider now the following gauge

$$\lambda_i = \epsilon_{ij} + H_1 - H_i\,; \qquad \mu_1 = 0\,; \tag{7.24}$$

$$\forall\ j \neq 1 \qquad \mu_j = \epsilon_{jj} - \lambda_j\,. \tag{7.25}$$

This gauge is such that $\epsilon'_{i1} = H_i - H_1$ for all $i \in [N]$, and, for optimality of M_1, these entries are all ≥ 0. Also, by construction, all entries ϵ'_{ii} on the optimal matching M_1 are zero. So, in order to prove that the gauge is Hungarian we only need to prove that ϵ'_{ij} for $i \neq j$ and $j \geq 2$ is never negative.

We do this by absurd. Suppose that $\epsilon'_{ij} < 0$. If (jj) is not in the loop

$$M_1 \triangle M_i = \big((i1), (ii), (i_2 i), (i_2 i_2), \ldots, (11)\big)\,,$$

(or if $i = 1$, so that $M_1 \triangle M_i = \emptyset$) then the matching $\hat{M}_j$ such that

$$M_1 \triangle \hat{M}_j = \big((j1), (jj), (ij), (ii), (i_2 i), (i_2 i_2), \ldots, (11)\big)$$

(i.e. $\big((j1), (jj), (1j), (11)\big)$ if $i = 1$) would violate the optimality condition on M_j, as we have

$$H_1 - H_i + \epsilon'_{i1} = \big(\epsilon'_{ii} - \epsilon'_{i_2 i} + \epsilon'_{i_2 i_2} - \ldots + \epsilon'_{11}\big) = 0\,, \tag{7.26}$$

so that we would get

$$H_1 - H_\epsilon(\hat{M}_j) + \epsilon'_{j1} = \epsilon'_{jj} - \epsilon'_{ij} + \big(\cdots\big) > 0 \tag{7.27}$$

because $\epsilon'_{jj} = 0$ and the expression in parenthesis coincide with the one in (7.26). Together with $H_1 - H_j + \epsilon'_{j1} = 0$, this causes an absurd, as would imply that M_j was not optimal.

If (jj) is in the loop $M_1 \triangle M_i$, the reasoning is similar, in fact in that case the symmetric difference $M_1 \triangle M_i$ is

$$\big((j1), (jj), (ij), (ii), (i_2 i), (i_2 i_2), \ldots, (11)\big)\,.$$

Consider now the matching $\hat{M}_i$ such that its symmetric difference with the matching M is

$$\big((i1), (ij), (i_{k+2} j), \ldots, (11)\big)\,.$$

its cost is $\epsilon_{i1} + \epsilon_{ij}$ then, from the hypothesis ($\epsilon_{i,j} < 0$) it is smaller than ϵ_{i1}, in contraddiction with the optimlity of M_i.

In order to prove that ϵ' has at least two zeroes per column $i \geq 2$, consider the minor of ϵ' obtained by deleting column 1 and row i. Then we have a matrix with non-negative entries, that contains a matching of zero cost (as $h_0 = H_1$ by construction of the gauge, and the removed element $\epsilon'_{i1} = H_i - H_1$, so this already saturates the lower bound H_i), and this implies that every column has at least one zero entry. This statement says nothing on columns $j \neq i$, where one still has the entry $\epsilon'_{jj} = 0$ by construction, but the entry ϵ'_{ii} is not in the minor, so there must be another zero in the same column.

If the instance is not degenerate, the set $Z_{\min}$ has cardinality N, and for any Hungarian gauge the graph Z has no loops. As we have proven that our gauge has at least $2N - 1$ edges, and a loop-free subgraph of a graph with $2N$ sites has at most $2N - 1$ edges (in this case being a spanning tree), also the final statement of the theorem follows.

□

Chapter 8

Some Variants of Assignment

In this chapter we define and describe some features of some very different problems which have the property to be defined in a similar way to the assignment problem. This is the case for the three-dimensional generalization of the Assignment. We will also describe the geometrical properties of the assignment solutions for the problem on a two dimensional lattice. We will see how the assignment problem can be used to solve fastly some instances of the Traveling Salesman Problem and how the investigation of the hard instances of this problem induces to the definition of a *NP-complete* variant of the Assignment.

8.1 Multi-index Matching

Consider the weighted hypergraph containing $n \cdot N$ vertices partitioned in n subset and N^n weighted hyperedges each one of coordination n, connecting vertices from different subsets. The problem of finding the optimal Matching (edge-subset of the hypergraph so that each vertex has coordination 1) is the multi-index matching problem. Remark that the Assignment problem is the $n = 2$ case of the multi-index matching problem.

Here and in the following, we describe a problem in the scheme

Problem name:

instance description.	feasible solution; condition to satisfy.

then the 3-Dimensional Matching Problem can be stated:

3-Dimensional Matching Problem:

$G(V_1, V_2, V_3; T)$	$M \subseteq T$;
with $T \subseteq V_1 \times V_2 \times V_3$	$\lvert M\rvert = n$,
and $\lvert V_1\rvert = \lvert V_2\rvert = \lvert V_3\rvert = n$.	$\forall v \in V_1 \cup V_2 \cup V_3 \quad \deg_T(v) = 1$.

The object G could be called a "tripartite hypergraph": indeed, instead of edges, it contains hyper-edges with three endpoints, one per set of vertices. Equivalently to what has been done in section 8.4.2, another representation turns out to be useful, in which the allowed elements for a matching are encoded in an array of zeroes and ones. As guessable, now the array is three-dimensional. Thus, define $W = \{w_{ijk}\}_{i,j,k=1,\dots,n}$ such that $w_{ijk} = 1$ if $(i, j, k) \in T$, with $i \in V_1$, $j \in V_2$ and $k \in V_3$, and $w_{ijk} = 0$ otherwise. A feasible matching M is then described by an array $X = \{x_{ijk}\}$, with $x_{ijk} = 0, 1$, such that $x_{ijk} = 1$ if the triangle $t = (i, j, k)$ is in M, and then exactly one element of the array is equal to 1 per i, j or k fixed, i.e.

$$\forall\, i = 1, \dots, n \qquad \sum_{j,k} x_{ijk} = 1\,; \tag{8.1a}$$

$$\forall\, j = 1, \dots, n \qquad \sum_{i,k} x_{ijk} = 1\,; \tag{8.1b}$$

$$\forall\, k = 1, \dots, n \qquad \sum_{i,j} x_{ijk} = 1\,. \tag{8.1c}$$

The cost function for M is then restated into

$$C_W(X) = \sum_{i,j,k} w_{ijk} x_{ijk}\,, \tag{8.2}$$

and X is a valid matching if $C_W(X) = n$. The numerical version is defined accordingly, just now the weights w_{ijk} are generic integer, and we have a threshold value for the cost (8.2).

The space of configurations of multi-index matching, for $n \geq 3$ at low temperatures requires replica simmetry breaking to be described. The presence of hard-constraints in this problem lead to a scheme different than for other optimization problem. In fact the system is correctly described by a *frozen*-1RSB Ansatz where states are made of single configurations. See [30] for further details.

8.2 Assignment on planar graphs

Let G be a planar weighted graph which admits a bipartition in odd vertices and even vertices (the square lattice is an example of such a graph). It is possible to associate to this graph an assignment problem for the even and the odd vertices only if the system has an equal number of even and odds vertices (this is not the case for the square lattice on a square domain of *odd*-vertices $\times$ *odd*-vertices).

Let G be such that the odd species has one vertex more the the even. So that $|odd| = N + 1$, $|even| = N$. Consider the $N + 1$ assignment problems obtained removing the i-th odd vertex, denote by π_i their solutions, a solution being a set of edges connecting the odd edges to the even ones. The symmetric difference of two of these solutions (i, j) is a walk from i to j plus a set of loops. If the weights on the tree are different so that the solutions do not admit degeneracy (as in hypothesis of chapter 7) then the symmetric difference of π_i and π_j is a self avoiding walk from i to j.

As a consequence of the existence of the special gauges introduced in section 7.6 it is possible to state that the union

$$\cup_{i,j=1}^{N+1}(\pi_i \Delta \pi_j) = \cup_{i=1}^{N+1}\pi_i \tag{8.3}$$

is a tree. To prove this statement we need to introduce an extra even vertex, the $N + 1$-th one. We connect this vertex to all the odd ones with $N + 1$ new edges. As we showed in section 7.6 a gauge such that in the *zeroes graph* the even $N + 1$-th vertex is connected to only one vertex and the whole *zeroes graph* is a tree exists. This special gauge has the property that any even vertices but the special one has coordination two. Then (after the remotion of the special even vertex and of one of the odd ones) a matching on the so given zeroes graph always exists. It is now easy to show that the zeroes graph coincides with the tree defined by the equation (8.3).

We studied some geometrical properties of the trees above for square lattices. We concentrated on the walk starting from the middle of the bottom boundary and ending in the middle of the upper boundary, the so defined curve is fractal and has some properties suggesting the presence of conformal invariance, anyway this curves are not distributed (in law) as the SLE [1] curves.

[1] Schramm-Loewner Evolution (SLE) is a stochastic process that describe the growht of a self-avoiding path in connected bidimensional domains, for a review see [2] and references therein.

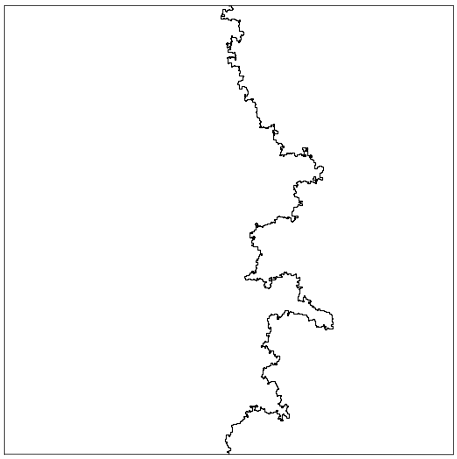


Figure 8.1: Double-Matching Instance on a 511×511 square lattice

The tree structure of the Assignment solutions as described in the context of the special gauges suggests some possible relation between the geometrical structure of domain walls in disordered systems excited states and SLE. Further investigations will consider other curves in the tree defined in (8.3).

