Algorithmic barriers in random constraint satisfaction problems

Louise Budzynski

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Ecole Normale Supérieure

Algorithmic barrier in random constraint satisfaction problems

Soutenue par
Louise BUDZYNSKI
Le 29 septembre 2020

École doctorale n°564
Ecole Doctorale Physique en Ile-de-France

Spécialité
Physique

Composition du jury :

Remi MONASSON Directeur de Recherche, Ecole Normale Supérieure	*Président*
Cristopher MOORE Professor, Santa Fe Institute	*Rapporteur*
Peter SOLLICH Professor, Universität Göttingen	*Rapporteur*
Lenka ZDEBOROVÁ Professor, Université Paris-Saclay	*Examinateur*
Alfredo BRAUNSTEIN Associate Professor, Politecnico di Torino	*Examinateur*
Guilhem SEMERJIAN MC1, Ecole Normale Supérieure	*Directeur de thèse*
Federico RICCI-TERSENGHI Full Professor, Università Sapienza	*Invité*

1794
ENS | PSL

Acknowledgments

First of all, I would like to thank Guilhem Semerjian for having accepted to supervise my doctoral studies. During these three years I had the opportunity to learn a lot scientifically, thanks to his very pedagogical explanations, his patience and his availability for discussions whenever I needed. More generally, I would like to thank him all his support, help and precious advices in all the aspects of this Ph.D., either from a practical or a scientific point of view. I would also like to thank him for the very efficient and careful reading of this manuscript.

I would also like to thank Federico Ricci-Tersenghi for his great supervision during my internship in Rome, that initiated this work, and for the collaboration we had during the Ph.D..

I warmly thank Alfredo Braunstein, Rémi Monasson, Cristopher Moore, Federico Ricci-Tersenghi, Peter Sollich and Lenka Zdeborovà for having accepted to participate to the jury of this defence. In particular, I would like to thank Cristopher Moore and Peter Sollich for having accepted to be the *rapporteurs* for my thesis.

Studying at the LPENS has been very exciting thanks to the numerous seminars and conferences organized here. I would like to thank in particular the professors at LPENS taking part in the Simon's Collaboration on Cracking the glass problem, Giulio Biroli, Jorge Kurchan and Francesco Zamponi for the organization of the weekly seminars. Similarly, I would like to thank Florent Krzakala and Lenka Zdeborovà for the organization of the Golosino seminars. More generally, I would like to thank all the pedagogical team of the École Normale Supérieure, for the great scientific formation I had the opportunity to follow. I would also like to thank Jean-François Allemand, Giulio Biroli, Jesper Jacobsen and Rémi Monasson for the precious advices they provided me along the way. From a more practical point of view, I would like to thank Viviane Sébille, Christine Chambon and Laura Baron-Ledez for their help in the administrative tasks.

I was happy to share my office with Aldo Battista, Dhruv Sharma and Emilio Trevisani, thank you for the funny moments in the office, and for all the nice discussions during lunch and coffee break. More generally I had the opportunity to met many students and postdoc at LPENS, thank you Ada Altieri, Alessandro Manacorda, Alia Abbara, Beatriz Seoane, Chen Liu, Davide Facoetti, Elisabeth Agoritsas, Félix Roy, Misaki Ozawa, Valentina Ros for the nice and friendly

atmosphere I found during these three years.

I also thank my family for their long-term support since the beginning of my studies, and my friends without whom these three years wouldn't have been as fun. Finally, my thoughts go to Émile, thank you for having been at my side in all the steps of this adventure.

Contents

Introduction

In a Constraint Satisfaction Problem (CSP), N discrete valued variables are subject to M constraints. Each of the constraints enforces some requirements on a subset of the variables. A solution of the CSP is an assignment of the variables that satisfies simultaneously all the constraints. Famous examples of CSPs are the k-satisfiability (k-SAT) problem and the graph q-coloring one (q-COL). In the first one the variables are Boolean and each constraint is the disjunction (OR) of k literals (a variable or its negation). In the second one the variables are placed on the vertices of a graph, they can take q possible values, to be interpreted as colors, and each edge of the graph enforce the constraint that the two vertices at its ends take different colors. In this Ph.D. we will be also interested in the bicoloring of k-hypergraph problem, that is similarly defined on a hypergraph, with hyperedges linking subsets of k (instead of two for a graph) vertices; the variables on the vertices can take two colors, and the constraint associated to each hyperedge is that both colors are present among its k adjacent vertices.

CSPs can be studied from several different perspectives; computational complexity theory [1, 2, 3] classifies them according to their worst-case difficulty, assessed by the existence or not of an efficient algorithm (running in a time polynomial in N, M) able to solve (i.e. to determine the existence or not of a solution) all their possible instances. Another perspective [4, 5, 6, 7, 8, 9, 10, 11, 12], that has been adopted in this Ph.D., consists of the characterization of the "typical" difficulty of CSPs, where typical is defined with respect to a random ensemble of instances. The most commonly studied random ensemble is obtained by drawing the M constraints uniformly at random. This corresponds to construct a $G(N, M)$ Erdős-Rényi random graph (or k-hypergraph). In this Ph.D., we will also consider a slightly different ensemble, the k-uniform $l + 1$-regular one, where the probability is uniform on the set of k-hypergraphs for which each vertex belongs to $l + 1$ hyperedges.

Random CSPs bear a formal similarity with models studied in statistical mechanics of disordered systems, and in particular with mean-field spin-glasses, the interactions induced by the constraints being of a frustrating nature while lacking a finite-dimensional structure thanks to the randomness in their construction. For instance, in the q-coloring problem on graphs, the variables can be considered as Potts spins, and the problem corresponds to find the ground state of an antiferromagnetic Potts model. If in the ground state all the links are

bicolored, then all the constraints are satisfied simultaneously, and the problem admits a solution.

A particularly interesting regime is the thermodynamic limit where $N, M \to \infty$ at a fixed ratio $\alpha = M/N$, the density of constraints per variable. Random CSPs exhibit threshold phenomena in this limit, the probability of some properties jumping abruptly from 1 to 0 as a function of the control parameter α. The most prominent of these phase transitions occurs at the satisfiability threshold α_{sat}, that depends on the parameter k, q of the problem. For $\alpha < \alpha_{sat}$ typical instances are satisfiable, i.e. admit configurations of variables that satisfy all constraints simultaneously, while for $\alpha > \alpha_{sat}$ a random instance is typically unsatisfiable. Thanks to the analogy between random CSPs and spin glasses, the application of the methods first developed in the context of statistical mechanics of disordered sytems, namely the replica and cavity method [4, 5, 6, 7, 8], has provided predictions of α_{sat} for many models, but also unveiled many other phase transitions for the structure of the set of solutions in the satisfiable phase. In addition, it has led to the proposal of new algorithms that exploit this detailed picture of the solution space. Many of these predictions have been confirmed rigorously later on [9, 10, 11, 12].

In this Ph.D. we give a particular attention to one of the phase transitions that occurs in the satisfiable phase, namely the clustering (or dynamic) transition that occurs at a critical density denoted α_d. This transition can be defined in various ways; the name clustering emphasizes the drastic change of the shape of the set of solutions, viewed as a subset of the whole configuration space. Below α_d the set of solutions of typical instances is rather well-connected, any solution can be reached from any other one by a rearrangement of a non-extensive number of variables. Above α_d the solution set splits in a large number of distinct groups of solutions, called clusters, that correspond to a pure state decomposition of the uniform measure over the set of solutions. The clusters are internally well-connected but well-separated one from the other. This transition marks also the birth of a specific type of long-range correlations between variables, known as point-to-set correlations, which implies the solvability of an information-theoretic problem called tree reconstruction [13]. These correlations forbid in turn the rapid equilibration of the stochastic processes that respect the detailed balance condition [14], hence the name dynamic given to α_d. The static properties of the model are instead not affected by the clustering transition and are only sensitive to a further condensation transition α_c that affects the number of dominant clusters [8]. In the cavity method [15] treatment of the random CSPs α_d can also be defined as the appearance of a non trivial solution of the one step of Replica Symmetry Breaking (1RSB) equation with Parisi breaking parameter $\mathcal{X} = 1$, see in particular [16] for the connection between this formalism and the reconstruction problem. In the large k limit the dynamic transition happens at a much smaller constraint density than the satisfiablity one. For instance, the asymptotic expansion of these two thresholds for the bicoloring of k-hypergraphs is $\alpha_d(k) \sim 2^{k-1} \ln k/k$ and $\alpha_{sat}(k) \sim 2^{k-1} \ln 2$.

Despite the rather detailed picture of the set of solutions of random CSPs, an important open problem remains to understand the behavior of algorithms

that attempt to find a solution in the satisfiable regime, where typical instances admit such configurations. In particular one would like to determine the algorithmic threshold $\alpha_{alg}(k)$ above which no algorithm is able to find a solution in polynomial time with high probability (assuming P$\neq$NP). For small values of k it is possible to design algorithms (see [6, 17, 18, 19, 20]) that are found through numerical simulations to be efficient at densities very close to the satisfiability threshold. The situation is quite different in the large k limit, where these algorithms cannot be studied numerically. One has to resort to analytical studies in this case, which can only be performed on simpler heuristics. The best result in this direction is the one of [21], which provides an algorithm that provably works in polynomial time up to densities of constraints coinciding at leading order with $\alpha_d(k)$. This leaves a wide range of α where typical instances have a non-empty set of solutions, but no known algorithm is able to find them efficiently (and where some families of algorithms have been proven to fail [22, 23, 24]).

One could hope to get some insight on the typical behavior of the algorithms and on the value of α_{alg} from the several phase transitions undergone by the set of solutions in the satisfiable phase. However the connection between these two aspects is delicate, because the algorithms considered are out-of-equilibrium, either because they explicitly break the detailed balance condition, or because their relaxation time scale is larger than the time scales accessible for practical experiments. Even if one cannot understand precisely $\alpha_{alg}(k)$ in terms of a structural phase transition one can reasonably state that the dynamic transition is a lower bound to the algorithmic one, $\alpha_d(k) \leq \alpha_{alg}(k)$. Indeed for $\alpha \leq \alpha_{\rm d}$ simulated annealing [25] should be able to equilibrate in polynomial time down to arbitrarily small temperatures, and hence sample uniformly the solution set. For α slightly larger than $\alpha_{\rm d}$ one expects simulated annealing to fall out-of-equilibrium on polynomial timescales but in many cases it should still be able to find (non-uniformly) solutions, hence the bound $\alpha_d(k) \leq \alpha_{alg}(k)$ is not tight in general.

The study of the structural phase transitions in the satisfiable regime, and in particular the definition of α_d in terms of long-range correlations, relies on the characterization of a specific probability law on the space of configurations, namely the uniform measure over solutions. In this Ph.D. we study probability measures over the set of solutions of random CSPs, for which not all solutions are equally probable. The same idea has been used in several articles, see in particular [26, 27, 28, 29, 30], with slightly different perspectives and results. For instance in [27, 28] solutions are weighted according to their local entropy, a quantity that counts the number of other solutions in their neighborhood. In [26] the solutions are weighted according to their number of frozen variables taking the same value in the whole cluster. In [29] hard sphere particles are considered as a CSP, with a bias due to an additional pairwise interaction between particles. In this Ph.D. we will study the modifications of the clustering threshold α_d induced by the non-uniformity between solutions. Focusing on the bicoloring on k-hypergraph problem, we will see that with an appropriate choice of the bias it is possible to delay the clustering threshold to higher densities of constraints, both for finite values of k and in the large k limit. For finite k values we will

check that this strategy has a positive impact on the performances of simulated annealing algorithms, in agreement with the discussion above.

This manuscript is based on the three papers [31, 32, 33] that have been produced during my Ph.D.. It is organized as follows. A survey describing the phase transitions in random CPSs and the search algorithms is presented in chapters 1-5, with a summary of some of the results obtained in the literature. In chapter 1 we start by defining the Constraint Satisfaction Problems and give some elements of the complexity theory. We also introduce the random ensembles of CSPs and explain their connection with statistical physics problems. Chapter 2 presents the several phase transitions undergone by the set of solutions in random CSPs in the thermodynamic limit. In chapter 3 we introduce the simulated annealing algorithm and other local search algorithms that perform a biased walk in the configuration space toward the set of solutions [17, 18, 19, 34]. In chapter 4 we introduce the message passing algorithms [6, 20, 35, 36, 37] inspired by statistical mechanics considerations. Chapter 5 describes the performances of the algorithms on random instances of CSPs. In particular it separates the regime of small values of k where some algorithms are efficient up to densities very close to the satisfiability threshold, and the large k limit, where there is a wide range of α in the satisfiable phase for which no provably efficient algorithm is known at present to find solutions.

In chapter 6 we introduce biased measures on the set of solutions. We give an explicit definition of the biased measures studied in this Ph.D., and describe the biased measures studied in the literature ([26, 27, 28, 29, 30]). Chapter 7 presents the cavity method that allows us to estimate the value of the thresholds for the several phase transitions. We describe the explicit derivation of these thresholds for the biased measures studied in the Ph.D..

In chapter 8 we present the results obtained for finite values of k for the bicoloring of k-hypergraphs. The sections 8.2-8.4 present the results obtained in the paper [31], that has been written in collaboration with Federico Ricci-Tersenghi. In [31] we have studied a simple implementation of the bias in the measure on the set of solutions of a hypergraph bicoloring instance, where the interactions induced by the bias can be factorized over the bicoloring constraints. We showed, for k between 4 and 6, that with well-chosen parameters such a bias allows us to increase α_d, and to improve the performances of simulated annealing. The section 8.5 presents part of the results obtained in [33], in which we introduce a more generic way of weighting the different solutions, that extends the one presented in [31] and incorporates interactions at larger range between variables. We show that for finite k this more generic biased measure allows a further increase of the dynamic threshold α_d.

The last two chapters study the large k regime. For the uniform measure over the set of solutions, some rigorous bounds are known on the asymptotic behavior of α_d [38, 39, 40]. In chapter 9 we present the results obtained in the paper [32], in which we provide an asymptotic expansion of the clustering threshold: we found that for the bicoloring on k-hypergraphs the clustering threshold occurs on the scale $\alpha \sim 2^{k-1}(\ln k + \ln\ln k + \gamma)/k$ with γ constant. We obtain a similar scaling $\alpha \sim (q/2)(\ln q + \ln\ln q + \gamma)$ for the q-coloring problem in

the large q limit, with exactly the same constant γ_d. We estimate this constant for the uniform measure $\gamma_{\mathrm{d,u}} \approx 0.871$, which falls into the range allowed by the previous bounds [38, 39, 40].

chapter 10 presents the results obtained in the paper [33]. In this paper we adapt the large k expansion of [32] to the biased measures studied in this Ph.D. and manage to assess the asymptotic effect of the bias on α_d. We show that the clustering threshold for our implementation of the bias arises at the same scale as in the uniform case, but with a constant γ_d that depends on the rescaled parameters describing the bias. We find that the factorized bias of [31] cannot improve on the constant γ_{d} in the asymptotic expansion with respect to $\gamma_{\mathrm{d,u}}$, while the bias with larger interaction range allows us to increase its value up to $\gamma_{\mathrm{d}} \approx 0.977$. Although this provides only a modest improvement, bearing on the third order of the asymptotic expansion of α_d, it opens the possibility to study further generalizations of the bias and to bring some light on the nature of the algorithmic gap between α_{alg} and α_{sat}.

Chapter 1

Definitions

1.1 Constraint Satisfaction Problems

1.1.1 Definitions

In a constraint satisfaction problem (CSP), a set of N variables are submitted to a set of M constraints (also called clauses). The variables x_i, $i \in \{1, \ldots, N\}$ take their values in a finite set χ. When χ is of size 2, the variables can either be viewed as Boolean variables: $\chi = \{0, 1\}$ or, to adopt the physics notation, as spins $\sigma_i \in \{-1, 1\}$, using the change of variable $\sigma_i = 2x_i - 1$. When the size of χ is an arbitrary integer q, the variables can be viewed as Potts spins, or colors: $\chi = \{1, \ldots, q\}$. We call $\underline{x} = \{x_1, \ldots, x_N\} \in \chi^N$ a configuration of the N variables. For a subset $S \subset \{1, \ldots, N\}$ of variables, we call $\underline{x}_S$ its configuration. The clauses c_a, $a \in \{1, \ldots, M\}$ involves a subset $\partial a \subset \{1, \ldots, N\}$ of variables, and impose a constraint on the value of their configuration $\underline{x}_{\partial a}$. They are functions $c_a : \chi^{|\partial a|} \to \{0, 1\}$ that evaluate to 1 when the constraint is satisfied, and to 0 otherwise. We call ∂i the subset of clauses in which the variable i appears. There is a natural cost function $E : \chi^N \to \mathbb{R}^+$ to define on a CSP, that counts the number of constraints that are unsatisfied (violated) by a configuration $\underline{x}$:

$$E(\underline{x}) = \sum_{a=1}^{M} (1 - c_a(\underline{x}_{\partial a})) \tag{1.1}$$

In the optimization version of the problem, one tries to find an optimal configuration that minimizes this cost function. In the decision version of the problem, one tries to answer the question: 'Given a threshold value E_0, is there a configuration $\underline{x}^*$ with a cost $E(\underline{x}^*) \leq E_0$?' One often takes the threshold to be 0, so that the question is 'Is there a configuration that satisfies all the constraints simultaneously ?'. Such a configuration is called a solution. Note that if the evaluation of the cost function is easy to perform, the decision version of the problem cannot be harder than the optimization version. Indeed once we have an optimal configuration, we only need to evaluate its cost and compare it to

E_0 in order to answer the decision problem.

We will adopt the following definitions. The constraint satisfaction problem is the general question to answer, defined by a set of parameters left unspecified. To be more precise, these parameters are for instance the set χ, the numbers N, M of variables and clauses, the function space in which the clauses c_a can take their values, the specific form of the cost function. An instance of the problem is obtained by setting a particular value to all the problem parameters. One says that an algorithm solves a constraint satisfaction problem if it produces an answer to any instance of the problem. A certificate is an assignment $\underline{x}$ that allows to check the answer of the problem.

There is another version of the CSPs, called the counting problem, that aims at counting all the solutions of a given instance. Generally this version is more difficult than the decision and optimization versions.

1.1.2 The satisfiability problem

As a first example, consider the satisfiability (SAT) problem. It is defined on a set of N Boolean variables. The clauses $\{c_1, \dots, c_M\}$ are logical constraints on the set of literals. For a variable x_i, its literal l_i is either equal to x_i or to its negation $\overline{x_i} = 1 - x_i$. Each clause is a disjunction (logical OR) on the literals formed from the variables in $\partial a = \{i(1), \dots, i(|\partial a|)\}$: $c_a = (l_{i(1)} \vee l_{i(2)} \vee \cdots \vee l_{i(|\partial a|)})$. The clause c_a evaluates to 1 if and only if at least one literal evaluates to 1. Among the $2^{|\partial a|}$ possible configurations that the variables in ∂a can take, there is only one that violates the clause, namely the one for which $l_{i(1)} = l_{i(2)} = \cdots = l_{i(|\partial a|)} = 0$. The satisfiability formula F is the conjunction (logical AND) over the set of clauses:

$$F = c_1 \wedge c_2 \wedge \cdots \wedge c_M \tag{1.2}$$

F evaluates to 1 if and only if all the clauses $\{c_1, \dots c_M\}$ evaluate to 1. We can call F a SAT formula, or a CNF formula (where CNF stands for conjunctive normal form). In the decision version of the SAT problem, a satisfying assignement, or solution, is then a configuration $\underline{x}$ such that $F = 1$. An instance of the SAT problem is thus defined by specifying the formula F, i.e. the value of N, M, then for each clause a the choice of the subset ∂a, and for each variable i in ∂a the choice of the literal l_i that appears in c_a. The k-SAT problem is the restriction of the SAT problem to the set of instances that have clauses of fixed length k. The optimization version of the SAT problem is called the MAX-SAT problem (or MAX-k-SAT when the clauses have length k).

Using the spins ($\sigma_i = 2x_i - 1$) the clauses can be re-written as follow:

$$c_a(\underline{\sigma}_{\partial a}) = 1 - \prod_{i \in \partial a} \mathbb{I}[\sigma_i = -J_i^a] \tag{1.3}$$

where J_i^a takes its values in $\{-1, +1\}$, with $J_i^a = 1$ if the literal is x_i, and $J_i^a = -1$ if the literal is $\overline{x_i}$, and $\mathbb{I}[A]$ is the indicator function of the event A.

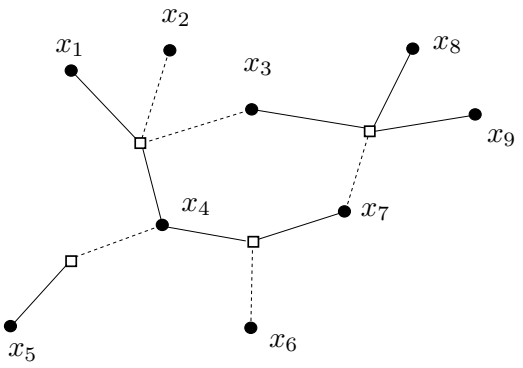

Figure 1.1: graphical representation of the SAT formula $F = (x_1 \vee \overline{x_2} \vee \overline{x_3} \vee x_4) \wedge (\overline{x_4} \vee x_5) \wedge (x_4 \vee \overline{x_6} \vee x_7) \wedge (x_3 \vee \overline{x_7} \vee x_8 \vee x_9)$

1.1.3 A graphical representation

There is a natural graphical representation of the instances of a CSP. For any instance, we associate a bipartite graph $G = (V, C, E)$. Figure 1.1 is an example of this representation for a SAT formula. The filled circles represent the set V of N vertices of the first type (called variable nodes), that is associated with the set of variables. The empty squares represent the set C of M vertices of the second type (called function nodes), that represent the set of clauses. E is the set of edges between variable nodes and function nodes. A link is drawn between the variable node i and the function node a if the variable i appears in the clause a: $i \in \partial a$. Note that for the SAT problem, an instance is not yet fully specified by this representation, since the information about the choice of the literal is missing. This can be specified by using two types of edges, for instance plain edges when $l_i = x_i$, and dashed edges when $l_i = \overline{x_i}$.

The bipartite graph G can be viewed equivalently as a hypergraph, where the variables are still represented as vertices, but the clauses are now hyper-edges, i.e. edges linking a subset of vertices $|\partial a|$ possibly greater than 2.

1.1.4 More examples of CSP

1. The k-XORSAT problem:

   Compared to the k-SAT problem, the disjunction $\vee$ (OR) is replaced by the eXclusive OR (XOR) (denoted by $\oplus$):

   $$c_a = (l_{i(1)} \oplus \cdots \oplus l_{i(k)}) \tag{1.4}$$

   This is equivalent to asking that the sum $\sum_{j=1}^{k} l_{i(j)}$ of the literals is equal to 1 modulo 2. This problem has a simpler structure than the k-SAT problem. In particular, one can solve this problem in polynomial time using algorithms based on Gaussian elimination.

2. The q coloring (q-col) problem:

   In this problem the variables can take q possible values. Each clause involves a pair of variables, and is satisfied when the two variables do not take the same color: $c_a(x_i, x_j) = \mathbb{I}[x_i \neq x_j]$.

3. The bicoloring problem on k-hypergraphs

   This problem is very similar to the q-coloring problem. Here $q = 2$: the variables are Booleans (or spins), and each constraint involve a k-tuple (with $k \geq 2$) of variables. The constraints forbids the configurations where all the variables take the same value. It is also called the Not All Equal (NAE) k-SAT problem, but note that compared to the k-SAT problem, the constraints forbids 2 configurations (namely the configuration where $x_i = 1$ for all $i \in \partial a$, and the opposite configuration where $x_i = 0$ for all $i \in \partial a$).

1.1.5 Other combinatorial optimization problems

The CSPs belongs to the family of combinatorial optimization problems. These problems consist of finding a configuration (in a finite set of possible configurations) that minimizes a cost function. We give few examples:

1. the assignment problem: given N agents and N tasks, and an affinity matrix M_{ij} representing the affinity of agent i for the task j, one looks for an assignment of each agent to a task in order to maximize the total affinity. The space of configurations is the space of permutations $\pi \in \mathfrak{S}(N)$. One defines the total cost function

$$E(\pi) = \sum_i C_{i\pi(i)} \tag{1.5}$$

2. the Hamiltonian cycle. This is a decision problem. Given a graph $G = (V, E)$, a Hamiltonian cycle is a path that visit each vertex of the graph exactly once, and gets back to this starting point. The question is then 'Does there exist a Hamiltonian cycle ?' The configurations are in the set of paths on G.

3. the traveling salesman problem: this is the optimization version of the Hamiltonian problem on a weighted graph. The cost function is defined as the sum of weights of the edges visited by the path. The question is then to find a Hamiltonian cycle that minimize the cost

1.2 Worst-case complexity

1.2.1 Time complexity

From an algorithmic point of view, one would be interested in finding an algorithm that solves a CSP (i.e. solves all its possible instances) in a reasonable

amount of time. Intuitively, an algorithm will require more time to provide an answer for large size instances. There are several ways to define the size of an instance. A natural definition is to take N the number of variables, but one could also have chosen $N+M$, or $|E|$ the total number of edges in the graphical representation of a CSP. Most of the time one works with instances for which $M = \alpha N$, with α a constant independent of N, therefore all these possible definitions are polynomially related. In computational complexity theory [1] it is therefore equivalent to work with any of them, and we will fix the size of the problem to be N.

One defines the time complexity function $T(N)$ of an algorithm as the largest amount of time needed for this algorithm to solve an instance of given size N. Note that this definition depends on the computer used to implement the algorithm. One will not enter into the details, and just state that one uses a 'realistic' computer model. Realistic means that such a model can run at most one (or a finite small number) of operations per unit time, where operations are for instance comparison, sum and multiplication. Some examples of realistic model are the one-tape (or multi-tape) Turing machines, and the random-access machines (RAMs). The algorithms can be split into classes according to the behavior of their time complexity function.

1. A polynomial time algorithm has a time complexity function $T(N) = O(p(N))$ for some polynomial function p.

2. An exponential time algorithm has a time complexity function that can be lower bounded by 2^{bn} for some constant $0 < b$.

This distinction is useful because exponential time algorithm have a time complexity that explodes when the size increases, therefore they become inefficient for large size instances. One says that a problem is intractable when it is so hard that no polynomial algorithm can solve it. Note that the time-complexity function is a worst-case measure, in the sense that is corresponds to the time needed for the algorithm to solve the hardest instance of a given size. It could happen that the majority of other instances require much less time, and the algorithm would be 'efficient' in a large proportion of the instances it encouters. As an example, the k-SAT problem is solved by an exponential time algorithm called DPLL, that one will describe further. As we shall see in the following, experimental observations and rigorous results show however that for instances such that $M = \alpha N$, with α small enough, this algorithm finds solutions very quickly for the vast majority of instances, even when N gets large.

1.2.2 Complexity classes

The theory of computational complexity classifies the difficulty of the decision problems according to the existence of an algorithm that solve a problem in polynomial time. This classification uses the notion of polynomial reduction. One says that a problem π_1 can be reduced to another problem π_2 when there is a function that maps an instance of π_1 into an instance of π_2. To be more

precise, let D_π be the set of instances of problem π. Let $Y_\pi \subseteq D_\pi$ be the subset of instances for which the answer to the decision problem is 'yes'. Then a reduction is a function $f : D_{\pi_1} \to D_{\pi_2}$ such that for all $I_1 \in D_{\pi_1}$, one has $I_1 \in Y_{\pi_1}$ if and only if $f(I_1) \in Y_{\pi_2}$. A polynomial time reduction is a reduction f that is computable by a polynomial time algorithm in the size of the problem π_1. One requires additionally that the size of $f(I_1)$ is polynomial in the size of I_1. The relation "π_1 can be reduced polynomially to π_2" is denoted $\pi_1 \propto \pi_2$. This relation is transitive:

Lemma 1 (Transitivity) *if $\pi_1 \propto \pi_2$ and $\pi_2 \propto \pi_3$, then $\pi_1 \propto \pi_3$.*

One defines the two following classes of problems:

1. The polynomial class P is the set of problems that can be solved by a polynomial time algorithm.

2. The non-deterministic polynomial class NP is the set of problems that have a polynomial time verifiability. This means when one provides a certificate that allows to check a 'yes' answer to the problem, the verification itself can be done in polynomial time.

The term non-deterministic comes from an alternative definition of the NP class, as the set of problems that can be solved by a polynomial time non-deterministic algorithm. A non-deterministic algorithm is composed of two stages. In the first stage, a certificate is proposed, and in the second stage, the certificate is verified. A non-deterministic algorithm is said to operate in polynomial time if the second stage is done in polynomial time, disregarding the first stage that might have required more time. Roughly speaking, a non-deterministic polynomial algorithm can pursue alternate paths or branches in parallel, and says 'yes' if any of these branches says 'yes'.

Note that if one has a polynomial reduction from π_1 to π_2, then π_1 cannot be harder than π_2. Suppose indeed that one has a polynomial algorithm that solves π_2. One can construct a polynomial algorithm to solve π_1 as follows. Given an instance $I_1 \in D_{\pi_1}$, this algorithm just applies the transformation f to get an instance $I_2 = f(I_1)$ of π_2, and applies the polynomial algorithm to solve the instance I_2. Therefore:

Lemma 2 *If $\pi_1 \propto \pi_2$, then $\pi_2 \in P$ implies $\pi_1 \in P$ (and equivalently $\pi_1 \notin P$ implies $\pi_2 \notin P$)*

Note that a problem π in P is automatically in NP. If there is a polynomial time algorithm that solves π, this algorithm can be converted into a checking algorithm that ignores the certificate that is provided, and simply returns its answer. Therefore P$\subseteq$NP.

A natural question is then to ask if P=NP, or if there exist problems in NP that are not in P. This is an open question, but it is widely conjectured that P$\neq$NP. This conjecture is supported by the existence of another class of NP problems gathering the hardest problems in NP. It is called the NP-complete class (NP-c). A problem π is in NP-c if it satisfies the two conditions:

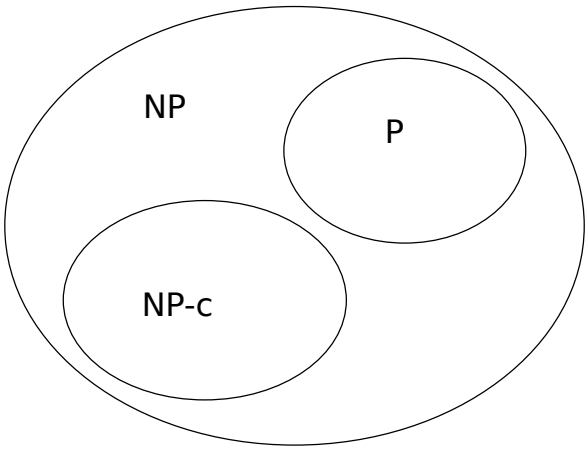

Figure 1.2: sketch of the classification of the decision problems

1. $\pi \in$NP
2. Any problem π' in NP can be polynomially reduced to it: $\pi' \propto \pi$

If one problem in NP-c could be solved in polynomial time, then all the problems in NP could be solved in polynomial time thanks to the polynomial reduction, and one would have P=NP. On the other hand, if one problem in NP is intractable, then all the problems in NP-c are intractables. Some problems in NP-c, such as the satisfiability problem, have been studied for decades without any proof that they belong to P, that is why one believes that P$\neq$NP. Figure 1.2 illustrates the conjecture on the structure of the sets NP, P and NP-c.

From the definition of the NP-c class, proving that a problem $\pi \in$NP belongs to NP-c requires to prove that any problem $\pi' \in$NP can be polynomially reduced to π. This seems difficult to achieve, but once one has proved that there is at least one problem in NP-c, the proof can be simplified thanks to this Lemma:

Lemma 3 *Let $\pi_1 \in$NP-c. If $\pi_2 \in$NP, and $\pi_1 \propto \pi_2$, then $\pi_2 \in$NP-c.*

Indeed since $\pi_1 \in$NP-c then for any problem $\pi' \in$NP one has $\pi' \propto \pi_1$, then by transitivity $\pi' \propto \pi_2$. The first problem that has been proved to be NP-complete is the satisfiability problem, by Stephen Cook in 1971 [41]

Theorem 1 (Cook,1971) *The satisfiability problem is NP-complete.*

We will not give the proof, but solely remark that the SAT problem has a very universal structure, that allows to re-write any decision problem into the SAT problem. Once we know that the SAT problem is in NP-c, one can apply the lemma 3 to prove that another problem π is NP-c. The method should follow these steps:

1. Show that $\pi \in$NP.
2. Select a known problem $\pi' \in$NP-c
3. Construct a reduction f from π' to π
4. Prove that f is a polynomial reduction.

Using this method, hundreds of problems have been shown to belong to NP-c. The table 1.1 gathers the problems that one has introduced above:

Table 1.1: Some decision problems belonging to NP or NP-c

P	NP-c
2-SAT	SAT
2-col	3-SAT
XORSAT	3-col
assignment	NAE-k-SAT with $k \geq 3$
	Hamiltonian circuit

1.3 An algorithm for the SAT problem

1.3.1 The Unit Clause Propagation procedure

We have seen that the 2-SAT problem is in P. We present here a polynomial algorithm that solves it. The algorithm is a sequential assignment procedure, which means that at each time step a variable is assigned to a given value. The algorithm ends either when all the variables are assigned, the assignment obtained being SAT, or when it has proven that the formula is UNSAT.

Each time a variable is assigned, one can simplify the initial CNF formula according to the following reduction procedure. Suppose that one has assigned the variable i to $x_i = 1$ (the case $x_i = 0$ is symmetric). Each clause containing the literal x_i is satisfied by this assignment, therefore can be removed from the formula. In each clause containing the opposite literal $\overline{x_i}$, one removes the literal $\overline{x_i}$ since it cannot satisfy the clause. The length of these clauses is then reduced by 1. One denotes $F|\{x_i = 1\}$ the simplified formula obtained with this procedure. This reduction procedure might produce a 0-clause, namely if in the formula there a unit clause (a 1-clause) of the form $c = (\overline{x_i})$, that is violated by the choice $x_i = 1$. One calls this a contradiction, it means that the simplified formula $F|\{x_i = 1\}$ is UNSAT, therefore to construct a SAT assignment for F, the choice $x_i = 0$ is necessary.

If at some step the formula obtained contains a unit clause, one is forced to satisfy it, and the simplified formula thus obtained might contain new unit clauses that one will have to satisfy in turn. This sequence of forced steps is called Unit Clause Propagation (UCP), and is done with the recursive UCP procedure (see algorithm 1). This procedure takes as input a CNF formula F

(possibly containing unit clauses), a partial assignment of variables A, and a set of variables V. In the first call we set F to be the initial CNF formula, and $A = \emptyset$, $V = \{1, \ldots, N\}$

Algorithm 1 UCP (F,A,V)

```
if there is a 0-clause in F then
    return UNSAT
end if
while there exist unit clauses do
    Pick a unit clause in F, say c = (l_i) and satisfy it
    Add the assignment of the variable x_i in A
    if there is a 0-clause in F then
        return UNSAT
    end if
end while
return A
```

The UCP procedure (1) returns all the assignments that were forced due to the presence of unit clauses. There is three possible outputs for the UCP procedure:

1. the output is an assignment of all the variables in V, then F is SAT and the assignment produced is a solution to F.

2. the output is a partial assignment. This partial assignment can be extended to a complete SAT assignment if and only if the input formula F is SAT. The simplified formula obtained does not contain unit clauses.

3. the output is UNSAT, therefore the input formula F is UNSAT.

The algorithm to solve 2-SAT uses the UCP procedure. It works as follows: given a 2-SAT formula F over a set of variables V, choose a variable $i \in V$, and fix $x_i = 1$. Then call the UCP procedure UCP($F|\{x_i = 1\}$,$\{x_i = 1\}$,$V \setminus i$). If the output has assigned all the variables, declare F satisfiable, and return the SAT assignment found. If it is only a partial assignment, keep it in memory, and let F', V' be the simplified formula and the set of not-yet assigned variables. Choose a new i' variable in V', and restart the UCP procedure UCP($F'|\{x_{i'} = 1\}$,$\{x_{i'} = 1\}$,$V' \setminus i'$).

If the output is UNSAT, it means that the initial choice has to be changed. Thus set $x_i = 0$ and restart the UCP procedure with UCP($F|\{x_i = 0\}$,$\{x_i = 0\}$,$V \setminus i$). If once again the procedure outputs UNSAT, then declare that the initial formula F is UNSAT. If not one can keep the partial assignment found and repeat the procedure. The step in which one chooses arbitrarily the value of a variable in a 2-clause is called a free step. By opposition, the assignment of a variable in a unit clause is called a forced step. The step in which we came back to the arbitrary choice done at the free step $x_i = 1$ to change it to $x_i = 0$ is called a backtracking step.

One can measure the complexity of this algorithm by the number of variable-fixing operations. For a 2-SAT formula, S. Cook showed that this algorithm has a polynomial complexity [41]. Note however that the algorithm presented above does not provide a method for solving the optimization problem MAX-2-SAT. In fact this problem has been shown to be in NP-c by M.R Garey, D.S Johnson and L Stockmeyer in [42] in 1976.

1.3.2 The DPLL algorithm for the SAT problem

The above algorithm can in fact be slightly modified to work on a CNF formula containing clauses of arbitrary length. In 2-SAT formulas, since any free choice produces unit clauses, one only needs to do backtracking steps on the last variable assigned on a free choice. All the other variables fixed at previous free steps are fixed, since they belong to a partial assignment compatible with at least one solution (if it exists). When there are clauses of length greater than 2 instead, one might have to do several free steps before having the possibility to apply the UCP procedure. If these choices lead to a contradiction, the backtracking then have to explore all the possible choices for these free variables before concluding that the formula is UNSAT. The number of backtracking steps might explode, leading to an exponential time complexity. This algorithm is called the Davis Putnam Logemann Loveland (DPLL) algorithm, from the authors names, and has been introduced in 1962 [43]. It is described by the algorithm 2 (see [3] p.726).

Algorithm 2 DPLL (INPUT: a SAT formula F, OUTPUT: is F satisfiable ?)

If F is empty **then return** SAT
If F contains and empty clause **then return** UNSAT
Select an unset variable x_i
If $F|\{x_i = 1\} = 1$ **then return** SAT
If $F|\{x_i = 0\} = 1$ **then return** SAT
Return UNSAT

The iterations of DPLL can be represented by a decision tree, as shown in figure 1.3 (the example is inspired from [44]).

It starts at the root of the tree, with the initial formula F containing all the variables to be assigned. A first variable is chosen and assigned to a value ($x_1 = 1$ in the example). The left-node of the first generation represent the simplified formula obtained with this choice. Since no unit clauses are produced, one needs to make another arbitrary choice: $x_2 = 1$ in the example. If there is a unit clause, the algorithm fixes the value of one variable present in a unit clause. When the algorithm finds a contradiction, it backtracks to the last variable assigned during a free step, and restarts the search from this node.

From this representation, one can measure the running time of DPLL on the formula F, by counting the number of nodes of the tree. It is intuitively clear why DPLL have an exponential complexity on SAT formulas with clauses of length greater than 2. If most of the variables have been fixed during a free

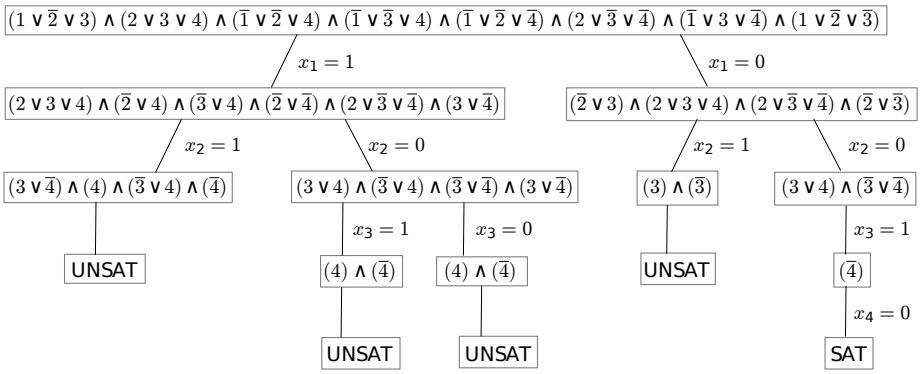

Figure 1.3: decision tree representing the DPLL algorithm solving the formula $F = (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_2 \vee x_3 \vee x_4) \wedge (\overline{x_1} \vee \overline{x_2} \vee x_4) \wedge (\overline{x_1} \vee \overline{x_3} \vee x_4) \wedge (\overline{x_1} \vee \overline{x_2} \vee \overline{x_4}) \wedge (x_2 \vee \overline{x_3} \vee \overline{x_4}) \wedge (\overline{x_1} \vee x_3 \vee \overline{x_4}) \wedge (x_1 \vee \overline{x_2} \vee \overline{x_3})$

step, and all the free choices need to be backtracked, the number of steps is roughly the number of nodes of a binary tree of depth N, which is $O(2^N)$. On a 2-SAT formula instead, each free choice is followed by a cascade of forced steps (possibly $O(N)$ of them). Even if all the variables have been fixed during a free step, and all the free steps have been backtracked, the complexity is still polynomially bounded by N^2.