8.3 Traveling Salesman Hard Instances

We introduced the traveling salesman problem (TSP) in section 4.1 as an *NP-complete* problem. Anyway an average case analysis shows that this problem is not hard: it exists an algorithm such that the average solution time is polinomially bounded.

Given a set of "distances" between "cities" the TSP consists in finding the shortest *circuit* that passes through all cities exactly once. The TSP can be formulated as a variant of the assignment in the following way. Given two cities i and j let $M_{i,j}$ be the distance between i and j. (In the non-symmetric TSP $M_{i,j}$ is not constrained to be equal to $M_{j,i}$). It is possible to associate a union of circuits to any assignment: consider a graph with N vertices and draw an edge connecting i to j if the element (i, j) belongs to the assignment. In this framework the TSP consists in finding the optimal assignment for the instances given by the matrix M such that the corresponding union of circuits consists of a single circuit.

8.3.1 Polynomial Algorithm for the Average-Case TSP

Given a TSP instance consisting of the distances-matrix M if we were able to examine all the $N!$ feasible assignments (say if they consists of a single loop or not) we should just pick the cheapest among the "good" assignments. Given an assignment for a given set of N cities the algorithm we can test if it consists of a single loop in linear time (with N steps). Also if the single-loop property is easy to test it is hard to keep efficently account of it in an algorithm consisting of a local search. In fact the single-loop property is not a local property (to test it we need to visit the whole circuit).

An efficent way to solve the TSP consists in the examination of the whole spectrum of the Assignment problem (see section 8.3.2) up to the first assignment configuration consisting of a single loop.

The number of different circuits visiting all the sites (cities) is a fraction $1/N$ of the number of feasible assignments and, since the single-loop circuits are distributed homogenously in the feasible assignments, then the time needed to solve a randomly chosen TSP instance is N^2 times the time needed to solve an Assignment instance (the "one-loop test" being subleading with respect to the Assignment problem).

The fact that the TSP problem is NP-complete (or more exactly NP-hard) but result to be polynomial in average means that there is a set of instances not fastly solved by the algorithm given above.

In the set of hard instances there are the instances such that a set S_1 of cities are close each to the other belonging to the set but they are far from any other out of this set. The lowest states for the assignment problem consists of some loops visiting the cities in S and some other loops visiting the other cities, a loop involving cities both in S both out of S will cost at least two times the minimum distance between a city in S and one out of S.

It is not hard to write algorithms able to find a structure like this when present and to solve the problem when there is such a structure.

Anyway it becomes more and more hard to write efficent algorithms when one restrict the set of random TSP instances to the set of instances containing several structures as the one described above (of cities close one to the other but very far from the other) or instances containing several nested structures of the kind described above.

In this context emerged the idea to study the set below of instances. This set contains only hard instances with respect to the presented algorithms (it is not possible to prove the existence of intrinsically hard

instances without proving NP≠P). Consider (for $1 \leq n \leq N/2$) instances with $M_{2n,2n-1}$ and $M_{2n-1,2n}$ smaller than N times the smallest M entries not in the set above.

The ground-state assignment consists of $N/2$ loops. The TSP solution via the visit of the spectrum of its related assignment problem requires an exponential number of visited assignment solutions.

We restrict ourselves to this set because we have argument to say that it contains a large fraction of hard instances, anyway the concept of hardness is algorithm dependent, all that we can prove in a formal way about the hardness of this set of instances is the *NP-Completeness* of the TSP restricted to this set of instances.

We introduce the One-in-two algorithm to show the NP-Completeness of the so restricted TSP problem and also because the One-in-two problem is a simplification of the TSP problem because there is not anymore the unlocal constraint asking to the assignment to consists of a single loop. In fact in the One-in-Two problem there are not unlocal constraints not yet present in the assignment problem.

8.3.2 The Assignment Spectrum

Consider the problem of finding the k-th cheapest Assignment among the feasible assignments. We refer to this problem as the visit of the Assignment spectrum. As we explained in **??**sect:AverageCaseTSP) it is part of an algorithm solving the TSP problem in an average time polynomial. Here we present an algorithm that is easily generalizable to the visit of the spectrum of other optimization problems (as the shortest path one). This algorithm is bounded (in the worst-case) by $k \cdot N$ times the time needed to solve an instance of the original problem plus the time needed by the sorting problem (subleading with respect to the assignment).

The algorithm is based on a successive partition of the space of feasible solutions. We first solve the assignment problem. We call this assignment (permutation) π_0, then for $1 \leq i \leq N$ we consider the N problems of finding the best assignment π_i such that $\pi_i(j) = \pi_0(j) \, \forall j < i$ and $\pi_i(i) \neq \pi_0(i)$. We sort this solutions (it could be useful to represent each of this solutions as a vertex of level one in a rooted tree where the root represents the optimal solution of the original problem), say i^* is the best one, we indicate the subsets belonging to the partition with S_i.

Since the subdivision of the space of solutions given above is a partition of the whole space of solutions then the second optimal solution

of the original problem is the best solution among the N vertex of level one in the constructed tree. Say it is the set $S_{i'}$

The step forward consists in the partition of $S_{i'}$: the space of feasible solutions containing the second best solution. After this partition we need to solve the assignment problem in all the subset of the new partition; then we sort this sets with respect to the cost of the optimal assignment therein and compare the minimum among these costs with the cost of the lowest cost matching in the sets S_i with $i \neq i^*$.

The so given algorithm, iterated so to obtain the whole spectrum up to the k-th level requires at each step the solution of the original problem (in our case the assignment one) at most N times and at each step gives a new element of the spectrum of the original problem we are interested in.

8.4 One in Two Problem

8.4.1 Motivations and a digression

In this section we describe a new NP-complete problem, the One-in-Two Matching, in the next section we will give a polynomial reduction from Boolean Satisfiability.

The existence of a new NP-complete problem is interesting by itself, in the idea of making the list of NP-complete problems still wider (a larger list of NP-complete problems gives a larger number of possible starting points for a reduction proof, and thus makes easier the task of determing whether a new problem is NP-complete). As we will see, as a side result we show that One-in-Two Matching Problem induces a chain of reductions from Boolean Satisfiability to the important 3-Dimensional Matching Problem (3DM in Garey-Johnson [18]) which avoids the complicated and size-demanding original reduction of Karp [23].

Another motivation, which was indeed the original one, is that One-in-Two Matching and Assignment give a hint on the structural reason why Hamiltonian Circuit (HC) and Traveling Salesman (TSP) are NP-complete, although their analogue Matching and Assignment are polynomial, and intrinsecally simple for what concerns the energetic landscape of configurations. Indeed, configurations of HC and TSP are permutations π composed of a single cycle. It is trivial to impose that π has no fixed points (just taking infinite weights on the diagonal), while the remaining restriction to have no cycles of length $2 \leq \ell < n$ must be at the root of the complexity discrepancy among the two problems.

In the definition of One-in-Two problems, put at infinity the non-diagonal block elements, $w_{2i-1,2i} = w_{2i,2i-1} = +\infty$, make the diagonal weight very favourite, $w_{ii} \to w_{ii} - \Delta$ with $\Delta \to +\infty$, and then transpose all the row pairs $(2i-1, 2i)$. Then, the One-in-Two constraints is equivalent to forbid all the length-2 cycles among $2i-1$ and $2i$ (which are a small subset of all the cycles forbidden in TSP). On the other side, the matching $\pi(2i) = 2i-1$, $\pi(2i-1) = 2i$ constitutes the obvious ground state of pure Assignment.

So, although *random* TSP instances have good heuristics and efficient approximants, based on the connection with Assignment [22], this new ensemble of random TSP instances would be hard in the average case, as, instead of just condensing a few ($\mathcal{O}(\ln n)$) relevant long cycles, it must first choice how to disentangle $\mathcal{O}(n)$ robust short cycles. Roughly speaking, as a cycle of length ℓ can be broken in ℓ points, and the set of cycle lengths is a partition of n, the average complexity of this procedure scales with $\exp(n/\overline{\ell}\,\ln\overline{\ell})$, and has a finite maximum for $\overline{\ell} = \mathcal{O}(1)$. Similar arguments are depicted at the end of section 8.4.3.

The reason lying behind the hardness of One-in-Two problem is the fact that just a change of a single "spin" determination behaves as a change of a whole row and of a whole column in the original instance. As we know the replica simmetry of problems is broken when small changes in the system induce the cross of energy (or free-energy) among local ground-state. The lack of a special simmetry allowing the design of a specific algorithm (as happens for Assignment or XOR-SAT) make of One-in-Two problem an hard problem both in the worst case (as we prove) both on average (as we guess).

Indeed, the existence of a NP-completeness proof for One-in-Two problems suggests that this narrow subset of the set of extra constraints of HC and TSP already contains the core of extra complexity of these problems. Also remark that, although the chain of reductions from SAT to Hamiltonian Circuit (3-SAT $\to$ Vertex-Covering $\to$ HC) is beautiful and elegant [18], our direct reduction to One-in-Two Matching is much cheaper.

8.4.2 Definition of 2-Dimensional Matching and Linear Assignment Problems

Given an unoriented bipartite graph $G(V_1, V_2; E)$, with $|V_1| = |V_2| = n$, the *Matching Problem* (discussed in chapter 7) asks for a subset $M \subseteq E$ of the edges, with cardinality n, such that each vertex has degree exactly

1 in $G|_M$, or for a certificate that such a set does not exist. A set M satisfying this requirement is called a *perfect matching* over G.

The scheme corresponding to Matching is

Matching Problem:

$G(V_1, V_2; E)$ with $\|V_1\| = \|V_2\| = n$.	$M \subseteq E$; $\|M\| = n, \quad \forall v \in V_1 \cup V_2 \quad \deg_M(v) = 1.$

This problem is a specific case of the more general Linear Assignment Problem, in which integer weights $w(e)$ are associated to the edges, a threshold value k is given, and the search is restricted to perfect matchings M such that the sum of weights on the edges of M is smaller than k [2].