In the algorithm 2, one has not specified how to choose the variable during a free step. One can use heuristics that aim at increasing the performances of DPLL. For instance, one can decide to pick a variable that will generate many unit clauses, or a variable that will satisfy the largest number of clauses. Since there is no good criterion to decide which heuristic to use, one would have to experiment them on the formula one has to solve.

1.4 Performances of the algorithms and random CSP ensembles

1.4.1 Complete and incomplete algorithms

An algorithm that solves a CSP, i.e. that provides an answer to all the possible instances of the CSP, is called a complete algorithm. DPLL is an example of complete algorithm for the SAT problem. By opposition, an incomplete algorithm is not guaranteed to provide an answer to any possible instances. In a decision problem, proving that a instance has a solution is often easier than proving that there is no solution. In the SAT problem for instance, a certificate for a 'yes' answer (to the question 'Is there a solution ?') is provided by a solution to the problem, that is an assignment of the N variable, therefore having size N. A certificate for the answer 'no' instead might have an exponential length, if

for instance one tries to exhibit all the possible assignments, showing that they all are UNSAT. The display of this certificate by itself have an exponential time complexity. An incomplete algorithm only tries to find a solution, and answers 'I don't know' whenever it is unable to provide one, or to prove that the problem has no solution.

We will present in the following chapters several incomplete algorithms for CSPs. Although they do not solve all the possible instances of the problem, it is observed experimentally that they are more efficient than the complete algorithms, on typical instances. To characterize their performances, one needs to precise what we mean by typical instances. This can be done by introducing random ensembles of instances. One then study the performances of algorithms on instances drawn from this ensemble.

1.4.2 Random CSP ensembles

We have seen that an instance of CSP can be represented with a hypergraph. Therefore one can use random graph ensembles to define the random CSP ensembles. The Erdős Rényi (ER) ensemble $G_N(k, M)$ is an example of random hypergraph ensemble. An instance of $G_N(k, M)$ is drawn by choosing independently for each clause a k-tuple of variables uniformly at random among the $\binom{N}{k}$ possible ones. Note that while the hyperedges $a \in C$ have a fixed degree k, the degree of the vertices $i \in V$ is not fixed.

We will also use another random hypergraph ensemble, called the k-uniform $(l+1)$-regular random hypergraph ensemble (the choice $(l+1)$ is made here for convenience when using the cavity method, see chapter 7). In this ensemble both the hyperedges and the vertices have a fixed degree k and $l+1$. All the hypergraphs with this property are equiprobable in this ensemble.

To define a random ensemble for the SAT problem, we need to specify how to draw the signs J_i^a. The random k-SAT ensemble is one of the most studied random ensembles for the k-SAT problem. An instance is obtained by drawing a hypergraph from $G_N(k, M)$, then independently for each variable in each k-tuple, one chooses a sign J_i^a with probability $1/2$.

It is useful to introduce the density of constraints $\alpha = M/N$, and to work with the ensemble $G_N(k, \alpha N)$. For the k-uniform $(l+1)$-regular random hypergraph ensemble, since N and M must satisfy the relation $N(l+1) = Mk$, the density of constraints is related to the degrees as $l + 1 = \alpha k$. Intuitively, increasing the density of constraints, the number of solutions should shrink, since it is harder to satisfy an overconstrained instance. At high density, we will see that most of the instances are in fact UNSAT.

We define the thermodynamic (large size) limit as $N, M \to \infty$ with a fixed ratio $\alpha = M/N$. In this limit, several properties of the random k-SAT ensemble concentrate around their mean value. This is called the self-averaging phenomenon. In particular, in this limit many random CSPs (including the random k-SAT ensemble) exhibit threshold phenomena. The probability of some property jumps abruptly from 1 to 0 when varying the control parameter α. We

say that a property of the instances is typical when its probability goes to 1 in the thermodynamic limit.

The most prominent threshold is the satisfiability threshold $\alpha_{sat}(k)$. Below $\alpha_{sat}(k)$, typical instances are SAT, while above $\alpha_{sat}(k)$, typical instances are UNSAT. In the satisfiability phase $\alpha < \alpha_{sat}(k)$, many other phase transitions concerning the properties of the set of solutions of a random instance are predicted by the so-called cavity method, that we shall describe in the next chapters.

Among them, the clustering transition describes a drastic change in the structure of the set of solutions in the space of configuration. Below the clustering threshold, the set of solutions is rather well-connected, any solution can be reached from any other one by nearby intermediate solutions, while above the clustering threshold the solution set splits in a large number of distinct groups of solutions, called clusters, which are internally well-connected but well separated one from the other.

1.4.3 Performances of DPLL on the random k-SAT ensemble

Our aim is to study and compare the performances of algorithms on a random ensemble of instances. We start by giving the experimental study of the DPLL algorithm on the random k-SAT ensemble.

For a fixed value k, one can plot the fraction of UNSAT instances found by DPLL as a function of α for several sizes N. The result for $k = 3$ is shown in figure 1.4 taken from the study of S. Kirkpatrick an B. Selman in 1994 [45] One can see that it is a increasing function of α, that goes from 0 to 1. As N increases, the drop becomes sharper around a threshold value. This is a numerical evidence of the satisfiability threshold.

One is interested in the running time of DPLL on random instances of k-SAT. On figure 1.5 taken from the study of D. Mitchell, B. Selman, H. Levesque in 1992 [46], the median running time is plotted against α. One can see that the running time has a peak around the satisfiability threshold. The height of this peak increases rapidly with N. At smaller and larger values of α instead, the running time is much smaller, and grows with N at a smaller rate. These observations indicate that the instances drawn from the random k-SAT ensemble are harder when α is close to $\alpha_{sat}(k)$. In the region of smaller α, the instances are underconstrained and therefore easy to solve. In the region of larger α, the instances are so overconstrained that DPLL finds quickly a contradiction showing that the instance is UNSAT.

In [47], P. Beame, R.Karp, T. Pitassi, and M. Saks show that the size of a certificate for an unsatisfiable instance of random 3-SAT is with high probability (w.h.p.) bounded from above by $2^{cN/\alpha}$, with c some contant, where "with high probability" means with probability going to one in the large size limit $N, M \to \infty$ at a fixed ratio $\alpha = M/N$. This results confirms the decreasing complexity observed in the unsatisfiable phase when α increases above the satisfiability threshold α_{sat}. In [48], A. Frieze and S. Suen show that in the satisfiable phase a

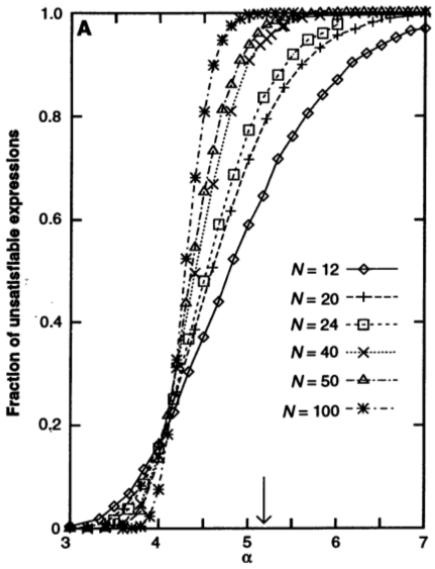

Figure 1.4: from [45] Fig 3.a, fraction of UNSAT formulas drawn from the random 3-SAT ensemble versus α for several sizes N.

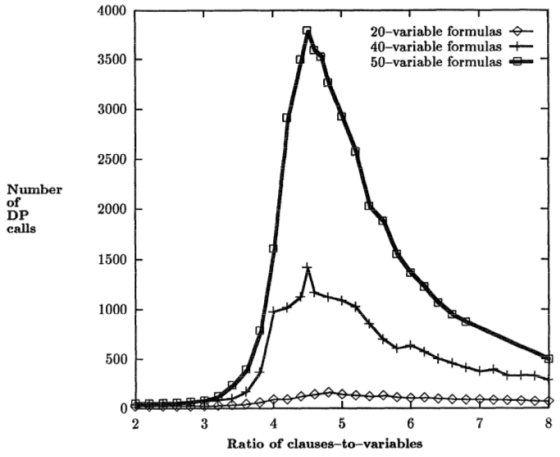

Figure 2: Median number of recursive DP calls for Random 3-SAT formulas, as a function of the ratio of clauses–to–variables.

Figure 1.5: from [46], median running time of DPLL on random instances of 3-SAT ensemble as a function of α, for several sizes N

modified version of DPLL without backtracking (with the unit clause branching rule) can find solutions efficiently for small enough densities: up to $\alpha \leq 3.003$ for $k = 3$, which is strictly smaller than the satisfiability threshold prediction $\alpha_{sat}(k = 3) = 4.267$ ([7]). This confirms the experimental observation that DPLL works in polynomial time for small enough densities. In [49], S. Cocco and R. Monasson provide a theoretical estimation of the algorithmic threshold for DPLL on random 3-SAT, introducing the random $(2 + p)$-SAT to study the evolution of the formula under the sequential assignments, with p the fraction of 3-clauses in the formula. They predict that above the density $\alpha \simeq 3.003$, the running time of DPLL with the unit clause branching rule becomes typically exponential.

1.4.4 The algorithmic barrier

The numerical study of the performances of DPLL on the random k-SAT ensemble indicates that in the region of α close to the satisfiability threshold the typical instances are hard to solve. In the next chapter we will introduce several incomplete algorithm that are designed to search for solutions, and we will compare their performances on the random ensembles. In particular, one is interested in determining the interval of α where it is possible to find an algorithm that is successful on typical instances. As we have seen this interval has to be in the satisfiable region $\alpha < \alpha_{sat}(k)$, but one could ask whether there exists a range in which typical instances have solutions, but no algorithm is able to find one. More precisely, one is interested in the putative algorithmic barrier $\alpha_{alg}(k)$, above which no algorithm is able to find a solution in polynomial time, assuming P$\neq$NP, and for typical instances. Is it possible to show that $\alpha_{alg}(k)$ coincides with $\alpha_{sat}(k)$, by exhibiting an algorithm efficient in the entire satisfiable region? Or are there limitations that force the strict inequality $\alpha_{alg}(k) < \alpha_{sat}(k)$? The structure of the set of solutions of typical instances undergoes a series of phase transitions in the satisfiable phase, that are predicted by the cavity method. It is therefore interesting to know if some of these phase transitions affect the performances of algorithms. For instance, algorithms such as Monte Carlo Markov Chains for the resolution of CSPs are affected by the clustering transition.

1.4.5 Some properties of the random hypergraph ensembles

In the ER ensemble $G_N(k, M)$ the degree of a uniformly chosen vertex i is a binomial random variable $B(M, k/N)$. In the thermodynamic limit therefore the degree of i obeys a Poisson law of mean αk.

A crucial property shared by the ER ensemble $G_N(k, \alpha N)$ and the k-uniform $(l+1)$-regular random hypergraph ensemble is their locally tree-like structure in the thermodynamic limit. Given a uniformly chosen vertex, the neighborhood within a finite distance is, with probability going to 1 as $N \to \infty$ (one says with high probability) acyclic. Consider a random instance of hyperhraph F, and a uniformly chosen vertex i of F, and let $B_{i,r}(F)$ be the subgraph containing the

vertices at distance smaller or equal than r from i. Then it is shown in [50] that $B_{i,r}(F)$ converges in distribution to the random tree ensemble $T_r(k,\alpha)$.

An instance of $T_r(k,\alpha)$ is constructed from the Galton Watson branching process: starting from the root, one draws its degree d from the distribution p_d. For the $G_N(k,\alpha k)$ ensemble one take the Poisson law $p_d = e^{-\alpha k}(\alpha k)^d/d!$, while for the regular ensemble one has $p_d = \mathbb{I}[d = l+1]$. We attach to the root d function nodes, and each of them are then connected to $k-1$ variable nodes, that constitute the first generation of the tree. Then for each variable node of the first generation, one draws an independent integer d with probability $r_d = \frac{p_d(d+1)}{\sum_{d'\geq 0} p_{d'}(d'+1)}$. The distribution r_d is called the residual degree distribution. It is the probability, given that the variable node i is connected to a given function node a, that i is connected to d other function nodes. For the $G_N(k,\alpha N)$ ensemble r_d is also distributed according to a Poisson law of mean αk, while for the k-uniform $(l+1)$-regular random hypergraph ensemble one has $r_d = \mathbb{I}[d = l]$. One reiterates the same process for each generation of the tree, until the r-th generation.

Since in the large size limit a random graph converges locally to a tree, one is then interested in the typical length of loops in these ensembles. It can be showed that for a random hypergraph F and a randomly chosen vertex i, the length l_i of the shortest loop going through i in F is $O(\log N)$, with high probability (w.h.p.). This scaling can be understood by the following heuristic argument. In $G_N(k,\alpha N)$, the average number of variable nodes in $B_{i,r}(F)$ is $(\alpha k(k-1))^r$. If one considers that loops appear when this number becomes comparable to the total number of nodes N (actually when it becomes comparable to $\sqrt{N}$ by the Birthday Paradox), then one obtains $r \propto \ln N/\ln(\alpha k(k-1))$.

1.5 Statistical physics and constraint satisfaction problems

1.5.1 Boltzmann probability measure

There is a natural mapping between optimization problems and statistical physics problems. In optimization problems, the goal is to minimize a cost function E defined on the set of possible configurations. For a CSP, there is a natural cost function $E : \chi^N \to \mathbb{R}^+$ that counts the number of clauses violated by a given assignment $\underline{x} \in \chi^N$ of the N variables: $E(\underline{x}) = \sum_{a=1}^{M}(1 - c_a(\underline{x}_{\partial a}))$ (equation (1.1)). For each instance F of a CSP, one introduces a Boltzmann probability distribution on the set of configurations:

$$\mu_\beta(\underline{x}) = \frac{1}{Z(\beta)} e^{-\beta E(\underline{x})} \tag{1.6}$$

the parameter β plays the role of an inverse temperature. In this setting the cost function E is also called the Hamiltonian, or energy function. The normalization $Z(\beta)$ is called the partition function. In the $\beta \to \infty$ limit, the

probability distribution μ_β concentrates on the configurations that minimize the energy. These configurations are called ground states in statistical physics, and corresponds to the optimal configurations of the optimization problem. In the decision version of the problem (with a threshold $E_0 = 0$), a solution is a zero-energy configuration. If the instance F admits solutions, then in the large β limit the probability distribution (1.6) converges to the uniform measure on the set of solutions, and the partition function counts the number of solutions. Using the expression (1.1) of the energy-cost function E, one can re-write the probability distribution μ_β:

$$\mu_\beta(\underline{x}) = \frac{1}{Z(\beta)} e^{-\beta \sum_{a=1}^{M} (1 - c_a(\underline{x}_{\partial a}))} \tag{1.7}$$

1.5.2 Statistical physics models

From the above expression, one sees the mapping from a CSP to a statistical physics problem. In this problem, the variables $x_i \in \chi$ interact through the local couplings $\psi_a(\underline{x}_{\partial a}) = 1 - c_a(\underline{x}_{\partial a})$. The graphical representation of the instance F can be viewed as the underlying graph representing the local interaction between variables.

Spin glass model are a generalization of the Ising model, in which the energy function reads

$$E(\underline{\sigma}) = -\sum_{a=1}^{M} J^a \prod_{i \in \partial a} \sigma_i \tag{1.8}$$

the variables σ_i are spins $\sigma_i \in \{+1, -1\}$, and the coupling J^a is either in $\mathbb{R}$ or in $\{-1, +1\}$. The p-spin model is is the restriction of the general spin glass model to p-body interaction. The 2-spin model is also called the Edward-Anderson model.

In the zero temperature limit, the spin glass model can be seen as a CSP: the Boltzmann measure μ_β is concentrated on the configurations which minimize the energy (1.8). A solution of the spin-glass problem should satisfy the following constraints. If $J^a > 0$ (resp. $J^a < 0$), then the product $\prod_{i \in \partial a} \sigma_i$ should be equal to 1 (resp. -1). For a two body interaction, it means that the variables should be aligned (resp. anti-aligned), therefore it is called a ferromagnetic (resp. antiferromagnetic) interaction.

Note that an instance of XORSAT can be transformed into an instance of the spin glass CSP with couplings J^a living in $\{-1, +1\}$. Recall that a XORSAT clause $c_a = (l_{i(1)} \oplus \cdots \oplus l_{i(|\partial a|)})$ equals 1 when the sum of literals is equal to 1 modulo 2. Let n_a be the number modulo 2 of variables that appear negated (the variables such that $l_i = \overline{x_i}$) in the clause a. Then the constraint $l_{i(1)} \oplus \cdots \oplus l_{i(|\partial a|)} = 1$ can be rewriten $x_{i(1)} \oplus \cdots \oplus x_{i(|\partial a|)} = b_a$, with $b_a = 1 - n_a$. Setting $J^a = 1 - 2b_a$ the above constraint is equivalent to ask for the constraint $J^a \sigma_{i(1)} \dots \sigma_{i(|\partial a|)} = 1$ to be satisfied.

Conversely, each instance of the spin glass problem (with couplings $J^a \in \{-1, +1\}$) can be transformed into a XORSAT instance. We conclude that the

decision version of the CSP associated with spin glasses at zero temperature is in P.

1.5.3 Graphical models

The measure (1.6) belongs to a family of measures described by graphical models, that we define here. Let $G = (V, C, E)$ be a bipartite graph, called a factor graph. V is called the set of variable nodes, C is the set of factor nodes, and E is the set of edges between variables and factors. For each variable node $i \in V$ we denote by $\partial i = \{a \in C : (ia) \in E\}$ the set of factor nodes connected to i, and for each factor node $a \in C$ we denote by $\partial a = \{i \in V : (ia) \in E\}$ the set of variable nodes connected to a. One defines also a space of configurations χ^N, with $N = |V|$, and χ a finite alphabet. Finally let $\{\omega_a\}_{a\in C}$ be a set of non-negative weights $\omega_a : \chi^{|\partial a|} \to \mathbb{R}^+$. A measure μ over χ^N described by the graphical model $G, \chi, \{\omega_a\}$ is defined as follows:

$$\mu(\underline{x}) = \frac{1}{Z} \prod_{a\in C} \omega_a(\underline{x}_{\partial a}); \quad Z = \sum_{\underline{x}} \prod_{a\in C} \omega_a(\underline{x}_{\partial a}) \tag{1.9}$$

The measure is well defined only if there exists at least one configuration for which all the weights ω_a are strictly positive.

Note that in the case of the measure (1.6) defined for a given instance F of CSP, the factor graph coincides with the graph representing F, and the weights $\{\omega_a\}$ are written: $\omega_a(\underline{x}_{\partial a}) = e^{-\beta(1-c_a(\underline{x}_{\partial a}))}$. The uniform measure on the set of solutions of F (assuming it is non-empty) is obtained by taking the limit $\beta \to \infty$ in μ_β. It can be written

$$\mu_u(\underline{x}) = \frac{1}{Z_u} \prod_{a\in C} (1 - c_a(\underline{x}_{\partial a})) = \frac{1}{Z_u} \prod_{a\in C} \mathbb{I}[\underline{x}_{\partial a} \text{ satisfies the clause } a] \tag{1.10}$$

where the subscript 'u' stands for 'uniform'.

We will have to define other measures associated with the instance F, for which the factor graph will be sometimes slightly different to the initial one representing F. Therefore it is more convenient to work with this general definition. We will generally denote by Θ the set of parameters needed to describe a given family of graphical models: i.e. the rules to construct the factor graph from the initial graph representing an instance F, and the function space in which the set of functions $\{\omega_a\}_{a\in C}$ belongs. For a given instance F, one defines $\Theta(F)$ to be the specification of these parameters to the instance F, and let $\mu_{\Theta(F)}$ be the measure associated with these parameters.

A measure μ described by a graphical model have the following decorrelation property called global Markov property. Let A, B, and S be three disjoint subset of V. One says that S separates A and B when there is no paths joining a node in A to a node in B without passing through S. Then (from [44] proposition 9.2):

$$\mathbb{P}_\mu[\underline{x}_A, \underline{x}_B|\underline{x}_S] = \mathbb{P}_\mu[\underline{x}_A|\underline{x}_S]\mathbb{P}_\mu[\underline{x}_B|\underline{x}_S] \tag{1.11}$$

one says that the variables $\underline{x}_A, \underline{x}_B$ are conditionally independent. In the chapter 4 we will describe the message-passing algorithms, that are built using the Markov property, and that allow to compute the properties of the measure μ when the factor graph is a tree.

Chapter 2

Phase transitions in random CSPs

We give now a more detailed picture of the phase transitions occuring in random CSPs when the density of constraints α is varied. We will take the particular example of the random k-SAT ensemble, but many other random CSP ensembles share the same qualitative picture.

2.1 The satisfiability transition

We say that an event occurs with high probability (w.h.p.) when its probability goes to 1 in the thermodynamic limit. We recall that the satisfiability threshold $\alpha_{sat}(k)$ separates a phase $\alpha < \alpha_{sat}(k)$ where random instances are SAT w.h.p. to a phase where random instances are UNSAT w.h.p. The cavity method [4, 5, 6, 7, 8] provides an estimate of the satisfiability threshold. However, the existence of the satisfiability transition is not yet proven for all values of k. It is summarized in the following conjecture:

Conjecture 1 *Let F be a random CNF formula with N variables and $M = \alpha N$ clauses, drawn from the random k-SAT ensemble. Let $P_N(k, \alpha)$ be the probability that the formula F is SAT. For any $k \geq 2$ there exists a constant $\alpha_{sat}(k)$ such that for all $\epsilon > 0$,*

$$\lim_{N\to\infty} P_N(k, \alpha_{sat}(k) - \epsilon) = 1, \text{ and } \lim_{N\to\infty} P_N(k, \alpha_{sat}(k) + \epsilon) = 0 \qquad (2.1)$$

This conjecture has been proven for $k = 2$, with $\alpha_{sat}(2) = 1$ by V. Chvatal, B. Reed (1992) [51], W. Fernandez de la Vega (1992) [52], and A. Goerdt (1996) [53]. In [12], J. Ding, A. Sly and N. Sun prove the satistfiability conjecture for large but finite k, and show that the value of $\alpha_{sat}(k)$ is given by the one-step symmetry breaking cavity method prediction. Proving the existence of the satisfiability threshold for all values of $k \geq 3$ is an open problem, however the

following theorem by E. Friedgut (1999) [54] provides a partial result in this sense:

Theorem 2 (Friedgut) *For every $k \geq 2$, there exists a sequence $\alpha_k(N)$ such that for all $\epsilon > 0$,*

$$\lim_{N\to\infty} P_N(k, \alpha_k(N) - \epsilon) = 1, \text{ and } \lim_{N\to\infty} P_N(k, \alpha_k(N) + \epsilon) = 0 \tag{2.2}$$

The above result states that the transition from SAT to UNSAT takes place in a window smaller than any fixed ϵ for N large enough. However it remains to prove the convergence of the sequence $\alpha_k(N)$ to some value $\alpha_{sat}(k)$ as $N \to \infty$, to prevent from possible oscillations. Upper and lower bounds have been established rigorously on the sequence $\alpha_k(N)$. In the following we present some of the methods used to derive these bounds.

2.1.1 Upper bounds

Upper bounds can be obtained with the first moment method. We describe the general strategy (presented in [44]). Define a function on the set of instances $U(F)$ such that

$$U(F) = \begin{cases} 0 \text{ if } F \text{ is UNSAT} \\ \geq 1 \text{ otherwise} \end{cases} \tag{2.3}$$

Then one applies the Markov inequality to $U(F)$ to obtain:

$$\mathbb{P}[F \text{ is SAT}] \leq \mathbb{E}[U(F)] \tag{2.4}$$

note that the equality is reached for $U(F) = \mathbb{I}[F \text{ is SAT}]$, but we do not know how to compute this quantity. One looks instead for a simpler function for which $\mathbb{E}[U(F)]$ can be computed, and has a vanishing limit when $N \to \infty$ for large enough α. As a first choice one can use $U(F) = Z(F)$, the number of solutions of F (see for instance [55]). By linearity of the expectation and by uniformity in the generation of the clauses the expectation $\mathbb{E}[Z(F)]$ is equal to the total number of assignments 2^N times the probability that a given assignment $\underline{\sigma}_0$ is SAT for F. This probability is the product over the clauses of the probability that each clauses is satisfied by $\underline{\sigma}_0$, therefore:

$$\mathbb{E}[Z(F)] = 2^N(1 - 2^{-k})^M = \exp[N(\ln 2 + \alpha \ln(1 - 2^{-k}))] \tag{2.5}$$

In the large N limit one gets:

$$\mathbb{E}[Z(F)] \to \begin{cases} 0 \text{ if } \alpha > \alpha_{UB}(k) \\ +\infty \ \alpha < \alpha_{UB}(k) \end{cases}, \ \alpha_{UB}(k) = -\frac{\ln 2}{\ln(1 - 2^{-k})} \tag{2.6}$$

and we have obtained the upper bound $\alpha_{UB}(k)$ on the conjectured satisfiability threshold $\alpha_{sat}(k)$. Intuitively, one can expect that this bound is not tight, because $Z(F)$ can take exponentially large values in N, and moreover its fluctuations can be exponentially large as well. In order to improve on this bound a

possible strategy is to choose a function $U(F)$ that counts the number of assignments in a small subclass of solutions. If the subclass is small enough one can hope to reduce the fluctuations of $U(F)$ and obtain a tighter upper bound. In [56] (1998), L. Kirousis, E. Kranakis, D. Krizanc, and Y. Stamatiou (1998) use the subclass of locally maximal SAT assignments, where a locally maximal SAT assignment is a SAT assignment such that for each variable i such that $x_i = 0$, flipping the variable to $x_i = 1$ leads to an UNSAT assignment. They define $U(F)$ to be the number of locally maximal SAT assignment and apply the first moment method to this function. They obtain the upper bound $\alpha'_{UB}(k)$, that is solution of the equation

$$\alpha \ln(1 - 2^k) + \ln\left(2 - \exp\left(-k\alpha/(2^k - 1)\right)\right) = 0 \tag{2.7}$$

2.1.2 Lower bounds

To derive a lower bound on $\alpha_k(N)$ one can use the second-moment method. The method was introduced by D. Achlioptas and C. Moore in 2006 [57]. Again the strategy is to use a function $U(F)$ that vanishes when the formula is UNSAT, and that is strictly positive otherwise. Applying the Cauchy-Schwarz inequality to $U(F)$ we get:

$$\mathbb{P}[F \text{ is SAT}] \geq \frac{\mathbb{E}[U(F)]^2}{\mathbb{E}[U(F)^2]} \tag{2.8}$$

One needs to find a function $U(F)$ for which the above ratio $\frac{\mathbb{E}[U(F)]^2}{\mathbb{E}[U(F)^2]}$ is not too difficult to compute and has a non vanishing limit when $N \to \infty$ limit for small enough α. One can show that applying this method to $Z(F)$ does not provide a useful bound, because the ratio $\frac{\mathbb{E}[Z(F)]^2}{\mathbb{E}[Z(F)^2]}$ vanishes for any non-zero value of α. It is however possible to apply this technique to another function U. In particular, one can choose U to be the size of a subset of the set of solutions. By choosing carefully this subset, one can hope to reduce the flutuations of $U(F)$ around its mean, so that the ratio $\frac{\mathbb{E}[U(F)]^2}{\mathbb{E}[U(F)^2]}$ is not vanishing in the large N limit.

Using this approach, D. Achlioptas and C. Moore show in [57] (Theorem 1.) that for $k \geq 3$:

$$\lim_{N\to\infty} P_N(k, \alpha) = 1 \text{ if } \alpha \leq 2^{k-1} \ln 2 - 2 \tag{2.9}$$

Note that in the large k limit, this result combined with the upper bound $\alpha_{UB}(k)$ provides the scaling $\alpha_{sat}(k) = O(2^k)$.

Another approach consists of analysing the behavior of some explicit algorithm that search for SAT assignments. If one can show that a particular algorithm finds a SAT assignment with positive probability when α is smaller than some value, then applying the Friedgut theorem 2 one can deduce that this happens w.h.p., therefore this value is a lower bound on the satisfiable threshold. More details on this analysis are given in the section 5.2.2.

2.2 Quenched and annealed averages

Consider a measure $\mu_{\Theta(F)}$, defined with a graphical model Θ, whose support is the set of solutions of a given instance F:

$$\mu_{\Theta(F)}(\underline{\sigma}) \begin{cases} = 0 \text{ if } \underline{\sigma} \text{ is not a solution} \\ > 0 \text{ if } \underline{\sigma} \text{ is a solution} \end{cases} \tag{2.10}$$

One might sometimes relax the condition $\mu_{\Theta(F)}(\underline{\sigma}) = 0$ when $\underline{\sigma}$ is not solution, and introduce a temperature β. The free entropy density is defined as follows:

$$\Phi(\Theta(F)) = \frac{1}{N} \ln Z(\Theta(F)) \tag{2.11}$$

When the instance F is randomly drawn (for instance from the random k-SAT ensemble), the measure $\mu_{\Theta(F)}$ becomes itself random. The goal of the cavity method is to determine the typical properties of the measure $\mu_{\Theta(F)}$ and of the free entropy density $\Phi(\Theta(F))$. To do so one defines the quenched free entropy density:

$$\Phi^q(\Theta) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}[\ln Z(\Theta(F))] \tag{2.12}$$

where the average $\mathbb{E}$ is taken on the random ensemble of instances. In many statistical physics systems, the free entropy density $\Phi(\Theta(F))$ concentrates thanks to the self-averaging phenomenon, therefore the quenched free-entropy density indeed contains information on the properties of the typical measures $\mu_{\Theta(F)}$ drawn from the random ensemble. Usually, the quenched free energy cannot be evaluated exactly for random CSPs. Its value can however be estimated with the cavity method, that we will describe in the next chapters. There is a natural upper bound on the quenched free entropy density, provided by the annealed free entropy density:

$$\Phi^a(\Theta) = \lim_{\mathcal{N}\to\infty} \frac{1}{N} \ln \mathbb{E}[Z(\Theta(F))] \tag{2.13}$$

From Jensen's inequality applied on the random variable $Z(\Theta(F))$ one obtains

$$\Phi^q(\Theta) \leq \Phi^a(\Theta) \tag{2.14}$$

2.3 Overview of the phase transitions in random CSPs

The cavity method predicts several phase transitions affecting the structure of the set of solutions of typical instances, in the satisfiable phase. We give in this section a qualitative description of the phase transition, and will explain the cavity computation by itself in the next chapters.

2.3.1 The clustering transition

The clustering transition occurs in the satisfiable phase $\alpha < \alpha_{sat}(k)$. It is also known as the reconstruction or the dynamical transition. This transition can be described from various perspectives.

Looking at the set of solutions of typical instances, the clustering threshold separates a regime where the set of typical solutions is rather well connected, any solution can be reached from any other by via nearby intermediate solutions. Moreover, this rearrangement is only on a non-extensive number of variables. Above $\alpha_d(k)$ the typical solutions breaks into an exponential number of clusters, or pure states, which are internally well connected, but separated one from each other by free-energy barriers.

Note that this definition concerns typical solutions with respect to the measure $\mu_{\Theta(F)}$ chosen to describe the set of solutions. In fact, the set of solutions could be connected, but by narrow, winding paths that a Markov chain would take exponential time to find, and changing the measure could raise or lower the total probability of the paths from one region of state space to the other, changing the mixing time from exponential to polynomial. Therefore the value of the clustering threshold depends on this choice, and we denote explicitly $\alpha_d(k, \Theta)$ the clustering threshold. Usually, the uniform measure is chosen to describe the set of solutions, but it is possible to introduce non-uniform measures, and study how the clustering threshold is affected by this change.

As already mentioned, the satisfiability threshold $\alpha_{sat}(k)$ separates the SAT phase where typical instances are satisfiables w.h.p., and the UNSAT phase where typical instances are unsatisfiables w.h.p.. This definition has to do with the whole set of solutions of the instance, not only typical solutions, therefore the value $\alpha_{sat}(k)$ does not depend on the specific choice of the measure μ_Θ.

The clustering transition can also be interpreted as the birth of the point-to-set correlation function under the probability measure μ chosen to describe the set of solutions. It was introduced in [14] by A. Montanari and G. Semerjian in 2006 (see also [58]). Given a variable node i and a set of variable nodes B, the point-to-set correlation function is defined for spin variables as follows (we used the definition in [59] equation (19)):

$$C(i, B) = \sum_{\underline{\sigma}_B} \mathbb{P}_\mu(\underline{\sigma}_B) \left(\sum_{\sigma_i} \mathbb{P}_\mu(\sigma_i | \underline{\sigma}_B) \sigma_i \right)^2 - \left(\sum_{\sigma_i} \mathbb{P}_\mu(\sigma_i) \sigma_i \right)^2 \tag{2.15}$$

Let $d(i, B)$ be the distance from i to B on the graph of interactions. This correlation measure how much information on the value of one spin variable i (the point) is provided by the observation of all spins in B (the set) at distance $d(i, B)$ from it. In the unclustered regime the point-to-set correlation $C(i, B)$ vanishes when $d(i, B)$ grows, while it does not decay to 0 in the clustered regime. In [14] A. Montanari and G. Semerjian show that this correlation implies the divergence of the relaxation time of local stochastic processes that respect the detailed balance condition (such as the Monte Carlo Markov chains), which justifies the terminology dynamic transition. One can understand this slow

down in the dynamics from the appearance of energetic barriers separating the clusters.

This transition also implies the solvability of the tree reconstruction problem introduced in [13] in 2003 by E. Mossel and Y. Peres. For the first discussion of the connection between 1RSB equations and the reconstruction problem, see the paper from M. Mézard and A. Montanari [16]. We will present this problem and explain how to compute the clustering threshold in its formalism in the chapter 7 (section 7.3).

In the cavity method treatment of random CSPs the clustering threshold is defined as the appearance of a non-trivial solution of the one step of Replica Symmetry Breaking (1RSB) equation with Parisi parameter $\mathcal{X} = 1$.

2.3.2 Cluster decomposition

In the unclustered phase, the typical solutions belongs to a single cluster. The thermodynamic properties of the measure μ are well described by the Replica Symmetric (RS) cavity method, that in particular provides an estimate Φ^{RS} of the quenched free entropy density $\Phi^q = \lim_{N\to\infty}(1/N)\mathbb{E}[\ln Z]$.

The RSB phenomenon described in the framework of spin glasses occurs in random CSPs precisely at the clustering transition. The assumption of the RS cavity method are not valid anymore, and one uses instead the one replica-symmetry-breaking (1RSB) cavity method. In this setting, one assumes that the solution set Ω splits into an exponential number of disjoint clusters (also called pure states) $\{\Omega_c\}$, and that the restriction μ_c of the measure μ to one of these clusters is efficiently described by the RS cavity method. For instance, the point-to-set correlation $C(i, B)$ associated with μ_c is vanishing in the large $d(i, B)$ limit, while μ has a non vanishing point-to-set correlation function. Let us write $\mu(\underline{\sigma}) = \omega(\underline{\sigma})/Z$, with $\omega(\underline{\sigma})$ the weight of the configuration $\underline{\sigma}$, and $Z = \sum_{\underline{\sigma}} \omega(\underline{\sigma})$ the partition function. The partition function Z is decomposed into the contribution of each clusters:

$$Z = \sum_{\underline{\sigma}} \omega(\underline{\sigma}) = \sum_{c} Z(c);\ Z(c) = \sum_{\underline{\sigma}\in\Omega_c} \omega(\underline{\sigma}) \tag{2.16}$$

where Ω_c is the set of solutions belonging to the cluster c. One can decompose the measure μ as follows:

$$\mu(\underline{\sigma}) = \sum_{c} p(c)\mu_c(\underline{\sigma}) \tag{2.17}$$

with μ_c the restriction of the measure μ to configurations in the cluster c:

$$\mu_c(\underline{\sigma}) = \frac{\mathbb{I}[\underline{\sigma} \in \Omega_c]\omega(\underline{\sigma})}{Z(c)} \tag{2.18}$$

and $p(c) = Z(c)/Z$ is the distribution over the clusters.

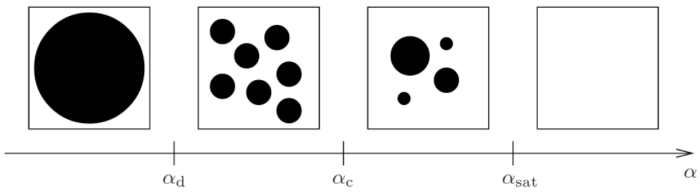

Figure 2.1: Sketch of the transitions undergone by the set of solutions. The set of solutions is in black, the rest of the configuration space is in white.

2.3.3 Complexity and the condensation transition

Let $\phi_c = (1/N) \ln Z(c)$ be the free entropy density associated with the configurations of the cluster c. At the leading order in N, the number of clusters with a given free entropy density ϕ is assumed to be exponential in N: $e^{N\Sigma(\phi)}$, where $\Sigma(\phi)$ is called the complexity. $\Sigma(\phi)$ is assumed to be self-averaging, and plays the role of an entropy density for the set of clusters. One can re-write the partition function Z as an integral over ϕ:

$$Z = \int \mathrm{d}\phi e^{N[\phi+\Sigma(\phi)]} \tag{2.19}$$

The total free entropy density $\Phi = (1/N) \ln Z$ can be evaluated via the Laplace method:

$$\Phi = \sup_{\phi:\Sigma(\phi)\geq 0} [\Sigma(\phi) + \phi] \tag{2.20}$$

the condition $\Sigma(\phi) \geq 0$ ensures that the clusters exist in the thermodynamic (large N) limit. The total free entropy thus results in a competition between clusters of large free entropy ϕ, and the most numerous clusters (with large $\Sigma(\phi)$). Let ϕ^* be the value achieving the supremum. Under the law μ, we say that a solution is typical if it belongs to a cluster of free entropy density ϕ^*. If $\Sigma(\phi^*) > 0$, there is an exponential number of clusters contributing to the total free entropy density. If $\Sigma(\phi^*) = 0$ instead, there is only a sub-exponential number of dominant clusters.

Increasing the value of α, there is a transition from a phase where $\Sigma(\phi^*) > 0$ toward a phase where $\Sigma(\phi^*) = 0$. It is called the condensation transition, and denoted $\alpha_c(k, \Theta)$, where we have emphasized the dependence of this transition on the choice of the measure describing the set of solutions μ_Θ.

Note that by construction one has $\alpha_d(k, \Theta) \leq \alpha_c(k, \Theta) \leq \alpha_{sat}(k)$. Figure 2.1 shows a sketch of the dynamical, condensation and satisfiability transitions on the set of solutions.

2.3.4 Computing the complexity function

In order to compute the condensation threshold α_c one needs to compute the complexity function Σ. To do so (see [60]), one introduces the following potential, called replicated entropy:

$$\Phi_1(\mathcal{X}) = \frac{1}{N} \ln Z_1(\mathcal{X});\ Z_1(\mathcal{X}) = \sum_c (Z(c))^{\mathcal{X}} \tag{2.21}$$

where $\mathcal{X}$ is a real parameter called the Parisi parameter (the subscript '1' is just here to remind that this function is used in the 1RSB framework). It plays the role of an inverse temperature for the modified distribution over the set of clusters:

$$p_{\mathcal{X}}(c) = \frac{(Z(c))^{\mathcal{X}}}{Z_1(\mathcal{X})} = \frac{e^{\mathcal{X} N \phi_c}}{Z_1(\mathcal{X})} \tag{2.22}$$

it allows to weight differently the clusters according to their free entropy. As we did before for the true partition function Z, one can re-write the modified partition function $Z_1(\mathcal{X})$ as an integral over ϕ:

$$Z_1(\mathcal{X}) = \int \mathrm{d}\phi e^{N[\mathcal{X}\phi + \Sigma(\phi)]} \tag{2.23}$$

Evaluating the integral via the Laplace method yields

$$\Phi_1(\mathcal{X}) = \sup_{\phi} [\Sigma(\phi) + \mathcal{X}\phi] = \Sigma(\phi_1^*(\mathcal{X})) + \mathcal{X}\phi_1^*(\mathcal{X}) \tag{2.24}$$

where we defined

$$\phi_1^*(\mathcal{X}) = \operatorname*{argmax}_{\phi} [\mathcal{X}\phi + \Sigma(\phi)] \tag{2.25}$$

Thus the Parisi potential $\Phi_1(\mathcal{X})$ and the complexity $\Sigma(\phi)$ are Legendre transforms of each other. One can invert this relation to get

$$\Sigma(\phi_1^*(\mathcal{X})) = \Phi_1(\mathcal{X}) - \mathcal{X} \frac{\mathrm{d}\Phi_1(\mathcal{X})}{\mathrm{d}\mathcal{X}} \tag{2.26}$$

The potential $\Phi_1(\mathcal{X})$ and its derivative can be estimated with the 1RSB cavity method, leading to an estimation of the complexity function. In most of the cases, the complexity function is a concave function of ϕ, as represented in figure 2.2

In the uncondensed regime $\alpha_d \leq \alpha \leq \alpha_c$, the expression (2.20) achieves its supremum for a value ϕ^* which satisfies the constraint $\Sigma(\phi) \geq 0$, therefore $\phi^* = \phi_1^*(\mathcal{X})$ and $\Phi = \Phi_1(\mathcal{X} = 1)$. It can be checked that the prediction $\Phi_1(\mathcal{X} = 1)$ coincides with the RS prediction Φ^{RS} of the free entropy denstity. Therefore in the clustered-uncondensed regime the thermodynamic prediction of the RS assumption is valid. The phase transition α_d does not affect thermodynamic quantities. As we have seen, this transition can be detected by the slow down of the dynamics, therefore this phase is also called the dynamic 1RSB (d1RSB)

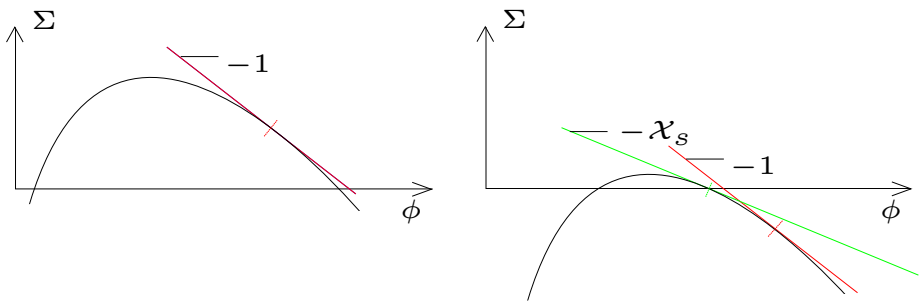

Figure 2.2: Left: uncondensed regime the supremum of (2.20) is achieved for ϕ^* such that $\Sigma(\phi^*) > 0$ and $(\mathrm{d}\Sigma(\phi)/\mathrm{d}\phi) = -1$ (the red tangent has slope -1). Right: condensed regime, the supremum of (2.20) is achieved for ϕ^* such that $\Sigma(\phi^*) = 0$ and $(\mathrm{d}\Sigma(\phi)/\mathrm{d}\phi) = -\mathcal{X}_s$ (green tangent), the value of ϕ such that $(\mathrm{d}\Sigma(\phi)/\mathrm{d}\phi) = -1$ has a negative complexity.

phase. The order parameter that detects this phase transition is the point-to-set correlation function.

In the condensed regime $\alpha_c \leq \alpha \leq \alpha_{sat}$, the value of ϕ such that $(\mathrm{d}\Sigma(\phi)/\mathrm{d}\phi) = -1$ has a negative complexity, and the supremum of (2.20) is achieved for a smaller value $\mathcal{X}_s < 1$: $\Phi = \Phi_1(\mathcal{X}_s)$. The free entropy density is not described anymore by the RS prediction Φ^{RS}, and the system undergoes a first-order phase transition. The condensed phase is therefore also called the static 1RSB (s1RSB) phase. Note that values of the Parisi parameter such that $\mathcal{X} > 1$ never contribute to the computation of the thermodynamic quantities (it can be however useful to look at these values if one is interested in atypical cluster sizes).

One can resume these observations with the following expression

$$\Phi = \inf_{\mathcal{X}\in[0,1]} \frac{\Phi_1(\mathcal{X})}{\mathcal{X}} \tag{2.27}$$

which is valid in all phases. In the unclustered phase $\alpha < \alpha_d(k,\Theta)$ there is only one cluster, one has $\Sigma(\phi) = 0$ for all ϕ, therefore from the definition of $Z_1(\mathcal{X})$ one gets $\Phi_1(\mathcal{X}) = \mathcal{X}\phi = \mathcal{X}\Phi$, and the equation (2.27) is correct. In the clustered phase, call $\mathcal{X}^*$ the minimizer of the expression (2.27). In the uncondensed phase one has $\mathcal{X}^* = 1$. Indeed from the expression (2.24) of $\Phi_1(\mathcal{X})$ one sees that

$$\frac{\mathrm{d}(\Phi_1(\mathcal{X})/\mathcal{X})}{\mathrm{d}\mathcal{X}} = -\frac{\Sigma(\phi(\mathcal{X}))}{\mathcal{X}^2} \tag{2.28}$$

hence in the uncondensed phase since $\Sigma(\phi(\mathcal{X})) = \Sigma(\phi^*(\mathcal{X})) > 0$ the function $\Phi_1(\mathcal{X})/\mathcal{X}$ is decreasing and reach its minimum at $\mathcal{X} = 1$. Therefore the expression (2.27) indeed gives $\Phi = \Phi_1(1)$. In the condensed regime, one has

$\Sigma(\phi(\mathcal{X}=1)) < 0$, therefore the equation (2.27) selects the value $\mathcal{X}_s$ such that $\Sigma(\phi(\mathcal{X}_s)) = 0$. Since $\Phi_1(\mathcal{X}_s) = \mathcal{X}_s\phi(\mathcal{X}_s)$ one gets back to the correct result: $\Phi = \Phi_1(\mathcal{X}_s)$.

2.3.5 Computing α_c and α_{sat} from the complexity

In practice, to compute the condensation threshold one needs to compute the value of the complexity function at $\mathcal{X}=1$: $\Sigma(\mathcal{X}=1) = \Sigma(\phi_1^*(\mathcal{X}=1)$ from the expression (2.26), and let α_c be the value at which $\Sigma(\mathcal{X}=1)$ vanishes.

The complexity function allows to predict the value of the satisfiability threshold. Note that at $\mathcal{X}=0$, the partition function $Z_1(\mathcal{X}=0)$ counts the number of clusters, independently of their size. Note also that for $\mathcal{X}=0$ one has $\Phi_1(\mathcal{X}=0) = \sup_\phi[\Sigma(\phi)] = \Sigma(\phi_1^*(\mathcal{X}=0))$. Therefore when $\Sigma(\phi_1^*(\mathcal{X}=0))$ becomes negative it means that for typical large instances all the clusters have disappeared, hence that there is no solution.

2.3.6 Rigidity and freezing transitions

From the notion of clusters, one defines the notion of frozen variable inside a cluster. They are variables that take the same value in all the solutions of the corresponding cluster. One defines the rigidity transition $\alpha_r(k,\Theta)$. Above the rigidity transition typical solutions (with respect to μ_Θ) contain an extensive number of frozen variables. One defines the freezing transition $\alpha_f(k)$, above which all solutions contain an extensive number of frozen variable. The freezing transition concern all solutions, not only typical one, and thus is harder to compute than the rigidity transition. The freezing threshold has been introduced in [26], its value has been computed using a large deviation derivation, for the bicoloring of hypergraphs. By definition the following inequality must hold: $\alpha_d(k,\Theta) \leq \alpha_r(k,\Theta) \leq \alpha_f(k) \leq \alpha_{sat}(k)$.

One says that a solution is frozen when it contains an extensive number of frozen variables. Experimentally, it is found that several incomplete algorithms (such as simulated annealing or focused algorithms that we will present in the next chapter) only found unfrozen solutions. Even in the rigidity phase $\alpha_r < \alpha < \alpha_f$ where typical solutions are frozen, it seems that it is easier for these algorithms to find untypical unfrozen solutions. The freezing transition thus appears as a good candidate for an upper bound on the algorithmic barrier $\alpha_{alg}(k)$. Indeed, one can hope that as long there is some untypical unfrozen solutions, an efficient algorithm should be able to find them. In [26], a biased measure on the set of solutions is introduced in order to turn to typical the unfrozen solutions that were untypical in the uniform measure μ_u. In the new biased measure, the rigidity threshold is increased up to densities close to the satisfiability threshold.

k	α_d	α_c	α_{sat} [7]	α_r
3	3.86	3.86	4.267	
4	9.38	9.547	9.931	9.88
5	19.16	20.80	21.117	
6	36.53	43.08	43.37	39.87 [62]

Table 2.1: clustering, condensation, satisfiability and rigidity thresholds for the random k-SAT ensemble.

q	l_d	l_c	l_{sat}
4	9	10	10
5	14	14	15
6	18	19	20

Table 2.2: clustering, condensation and satisfiability thresholds for the q-COL $(2, l+1)$-regular ensemble.

2.3.7 Some values of the thresholds

We now give some values of the thresholds introduced above. In the table 2.1 the threshold values for the random k-SAT ensemble for the uniform measure over the set of solutions (see [59] Table 1.). Note the qualitative change between $k = 3$ and $k \geq 4$. At $k = 3$ the clustering threshold equals the condensation threshold. In that specific case, the phase transition separates the unclustered phase from a full-RSB phase, and the complexity computed under the 1RSB approximation (see section 2.3.4) is not valid. For $k \geq 4$ instead there a clustered-uncondensed phase. The thresholds for the q coloring problem defined on the random $(l+1)$-regular graph ensemble (with $k = 2$) are shown in the table 2.2 (see [61]). The convention for l_{sat} is that the random graphs are w.h.p. uncolorable if $l \geq l_{sat}(q)$. Similarly, the clustered (resp. condensed) phase is located in the region $l \geq l_d$ (resp. $l \geq l_c$). In the table 2.3 we give some of the thresholds for the bicoloring problem on the random $(k, l+1)$-regular ensemble. These results are taken from [26]. Note that even if the model is only defined for integer values of l, the threshold values of the rigidity l_r and the condensation l_c are given as real numbers, since these two thresholds can be expressed as analytic functions of l. One has to round this real threshold to its two nearest integers, to find the largest (resp. smallest) l such that the property (the presence of frozen variables, or the condensation of the set of solutions) is true (resp. false) in the thermodynamic limit. Note that we have made the arbitrarily choice to show the thresholds of the k-SAT problem on the Erdős Rényi ensemble, and the thresholds of the q-coloring problem on the regular ensemble, but in practice there is no technical limitation to compute the thresholds of these two problems on any of these two ensembles.

k	l_d	l_r	l_c	l_{sat}
3	5	6.673	5	5.742
4	17	20.64	19	18.89
5	47	51.45	51	51.50
6	108	117.16	128	128.50
7		255.10		305.34
8		541.99		703.99

Table 2.3: clustering, rigidity, condensation and satisfiability thresholds for the bicoloring problem on the $(k, l+1)$-regular ensemble.

Chapter 3

Local Search Algorithms

We will present now a family of incomplete algorithms that search for solutions of an instance of CSP. Starting from an initial configuration $\underline{x}_0$, they perform a walk in the configuration space, with a bias toward the set of solutions. The term local means that at each step, only one variable (or sometimes a small number with respect to N) is flipped.

3.1 Simulated Annealing

We have seen that one can re-write a CSP into the form of a statistical physics problem. We can now apply the tools developed for physical problems to CSPs. In particular, the Monte Carlo method allows to sample from a Boltzmann distribution μ_β as defined in (1.6) with finite inverse temperature β. In the zero-temperature limit ($\beta \to \infty$) the Monte Carlo method cannot be applied directly, because the convergence hypotheses are not fulfilled (specifically irreducibility). To sample from the Boltzmann distribution with $\beta \to \infty$, and thus extract the solution of the instance F, one can use the Simulated Annealing algorithm. This method performs several steps of Monte Carlo, decreasing the temperature at each step. The application of Simulated Annealing for optimization problem has been introduced in [25] by S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi in 1983.

3.1.1 Monte Carlo method

This method is an simple way to sample from a Boltzmann distribution at finite temperature, when the number N of variables gets large. One constructs a Monte Carlo Markov Chain which is guaranteed to converge to the target distribution μ_β. The transition rules $\omega(\underline{x} \to \underline{x}')$ of the Markov Chain must satisfy the following conditions in order to ensure the convergence:

1. irreducibility: for any pair of configuration $\underline{x}, \underline{x}'$ there exists a path $(\underline{x}_0, \ldots, \underline{x}_n)$ of length n with $\underline{x}_0 = \underline{x}, \underline{x}_n = \underline{x}'$, and with non-zero transition probabilities: $\omega(\underline{x}_i \to \underline{x}_{i+1}) > 0$ for all $i = 0, \ldots, n-1$

2. aperiodicity: for any pair of configurations $\underline{x}$ and $\underline{x}'$ there exists a positive integer $n(\underline{x}, \underline{x}')$ such that for any $n \geq n(\underline{x}, \underline{x}')$ there exists a path of length n connecting $\underline{x}$ to $\underline{x}'$ with non-zero probability. (This is automatically verified for an irreducible chain if one enforces the additional condition $\omega(\underline{x} \to \underline{x}) > 0$)

3. stationarity: the distribution $\mu_\beta(\underline{x})$ is stationary with respect to the transition probabilities $\omega(\underline{x} \to \underline{x}')$:

$$\sum_{\underline{x}} \mu_\beta(\underline{x}) \omega(\underline{x} \to \underline{x}') = \mu_\beta(\underline{x}') \tag{3.1}$$

The last condition is also called the global balance condition. A stronger condition is often enforced, called the detailed balance condition. For any pair $\underline{x}, \underline{x}'$ such that either $\omega(\underline{x} \to \underline{x}') > 0$ or $\omega(\underline{x}' \to \underline{x}) > 0$ one requires:

$$\mu_\beta(\underline{x}) \omega(\underline{x} \to \underline{x}') = \mu_\beta(\underline{x}') \omega(\underline{x}' \to \underline{x}) \tag{3.2}$$

Note that the detailed balance condition implies the global balance condition. In many cases the detailed balance is used as an easy way to prove global balance.

Under the above conditions on the transition probabilities, the MCMC converges to the target distribution μ (see [44] theorem 4.12):

Theorem 3 *Assume that the rates $\omega(\underline{x} \to \underline{x}')$ satisfy the three conditions mentionned above. Let $\underline{X}_0, \ldots, \underline{X}_t, \ldots$ be random variables distributed according to the Markov Chain with rates $\omega(\underline{x} \to \underline{x}')$, and initial condition $\underline{X}_0 = \underline{x}_0$. Let $f : \chi^N \to \mathbb{R}$ be a real valued function, then:*

1. *The probability distribution of $\underline{X}_t$ converges to the stationary distribution:*

$$\lim_{t \to \infty} \mathbb{P}[\underline{X}_t = \underline{x}] = \mu_\beta(\underline{x}) \tag{3.3}$$

2. *Time averages converge to averages over the stationary distribution*

$$\lim_{t \to \infty} \frac{1}{t} \sum_{s=1}^{t} f(\underline{X}_s) = \sum_{\underline{x}} \mu_\beta(\underline{x}) f(\underline{x}) \text{ almost surely} \tag{3.4}$$

The Monte Carlo method consists of defining the transition rules of a Monte Carlo Markov Chain that converges toward the desired distribution, and then simulating it on a computer until convergence. The Metropolis and the Glauber dynamics are two examples of MCMC adapted to a Boltzmann measure $\mu_\beta(\underline{\sigma}) = e^{-\beta E(\underline{\sigma})}/Z(\beta)$ defined over a set of spin variables.

3.1.2 Metropolis algorithm

Let $\underline{\sigma}^{(i)}$ be the configurations that coincides with $\underline{\sigma}$ except for the variable i: $\sigma_i^{(i)} = -\sigma_i$. Let $\Delta_i = E(\underline{\sigma}^{(i)}) - E(\underline{\sigma})$. At each step, a site i is chosen uniformly at random, and its spin variable is flipped with probability

$$\omega_i(\underline{\sigma}) = \exp[-\beta \max(\Delta_i, 0)] \tag{3.5}$$

The transition probabilities can be expressed in terms of the ω_i:

$$\omega(\underline{\sigma} \to \underline{\tau}) = \frac{1}{N} \sum_{i=1}^{N} \omega_i(\underline{\sigma}) \mathbb{I}[\underline{\tau} = \underline{\sigma}^{(i)}] + \left[1 - \frac{1}{N} \sum_{i=1}^{N} \omega_i(\underline{\sigma})\right] \mathbb{I}[\underline{\tau} = \underline{\sigma}] \tag{3.6}$$

One can check that $\omega(\underline{\sigma} \to \underline{\tau})$ satisfy all the conditions that ensure the convergence of the Markov Chain towards the measure μ_β. The Algorithm 3 shows an implementation of the Metropolis algorithm.

Algorithm 3 Metropolis Algorithm

Generate $\underline{\sigma}_0$ uniformly at random from $\{+1, -1\}^N$
for $t = 1$ to $t = T$ **do**
  Draw i uniformly at random in $\{1, \ldots, N\}$
  Compute $\omega_i(\underline{\sigma}_{t-1})$ from (3.5)
  Draw a random number p uniformly in $[0, 1]$
  if $p < \omega_i(\underline{\sigma}_{t-1})$ **then**
    Set $\underline{\sigma}_t = \underline{\sigma}_{t-1}^{(i)}$
  else
    Set $\underline{\sigma}_t = \underline{\sigma}_{t-1}$
  end if
end for
return $\underline{\sigma}_T$

3.1.3 Heat bath algorithm

Another possibility is to use the Glauber dynamics. Compared to the Metropolis algorithm, the only change is in the expression of ω_i:

$$\omega_i(\underline{\sigma}) = \frac{\mu_\beta(\underline{\sigma}^{(i)})}{\mu_\beta(\underline{\sigma}) + \mu_\beta(\underline{\sigma}^{(i)})} = \frac{1}{2}\left[1 - \tanh\left(\frac{\beta \Delta_i}{2}\right)\right] \tag{3.7}$$

3.1.4 Relaxation time scale of the Monte Carlo dynamics

The performance of the Monte Carlo method is affected by the phase transitions occuring in statistical physics models. In the low temperature (or high density) phase, the dynamics become very slow, and the time needed for the Markov process to converge increases. In order to quantify this slow down, it

is useful to introduce a time scale for the equilibrium dynamics. Starting from a configuration $\underline{\sigma}(0)$ drawn from the stationary distribution μ_β, one defines for each variable i the time correlation function:

$$D_i(t) = \langle \sigma_i(0)\sigma_i(t)\rangle - \langle \sigma_i\rangle^2 \tag{3.8}$$

the average $\langle\cdot\rangle$ is taken with respect to the realization of the Monte Carlo dynamics and to the choice of $\underline{\sigma}(0)$. For finite size systems, this time correlation function is expected to be positive and decreasing, with $D_i(0) = 1$ and a vanishing limit when $t \to \infty$. In the large t limit, the Monte-Carlo Markov chain has indeed converged to the stationary distribution μ_β, and $\sigma_i(0)$ and $\sigma_i(t)$ are decorrelated.

From the time correlation function $D_i(t)$, one can extract the relaxation time scale, that is the smallest integer τ such that $D_i(t)$ decreases below a given threshold $\epsilon > 0$:

$$\tau_i(\epsilon) = \min\{\tau > 0 \text{ such that } D_i(t) \leq \epsilon \ \forall t \geq \tau\} \tag{3.9}$$

In [63], A. Montanari and G. Semerjian show that this time scale can be related to the point-to-set correlation function:

$$C_i(r) = C(i, B(i,r)) = |\langle \sigma_i \langle \sigma_i\rangle_{B_i(r)}\rangle - \langle \sigma_i\rangle^2| \tag{3.10}$$

where $B(i,r)$ is the subset of variable nodes j at distance from i smaller than the integer r: $d_{ij} \leq r$. Using the notation $\underline{\sigma}_{\sim i,r} = \underline{\sigma}_{V\setminus B(i,r)}$, the average $\langle\cdot\rangle_{B_i(r)}$ is taken with respect to the conditional distribution $\mathbb{P}_{\mu_\beta}[\cdot|\underline{\sigma}_{\sim i,r}]$. The correlation length is the smallest integer l such that $C_i(r)$ decreases below ϵ:

$$l_i(\epsilon) = \min\{l \geq 0 : \quad C_i(r) \leq \epsilon, \ \forall r \geq l\} \tag{3.11}$$

It is shown in [63] that for the Glauber dynamics at finite β on a graphical model (with local interactions) as defined in the paragraph 1.5.3, the following inequality holds:

$$l_i(2\sqrt{2}\epsilon) \leq C\tau_i(\epsilon) \tag{3.12}$$

with $C = 2ke$ a constant.

In the unclustered regime $\alpha < \alpha_d$, the point-to-set correlation function $C_i(r)$ for typical large instances decreases to 0 when r gets large. The length scale $l_i(\epsilon)$ is thus finite in this phase. In the clustered regime $\alpha > \alpha_d$, the point-to-set correlation function has instead a non vanishing limit. The length scale $l_i(\epsilon)$ is set by convention to be equal to the maximum distance from i to a vertex in G, that scales as $O(\ln N)$ for typical instances of the ER or the regular ensemble. Therefore the relaxation time diverges with N for typical instances. It is also shown in [63] that for the p-spin glass model, in the low temperature phase, that the relaxation time has an exponential scaling in N. One can also expect a similar scaling for the random k-SAT.

Intuitively, the divergence of the relaxation time can be understood from the presence of free-energy barriers. In the clustered phase, the measure μ_β

decomposes into an exponential number of clusters. These clusters are well separated by free-energy barriers: the total weight of all configurations that are not in a cluster is exponentially small in N. Looking at the equilibrium dynamics, one starts from a configuration $\underline{\sigma}(0)$ inside a cluster c. Because of the presence of the exponentially large barriers separating the clusters, the dynamic will stay inside the cluster for a long time (assumed to be exponential), before exploring other clusters at larger time scales.

3.1.5 Cooling scheme

The global balance (stationarity) condition for the Boltzmann distribution at finite β writes:

$$\sum_{\underline{x}} e^{-\beta(E(\underline{x})-E(\underline{x}'))}\omega(\underline{x} \to \underline{x}') = 1 \tag{3.13}$$

Since each of the terms are smaller or equal than 1, one has $\omega(\underline{x} \to \underline{x}') \leq e^{-\beta(E(\underline{x}')-E(\underline{x}))}$. Therefore in the $\beta \to \infty$ limit, for any $\underline{x}, \underline{x}'$ such that $E(\underline{x}') > E(\underline{x})$ one has $\omega(\underline{x} \to \underline{x}') \to 0$. This means that the paths are forced to be non-increasing in energy, i.e. the Markov Chain is not irreducible. Intuitively, this means that the algorithm can get trapped in a local minima and cannot escape from it since it cannot climb up in energy.

Since the Markov Chain is irreducible for any finite temperature, another idea would then be to work with a fixed finite temperature β which is small enough. One uses a MCMC to sample from it, and since the set of solutions is included in the support of the measure μ_β, a strategy would be to wait until one generates a solution. However this strategy might also fail, because the probability under μ_β of having a zero-energy configuration is exponentially small in N. Indeed one has $\mathbb{P}_{\mu_\beta}[E(\underline{\sigma}) = 0] = \exp[N(\lim_{\beta' \to \infty} \phi_N(\beta') - \phi_N(\beta)]$, with $\phi_N(\beta) = (1/N) \ln Z(\beta)$ the free entropy density associated with μ_β, which is a decreasing function of β.

To overcome these difficulties, the Simulated Annealing algorithm uses a cooling scheme. It performs n steps of Monte Carlo Markov Chain (MCMC), with temperatures $\beta_1 < \beta_2 < \cdots < \beta_n$. The final configuration of the step i is used as the initial configuration of step $i + 1$. At the beginning, small values of β (high temperature) allows the dynamics to escape from local minima and equilibrate on the space of configurations. At the end, the large value of β (small temperature) allows to find the minima (hoping it is a global one) inside the basin where the dynamics has falled.

3.2 Focused algorithms

In the Metropolis algorithm 3, the new variable to be flipped is chosen uniformly at random among the set of N variables. One can try to enhance the performances by choosing it according to some heuristics. In focused search algorithms, one chooses the new variable at random among the variables that appears in violated clauses.

3.2.1 RandomWalkSAT algorithm

We present first the RandomWalkSAT algorithm 4 that performs a simple random walk. This algorithm has been proposed by C.H. Papadimitriou in [64]). It takes as input the CNF formula F and a maximum number T of steps and return either a SAT formula or 'not found'. The algorithm starts with a random initial configuration, for instance independently for each variable one can take x_i uniformly in $\chi = \{-1, +1\}$. In [64], C.H. Papadimitriou showed that RandomWalkSAT solves w.h.p. any satisfiable instance of the random 2-SAT ensemble, in a number of steps $T = O(N^2)$. In [65], U. Schöning showed the following result, true for any instance F of the 3-SAT problem. Let R be the number of restarts of the RandomWalkSAT algorithm, each time with an initial condition drawn uniformly at random. Let $T = 3N$ be the number of flips of each restarts. If none of the R successive runs of the RandomWalkSAT on the instance F has provided a solution, then the probability that F is satisfiable is lower than $e^{-(3/4)^N R}$. Therefore after many restarts $R \gg (4/3)^N$, if there are solutions, the probability that RandomWalkSAT has not found one of them is very small.

Algorithm 4 RandomWalkSAT(F,T)

```
Initialize σ to a random assignment
for T times do
    if σ is SAT then
        return σ and stop
    else
        choose an unsatisfied clause a uniformly at random
        choose an index i uniformly at random in ∂a
        Set σ = σ^(i)
    end if
end for
return 'not found'
```

3.2.2 WalkSAT algorithm

The algorithm RandomWalkSAT 4 makes essentially random steps although it focuses on unsatisfied clauses. It is possible instead to choose the new variable according to some specific properties. For instance one can use the energy-cost function $E(\underline{x}) = \sum_{a=1}^{M}(1 - c_a(\underline{x}_{\partial a}))$ defined in equation (1.1), and decide to flip the variable that lead to the largest decrease in energy. An algorithm based on this rule is called a greedy algorithm. However the risk with this greedy choice is to be trapped in local minima (similarly as the case of Metropolis algorithm with zero-temperature). A good strategy is to design a mixed algorithm that performs both random and greedy steps. The idea is that the greedy steps drive the assignments toward the configurations with low energy, while the random steps allow to escape from local minima. The algorithm WalkSAT has been

proposed in 1996 by B. Selman, H. Kautz, B. Cohen [66], and is presented in the algorithm 5. As for RandomWalkSAT, it takes as input a CNF formula F, a maximum number T of iterations, and additionally a probability $p \in [0,1]$ of performing a random step (with probability p) or a greedy step (with probability $1-p$). One can then optimize on p in order to increase the performances.

Algorithm 5 WalkSAT (F, T, p)

```
Initialize σ̲ to a random assignment
for T times do
    if σ̲ is SAT then
        return σ̲ and stop
    else
        Let r be uniform at random in [0,1]
        if r < 1 − p then
            For each i let Δ_i = E(σ̲^(i)) − E(σ̲)
            Flip a variable σ_i for which Δ_i is minimal
        else
            Choose a violated clause a uniformly at random
            Flip uniformly at random a variable σ_i, i ∈ ∂a
        end if
    end if
end for
```

3.2.3 Focused Metropolis algorithm

Let us mention that the Metropolis algorithm 3 can be turned into a focused algorithm 6. The Focused Metropolis Algorithm (FMS) has been introduced in 2005 by S. Seitz, M. Alava and P. Orponen in [67].

Algorithm 6 Focused Metropolis Search

```
S = random assignment of values to the variables
while S is not a solution do
    C = a clause not satisfied by S selected uniformly at random
    V = a variable in C selected uniformly at random
    ΔE = change in number of unsat clauses if V is flipped in S
    if ΔE ≤ 0 then
        flip V in S
    else
        Draw p ∈ [0,1] uniformly at random
        if p < η^ΔE then
            flip V in S
        end if
    end if
end while
```

The heuristic for the choice of the new variable to be assigned, and the rule to decide if a flip is performed leads to several other variations around WalkSAT. It is hard to predict in advance which algorithm will be the most efficient. In the next chapters we will describe some of the numerical experiments that compare their performances on the random k-SAT ensemble. The focused algorithms are said to be out-of-equilibrim, because they do not satisfy the global balance condition with respect to the Boltzmann measure μ_β, therefore are not guaranteed to converge to it. However, since our goal is to find only one solution to the instance F, it is not necessary to sample correctly the distribution defined on the set of solutions.

Chapter 4

Message Passing Algorithms

4.1 Belief propagation

Consider a measure μ defined with a graphical model, as presented in chapter 1 (equation (1.9)). When the associated factor graph is a tree, Belief Propagation (BP) is a method that allows to compute the partition function Z, and the marginal with respect to μ of a variable x_i, or any subset of variables. It also provides an efficient way to sample a configuration $\underline{x}$ from μ. Moreover, all these tasks can be achived in polynomial time in the size N. In the Erdős Rényi hypergraph ensemble $G_N(k, M)$ and in the (k, l)-regular hypergraph ensemble, typical large instances are locally tree-like. In practice, the BP method is also used as an approximation on instances randomly drawn from one of these ensembles.

4.1.1 BP messages

For each edge $(i, a) \in E$ of the factor graph, define two messages $\eta_{i\to a}, \widehat{\eta}_{a\to i}$. These messages are marginal distributions on a modified factor graph in which some nodes have been removed. More precisely, $\eta_{i\to a}$ is the marginal of the variable x_i when the function node a has been removed, and $\widehat{\eta}_{a\to i}$ is the marginal of the variable x_i when all the function nodes in $\partial i \setminus a$ have been removed. Using the Markov Property, it can be shown that on a tree, the set of messages $\{\eta_{i\to a}, \widehat{\eta}_{a\to i}\}_{(i,a)\in E}$ satisfies the following set of local equations:

$$\eta_{i\to a}(x_i) = \frac{1}{z_{i,a}} \prod_{b\in\partial i\setminus a} \widehat{\eta}_{a\to i}(x_i) \tag{4.1}$$

$$\widehat{\eta}_{a\to i}(x_i) = \frac{1}{\widehat{z}_{i,a}} \sum_{\underline{x}_{\partial a\setminus i}} \omega_a(\underline{x}_{\partial a}) \prod_{j\in\partial a\setminus i} \eta_{j\to a}(x_j) \tag{4.2}$$

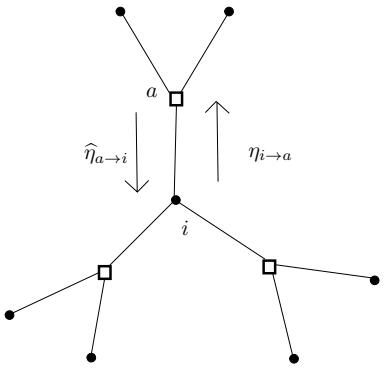

Figure 4.1: Messages $\eta_{i\to a}, \widehat{\eta}_{a\to i}$ on a small factor graph with $N = 7$ vertices and $M = 3$ hyperedges with $k = 3$.

where the $z_{i,a}, \widehat{z}_{i,a}$ are constants that ensure the normalization of the messages. These equations are called the Belief Propagation equations. Note that the message $\eta_{i\to a}$ going from $i \to a$ depends only on the messages $\{\widehat{\eta}_{b\to i}\}_{b\in\partial i\setminus a}$ coming to the variable node i, as illustrated in figure 4.1. Similarly, the message $\widehat{\eta}_{a\to i}$ going from $a \to i$ depends only on the messages $\{\eta_{j\to a}\}_{j\in\partial a\setminus i}$ coming to the variable node i. We will sometimes use the shorthand notation:

$$\eta_{i\to a} = f(\{\widehat{\eta}_{b\to i}\}_{b\in\partial i\setminus a}); \quad \widehat{\eta}_{a\to i} = g(\{\eta_{j\to a}\}_{j\in\partial a\setminus i}) \tag{4.3}$$

to denote the BP equations. The BP equations have been introduced by J. Pearl in 1988 ([68]), in the context of probabilistic inference, and by R.G. Gallager in 1962 ([69]) for decoding low density parity check codes. See also [70, 71] for a presentation of the BP algorithm.

4.1.2 BP iterations

On a finite tree, the BP equations (4.3) admit a unique solution. One can solve them using a recursive procedure. Let $\{\eta^{(t)}_{i\to a}, \widehat{\eta}^{(t)}_{a\to i}\}_{(i,a)\in E}$ be the set of messages guessed at step t. One then considers the BP equations as the fixed-point equations of the following recursive equations:

$$\eta^{(t+1)}_{i\to a} = f(\{\widehat{\eta}^{(t)}_{b\to i}\}_{b\in\partial i\setminus a}); \quad \widehat{\eta}^{(t)}_{a\to i} = g(\{\eta^{(t)}_{j\to a}\}_{j\in\partial a\setminus i}) \tag{4.4}$$

For the initial condition, a simple choice is to take uniform messages: $\eta^{(0)}_{i\to a}(x_i) = 1/|\chi|$. One can also define a probability measure P on the space of distribution over χ, and extract the initial messages i.i.d from P. The algorithm 7 (see [44]) describe the iterative procedure to find a solution to the BP equations (4.3). It takes as input a factor graph G, the set of function nodes $\{\omega_a\}_{a\in C}$ defined on G, a number t_{max} maximum of iterations, and a precision threshold ϵ. It outputs either a set of messages, or state that is has not converged.

Algorithm 7 Belief Propagation(G,$\{\omega_a\}_{a\in C}$,t_{max},ϵ)

For each edge (a, i), initialize the message $\eta_{i\to a}$
for $t = 1$ to t_{max} **do**
  Compute the messages $\{\widehat{\eta}^{(t)}_{a\to i}\}_{(i,a)\in E}$, then $\{\eta^{(t+1)}_{i\to a}\}_{(i,a)\in E}$ from (4.4)
  Let Δ be the maximum message change
  if $\Delta < \epsilon$ **then return** the set of messages $\{\eta^{(t+1)}_{i\to a}, \widehat{\eta}^{(t)}_{a\to i}\}$
  end if
end for
return "not converged"

On a tree of diameter t_{max} the algorithm 7 is guaranteed to find the set of messages solution of (4.3), independently of the choice of the initialization and of the updating scheme (4.4). In this algorithm, we have not specified the ordering of the edges (i, a) for the update of the messages. When BP is used as a heuristic on factor graphs with loops (without guarantee of convergence), one usually reshuffles the ordering, taking a random permutation of the edges before each update.

4.1.3 Marginals

Once we have the solution of the BP equations, the marginal $\mu_i(x_i)$ of a variable x_i can be constructed from the messages $\{\widehat{\eta}_{a\to i}\}_{a\in\partial i}$. Using the tree structure of the factor graph and the Markov property one can write:

$$\mu_i(x_i) \propto \prod_{a\in\partial i} \left(\sum_{\underline{\sigma}_{\partial b\setminus i}} \omega_b(\underline{\sigma}_{\partial b}) \prod_{j\in\partial b\setminus i} \eta_{j\to b}(\sigma_j) \right) \tag{4.5}$$

using the BP equation (4.2) one gets:

$$\mu_i(x_i) = \frac{1}{z_i} \prod_{a\in\partial i} \widehat{\eta}_{a\to i}(x_i) \tag{4.6}$$

where z_i is a normalization constant. One can write similarly the marginal of a subset of variables. Let C_R be a subset of function nodes, and V_R be the subset of variable nodes connected to any of the function nodes in C_R, and let R be the induced subgraph. Suppose that R is connected. Finally let ∂R be the subset of function nodes that are not in C_R but are connected to a variable node in V_R. Then the marginal μ_R of the variables $\underline{x}_R$ in V_R is

$$\mu_R(\underline{x}_R) = \frac{1}{z_R} \prod_{a\in C_R} \omega_a(\underline{x}_{\partial a}) \prod_{a\in\partial R} \widehat{\eta}_{a\to i(a)}(x_{i(a)}) \tag{4.7}$$

where for each $a \in \partial R$, $i(a)$ is the unique variable node $i \in \partial a \cup V_R$. For the particular case $C_R = \{a\}$, using the BP equation (4.1) this expression reads:

$$\mu_a(\underline{x}_{\partial a}) = \frac{1}{z_a}\omega_a(\underline{x}_{\partial a}) \prod_{i \in \partial a} \eta_{i \to a}(x_i) \tag{4.8}$$

Conditional laws

Let $\underline{x}_{\sim i}$ be the configuration of the variables in $V \backslash i$. The conditional distribution with respect to μ of $\underline{x}_{\sim i}$ given that $x_i = z$ takes the form:

$$\mathbb{P}_\mu[\underline{x}_{\sim i} | x_i = z] \propto \prod_{a \in C} \omega_a(\underline{x}_{\partial a}) \mathbb{I}[x_i = z] \tag{4.9}$$

Therefore the factor graph associated with this conditional measure is the initial factor graph, on which one has added an extra function node $\mathbb{I}[x_i = z]$ of degree 1, attached to the variable node i. One can run the BP algorithm 7 on this modified factor graph to obtain the marginal law $\mathbb{P}_\mu[x_j | x_i = z]$, for a given variable node $j \neq i$.

Consider the particular case where one wants to compute the marginal $\mathbb{P}_\mu[\{\underline{x}_{\partial a \backslash i}\}_{a \in \partial i} | x_i]$. On a tree factor graph, when the variable node i is removed, the graph splits into $|\partial i|$ disconnected subtrees $\{T_{a \to i}\}_{a \in \partial i}$, where $T_{a \to i}$ is the subtree whose root is a. Since the value of x_i is fixed, the subsets of variables $\{\underline{x}_{T_{a \to i}}\}_{a \in \partial i}$ become independent of each other:

$$\mathbb{P}_\mu[\{\underline{x}_{\partial a \backslash i}\}_{a \in \partial i} | x_i = z] \propto \prod_{a \in \partial i} \mathbb{P}_\mu[\underline{x}_{\partial a \backslash i} | x_i = z] \tag{4.10}$$

and $\mathbb{P}_\mu[\underline{x}_{\partial a \backslash i} | x_i = z]$ can be expressed in terms of the BP equation written for $\mathbb{P}_\mu[\underline{x}_{T_{a \to i}}]$:

$$\mathbb{P}_\mu[\underline{x}_{\partial a \backslash i} | x_i = z] \propto \omega_a(\underline{x}_{\partial a}) \prod_{j \in \partial a \backslash i} \eta_{j \to a}(x_j) \tag{4.11}$$

4.1.4 Bethe free entropy and Entropy

When the factor graph is a tree, one can express the free entropy density $\Phi = (1/N) \ln Z$ from the set of messages that is solution of the BP equation (4.3):

$$\Phi^{Bethe} = \frac{1}{N} \sum_{a \in C} \ln Z_a(\{\eta_{i \to a}\}_{i \in \partial a}) + \frac{1}{N} \sum_{i \in V} \ln Z_i(\{\widehat{\eta}_{a \to i}\}_{a \in \partial i}) \tag{4.12}$$

$$- \frac{1}{N} \sum_{(i,a) \in E} \ln Z_{i,a}(\eta_{i \to a}, \widehat{\eta}_{a \to i}) \tag{4.13}$$

where

$$Z_a(\{\eta_{i\to a}\}_{i\in\partial a}) = \sum_{\underline{\sigma}_{\partial a}} \omega_a(\underline{\sigma}_{\partial a}) \prod_{i\in\partial a} \eta_{i\to a}(\sigma_i) \tag{4.14}$$

$$Z_i(\{\widehat{\eta}_{a\to i}\}_{a\in\partial i}) = \sum_{\sigma_i} \prod_{a\in\partial i} \widehat{\eta}_{a\to i}(\sigma_i) \tag{4.15}$$

$$Z_{(i,a)}(\eta_{i\to a}, \widehat{\eta}_{a\to i}) = \sum_{\sigma_i} \eta_{i\to a}(\sigma_i)\widehat{\eta}_{a\to i}(\sigma_i) \tag{4.16}$$

this expression is called the Bethe free entropy density. It has been introduced by H. A. Bethe in 1935 in [72] for the ferromagnetic Ising model. Note that one can alternatively consider the Bethe free entropy density as a functional on the set of messages $\Phi^{Bethe}[\{\eta_{i\to a}, \widehat{\eta}_{a\to i}\}_{(i,a)\in E}]$. It can be shown (see [73]) that the stationary points of the Bethe free entropy are in one-to-one correspondence with the solutions of the BP equation.