Linear Assignment Problem:

$G(V_1, V_2; E)$ with $\|V_1\| = \|V_2\| = n$; $w : E \to \mathbb{Z}$; $k \in \mathbb{Z}$.	$M \subseteq E$; $\|M\| = n, \quad \forall v \in V_1 \cup V_2 \quad \deg_M(v) = 1,$ $\sum_{e \in M} w(e) \leq k.$

More precisely, the weights could be also infinite, i.e. $w : E \to \mathbb{Z} \cup \{+\infty\}$, with the natural formal rules $n + \infty = +\infty + n = +\infty + \infty = +\infty$ and $+\infty > k$. Then, one can assume without loss of generality that G is the complete balanced bipartite graph.

In traditional notations, Matching Problem is resumed in the case $w(e) = 1$ for edges in $E(G)$ and $w(e) = 0$ otherwise, $\leq$ being replaced by $\geq$, and $k = n$.

A convenient representation of these problems is via the $n \times n$ matrix $W = \{w_{ij}\}$ of the weights. A feasible matching M is then described by a matrix $X = \{x_{ij}\}$, with $x_{ij} = 0, 1$ and exactly one element equal to 1 per row and per column, such that $x_{ij} = 1$ if edge $e = (i, j)$ is in M. The cost function for M is restated into

$$C_W(X) = \sum_{i,j} w_{ij} x_{ij} = \mathrm{tr} W X^{\mathrm{T}} . \tag{8.4}$$

Another convenient representation of feasible matchings is via permutations in the symmetric group over n elements, $\pi \in \mathcal{S}_n$, where $\pi(i) = j$

[2] Equivalently, one can restrict to consider the complete balanced bipartite graph with $2n$ vertices, $\mathcal{K}_{n,n}$, and set $w(e) = +\infty$ for edges in $E(\mathcal{K}_{n,n}) \setminus E(G)$.

if edge (i, j) is in M. In this notation the cost function reads

$$C_W(\pi) = \sum_i w_{i\pi(i)} \,. \tag{8.5}$$

An example of problem instance and solution could be (on the left, items $w_{i\pi(i)}$ are written in bold)

$$W = \begin{pmatrix} 3 & 7 & \boxed{\mathbf{2}} & 4 & 1 & 1 \\ \boxed{\mathbf{1}} & 6 & 1 & 7 & 8 & 2 \\ 3 & \boxed{\mathbf{3}} & 2 & 5 & 6 & 3 \\ 4 & 2 & 8 & 6 & \boxed{\mathbf{2}} & 5 \\ 5 & 5 & 1 & 6 & 3 & \boxed{\mathbf{4}} \\ 4 & 9 & 8 & \boxed{\mathbf{1}} & 4 & 3 \end{pmatrix} \quad \text{with } k = 15; \qquad \begin{array}{l} \{\pi(i)\}_{i=1,\dots,6} = \{3, 1, 2, 5, 6, 4\}\,; \\ C(\pi) = 13. \end{array}$$

Given a whatever Assignment instance W, many algorithms allow to find in polynomial time the optimal assignment π^*, and its cost $C^* = C(\pi^*)$, the most famous being probably the *Hungarian Algorithm* [28, 29].

8.4.3 Definition of One-in-Two Matching and One-in-Two Assignment Problems

Now we can define the *One-in-Two Matching* and *Assignment Problems* as the variants of Matching (resp. Assignment) Problem in which, assumed that the dimension $2n$ of the matrix is even, the set of allowed partitions π is restricted to include only the ones such that, for each $i = 1, \dots, n$, either $\pi(2i-1) = 2i-1$ or $\pi(2i) = 2i$.

Thus the description of One-in-Two Matching and Assignment could be resumed in the tables

One-in-Two Matching Problem:

$G(V_1, V_2; E)$	$M \subseteq E$;
with $\lvert V_1\rvert = \lvert V_2\rvert = n$;	$\lvert M\rvert = n, \quad \forall v \in V_1 \cup V_2 \quad \deg_M(v) = 1,$
partition of $\{V_1; V_2\}$ into	
quadruplets $q = (v, v'; u, u')$.	$\forall q \quad \big((v, u) \in M\big) \,\dot\vee\, \big((v', u') \in M\big).$

One-in-Two Assignment Problem:

$G(V_1, V_2; E)$ with $\lvert V_1\rvert = \lvert V_2\rvert = n$; partition of $\{V_1; V_2\}$ into quadruplets $q = (v, v'; u, u')$; $w : E \to \mathbb{Z}$; $k \in \mathbb{Z}$.	$M \subseteq E$; $\lvert M\rvert = n, \quad \forall v \in V_1 \cup V_2 \quad \deg_M(v) = 1,$ $\forall q \quad \big((v,u) \in M\big) \dot\vee \big((v',u') \in M\big),$ $\sum_{e\in M} w(e) \le k.$

Remark that, when proven that One-in-Two Assignment is NP-complete, we will also have a proof that the variant with $\big((v,u) \in M\big) \vee \big((v',u') \in M\big)$ instead of $\big((v,u) \in M\big) \dot\vee \big((v',u') \in M\big)$ is NP-complete. Indeed, given an instance of the OR problem, shifting the diagonal weights to $w_{ii} \to w_{ii} + \Delta$, and $k \to k + n\Delta$, we have that $C_W(M) - k = n'\Delta$, with n' the number of blocks with two matched elements, and in the limit $\Delta \to +\infty$ we recover the analogous XOR problem. On the contrary, in the limit $\Delta \to -\infty$ the problem becomes trivial.

The costs of the diagonal elements do not play any role, and can be fixed to zero. Indeed, if for some $i \le n$ we have $w_{2i,2i} = w_{2i-1,2i-1} = +\infty$, no finite-cost assignment exists, while if only one of the two is infinite (say, $w_{2i-1,2i-1}$), we are forced to fix the permutation on the other one ($w_{2i,2i}$), and thus the elements with i or j equal to $2i$ never play a role: we would have had an identical cost function if $w_{2i-1,2i-1}$ were zero, and $w_{2i,j} = w_{j,2i} = +\infty$ for each $j \neq 2i$:

$$W = \left(\begin{array}{cc|cc} +\infty & w_{12} & w_{13} & \dots \\ w_{21} & \boxed{\boldsymbol{w_{22}}} & w_{23} & \dots \\ \hline w_{31} & w_{32} & w_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{array}\right) \equiv \left(\begin{array}{cc|cc} 0 & +\infty & w_{13} & \dots \\ +\infty & \boxed{\boldsymbol{w_{22}}} & +\infty & \dots \\ \hline w_{31} & +\infty & w_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{array}\right). \quad (8.6)$$

Thus, without loss of generality, we can assume that w_{ii} is finite for each i. Then, from the invariance of Linear Assignment, one easily convince himselfs that an equivalent instance (i.e. an instance with identical cost function, up to an overall constant) can be produced, with $w_{ii} = 0$ for each $i \le 2n$, and that the values $w_{2i,2i-1}$ and $w_{2i-1,2i}$ never appear in allowed matchings. For this reason, we will assume in the following that $w_{ii} = 0$, and denote the four elements in the n diagonal blocks of size 2 with special symbols $*$ and $\cdot$, instead that with a weight value. For

example, an instance with $2n = 6$ could be

$$W = \left(\begin{array}{cccccc} * & \cdot & 2 & 4 & 1 & 1 \\ \cdot & * & 1 & 7 & 8 & 2 \\ 3 & 3 & * & \cdot & 6 & 3 \\ 4 & 2 & \cdot & * & 2 & 5 \\ 5 & 5 & 1 & 6 & * & \cdot \\ 4 & 9 & 8 & 1 & \cdot & * \end{array}\right) \text{ with } k = 10;$$

The choice of representation with $*$s is done for mnemonic reasons: at sight, one knows that a valid matching should use exactly one $*$ per block. For example, a valid matching with weight 9 could be the following (on the right side, elements i such that $\pi(i) = i$ are underlined in order to highlight the one-in-two constraint satisfaction)

$$\left(\begin{array}{cccccc} * & \cdot & \mathbf{2} & 4 & 1 & 1 \\ \cdot & \mathbf{*} & 1 & 7 & 8 & 2 \\ 3 & 3 & * & \cdot & 6 & \mathbf{3} \\ 4 & 2 & \cdot & \mathbf{*} & 2 & 5 \\ 5 & 5 & 1 & 6 & \mathbf{*} & \cdot \\ \mathbf{4} & 9 & 8 & 1 & \cdot & * \end{array}\right) \qquad \begin{array}{l} \{\pi(i)\}_{i=1,\ldots,6} = \{3, \underline{2}; 6, \underline{4}; \underline{5}, 1\}\,; \\ C(\pi) = 9. \end{array}$$

Clearly, a one-in-two matching can be described by a choice of the elements kept fixed by the permutation (i.e. the $*$ chosen in each block), times an allowed choice of assignment in the n-dimensional minor matrix resulting from the removal of the fixed rows and columns. Thus, allowed matchings are in bijection with pairs $(\vec{\sigma}, \pi)$, where $\vec{\sigma} \in \{0,1\}^n$ and $\pi \in \mathcal{S}_n$, with all $\pi(i) \neq i$. For example, the previous configuration could be described as $(\vec{\sigma}, \{\pi(i)\}_{i=1,\ldots,3}) = ((0,0,1), \{2,3,1\})$. This fact suggests a naive interpretation for the potential hardness of this variant of Assignment: for any choice of fixed elements (the $*$s), the problem of finding the optimal assignment is polynomial (just put $+\infty$ on the remaining $*$s, and use Hungarian Algorithm on the resulting Linear Assignment instance), nonetheless one should perform a search among these 2^n possible choices, which, in absence of a sufficiently strong correlation or a skill mathematical structure, could make the search exponential in size.

8.5 Proof of linear reduction from Boolean Satisfiability problems

Here we prove that also One-in-Two Matching, less general w.r.t. the analogue One-in-Two Assignment, allows for linear reduction from arbitrary instances of SAT, 3-SAT or NAE-3-SAT Problems.

A SAT (or NAE-SAT) instance with n literals and m clauses can be encoded into a bipartite graph $G(V_\ell, V_c; E)$, with V_ℓ being the set of literals $\{u_i\}_{i=1,\dots,n}$ and V_c the set of clauses $\{C_a\}_{a=1,\dots,m}$, and a map $s : E \to \{\pm 1\}$ which states whether the literal enters negated or not, i.e. if $u_i \in C_a$ then $s(i,a) = +1$, while if $\overline{u}_i \in C_a$ then $s(i,a) = -1$.