From the expression of the free entropy, one can derive the expression of the entropy in terms of the set of messages. Define the Shannon entropy associated to the measure μ:

$$S[\mu] = -\sum_{\underline{x}} \mu(\underline{x}) \ln \mu(\underline{x}) \tag{4.17}$$

The following general relation holds between $S[\mu]$ and the free entropy density Φ:

$$S[\mu] = N\Phi + U \tag{4.18}$$

where U is the internal energy:

$$U = -\sum_{\underline{x}} \mu(\underline{x}) \sum_{a\in A} \ln \omega_a(\underline{x}_{\partial a}) \tag{4.19}$$

one can re-write this expression in terms of the marginals $\mu_a(\underline{x}_{\partial a}) = \mathbb{P}_\mu[\underline{x}_{\partial a}]$

$$U = -\sum_{a\in C} \sum_{a\in A} \mu_a(\underline{x}_{\partial a}) \ln \omega_a(\underline{x}_{\partial a}) \tag{4.20}$$

and use the expression (4.8) of the marginal μ_a in terms of BP messages to finally obtain an expression of $S[\mu]$ in terms of the messages:

$$S[\mu] = N\Phi^{Bethe} - \sum_{a\in C} \frac{\sum_{\underline{x}_{\partial a}} \omega_a(\underline{x}_{\partial a}) \ln(\omega_a(\underline{x}_{\partial a})) \prod_{i\in\partial a} \eta_{i\to a}(x_i)}{\sum_{\underline{x}_{\partial a}} \omega_a(\underline{x}_{\partial a}) \prod_{i\in\partial a} \eta_{i\to a}(x_i)} \tag{4.21}$$

4.1.5 Hard-fields

One says that a BP message $\eta_{i\to a}$ is a hard field when it forces the variable x_i to take a particular value $z \in \chi$: $\eta_{i\to a}(x_i) = \eta_z^{\text{hard}}(x_i)$, with $\eta_z^{\text{hard}}(x_i) = \mathbb{I}[x_i = z]$. A similar definition holds for the message $\widehat{\eta}_{a\to i}$: $\widehat{\eta}_{a\to i}(x_i) = \widehat{\eta}_z^{\text{hard}}(x_i) = \mathbb{I}[x_i = z]$. When the measure μ is supported on the set of solutions of a CSP instance

F (assuming that F is satisfiable), and for tree factor graphs, the presence of hard fields in the set of BP messages solving the BP equations (4.3) indicates that some variables should take one particular value in order to construct a SAT assignment (i.e. in all the solutions to F, these variables take the same value). Indeed from the expression of the marginal μ_i (equation (4.6)), we see that whenever there is a hard fields $\widehat{\eta}_{a\to i}$: $\widehat{\eta}_{a\to i} = \eta_z^{\text{hard}}$, then we get $\mu_i(x_i) = \mathbb{I}[x_i = z]$.

4.1.6 Warning Propagation

One defines the Warning Propagation (WP) messages to be the projection of the BP messages on their hard part:

$$\nu_{i\to a}(x_i) = \mathbb{I}[\eta_{i\to a}(x_i) > 0]; \quad \widehat{\nu}_{a\to i}(x_i) = \mathbb{I}[\widehat{\eta}_{a\to i}(x_i) > 0] \tag{4.22}$$

these messages have the following interpretation. When $\nu_{i\to a}(z) = 1$, then according to the set of constraints in $\partial i \setminus a$, the variable x_i is allowed to take the value z. When $\nu_{i\to a}(z) = 0$ instead, the set of constraints in $\partial i \setminus a$, forbid the variable x_i to take the value z. Similarly, when $\widehat{\nu}_{a\to i}(z) = 1$, then the constraint a allows the variable x_i to take the value z, and forbid it when $\widehat{\nu}_{a\to i}(z) = 0$. The Warning Propagation equations are the projection of the BP equations (4.3) on the WP messages. We will use the shorthand notation:

$$\nu_{i\to a} = f^{\text{WP}}(\{\widehat{\nu}_{b\to i}\}_{b\in\partial i\setminus a}); \quad \widehat{\nu}_{a\to i} = g^{\text{WP}}(\{\nu_{j\to a}\}_{j\in\partial a\setminus i}) \tag{4.23}$$

The first BP equation (4.1) has the same form for all CSP. Projected on the WP messages this gives:

$$\nu_{i\to a}(\sigma_i) = \begin{cases} 0 \text{ if } \exists a \in \partial i \setminus a : \ \widehat{\nu}_{a\to i}(\sigma_i) = 0 \\ 1 \text{ otherwise} \end{cases} \tag{4.24}$$

The second BP equation depends on the choice of the function nodes $\{\omega_a\}_{a\in C}$. Take for instance the k-SAT clauses $\omega_a(\underline{x}_{\partial a}) = c_a(\underline{x}_{\partial a}) = 1 - \prod_{i\in\partial a} \mathbb{I}[\sigma_i = -J_i^a]$. Then the second WP equation states:

$$\widehat{\nu}_{a\to i}(-J_i^a) = \begin{cases} 0 \text{ if } \forall j \in \partial a \setminus i : \ \widehat{\nu}_{a\to j}(J_j^a) = 0 \\ 1 \text{ otherwise} \end{cases} \tag{4.25}$$

$$\widehat{\nu}_{a\to i}(J_i^a) = 1 \tag{4.26}$$

Another example is the bicoloring on k-hypergraphs (NAE-SAT) for which $\omega_a(\underline{\sigma}_{\partial a}) = \mathbb{I}[\sigma_{i(1)}, \ldots, \sigma_{i(|\partial a|)} \text{ n.a.e}]$, where 'n.a.e' stands for 'not-all-equal'. In that case the second WP equation states:

$$\widehat{\nu}_{a\to i}(\sigma_i) = \begin{cases} 0 \text{ if } \forall i \in \partial a \setminus i : \ \widehat{\nu}_{a\to i}(-\sigma_i) = 0 \\ 1 \text{ otherwise} \end{cases} \tag{4.27}$$

In order to solve the WP equations, one can run an algorithm similar to the algorithm 7, replacing the BP iterations (4.4) by the WP iterations obtained by introducing an uptdate scheme on the WP equations (4.23):

$$\nu_{i\to a}^{(t+1)} = f^{\mathrm{WP}}(\{\widehat{\nu}_{b\to i}^{(t)}\}_{b\in\partial i\setminus a}); \quad \widehat{\nu}_{a\to i}^{(t)} = g^{\mathrm{WP}}(\{\nu_{j\to a}^{(t)}\}_{j\in\partial a\setminus i}) \tag{4.28}$$

When the WP messages are initialized with the permissive value $\nu_{i\to a}(x_i) = 1$ on all the edges, these iterations are guaranteed to converge. On a tree, the set of WP messages obtained is the unique solution to the WP equations. From the WP solution, one can construct the equivalent of the marginal μ_i, but for WP messages:

$$\nu_i(x_i) = \prod_{a\in\partial i} \widehat{\nu}_{a\to i}(x_i) \tag{4.29}$$

One calls ν_i a local field. When $\nu_i(z) = 1$, this means that the value $x_i = z$ is allowed by the set of messages $\{\widehat{\eta}_{a\to i}\}_{a\in\partial i}$, while it is forbidden when $\nu_i(z) = 0$. For a measure whose support is the set of solution of a given CSP instance F, the set of local fields provides useful information about the satisfiability of this instance. One says that a contradiction occurs for the variable x_i when $\nu_i(x) = 0$ for all $x \in \chi$. If such a contradiction occurs, it means that for any value that the variable x_i can take, there will be clauses that will be violated. Therefore the instance F has no solutions. If no contradictions occur, then the set of variables for which there is a unique value $x_i \in \chi$ such that $\nu_i(x_i) = 1$ corresponds to the set of variables that are forced to take this specific value in all the solutions of the instance F (if there is solutions). For the SAT problem, it is shown in [74] that the WP equations correspond to the UCP rule. In UCP, a contradiction corresponds to the appearance of a zero-length clause, indicating that two unit-clauses are contradicting each other. A contradiction in UCP occurs if and only if there is a contradiction in the WP messages. If no contradiction has occured, then the partial assignment found by the UCP algorithm corresponds to the set of variables with $\nu_i(x_i) = 1$ for some $x_i \in \chi$.

Note that if no contradiction has occured, it does not mean that the instance F admits a solution (otherwise UCP/WP would solve the decision problem on $k \geq 3$-SAT and since it works in polynomial time one would have incidentally P=NP). Moreover, when F admits solutions, it can happen that a variable that is forced to take a given value in all solutions is not unveiled by the WP iterations. This may cause the failure of the message-passing algorithms that we will present below.

In the case where the factor graph is a tree, however, it can be shown that if no contradiction has occured, then the problem admits solutions (see [35] Theorem 1 for the SAT problem). In that case, the information contained in the local fields can be used to construct a solution.

4.1.7 Survey Propagation

When a factor graph contains loops, the BP equation may admit more than one solution. In the RSB framework described in chapter 2, above the clustering

threshold, for large typical instances, the measure μ is assumed to split into an exponential number of states $\{\mu_c\}$, each of these states being the restriction of the measure μ to one cluster c:

$$\mu(\underline{x}) = \sum_c p(c)\mu_c(\underline{x}) \tag{4.30}$$

where $p(c)$ is the distribution over the set of clusters. In the 1RSB cavity method, one assumes that each state μ_c is associated with one solution of the BP equation, to be denoted $\{\eta^c_{i\to a}, \widehat{\eta}^c_{a\to i}\}_{(i,a)\in E}$. The 1RSB cavity method then introduces a set of probability distributions $\{P_{i\to a}, \widehat{P}_{a\to i}\}_{(i,a)\in E}$. The probability distribution $P_{i\to a}$ is the probability to observe the BP message $\eta^c_{i\to a}$ when the cluster c is chosen randomly according to the distribution $p(c)$. A similar definition holds for $\widehat{P}_{a\to i}$. These probability distributions satisfy a set of equations, called the 1RSB cavity equations, that we denote using the shorthand:

$$P_{i\to a}(\eta_{i\to a}) = F(\{\widehat{P}_{b\to i}\}_{b\in\partial i\setminus a}), \quad \widehat{P}_{a\to i}(\widehat{\eta}_{a\to i}) = G(\{P_{j\to a}\}_{j\in\partial a\setminus i}). \tag{4.31}$$

We will postpone the derivation of these equations to chapter 7 dedicated to the cavity method (see section 7.6). Note that the $\{P_{i\to a}, \widehat{P}_{a\to i}\}_{(i,a)\in E}$ are probability distributions over the space of messages. For instance, a message $\eta(\sigma)$ over a spin variable can be parametrized by its mean value h: $\eta(\sigma) = (1 + h\sigma)/2$. Therefore the space of this message is already $[0, 1]$. It is more tractable to work with the projection of the distributions $\{P_{i\to a}, \widehat{P}_{a\to i}\}_{(i,a)\in E}$ on their contribution to hard fields. One uses the following decomposition:

$$P_{i\to a}(\eta_{i\to a}) = \sum_{x\in\chi} h_{i\to a,x}\delta[\eta_{i\to a} - \eta^{\rm hard}_x] + Q_{i\to a}(\eta_{i\to a}) \tag{4.32}$$

$$\widehat{P}_{a\to i}(\widehat{\eta}_{a\to i}) = \sum_{x\in\chi} \widehat{h}_{a\to i,x}\delta[\widehat{\eta}_{a\to i} - \widehat{\eta}^{\rm hard}_x] + \widehat{Q}_{a\to i}(\widehat{\eta}_{a\to i}) \tag{4.33}$$

When the Parisi parameter $\mathcal{X}$ is equal to the special value 0 or 1, one can obtain closed equations on the weight of hard fields $\{h_{i\to a,x}, \widehat{h}_{a\to i,x}\}_{(i,a)\in E, x\in\chi}$. For the special case $\mathcal{X} = 0$ the equations obtained are called (SP) Survey Propagation equations:

$$h_{i\to a,x} = F^{\rm SP}(\{\widehat{h}_{b\to i,x'}\}_{b\in\partial i\setminus a,x'\in\chi}),\ \widehat{h}_{a\to i,x} = G^{\rm SP}(\{h_{j\to a,x'}\}_{j\in\partial a\setminus i,x'\in\chi}) \tag{4.34}$$

We recall that setting the Parisi parameter $\mathcal{X} = 0$ corresponds to putting equal probability on all clusters, regardless of their size, while $\mathcal{X} = 1$ weights them proportional to their size, i.e., corresponding to uniformly random solutions. The SP message $h_{i\to a,x}$ can be interpreted as the probability (over the choice of cluster) that the set of clauses in $\partial i \setminus a$ send a hard field $\eta_{i\to a} = \eta^{\rm hard}_x$ forcing x_i to take the value x. Similarly, the SP message $\widehat{h}_{a\to i}$ can be interpreted as the probability that the clause a send a hard field $\widehat{\eta}_{a\to i} = \widehat{\eta}^{\rm hard}_x$.

Note that on tree problems, since there is only one solution to the BP equations, the probability distributions $\{P_{i\to a}, \widehat{P}_{a\to i}\}_{(i,a)\in E}$ are Dirac delta's on the

unique solution to the BP equations. The SP messages are either equal to 0 or to 1, and thus reduce to the WP messages, with the following correspondence:

- If there is a $x \in \chi$ such that $h_{i\to a,x} = 1$ then it translates to $\nu_{i\to a}(x) = 1$ and $\nu_{i\to a}(x') = 0\ \forall x' \neq x$
- if for all values $x \in \chi$ one has $h_{i\to a,x} = 0$, then $\nu_{i\to a}(x) = 1$ for all x.

From the SP messages, one can construct a local field $\nu_i^{\mathrm{SP}}(z)$, that can be interpreted as the probability that the variable i receives a warning that forces it to take the value $x_i = z$. Let $P_i(\mu_i)$ be the probability distribution of the marginal μ_i in the 1RSB framework. One can express P_i as a function of the $\{\widehat{P}_{a\to i}\}_{a\in\partial i}$:

$$P_i(\mu_i) = \int \prod_{a\in\partial i} \mathrm{d}\widehat{P}_{a\to i}(\widehat{\eta}_{a\to i})\delta[\mu_i - f^{\mathrm{marg}}(\{\widehat{\eta}_{a\to i}\})] \tag{4.35}$$

where $\mu_i = f^{\mathrm{marg}}(\{\widehat{\eta}_{a\to i}\})$ is a shorthand for the equation (4.6). We then project on the decomposition

$$P_i(\mu_i) = \sum_{x\in\chi} \nu_i^{\mathrm{SP}}(x)\delta[\mu_i - \mathbb{I}[\bullet - x]] + Q_i(\mu_i) \tag{4.36}$$

to get an expression for ν_i^{SP}:

$$1 - \nu_i^{\mathrm{SP}}(z) = \prod_{a\in\partial i} (1 - \widehat{h}_{a\to i,z}) \tag{4.37}$$

4.2 Algorithms

4.2.1 Sampling procedure

The resolution of the BP equations on a tree allows to compute the marginals and conditional probability laws of μ. From this knowledge, one can sample from μ using a sequential assignment procedure. This procedure assign a value to each variable recursively. Suppose that at time t, one has assigned the values of a subset D of variables according to their marginal law $\mathbb{P}_\mu[\underline{x}_D]$, leading to a partial assignment $\underline{x}_D = \underline{z}_D$. One defines the conditional probability $\mathbb{P}_\mu[\underline{x}_{V\setminus D}|\underline{z}_D]$ on the subset of variables that are still not assigned. Then one chooses a new variable $i \in V \setminus D$, computes its marginal from the conditional measure, and draws its value from this marginal. If at each step of the algorithm the marginal thus computed is exact, then this procedure leads to an exact sampling from μ. However, computing marginals of a measure μ over a set in N dimensions is computationnaly intractable. One can use instead a heuristic to estimate the marginals. The BP-guided decimation is a sampling procedure that uses the BP estimation of the marginals.

4.2.2 BP-guided decimation

The BP-guided decimation is described in the algorithm 8. It takes as input the graphical model that describe the measure μ (the factor graph G, and the set of functions $\{\omega_a\}_{a\in C}$). It outputs either a configuration $\underline{z}$ (that is a solution to the associated instance F), or returns "not-found".

Algorithm 8 BP-guided sampling(G, $\{\omega_a\}_{a\in C}$)

```
initialize D = ∅
for t = 1..., N do
    Run BP for P_μ[x|z_D]
    if BP does not converge then
        return "not found" and exit
    else if there is a contradiction then
        return "not found" and exit
    else
        choose i ∈ V \ D
        compute the BP marginal P_μ(x_i|z_D)
        draw x_i from P_μ(x_i|z_D)
        Set x_i = z_i and D → D ∪ {i}
        Add a factor I[x_i = z_i] to the graphical model
    end if
end for
return z_D
```

When the CSP instance F does not admit a solution, the measure whose support is the set of solution is ill-defined, but one can still define the BP messages, and run the BP-decimation algorithm 8. The algorithm will either find a contradiction, concluding that the instance is not satisfiable, or the BP iterations will not converge. As already mentioned, it can happen that a variable x_i which is forced to take one particular value (either in the initial instance F, or in the simplified formula obtained after t steps) is not detected by the BP iterations. In that case the marginal μ_i can have non-zero weights on other values, leading to a wrong assignment of the variable x_i. The resulting simplified instance therefore does not admit solutions, and the algorithm fails. It is not specified in the algorithm 8 how the new variable to be assigned is chosen. One possibility is to take the variable which is the most biased toward one value, i.e. the variable i which has the largest $\text{argmax}_z\{\mu_i(z)\}$.

From this algorithm, it is possible to construct a similar version in which the BP messages are replaced by the WP messages. At each step, the WP iterations are run in order to determine the local fields of the not-yet assigned variables. Each variable that is forced to take one value is assigned to this value. The formula is simplified, and if there is no contradiction, the WP iterations are run for the next step. If there is no local field forcing a variable, then one picks at random a variable and fix it to an arbitrary value. This algorithm is equivalent to the UCP algorithm presented in chapter 1, with an extra free step (or to

the DPLL without backtracking). Note that since in the WP messages one has discarded some information about the tendency of a variable to take a particular value, the BP-guided decimation should be in principle more efficient than the WP-guided decimation, because the BP-guided decimation algorithm might be closer to choosing a uniformly random solution, by giving variables values with the right marginal probability.

4.2.3 SP-guided decimation

The WP and BP-guided decimation algorithms are guaranteed to work on satisfiable instance whose underlying graph is a tree. When the factor graph has loops, they can be used as heuristics, but without a guarantee of convergence. The SP-guided decimation is a similar procedure, that uses the SP messages $\{h_{i\to a,x}, \widehat{h}_{a\to i,x}\}_{(i,a)\in E, x\in\chi}$ defined in (4.32). It is also a heuristic, there is no guarantee of convergence, but it aims at taking into account the presence of loops in the factor graph. The SP-guided decimation is described in the algorithm 9. It has been introduced in 2002 by M. Mézard and R. Zecchina, in [75]. An algorithmic description is given in [35]. It takes as input the graphical model $(G, \{\omega_a\}_{a\in C})$ and a threshold $\epsilon > 0$. The new variable to be assigned is the one with a largest bias π_i toward a given value, with $\pi_i = \max_{x\in\chi}\{\nu_i^{\mathrm{SP}}(x)\}$. The value is assigned according to π_i: $x_i = \mathrm{argmax}_x\{\nu_i^{\mathrm{SP}}\}$. If at some point all the variables have a bias close to 0, it means that there is no variables that is forced to take a particular value. In this case the problem is simpler, and the algorithm calls WalkSAT to solve the instance.

Algorithm 9 SP-guided decimation(F,$\{\omega_a\}_{a\in C}$,ϵ)

```
Set D = ∅
for t = 1, ..., N do
    Run SP
    if SP does not converge then
        return "not found"
    else
        For each i ∈ V \ D, compute the bias π_i = max_{x∈χ}{ν_i^SP(x)}
        Let j ∈ V \ D be the variable with the largest π_i
        if π_i ≤ ε then
            Call WalkSAT
        else
            Fix x_j according to π_j
            Simplify the formula according to this choice
        end if
    end if
end for
return the assignment found
```

We will see in the next chapter that this algorithm outperforms the BP-

guided decimation and the local search algorithms defined in the previous chapter, on random k-SAT instances, especially in the clustered phase. There is several possibilities to improve this algorithm. One can introduce backtracking: if the algorithm fail at some point, one releases the value of some variables that were fixed. The performances of SP-guided decimation with backtracking have been studied in [20]. Moreover, the 1RSB cavity method provides a way to study the cluster decomposition. In particular, the complexity $\Sigma(\mathcal{X})$ at $\mathcal{X} = 0$ is an estimation of the logarithm of the number of clusters, for typical large instances. One can use this quantity to choose the next variable to be assigned. If a choice lead to a huge decrease in $\Sigma(\mathcal{X} = 0)$, this would mean that a lot of clusters have been killed by this choice because they are not compatible with it. One can for instance choose the variable that maximizes the number of clusters still present at the next step.

Chapter 5

Performance of the algorithms on random Constraint Satisfaction Problems

In this chapter we compare the performances of the algorithms that search for solutions on the random CSP ensembles. We will present some of the numerical experiments and theoretical results that have been obtained in the literature. As we have seen, these algorithms can be divided in two broad families. In the first one, the algorithms perform a biased random walk in the space of configurations towards the solutions. In the second one, the algorithms assign sequentially the value of the variables: at a given step, only a subset of variables has been assigned, the others are free. The algorithm then chooses a new variable and its value according to some heuristic. These heuristics can be based on simple properties of the not-yet assigned variables, for instance the number of clauses that will be satisfied by this choice. Or they can use the information provided by message passing algorithms. In the first chapter, we have seen that the performances of DPLL on the random k-SAT ensemble depends on the density of constraints α. Close to the satisfiability transition $\alpha_{sat}(k)$, the typical instances are hard to solve, and the running time of DPLL grows exponentially in N. We would like to study the performances of the other algorithms, and determine the range of α for wich their running time grows polynomially in N. In particular, it would be interesting to characterize the putative algorithmic barrier $\alpha_{alg}(k) \leq \alpha_{sat}(k)$ above which no algorithm is able to find a solution in polynomial time, for a typical random instance (assuming P$\neq$NP). If it is possible to design an algorithm able to run polynomially in the whole satisfiable phase, then $\alpha_{alg}(k) = \alpha_{sat}(k)$, but it is also possible that some intrinsic properties of the set of solutions implies a strict inequality. Up to now, the precise location of $\alpha_{alg}(k)$ is still

unknown, especially for large values of k. For small values of the connectivity k, numerical experiments can be performed in reasonable time. They suggest that some local search algorithms and message-passing inspired algorithms can work up to densities very close to the satisfiability threshold, thus setting lower bounds on $\alpha_{alg}(k)$ almost coinciding with the upper bound $\alpha_{sat}(k)$. When k increases, the numerical experiments become heavier, but simple enough algorithms can be studied analytically for all k. In this regime there is a wide range of α for which instances admit solutions w.h.p., yet no known algorithm is able to find one in polynomial time. The cavity method applied to random k-SAT provides a very detailed description of the structure of the set of solutions. It is an interesting open problem to determine whether some of the phase transitions undergone by the set of solutions affect the perfomances of the algorithms, and therefore could be related to the algorithmic barrier $\alpha_{alg}(k)$.

5.1 Small connectivity k

5.1.1 Two numerical experiments on local search algorithms at small k

In [76], B. Selman, H. Levesque and D. Mitchell compare the performances of the complete algorithm DPLL (with the unit clause branching rule) with a variation of the WalkSAT algorithm, called GSAT, in which there is only greedy steps (setting $p = 0$ in the WalkSAT algorithm). They run these two algorithms on random formulas drawn from the random 3-SAT ensemble in the hard regime $\alpha \simeq 4.3$, with increasing number of variables N. They show that the GSAT algorithm is faster and therefore able to run on larger instances than the DPLL algorithm with the unit clause branching rule. This study provides evidence that the local search algorithm GSAT performs better than the complete algorithm DPLL on random 3-SAT in the hard regime. In [17], B. Selman, H.A. Kautz and B. Cohen compare the performances of four local search algorithms: simulated annealing, GSAT, WalkSAT with a non-zero probability p of random steps, and an algorithm called random noise. This algorithm is a variation of the WalkSAT algorithm, in which the random step is not focused: the variable to be flipped is chosen uniformly at random among the N variables. While GSAT performs only greedy steps, the three other algorithms can escape from local minima by making uphill moves. They show that GSAT performs less well than algorithms with escape strategies, and that WalkSAT outperforms the other algorithms.

5.1.2 Overview of the algorithmic performances for small values of k

In order to compare the performances of the algorithms, it is convenient to introduce the algorithmic threshold for each algorithm $\alpha_a^{ALG}(k)$ (using the notations of [20]) above which the probability of finding a solution in polynomial time vanishes. More precisely, the performance of an algorithm depends on the

density of constraints α, on the size N of the problem, and on the cutoff T on the time we let it run. Let $p_{\rm succ}^{ALG}(\alpha, T, N)$ be the probability (with respect to the random instance generation and the stochasticity of the algorithm) that the algorithm ALG has found a solution after T steps on a random instance of size N with density α. Following the definition given in [31], section V.A, the algorithmic threshold $\alpha_a^{ALG}(k)$ is the smallest α such that:

$$\lim_{N\to\infty} p_{\rm succ}^{ALG}(\alpha, T = N^c, N) = 0 \tag{5.1}$$

for all $c > 0$. In practice the large N limit is inaccessible, and one has to extrapolate from the results obtained at finite N to estimate the algorithmic threshold. Note however that as long as N is finite, any running time T can be considered polynomial if the parameter c is arbitrary. Moreover as N increases the only interesting regime is the linear regime. One then defines the algorithmic threshold as the smallest value of α such that $\lim_{N\to\infty} p_{\rm succ}^{ALG}(\alpha, T = \kappa N, N)$, with κ a constant that can be sent to infinity after the large N limit.

For the RandomWalkSAT algorithm on the random 3-SAT ensemble, the algorithmic threshold is located at $\simeq 2.7$. This value has been determined both numerically and analytically in [34] by R. Monasson and G. Semerjian and in [77] by W. Barthel, A. K. Hartmann and M. Weigt.

In [67] S. Seitz, M. Alava and P. Orponen show that WalkSAT with an optimized parameter p has a linear regime up to $\alpha \approx 4.20$ on random 3-SAT. They also found that the FMS algorithm (see algorithm 6 presented in chapter 3) works in linear time up to $\alpha \approx 4.20$. We recall that the clustering threshold for the uniform measure is $\alpha_d(3) = 3.86$. Therefore it does not seems that the clustering threshold is relevant to understand the limitations of the local search algorithms for $k = 3$. Moreover, the result obtained for FMS and WalkSAT is very close to the satisfiability threshold $\alpha_{sat}(3) = 4.267$. (Note that a similar observation can be made for the 3-coloring on graph: in [78], D. Achlioptas and C. Moore show that a simple algorithm works up to 4.03, while the clustering threshold is at 4.00.) Above $k = 3$, the set of solution undergoes a qualitative change: the condensation threshold separates from the clustering threshold, $\alpha_d < \alpha_c$. It is not clear however if this affects the behavior of the local search algorithms: in [19], the authors show that $\alpha_a^{FMS}(4) = 9.6$ for FMS, which is above the clustering and the condensation threshold $\alpha_d(4) = 9.38$, $\alpha_c(4) = 9.547$. In [79], S. S. Seitz and P. Orponen study experimentally the focused record-to-record travel (FRRT) algorithm. It is a local search algorithm that flips at each step a variable involved in an unsatisfied clause. The flip is accepted if the increase in the number of unsatisfied clauses is smaller than a threshold integer d. They show that this algorithm works in lineal time up to $\alpha \simeq 4.23$ on the random 3-SAT ensemble.

These results have to be compared with the performances of message-passing algorithms. For the BP-guided decimation (BPD) one gets $\alpha_a^{BPD}(3) \simeq 3.8$, and $\alpha_a^{BPD}(4) = 9.05$ ([37]), while SP-guided decimation (SPD) provides better results : $\alpha_a^{SPD}(3) \simeq 4.252$ ([20],[80]), and $\alpha_a^{SPD}(4) \simeq 9.73$ ([20]). Although very close to the satisfiability threshold, the algorithmic threshold for SPD is still

smaller. As expected, SPD performs better than BPD. This can be explained from the fact that the SP messages take into account the clustering phenomenon. Indeed above the condensation threshold, even if the BP iterations converge to a fixed point, the BP marginals thus obtained are likely to provide a bad estimate of the exact marginals, thus leading to wrong assignments in the decimation. The running time of BP and SP-guided decimation is determined by the number of iterations t_{max} needed for the Message-Passing iterations to converge. Each Message-Passing iteration requires $O(N)$ operations, therefore the convergence is achieved in $O(t_{max}N)$ operations. For the whole decimation algorithm, one obtains a complexity $O(N^2 t_{max})$. One can estimate the order of magnitude of t_{max} as the typical diameter of the graphs: $t_{max} = O(\ln N)$. In practice it is observed ([20],[37],[35]) that the number of iterations t_{max} is growing slowly in N. In [20] R. Marino, G. Parisi and F. Ricci-Tersenghi study a version of the SP guided decimation algorithm including backtracking (BSP) on the random k-SAT ensemble with $k = 3$ and $k = 4$. They obtain the best algorithmic threshold for $k = 3$ and $k = 4$:$\alpha_a^{BSP}(3) = 4.26$, and $\alpha_a^{BSP}(4) = 9.9$ which is very close to the satisfiability threshold.

All these results can be compared with the performances of the algorithms using simpler heuristics. For some of them, the rate of convergence can be proven rigorously. The Generalized Unit Clause (GUC) algorithm has been proven to work up to $\alpha_a^{GUC}(3) = 3.003$ on random 3-SAT, and to $\alpha_a^{GUC}(4) = 5.54$ on random 4-SAT (see [48]). GUC is a sequential algorithm: at each step, the algorithm classifies the clauses of the simplified formula according to their length. Let $1 \leq q \leq k$ be the size of the smallest clauses. The algorithm chooses uniformly at random a clause of length q, then chooses uniformly at random a literal in this clause and assign the value of the underlying variable so that the clause is satisfied. The behavior of this algorithm can be analyzed for any k, and in particular on can study its performances in the large k limit (see part 5.2.2). In [81] and [82], the authors independently proved that this bound can be rigorously improved for random 3-SAT, using an algorithm that sets variables with high degree and a large imbalance between positive and negative appearances. They prove that this algorithm succeeds up to densities of constraints $\alpha \leq 3.52$. From these comparisons one can conclude that using the more refined heuristics presented above helps to increase the algorithmic performances. It is also interesting to compare these results with the performances of DPLL, i.e. when one allows backtracking. As already mentioned in the section 1.4.3, the algorithmic threshold for DPLL is estimated at 3.003 for the random 3-SAT ensemble (see [49]), thus it seems that allowing backtracking does not improve on the algorithmic threshold in this case.

5.1.3 Frozen variables and whitening dynamics

It is interesting to study the properties of the solutions found by the local search algorithms. In particular, one would like to know whether these solutions contain frozen variables. Recall that frozen variables are variables that take the same value in all the solutions of a cluster. One can use the correspondence

between clusters and solutions of the BP equations, and say that a variable is frozen in the cluster c if in the corresponding BP fixed point the variable receives a hard message forcing it to a particular value. This information is contained in the solutions of the WP equations, which are the projection of the BP equations on hard messages. Therefore to determine the set of frozen variables in a solution $\underline{\sigma}$, one can run the WP iterations, starting from the initial condition

$$\nu_{i\to a}^{(t=0)}(\tau_i) = \sigma_i \tag{5.2}$$

It can be shown that this dynamics is monotonic in time, and never produces contradictions, so that the WP iterations converge toward a unique fixed point in a finite number of iterations. A variable i is then declared frozen if it receives a hard message $\widehat{\nu}_{a\to i}$ from at least one clause a that forces it to take one particular value.

For the k-SAT problem and for the NAE-k-SAT problem, the WP iterations with $\underline{\sigma}$ as initial condition is equivalent to the whitening dynamics. Starting from the initial condition $\underline{\sigma}^{(t=0)} = \underline{\sigma}$ (with $\underline{\sigma}$ a solution of the instance), the variables are iteratively whitened (set to $\sigma_i = 0$), if they are not constrained. A variable i such that $\sigma_i^{(t)} \neq 0$ is declared white at time $t+1$ if the partial configuration obtained by flipping it and keeping the other non-white variables j (such that $\sigma_j^{(t)} \neq 0$) unchanged can be extended to a solution by an assignment of the white variables. In other words, one has to check if the variable belongs only to clauses which either involve a white variable, or are satisfied by another variable. The procedures stops when either all variables are white, or a fixed point is reached. The variables that are non-white at the end of the run are then declared frozen.

In [19], the authors checked that the solutions found by the FMS algorithm were completely white, i.e. there were no frozen variables in the solutions they found. It is also the case for the solutions of the BSP algorithm [20], and more generally for both local search algorithm and message-passing guided decimation algorithms. This results provide the conjecture that the algorithmic barrier $\alpha_{alg}(k)$ is upper bounded by the freezing threshold $\alpha_f(k)$, above which all solution contain a finite fraction of frozen variables. The rigidity threshold is defined as the value of α above which the typical solutions (with respect to the measure μ defined to describe the set of solutions) contains a finite fraction of frozen variables. In the region $\alpha_r \leq \alpha < \alpha_f$ the algorithms find untypical unfrozen solutions ([20, 83, 84, 85]).

5.1.4 Analytical study of the BP-guided decimation

In [36] and [37], a theoretical framework is developed to study the running time of the BP-guided decimation algorithm. Using the cavity method, the authors analyse the behavior of an idealized algorithm where marginal are computed exactly. To do so, they introduce an ensemble of CSP parametrized by α, θ, that allows to study the statistics of the simplified formulas obtained after $t = \theta N$ decimation steps. An instance of this ensemble is generated as follows:

1. draw a satisfiable instance with parameter $\alpha < \alpha_{sat}(k)$
2. draw a uniform solution $\underline{\tau}$ of this instance
3. construct a set D by including in D each variable independently with probability θ
4. consider the residual formula on the variables in $V \setminus D$, obtained from the initial formula by imposing $\sigma_i = \tau_i$ for $i \in D$.

The initial ensemble is recovered with $\theta = 0$. Note that the formula thus obtained cannot be seen as a formula drawn from the initial random ensemble, since there are non-trivial correlations between variables in the measure μ_D induced by the choice of $\underline{\tau}$. In order to study the statistical properties of the formulas obtained from the ensemble defined above, they compute the quenched averaged residual entropy:

$$\omega(\theta) = \lim_{N\to\infty} \mathbb{E}_F \mathbb{E}_{\underline{\tau}} \mathbb{E}_D[\ln Z(\underline{\tau}_D)] \tag{5.3}$$

with

$$Z(\underline{\tau}_D) = \sum_{\underline{\sigma}} \prod_{a=1} \mathbb{I}[\underline{\sigma}_{\partial a} \text{ satisfies the clause } a] \mathbb{I}[\underline{\sigma}_D = \underline{\tau}_D] \tag{5.4}$$

being the number of solutions compatible with the partial assignment $\underline{\tau}_D$. They also introduce the average fraction of variables which have been explicitly assigned or which can be logically deduced from these assignments

$$\phi(\theta) = \theta + \lim_{N\to\infty} \mathbb{E}_F \mathbb{E}_{\underline{\tau}} \mathbb{E}_D[\text{nb. of directly implied variables}] \tag{5.5}$$

Note that $(\mathrm{d}\phi/\mathrm{d}\theta) - 1$ is the average number of newly implied variables appearing during one step. If this number diverges (signaled by a discontinuity in $\phi(\theta)$), it might be the sign of potential errors that the decimation procedure will make if the marginals are not computed exactly. Both $\omega(\theta)$ and $\phi(\theta)$ are computed with the RS cavity method. They provide an estimation of the free entropy and fraction of variables at step $t = \theta N$, in an idealized algorithm in which the marginals are computed exactly. For $k = 4$, they obtain the following results. The quenched residual free entropy is a decreasing function of θ. Below a threshold value $\alpha_* = 9.05$ the function is smoothly decreasing, while above this value the function $\omega(\theta)$ undergoes a discontinuity in its derivative at $\theta_c(\alpha)$, corresponding to a condensation transition. Above $\theta_c(\alpha)$ the total free entropy $\omega(\theta)$ is dominated by a sub-exponential number of clusters, therefore it might be harder to find solutions. For $\phi(\theta)$, below the threshold value $\alpha'_* = 8.05$, $\phi(\theta)$ is a smoothly increasing function of θ. Above α'_*, $\phi(\theta)$ is discontinuous at some value of θ. This discontinuity is the sign of an avalanche of directly implied variables during the decimation procedure. These estimations are compared with numerical experiments. They compute the empirical probability of success for BPD, and see that it vanishes when α is close to $\alpha_* = 9.05$ ($k = 4$). The performances of the algorithm seems therefore to be related to the phase transition undergone by the residual free entropy $\omega(\theta)$. While $\phi(\theta)$ does not seem

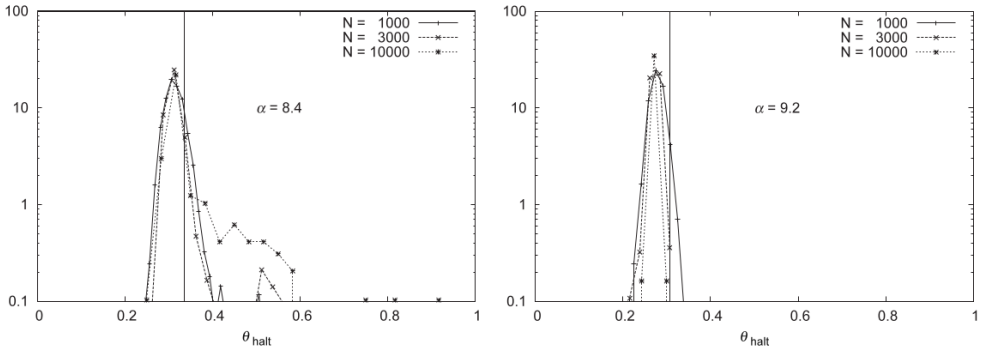

Figure 11. Distribution of the halting time of the BP guided decimation algorithm on 4-sat random formulae with $\alpha = 8.4$ (left panel) and $\alpha = 9.2$ (right panel). Vertical lines show the value of $\theta'_+(\alpha)$ where $\phi(\theta)$ is discontinuous.

Figure 5.1: From [37]

to influence the algorithmic threshold, it has an influence on the running time. Let θ_{halt} be the fraction of variables assigned before the discovery of a contradiction, during unsuccessful runs. In figure 5.1 (see [37] figure 11.) the halting time θ_{halt} is plotted for two values of α above the threshold $\alpha'_* \simeq 8.1$. The maximum of the distribution of θ_{halt} is close (a bit smaller) to the discontinuity point of $\phi(\theta)$. It is quite surprising that the jump of $\phi(\theta)$ does not affect the probability of finding a sat assignment, because one would have thought that an $O(N)$ avalanche of forced variables produces a contradiction in the messages w.h.p. Instead the BP-guided decimation algorithm is able to fix a finite $O(N)$ fraction of variables without contradiction. The success probability falls when entering the condensation phase $\alpha > \alpha_*$.

In [36] A. Montanari, F. Ricci-Tersenghi and G. Semerjian provide the asymptotic expansion of the value $\alpha'_*(k)$ at which ϕ starts to be discontinuous:

$$\alpha_{spin}(k) = e\frac{2^k}{k}(1 + O(k^{-1})) \tag{5.6}$$

This provides an estimated lower bound on the algorithmic threshold for the BP-guided decimation algorithm. In [86], A. Coja-Oghlan proves that BP-guided decimation algorithm fails to solve random k-SAT formulas already for $\alpha = O(2^k/k)$.

5.2 Large connectivity k

5.2.1 Asymptotic expansion of the thresholds

In the large k limit the satisfiability threshold for the random k-SAT ensemble scales as:

$$\alpha_{sat}(k) \sim 2^k \ln 2. \tag{5.7}$$

This result has been obtained rigorously in [87] by D. Achlioptas, A. Naor and Y. Peres. Using the first and second moment method, they obtain that the satisfiability threshold is located in the interval $(2^k \ln 2 - k, 2^k \ln 2)$. In [7], S. Mertens, M. Mézard and R. Zecchina provide an asymptotic expansion in powers of 2^{-k} of the satisfiability threshold in the large k limit, using the cavity metod. In [12] this prediction for the asymptotic expansion is proven to be exact for large but finite k. In [88], D. Achlioptas and C. Moore prove the scaling of the satisfiability threshold for the bicoloring on k-hypergraphs:

$$\alpha_{sat}(k) \sim 2^{k-1} \ln 2 \ . \tag{5.8}$$

For the q-coloring problem on graphs one has (see [89]):

$$c_{sat}(q) \sim 2q \ln q \tag{5.9}$$

where $c = 2\alpha$ is the average degree. The asymptotic expansions of the rigidity and clustering thresholds have also been established for several CSPs, for the uniform measure. It is shown in [10] that the clustering threshold in the large k, q limit is upper-bounded by:

$$\alpha_d(k) \leq 2^k (\ln k)/k \tag{5.10}$$

for the k-SAT problem, by:

$$\alpha_d(k) \leq 2^{k-1} (\ln k)/k \tag{5.11}$$

for the bicoloring on k-hypergraphs, and by:

$$c_d(q) \leq q \ln q \tag{5.12}$$

for the q-coloring on graphs. Moreover, the authors show that the rigidity threshold arise at the same scale. It is possible to derive the following subdominant terms for the rigidity threshold, using the naive reconstruction procedure (see [62] and [38]). This gives for the bicoloring on k-hypergraphs:

$$\alpha_r(k) = \frac{2^{k-1}}{k} (\ln k + \ln \ln k + \gamma_r + o(1)); \ \gamma_r = 1 \tag{5.13}$$

and for the q-coloring on graphs (see also [11]):

$$c_r(q) = q(\ln q + \ln \ln q + \gamma_r + o(1)); \ \gamma_r = 1 \tag{5.14}$$

In general, the analytic determination of the rigidity threshold is simpler than for the clustering threshold, because it corresponds to the bifurcation of a scalar (instead of functional) fixed-point equation, as we will see in chapter devoted to the cavity method. For these two problems, it is conjectured that the clustering threshold arise at the same scale:

$$\alpha_d(k) = \frac{2^{k-1}}{k}(\ln k + \ln\ln k + \gamma_d + o(1));\ c_d(q) = q(\ln q + \ln\ln q + \gamma_d + o(1)) \quad (5.15)$$

In [39], the leading term is proven rigorously for these two problems. For the coloring problem, the inequality $1 - \ln 2 \leq \gamma_d \leq \gamma_r = 1$ has been established rigorously in [38], and the strict inequality $\gamma_d < 1$ was later obtained in [40]. In [32] we characterize the value of γ_d for the two models in terms of a functional equation, that is solved numerically, yielding to the estimate $\gamma_d \simeq 0.871$ for both models. We also provide an analytic lower bound $\gamma_d \geq 1 + \ln(2(\sqrt{2}-1)) \simeq 0.812$.

The freezing transition has been predicted in [26] to arise at $\alpha_f(k) \sim \alpha_{sat}(k)/2$ for the NAE-k-SAT problem.

The condensation threshold is also expected to arise close to the satisfiability threshold. In [59] A. Montanari, F. Ricci-Tersenghi and G. Semerjian obtain the following asymptotic expansion for the random k-SAT problem, using the uniform measure:

$$\alpha_c(k) = 2^k \ln 2 - \frac{3\ln 2}{2} + O(2^{-k}) \quad (5.16)$$

5.2.2 Algorithmic performances in the large k limit

Numerical experiments cannot access directly the large k limit, but simple enough algorithms can be studied analytically for all k; sequential assignment algorithms that use simple heuristics to guide their choices can be described in terms of differential equations. In [90], M-T. Chao and J. Franco study the unit clause algorithm defined in Algorithm 10 and proved that a satisfiable assignment can be found with positive probability when $N \to \infty$ whenever $\alpha < \alpha_*(k)$, where

$$\alpha_*(k) = \frac{1}{2}\left(\frac{k-1}{k-2}\right)^{k-2}\frac{2^k}{k}. \quad (5.17)$$

Algorithm 10 Unit clause(F)

if there is any unit clause (forced step) **then**
  Pick a unit clause uniformly at random and satisfy it.
else (free step)
  Pick a not-yet assigned variable uniformly at random and assign it to 0/1 uniformly at random
end if

Although we will not explain the proof, it is interesting to mention that it relies on the uniform randomness of the simplified formula over the set $V(t)$ of the not-yet assigned variables obtained at step t:

Lemma 4 (uniform randomness [91]) *For every $0 \leq t \leq N$, and $1 \leq s \leq k$, the set $S_s(t)$ of s-clauses form a random s-SAT formula drawn from the random s-SAT ensemble with $|V(t)|$ variables and $C_s(t)$ clauses. Moreover these s-SAT formulas are independent of each other.*

Note that this lemma does not hold for the BP-guided decimation: the decimation procedure has introduced non-trivial correlations between the variables in $V(t)$. The proof then uses the fact that the formula composed of the 2-clauses at each step t is distributed from the random 2-SAT ensemble. Looking at its density, whenever it crosses the value $\alpha_{sat}(2) = 1$, then the algorithm will fail w.h.p., while it can be proven that if the density of 2-clauses is bounded away from 1, then the algorithm succeeds with positive probability. The evolution of the 2-clause density conditioned on the previous steps can be studied by means of differential equations on the mean value of the conditional s-clause densities.

This method has also been applied in [51] to improve this result. They study the Shortest Clause algorithm. This algorithm is a variation of the Unit Clause algorithm 10. Instead of looking at unit-clause only, this algorithm also look at clauses of length 2 (2-clauses). At each step, if there are unit-clauses, then the algorithm performs a forced step. If there is no unit-clauses but there are 2-clauses, then the algorithm picks one of them at random and satisfy at random a literal in it. Otherwise it performs a free step. [51] proves that this algorithm finds a solution w.h.p. for α smaller than

$$\frac{1}{8}\left(\frac{k-1}{k-3}\right)^{k-3}\frac{2^k}{k} \tag{5.18}$$

In [48] this bound is improved by studying the Shortest Clause algorithm with limited backtracking. They show that the algorithm find a solution w.h.p. when $\alpha < \alpha_*(k)$, where $\alpha_*(k) \simeq 1.817(2^k/k)$. In the large k limit, all these algorithms are performant for densities smaller than $c2^k/k$, with a constant c depending on the algorithm. In [36], a similar scaling is obtained: the BP-guided decimation algorithm is conjectured to fail above the threshold $e2^k/k$ corresponding to the appearance of cascades of logically implied variables during the decimation process, and in [86], A. Coja-Oghlan proves that BP-guided decimation fails to solve random k-SAT formulas already for $\alpha = O(2^k/k)$.

This scaling is improved in [21], where an algorithm is shown to work up to densities of constraints coinciding at leading order with $\alpha_d(k) \sim 2^k \ln k/k$. Up to now, there is no algorithm that has been proven to find solution in polynomial time above this scaling. This leaves a multiplicative gap of order k (neglecting the sub-dominant logarithmic correction) from the satisfiability transition, hence a wide range of parameters α for which typical instances are known to have solutions, yet no provably efficient algorithm is known at present to find them.

Some negative results have also been obtained. In [23], it is shown that the WalkSAT algorithm is ineffective w.h.p. for densities of constraints larger than $c2^k(\ln k)^2/k$, with $c > 0$ a constant, while it has been shown in [92] that this algorithm finds satisfying assignment in linear time w.h.p. if $\alpha < c'2^k/k$, with

$c' > 0$ another constant. In [22], it is shown that for the NAE-k-SAT problem, a certain class of local algorithms fail to find solutions for $\alpha > 2^{k-1}(\ln k)^2/k$, for large k.

Chapter 6

Biased Measures for random Constraint Satisfaction Problems

In this chapter we introduce biased probability measures over the set of solutions of random instances of CSPs, i.e. probability measures for which not all solutions are equiprobable. This perspective has been used in previous works [26, 30, 27, 28, 29] (in [27, 28] the local entropy, or density of solutions in configuration space, is used to weight differently the solutions, in [26, 30] this role is played by the number of frozen variables, while in [29] hard sphere particles are considered as a CSP, with a bias due to an additional pairwise interaction between particles). Indeed the structural phase transitions (clustering, condensation, rigidity) introduced in chapter 2 depend on the choice of the measure describing the set of solutions. In [26] it has been demonstrated that the threshold for properties that are typical in the uniform ensemble (in particular the existence of frozen variables) can be significantly moved by an appropriate bias. This opens some hope to diminish the algorithmic gap, by giving more weight to solutions that are "easier" to find, for instance because they contain less frozen variables, and to turn atypical properties of the uniform measure into typical ones of the biased measure.

We will present the form of the bias used during this Ph.D., that was studied in [31] and in [33]). We will also present the biases introduced in the papers mentioned above. In this Ph.D., we have focused on the increase of the dynamic threshold α_d that results from a well-chosen bias between solutions. The algorithmic motivation for the study of this threshold comes from the Simulated Annealing [25] procedure, as explained in chapter 3. Below α_d a Markov Chain reversible with respect to a finite-temperature probability distribution should be able to equilibrate in polynomial time, hence to find solutions once the temperature is lowered slowly enough (if there is no reentrance in temperature). Although it is true that many algorithm do not respect detailed balance (focused

algorithm), therefore are not affected by this transition, and that Simulated Annealing algorithm can find solutions even out-of-equilibrium, we observed that the bias introduced in [31] had positive impact on the performances of Simulated Annealing. We have focused on the bicoloring on k-hypergraph, that exhibits the same threshold phenomena as k-SAT, but for which the computations are a bit simpler.

6.1 Definitions

6.1.1 Biased measure over the set of solutions

Let F be an instance of CSP, with $S(F)$ the set of its solutions. Assuming that it is non-empty, one can define the uniform measure over the set of solutions:

$$\mu_u(\underline{\sigma}) = \frac{1}{Z_u} \times \begin{cases} 1 \text{ if } \underline{\sigma} \in S(F) \\ 0 \text{ if } \underline{\sigma} \notin S(F) \end{cases} \tag{6.1}$$

where the subscript u stands for "uniform". The variables σ_i ($i \in \{1, \dots, N\}$) leave in a discrete space χ ($\chi = \{-1, +1\}$ for spin variables). In the most generic form, a biased measure over the set of solutions can be writen:

$$\mu_b(\underline{\sigma}) = \frac{b(\underline{\sigma})}{\sum_{\underline{\sigma}'} b(\underline{\sigma}')} \tag{6.2}$$

where $b : \chi^N \to \mathbb{R}$ is a non-negative function:

$$b(\underline{\sigma}) \begin{cases} > 0 \text{ if } \underline{\sigma} \in S(F) \\ = 0 \text{ if } \underline{\sigma} \notin S(F) \end{cases} \tag{6.3}$$

There are infinitely many possible choices for the function b, we will present the choices that we made during the Ph.D., and the choices made in the papers [26, 30, 27, 28, 29]. Note that we need to impose some locality requirements on μ, so that the Monte Carlo simulations are tractable. Indeed at each step one needs to compute the change in energy when a variable is flipped. For measures defined with graphical models (such as (1.7)), computing the change in energy requires only a small number of operations. In the next chapters, we will apply the cavity method to study the typical properties of the measure describing the solution set $S(F)$, when F is drawn from a random CSP ensemble. This method requires that the measure can be described as a graphical model, and that the underlying graph representing the measure is locally tree-like in the thermodynamic limit. As we have seen, the uniform distribution μ_u can be described by a graphical model whose underlying graph is the initial graph representing the instance F. We recall here its formulation as a graphical model (equation (1.10))

$$\mu_u(\underline{\sigma}) = \frac{1}{Z_u} \prod_{a \in C} (1 - c_a(\underline{\sigma}_{\partial a})) = \frac{1}{Z_u} \prod_{a \in C} \mathbb{I}[\underline{\sigma}_{\partial a} \text{ satisfies the clause } a] \tag{6.4}$$

In the rest of this section we will introduce biased measures that introduce local interactions (with respect to the notion of distance induced by the hypergraph representing F) between variables. For simplicity we will consider only instances defined on k-hypergraphs.

6.1.2 Intra-clause bias

We introduce a biased function b that factorizes on the clauses:

$$b(\underline{\sigma}) = \prod_{a \in C} \omega(\underline{\sigma}_{\partial a}) \tag{6.5}$$

The function $\omega : \chi^k \to \mathbb{R}$ vanishes when the subset of variables does not satisfy the constraint a, and is strictly positive when it does. In that way the set of solutions $S(G)$ is not modified. Note that the underlying graph representing this graphical model coincides with the initial graph representing the instance F.

Application to the bicoloring on k-hypergraphs

We study this form of bias on the bicoloring of k-hypergraphs in [31]. We recall that in that case, the variables are spins $\sigma \in \{+1, -1\}$, and each constraint a forbids the variables $\underline{\sigma}_{\partial a}$ to take the same value:

$$1 - c_a(\underline{\sigma}_{\partial a}) = \mathbb{I}[\underline{\sigma}_{\partial a} \ n.a.e] \tag{6.6}$$

where $n.a.e$ stands for 'not all equal'. The biased measure reads:

$$\mu(\underline{\sigma}) = \frac{1}{Z(G)} \prod_{a \in C} \omega(\underline{\sigma}_{\partial a}) \tag{6.7}$$

We assume that ω is invariant under all permutations of its k elements. As the latter are binary variables, ω depends only on the number p of -1 among its arguments, and we will denote $\omega_p \geq 0$ the value it then assumes. This translates into the formula:

$$\omega(\sigma_1, \dots, \sigma_k) = \omega_p \qquad \text{if} \qquad \sum_{i=1}^{k} \sigma_i = k - 2p \ . \tag{6.8}$$

The uniform measure over the solutions of the bicoloring problem is recovered for the choice $\omega_0 = \omega_k = 0$, $\omega_1 = \cdots = \omega_{k-1} = 1$. If one chooses instead ω_p to depend on p for $p \in \{1, \dots, k-1\}$, while keeping $\omega_0 = \omega_k = 0$, one obtains a probability measure μ that is still supported solely on the proper bicolorings of G, but is not uniform anymore. We will sometimes relax the constraint $\omega_0 = \omega_k = 0$, to model the effect of a positive temperature that allows some constraints to be violated, as it will be used in Simulated Annealing. We will in any case always assume that $\omega_p = \omega_{k-p}$: this ensures that the global spin-flip symmetry $\mu(-\underline{\sigma}) = \mu(\underline{\sigma})$ (which is indeed a property of the set of proper bicolorings) is preserved.

We have worked in particular with the specific choice:

$$\omega_0 = \omega_k = 0, \quad \omega_1 = \omega_{k-1} = 1 - \epsilon, \quad \omega_2 = \cdots = \omega_{k-2} = 1 \tag{6.9}$$

i.e. we weight solutions according the their number of "almost monochromatic" clauses (clauses with only one variable that takes a different value from the others). These "almost monochromatic" clauses are responsible for the existence of frozen variables. We recall that the frozen variables are variables that takes the same values in all the solutions of a cluster. This is therefore one of the mechanism of the RSB phenomenon, although not the only one. Indeed for instance when $k \geq 4$, the rigidity transition α_r is strictly greater than the clustering transition $\alpha_d < \alpha_r$, hence there in the range $\alpha_d \leq \alpha < \alpha_r$ the set of solution is clustered yet typical solutions do not contain a positive fraction of frozen variables.

6.1.3 Bias with interactions at distance 1

We consider the following measure that introduces the shortest non-trivial interactions that allows the coupling of variables from different hyper-edges:

$$\mu(\underline{\sigma}) = \frac{1}{Z} \prod_{a \in C} \omega(\underline{\sigma}_{\partial a}) \prod_{i \in V} \varphi_i(\sigma_i, \{\underline{\sigma}_{\partial a \setminus i}\}_{a \in \partial i}) \tag{6.10}$$

in that case, the function ω is just the indicator function of the event "$\sigma_1, \ldots, \sigma_k$ satisfy the constraint". The biasing function $\varphi_i > 0$ couples the i-th variable with its $|\partial i|(k-1)$ neighbors at distance 1.

Application to the bicoloring on k-hypergraphs

We study this form of bias on the bicoloring problem of k-hypergraph in [33]. For simplicity we restrict ourselves to regular hypergraphs, where every vertex has the same degree $|\partial i| = l + 1$. We used the same function φ on all the vertices. There is still a vast freedom in the choice of φ. We restrict this choice by imposing its invariance under a global spin-flip of its arguments (that enforces the global spin-flip symmetry $\underline{\sigma} \to -\underline{\sigma}$ of the set of solution). We also impose its invariance under the permutations of the $l + 1$ hyperedges around i and of the $k - 1$ neighboring variables inside each of these hyperedges. This amounts to take

$$\varphi(\sigma_i, \{\underline{\sigma}_{\partial a \setminus i}\}_{a \in \partial i}) = \widehat{\varphi}(\{m_{a \to i}\}_{a \in \partial i}) \quad \text{with } m_{a \to i} = \sum_{j \in \partial a \setminus i} \frac{1 + \sigma_i \sigma_j}{2} \tag{6.11}$$

and $\widehat{\varphi}$ invariant under the permutations of its $l + 1$ arguments. $m_{a \to i}$ counts the number of variables in $\partial a \setminus i$ that are of the same color as σ_i. In [33] we discard part of the information contained in $m_{a \to i}$ and only distinguish between the cases $m_{a \to i} = 0$ and $m_{a \to i} > 0$. This is indeed a relevant information about the solution $\underline{\sigma}$: in the former case σ_i is the only variable of its color in the a-th

hyperedge, hence σ_i cannot be flipped without violating the a-th monochromatic constraint. One says that a forces i in such a situation. On the contrary when $m_{a\to i} > 0$ the variable σ_i is not forced to its value by the a-th hyperedge. With this simplification, and because of the invariance by permutation of the arguments of φ, the weight of the variable i becomes a function of the number of constraints forcing it. As $\underline{\sigma}$ is a solution, the event $m_{a\to i} = 0$ is equivalent to "the variables in $\underline{\sigma}_{\partial a\setminus i}$ are all equal (a.e)", and the biased measure becomes thus:

$$\mu(\underline{\sigma}) = \frac{1}{Z}\prod_{a\in C}\omega(\underline{\sigma}_{\partial a})\prod_{i\in V}\psi\left(\sum_{a\in\partial i}\mathbb{I}[\underline{\sigma}_{\partial a\setminus i}\ a.e]\right) \tag{6.12}$$

where $\psi(p)$ is the weight attributed to a variable contained in $p \in \{0,\dots,l+1\}$ forcing hyperedges. With this choice, there are $l+1$ free parameters to describe the function ψ (its argument can take $l+2$ values, but a global multiplicative constant gets absorbed in the normalization Z).

As mentioned, the presence of forcing clauses is one mechanism of the RSB phenomenon, but cannot explain it on its own. In [33] we showed that bias defined in (equation 6.12) was already giving positive results on the clustering threshold, both for small connectivity k and in the large k limit. In the future it would be interesting to study more general form of biases, keeping all the information in $m_{a\to i}$.

In [33] we give a particular attention to the specific form:

$$\psi(0) = 1, \quad \psi(p) = b(1-\epsilon)^p \text{ for } 1 \le p \le l+1 \tag{6.13}$$

which contains the two parameters $b > 0$ and $\epsilon < 1$. This form actually encompasses the two following cases:

- when $\epsilon = 0$, in such a way that:

$$\psi(p) = \begin{cases} 1 \text{ if } p = 0 \\ b \text{ if } p > 0 \end{cases} \tag{6.14}$$

  one recovers a measure studied in [26], that we will present in the next section. Indeed the weight of a solution $\underline{\sigma}$ is then b raised to a power equal to the number of variables that are forced by at least one constraint, i.e. those that are not "whitened" after $T = 1$ step of a coarsening algorithm used in [9, 93, 35, 94, 95, 84, 83], whose large deviations on atypical solutions were studied in [26] for arbitrary values of T (with an unfortunate conflict of notation the parameter called b here is denoted e^ϵ in [26]).

- when $b = 1$ one has

$$\psi(p) = (1-\epsilon)^p \tag{6.15}$$

  the weight of a solution $\underline{\sigma}$ is thus $(1-\epsilon)$ raised to the number of forcing constraints, hence we recover the bias introduced in the previous paragraph (6.1.2), and studied in [31]. Indeed in a solution every constraint

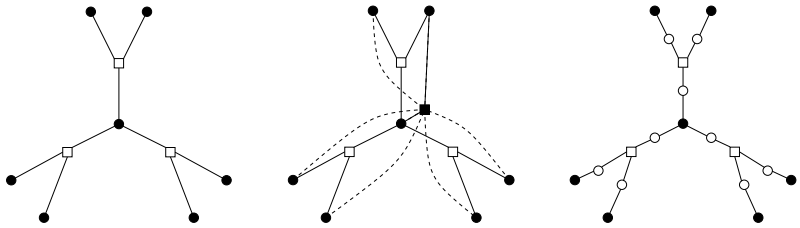

Figure 6.1: Left panel: an example of a hypergraph G with $N = 7$ vertices represented by black circles, and $M = 3$ hyperedges linking $k = 3$ vertices, drawn as white squares. Center panel: the introduction of an interaction, represented as a black square, between all the vertices at distance 1 from the central vertex i, generates short loops even if G is a tree. Right panel: the factor graph representation of the probability measure (6.16), the white circles stand for the variable nodes $v_{(i,a)}$, the black circles (resp. white squares) are the interaction factors $\widetilde{\psi}$ (resp. $\widetilde{\omega}$).

a forces at most one of its variable $i \in \partial a$, when σ_i is the unique representant of its color in $\underline{\sigma}_{\partial a}$, hence there is no double counting of forcing constraints in the product over variables in (6.12).

Obviously the uniform measure (1.10) is recovered when $b = 1$ and $\epsilon = 0$.

Factor graph and auxiliary variables

The factor graph representing the measure (6.12) is not the underlying graph G representing the instance F. Instead the function ψ has introduced short loops, as is it represented in figure 6.1 (middle panel). This prevents from a direct application of the cavity method, and requires a preliminary step in order to get rid of these short loops. To achieve this we introduce some auxiliary, redundant variables in the following way: for each edge $(i, a) \in E$ between a vertex i and one of its incident hyperedge $a \in \partial i$ we introduce two variables, $w_{a\to i} \in \{0, 1\}$ and $\sigma_i^a \in \{-1, +1\}$ which are deterministic functions of the original configuration $\underline{\sigma}$, according to $w_{a\to i} = \mathbb{I}[\underline{\sigma}_{\partial a\setminus i} a.e]$ and $\sigma_i^a = \sigma_i$. We call $v_{(i,a)} = (\sigma_i^a, w_{a\to i})$ the value of these two auxiliary variables, and $\underline{v} = \{v_{(i,a)}\}_{(i,a)\in E}$ their global configuration. Consider now the following probability law for $\underline{v}$:

$$\mu(\underline{v}) = \frac{1}{Z} \prod_{i\in V} \widetilde{\varphi}(\{v_{(i,a)}\}_{a\in\partial i}) \prod_{a\in C} \widetilde{\omega}(\{v_{(i,a)}\}_{i\in\partial a}) \tag{6.16}$$

where

$$\widetilde{\varphi}(\sigma_1, w_1, \ldots, \sigma_{l+1}, w_{l+1}) = \psi\left(\sum_{i=1}^{l+1} w_i\right) \mathbb{I}[\sigma_1 = \cdots = \sigma_{l+1}] \tag{6.17}$$

$$\widetilde{\omega}(\sigma_1, w_1, \ldots, \sigma_k, w_k) = \omega(\sigma_1, \ldots, \sigma_k) \prod_{i=1}^{k} \mathbb{I}[w_i = \mathbb{I}[\{w_i = \mathbb{I}[\{\sigma_j\}_{j\neq i} a.e]\}]] \tag{6.18}$$

One realizes easily that for a configuration $\underline{v}$ in the support of (6.16) σ_i^a is independent of a, and the marginal law of $\underline{\sigma}$ is nothing but (6.16). The partition function Z is the same in the two expressions (6.12) and (6.16), and in the support of $\mu(\underline{v})$ the variables $w_{a\to i}$ are the deterministic functions of $\underline{\sigma}$ defined above. This equivalent rewriting with redundant variables has an important advantage: as shown in the right panel of Fig. 6.1 the graphical model corresponding to (6.16), with variables v on the edges (i, a) of G and interaction nodes both on the original hyperedges ($\widetilde{\omega}$) and on the original vertices ($\widetilde{\varphi}$), respects the topology of G, and is thus (locally) a tree if G is.

6.1.4 Bias with interactions at larger distance

We have presented in the two previous paragraphs (6.1.2 and 6.1.3) the form of the bias studied in this Ph.D.. Although we did not investigate it, it is natural to extend these definitions, introducing interactions between variables at larger distance. We show in this paragraph that for some specific form of the bias, it is possible to apply a similar transformation as the one described above that introduce auxiliary variables, to recover a locally tree-like factor graph. This approach has been studied in [26], with a bias counting the number of frozen variables in a solution.

Let $B(i, r)$ be the set of variables at distance $\leq r$ from i. One would like to study the following probability measure:

$$\mu(\underline{\sigma}) = \frac{1}{Z} \prod_{a\in C} \omega(\underline{\sigma}_{\partial a}) \prod_{i\in V} \varphi_i(\sigma_i, \underline{\sigma}_{B(i,r)}) \tag{6.19}$$

This measure cannot in general be represented by a simple graphical model, but it is possible to apply a similar transformation as the one used in the paragraph 6.1.3, to recover a locally tree-like underlying graph, if we assume that the function φ_i is constructed in the following way. One each edge (i, a), one introduces a set of auxiliary variables $\{g_{i\to a}^{(s)}, \widehat{g}_{a\to i}^{(s)}\}_{s\in\{1,\ldots,r\}}$ that are determined uniquely from the solution $\underline{\sigma}$:

$$g_{i\to a}^{(1)} = \sigma_i \tag{6.20}$$

$$\widehat{g}_{a\to i}^{(s)} = \widehat{G}_s(\{g_{j\to a}^{(s)}\}_{i\in\partial a\setminus i}) \quad \text{where } s \in \{1, \ldots, r\} \tag{6.21}$$

$$g_{j\to a}^{(s+1)} = G_{s+1}(\{\widehat{g}_{b\to j}^{(s)}\}_{b\in\partial a\setminus j}) \quad \text{where } s \in \{1, \ldots, r-1\} \tag{6.22}$$

where the $\{G_s, \widehat{G}_s\}$ are updating functions. One can then choose φ_i to be a function of the auxiliary variables $\{\widehat{g}^{(s)}_{a\to i}\}^{s\in\{1,\dots,r\}}_{a\in\partial i}$ surrounding it:

$$\varphi_i(\sigma_i, \underline{\sigma}_{B(i,r)}) = \phi_i(\{\widehat{g}^{(s)}_{a\to i}\}^{s\in\{1,\dots,r\}}_{a\in\partial i}) \tag{6.23}$$

Introducing the auxiliary variable $\sigma^a_i = \sigma_i$ as before, and defining for each edge (i,a) the auxiliary variable $v_{(i,a)} = (\sigma^a_i, \{g^{(s)}_{i\to a}, \widehat{g}^{(s)}_{a\to i}\}^{s\in\{1,\dots,r\}})$, one can consider the probability law for $\underline{v}$:

$$\mu(\underline{v}) = \frac{1}{Z}\prod_{i\in V}\widetilde{\varphi}_i(\{v_{(i,a)}\}_{a\in\partial i})\prod_{a\in C}\widetilde{\omega}(\{v_{(i,a)}\}_{i\in\partial a}) \tag{6.24}$$

where

$$\begin{aligned}\widetilde{\varphi}_i(\{v_{(i,a)}\}_{a\in\partial i}) = \sum_{\sigma\in\chi}\phi_i(\{\widehat{g}^{(s)}_{a\to i}\}^{s\in\{1,\dots,r\}}_{a\in\partial i})\prod_{a\in\partial i}\mathbb{I}[\sigma^a_i = g^{(1)}_{i\to a} = \sigma] \\ \times\prod_{s=1}^{r-1}\mathbb{I}[g^{(s+1)}_{i\to a} = G_s(\{\widehat{g}^{(s)}_{b\to i}\}_{b\in\partial i\setminus a})]\end{aligned} \tag{6.25}$$

$$\widetilde{\omega}(\{v_{(i,a)}\}_{i\in\partial a}) = \omega(\{\sigma^a_i\}_{i\in\partial a})\prod_{i=1}^{k}\prod_{s=2}^{r}\mathbb{I}[\widehat{g}^{(s)}_{a\to i} = \widehat{G}_s(\{g^{(s)}_{j\to a}\}_{j\in\partial a\setminus i})] \tag{6.26}$$

As before, one can check that for a configuration $\underline{v}$ in the support of this measure, σ^a_i is independent of a, and the marginal of $\underline{\sigma}$ is the initial biased measure. The partition function is the same in the two expressions.

This transformation can be applied for any definition of the auxiliary variables $\{g^{(s)}_{i\to a}, \widehat{g}^{(s)}_{a\to i}\}^{s\in\{1,\dots,r\}}_{a\in\partial i}$ and update rules $G_s, \widehat{G}_s$. In the cavity method that we will present in the next chapter, we study the typical properties of the measure μ chosen to describe the set of solution, when the graph G is drawn from a given random graph ensemble. One needs to introduce several layers of probability measures over the set of variables, therefore the computations are in general not tractable if the variables do not live in a small space. In practice it is needed that the auxiliary variables $\{g^{(s)}_{i\to a}, \widehat{g}^{(s)}_{a\to i}\}^{s\in\{1,\dots,r\}}_{a\in\partial i}$ live in a discrete space.

In [26], A. Braunstein, L. Dall'Asta, G. Semerjian and L. Zdeborova have used the WP messages $\nu^{(s)}_{i\to a}, \widehat{\nu}^{(s)}_{a\to i}$ defined in chapter 4, obtained after s iteration of the WP equations, with initial condition $\nu^{(1)}_{i\to a} = \delta_{\bullet,\sigma_i}$. In this paper, the bias is giving different weights to the configurations according to the fraction of variables that are still frozen after s iterations of the WP/whitening procedure. We will give more details about this bias in the next section.

A natural choice would have been to take for $\{g^{(s)}_{i\to a}, \widehat{g}^{(s)}_{a\to i}\}^{s\in\{1,\dots,r\}}_{a\in\partial i}$ the BP messages $\{\eta^{(s)}_{i\to a}, \widehat{\eta}^{(s)}_{a\to i}\}^{s\in\{1,\dots,r\}}_{a\in\partial i}$ obtained after s iterations, similarly as it was done for the WP messages in [26]. Of course the space of the BP messages is too big to be handled by the cavity method, even for messages over spin variable

that can be described from their mean value $m = \langle\sigma\rangle_\eta \in [-1, 1]$. One could have chosen the bias to depend on the marginal obtained after s iterations:

$$\varphi(\sigma_i, \underline{\sigma}_{B_{i,r}}) = \phi_i(\{\eta_i^{(s)}\}_{s\in\{1,\dots,r\}}) \tag{6.27}$$

with

$$\eta_i^{(s)}(\sigma_i) = \frac{\prod_{a\in\partial i} \widehat{\eta}_{a\to i}^{(s)}(\sigma_i)}{\sum_\sigma \prod_{a\in\partial i} \widehat{\eta}_{a\to i}^{(s)}(\sigma)} \tag{6.28}$$

In particular, one can think of a bias that would favor configurations for which the point-to-set correlation

$$C_i^{(s)} = \left| \frac{\sum_\sigma \sigma \eta_i^{(s)}(\sigma)}{\sum_\sigma \eta_i^{(s)}(\sigma)} \right| \tag{6.29}$$

is small. Note however that this is the point-to-set correlation function corresponding to the uniform measure, hence trying to minimize it is not guaranteed to increase the clustering threshold associated with the biased measure. With the WP messages this problem does not happen, since they are the same for both the uniform and the biased measure.

6.2 Biased measures studied in the literature

6.2.1 Bias counting the frozen variables

We now describe more precisely the bias introduced in [26], defined on the bicoloring of k-hypergraphs. The biased measure weights the solutions according to their fraction of frozen variables during the whitening procedure. One can find the frozen variables running the WP iterations defined in chapter 4 (in paragraph 4.1.6). For spin variables, it is suitable to use the notations, defined from the set of BP messages $\{\eta_{i\to a}, \widehat{\eta}_{a\to i}\}$:

$$h_{i\to a} = \begin{cases} +1 \text{ if } \eta_{i\to a}(\sigma_i) = \delta_{\sigma_i,+1} \\ -1 \text{ if } \eta_{i\to a}(\sigma_i) = \delta_{\sigma_i,-1} \\ 0 \text{ otherwise} \end{cases} \tag{6.30}$$

and similarly:

$$u_{a\to i} = \begin{cases} +1 \text{ if } \widehat{\eta}_{a\to i}(\sigma_i) = \delta_{\sigma_i,+1} \\ -1 \text{ if } \widehat{\eta}_{a\to i}(\sigma_i) = \delta_{\sigma_i,-1} \\ 0 \text{ otherwise} \end{cases} \tag{6.31}$$

For the bicoloring problem the WP equations reads:

$$u_{a\to i}^{t+1} = \begin{cases} +1 \text{ if } \forall j \in \partial a \setminus i : \ h_{j\to a}^{t} = -1 \\ -1 \text{ if } \forall j \in \partial a \setminus i : \ h_{j\to a}^{t} = +1 \\ 0 \text{ otherwise} \end{cases} \tag{6.32}$$

$$h_{i\to a}^{t+1} = \begin{cases} +1 \text{ if } \exists b \in \partial i \setminus a : \ u_{b\to i}^{t+1} = +1 \\ -1 \text{ if } \exists b \in \partial i \setminus a : \ u_{b\to i}^{t+1} = -1 \\ 0 \text{ otherwise} \end{cases} \tag{6.33}$$

To find the frozen variables in a solution $\underline{\sigma}$, one runs the WP iterations, starting from the initial condition:

$$h_{i\to a}^{0} = \sigma_i \tag{6.34}$$

One can show that the dynamical evolution of the WP messages is monotonic, and therefore guaranteed to converge to a fixed point, independently of the update scheme. One can translate the WP iterations into an evolution of configurations: $\underline{\sigma}^{t,WP} \in \{-1,+1,0\}^N$:

$$\sigma_i^{0,WP} = \sigma_i, \quad \text{and} \quad \sigma_i^{t,WP} = \begin{cases} +1 \text{ if } \exists b \in \partial i \setminus a : \ u_{b\to a}^{t+1} = +1 \\ -1 \text{ if } \exists b \in \partial i \setminus a : \ u_{j\to a}^{t+1} = -1 \\ 0 \text{ otherwise} \end{cases} \tag{6.35}$$

One says that the variable σ_i is whitened after time t if $\sigma_i^{t,WP} = 0$. Let $t_i(\underline{\sigma})$ the whitening time of the variable σ_i:

$$t_i(\underline{\sigma}) = \inf\{t : \sigma_i^{t,WP} = 0\} \tag{6.36}$$

if $t_i(\underline{\sigma}) = \infty$ then σ_i is declared frozen. Define the empirical cumulative distribution of the whitening times:

$$P_t(\underline{\sigma}, G) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}[t_i(\underline{\sigma}) \geq t+1] \tag{6.37}$$

$P_t(\underline{\sigma}, G)$ is the fraction of still frozen variables at time t in $\underline{\sigma}$. The authors have studied in particular the following biased measure:

$$\mu(\underline{\sigma}) = \frac{1}{Z(\epsilon, T, G)} \sum_{\underline{\sigma}} \prod_{a=1}^{M} \omega_a(\underline{\sigma}_{\partial a}) \prod_{i=1}^{N} \exp\left[N\epsilon P_T(\underline{\sigma}, G)\right] \tag{6.38}$$

with a fixed time $T \in \{1, \ldots, \infty\}$. When $\epsilon > 0$, this bias favors configurations with a large fraction $P_T(\underline{\sigma}, G)$ of frozen variables at time T, while when $\epsilon < 0$, this favors configurations with a small fraction of frozen variables after T iterations of the whitening procedure. This measure can also be writen in terms of the whitening times of each variables:

$$\mu(\underline{\sigma}) = \frac{1}{Z(\epsilon, T, G)} \sum_{\underline{\sigma}} \prod_{a=1}^{M} \omega_a(\underline{\sigma}_{\partial a}) \prod_{i=1}^{N} b(t_i(\underline{\sigma})) \tag{6.39}$$

with $b(t) = e^{\epsilon \mathbb{I}[t \geq T+1]}$. This bias introduce interactions between variables at distance T. Note that they also studied a biased measure that weight solutions according to their fraction of frozen variables for all times:

$$\mu(\underline{\sigma}) \frac{1}{Z(\{\epsilon_t\}, G)} \sum_{\underline{\sigma}} \prod_{a=1}^{M} \omega_a(\underline{\sigma}_{\partial a}) \exp\left[\sum_{t=1}^{\infty} N \epsilon P_t(\underline{\sigma}, G)\right], \tag{6.40}$$

thus introducing interactions at all distances in the graph. As already mentioned, the case $T = 1$ in the probability measure (6.38) corresponds to the bias introduced in section 6.1.3, with the correspondence $b = e^{\epsilon}$. The effect of the biased measure (6.38) (with $T = 1$) on the clustering transition is studied in [26]. For $k = 6$, they observed that is could not improve on the value of α_d. In [33], we showed that with another choice for the function $\psi(p)$ defined in (6.12) we could improve on its value, for $k = 4, 5$.