It is customary to graphically represent a Satisfiability instance by mean of a factor graph, i.e. the bipartite graph above, where "literal" vertices in V_ℓ are denoted by small circles, and "clause" vertices in V_c by small squares, and an edge is drawn in solid line if $s(e) = +1$, and dashed if $s(e) = -1$. The instance is satisfied by a given boolean assignment if for each clause vertex there is at least one solid-edge true neighbour or one dashed-edge false neighbour. In NAE-SAT the same as above holds, but also not all the neighbours must satisfy the clause.

In order to perform our reduction, it is easier to first perform a decoration on the graph: for each variable, introduce an auxiliary variable per incident edge (in a sense, the literal "as seen from the clause"), then substitute the original variable node by a "consistency check" clause (drawn as a small triangle), which ensures that the boolean values on the copies of the variable coincide. A small example of SAT factor graph, and the corresponding decorated graph, could be the following: In our reduction, we have two 2×2 blocks $\left(\begin{smallmatrix} * & \cdot \\ \cdot & * \end{smallmatrix}\right)$ per edge $(i,a) \in E(G)$, and thus the entries of the matrix W are labeled by an index $(i,a)^{\pm}_{1,2}$, where $(1,2)$ stands for the first and second block, and $\pm$ stands for the first and second index inside the block. Suppose to order arbitrarily the edges $e_\alpha = (i_\alpha, a_\alpha)$, then for helping visualization, we will assume that the entries of W are ordered as

$$\begin{aligned}((i_1,a_1)^+_1, (i_1,a_1)^-_1, \dots, (i_k,a_k)^+_1, (i_k,a_k)^-_1,\\ (i_1,a_1)^+_2, (i_1,a_1)^-_2, \dots, (i_k,a_k)^+_2, (i_k,a_k)^-_2)\end{aligned} \tag{8.7}$$

that is, first all the 1-blocks, then, in the same order, all the 2-blocks.

The choose of $*$s in the matching correspond to the sequence of boolean assignments for the literals, "as they are seen from the clause",

i.e. for the literals in the decorated instance. So we need a "truth-setting" structure, which ensures that all these values coincide (in other words, implements the "consistency check" clause), and a "satisfaction-testing" structure, which checks that each boolean clause in the original formula is satisfied. The truth-setting structure is encoded in the set of 1s in the entry pairs with $i = j$, while the satisfaction-testing structure is encoded in the set of 1s in the entry pairs with $a = b$. More precisely, entries $w = 1$ can appear in the off-block matrix elements at index pairs $((i,a)^{\pm}_{\alpha}, (j,b)^{\pm}_{\beta})$ only in one of the two cases:

$i = j$, $\alpha = 2$ and $\beta = 1$ (truth-setting structure)
$a = b$, $\alpha = 1$ and $\beta = 2$ (satisfaction-testing structure)

We start describing the truth-setting structures. For each variable i, call $A(i)$ the set of adjacent clauses, and choose an arbitrary cyclic ordering on this set. For $i \in V_\ell$ and $a, b \in A(i)$ we state

$$w_{(i,a)^{\sigma}_{\alpha},(i,b)^{\tau}_{\beta}} = \begin{cases} 1 & a = b, \quad \sigma = \tau = -, \quad \alpha = 2, \quad \beta = 1; \\ & a = b - 1, \quad \sigma = \tau = +, \quad \alpha = 2, \quad \beta = 1; \\ 0 & \text{otherwise.} \end{cases} \tag{8.8}$$

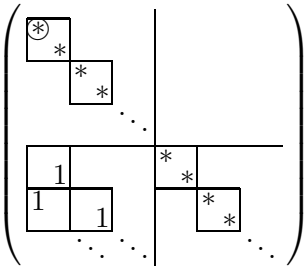
that is, the minor of W restricted to indices with fixed i looks like

$$W|_{\text{fixed } i} = \left(\begin{array}{c|c} I_* & 0 \\ \hline W' & I_* \end{array}\right);$$

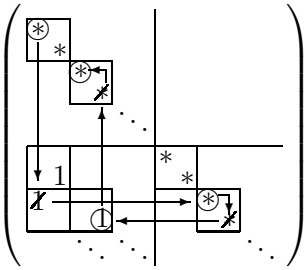
$$I_* = \begin{pmatrix} \begin{smallmatrix} * & \cdot \\ \cdot & * \end{smallmatrix} & 0 & 0 & \dots & 0 \\ 0 & \begin{smallmatrix} * & \cdot \\ \cdot & * \end{smallmatrix} & 0 & & 0 \\ 0 & 0 & \begin{smallmatrix} * & \cdot \\ \cdot & * \end{smallmatrix} & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \begin{smallmatrix} * & \cdot \\ \cdot & * \end{smallmatrix} \end{pmatrix} \qquad W' = \begin{pmatrix} \begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix} & 0 & 0 & \dots & \begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix} \\ \begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix} & \begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix} & 0 & & 0 \\ 0 & \begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix} & \begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix} & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots\dots & \begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix} & \begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix} & \end{pmatrix}. \tag{8.9}$$

Remark that in all rows with sub-index 2 and columns with sub-index 1 (i.e. all rows ands columns in W') we have exactly one allowed entry beyond the $*$ element on the diagonal. In order to see how the truth-setting procedure works, consider what happens if we choose the top-left $*$ in the first block (i.e. we choose "$\sigma_{(i,a_1)} = 1$" in the string of $\vec{\sigma}_{(i,a)}$ for the "variables seen from the clauses"). At the beginning the matrix is (a circle means "element chosen in the matching", a bar means "element

not choosen in the matching")

$W =$ 

while, after six logical implication we have

$W =$ 

Remarking that the choice of 1s in the matrix has a cyclic ordering w.r.t. clause indices, with a simple induction one deduces that the boolean assignments are "$\sigma_{i,a_k} = 1$" for all clauses a_k incident with variable i. A similar statement can be done in the case "$\sigma_{i,a_1} = 0$" (just exchange circles with bars in the diagrams above). So, we have determined that only consistent boolean assignment of literals are allowed.

Now we can build the structures for the satisfaction testing. We are interested in the restriction of W to a given fixed index a, which is of the form

$$W|_{\text{fixed } a} = \left(\begin{array}{c|c} I_* & W'' \\ \hline 0 & I_* \end{array}\right) . \tag{8.10}$$

We describe the matrix W'' for a clause of length k with all unnegated literals, $C_a = u_{i_1} \vee \ldots \vee u_{i_k}$. All the other cases can be trivially inferred [3].

[3]Define the transposition matrix $(T^{(k)})_{ij} = 1$ if $i = j \notin \{2k, 2k-1\}$, if $i = 2k$ and $j = 2k-1$ or if $i = 2k-1$ and $j = 2k$, and zero otherwise. I.e., matrix $T^{(k)}$, acting on the left, transpose rows $2k-1$ and $2k$, while acting on the right transpose the corresponding columns. Then, if matrix W'' encodes a clause involving $(u_1, \ldots, u_k, \ldots, u_\ell)$, the matrix $T^{(k)} W'' T^{(k)}$ encodes the same clause on literals $(u_1, \ldots, \overline{u}_k, \ldots, u_\ell)$.

Choose a whatever literal index (say, i_1) among the neighbours of the clause a, and set

$$w_{(i,a)_1^\sigma,(j,a)_2^\tau} = \begin{cases} & \text{if } i = j \text{ and } \sigma = \tau = -1; \\ 1 & \text{if } i = j \neq i_1 \text{ and } \sigma = \tau = +1; \\ & \text{if } i = i_1,\ j \neq i_1,\ \sigma = +1 \text{ and } \tau = -1; \\ & \text{if } i \neq i_1,\ j = i_1,\ \sigma = -1 \text{ and } \tau = +1; \\ 0 & \text{otherwise} \end{cases} \tag{8.11}$$

that is, in an extensive representation of the matrix,

$$W'' = \left(\begin{array}{cc|cccc|c} 0 & 0 & 0 & 1 & 0 & 1 & \dots \\ 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & \\ 1 & 0 & 0 & 0 & 0 & 1 & \\ \vdots & \vdots & \vdots & \vdots & & & \ddots \end{array}\right) \tag{8.12}$$

which indeed makes the game, as one can easily check. Indeed, if all literals are negated, we are left with matrix minor

$$W''_{(\mathrm{F},\mathrm{F},\dots,\mathrm{F})} = \left(\begin{array}{c|c} 0 & \mathbf{0} \\ \hline \mathbf{0} & I_{k-1} \end{array}\right),$$

which clearly does not allow for any valid matching, while, if the first literal is true we have

$$W''_{(\mathrm{T},\dots)} = \left(\begin{array}{c|c} 1 & \cdot \\ \hline \cdot & I_{k-1} \end{array}\right),$$

which allows at least for the diagonal matching, $\pi(j) = j$ for all $j = 1, \dots, k$, and if the first literal is false, but one of the others (say, the h-th) is true, we have

$$W''_{(\mathrm{F},\dots,\mathrm{T},\dots)} = \left(\begin{array}{c|ccc} 0 & \cdot & 1 & \cdot \\ \hline \cdot & I_{h-2} & \mathbf{0} & 0 \\ 1 & \mathbf{0} & 1 & \mathbf{0} \\ \cdot & 0 & \mathbf{0} & I_{k-h} \end{array}\right),$$

which allows at least for the matching $\pi(1) = h$, $\pi(h) = 1$ and $\pi(j) = j$ otherwise. Similarly, for a NAE-k-SAT clause $C_a = (u_{i_1} \vee \dots \vee u_{i_k}) \wedge$

$(\overline{u}_{i_1} \vee \ldots \vee \overline{u}_{i_k})$ we can choose

$$w_{(i,a)_1^\sigma,(j,a)_2^\tau} = \begin{cases} 1 & \begin{array}{l} \text{if } i = j \neq i_k \text{ and } \sigma = \tau = -1; \\ \text{if } i = j \neq i_1 \text{ and } \sigma = \tau = +1; \\ \text{if } i = i_1,\ j \neq i_1,\ \sigma = +1 \text{ and } \tau = -1; \\ \text{if } i \neq i_1,\ j = i_1,\ \sigma = -1 \text{ and } \tau = +1; \\ \text{if } i = i_k,\ j \neq i_k,\ \sigma = -1 \text{ and } \tau = +1; \\ \text{if } i \neq i_k,\ j = i_k,\ \sigma = +1 \text{ and } \tau = -1; \end{array} \\ 0 & \text{otherwise} \end{cases} \tag{8.13}$$

corresponding to the matrix minor

$$W'' = \left(\begin{array}{cc|cc|cc|ccc|cc} 0 & 0 & 0 & 1 & 0 & 1 & \ldots & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & & \ddots & \vdots & \vdots \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & \cdots & 0 & 0 \end{array}\right) \tag{8.14}$$

This completes the reduction proof for SAT and NAE-SAT (and then, in particular, for 3-SAT and NAE-3-SAT). Indeed, the proposed encodings of a SAT and NAE-SAT clause are a special case of the most general k-literal clause, in which the number of true literals must be in the range $\{h_{\min}, \ldots, h_{\max}\}$, (also having as a special case the 1-in-3-SAT problem, which is NP-complete [18]), for which a general encoding is possible

It is common in reduction proofs that multiple appearence of a literal in a clause requires some care (truth-setting and satisfaction-testing structures could damagely interfere), and one should make some standard comment on the fact that this case can be excluded with small effort from any SAT instance. This does not happen in our case. One can understand this from the fact that the 1s in the two structures appear in different rows and columns, and logical implications which allow to test the performance of the structures involve only these rows/columns.