Under the RS assumption, the authors have computed the entropy $s(T, \theta, k, l)$ counting the number of solutions that have a fraction θ of frozen variables at time T:

$$s(T, \theta, k, l) = \lim_{N \to \infty} \mathbb{E}_G[s(T, \theta, G)], \quad \text{with} \tag{6.41}$$

$$s(T, \theta, G) = \frac{1}{N} \lim_{N \to \infty} \ln \left(\sum_{\underline{\sigma} \in S(G)} \mathbb{I}[P_T(\underline{\sigma}, G) = \theta] \right) \tag{6.42}$$

They have also computed the dominant path $P_t(T, \theta, k, l)$ which is the fraction of frozen variables at time t conditioned on $P_T = \theta$.

$$P_t(T, \theta, k, l) = \lim_{N \to \infty} \mathbb{E}_G \left[\frac{\sum_{\underline{\sigma} \in S(G)} P_t(\underline{\sigma}, G) \mathbb{I}[P_T(\underline{\sigma}, G) = \theta]}{\sum_{\underline{\sigma} \in S(G)} \mathbb{I}[P_T(\underline{\sigma}, G) = \theta]} \right] \tag{6.43}$$

If it tends to 0 as $t \to \infty$, then typical solutions under the constraint $P_T = \theta$ are unfrozen. They observed the existence of a threshold value θ^{tip} called the tipping point: for T, k, l fixed, below (resp. above) θ^{tip} the constraint $P_T = \theta$ selects unfrozen (resp. frozen) solutions. For fixed k, T, they define the threshold $l_T(k)$ above which the entropy of the tipping point $s(T, \theta^{tip}(T, k, l), k, l)$ becomes negative. Within the RS framework, this corresponds to disappearance of solutions such that $P_T = \theta^{tip}(k, l)$. In fact above $l_T(k)$, all the solutions have a fraction $\theta > \theta^{tip}$ of frozen variables at time T, thus contain a positive fraction of frozen variables in the large t limit. The threshold $l_T(k)$ can be interpreted as follows: below $l_T(k)$, it is possible to find a value of ϵ for which the typical solutions under the measure (6.38) are unfrozen, while for $l > l_T(k)$ it is not possible, and the typical solutions of (6.38) are frozen for any value of ϵ. They showed that $l_{T=1}(k) > l_r(k)$, where $l_r(k)$ is the rigidity transition under the uniform measure. They also showed that l_T is growing with T. l_f is interpreted as the limit $T \to \infty$ of l_T. They obtain the following large k scalings for this set of thresholds:

$$l_T(k) = \frac{2^{k-1} k \ln 2}{\ln^{\circ T}(k)} \left(1 + O\left(\frac{\ln^{\circ (T+1)}(k)}{\ln^{\circ T}(k)}\right)\right) \tag{6.44}$$

where $\ln^{\circ T}(k)$ is the T-times iterated logarithmic function: $\ln^{\circ(T+1)}(k) = \ln(\ln^{\circ T}(k))$. They estimated the scaling freezing transition to occur at $l_f(k) \sim 2^{k-1} k \ln 2/2 \sim l_{sat}(k)/2$. We recall that the scaling for $l_r(k)$ is

$$l_r(k) = 2^{k-1}(\ln k + \ln\ln k + 1 + O(\ln\ln k/\ln k)) \tag{6.45}$$

In conclusion, with the biased measure (6.38) at finite T, it is possible to obtain typical unfrozen solutions (tuning the value of the parameter ϵ), for densities of constraint significantly larger than the rigidity threshold corresponding to the uniform measure

In [30], H. Zhao and H. Zhou study a similar biased measure on the random k-SAT problem. They introduce a bias that weight the solutions according to the number of variables that are needed to satisfy the clauses. The others variables are white, in the sense that when they are flipped, the new configuration is still a solution. In this formalism, the variables σ_i can take 3 values (the white state being $\sigma_i = 0$). A configuration $\underline{\sigma}$ lives in $\{-1, +1, 0\}^N$ The partition function associated with the measure is

$$Z = \sum_{\underline{\sigma}} \prod_{a=1}^{M} \omega(\underline{\sigma}_{\partial a}) \prod_{i=1}^{N} e^{-\beta \delta^0_{\sigma_i}} \tag{6.46}$$

where $\omega(\underline{\sigma}_{\partial a})$ ensure the constraint a to be satisfied:

$$\omega(\underline{\sigma}_{\partial a}) = 1 - \prod_{i \in \partial a} \left(1 - \delta^{J^a_i}_{\sigma_i}\right) \ . \tag{6.47}$$

The parameter $\beta > 0$ favors configurations $\underline{\sigma} \in \{-1, +1, 0\}^N$ with a maximal number of variables that are equal to 0, therefore that do not contribute to the satisfaction of the clauses. They apply the BP-guided decimation procedure on this biased measure, and compute the probability of success for random 3-SAT. They observe a small positive probability of success for $\alpha \simeq 4.15$, while for the BPD algorithm on uniform measure they observe null probability of success above $\simeq 4.15$, thus indicating that the biased measure allows for better performances. Note that increasing the value of β, one might enter in a condensed phase where the BP marginals do not provide a good estimation on the exact marginals. It would also be interesting to estimate numerically the algorithmic threshold for this algorithm, as defined in chapter 5.

6.2.2 Local entropy

In [27, 28] the authors study a measure that favors solutions that are in regions with a high density of solutions. These regions are believed to be more accessible by the search algorithms, because they correspond to wide minima in the energy landscape. They weight each solution $\underline{\sigma}$ according to its local entropy, i.e. a measure of the number of solutions $\underline{\sigma}'$ whithin a given distance from the reference solution $\underline{\sigma}$. The local entropy is defined as follows:

$$F(\underline{\sigma}, \gamma) = \frac{1}{N} \log \mathfrak{N}(\underline{\sigma}, \gamma) \tag{6.48}$$

where

$$\mathfrak{N}(\underline{\sigma},\gamma)=\sum_{\underline{\sigma}'}\prod_{a\in C}\omega(\underline{\sigma}'_{\partial a})e^{-\gamma d(\underline{\sigma},\underline{\sigma}')} \tag{6.49}$$

$\mathfrak{N}(\underline{\sigma},\gamma)$ is a weighted sum on the solutions $\underline{\sigma}'\in S(G)$, with a weight that depends on the Hamming distance from the reference configuration $\underline{\sigma}$. γ is a coupling strength, in the limit $\gamma\to\infty$ this sum is dominated by the reference solution. When γ is small the terms selected in the sum belongs to wider regions around $\underline{\sigma}$. They show that this quantity can be used as a cost function to optimize on. More precisely, they introduce the following measure:

$$\mu(\underline{\sigma})=\frac{1}{Z(y,\gamma)}e^{yF(\underline{\sigma},\gamma)} \tag{6.50}$$

Note that they do not impose the condition $\underline{\sigma}\in S(F)$, and allow positive weights for configurations $\underline{\sigma}$ that are not solutions. In fact they expect that configurations that are surrounded by a dense region of solutions are themselves solutions. In practice they observe that this is the case experimentally: approximate sampling procedures on this measure allow to find solutions. The local entropy plays the role of the energy, and the parameter y plays the role of the inverse temperature. When y is large this favors configurations $\underline{\sigma}$ with large local entropy, i.e. surrounded by a a dense region of solutions. In [28] they also consider an additional parameter β, looking at the following distribution:

$$\mu(\underline{\sigma})=\frac{1}{Z(\beta,y,\gamma)}e^{y\phi(\underline{\sigma},\beta,\gamma)};\ \phi(\underline{\sigma},\beta,\gamma)=\ln\left(\sum_{\underline{\sigma}'}e^{-\beta E(\underline{\sigma}')-\gamma d(\underline{\sigma},\underline{\sigma}')}\right) \tag{6.51}$$

$\phi(\underline{\sigma},\beta,\gamma)$ is called the local free entropy, taking its $\beta\to\infty$ limit, one recovers the local entropy $F(\underline{\sigma},\gamma)$. With low value of γ, the entropy landscape is smoother than the energy landscape, its local minima corresponding to dense regions of local minima in the energy landscape. This corresponds to coarse-grain the energy landscape, with γ controlling the granularity. The goal with this approach is to avoid being trapped in the local minima of the energy landscape. Increasing the value of γ, one focus on the configurations $\underline{\sigma}'$ that are closer to the reference solution $\underline{\sigma}$, therefore this allows to focus on configurations $\underline{\sigma}$ which are inside a narrow local minimum. If β is large enough this is likely to be a global minimum. When $\gamma\to\infty$ the only configuration selected in $\phi(\underline{\sigma},\beta,\gamma)$ is $\underline{\sigma}'=\underline{\sigma}$, and one recovers the initial Gibbs measure, with inverse temperature βy, and the entropy landscape tends to the initial energy landscape.

They perform approximate sampling procedures, such as Simulated Annealing, and Message-Passing decimation procedures to sample from this distribution, with cooling schemes that progressively increase the value of the parameters β, γ and y. Note that the local free entropy and the local entropy are not easy to compute, compared to the energy function.

To obtain this quantity, a first approach developed in [27] is to study the system defined in $\mathfrak{N}$, in which each variable i is coupled to an external field $e^{\gamma\sigma'_i}$ (using the expression of the Hamming distance: $d(\underline{\sigma},\underline{\sigma}')=\sum_{i=1}^{N}(1-\sigma_i\sigma'_i)/2)$).

The parameter β is not considered in [27] ($\beta = \infty$). $F(\underline{\sigma}, \gamma)$ can be estimated with the Bethe Peirls approximation, running the BP equations for this modified system, and computing the Bethe free energy from the messages obtained. Note that this approach might fail in the condensed phase, because the BP iterations might not converge, or the Bethe estimation might not be correct. They apply a Simulated Annealing algorithm defined as follows: $\underline{\sigma}$ is initialized at random. At each step $F(\underline{\sigma}, \gamma)$ is computed. Then a random local flip is accepted or rejected using Metropolis rule at fixed temperature $1/y$. More precisely, let ΔF be the difference in local entropy. If there is an increase in ΔF, then the flip is always accepted. If there is a decreese in ΔF, then the flip is accepted w.p $e^{y\Delta F}$, and rejected otherwise. Applying this algorithm on the perceptron and on the k-SAT problem, they observe that this algorithm performs better than the usual Simulated Annealing algorithm that uses the energy function (as defined in chapter 3).

In [28] they use another approach to compute the local free entropy. Instead of using BP, consider that y is a non-negative integer, and rewrite the partition function

$$\begin{aligned} Z(\beta, y, \gamma) &= \sum_{\underline{\sigma}} e^{yF(\underline{\sigma}, \beta, \gamma)} \\ &= \sum_{\underline{\sigma}, \underline{\sigma}^1, \dots, \underline{\sigma}^y} e^{-\beta \sum_{a=1}^{y} E(\underline{\sigma}^a) - \gamma \sum_{a=1}^{y} d(\underline{\sigma}, \underline{\sigma}^a)} \\ &= \sum_{\underline{\sigma}^1, \dots, \underline{\sigma}^y} e^{-\beta(\sum_{a=1}^{y} E(\underline{\sigma}^a) + A(\{\underline{\sigma}^a, \beta, \gamma\}))} \end{aligned} \tag{6.52}$$

where in the last equation the reference configuration has been traced out

$$A(\{\underline{\sigma}^a, \beta, \gamma\}) = -\frac{1}{\beta} \ln \sum_{\underline{\sigma}} e^{-\gamma \sum_{a=1}^{y} d(\underline{\sigma}, \underline{\sigma}^a)} \tag{6.53}$$

They apply Simulated Annealing algorithm to sample from this measure, with increasing γ, decreasing β. At each step, a replica $a \in \{1, \dots, y\}$ is randomly chosen, and the randomly chosen variable inside this replica is flipped according to the Metropolis rule computed from the change in the energy function defined from the probability law with partition function (6.52). The algorithm stops either when one solution is found in one of the replicas, or when the maximal number of steps is reached. They observe a better performance using this method on the perceptron, compared to the non interacting case ($\gamma = 0$), which correponds to running the energetic Simutaled Annealing procedure in parallel on the y independent copies.

6.2.3 Biased interactions in the hard sphere packing problem

In [96] and [29] the authors use a biased measure to sample solutions for the problem of packing hard spheres. In this problem, a set of N spheres of fixed

radius live in a d-dimensional volume V. The variables correspond to the position of each sphere. The constraint is that for each pair of spheres, the distance between their centers is larger than their diameter, so that they do not overlap, hence the name hard sphere. A solution is a assignment of the positions of the spheres so that there is no overlapping. The parameter controlling the hardness of this problem is the packing fraction $\varphi = Nv_d/V$, v_d being the volume of a unit-diameter sphere. The thermodynamic limit is $N, V \to \infty$ with φ constant. As φ increases, one encounters a dynamical transition φ_d similar to the one found in random k-SAT. Below φ_d, the solutions can be sampled uniformly in polynomial time in N, using Monte Carlo methods, or molecular dynamics, while above this requires an exponential time. This result is valid in the large d limit (taken after the thermodynamic limit).

The uniform measure over the set of solutions can be viewed as a Gibbs-Boltzmann distribution describing a set of particles that interact with each other through hard interactions. The authors introduce a bias in the interactions: keeping the hard constraint (i.e. forbidding the spheres to overlap), they add a soft interaction at larger distances. They look at the change of φ_d under this bias, trying to maximize its value. The biased pair-wise interaction $v(r)$ between two particles at distance r takes the following form:

$$v(r) = \begin{cases} \infty, & r < 1 \\ v_+(r), & r > 1 \end{cases} \tag{6.54}$$

where we have considered spheres of unit diameter, so that two spheres are overlapping when their distance r is smaller than 1. $v_+(r)$ is an arbitrary function that decrease to 0 as $r \to \infty$. The uniform case is recovered for $v_+(r) = 0$. More precisely, in [96], they have shown that with this specific form:

$$v_+(r) = \begin{cases} -U_0, & r \leq r_0 \\ 0, & r > r_0 \end{cases} \tag{6.55}$$

for some $r_0 > 1$ and $U_0 > 0$, it is possible to increase the dynamical transition φ_d, thus to increase the range for which it is possible to sample solutions in polynomial time in N. This interaction introduces short range attractive interaction, just after the hard wall. Intuitively, this means that a measure that favor solutions that are packed more closely can be sampled in polynomial time for higher packing fractions than the uniform measure.

In [29] a more detailed soft interaction with more than one step is studied:

$$v_+(r) = \begin{cases} -U_0, & r \leq r_0 \\ -U_1, & r_0 < r \leq r_0 + r_1 \\ \dots & \\ -U_{n-1}, & \sum_{i=0}^{n-2} r_i < r \leq \sum_{i=0}^{n-1} r_i \\ 0, & r > \sum_{i=0}^{n-1} r_i \end{cases} \tag{6.56}$$

Optimizing on the several parameters, they could improve on the dynamical packing fraction φ_d. They observe that the profile that maximises the value of

φ_d should have a short-range attractive part, then a weak repulsive tail at larger r.

Chapter 7

Cavity Method

In this chapter we present the cavity method treatment of random CSPs. We apply this method to the bicoloring on k-hypergraphs, and study the typical properties of the two biased measures (defined in (6.7) and in (6.12)) that describe the set of solutions. The replica and the cavity method have been first developed in the context of statistical mechanics of disordered systems. Random CSPs bear a formal similarity with mean-field spin glasses, the interactions induced by the constraints being of a frustrating nature while lacking a finite-dimensional structure thanks to the randomness in their construction. This analogy has allowed to apply these techniques to random CSPs [4, 5, 6, 7, 8]. This line of study has provided predictions of $\alpha_{sat}(k)$ for many models [97, 44], but also unveiled many other phase transitions for the structure of the set of solutions in the satisfiable phase [98] (see chapter 2 for the description of some of them). Moreover, it has lead to the proposal of new algorithms that exploit this detailed picture of the solution space (see chapter 5 for the presentation of some of them). Many of these (heuristic) predictions have been confirmed rigorously later on [9, 10, 11, 12].

We will pay a particular attention to the clustering phase transition $\alpha_d \leq \alpha_{sat}$, which is also known as the dynamical or reconstruction transition. This transition can be defined in several ways. Below α_d the set of solutions of typical instances is rather well-connected, any solution can be reached from another through a path consituted of nearby solutions. Above α_d the solution set splits into an exponential number of isolated subsets of solutions, called clusters, which are internally well-connected but separated from each other by regions without solutions. This is called the clustering phenomenon (see section 2.3.1). This transition is also defined as the appearance of long range correlations between variables, known as the point-to-set correlation [14]. It was shown in [14] that these correlations forbid the rapid equilibration of the stochastic processes that respect the detailed balance condition, such as Simulated Annealing algorithm (see section 3.1.4). This property motivates the name dynamic transition. In the cavity method [15] treatment of the random CSPs, we define α_d as the appearance of a non trivial solution of the one step of Replica Symmetry Breaking

(1RSB) equation with Parisi breaking parameter $\mathcal{X} = 1$. The appearance of a non-zero point-to-set correlation also implies the solvability of an information theoretic problem called hypertree reconstruction [13], see [16] for the connection between 1RSB equations at $\mathcal{X} = 1$ and the reconstruction problem, that will be discussed in Section 7.3. In this setting, α_d corresponds to the hypertree reconstruction threshold, above which the problem becomes solvable.

The study of the structural phase transitions in the satisfiable regime, and in particular the definition of α_d in terms of long-range correlations, relies on the characterization of a specific probability law on the space of configurations, namely the uniform measure over solutions. In this chapter we will compute the clustering threshold for the two biased measures defined in the previous chapter (in (6.7) and in (6.12)). In chapter 8 we will show that for finite values of k optimizing on the parameters describing the biases allows us to increase α_d, and therefore to improve the performances of simulated annealing (we checked it for the biased measure (6.7)). In chapter 10 we will study the effect of the bias on the clustering threshold in the large k limit.

This chapter is organized as follows. In Section 7.1 we recall the definitions of the two biased measures, and write the BP equations (see section 4.1.2). In Section 7.2 we present the simplest version of the cavity method, called replica symmetric (RS) that provides predictions for the typical properties of the measures. In Section 7.3 we derive the formalism that allows us to compute the clustering threshold, exploiting its definition in framework of the reconstruction problem. In Section 7.4 we explain how to compute the rigidity threshold α_r, that provides an upper bound on the clustering threshold $\alpha_d \leq \alpha_r$. In section 7.5 we derive recursive distributional equations that will be used in the large k expansion of α_d (see chapters 9 and 10). In Section 7.6 we present the 1RSB formalism, that is the first non-trivial level of the cavity method, which allows to deal with the RSB phenomenon, and provides estimates for several structural phase transitions (such as the clustering, condensation and satisfiability transitions). Finally in Section 7.7 we compute an analytical upper-bound on α_d, the Kesten-Stigum threshold. This chapter summarizes part of the derivations performed in [31], [32] and [33].

7.1 Definition of the model

7.1.1 Graphical model

We consider in this chapter the two graphical models representing the biased measures introduced in chapter 6. We will denote by Θ_0 the graphical model for the intra-clause measure (see equation (6.7)), that we recall here:

$$\mu_{\Theta_0}(\underline{\sigma}) = \frac{1}{Z_{\Theta_0}(G)} \prod_{a=1}^{M} \omega(\underline{\sigma}_{\partial a}) \tag{7.1}$$

Θ_0 gathers both the underlying factor graph G_{Θ_0} (which in this case coincides with the initial factor graph $G = (V, C, E)$ representing the instance F), and the

choice of the function ω (defined by the parameters $\{\omega_p\}_{p=0,\dots,k}$ as in equation (6.8)). Note that the uniform measure is recovered in this setting by choosing $\omega_0 = \omega_k = 0$ and $\omega_1 = \cdots = \omega_{k-1} = 1$.

As explained in the section 6.1.3, the factor graph of the biased measure with interactions at distance 1 (see equation (6.12)) contains small loops compared to the initial factor graph representing F. The measure (6.12) has been modified, introducing auxiliary variables $v_{(i,a)}$ on the edges $(i,a) \in E$, where $v = (\sigma, w) \in \{-1,+1\} \times \{0,1\}$. We will denote by Θ_1 the graphical model for the measure over the set of variables $\underline{v} = \{v_{(i,a)}\}_{(i,a)\in E}$ (see equation (6.16)). Θ_1 gathers the modified factor graph G_{Θ_1} and the choice of the function ψ (introduced right after the equation (6.12)) that counts the number of forcing clauses surrounding each vertex $i \in V$. In this setting, the uniform measure is recovered by choosing $\psi(p) = 1$ for all p. The new factor graph G_{Θ_1} lies on the initial factor graph G, the new variables $v_{i,a}$ sit on the initial edges $(i,a) \in E$, and there are two types of function nodes, living on the vertices V and the k-hyperedges C respectively. We recall the expression of the measure μ_{Θ_1} introduced in (6.16):

$$\mu_{\Theta_1}(\underline{v}) = \frac{1}{Z_{\Theta_1}(G)} \prod_{i\in V} \widetilde{\varphi}(\{v_{(i,a)}\}_{a\in\partial i}) \prod_{a\in C} \widetilde{\omega}(\{v_{(i,a)}\}_{i\in\partial a}) \tag{7.2}$$

We recall the expression of the function nodes:

$$\widetilde{\varphi}(\sigma_1, w_1, \dots, \sigma_{l+1}, w_{l+1}) = \psi\left(\sum_{i=1}^{l+1} w_i\right) \mathbb{I}[\sigma_1 = \cdots = \sigma_{l+1}] \tag{7.3}$$

$$\widetilde{\omega}(\sigma_1, w_1, \dots, \sigma_k, w_k) = \omega(\sigma_1, \dots, \sigma_k) \prod_{i=1}^{k} \mathbb{I}[w_i = \mathbb{I}[\{w_i = \mathbb{I}[\{\sigma_j\}_{j\neq i} a.e]\}]] \tag{7.4}$$

Assume that the underlying factor graph $G = (V, C, E)$ is distributed according to a random ensemble, such as the k-uniform $l+1$-regular random graph ensemble, or the Erdős Rényi random graph ensemble $G_N(k, M)$, that have been defined in chapter 1, part 1.4.2. As explained in chapter 1 (part 1.4.5), these structures are locally tree-like. The aim of the cavity method is to determine the typical properties of the graphical models μ_Θ (with $\Theta \in \{\Theta_0, \Theta_1\}$) defined on this graph, and of the free energy density $\ln Z_\Theta(G)/N$ in the thermodynamic limit. The first step of the cavity method amounts to study such models on finite trees. In that case we have seen in chapter 4 that one can exploit the recursive nature of trees to derive an exact description of μ_Θ in terms of its marginals (see part 4.1.3), using the Belief Propagation messages (defined part 4.1.1). The free energy can be expressed in terms of these messages (see part 4.1.4).

7.1.2 BP equations and Bethe free-energy

We recall in this section the Belief Propagation equations introduced in chapter 4 (see equations (4.1) and 4.2)). We write the BP equations for the two graphical models Θ_0 and Θ_1 representing the biased measures.

Intra-clause bias

We introduce the messages $\eta_{i\to a}$ and $\widehat{\eta}_{a\to i}$ on each edge (i,a) of the factor graph $G_{\Theta_0} = G$, that are the marginal probability laws of σ_i in amputated graphs where some interactions are discarded; $\eta_{i\to a}$ is the marginal of σ_i in the factor graph where one removes the hyperedge a, and $\widehat{\eta}_{a\to i}$ is the marginal of σ_i in the factor graph where one removes all the hyperedges in $\partial i \setminus a$. Removing an interaction in a tree breaks it into independent subtrees, which allows to write recursive equations between these messages:

$$\eta_{i\to a}(\sigma_i) = \frac{1}{z^{ia}} \prod_{b\in\partial i\setminus a} \widehat{\eta}_{b\to i}(\sigma_i) \tag{7.5}$$

$$\widehat{\eta}_{a\to i}(\sigma_i) = \frac{1}{\widehat{z}^{ai}} \sum_{\underline{\sigma}_{\partial a\setminus i}} \omega(\underline{\sigma}_{\partial a}) \prod_{j\in\partial a\setminus i} \eta_{j\to a}(\sigma_j) \ , \tag{7.6}$$

where the constants z^{ia} and $\widehat{z}^{ai}$ are normalizing factors. These equations are valid for any (discrete) domain of the spins σ_i; as we are studying the case where $\sigma_i = \pm 1$, we can parametrize the probability laws $\eta_{i\to a}$ and $\widehat{\eta}_{a\to i}$ by their mean values, defining $\eta_{i\to a}(\sigma_i) = (1 + h_{i\to a}\sigma_i)/2$ and $\widehat{\eta}_{a\to i}(\sigma_i) = (1 + u_{a\to i}\sigma_i)/2$, with $h_{i\to a}, u_{a\to i} \in [-1, 1]$. The recursive equations can be rewritten with this parametrization as

$$h_{i\to a} = f_{\Theta_0}(\{u_{b\to i}\}_{b\in\partial i\setminus a}) \ , \qquad u_{a\to i} = \widehat{f}_{\Theta_0}(\{h_{j\to a}\}_{j\in\partial a\setminus i}) \ , \tag{7.7}$$

where the functions f_{Θ_0} and $\widehat{f}_{\Theta_0}$ read explicitly

$$f_{\Theta_0}(u_1,\dots,u_d) = \frac{\prod_{i=1}^{d}(1+u_i) - \prod_{i=1}^{d}(1-u_i)}{z_{0,\Theta_0}(u_1,\dots,u_d)} \tag{7.8}$$

$$z_{0,\Theta_0}(u_1,\dots,u_d) = \prod_{i=1}^{d}(1+u_i) + \prod_{i=1}^{d}(1-u_i) \tag{7.9}$$

$$\widehat{f}_{\Theta_0}(h_1,\dots,h_{k-1}) = \frac{\sum_{\sigma_1,\dots,\sigma_k} \omega(\sigma_1,\dots,\sigma_k)\sigma_k \prod_{i=1}^{k-1}(1+h_i\sigma_i)}{\widehat{z}_{0,\Theta_0}(h_1,\dots,h_{k-1})} \tag{7.10}$$

$$\widehat{z}_{0,\Theta_0}(h_1,\dots,h_{k-1}) = \sum_{\sigma_1,\dots,\sigma_k} \omega(\sigma_1,\dots,\sigma_k) \prod_{i=1}^{k-1}(1+h_i\sigma_i) \tag{7.11}$$

The function f_{Θ_0} has been written here for a vertex of degree $d+1$.

On a tree the equations (7.7) admit a single solution, that can be found by iterating them from the leaves towards the interior of the graph. Once this solution is determined one can easily compute the marginal probability of σ_i under μ_{Θ_0} using the formula in (7.5) with all messages incoming onto i (see section 4.1.3), as well as the partition function $Z_{\Theta_0}(G)$. The Bethe free entropy

density introduced in section 4.1.4 provides an exact expression for the free entropy density: $(1/N)\ln Z_{\Theta_0}(G)) = \Phi^{Bethe}_{\Theta_0}$ where

$$\Phi^{Bethe}_{\Theta_0} = \frac{1}{N}\sum_{i=1}^{N} \ln \mathcal{Z}^{\rm v}_{0,\Theta_0}(\{u_{a\to i}\}_{a\in\partial i}) + \frac{1}{N}\sum_{a=1}^{M} \ln \mathcal{Z}^{\rm c}_{0,\Theta_0}(\{h_{i\to a}\}_{i\in\partial a}) - \frac{1}{N}\sum_{(i,a)} \ln \mathcal{Z}^{\rm e}_{0,\Theta_0}(h_{i\to a}, u_{a\to i}) \ , \tag{7.12}$$

where the last sum runs over the edges of the factor graph, and the local partition functions are defined as:

$$\mathcal{Z}^{\rm v}_{0,\Theta_0}(u_1,\dots,u_d) = \sum_{\sigma}\prod_{i=1}^{d}\left(\frac{1+\sigma u_i}{2}\right) \ , \tag{7.13}$$

$$\mathcal{Z}^{\rm c}_{0,\Theta_0}(h_1,\dots,h_k) = \sum_{\sigma_1,\dots,\sigma_k}\omega(\sigma_1,\dots,\sigma_k)\prod_{i=1}^{k}\left(\frac{1+\sigma_i h_i}{2}\right) \ , \tag{7.14}$$

$$\mathcal{Z}^{\rm e}_{0,\Theta_0}(h,u) = \sum_{\sigma}\left(\frac{1+\sigma h}{2}\right)\left(\frac{1+\sigma u}{2}\right) \ . \tag{7.15}$$

Finally, the Shannon entropy of the measure μ_{Θ_0} can be expressed from the BP messages, see equation (4.21):

$$S[\mu_{\Theta_0}] = N\Phi^{Bethe}_{\Theta_0} - \sum_{a\in C}\frac{\sum_{\underline{\sigma}_{\partial a}}\omega(\underline{\sigma}_{\partial a})\ln(\omega(\underline{\sigma}_{\partial a}))\prod_{i\in\partial a}\eta_{i\to a}(\sigma_i)}{\sum_{\underline{\sigma}_{\partial a}}\omega(\underline{\sigma}_{\partial a})\prod_{i\in\partial a}\eta_{i\to a}(\sigma_i)} \tag{7.16}$$

Bias with interactions at distance 1

We now specify the BP equations and the expression of the thermodynamic quantities for the measure μ_{Θ_1} (equation (7.2)). The BP messages are $\eta_{i\to a}$ and $\widehat{\eta}_{a\to i}$, the marginal laws of the variable $v_{(i,a)}$ that is placed on the edge (i,a) of G, in graphs where one has removed the interactions a and i, respectively. These definitions are illustrated in the right panel of figure 7.1. Note that in a literal application of the BP algorithm one would have introduced messages from every variable node to every interaction node, for instance $\eta_{i\to(i,a)}$ and $\eta_{(i,a)\to a}$; as the variable nodes are of degree two these two messages are actually equal, we denoted their common value $\eta_{i\to a}$ to lighten the notations. The BP equations between these messages are of the form

$$\eta_{i\to a} = f_{\Theta_1}(\{\widehat{\eta}_{b\to i}\}_{b\in\partial i\setminus a}), \quad \widehat{\eta}_{a\to i} = \widehat{f}_{\Theta_1}(\{\eta_{j\to a}\}_{j\in\partial a\setminus i}) \tag{7.17}$$

where the functions f_{Θ_1} and $\widehat{f}_{\Theta_1}$ derive from the interaction nodes $\widetilde{\varphi}$ and $\widetilde{\omega}$ stated in equations (7.3,7.4). The relation $\eta = f_{\Theta_1}(\widehat{\eta}_1,\dots,\widehat{\eta}_l)$ is thus found to mean

$$\eta(\sigma,w) = \frac{1}{z_{0,\Theta_1}(\widehat{\eta}_1,\dots,\widehat{\eta}_l)}\sum_{w_1,\dots,w_l}\psi\left(w+\sum_{i=1}^{l}w_i\right)\prod_{i=1}^{l}\widehat{\eta}_i(\sigma,w_i) \tag{7.18}$$

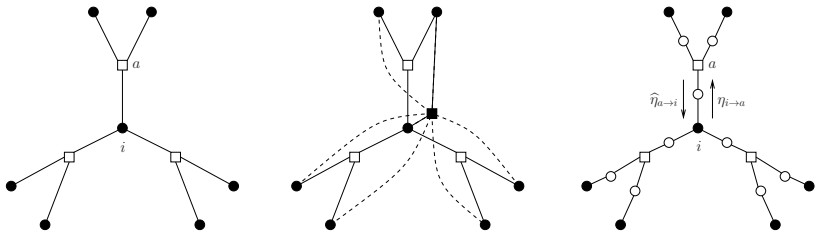

Figure 7.1: Left panel: an example of a hypergraph G with $N = 7$ vertices represented by black circles, and $M = 3$ hyperedges linking $k = 3$ vertices, drawn as white squares. Center panel: the introduction of an interaction, represented as a black square, between all the vertices at distance 1 from the central vertex i, generates short loops even if G is a tree. Right panel: the factor graph representation of the probability measure (7.2), the white circles stand for the variable nodes $v_{(i,a)}$, the black circles (resp. white squares) are the interaction factors $\widetilde{\psi}$ (resp. $\tilde{\omega}$). The messages $\eta_{i\to a}$ and $\widehat{\eta}_{a\to i}$ obey the Belief Propagation equations (7.17).

where $z_{0,\Theta_1}(\widehat{\eta}_1, \dots, \widehat{\eta}_l)$ is a normalization constant. Similarly $\widehat{\eta} = \widehat{f}_{\Theta_1}(\eta_1, \dots, \eta_{k-1})$ stands for:

$$\widehat{\eta}(\sigma, w) = \frac{1}{\widehat{z}_{0,\Theta_1}(\eta_1, \dots, \eta_{k-1})} \sum_{\substack{\sigma_1,\dots,\sigma_{k-1} \\ w_1,\dots,w_{k-1}}} \omega(\sigma, \sigma_1, \dots, \sigma_{k-1}) \mathbb{I}[w = \mathbb{I}[\sigma_1, \dots, \sigma_{k-1} a.e]] \prod_{i=1}^{k-1} \eta_i(\sigma_i, w_i) \mathbb{I}[w_i = \mathbb{I}[\sigma, \sigma_1, \dots, \sigma_{i-1}, \sigma_{i+1}, \dots, \sigma_{k-1} a.e]] \tag{7.19}$$

with $\widehat{z}_{0,\Theta_1}(\eta_1, \dots, \eta_{k-1})$ a normalization constant. More explicitly one has:

$$\widehat{\eta}(\sigma, 1) = \frac{1}{\widehat{z}_{0,\Theta_1}(\eta_1, \dots, \eta_{k-1})} \prod_{i=1}^{k-1} \eta_i(-\sigma, 0) \tag{7.20}$$

$$\widehat{\eta}(\sigma, 0) = \frac{1}{\widehat{z}_{0,\Theta_1}(\eta_1, \dots, \eta_{k-1})} \Big[\sum_{i=1}^{k-1} \eta_i(-\sigma, 1) \prod_{j \neq i} \eta_j(\sigma, 0) + \sum_{\substack{I \subset \{1,\dots,k-1\} \\ 2 \leq |I| \leq k-2}} \prod_{i \in I} \eta_i(-\sigma, 0) \prod_{i \notin I} \eta_i(\sigma, 0) \Big] \tag{7.21}$$

as there is at most one variable which is the unique representant of its color in a set of $k \geq 3$ binary variables that is not monochromatic.

The Bethe free energy takes the following form:

$$\Phi^{Bethe}_{\Theta_1} = \frac{1}{N}\sum_{i=1}^{N} \ln \mathcal{Z}^{\rm v}_{0,\Theta_1}(\{\widehat{\eta}_{a\to i}\}_{a\in\partial i}) + \frac{1}{N}\sum_{a=1}^{M} \ln \mathcal{Z}^{\rm c}_{0,\Theta_1}(\{\eta_{i\to a}\}_{i\in\partial a}) - \frac{1}{N}\sum_{(i,a)} \ln \mathcal{Z}^{\rm e}_{0,\Theta_1}(\eta_{i\to a},\widehat{\eta}_{a\to i}) \ , \tag{7.22}$$

where:

$$\mathcal{Z}^{\rm v}_{0,\Theta_1}(\widehat{\eta}_1,\dots,\widehat{\eta}_d) = \sum_{\sigma}\sum_{w_1,\dots,w_d} \psi\left(\sum_{i=1}^{d} w_i\right)\prod_{i=1}^{d}\widehat{\eta}_i(\sigma,w_i) \ , \tag{7.23}$$

$$\mathcal{Z}^{\rm c}_{0,\Theta_1}(\eta_1,\dots,\eta_k) = \sum_{\substack{\sigma_1,\dots,\sigma_k\\ w_1,\dots,w_k}} \omega(\sigma_1,\dots,\sigma_k)\prod_{i=1}^{k}\eta_i(\sigma_i,w_i)\mathbb{I}\left[w_i = \mathbb{I}[\{\sigma_j\}_{j\neq i}a.e]\right] \ , \tag{7.24}$$

$$\mathcal{Z}^{\rm e}_{0,\Theta_1}(\eta,\widehat{\eta}) = \sum_{\sigma,w}\eta(\sigma,w)\widehat{\eta}(\sigma,w) \ . \tag{7.25}$$

For the Shannon entropy of the measure μ_{Θ_1} we get the expression:

$$S[\mu_{\Theta_1}] = N\Phi^{Bethe}_{\Theta_1} - \sum_{i\in V}\frac{\sum_{\sigma,\{w_(i,a)\}_{a\in\partial i}}\psi\left(\sum_{a\in\partial i}w_{(ia)}\right)\ln\left(\psi\left(\sum_{a\in\partial i}w_{(ia)}\right)\right)\prod_{i\in\partial a}\eta_{i\to a}(\sigma,w_{(i,a)})}{\sum_{\sigma,\{w_{(i,a)}\}_{a\in\partial i}}\psi\left(\sum_{a\in\partial i}w_{(ia)}\right)\prod_{i\in\partial a}\eta_{i\to a}(\sigma,w_{(i,a)})} \tag{7.26}$$

7.2 Replica symmetric cavity method

The aim of the cavity method is to characterize the properties of the measure μ_Θ for typical random graphs in the thermodynamic limit. In particular one would like to estimate the value of the quenched free-entropy density and of the averaged Shannon entropy density

$$\Phi(k,\alpha,\Theta) = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}[\ln Z_\Theta(G)], \quad s(k,\alpha,\Theta) = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}[S[\mu_\Theta]] \tag{7.27}$$

thanks to the self-averaging phenomenon, the free entropy density $(\ln Z_\Theta(G))/N$ and the entropy density are expected to concentrate on these values. There are different versions of the cavity method, that rely on self-consistent hypotheses of various complexity on the effect of long loops. In the simplest version, called replica symmetric (RS), one assumes a fast decay of the correlations between distant variables in the measure μ_Θ, in such a way that the BP equations converge to a unique fixed point on a typical large instance, and that the measure is well decribed by the locally tree-like approximation. Increasing α above the clustering threshold $\alpha_d(k,\Theta)$ causes the appearance of long-range correlations

between distant variables under the measure μ_Θ. In such a case the RS hypothesis breaks down, we will present in the section 7.6 the one-step replica symmetry breaking formalism, that is able to deal with this RSB phenomenon [15].

7.2.1 RS cavity equations

Consider an uniformly chosen directed edge $i \to a$, in a random hypergraph, and call $\mathcal{P}^{RS}$ the probability law of the message $\eta_{i\to a}$ solution of the BP equation. Similarly let $\widehat{\mathcal{P}}^{RS}$ be the probability law of the message $\widehat{\eta}_{a\to i}$. The RS cavity method assumes the decorrelation of the incoming messages on a uniformly chosen vertex i (resp hyperedge a). The incoming messages are i.i.d with the law $\widehat{\mathcal{P}}^{RS}$ (resp. $\mathcal{P}^{RS}$). The distributions obey the following self-consistent equations:

$$\mathcal{P}^{RS}(\eta) = \sum_{d=0}^{\infty} r_d \int \prod_{i=1}^{d} \mathrm{d}\widehat{\mathcal{P}}^{RS}(\widehat{\eta}_i)\delta[\eta - f_\Theta(\widehat{\eta}_1, \dots, \widehat{\eta}_d)] \tag{7.28}$$

$$\widehat{\mathcal{P}}^{RS}(\widehat{\eta}) = \int \prod_{i=1}^{k-1} \mathrm{d}\mathcal{P}^{RS}(\eta_i)\delta[\widehat{\eta} - \widehat{f}_\Theta(\eta_1, \dots, \eta_{k-1})] \tag{7.29}$$

Where $r_d = \mathbb{I}[d = l]$ for the random k-uniform, $l+1$-regular graph, and $r_d = e^{-\alpha k}(\alpha k)^d/d!$ for the ER ensemble. These equations can be equivalently written as equalities in distribution between random variables (we use the symbol $\stackrel{d}{=}$). For instance (7.28) means

$$\eta \stackrel{d}{=} f_\Theta(\widehat{\eta}_1, \dots, \widehat{\eta}_d) \tag{7.30}$$

with η drawn from $\mathcal{P}^{RS}$, d is a random variable drawn from r_d, and the $\widehat{\eta}_i$'s are i.i.d. copies of a random variable of law $\widehat{\mathcal{P}}^{RS}$.

In general, there might be non-uniqueness of the solution of the RS equations (7.28,7.29), and the determination of the relevant solution (that describes properly the typical properties of the measure) is not obvious. The solution of these equations can be solved using population dynamics, that one will explain in the next chapter, section 8.1. We will see however that for the measure μ_{Θ_0} the relevant solution can be found by symmetry considerations. For the measure μ_{Θ_1}, it is also possible to guess the solution, when one focuses on the k-regular $l+1$-uniform ensemble.

7.2.2 Intra-clause bias

As we assume that $\omega_p = \omega_{k-p}$, i.e. that the model is invariant under the global spin-flip symmetry, the RS equations admit as a solution the uniform distributions $\mathcal{P}^{RS}(h) = \delta(h)$, $\widehat{\mathcal{P}}^{RS}(u) = \delta(u)$. For frustrated models with an antiferromagnetic character this is the relevant solution (the local order that

might emerge from a non-trivial solution is incompatible with the long loops of all sizes).

The RS cavity prediction for the free-entropy (7.27) is then obtained by averaging the Bethe expression (7.12) with respect to the message distributions $\mathcal{P}^{RS}$ and $\widehat{\mathcal{P}}^{RS}$, which yields:

$$\begin{aligned}\Phi^{RS}(k,\alpha,\{\omega_p\}) = \sum_{d=0}^{\infty} p_d \int \left(\prod_{i=1}^{d} \mathrm{d}u_i \widehat{\mathcal{P}}^{RS}(u_i)\right) \ln \mathcal{Z}_0^{\mathrm{v}}(u_1,\dots,u_d) \\ + \alpha \int \left(\prod_{i=1}^{k} \mathrm{d}h_i \mathcal{P}^{RS}(h_i)\right) \ln \mathcal{Z}_0^{\mathrm{c}}(h_1,\dots,h_k) \\ - \alpha k \int \mathrm{d}h\mathrm{d}u \mathcal{P}^{RS}(h)\widehat{\mathcal{P}}^{RS}(u) \ln \mathcal{Z}_0^{\mathrm{e}}(h,u) \ . \end{aligned} \tag{7.31}$$

One obtains explicity the value of the free-entropy by inserting the trivial solution of the RS equations into (7.31):

$$\begin{aligned}\Phi^{RS}(k,\alpha,\{\omega_p\}) &= \ln 2 + \alpha \ln \left(\frac{1}{2^k} \sum_{\sigma_1,\dots,\sigma_k} \omega(\sigma_1,\dots,\sigma_k)\right) \\ &= \ln 2 + \alpha \ln \left(\frac{1}{2^k} \sum_{p=0}^{k} \binom{k}{p} \omega_p\right) \ . \end{aligned} \tag{7.32}$$

Note that this expression actually coincides with the annealed (first moment) computation $\lim_{N\to\infty}(\ln \mathbb{E}[Z(G)])/N$.

In the special case $\omega_0 = \omega_k = 0$, $\omega_1 = \cdots = \omega_{k-1} = 1$, for which $\mu_{\Theta_0}(\underline{\sigma})$ corresponds to the uniform measure over proper bicolorings, the partition function $Z_{\Theta_0}(G)$ counts the number of solutions, hence the free-entropy $\ln Z_{\Theta_0}(G)$ is equal to the entropy of the uniform measure. The prediction of the RS cavity method for the entropy density is thus (using a subscript 'u' for uniform):

$$s_{\mathrm{u}}^{RS}(k,\alpha) = \Phi_{\mathrm{u}}^{RS}(\alpha) = \ln 2 + \alpha \ln \left(1 - \frac{1}{2^{k-1}}\right) \ . \tag{7.33}$$

For a generic choice of parameters $\{\omega_p\}$ the free-entropy $\ln Z_{\Theta_0}$ differs from the (Shannon) entropy of the measure $\mu_{\Theta_0}(\underline{\sigma})$. Averaging over the expression (7.16), and using the expression of the joint marginal of the variables around a constraint (see equation (4.8)) one thus obtains the RS prediction for the entropy density

$$s^{RS}(k,\alpha,\{\omega_p\}) = \ln 2 + \alpha \ln \left(\frac{1}{2^k} \sum_{p=0}^{k} \binom{k}{p} \omega_p\right) - \alpha \, \frac{\sum_{p=0}^{k} \binom{k}{p} \omega_p \ln \omega_p}{\sum_{p=0}^{k} \binom{k}{p} \omega_p} \ . \tag{7.34}$$

This quantity is a decreasing function of α and becomes negative for $\alpha > \alpha^{s=0}$, with

$$\alpha^{s=0}(k,\{\omega_p\}) = \frac{\ln 2}{\frac{\sum_{p=0}^{k}\binom{k}{p}\omega_p \ln \omega_p}{\sum_{p=0}^{k}\binom{k}{p}\omega_p} - \ln\left(\frac{1}{2^k}\sum_{p=0}^{k}\binom{k}{p}\omega_p\right)} \ . \tag{7.35}$$

The negativity of the entropy for $\alpha > \alpha^{s=0}$ is a clear sign of the failure of the RS assumptions, the Shannon entropy of a discrete probability measure being non-negative. This is however not the only mechanism for the appearance of Replica Symmetry Breaking (RSB), as we shall see this phenomenon can occur in a phase with $s > 0$.

7.2.3 Bias with interaction at distance 1

In this case, it is a priori not obvious to guess the solution of the RS equation (7.28,7.29) from symmetry arguments. Indeed, the dependence in w of the BP messages is non-trivial, and may vary from one vertex to another. In the general case, one needs to solve numerically the equations (7.28,7.29). This can be done using population dynamics, a method that we will explain in the section 8.1. However one might find more than one solution, and it remains difficult to determine which solution is relevant to estimate the typical properties of the measure μ_{Θ_1}. The situation gets simplified in the case of the k-uniform $l+1$-regular hypergraph. In this case, the local neighborhood of every vertex is the same, it is thus natural to look for a translationally invariant solution of the BP equations. Moreover the probability measure we are studying is invariant under the spin-flip symmetry $\underline{\sigma} \to -\underline{\sigma}$, we can thus further restrict ourselves to a solution of the BP equation that respects this invariance. This amounts to take $\eta_{i\to a}(\sigma, w) = \eta_*(w)$, $\widehat{\eta}_{a\to i}(\sigma, w) = \widehat{\eta}_*(w)$ for all edges (i, a). Plugging this form into (7.17) yields the equations satisfied by η_* and $\widehat{\eta}_*$:

$$\eta_*(w) = \frac{1}{z}\sum_{p=0}^{l}\binom{l}{p}\psi(p+w)\widehat{\eta}_*(0)^{l-p}\widehat{\eta}_*(1)^p \ , \tag{7.36}$$

$$\widehat{\eta}_*(1) = \frac{1}{\widehat{z}}\eta_*(0)^{k-1} \ , \tag{7.37}$$

$$\widehat{\eta}_*(0) = \frac{1}{\widehat{z}}\left[(k-1)\eta_*(1)\eta_*(0)^{k-2} + (2^{k-1}-k-1)\eta_*(0)^{k-1}\right] \ . \tag{7.38}$$

Introducing the ratio of probabilities

$$y = \frac{\eta_*(0)}{\eta_*(1)} \ , \qquad \widehat{y} = \frac{\widehat{\eta}_*(0)}{\widehat{\eta}_*(1)} \ , \tag{7.39}$$

one can get rid of the normalization factors z and $\widehat{z}$ and rewrite (7.36,7.37,7.38) more simply

$$y = \frac{\sum_{p=0}^{l} \binom{l}{p} \psi(p) \widehat{y}^{-p}}{\sum_{p=0}^{l} \binom{l}{p} \psi(p+1) \widehat{y}^{-p}} , \qquad \widehat{y} = 2^{k-1} - k - 1 + \frac{k-1}{y} . \tag{7.40}$$

It turns out that for any choice of the parameters k, l, and ψ there exists a unique solution $(y, \widehat{y})$ to the equations (7.40), which might not be obvious at first sight; a proof of this existence and uniqueness is provided in Appendix A.

One obtains the RS prediction for the free entropy density and the entropy density by averaging the expressions (7.22) and (7.26); on the translationally invariant solution this yields:

$$\begin{aligned} \Phi^{RS}(k,l,\psi) = &\left(1 - \frac{(l+1)(k-1)}{k}\right) \ln 2 - (l+1) \ln \left(1 + \frac{1}{y\widehat{y}}\right) \\ &+ \ln \left(\sum_{p=0}^{l+1} \binom{l+1}{p} \psi(p) \widehat{y}^{-p} \right) + \frac{l+1}{k} \ln \left(2^{k-1} - k - 1 + \frac{k}{y} \right) , \end{aligned} \tag{7.41}$$

$$s^{RS}(k,l,\psi) = \Phi^{RS}(l,\psi) - \frac{\sum_{p=0}^{l+1} \binom{l+1}{p} \psi(p) \widehat{y}^{-p} \ln \psi(p)}{\sum_{p=0}^{l+1} \binom{l+1}{p} \psi(p) \widehat{y}^{-p}} . \tag{7.42}$$

One can see again for some choices of the parameters, in particular when l gets large enough, this expression of the entropy becomes negative. We denote $l^{s=0}(k,\psi)$ the first integer value such that the entropy becomes negative.

7.3 The dynamic transition

We shall now present the formalism that allows to compute the location of the dynamic transition which, as explained in the introduction, manifests itself in different ways. Here we shall exploit its definition in terms of the existence of long-range point-to-set correlations in the probability measure μ_Θ [16, 14], that are related to the solvability of a tree reconstruction problem [13]. In the section 7.6 we will explain the connection with the 1RSB formalism.

The point-to-set correlation function, or overlap, has been introduced in chapter 2 (see equation (2.15)). The function $C(i, B)$ measures the correlation between the variable at the point i and those in the set B. Let us define the point-to-set correlation function averaged over the graph ensemble, at distance n, as follows:

$$C_n = \lim_{N\to\infty} \mathbb{E}[C(0, B_n)] = \lim_{N\to\infty} \mathbb{E}[\langle \sigma_0 \langle \sigma_0 \rangle_{\underline{\sigma}_{B_n}} \rangle - \langle \sigma_0 \rangle^2] , \tag{7.43}$$

where 0 is an arbitrary reference vertex and B_n is the set of vertices at distance n from 0; $\langle\cdot\rangle$ denotes the expectation with respect to μ_Θ, while $\langle\cdot\rangle_{\underline{\sigma}_{B_n}}$ is the conditional average with the law $\mu_\Theta(\cdot|\underline{\sigma}_{B_n})$. Note that the second term in C_n actually vanishes thanks to the invariance of μ_Θ under the spin-flip transformation. For the biased measure μ_{Θ_1} with interactions at distance 1, because the interactions in the biased measure couple spins belonging to neighboring hyperedges it is not enough to take for B_n the set of variables at distance exactly n from the root; it is however equivalent to include in B_n all the variables are distance at least n, or at distances n and $n+1$. The dynamic transition separates an underconstrained (Replica Symmetric, RS) regime in which $C_n \to 0$ as $n \to \infty$, and an overconstrained (Replica Symmetry Breaking, RSB) one in which C_n remains strictly positive at large distances. To compute C_n we first remark that the local neighborhood of the vertex 0, up to any finite distance, converges when $N \to \infty$ to a regular tree structure. Moreover the marginal law of μ on any finite neighborhood of G converges, within the hypothesis of the RS solution described in Sec. 7.2, to a measure that admits an explicit description in terms of a broadcast process that we describe below for the two measures μ_{Θ_0} and μ_{Θ_1}.

7.3.1 The broadcast process

Generating a configuration with the law of μ in a finite neighborhood of a root vertex 0 amounts indeed to do a sequential sampling, starting from the root variable σ_0, which is drawn uniformly at random in $\{-1,+1\}$. The value of the variables in the first generation $\{\underline{\sigma}_{\partial a\setminus 0}\}_{\partial 0}$ are then drawn from the conditional law $\mu(\{\underline{\sigma}_{\partial a\setminus 0}\}_{\partial 0}|\sigma_0)$. In chapter 4 (section 4.1.3) we have seen how to compute the conditional law from the BP messages that are solutions of the BP equations (see equation (4.11)). Then the value of the variables of the next generations are chosen according to the conditional law knowing the value of the previous variables. We first specify this broadcast process for the measure μ_{Θ_0} representing the intra-clause biased measure, then describe it for the measure μ_{Θ_1} with interactions at distance 1, in the k-uniform $l+1$-regular graph ensemble.

Broadcast process for the measure with intra-clause bias μ_{Θ_0}

In this case we have seen that the solution of the RS equation (7.28,7.29) is $\mathcal{P}^{RS}(h) = \delta(h), \widehat{\mathcal{P}}^{RS}(u) = \delta(u)$. On a finite set of variables the measure μ_{Θ_0} is then described (within the RS assumption) by the set of messages $\{h_{i\to a} = 0, u_{a\to i} = 0\}$ for (i,a) in the subgraph containing the subset of variables. One can then use this solution to sample a configuration from μ_{Θ_0}, using the following broadcast process. Choose σ_0 to ± 1 with probability $1/2$. Then, independently for each edge a adjacent to the root, draw the configuration of the $k-1$ other variables with the conditional probability $\widetilde{p}(\underline{\sigma}_{\partial a\setminus 0}|\sigma_0)$, where

$$\widetilde{p}(\sigma_1,\dots,\sigma_{k-1}|\sigma) = \frac{\omega(\sigma_1,\dots,\sigma_{k-1},\sigma)}{\sum_{\sigma'_1,\dots,\sigma'_{k-1}} \omega(\sigma'_1,\dots,\sigma'_{k-1},\sigma)} \ . \tag{7.44}$$

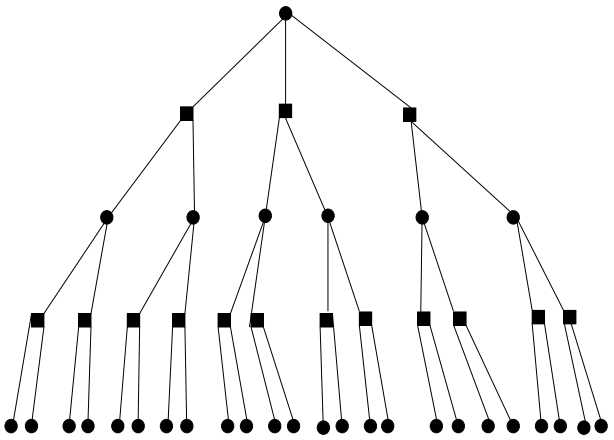

Figure 7.2: An example of a hypertree, vertices being represented by circles, while hyperedges are drawn as squares linked to k vertices, with $k = 3$ on figure.

Once the values of all the spins at distance 1 from the root have been set in this way the same process can be iterated, each of these vertices of the first generation being in turn considered as the root of the subtree lying below it. On figure 7.2 is represented a hypertree with a root on the top and two generations of vertices below it, on which the broadcast process is performed.

Broadcast process for the bias at distance 1

In this case we have seen that the solution of the RS equation (7.28,7.29) is $\mathcal{P}^{RS}(\eta) = \delta[\eta - \eta_*], \widehat{\mathcal{P}}(\widehat{\eta}) = \delta[\widehat{\eta} - \widehat{\eta}_*]$, where $\eta_*(v), \widehat{\eta}_*(v)$ are solution of the equation (7.36-7.38). The broadcast process takes the following form:

- choose $\sigma_0 = \pm 1$ with equal probability 1/2.
- draw the $l+1$ variables $v_1 = (\sigma_1, w_1), \ldots, v_{l+1} = (\sigma_{l+1}, w_{l+1})$ adjacent to the root with the probability

$$p_0(v_1, \ldots, v_{l+1}|\sigma_0) = \frac{\psi\left(\sum_{i=1}^{l+1} w_i\right) \prod_{i=1}^{l+1} \widehat{\eta}_*(w_i)}{\sum_{w'_1, \ldots w'_{l+1}} \psi\left(\sum_{i=1}^{l+1} w'_i\right) \prod_{i=1}^{l+1} \widehat{\eta}_*(w'_i)} \prod_{i=1}^{l+1} \mathbb{I}[\sigma_i = \sigma_0] \ . \tag{7.45}$$

- consider each of the $v_1, \ldots, v_{l+1}$ variables of the first generation as the root of the subtree lying below it, and draw the value of the descendents

$v_1, \dots v_{k-1}$ of a variable equal to v from the conditional law

$$\widehat{p}(v_1, \dots, v_{k-1}|v) = \frac{\omega(\sigma, \sigma_1 \dots \sigma_{k-1}) \mathbb{I}[w = \mathbb{I}[\{\sigma_i\} \text{ a.e.}]] \prod_{i=1}^{k-1} \eta_*(w_i) \mathbb{I}[w_i = \mathbb{I}[\sigma, \{\sigma_j\}_{j \neq i} \text{ a.e.}]]}{\sum_{\{\sigma'_i, w'_i\}} \omega(\sigma, \sigma'_1 \dots \sigma'_{k-1}) \mathbb{I}[w = \mathbb{I}[\{\sigma'_i\} \text{ a.e.}]] \prod_{i=1}^{k-1} \eta_*(w'_i) \mathbb{I}[w'_i = \mathbb{I}[\sigma, \{\sigma'_j\}_{j \neq i} \text{ a.e.}]]} . \tag{7.46}$$

- consider again the variables of the second generation as roots, and extract the value of their descendents from the conditional law

$$p(v_1, \dots, v_l|v) = \frac{\psi\left(w + \sum_{i=1}^{l} w_i\right) \prod_{i=1}^{l} \widehat{\eta}_*(w_i)}{\sum_{w'_1, \dots w'_l} \psi\left(w + \sum_{i=1}^{l} w'_i\right) \prod_{i=1}^{l} \widehat{\eta}_*(w'_i)} \prod_{i=1}^{l} \mathbb{I}[\sigma_i = \sigma] \ . \tag{7.47}$$

- iterate the last two steps until all the variables in the target neighborhood have been assigned.

On figure 7.3 is represented the hypertree with the root 0 at the top on which lives the root variable σ_0, and three generation of variables below it. This broadcast procedure, that must be performed on the v variables and not only on the σ's to preserve the Markov structure of the tree, can be interpreted as the transmission of an information (the value σ_0 at the root) through noisy channels (the conditional laws $p_0, \widehat{p}$ and p defined in (7.45,7.46,7.47)), that we will describe below.

7.3.2 The reconstruction problem and its recursive distributional equations

This description of the sampling from μ_Θ as a broadcast process naturally calls for an interpretation as the transmission of an information, the value of the spin at the root, through noisy channels, towards a set of receivers, the vertices at a certain distance n from the root, to be denoted B_n. For the measure μ_{Θ_0} the noisy channels are represented by the hyperedges and correspond to the conditional probability expressed in (7.44). For the measure μ_{Θ_1} the noisy channels are the conditional laws $p_0, \widehat{p}$ and p defined in (7.45,7.46,7.47) In this information theoretic perspective the tree reconstruction problem asks the following question: given only the values $\underline{\sigma}_{B_n}$ of the spins at distance n from the root in a configuration $\underline{\sigma}$ generated as above, how much information is available on the value of the root σ_0, in the sense that the observation of $\underline{\sigma}_{B_n}$ allows to infer the value of the root σ_0 with a success probability larger than the one expected from a random guess ? Does a non-vanishing amount of information on σ_0 survives in the $n \to \infty$ limit ? In this Bayesian setting the information-theoretical optimal strategy of an observer having to reconstruct σ_0 from $\underline{\sigma}_{B_n}$ is to compute

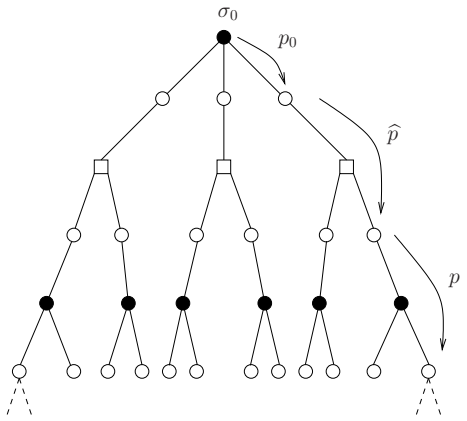

Figure 7.3: The tree structure considered for the computation of C_n, represented here for $k = l + 1 = 3$. The generation of a configuration from the law ρ is performed in a broadcast fashion, the root σ_0 being ± 1 with probability $1/2$, this information is then propagated down the tree with transmission channels p_0, $\widehat{p}$ and p.

the posterior probability $\mu_\Theta(\sigma_0|\underline{\sigma}_{B_n})$, which for Ising variables is completely described by the magnetization $\langle\sigma_0\rangle_{\underline{\sigma}_{B_n}}$. The correlation function C_n is a possible way of quantifying this amount of information, the tree reconstruction problem being solvable if and only if C_n remains strictly positive in the large n limit. Thanks to the tree structure of the interaction graph this strategy is actually computationally feasible and can be performed in a recursive way, that we now specify for the two measures $\mu_{\Theta_0}, \mu_{\Theta_1}$.

Reconstruction for the bias with intra-clause interactions

Once supplemented by the boundary conditions

$$h_{i\to a} = \sigma_i \quad \text{for} \quad i \in B_n \ , \tag{7.48}$$

the BP equations (7.7) admit a unique solution for any finite tree, and the posterior probabilities $\mu_{\Theta_0}(\sigma_i|\underline{\sigma}_{B_n})$ can be determined, for $i \notin B_n$, from these messages according to

$$\mu_{\Theta_0}(\sigma_i|\underline{\sigma}_{B_n}) = \frac{\prod_{a\in\partial i} 1 + u_{a\to i}\sigma_i}{\sum_\sigma \prod_{a\in\partial i} 1 + u_{a\to i}\sigma} \ , \tag{7.49}$$

note that in order to lighten the notations we have kept implicit the dependency of the messages on the observed variables $\underline{\sigma}_{B_n}$. The average value of the spin at the root, in the posterior probability measure conditional on the observations

of the spins at distance n, reads

$$\langle \sigma_0 \rangle_{B_n} = h_0 = \sum_{\sigma_0} \sigma_0 \mu_{\Theta_0}(\sigma_0 | \underline{\sigma}_{B_n}) = f_{\Theta_0}(\{u_{a\to 0}\}_{a\in\partial 0}) \ . \tag{7.50}$$

(with f_{Θ_0} defined in the equation (7.8)). It is this value which has to be compared with the actual value of σ_0 in the broadcast process that led to $\underline{\sigma}_{B_n}$ in order to assess the amount of information transmitted between 0 and B_n. Once averaged over $\mu_{\Theta_0}(\underline{\sigma})$ this yields the point-to-set correlation function $C(0, B_n) = \langle \sigma_0 \langle \sigma_0 \rangle_{B_n} \rangle$.

The computation descrived above was performed for a given tree. We now want to take the average $\mathbb{E}$ over the graph ensemble. We recall that the finite set B_n converges in distribution to a tree drawn as an instance of a Galton-Watson process with n generations, in which every vertex has a random offspring (number of descendent hyperedges) distributed as a Poisson law of parameter αk (we consider here the ER ensemble $G_N(k, \alpha N)$). Let us call as above $h \in [-1, 1]$ the posterior magnetization of the root given the observations of the spins on the vertices of the n-th generation (the root corresponding to the generation $n = 0$). h is a random variable because of the randomness in the generation of the Galton-Watson tree on the one hand, and in the broadcast process yielding $\underline{\sigma}_{B_n}$ on the other hand. We shall denote $P_{\sigma,n}(h)$ its distribution when the broadcast is conditioned to the root being equal to σ. As both the broadcast process (from the root downwards) and the resolution of the BP equations (from the leaves upwards) decompose recursively along the branches of the tree, it is possible to obtain from the above analysis an inductive formula relating these distributions for trees of depth n and $n + 1$, namely

$$P_{\sigma,n+1}(h) = \sum_{l=0}^{\infty} e^{-\alpha k} \frac{(\alpha k)^l}{l!} \int \prod_{i=1}^{l} \mathrm{d}\widehat{P}_{\sigma,n}(u_i)\, \delta(h - f_{\Theta_0}(u_1, \ldots, u_l)) \ , \tag{7.51}$$

$$\widehat{P}_{\sigma,n}(u) = \sum_{\sigma_1,\ldots,\sigma_{k-1}} \widetilde{p}(\sigma_1, \ldots, \sigma_{k-1} | \sigma) \int \prod_{i=1}^{k-1} \mathrm{d}P_{\sigma_i,n}(h_i)\, \delta(u - \widehat{f}_{\Theta_0}(h_1, \ldots, h_{k-1})) \ , \tag{7.52}$$

with the initial condition $P_{\sigma,0}(h) = \delta(h - \sigma)$ that expresses the observation of the variables at the boundary of the tree. These are examples of Recursive Distributional Equations (RDEs), as they define by recursion a sequence of probability distributions. They can be equivalently written as equalities in distribution between random variables, for instance (7.51) means $h \stackrel{d}{=} f_{\Theta_0}(u_1, \ldots, u_l)$ with h drawn from $P_{\sigma,n+1}$, l is a Poisson distributed random variable with mean αk, and the u_i's are i.i.d. copies of a random variable of law $\widehat{P}_{\sigma,n}$.

One can slightly simplify these equations by noticing that the invariance of the bicoloring problem under a global spin-flip of all its variables implies that $P_{-\sigma,n}(h) = P_{\sigma,n}(-h)$ and $\widehat{P}_{-\sigma,n}(u) = \widehat{P}_{\sigma,n}(-u)$; this can be checked by induction from (7.51,7.52), using $\widetilde{p}(\sigma_1, \ldots, \sigma_{k-1} | \sigma) = \widetilde{p}(-\sigma_1, \ldots, -\sigma_{k-1} | -\sigma)$ and the fact that f_{Θ_0} and $\widehat{f}_{\Theta_0}$ change sign when all their arguments are multiplied

by -1. It is thus redundant to track the evolution with n of the distributions with both $\sigma = +1$ and $\sigma = -1$, we shall instead define $P_n(h) = P_{+1,n}(h)$ and $\widehat{P}_n(u) = \widehat{P}_{+1,n}(u)$ and close the equations on these two sequences of distributions:

$$P_{n+1}(h) = \sum_{l=0}^{\infty} e^{-\alpha k} \frac{(\alpha k)^l}{l!} \int \prod_{i=1}^{l} \mathrm{d}\widehat{P}_n(u_i)\, \delta(h - f_{\Theta_0}(u_1, \ldots, u_l)) \;, \tag{7.53}$$

$$\widehat{P}_n(u) = \sum_{\sigma_1,\ldots,\sigma_{k-1}} \widetilde{p}(\sigma_1, \ldots, \sigma_{k-1}| + 1) \int \prod_{i=1}^{k-1} \mathrm{d}P_n(h_i)\, \delta(u - \widehat{f}_{\Theta_0}(h_1\sigma_1, \ldots, h_{k-1}\sigma_{k-1})) \;. \tag{7.54}$$

The answer to the reconstructibility question raised above can be read off from the behavior of $P_n(h)$ in the large n limit: if it tends to the trivial distribution $\delta(h)$ (which is always a stationary solution of (7.53,7.54)), then all the information on the value of the root has been washed out and the reconstruction problem is not solvable. On the contrary if the limit of $P_n(h)$ is non-trivial then the observation of $\underline{\sigma}_{B_n}$ contains some information on the root and the problem is said to be solvable. The occurence of these two situations depend on the parameters k, α and $\{w_p\}$ of the model; increasing α gives rise to a larger number of spin variables to be observed, which makes the inference problem easier. Hence it is natural to expect the existence of a threshold $\alpha_d(k, \{w_p\})$ such that the problem is unsolvable (resp. solvable) for $\alpha < \alpha_d$ (resp. $\alpha > \alpha_d$). We will see in the section 7.6.2 that this threshold coincides with the clustering threshold for the random bicoloring on k-hypergraphs. Indeed, the cavity formalism leads to the same distributional equations as equations (7.51,7.52). More precisely, in the cavity framework they corresponds to the 1RSB equations with Parisi parameter $\mathcal{X} = 1$ [16].

It is actually possible to describe this transition by the point-to-set correlation function, which constitute a scalar order parameter (instead of the functional one P_n). It can be expressed as follows

$$C_n(\alpha, k, \{w_p\}) = \sum_{\sigma} \frac{1}{2} \int \mathrm{d}P_{\sigma,n}(h) \sigma h \;. \tag{7.55}$$

Using the symmetry properties described above this expression can be simplified:

$$C_n(\alpha, k, \{w_p\}) = \int \mathrm{d}P_n(h)\, h \;. \tag{7.56}$$

The solvability threshold is determined by the large n limit of the point-to-set correlation:

$$C(\alpha, k, \{w_p\}) = \lim_{n\to\infty} C_n(\alpha, k, \{w_p\}) \;. \tag{7.57}$$

This is indeed an order parameter for the transition in the sense that $C(\alpha, k, \{w_p\}) > 0$ if and only if $\alpha > \alpha_d(k, \{w_p\})$. This equivalence is a consequence of some symmetry properties of the distributions P_n that we will describe now.

Let us call $T_n(h)$ the distribution of the posterior magnetization h of the root, in a broadcast process which is not conditioned on the value of the root in $\underline{\sigma}$, i.e. $T_n(h) = (P_{+,n}(h) + P_{-,n}(h))/2$. A consequence of Bayes theorem applied to the joint law between the spins at the root and on the boundary vertices is the following converse relation between conditional and unconditional laws, $P_{\sigma,n}(h) = (1 + \sigma h)T_n(h)$ (see [16, 39, 99] for more details on this property and its consequences). In addition the invariance of the problem under a global spin-flip implies that $T_n(-h) = T_n(h)$, which yields a symmetry constraint on the distributions P_n,

$$P_n(-h) = \frac{1-h}{1+h}\, P_n(h) \ , \qquad \int \mathrm{d}P_n(h)\, b(h) = \int \mathrm{d}P_n(h)\, b(-h)\, \frac{1-h}{1+h} \ , \quad (7.58)$$

the second equality being valid for any function $b(h)$ such that the integrals exist. This symmetry (sometimes called Nishimori symmetry [100], and also fulfilled by $\widehat{P}_n(u)$) implies several identities between the moments of h that can be derived by appropriate choices of the test function b [39, 99]; here we shall only state the simplest one, namely

$$C_n(\alpha, k, \{w_p\}) = \int \mathrm{d}P_n(h)\, h = \int \mathrm{d}P_n(h)\, h^2 \ , \quad (7.59)$$

that can be obtained from (7.58) with $b(h) = h(1-h)$. This is enough to justify our statement of the characterization of the reconstruction transition via the behavior of C_n: if the latter vanishes in the $n \to \infty$ limit this implies that both the average and the variance of P_n go to zero, hence P_n tends to the trivial distribution $\delta(h)$.

Reconstruction for the bias with interactions at distance 1

We now give the explicit derivation of the posterior probability $\mu_{\Theta_1}(\sigma_0|\underline{\sigma}_{B_n})$, adapting the computations explained in the previous section. In this case we look at the k-uniform $l+1$-regular graph ensemble. The conditional measure $\mu_{\Theta_1}(\cdot|\underline{\sigma}_{B_n})$ is exactly described in terms of the solution of the BP equations (7.17), supplemented with the boundary condition $\widehat{\eta}_{i\to a}(v) = \delta_{v,v_{(i,a)}}$ on the edges at distance larger than n of the root, $v_{(i,a)}$ being the value taken by the variable during the broadcast. These BP messages, directed towards the root, are thus random variables because of the randomness in the boundary condition $\underline{\sigma}_{B_n}$. One can nevertheless write recursion equations on their distributions, their law depending only on their distance from the boundary, as it was done in the previous section. We shall denote $P_{v,n}(\eta)$ the law of the message η on an edge at distance n from the boundary, conditional on the value of the variable on this edge being v in the broadcast, and similarly $\widehat{P}_{v,n}(\widehat{\eta})$ for the law of the $\widehat{\eta}$ messages. Putting together all the above observations leads to the following

recursion equations:

$$P_{v,n+1}(\eta) = \sum_{v_1,\dots,v_l} p(v_1,\dots,v_l|v) \int \prod_{i=1}^{l} \mathrm{d}\widehat{P}_{v_i,n+1}(\widehat{\eta}_i)\, \delta(\eta - f_{\Theta_1}(\widehat{\eta}_1,\dots,\widehat{\eta}_l)) \ , \tag{7.60}$$

$$\widehat{P}_{v,n+1}(\widehat{\eta}) = \sum_{v_1,\dots,v_{k-1}} \widehat{p}(v_1,\dots,v_{k-1}|v) \int \prod_{i=1}^{k-1} \mathrm{d}P_{v_i,n}(\eta_i)\, \delta(\widehat{\eta} - \widehat{f}_{\Theta_1}(\eta_1,\dots,\eta_{k-1})) \ , \tag{7.61}$$

where f_{Θ_1}, $\widehat{f}_{\Theta_1}$, $\widehat{p}$ and p have been defined in (7.17,7.46,7.47), respectively, and with the initial condition for $n = 0$:

$$\widehat{P}_{v,0}(\widehat{\eta}) = \delta(\widehat{\eta}(\cdot) - \delta_{v,\cdot}) \ . \tag{7.62}$$

The point-to-set correlation function is then computed as

$$C_n = \frac{1}{2} \sum_{\sigma_0} \sum_{v_1,\dots,v_{l+1}} p_0(v_1,\dots,v_{l+1}|\sigma_0) \int \prod_{i=1}^{l+1} \mathrm{d}\widehat{P}_{v_i,n}(\widehat{\eta}_i)\, \sigma_0\, m(\widehat{\eta}_1,\dots,\widehat{\eta}_{l+1}) \ , \tag{7.63}$$

where p_0 is the law defined in (7.45), and with the expression

$$m(\widehat{\eta}_1,\dots,\widehat{\eta}_{l+1}) = \frac{m_+(\widehat{\eta}_1,\dots,\widehat{\eta}_{l+1}) - m_-(\widehat{\eta}_1,\dots,\widehat{\eta}_{l+1})}{m_+(\widehat{\eta}_1,\dots,\widehat{\eta}_{l+1}) + m_-(\widehat{\eta}_1,\dots,\widehat{\eta}_{l+1})} \ , \tag{7.64}$$

$$m_\sigma(\widehat{\eta}_1,\dots,\widehat{\eta}_{l+1}) = \sum_{w_1,\dots,w_{l+1}} \psi\left(\sum_{i=1}^{l+1} w_i\right) \prod_{i=1}^{l+1} \widehat{\eta}_i(\sigma, w_i) \ , \tag{7.65}$$

for the conditional average magnetization of the root.

The recursion equations (7.60,7.61), (which are equivalent to the 1RSB equations with Parisi breaking parameter equal to $\mathcal{X} = 1$), always admit the trivial solution $P_v(\eta) = \delta(\eta - \eta_*)$, $\widehat{P}_v(\widehat{\eta}) = \delta(\widehat{\eta} - \widehat{\eta}_*)$ as a stationary fixed point. In the non-reconstructible (RS) phase this is the limit reached by $P_{v,n}$ and $\widehat{P}_{v,n}$ in the large n limit, and then $C_n \to 0$. On the contrary in the reconstructible (RSB) phase the limit of $P_{v,n}$ and $\widehat{P}_{v,n}$ is a non-trivial fixed point, and C_n remains strictly positive. For a given choice of the parameters k and ψ we define the dynamic transition l_d as the threshold separating these two behaviors. As l is here an integer parameter we will say more precisely that $l < l_\mathrm{d}$ is the RS phase, $l \geq l_\mathrm{d}$ the RSB phase, i.e. $l_\mathrm{d}(k,\psi)$ is the smallest integer value of l such that RSB occurs.

Simplifications and symmetries

The recursion equations (7.60,7.61) bear, for each value of n, on eight distributions $P_{v,n}$, $\widehat{P}_{v,n}$, as the variable $v = (\sigma, w)$ takes four different values. This

number can however be divided by two thanks to the invariance of the problem under the spin-flip symmetry $\underline{\sigma} \to -\underline{\sigma}$. To state its consequences let us define the flip transformation $\eta \to \eta^f$ between messages, according to $\eta^f(\sigma, w) = \eta(-\sigma, w)$ (and similarly $\widehat{\eta}^f(\sigma, w) = \widehat{\eta}(-\sigma, w)$). The channels p and $\widehat{p}$ being invariant under a global spin-flip, one can check that

$$P_{(-,w),n}(\eta) = P_{(+,w),n}(\eta^f) \ , \quad \widehat{P}_{(-,w),n}(\widehat{\eta}) = \widehat{P}_{(+,w),n}(\widehat{\eta}^f) \ , \tag{7.66}$$

which allows to close (7.60,7.61) on the four distributions $\{P_{(+,w),n}, \widehat{P}_{(+,w),n}\}_{w=0,1}$, that we shall denote for simplicity $\{P_{w,n}, \widehat{P}_{w,n}\}_{w=0,1}$. Using this property, as well as the invariance of f_{Θ_1}, $\widehat{f}_{\Theta_1}$ under a permutation of their arguments and a more explicit version of the expressions (7.46,7.47) of $\widehat{p}$ and p, one can simplify (7.60,7.61) into:

$$P_{w,n+1}(\eta) = \sum_{p=0}^{l} \frac{\binom{l}{p}\psi(p+w)\widehat{y}^{-p}}{\sum_{p'=0}^{l}\binom{l}{p'}\psi(p'+w)\widehat{y}^{-p'}} \int \prod_{i=1}^{p} \mathrm{d}\widehat{P}_{1,n+1}(\widehat{\eta}_i)$$
$$\times \prod_{i=p+1}^{l} \mathrm{d}\widehat{P}_{0,n+1}(\widehat{\eta}_i)\, \delta(\eta - f_{\Theta_1}(\widehat{\eta}_1, \dots, \widehat{\eta}_l)) \ , \tag{7.67}$$

$$\widehat{P}_{1,n+1}(\widehat{\eta}) = \int \prod_{i=1}^{k-1} \mathrm{d}P_{0,n}(\eta_i)\, \delta(\widehat{\eta} - \widehat{f}_{\Theta_1}(\eta_1^f, \dots, \eta_{k-1}^f)) \ , \tag{7.68}$$

$$\widehat{P}_{0,n+1}(\widehat{\eta}) = \frac{k-1}{y\widehat{y}} \int \mathrm{d}P_{1,n}(\eta_1) \prod_{i=2}^{k-1} \mathrm{d}P_{0,n}(\eta_i)\, \delta(\widehat{\eta} - \widehat{f}_{\Theta_1}(\eta_1^f, \eta_2, \dots, \eta_{k-1}))$$
$$+ \frac{1}{\widehat{y}} \sum_{t=2}^{k-2} \binom{k-1}{t} \int \prod_{i=1}^{k-1} \mathrm{d}P_{0,n}(\eta_i)\, \delta(\widehat{\eta} - \widehat{f}_{\Theta_1}(\eta_1^f \dots \eta_t^f, \eta_{p+1}, \dots, \eta_{k-1})) \ . \tag{7.69}$$

The expression (7.63) of the correlation function C_n can similarly be rewritten as:

$$C_n = \sum_{p=0}^{l+1} \frac{\binom{l+1}{p}\psi(p)\widehat{y}^{-p}}{\sum_{p'=0}^{l+1}\binom{l+1}{p'}\psi(p')\widehat{y}^{-p'}} \int \prod_{i=1}^{p} \mathrm{d}\widehat{P}_{1,n}(\widehat{\eta}_i) \prod_{i=p+1}^{l+1} \mathrm{d}\widehat{P}_{0,n}(\widehat{\eta}_i)\, m(\widehat{\eta}_1, \dots, \widehat{\eta}_{l+1}) \ . \tag{7.70}$$

As already mentioned in the previous section, a further symmetry constrains the distributions $P_{v,n}$; let $T_n(\eta)$ the distribution of η in a broadcast process wich is not conditioned on the value of the root, i.e.:

$$P_n(\eta) = \sum_{v=(\sigma,w)} \eta_*(w) P_{v,n}(\eta) \ , \tag{7.71}$$

where η_* is normalized in such a way that $\eta_*(0) + \eta_*(1) = 1/2$. Applying Bayes theorem to express the joint law of the variable at the root and those at the

boundary one obtains [16]

$$P_{v,n}(\eta) = \frac{\eta(v)}{\eta_*(w)} P_n(\eta) \ . \tag{7.72}$$

This yields a relation between $P_{w,n} = P_{(+,w),n}$ for the two values of w, namely

$$P_{1,n}(\eta) = y \frac{\eta(+,1)}{\eta(+,0)} P_{0,n}(\eta) \ , \tag{7.73}$$

where we recall that $y = \eta_*(0)/\eta_*(1)$ was defined in Eq. (7.39). Moreover the spin-flip symmetry implies the invariance of P_n, i.e. $P_n(\eta) = P_n(\eta^f)$. This property, combined with (7.72), allows to relate $P_{w,n}$ in η and η^f through a change of density, namely

$$P_{0,n}(\eta^f) = \frac{\eta(-,0)}{\eta(+,0)} P_{0,n}(\eta) \ , \qquad P_{1,n}(\eta^f) = \frac{\eta(-,1)}{\eta(+,1)} P_{1,n}(\eta) \ . \tag{7.74}$$

These symmetry relations, as well as the similar ones that hold for $\widehat{P}_{w,n}$ modulo the replacement of y by $\widehat{y}$ in (7.73), will be particularly useful in the treatment of the large k limit presented in chapter 10. They imply a variety of identities between average observables, and in particular they can be used to rewrite the correlation function as

$$C_n = \sum_{p=0}^{l+1} \frac{\binom{l+1}{p} \psi(p) \widehat{y}^{-p}}{\sum_{p'=0}^{l+1} \binom{l+1}{p'} \psi(p') \widehat{y}^{-p'}} \int \prod_{i=1}^{p} \mathrm{d}\widehat{P}_{1,n}(\widehat{\eta}_i) \prod_{i=p+1}^{l+1} \mathrm{d}\widehat{P}_{0,n}(\widehat{\eta}_i) \, m(\widehat{\eta}_1, \ldots, \widehat{\eta}_{l+1})^2 \ , \tag{7.75}$$

which obviously shows that $C_n \geq 0$. This alternative form of C_n can be derived by first checking that

$$\sum_{p=0}^{l+1} \binom{l+1}{p} \psi(p) \widehat{y}^{-p} \int \prod_{i=1}^{p} \mathrm{d}\widehat{P}_{1,n}(\widehat{\eta}_i) \prod_{i=p+1}^{l+1} \mathrm{d}\widehat{P}_{0,n}(\widehat{\eta}_i) \, A(\widehat{\eta}_1, \ldots, \widehat{\eta}_{l+1}) \tag{7.76}$$

$$= \sum_{p=0}^{l+1} \binom{l+1}{p} \psi(p) \widehat{y}^{-p} \int \prod_{i=1}^{p} \mathrm{d}\widehat{P}_{1,n}(\widehat{\eta}_i) \prod_{i=p+1}^{l+1} \mathrm{d}\widehat{P}_{0,n}(\widehat{\eta}_i) \, \frac{m_-(\widehat{\eta}_1, \ldots, \widehat{\eta}_{l+1})}{m_+(\widehat{\eta}_1, \ldots, \widehat{\eta}_{l+1})} A(\widehat{\eta}_1^f, \ldots, \widehat{\eta}_{l+1}^f) \tag{7.77}$$

for an arbitrary function A which is invariant under the permutation of its arguments, and such that the integrals are well-defined, and then applying this identity with the test function $A = m(1-m)$.