Also remark that the reduction is linear not only in matrix size w.r.t. the original factor-graph size (the dimension n of the Matching matrix is four times the number of edges in the factor graph), but also in the number $\mathcal{N}$ of non-zero entries in the matrix, which is indeed very

sparse. Each literal of coordination k requires a truth-setting structure with $2k$ entries, while each clause of length k requires a satisfaction-testing structure with $4k-3$ entries ($6k-8$ for a NAE clause), thus we have that, for a factor graph $G(V_\ell, V_c; E)$

$$\mathcal{N} = 6|E(G)| - 3|V_c(G)| \qquad \text{SAT problem;}$$
$$\mathcal{N} = 8|E(G)| - 8|V_c(G)| \qquad \text{NAE-SAT problem.}$$

Finally, we also remark that the instances of One-in-Two Matching obtained as reduction from Satisfiability Problems are of a particular kind: the 1s are contained only in the top-right and bottom-left $2n \times 2n$ quadrants (the matrices W' and W''). We call *Bipartite* One-in-Two Matching this specialized problem.

8.6 Proof of linear reduction from One-in-Two Matching to 3 Dimensional Matching

Now we describe the reduction from One-in-Two Matching to 3DM. An identical reduction goes from One-in-Two Assignment to Numerical 3DM. Call $W = \{w_{ij}\}$ our One-in-Two Matching instance of dimension $2n$ to be encoded, with

$$w_{ij} = \begin{cases} * & i = j \leq 2n; \\ \cdot & \begin{array}{l} i = 2h,\ j = 2h-1,\ h \leq n \\ \text{or } i = 2h-1,\ j = 2h,\ h \leq n; \end{array} \\ w_{ij} \quad (\in \{0,1\}) & \text{otherwise.} \end{cases} \tag{8.15}$$

and $W^{(3)}$ our suggested output 3DM instance, also of dimension $2n$. Our formal reduction is, calling $A(k) = \{2(k-n)-1, 2(k-n)\}$ for $k = n+1, \ldots, 2n$,

$$w^{(3)}_{ijk} = \begin{cases} 1 & \begin{array}{l} i = j = 2k-1, \\ i = j = 2k; \end{array} \\ w_{ij} & i \in A(k),\ j \notin A(k); \\ 0 & \text{otherwise.} \end{cases} \tag{8.16}$$

or, more pictorially, call $\vec{e}_i$ and $\vec{w}_i$ the vectors

$$\vec{e}_i = (0, \ldots, 0, \overset{i\text{-th}}{1}, 0, \ldots, 0)\,; \tag{8.17a}$$
$$\vec{w}_{2i-1} = (w_{2i-1,1}, \ldots, w_{2i-1,2i-2}, 0, 0, w_{2i-1,2i+1}, \ldots, w_{2i-1,2n})\,; \tag{8.17b}$$
$$\vec{w}_{2i} = (w_{2i,1} \quad, \ldots, w_{2i,2i-2} \quad, 0, 0, w_{2i,2i+1} \quad, \ldots, w_{2i,2n} \quad)\,; \tag{8.17c}$$

then $W^{(3)}$, written as a matrix on indices (i,k), of vectors on index j, looks like

$$W^{(3)} = \begin{pmatrix} \vec{e}_1 & 0 & \dots & \vec{w}_1 & 0 & \dots \\ \vec{e}_2 & 0 & \dots & \vec{w}_2 & 0 & \dots \\ 0 & \vec{e}_3 & \dots & 0 & \vec{w}_3 & \dots \\ 0 & \vec{e}_4 & \dots & 0 & \vec{w}_4 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{pmatrix} . \tag{8.18}$$

Indeed, remark that in the planes (i,j) for $k = 1, \dots, n$ there are only two allowed entries, whose (i,j) coordinates correspond to the ones of the $*$s in the k-th block of the original instance. Mimicking the One-in-Two constraint, for each block k we are forced to choose a value $\sigma_k \in \{0,1\}$, with (say) 0 and 1 selecting respectively the entry with even and odd indices. Then, in all (i,j) layers with $k = n+1, \dots, n$ there are only two non-empty i-rows, (which are empty in all the other layers with $k = n+1, \dots, 2n$). Because of the choice of the vector $\vec{\sigma}$, exactly one of them is now forbidden. So, the three-dimensional constraint of choosing one element per index i, j and k is at this point reduced to a traditional two-dimensional matching constraint, as there is a bijection between unmatched layers k and non-empty unmatched rows i. I.e. the 3-dimensional $2n \times 2n \times 2n$ array of equation (8.18) is now restricted to the $n \times n \times n$ array

$$W'^{(3)}(\vec{\sigma}) = \begin{pmatrix} \vec{w}^{(\vec{\sigma})}_{2-\sigma(1)} & 0 & 0 & \dots \\ 0 & \vec{w}^{(\vec{\sigma})}_{4-\sigma(2)} & 0 & \dots \\ 0 & 0 & \vec{w}^{(\vec{\sigma})}_{6-\sigma(3)} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} ; \tag{8.19}$$

$$\left(\vec{w}_i^{(\vec{\sigma})}\right)_j = \left(\vec{w}_i\right)_{2j-\sigma(j)} . \tag{8.20}$$

It is easily understood that the set of forbidden indices j after the choice of vector $\vec{\sigma}$ and the disposition of the entries w_{ij} are in accord with the picture of section 8.4.3, where a vector $\vec{\sigma}$ determines a n-dimensional minor of the original $2n$-dimensional instance, with forbidden entries on the diagonal.

As a corollary of the construction we have that the reduction is linear not only for what concerns the size of the array (an $n \times n$ One-in-Two instance goes into an $n \times n \times n$ 3DM array), but also on the number of non-zero entries in the instance [4], which is proportional to the length of

[4] More precisely, the cardinality of T for 3DM equals the one of E for One-in-Two

the bit-encoding of the instance.

Putting together this result with the one of section 8.5, we have in turn a *linear* reduction from SAT problems to 3DM,

$$\mathcal{N}_{\mathrm{3DM}} = 10|E(G)| - 3|V_c(G)| \qquad \text{SAT problem;}$$
$$\mathcal{N}_{\mathrm{3DM}} = 12|E(G)| - 8|V_c(G)| \qquad \text{NAE-SAT problem.}$$

which is much more economic of the cubic one first presented in Karp seminal '72 paper [23] (see also [18]), and maybe (depending from the tastes) technically simpler.

Matching, plus the $2n$ "deterministic" entries of vectors $\vec{e}_i$, i.e. inside a factor 2 if we understand that 3DM instances having planes with only one valid entry can be trivially reduced in size.

Chapter 9

Conclusions

In this thesis we discussed the relation between optimization problems and the physics of disordered systems, and in particular of cavity methods.

In chapter 6 we discussed a classical disordered systems problem: the spin glass model. In particular we concentrate on the ground-state (finding it is an *NP-Complete* problem) and on the determination of its energy. We propose an algorithm that makes use of cavity fields to determine an upper and a lower bound for the energy of such a system. The algorithm we gave has been implemented and studied on this model, anyway it can be extended to a larger class of disordered systems.

We presented in chapter 7 an algorithm to solve a classical (and also useful) problem in Combinatorial Optimization: the Assignment. The algorithm we presented uses the cavity fields and equations to work. This algorithm has the advantage to be easier to implement on hardware device than the well-known (and well-performing) Hungarian Algorithm. Surprisingly this algorithm converge almost always to the optimal solution for the assignment problem. In chapter 7 we find a bound on the convergence time.

In chapter 8 we presented some problems related to the Assignment problem. Many of them also if slightly differently stated are *NP-complete*. For the one-in-two problem the *NP-complteness* is proven. The one-in-two problem is presented here because of the low-dimensional geometry of the space where it can be defined, because of the small-encoding of reduction to and from other hard problems, and because of the locality of its interactions. The reduction is relatively simple and specially compact. As a side result, combination with a reduction proof

from one-in-two Matching to 3-Dimensional Matching (3DM) provides a simple reduction proof for 3DM.

We worked always on cavity fields and equations at zero temperature, this approach allows to simplify the discussion of the cavity method in several aspects and allows to concentrate on the ground state of the problems. When there exists many local ground states one of the aspects that make hard to find the global one is the crossing between their energies when adding a new variable. When working in this regime the physics of disordered systems (replica simmetry and its breaking) comes into play.

Further directions of investigation will include

- A deeper comprehension of the behaviour of the algorithm used in chapter 6 to find lower-bounds to Ising Spin Glasses Ground-State energy.
- The extension of the algorithm introduced in chapter 6 to other optimization problems.
- A study of the efficency of the cavity algorithm for the assignment when used to find approximated solutions.
- A statistical investigation of the properties of the random version of one-in-two problem.
- An investigation of One-In-Two problem via cavity method techniques also in relation to the important question of determining the geometrical characteristics of the hard instances of the Traveling Salesman Problem.

The approach to combinatorial optimization via the physics of cavity fields have many applications, we expect that these methods allows further investigations, both for a deeper comprehension of the nature of complexity and for the design of faster algorithms, both exact and approximated.

Appendix A

Graph Theory

Graph Theory is a branch of mathematics, its origin probably date back to Euler's work on the Königsberg bridges problem in 1766. Graph theory can be used to solve optimization problems on networks.

A *graph* G is an ordered pair $(V(G), E(G))$ where $V(G)$ and $E(G)$ are disjoint sets called the *vertices* and *edges* of G respectively, together with an incidence function f which associates an unordered pair of vertices $\{u, v\}$ with each edge e of G.

We will restrict to "proper" graphs (graphs such that the endpoints of each edge are distinct).

Graphs are often represented drawing in the plane vertices represented by points and edges represented by lines between the two points corresponding to the endpoints.

Another representation is through the incidence matrix $M(G)$: a square matrix with $N = |V|$ rows, such that the entry $M_{i,j}$ is one if there is an edge incident both on i and j, and is zero otherwise.

- An *undirected graph* $G = (V, E)$ is given by its vertices $i \in V$ and its undirected edges $i, g \in E$
- The number $|V|$ of vertices (*nodes*) is the order and $|E|$ the number of edges (*arcs*) is the size of the graph.
- Two vertices are *adjacent* (*neighboring*) if $i, j \in E$.
- The edge i, j is *incident* on its ends i and j.
- The *degree* (coordination) of a vertex is the number of adjacent vertices. The vertices of zero degree are called *isolated*, a vertex with degree one is called *leaf*.

- A graph $G' = (V', E')$ is a *subgraph* of G if $V' \subset V$ and $E' \subset E$.
- A *walk* is a sequence of edges $i_0, i_1, i_1, i_2, \ldots, i_{l-1}, i_l$. A *path* is a self avoiding walk.
- A path with $i_0 = i_l$ is called a *cycle*.
- A walk with $i_0 = i_l$ is called *circuit*.
- A graph G is *connected* if for every pairs i, j of vertices in V there is at least a path with endpoints in i and j.
- A graph with no cycles is a *forest*.
- A *tree* is a forest consisting of a single connected component. A forest is the union of disconnected trees. For a tree $|V| = |E| + 1$ holds. For a forest $|V| = |E|+$ number of connected components.
- Given a path or a walk, the number of edges is its length.
- The *girth* of a graph is the length of the shortest cycle contained in the graph.
- The *complete* graph of vertices V is the graph such that for all pairs i, j the edge $i, j \in E$
- A graph is *weighted* if there is a function that associates a weight to every edge.
- A graph G is *bipartite* with bipartition $\{X, Y\}$ if $\{X, Y\}$ is a partition of the vertices $V(G)$ and all the edges of G join vertices of X to vertices of Y.
- A circuit that uses every edge exactly once is an *Eulerian circuit*.
- A circuit that visits every vertex exactly once is an *Hamiltonian cycle*.
- A *directed graph* $G = (V, E)$ is given by its vertices $i \in V$ and its directed edges $i, j \in V \times V$.
- A vertex i has an *outdegree* equals to the number of outgoing edges (i, j) and an *incoming degree* equals to the number of ingoing edges (j, i).

- A directed graph D is strongly connected if, for each ordered pair of vertices u and v of D, there is a directed walk from u to v in D.

- Let e be an edge of a graph G. The e is a *bridge* of G if $G \setminus e$ has more connected components than G.

  Let e be a bridge of a connected graph G and let u and v be the end vertices of e. Then $G \setminus e$ has exactly two components H_1 and H_2 with $u \in V(H_1)$ and $v \in V(H_1)$. Let e be an edge of a graph G. Then e is a bridge if and only if e is not contained in any cycle of G. Remark also that a connected graph is a tree if and only if every edge of G is a bridge. Let T be a tree then all edges are bridges.

Some further remarks: let G_1, G_2 be connected subgraphs of a graph G such that $V(G_1) \cap V(G_2) \neq \emptyset$. Then $G_1 \cup G_2$ is connected. Similarly, let D_1, D_2 be strongly connected subdigraphs of a digraphs D such that $V(D_1) \cap V(D_2) \neq \emptyset$. Then $D_1 \cup D_2$ is strongly connected. Let G be a graph and u, v be distinct vertices of G. If G has a walk from u to v then G has a path from u to v. In every graph G we have $\sum_{v \in V(G)} d(v) = 2|E(G)|$.

A.0.1 Cayley Graphs

Often we need to work on graphs with large girth, as for example happens in information theory. Then an important question is whether large *girth* graphs with high connectivity exists. The answer is yes, here we will explain how to construct some high connectivity large girth graphs.

We will see another example of large girth graphs in the context of Random Graphs in appendix B.

A special class of graphs are the Cayley graphs. To define them we only need a group and a subset of elements of this group.

Consider a group $\mathcal{G}$ with a given operation $+$. Consider a subset $\mathcal{S} \subset \mathcal{G}$. The directed Cayley Graph $C(\mathcal{G}, \mathcal{S})$ has a vertex for each element of the group $\mathcal{G}$ and a directed edge joining two vertices v w if exists s in $\mathcal{S}$ such that $w = v+s$. If the set $\mathcal{S}$ is symmetric (for each element $s \in \mathcal{S}$ also $s^{-1} \in \mathcal{S}$) then for every directed edge (v, w) in the Cayley graph there is (w, v) and then we can define the undirected Cayley Graph replacing each couple of directed edges (v, w), (w, v) with a single undirected edge.

Remark that a large girth Caley Graph has o be defined through a non-abelian group $\mathcal{G}$, in fact an abelian group gives Cayley graphs with girth four.

A.1 Basic Graph Algorithms

A.1.1 Euler Tours

A graph G is eulerian (admits an *Eulerian circuit*) if and only if it is connected and all its vertices have even degree. This statement can be shown by an algorithm for the construction of Euler's tour.

Consider a graph with even degree on all vertices. Choose arbitrarly a starting point i_s, then construct an auxiliar circuit $\mathcal{C}$ by increasing a walk passing through the starting vertex. If the circuit does not passes through each arc then on the circuit there are some non-saturated vertices (vertices such that not all the incident edges are in the circuit).

We choose arbitrarly one of such non-satuarated vertices i_v on the circuit as starting point for the search of a circuit $\widetilde{\mathcal{C}}$ on the graph $G \setminus G'$ obtained by deleting the edges in $\mathcal{C}$. Then we use as auxiliar circuit the circuit given by a walk $W_{\mathcal{C}} \subset \mathcal{C}$ going from i_s to $i_v \cup \widetilde{\mathcal{C}} \cup (\mathcal{C} \setminus W_{\mathcal{C}})$.

It is easy to see that this iterative algorithm ends (in a finite time) with an *eulerian circuit*.

Remark that a circuit has degree even on all the vertices. Then if we delete a circuit from a given graph we do not change the parity of the degree of its edges. Then a graph with at least one vertex with odd degree cannot admit an eulerian circuit.

A.1.2 Shortest path

Consider an undirected network N in which all edges have non-negative weights. We will describe the Dijkstra's Algorithm: an algorithm for finding the shortest paths in N from a given vertex to every vertex. Let P be a path; $w(P)$ is the sum of the weights of the edges of P.

We start with a tree T_1 that contains only the starting site x_1. In the i-th step we have a tree T_i with vertices $\{x_1, x_2, \ldots, x_i\}$ and a function $d_{T_i}(x_1, x)$ of the vertices $x \in T_i$ (This is the distance between x_1 and x on the tree T_i). We grow at each step the tree by choosing the edge x, y in the graph from T_i to $G \setminus T_i$ such that $d_{T_i}(v, x) + w(x, y)$ is as small as possible.

It easy to show, by induction, that this algorithm works (remark that the proof uses the fact that the edges have positive weight): at each step the best walk from x_1 to an element of T_i is in T_i.

A.1.3 Maximum flows in networks

Consider a directed graph D with weights on the edges non-negative integer. Given a vertex v of D, we denote the set of arcs leaving v by $a_D^+(v)$ and the set of arcs entering v by $A_D^-(v)$. Associate to each edge a weight $c(e)$ called its capacity. and two special points x and y.

Suppose f is a given function that associates a non-negative integer with each arc of D. For each vertex define the outgoing flow f^+ associated to f: $f^+(v) = \sum_{e \in A_D^+(v)} f(e)$ and $f^+(v) = \sum_{e \in A_D^+(v)} f(e)$ the incoming flow. We say that f is a flow in D if:

- $0 \leq f(e) \leq c(e)$ for all edges e.
- $f^+(v) = f^-(v)$ for all vertices in $V(D) \setminus x, y$.

The value of a given flow f is $f^+(x) - f^-(x)$.

In a pictorical representation the capacity of an edge can represent the capacity of a pipe between the two endpoints. We have a spring in x and a sink in y. We look for the flow such to maximize the quantity of fluid we can move from x to y.

Consider the problem of finding a flow with maximum value.

Given a subset U of D we consider the arc-cut A_D^+. Its capacity is $c^+(U) = \sum_{e \in A_D^+} c(e)$. Similarly $c^-(U) = \sum_{e \in A_D^-} c(e)$; $f^+(U) = \sum_{e \in A_D^+} f(e)$ and $f^-(U) = \sum_{e \in A_D^-} f(e)$. Remark that if U is such that x is in U and y is in U's complement then the value of a flow f is equal to $f^+(U) - f^-(U)$. It follows that the value of every flow is smaller of equal of $c^+(U)$. Then if we have a flow f and a set such that the value of f is equal to $c^+(U)$ then f is optimal.
Given a flow f, an f-unsaturated path is a path P satisfying the following two conditions:

- for each forward directed arc e of P, $f(e) < c(e)$, and
- for each backward directed arc of P $f(e) > 0$

If f contains an unsaturated path from x to y then there is a g flow with value bigger then f.

To solve this problem Ford and Fulkerson gave an algorithm in 1956. The algorithm starts from a flow in D, then at each step construct a unsaturated flow tree rooted on x untill it touches y (and then another flow with higher value is constructed) or gives a set U such that $c^+(U)$ is equal to the value of f (then a certificate of optimality is given)

Remark that the time needed by the so given Ford-Folkerson algorithm can be equal to the maximum value of the capacity, then this algorithm is not a polynomial one. There is a refinment of the construction of the unsaturated flow tree such that the algorithm is strongly polynomial. Something similar happens also for the Assignment [16].

A.1.4 Matching

Given a graph G a matching M is a subset of the edges of G such that no two edges of M have a common end-vertex.

A maximum matching is a matching with maximal cardinality over all matchings in G.

Let M be a matching in a graph G. An M-alternating path in G is a path whose edges alternate between M and $E(G) \setminus M$. An *M-augmenting path* in G is an M-alternating path whose end vertices are M-unsatured.

Let M be a matching in a graph G. Then M is a maximum matching in G if and only if G has no M-augmenting path.

When the graph is bipartite we refer to the matching problem as the assignment problem. On the assignment problem there is a special way to modify the weights of the problem so to have as the result of the iterative steps of an algorithm spanning subgraphs of zero-weight edges. We refer to these subgraphs as zeroes subgraphs. The optimal assignment problem is solved by the iteration of two steps: the construction of bigger zeroes graph as possible using Egerváry's theorem and the search for an optimal matching in the zeroes graph using Kőnig's theorem. For more details see chapter 7.

Appendix B

Random Graphs

Let n be a positive integer, $0 \leq p \leq 1$. The random graph $G(n,p)$ is a probability space over the set of graphs on the vertex set $\{1,\ldots,n\}$ where the probability for every edge $\{i,j\}$ to be in G is p and these probabilities are mutully independent. This kind of graphs has been studied and introduced by Erdos and Rényi in 1960 ([14]).

It is useful to introduce a dynamic model for the construction of random graphs. Let $x_{i,j}$ be a set of i.i.d. random variables drawn uniformly in $[0,1]$ associated to the edges of the complete graph with n vertices. Now let p go from 0 to 1. Fixed a real p the edges such that $p < x_{i,j}$ are turned-off and the other ones are turned-on. As p increases the graph evolves from empty to full.

Another set of random graphs is often defined: $G(n,e)$, that is the probability measure such that only the graphs with exactly e edges have non-zero probability and they have all the same probability. Also for this set is possible to define an evolution model: start with the empty graph and randomly add one edge until the graph become full.

Generally $G(n,e)$ and $G(n,p)$ with $p = e/\binom{n}{2}$ have several common features.

Given a property A there is a probability that $G(n,p)$ satisfies A. We say that A is monotone if the probabilty for a graph in $G(n,p)$ to satisfy A is a monotone function in p. As an example we can consider the event 'G is triangle free'. Let X be the number of triangles of a graph, the expectation is $E[X] = \binom{n}{3}p^3$ and then in the limit $n \to \infty$

$$\lim_{n\to\infty} E[X] = (np)^3/6$$

Asymptotically the distribution of X is poissonian, so the limit for n

large for the probability to realize the event A is $e^{-(np)^3/6}$. Remark that when $p \ll 1/n$ this probability is almost 1 and when $p \gg 1/n$ the probability is almost 0.

We say that a function $r(n)$ is a threshold function, meaning that

- when $p(n) \ll r(n) \lim_{n\to\infty}$ of the Probability for A to be satisfied by the random graph $\mathcal{G}(n,p)$ is zero.
- when $p(n) \gg r(n) \lim_{n\to\infty}$ of the Probability for A to be satisfied by the random graph $\mathcal{G}(n,p)$ is one.

In our example $r(n) = 1/n$ is a threshold function but also $10/n$ is a threshold function: the threshold function, when it exists is not unique.

We could approach the problem of triangle freeness by considering every set S of three vertices. Let B_S the event that S is a triangle, then Prob$[B_S]= p^3$. If the B_S where mutually independent we would have thet the probability to do not have triangles in $G(n,p)$ is $e^{-(np)^3/6}$. In reality, B_s and B_T are mutually independente only if the edges e_S connecting the three vertices S are disjoint by e_T. This situation appears often in the study of random graphs,

B.1 Small Subgraphs

Let H be a graph with v vertices and e edges. Let $\rho(H) := e/v$ be the density of H. We call H balanced if every subgraph H' has $\rho(H') \leq \rho(H)$. We call H strictly balanced if every proper subgraph H' has $\rho(H') < \rho(H)$. Then cliques and loops are strictly balanced.

Theorem 5 *Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph of G. Then $p = n^{-v/e}$ is the threshold function for A.*

Let X_S be the indicator the random variable for A_S, then the indicator X for the total number of appearence of the graph H in G is the sum over all the choices of v vertices of X_S

$$X = \sum_{|S|=v} X_S$$

The expectation of X is, by linearity $E[X] = O(n^v p^e)$. If $p << n^{-v/e}$ then $E[X] = o(1)$ and so $X = 0$ almost surely (in general $Pr[X >$

$0] \leq E[X]$ holds). If $p >> n^{-v/e}$ then $E[X] \to \infty$ (If $E[X] \to \infty$ and $\Delta^* = o(E[X])$ then $X > 0$ almost always and $X \sim E[X]$ almost always; where $\Delta^* = \sum_{Correlated\{i,j\}} Pr[A_j|A_i]$ and the variables are symmetric: There is automophsm that send event in A_i in A_j). □

As a Corollary it is easy to show that if H is not balanced then $p = n^{-v/e}$.

B.2 Optimization Problems on Random Graphs

As we said in section 4.3 often we are interested in the behaviour of algorithms on random instances. A lot of problems are defined on graphs (weighted or not) then we need a measure for the weights and for the graphs.

It is useful to study Optimization Problems on different set of random ensembles. A natural set which has been often studied is the set of Erdős Rényi random graphs, since it is a big set of graphs and its edges are almost independent.

Moreover Erdős Rényi random graphs have the remarkable feature to be large girth graphs. This feature is very useful when we want to test the validity of cavity methods that are exacts on trees.

Many different problems have been studied on different kind of samples, one example is the *coloring problem*. Given a graph and q colors the coloring problem consists in determining the way to color the vertices of the graph that minimizes the number of neighbour vertices with the same color. The graph is called q-colorable if exists a coloring such that the number of neighbour vertices with the same color is zero.

The Hamiltonian that corresponds to this problem is the Pott's one;

$$H[\sigma] = \sum_{<i,j>} \delta_{\sigma_i,\sigma_j}$$

where σ_i has value in $0, q-1$

The coloring problem has been shown ([27]) to undergo several phase transitions while p is changed.

B.3 Tree Approximation

The difficulties in the solution of optimization problems come from the complexity of the energy landscape linked to the presence of frustration.

An optimization problem defined on a tree (such that its factor graph has no loops) is not hard.

One of the reasons that induced to study the Optimization Problems on Erdős Rényi Graphs is the fact that (for finite average connectivities) the structure of such graphs is locally a tree structure. In fact in the large N limit the size of the loops is logarithmic in the size.

Given a fixed vertex it is easy to see that for any fixed n the probability to have at least one loop long n goes to zero in the large N limit. It is also possible to show that the typical length of loops is $O(\ln N)$. In fact the average number of vertices connected to a fixed vertex after n steps is k^n. A loop will be closed when this number is comparable with N.

By increasing the parameter p from 0 to 1 we can observe the evolution of geometrical structures in the ensemble of random graphs. It is not exact (for any finite size) to say that the connectivity distribution of each vertex is Poissonian; anyway as we are interested in the thermodynamic limit we will always assume to be in this limit so that such (asymptotic) statements are true. At the beginning ($p = 0$) we have N isolated clusters, then there will be small clusters of small size. One of the first questions we want to answer is what is the value of p such that finite-size clusters appear. By finite-size cluster we mean a cluster such that its sites are a finite fraction of the sites in G. Such value for the parameter p is called percolation threshold.

In fact the problem of the giant component is equivalent to the percolative problem on the complete graph; then it can be studied in the Fortuin-Kastelyin representation as a statistical mechanics problem. The appearence of a gigantic component (a finite-size connected component) shows non-analytic behaviour in the *thermodynamic limit* because it has a phase transition. The appearence of the gigantic component correspond to the break of the simmetry under change of the spin values of the variables (up-down for the $q = 2$ case)).

We can analyze this problem just by analyzing it in the tree approximation. It is possible because when there is not a gicantic component then there are $O(N)$ isolated clusters. If we randomly put another edge it has a probability $\sim 1/N$ to connect a vertex of a component to another of the same component. Then, in absence of a gigantic component, the isolated clusters are (typically) trees.

Given a random tree, if we remove a vertex i with connectivity k_i we have k_i disconnected trees and each vertex on them has the same

distribution of a randomly drawn vertex except for the vertices originally linked to i. Remark that if the distribution of the connectivity is poissonian on the vertices in the graph then if the tree is drawn by choosing randomly the vertex to be eliminate in G, the connectivity of its neighbours is distributed according to $p(x) = Poiss_k(x-1)$ for $x \neq 0$ and $p(0) = 0$, then after the remotion of i the graph remains Poissonian.

In general a vertex is in the gigantic component of a graph if its neighbours are there. A (finite connectivity) vertex belonging to a tree is in the gigantic component if, after its remotion, at least one of its neighbours are in a gigantic component. This maybe is the easier example of use of the cavity method! The cavity consists in the remotion of the vertex i. The cavity systems is the original graph without the i vertex; we are considering the observable "belong to the gigantic component", and we are supposing that the probability which i's neighbours have this property is independent of the fact that i itself has this property. Let η be this probability

$$1 - \eta_i = \prod_{j \in \mathcal{V}(i)} (1 - \eta_j)$$

This equation for the Erdős Rényi graphs is

$$\eta = 1 - e^{-\gamma\eta}$$

This equation has a null solution for every γ and a not null solution for $\gamma > \gamma_c$ with $\gamma_c = 1$.

For $\gamma > 1$ the not null solution is the physical one. A poissonian graph has not a gigantic component for $\gamma < 1$ and has (almost surely) one gigantic component if $\gamma > 1$. The size of the gigantic component is $\eta \cdot N$ and is continue in the parameter γ.

B.3.1 k-core

The k-core is the subgraph of maximum size such that all its vertices have connectivity at least k. The characterization of the k-core has a crucial importance in the study both theoretical and algorithmitic of several Optimization Problems. Also the k-core can be studied in the tree approximation, it exhibit a discontinuous phase transition. For a spin-glass it is easy to see that is possible to reduce the problem just removing all the part of the graph not belonging to the 2-core and substituting it with some easy to find (in complexity) external fields.

The algorithmic problem of finding the k-core of a given graph is easy. A way to find the k-core of a graph is the leaf removal. Begin with the whole graph, and perform the following iterative step:

- if there are not vertices with coordination smaller than k then stop, else remove one among these vertices

Also the size of the k-core of random graphs is affected by phase transitions.

Leaf removal algorithms can be used also to find the hard-part of problems easily described by hypergraphs. One example are the satisfaction problems. The clauses in satisfaction problems involves often more than two variables, this clauses can be represented by hyperedges and the set of vertices and hyperedges is called hypergraphs. Hyperedges involving vertices with connectivity one correspond to satisfiable clauses in SAT and XORSAT problems then in hypergraphs the vertices with coordination 1 play the same role of the leaves in graphs, for this reason we refer to them as leaf. The leaf removal described above works for hypergraphs as well as for graphs. Remark that the leaf removal algorithm is not able to remove the hyperloops (sub-hypergraphs such that every site has even connectivity).

We can analyze the 2-core properties by using a random leaf removal algorithm: the algorithm such that, given the set of removable vertices at the iterative step, removes randomly one of them with the same probability. It is possible to extend, by using the same algorithm, the analysis also to hypergraphs.

Such an algorithm induce a dynamic on the fraction of sites with given connectivity. This dynamic in the thermodynamic limit is deterministic and solvable, it stops when no more sites with connectivity one are found.

In the case of graphs the leaf removal transition coincides with the percolation one: when $\gamma = 1$ a k-core emerges and its size is continuous in γ. This critical value coincides with the percolative value because, when a finite fraction of sites belongs to a single connected component then a random new edge close a loop with finite probability.

For random hypergraphs with hyperedges of fixed coordination p, the percolative critical average coordination $\gamma_p = \frac{1}{p-1}$ is smaller than the core-critical γ_c one. Moreover the transition is not continuous: immediately above the critical value γ_c the size of the core is finite.

A deeper statistical description of the core in random graphs and hypergraphs is given in [35]

Acknowledgments

I wish to express my deepest gratitude to Sergio Caracciolo, Andrea Montanari and Andrea Sportiello who patiently guided me during these years of fruitful scientific collaboration.

Bibliography

[1] H. Bauke S. Franz and S. Mertens, *Number Partitioning as Random Energy Model* J.Stat.Mech P04003(2004)

[2] B. Bauer and D. Bernard, *2D growth processes: SLE and Loewner chains* Phys Rept (**432**) 2006 115-221 `math-ph/0602049.`

[3] M. Bayati, D. Shah, M. Sharma *Maximum Weight Matching via Max-Product Belief Propagation* Proc of the 2005 IEEE Intern. Symposium on Information Theory, `cs.IT/0508101`

[4] M. Bayati, C. Borgs, J.Chayes, R. Zecchina, *Belief-Propagation for weighted b-Matchings on Arbitrary Graphs and its Relation to Linear Programs with Integer Solutions*

[5] S. Boettcher *Numerical Results for Ground States of Spin Glasses on Bethe Lattices* Eur. Phys. J. (B **31**) 29-39 (2003)

[6] A. Braunstein, M. Mezard, R. Zecchina *Survey propagation: an algorithm for satisfiability* Random Structure and Algorithms (**27**), 201-226 (2005)

[7] S. Caracciolo A. Sportiello *An exactly solvable random satisfiability problem* J. Phys. (A **35**) 7661-7688 (2002)

[8] T. Castellani, F. Krząkała, F. Ricci-Tersenghi *Spin Glasses with ferromagnetically biased couplings on the Bethe lattice: analytic solutions and numerical simulations.* Eur. Phys. J. B **47**, 99 (2005) `cond-mat/0403053v3`

[9] F. Castellano *A Cavity Algorithm for Optimal Assignment* Università degli Studi di Milano (2005-2006)

[10] S.A. Cook *The complexity of theorem-proving procedures* Proc. 3rd Ann. ACM Symp. on Theory of Computing 151-158. ACM New York (1971)

[11] B. Deridda *Random Energy Model: Limit of a Family of Disordered Models*, Phys. Rev. Lett **45**, (1980).

[12] E.A. Dinic *Algorithm for solution of a problem of maximum flow in a network with power estimation* Soviet. Math. Doklady **11** 1277-1280 (1970)

[13] J. Edmonds and R.M. Karp *Theoretical improvements in algorithmic efficiency for network flow problems* J. ACM **19** (2) 248-264 (1972)

[14] P. Erdős A. Rényi *On Random Graphs* Publ.Math. Debrecen **6** 290-297 (1959)

[15] L.R. Ford D.R. Fulkerson (1956) *A simple algorithm for Finding Maximal Network Flows and an Application to the Hitchcock Problem* Canad. J. Math **9** (1957)

[16] A. Frank, *On Kuhn's Hungarian method, a tribute from hungary* Tech. Rep. TR2004-14, Egerváry Research Group, Budapest (2004)

[17] S. Franz M. Leone *Replica Bounds for Optimization problems and Diluted spin systems* `cond-mat/0208280`

[18] M.R. Garey and D.S. Johnson, *Computers and Intractability* W.H.Freeman, San Francisco, 1979

[19] M.X. Goemans D.P. Williamson *Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming* J.Assoc.Comput.Mach (1995)

[20] C. Grosso, *Cavity Methods Analysis for Random Assignment Problems,* Università degli Studi di Milano (2002-2003)

[21] I. Kanter and H. Sompolinski *Mean-Field theory of spin-glasses with finite coordination number* Phys. Rev. Lett **164** (1987)

[22] R.M. Karp *A Patching algorithm for the nonsymmetric Traveling Salesman Problem* SIAM J.Computing **8** n.4 (1979)

[23] R.M. Karp *Reducibility among combinatorial optimization problems,* Complexity of COmputer Computations, 85-103, Plenum Press, NY, 1972.

[24] S. Kirkpatrick C. D. Gelatt M. P. Vecchi, *Optimization by Simulated Annealing*, Science vol.**220** n4598 pages 671-680 (1983)

[25] H.Kitani H.Nishimori A.Aoki *Inequalities for the Local Energy of Random Ising Models* `cond-mat0703199v1` (2007)

[26] D.Knuth *The Stanford GraphBase: A Platform for Combinatorial Computing* Addison-Wesley (1993)

[27] F. Krząkała L. Zdeborova *Phase transitions in the Coloring of Random Graphs* Phys. Rev. E **76**, 031131 (2007) `arXiv:0704.1269v2`

[28] H.W. Kuhn *The Hungarian Method for the Assignment Problem,* Naval Research Logistic Quarterly, **2**:83-97, (1955)

[29] L. Lovász, M.D. Plummer, *Matching Theory* North-Holland Amsterdam - New York , (1986)

[30] O.C. Martin, M. Mézard, O. Rivoire *Random multi-index matching problems,* J.Stat.Mech P09006 (2005), cond-mat/0507180

[31] M. Mezard A. Montanari *Information, Physics and Computation* Oxford University Press (2008)

[32] M. Mezard, G. Parisi *The Cavity Method at Zero temperature* J.Stat.Phys **111** (2003) `cond-mat/0207121`

[33] M. Mezard, G. Parisi *The Bethe lattice spin glass revisited* Eur. Phys. J B**20**, 217 (2001) `cond-mat/0009418`

[34] M. Mézard, G. Parisi, M.A. Virasoro, *Spin Glasses Theory and Beyond*, Word Scientific, Singapore, (1987)

[35] M. Mezard F. Ricci-Tersenghi R. Zecchina *Alternative solutions to diluted p-spin models and XORSAT problems* J. Stat. Phys. **111**, 505 (2003) `cond-mat/0207140`

[36] M. Mézard, L. Zdeborova *The number of matchings in random graphs*, J.Stat.Mech. P05003 (2006) `cond-mat/0603350`

[37] M. Mézard R. Zecchina *The random K-satisfiability problem: from an analytic solution to an efficent algorithm* Phys. Rev. E **66** (2002) 056126, `cond-mat/0207194`

[38] *Introduction to phase transitions in random optimization problems* Les Houches Summer School Proceedings Volume **85**, pages 1-65 (2007)

[39] A.Montanari, T.Rizzo *How to compute loop corrections to the Bethe approximation*, J. Stat. Mech. (2005) P10011 `cond-mat/0506769`

[40] J. Munkres *Algorithms for the Assignment and Transportation Problems* J. Soc. Ind. and Appl. Math. **5**(1) (1957)

[41] C. Nair, B. Prabhakar, M. Sharma, *Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures,* Random Struct. Algorithms, **27**, n. 4, (2005)

[42] L. Onsager *A Two-Dimensional Model with a Order-Disorder Transition* Phys.Rev. **65**, 117,149 (1944)

[43] A. Pagnani G. Parisi, M. Ratieville, *Near optimal configurations in mean field disordered systems*, Phys.Rev.E **68** 0406706 (2003) `cond-mat/0307569`

[44] A.G. Percus O.C. Martin *The stochastic traveling salesman problem: Finite size scaling and the cavity prediction* J. Stat. Phys. **94**:5/6 (1999) 739-758 `cond-mat/9802295 (1998)`

[45] D. Sherrington, S. Kirckpatrick *Solvable Model of a Spin Glass*, Phys,Rev.Lett. **35** 1792-1796 (1975)

[46] N. Sourlas *Soft annealing: a new approach to difficult computational problems.* Phys. Rev. Lett **94** (2005) 070601 `cond-mat/0408082`

[47] L. Viana, A.J. Bray *Phase Diagram for Diluite Spin Glasses*, J.Phys. C **18**, 3037 (1985)

[48] J. S. Yedidia, W. T. Freeman, Y. Weiss *Generalized Belief Propagation*, Advances in Neural Information Processing Systems NIPS **13**, 689-695 (2000)