7.4 Hard fields and the naive reconstruction

The tree reconstruction problem considered above asks whether the observation of $\underline{\sigma}_{B_n}$ gives some information on the value of the root σ_0, as quantified by the

correlation function C_n. Answering this question requires to solve the functional recursion relations (7.53,7.54 for μ_{Θ_0}, and (7.67-7.68) for μ_{Θ_1}). The strategy is based on the optimal inference algorithm, namely the computation of the posterior probability $\mu_\Theta(\sigma_i|\underline{\sigma}_{B_n})$ via the BP equations ((7.7) for μ_{Θ_0},(7.17) for μ_{Θ_1}). We shall now consider a more drastic question, namely whether $\underline{\sigma}_{B_n}$ allows to infer σ_0 with perfect certainty, i.e. is σ_i uniquely determined in the measure $\mu_\Theta(\cdot|\underline{\sigma}_{B_n})$, or in other words is σ_i constant in all the proper bicolorings of the graphs that take the values $\underline{\sigma}_{B_n}$ on the boundary vertices? This corresponds to a projection of the BP algorithm towards its Warning Propagation [35] version introduced in the section 4.1.6 that only keeps sure beliefs and discards the other ones. It is clear that if the answer to this new question is yes then the inferred value of σ_i is the correct one that was used in the broadcast, hence if reconstruction is possible in this strong sense of certain inference it is also possible in the definition introduced in the section 7.3.2. We call H_n the probability of this event. It turns out that H_n is much simpler to compute than C_n, with scalar recursions instead of functional ones, and that H_n is a lower bound for C_n; this last fact is quite intuitive, if $\underline{\sigma}_{B_n}$ implies the value of σ_0 it certainly conveys information about it.

The computation of H_n requires to project the BP equations on the WP equations, and to compute from the distribution over the BP messages (denoted $P_n(h)$ for μ_{Θ_0}, and $P_{w,n}(\eta)$ for μ_{Θ_1}) the probability of having a hard message. We will precise the definition of hard message in the case μ_{Θ_1}. To explain the computation of H_n let us first remark that σ_0 is implied by $\underline{\sigma}_{B_n}$ if and only if all the proper bicolorings of the tree that coincides with $\underline{\sigma}_{B_n}$ on the boundary take the same value at the root. By definition we only consider biased measures that do not strictly forbid any solution ($\omega_p > 0$ for $p \in \{1, \dots, k-1\}$ for μ_{Θ_0}, and $\psi(p) > 0$ for all p for μ_{Θ_1}), hence the certain determination of σ_0 can only arise from the bicoloring constraints acting on the spin variables. This observation can be turned into an algorithm, called the naive reconstruction procedure: consider all the hyperedges at the boundary, and declare them "forcing to the value σ" if their $k-1$ variables at distance n from the root are all equal to $-\sigma$, and "not forcing" otherwise. Now the variables at distance $n-1$ are assigned the value σ if at least one of their incident boundary hyperedge is forcing to this value (by construction of the broadcast process there cannot be conflicting forcings to $+$ and $-$ on the same variable), and a "white" value 0 if all the hyperedges are not forcing. This process can be iterated from the boundary towards the root, hyperedges being forcing if and only if $k-1$ among their variables have been assigned the same value $+1$ or -1. H_n is thus the probability that this successive forcing mechanism percolates from the boundary to the root, with at least one of its incident hyperedge forcing it.

7.4.1 Intra-clause bias

To incorporate naturally this computation in the one presented in the section 7.3.2 we decompose the field distributions as follows:

$$P_n(h) = H_n\,\delta(h-1)+(1-H_n)\,Q_n(h)\;,\quad \widehat{P}_n(u) = \widehat{H}_n\,\delta(u-1)+(1-\widehat{H}_n)\,\widehat{Q}_n(u)\;, \tag{7.78}$$

where Q_n and $\widehat{Q}_n$ are probability measures with no atom in 1. H_n is the weight of the "hard field" $h = 1$, that constrain completely the variable at the root and makes it a frozen variable under the boundary condition $\underline{\sigma}_{B_n}$. One also says that $h = 1$ is forcing the value of the root to $\sigma = 1$. H_n is precisely the probability of success of the naive reconstruction procedure. In the following we shall call Q_n and $\widehat{Q}_n$ the distributions of soft fields; note that they do not contain atoms in -1 (because a variable σ_i cannot be forced by $\underline{\sigma}_{B_n}$ to another value that it had in the broadcast), and that they enjoy the same Bayes symmetry (7.58) as the complete distributions P_n and $\widehat{P}_n$. Inserting this decomposition in (7.53,7.54), and considering the possible combinations of arguments of the functions $f_{\Theta_0}, \widehat{f}_{\Theta_0}$ in (7.7) that yields $h, u = 1$ one easily obtains the following evolution equations for H_n and $\widehat{H}_n$:

$$H_{n+1} = 1 - e^{-\alpha k \widehat{H}_n}\;,\qquad \widehat{H}_n = \widetilde{p}(-,\dots,-|+)H_n^{k-1}\;. \tag{7.79}$$

The first one expresses the fact that a spin is perfectly recovered as soon as one of its neighboring interactions forces it to its correct value, while the second one shows that this latter event happens when in the broadcast the $k-1$ variables adjacent to it have been given the same value, and that they all have been perfectly recovered.

Assembling these two equations we obtain a simple recursive equation on H_n,

$$H_{n+1} = 1 - e^{-\Gamma(\alpha,k,\{w_p\})H_n^{k-1}}\;,\qquad \Gamma(\alpha,k,\{w_p\}) = \alpha k\widetilde{p}(-,\dots,-|+)\;, \tag{7.80}$$

with the initial condition $H_0 = 1$. By a numerical inspection of the shape of this recursion function one easily realizes that for $k \geq 3$ the fixed point reached by H_n for $n \to \infty$ undergoes a discontinuous bifurcation from zero to a strictly positive value when Γ crosses a critical value $\Gamma_{\rm r} = \Gamma_{\rm r}(k)$. The latter can be determined by noting that at such a bifurcation the derivative of the recursion function must be equal to 1, hence that $\Gamma_{\rm r}$ and $H_{\rm r}$, the fixed point at the bifurcation, are solutions of

$$H_{\rm r} = 1 - e^{-\Gamma_{\rm r}H_{\rm r}^{k-1}}\;,\qquad 1 = (k-1)H_{\rm r}^{k-2}\Gamma_{\rm r}e^{-\Gamma_{\rm r}H_{\rm r}^{k-1}}\;. \tag{7.81}$$

One can close these two equations on a single one that determines $H_{\rm r}$: $1 = (k-1)\ln(1-H_{\rm r})(1-1/H_{\rm r})$, from which one obtain $\Gamma_{\rm r}(k) = -\ln(1-H_{\rm r})/(H_{\rm r}^{k-1})$,

and in terms of the original parameter α this "rigidity" transition occurs at

$$\alpha_r(k,\{w_p\}) = \frac{\Gamma_{\rm r}}{k\widetilde{p}(-,\ldots,-|+)} = \frac{1}{k}\Gamma_r(k)\frac{\sum_{p=1}^{k-1}\binom{k-1}{p}\omega_p}{\omega_1} = \frac{1}{k}\Gamma_r(k)\frac{\sum_{p=1}^{k-1}\binom{k}{p}\omega_p}{2\omega_1} \ . \tag{7.82}$$

We shall denote $\alpha_{r,u}$ the value of this threshold for the uniform case $\omega_1 = \cdots = \omega_{k-1} = 1$, i.e. when all proper bicolorings are weighted equally, in such a way that

$$\alpha_{r,u}(k) = \frac{1}{k}\Gamma_r(k)(2^{k-1}-1) \ . \tag{7.83}$$

From the intuitive interpretation of this computation as the analysis of a suboptimal reconstruction algorithm it is clear that α_r should be an upper bound on α_d, and indeed if $H = \lim_n H_n$ is strictly positive then P_n cannot converge to $\delta(h)$. The inequality $\alpha_d \leq \alpha_r$ is in general strict, i.e. there can be non-trivial solution of the equations (7.53,7.54) that do not contain any hard-field; this has been seen numerically in many problems, and proven rigorously for the graph q-coloring problem in the large q limit in [38, 40]. Moreover this probability H of perfect reconstruction is a lower bound on the correlation function C, as for any value of n the fraction of hard fields H_n is a lower bound on C_n:

$$C_n(\alpha,k) = \int {\rm d}P_n(h)h = H_n + (1-H_n)\int {\rm d}Q_n(h)h \tag{7.84}$$

$$\text{and} \quad \int {\rm d}Q_n(h)h = \int {\rm d}Q_n(h)h^2 \geq 0 \ , \tag{7.85}$$

because of the consequence stated in (7.59) of the Bayes symmetry, that is also enjoyed by Q_n. An illustration of the bound $H(\alpha,k) \leq C(\alpha,k)$ can be found for $k=5$ in the bottom panel of Fig. 9.2. We can see on this figure that for finite k this bound, as well as its consequence $\alpha_d(k) \leq \alpha_r(k)$, are not tight. There is an intermediate regime $\alpha_d < \alpha < \alpha_r$ where reconstruction is possible but naive reconstruction is not, all the relevant information on the value of the root is asymptotically contained in the soft fields distribution. Another view on this phenomenon is given in Fig. 7.4 where one sees the fraction H_n of hard fields fall to zero as n grows while C_n remains at a positive plateau value. Also when $\alpha > \alpha_r$ one can see on the bottom panel of Fig. 9.2 that $C > H$ for $k=5$, i.e. the soft fields do bring some additional information on the value of the root even in the presence of hard fields.

Let us also underline that among all the parameters $\{\omega_p\}$ that define the bias among proper bicolorings $\omega_1 = \omega_{k-1}$ plays a special role in the expression (7.82) of the rigidity threshold. Indeed hard fields are propagated along constraints that are "almost violated", in the sense that they contain a single variable of a given color. Penalizing such "almost monochromatic" hyperedges tends thus to avoid the percolation of frozen variables.

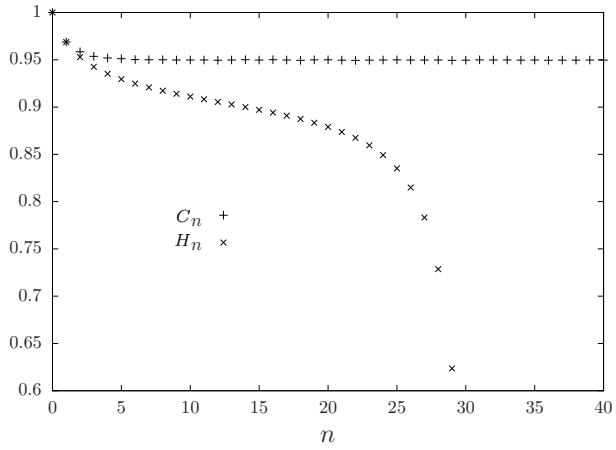

Figure 7.4: The evolution of C_n and H_n for $k = 5$, $\alpha = 10.4$, in the intermediate regime between α_d and α_r: the fraction H_n of hard fields falls to zero (the vertical range has been reduced for the sake of readability) while the correlation function C_n tends to a positive constant.

7.4.2 Bias with interactions at distance 1

To embed the analysis of this naive reconstruction algorithm into the formalism defined above for the measure μ_{Θ_1} we first introduce some terminology to classify the messages $\eta, \widehat{\eta}$; we will say that

- η is forcing to $\sigma = \pm 1$ iff $\eta(-\sigma, 0) = \eta(-\sigma, 1) = 0$, $\eta(\sigma, 0) > 0$ and $\eta(\sigma, 1) > 0$.
- η is non-forcing iff $\eta(\sigma, w) > 0$ for all σ and w.
- $\widehat{\eta}$ is forcing to $\sigma = \pm 1$ iff $\widehat{\eta}(\sigma, 1) = 1$, $\widehat{\eta}(\sigma, 0) = \widehat{\eta}(-\sigma, 0) = \widehat{\eta}(-\sigma, 1) = 0$; we write then $\widehat{\eta} = \widehat{\eta}^{\sigma}$.
- $\widehat{\eta}$ is non-forcing iff $\widehat{\eta}(+, 0) + \widehat{\eta}(+, 1) > 0$ and $\widehat{\eta}(-, 0) + \widehat{\eta}(-, 1) > 0$.

We will also use the term hard (resp. soft) field for the forcing (resp. non-forcing) BP messages. Inserting these definitions in the BP equations (7.17) one can check the combination rules argued for above: $\eta = f(\widehat{\eta}_1, \dots, \widehat{\eta}_l)$ is forcing to σ iff at least one $\widehat{\eta}_i$ is forcing to σ and none forcing to $-\sigma$, η is non-forcing otherwise. Similarly $\widehat{\eta} = g(\eta_1, \dots, \eta_{k-1})$ is forcing to σ iff all the η_i are forcing to $-\sigma$, and non-forcing otherwise.

We decompose now the distributions $P_{w,n}, \widehat{P}_{w,n}$ between the contributions of the hard and of the soft fields, defining

$$P_{w,n}(\eta) = h_{w,n}\, R_{w,n}(\eta) + (1 - h_{w,n})\, Q_{w,n}(\eta) \ , \tag{7.86}$$

$$\widehat{P}_{w,n}(\widehat{\eta}) = \widehat{h}_{w,n}\, \delta(\widehat{\eta} - \widehat{\eta}^{+}) + (1 - \widehat{h}_{w,n})\, \widehat{Q}_{w,n}(\widehat{\eta}) \ , \tag{7.87}$$

where $h, \widehat{h} \in [0,1]$ are the total weights of hard fields in the corresponding distributions, the R are normalized distributions on η's forcing to $+$, and Q and $\widehat{Q}$ are probability laws supported on non-forcing messages. By construction there are no messages forcing to $-$ in $P_{w,n} = P_{(+,w),n}$.

Inserting these decompositions in the recursion equations (7.67-7.69) we see that the evolution of the hard fields weights decouple; in particular from (7.68) we obtain $\widehat{h}_{1,n+1} = (h_{0,n})^{k-1}$ and from (7.69) $\widehat{h}_{0,n+1} = 0$, we shall thus write more simply $\widehat{h}_n$ instead of $\widehat{h}_{1,n}$. The equation (7.67) yields

$$h_{w,n+1} = 1 - \frac{\sum_{p=0}^{l} \binom{l}{p} \psi(p+w) \widehat{y}^{-p} (1-\widehat{h}_{n+1})^p}{\sum_{p=0}^{l} \binom{l}{p} \psi(p+w) \widehat{y}^{-p}} \ , \tag{7.88}$$

the recursion can thus be closed on $h_{0,n}$ and $\widehat{h}_n$, and solved starting from the initial condition $h_{0,0} = 1$. Finally H_n can be read off from the expression (7.70) of C_n by isolating the contribution with at least one forcing field $\widehat{\eta}$ around the root, which gives

$$H_n = 1 - \frac{\sum_{p=0}^{l+1} \binom{l+1}{p} \psi(p) \widehat{y}^{-p} (1-\widehat{h}_n)^p}{\sum_{p=0}^{l+1} \binom{l+1}{p} \psi(p) \widehat{y}^{-p}} \ . \tag{7.89}$$

Depending on the choice of the parameters (l, k, ψ) the sequence $h_{0,n}$ (or equivalently H_n) either decays to 0 or to a strictly positive fixed point. The so-called rigidity threshold $l_{\mathrm{r}}(k, \psi)$ introduced in the previous section separates these two behaviors: $H_n \to 0$ when $l < l_{\mathrm{r}}$ whereas it remains strictly positive in the large n limit for $l \geq l_{\mathrm{r}}$. In this latter case there is a positive probability for the observation of a far away boundary to completely determine the root (the naive reconstruction problem is solvable), hence there is certainly information about the value of the root (the usual reconstruction problem is also solvable).

The correlation function C_n can also be decomposed from (7.70) as the sum of H_n and a soft contribution:

$$C_n = H_n + \sum_{p=0}^{l+1} \frac{\binom{l+1}{p} \psi(p) \widehat{y}^{-p} (1-\widehat{h}_n)^p}{\sum_{p'=0}^{l+1} \binom{l+1}{p'} \psi(p') \widehat{y}^{-p'}} \int \prod_{i=1}^{p} \mathrm{d}\widehat{Q}_{1,n}(\widehat{\eta}_i) \prod_{i=p+1}^{l+1} \mathrm{d}\widehat{Q}_{0,n}(\widehat{\eta}_i)\, m(\widehat{\eta}_1, \dots, \widehat{\eta}_{l+1}) \ . \tag{7.90}$$

Exploiting the symmetry relations (7.73,7.74) one can rewrite the second term in this equation with the integrand squared, exactly as we did in the expression (7.75) of C_n, which proves the bound $C_n \geq H_n$ and confirms the intuition that the reconstruction problem is solvable if the naive reconstruction is.

7.5 The distribution of soft-fields

7.5.1 Uniform measure

We will now derive the recursion equations on the distribution of the soft fields Q_n and $\widehat{Q}_n$ introduced in (7.78), that complete the equations (7.79) for the evolution of the weights H_n and $\widehat{H}_n$ of hard fields. Both are obtained by plugging the decomposition (7.78) into the recursion equations (7.53,7.54) on P_n and $\widehat{P}_n$. We will focus on the uniform distriution (obtained by setting $\omega_0 = \omega_k = 0$, and $\omega_1 = \cdots = \omega_{k-1} = 1$ in equation (6.8)). The recursion equations that we will obtain in this section will be used in the large k analysis of the clustering threshold for the uniform measure, presented in [32] (see chapter 9).

As $f_{\Theta_0}(u_1, \ldots, u_l) = 1$ as soon as one of the arguments is equal to 1, i.e. as soon as one of the neighboring interactions forces the root variable, it is easy to see from (7.53) that the soft part of P_n arises from the combination of only soft u's, namely

$$Q_{n+1}(h) = \sum_{l=0}^{\infty} e^{-\alpha k(1-\widehat{H}_n)} \frac{(\alpha k(1-\widehat{H}_n))^l}{l!} \int \prod_{i=1}^{l} \mathrm{d}\widehat{Q}_n(u_i)\, \delta(h - f_{\Theta_0}(u_1, \ldots, u_l)) \; . \tag{7.91}$$

The treatment of (7.54) requires a little bit more work in order to put the resulting equation into the convenient form used in [32]. Denoting p the number of soft fields picked in the r.h.s. of (7.54), we first rewrite this equation as

$$\widehat{P}_n(u) = \sum_{p=0}^{k-1} \binom{k-1}{p} H_n^{k-1-p}(1-H_n)^p \sum_{\sigma_1 \ldots \sigma_{k-1}} p(\sigma_1 \ldots \sigma_{k-1}|+) \int \mathrm{d}Q_n(h_1) \ldots \mathrm{d}Q_n(h_p)\, \delta(u - g(\sigma_1 h_1 \ldots \sigma_p h_p, \sigma_{p+1} \ldots, \sigma_{k-1})) \; . \tag{7.92}$$

As explained above the term $\delta(u-1)$ arises solely from the term $p = 0$, $(\sigma_1 \ldots \sigma_{k-1}) = (-, \ldots, -)$. Moreover one realizes by inspection of the expression of $\widehat{f}_{\Theta_0}$ in (7.10) that $u = 0$ as soon as among the hard fields $(\sigma_{p+1} \ldots, \sigma_{k-1})$ at least one is positive and at least one is negative; this expresses the fact that in such a situation the bicoloring condition is satisfied whatever the value of the spin at the root. The initial condition for this evolution of the soft fields distribution turns out to be $Q_1(h) = \delta(h)$, as can be easily realized by an explicit computation of $\widehat{P}_0(u)$ and $P_1(h)$ starting from $P_0(h) = \delta(h-1)$.

Combining the expression (7.44) of the broadcasting probability $\widetilde{p}(\sigma_1 \ldots \sigma_{k-1}|+)$ with the examination of the possibles outcomes of (7.10), one obtains after a

short computation the following expression for the soft part of $\widehat{P}_n$,

$$(1-\widehat{H}_n)\widehat{Q}_n(u) = \frac{2^{k-1}}{2^{k-1}-1}\left[1+(1-H_n)^{k-1}-2\left(1-\frac{H_n}{2}\right)^{k-1}\right]\delta(u)$$
$$+\sum_{p=1}^{k-2}\binom{k-1}{p}(1-H_n)^p H_n^{k-1-p}\int\prod_{i=1}^{p}\mathrm{d}Q_n(h_i)\frac{1}{2^{k-1}-1}\sum_{\sigma_1,\dots,\sigma_p}\delta(u-g_p(\sigma_1 h_1,\dots,\sigma_p h_p))$$
$$+\sum_{p=1}^{k-2}\binom{k-1}{p}(1-H_n)^p H_n^{k-1-p}\int\prod_{i=1}^{p}\mathrm{d}Q_n(h_i)\frac{1}{2^{k-1}-1}\sum_{\sigma_1,\dots,\sigma_p}{}'\,\delta(u+g_p(\sigma_1 h_1,\dots,\sigma_p h_p))$$
$$+\frac{(1-H_n)^{k-1}}{2^{k-1}-1}\sum_{\sigma_1,\dots,\sigma_{k-1}}{}'\int\prod_{i=1}^{k-1}\mathrm{d}Q_n(h_i)\delta(u-g(\sigma_1 h_1,\dots,\sigma_{k-1}h_{k-1}))\ , \quad (7.93)$$

where the primed sums $\sum'$ exclude the configuration with all the spin arguments equal to $+1$, and where g_p is defined as

$$g_p(h_1,\dots,h_p) = -g(h_1,\dots,h_p,+1,\dots,+1) = \frac{\prod_{i=1}^{p}\frac{1+h_i}{2}}{2-\prod_{i=1}^{p}\frac{1+h_i}{2}}\ . \quad (7.94)$$

7.5.2 Bias with interaction at distance 1

We also give here the recursion equation for the soft fields distributions, obtained by inserting the decompositions (7.86,7.87) into (7.67-7.69):

$$Q_{w,n+1}(\eta) = \sum_{p=0}^{l}\frac{\binom{l}{p}\psi(p+w)\widehat{y}^{-p}(1-\widehat{h}_{n+1})^p}{\sum_{p'=0}^{l}\binom{l}{p'}\psi(p'+w)\widehat{y}^{-p'}(1-\widehat{h}_{n+1})^{p'}}\int\prod_{i=1}^{p}\mathrm{d}\widehat{Q}_{1,n+1}(\widehat{\eta}_i)$$
$$\times\prod_{i=p+1}^{l}\mathrm{d}\widehat{Q}_{0,n+1}(\widehat{\eta}_i)\,\delta(\eta-f_{\Theta_1}(\widehat{\eta}_1,\dots,\widehat{\eta}_l))\ , \quad (7.95)$$
$$\widehat{Q}_{1,n+1}(\widehat{\eta}) = \sum_{u=1}^{k-1}\frac{\binom{k-1}{u}(h_{0,n})^{k-1-u}(1-h_{0,n})^u}{1-(h_{0,n})^{k-1}}\int\prod_{i=1}^{u}\mathrm{d}Q_{0,n}(\eta_i)$$
$$\times\prod_{i=u+1}^{k-1}\mathrm{d}R_{0,n}(\eta_i)\,\delta(\widehat{\eta}-\widehat{f}_{\Theta_1}(\eta_1^f,\dots,\eta_{k-1}^f)) \quad (7.96)$$
$$\widehat{Q}_{0,n+1}(\widehat{\eta}) = \widehat{P}_{0,n+1}(\widehat{\eta})\ . \quad (7.97)$$

These recursive equations will be used in chapter 10, that deals with the asymptotic expansion of the clustering threshold for biased measures, and that gathers the results obtained in the paper [33]. The last equation comes from the absence of hard fields in $\widehat{P}_0$, one can thus take the expression (7.69) and insert in

its right hand side the decomposition (7.86) for P_1 and P_0 to have an equation involving only the soft fields distributions; this is relatively cumbersome notationally in general, we shall only write the corresponding equation in a special case in chapter 10.

7.6 1RSB formalism

We now introduce the 1RSB formalism, and explain the connection between the reconstruction problem on trees and the 1RSB cavity equations at $\mathcal{X} = 1$. The hypothesis underlying the RS cavity method must break down when the density of interactions per variable α becomes too large; a first hint of this phenomenon, called Replica Symmetry Breaking (RSB), is the negativity of the RS entropy at large enough α, which is impossible for a system with discrete degrees of freedom. In chapter 2, we have introduced the RSB phenomenon, and related it to the clustering of the solution set (see sections 2.3.1 and 2.3.2). As a matter of fact RSB can occur before $\alpha^{s=0}$; increasing α above a certain threshold causes the appearance of the long-range point-to-set correlation between distant variables under the measure μ, which contradicts the RS hypothesis. The point-to-set correlation function has been defined in equation (2.15), and has been estimated in the section 7.3.2 in the framework of the tree reconstruction problem. When this long-range correlation appears it becomes necessary to use more refined versions of the cavity method, that are able to deal with this RSB phenomenon [15]. At its first non-trivial level, called 1RSB for one step of RSB, the cavity method postulates the existence of a partition of the configuration space $\{-1, 1\}^N$ into "pure states", or clusters, such that the restriction of the measure μ to a pure state has good decorrelation properties. This restricted measure can then be treated as the full measure in the RS cavity method, i.e. with BP equations to describe its marginal probabilities, and the Bethe free-entropy to compute its partition function.

To be more quantitative let us index with c the partition of the configuration space into clusters, and denote $Z(c)$ the contribution to the partition function of the c-th cluster, as it was done in the section 2.3.2 (see equation (2.16)). We also denote $\{\widehat{\eta}^c_{a\to i}, \eta^c_{i\to a}\}$ the solution of the BP equations that describe it. The 1RSB cavity method aims at computing the potential (2.20) that we recall here:

$$\Phi_1(\mathcal{X}) = \lim_{N\to\infty} \frac{1}{N} \ln \left(\sum_c (Z(c))^{\mathcal{X}} \right), \tag{7.98}$$

where the Parisi parameter $\mathcal{X}$ allows to weight differently the various pure states, according to their relative weights. This quantity contains precious informations about the pure-state decomposition; suppose indeed that, at the leading exponential order, there are $e^{N\Sigma(\phi)}$ pure states c with $Z(c) = e^{N\phi}$ (again neglecting sub-exponential corrections). The complexity $\Sigma(\phi)$ plays thus the role of an entropy density, with pure states replacing usual configurations, and captures the RSB phenomenon quantitatively. The potential $\Phi_1(\mathcal{X})$ and the complexity $\Sigma(\phi)$ are Legendre transforms of each other [60]; evaluating (7.98) via the

Laplace method yields indeed

$$\Phi_1(\mathcal{X}) = \sup_{\phi} \left[\Sigma(\phi) + \mathcal{X}\phi\right] , \tag{7.99}$$

which can be inverted in terms of the conjugated parameter as

$$\Sigma(\mathcal{X}) = \Phi_1(\mathcal{X}) - \mathcal{X}\frac{\mathrm{d}}{\mathrm{d}\mathcal{X}}\Phi_1(\mathcal{X}) . \tag{7.100}$$

7.6.1 1RSB cavity equations

In order to compute Φ_1 one introduces, for a given sample and a given edge (i,a) of the factor graph, two distributions $P_{i\to a}$ and $\widehat{P}_{a\to i}$, that encode the laws of $h^c_{i\to a}$ and $u^c_{a\to i}$ when the pure state c is chosen randomly with a probability proportional to $Z(c)^{\mathcal{X}}$. These distributions are found to obey self-consistent equations of the form

$$P_{i\to a} = F_\Theta(\{\widehat{P}_{b\to i}\}_{b\in\partial i\setminus a}) , \qquad \widehat{P}_{a\to i} = \widehat{F}_\Theta(\{P_{j\to a}\}_{j\in\partial a\setminus i}) , \tag{7.101}$$

where $P = F_\Theta(\widehat{P}_1,\dots,\widehat{P}_d)$ is a shorthand for

$$P(\eta) = \frac{1}{z_{1,\Theta}(\widehat{P}_1,\dots,\widehat{P}_d)} \int \left(\prod_{i=1}^{d} \mathrm{d}\widehat{P}_i(\widehat{\eta}_i)\right) \delta(\eta - f_\Theta(\widehat{\eta}_1,\dots,\widehat{\eta}_d))\, z_{0,\Theta}(\widehat{\eta}_1,\dots,\widehat{\eta}_d)^{\mathcal{X}} , \tag{7.102}$$

and $\widehat{P} = \widehat{F}_\Theta(P_1,\dots,P_{k-1})$ means

$$\widehat{P}(\widehat{\eta}) = \frac{1}{\widehat{z}_{1,\Theta}(P_1,\dots,P_{k-1})} \int \left(\prod_{i=1}^{k-1} \mathrm{d}P_i(\eta_i)\right) \delta(\widehat{\eta} - \widehat{f}_\Theta(\eta_1,\dots,\eta_{k-1}))\, \widehat{z}_{0,\Theta}(\eta_1,\dots,\eta_{k-1})^{\mathcal{X}} ; \tag{7.103}$$

the functions f_Θ, $z_{0,\Theta}$, and $\widehat{f}_\Theta$, $\widehat{z}_{,\Theta}$ were defined in Eqs. (7.8 -7.11) for the measure μ_{Θ_0}, and (7.17) for the measure μ_{Θ_1}. The factors $z_{1,\Theta}$ and $\widehat{z}_{1,\Theta}$ ensure the normalization of the distributions $P(\eta)$ and $\widehat{P}(\widehat{\eta})$.

In order to deal with random hypergraphs one introduces the probability distributions over the 1RSB messages $\mathcal{P}^{1RSB}(P)$ and $\widehat{\mathcal{P}}^{1RSB}(\widehat{P})$ that obeys the consistency relations similar to (7.28,7.29),

$$\mathcal{P}^{1RSB}(P) = \sum_{d=0}^{\infty} p_d \int \left(\prod_{i=1}^{d} \mathrm{d}\widehat{\mathcal{P}}^{1RSB}(\widehat{P}_i)\right) \delta[P - F_\Theta(\widehat{P}_1,\dots,\widehat{P}_d)] , \tag{7.104}$$

$$\widehat{\mathcal{P}}^{1RSB}(\widehat{P}) = \int \left(\prod_{i=1}^{k-1} \mathrm{d}\mathcal{P}^{1RSB}(P_i)\right) \delta[\widehat{P} - \widehat{F}_\Theta(P_1,\dots,P_{k-1})] .$$

The 1RSB potential for typical random hypergraphs can then be computed from

the solution of these equations as

$$\Phi_{1,\Theta}(\mathcal{X}) = \sum_{d=0}^{\infty} p_d \int \left(\prod_{i=1}^{d} \mathrm{d}\widehat{P}_i \widehat{\mathcal{P}}^{1RSB}(\widehat{P}_i) \right) \ln \mathcal{Z}^{\mathrm{v}}_{1,\Theta}(\widehat{P}_1, \ldots, \widehat{P}_d) \tag{7.105}$$

$$+ \alpha \int \left(\prod_{i=1}^{k} \mathrm{d}P_i \mathcal{P}^{1RSB}(P_i) \right) \ln \mathcal{Z}^{\mathrm{c}}_{1,\Theta}(P_1, \ldots, P_k)$$

$$- \alpha k \int \mathrm{d}P \mathrm{d}\widehat{P} \mathcal{P}^{1RSB}(P) \widehat{\mathcal{P}}^{1RSB}(\widehat{P}) \ln \mathcal{Z}^{\mathrm{e}}_{1,\Theta}(P, \widehat{P}) \ , \tag{7.106}$$

with:

$$\mathcal{Z}^{\mathrm{v}}_{1,\Theta}(\widehat{P}_1, \ldots, \widehat{P}_d) = \int \left(\prod_{i=1}^{d} \mathrm{d}\widehat{P}_i(\widehat{\eta}_i) \right) \left(\mathcal{Z}^{\mathrm{v}}_{0,\Theta}(\widehat{\eta}_1, \ldots, \widehat{\eta}_d) \right)^{\mathcal{X}} \ , \tag{7.107}$$

$$\mathcal{Z}^{\mathrm{c}}_{1,\Theta}(P_1, \ldots, P_k) = \int \left(\prod_{i=1}^{k} \mathrm{d}P_i(\eta_i) \right) \left(\mathcal{Z}^{\mathrm{c}}_{0,\Theta}(\eta_1, \ldots, \eta_k) \right)^{\mathcal{X}} \ , \tag{7.108}$$

$$\mathcal{Z}^{\mathrm{e}}_{1,\Theta}(P, \widehat{P}) = \int \mathrm{d}P(\eta) \mathrm{d}\widehat{P}(\widehat{\eta}) \left(\mathcal{Z}^{\mathrm{e}}_{0,\Theta}(\eta, \widehat{\eta}) \right)^{\mathcal{X}} \ . \tag{7.109}$$

The expressions of $\mathcal{Z}^{\mathrm{v}}_{0,\Theta}, \mathcal{Z}^{\mathrm{c}}_{0,\Theta}, \mathcal{Z}^{\mathrm{e}}_{0,\Theta}$ have been specified in (7.13-7.15) for μ_{Θ_0}, and in (7.23-7.25) for μ_{Θ_1}. Finally the 1RSB prediction for the free-entropy is given in the equation (2.27) that we recall here:

$$\Phi^{1RSB}_{\Theta} = \inf_{\mathcal{X} \in [0,1]} \frac{\Phi_{1,\Theta}(\mathcal{X})}{\mathcal{X}} \ . \tag{7.110}$$

Note that the 1RSB equations always admit the RS solution as a special case, when the distributions P in the support of $\mathcal{P}^{1RSB}$ are Dirac measures. In most models this trivial solution of the 1RSB equations is the only one at small values of α; then $\Phi_1(\mathcal{X}) = \mathcal{X}\Phi^{RS}$, and the thermodynamic prediction of the RS and 1RSB versions of the cavity method coincides. Increasing the number of constraints of the system non-trivial solution of the 1RSB equations can appear; the dynamic (clustering) threshold $\alpha_d(k, \Theta)$ is defined as the smallest value of α for which the 1RSB equations with $\mathcal{X} = 1$ admit a solution distinct from the RS one. A further distinction has then to be made: if the associated complexity $\Sigma(\mathcal{X} = 1)$ is positive the extremum in (7.110) is reached for $\mathcal{X} = 1$ and $\Phi^{1RSB} = \Phi^{RS}$. In such a "dynamic 1RSB" situation the typical configurations of the Gibbs measure are supported on an exponentially large number of pure states, in such a way that the total free-entropy (or any correlation functions between a finite number of variables) is unable to detect the difference with a RS situation. On the contrary when $\Sigma(\mathcal{X} = 1) < 0$ the extremum in (7.110) selects a non-trivial value $\mathcal{X}_{\mathrm{s}} < 1$ of the Parisi parameter, the Gibbs measure condenses on a sub-exponential number of clusters, and correlations between finite sets of variables unveil the RSB phenomenon. One calls condensation threhold α_c the smallest value of α for which a solution of the 1RSB equations with $\Sigma(\mathcal{X} =$

$1) < 0$ exists, which corresponds to a point of non-analyticity of the free-entropy density. We will give in section 7.6.3 an explicit expression of the complexity $\Sigma(\mathcal{X} = 1)$ for the biased measure with intra-clause bias μ_{Θ_0}.

7.6.2 Simplifications for $\mathcal{X} = 1$

The complete 1RSB equations have a rather intricate structure, as they are self-consistent equations for probability distributions over probability distributions, which make in particular their numerical resolution rather cumbersome. Fortunately for special values of the parameter $\mathcal{X}$ (i.e. $\mathcal{X} = 0$ and $\mathcal{X} = 1$) they can be largely simplified. We shall sketch here this simplification procedure for $\mathcal{X} = 1$, which as explained above is the important one for the determination of the dynamic and condensation phase transitions; for further details the reader is referred to [59] where the simplifications are explained in more details and in a general setting. We will see that the 1RSB equations at $\mathcal{X} = 1$ coincide with the recursive distributional equations obtained for the tree-reconstruction problem, hence showing that the threshold for solvability and the clustering threshold are equal. We recall that in our settings the solution of the RS equations (7.28,7.29) is a trivial delta function $\mathcal{P}^{RS}(\eta) = \delta[\eta - \eta_*], \widehat{\mathcal{P}}^{RS}(\widehat{\eta}) = \delta[\widehat{\eta} - \widehat{\eta}_*]$. For the measure with a bias that factorises on the clauses (see equation (7.1)), one has seen in section 7.2.2 that $\eta_*(\sigma) = 1/2, \widehat{\eta}_*(\sigma) = 1/2$. For the biased measure with interaction at distance 1, for the k-regular $l + 1$-uniform graph ensemble, $\eta_*(\sigma), \widehat{\eta}_*(\sigma)$ is solution of the equations (7.36-7.38). This will allow for additional simplifications compared to the general setting described in [59].

The crucial technical property of the equations that opens the door to simplifications at $\mathcal{X} = 1$ is the fact that, for this value, the normalization constant $z_{1,\Theta}$ in (7.102) does not depend on the whole distributions $\widehat{P}_1, \ldots, \widehat{P}_d$, but only on their average values $\int \mathrm{d}\widehat{P}_i(\widehat{\eta}_i)\widehat{\eta}_i$ (a similar statement holds for $\widehat{z}_{1,\Theta}$ in (7.103)). More precisely, let

$$\overline{\eta}(P) = \int \mathrm{d}P(\eta)\eta, \quad \widehat{\overline{\eta}}(\widehat{P}) = \int \mathrm{d}\widehat{P}(\widehat{\eta})\widehat{\eta}. \tag{7.111}$$

Then one has:

$$z_{1,\Theta}(\widehat{P}_1, \ldots, \widehat{P}_d) = z_{0,\Theta}(\widehat{\overline{\eta}}(\widehat{P}_1), \ldots, \widehat{\overline{\eta}}(\widehat{P}_d)) \tag{7.112}$$

and

$$\widehat{z}_{1,\Theta}(P_1, \ldots, P_{k-1}) = \widehat{z}_{0,\Theta}(\overline{\eta}(P_1), \ldots, \overline{\eta}(P_{k-1})). \tag{7.113}$$

Furthermore, one can check that the random variable $\overline{\eta}$ (resp. $\widehat{\overline{\eta}}$) induced by the definition (7.111) with P (resp $\widehat{P}$) drawn from $\mathcal{P}^{1RSB}$ (resp. $\widehat{\mathcal{P}}^{1RSB}$) obey the RS self-consistency equations (7.28,7.29):

$$\overline{\eta} \stackrel{d}{=} f(\widehat{\overline{\eta}}_1, \ldots, \widehat{\overline{\eta}}_d); \quad \overline{\eta} \stackrel{d}{=} \widehat{f}(\overline{\eta}_1, \ldots, \overline{\eta}_{k-1}) \tag{7.114}$$

where d is drawn from the probability law r_d, and $\widehat{\overline{\eta}}_1, \ldots, \widehat{\overline{\eta}}_d$ are obtained by independenlty drawing $\widehat{P}_i$ from $\widehat{\mathcal{P}}^{1RSB}$, then constructing $\widehat{\overline{\eta}}_i = \widehat{\overline{\eta}}(\widehat{P}_i)$. Similarly

to obtain $\overline{\eta}_1, \ldots, \overline{\eta}_{k-1}$, one draws $P_1, \ldots, P_{k-1}$ i.i.d from $\mathcal{P}^{1RSB}$, then one sets $\overline{\eta}_i = \overline{\eta}(P_i)$.

Conditional on this average values F is thus a multilinear function of its arguments, the equations (7.104) can then be averaged and closed on the mean distributions Q and $\widehat{Q}$ defined as:

$$T(\eta) = \int \mathrm{d}\mathcal{P}^{1RSB}(P) P(\eta) \tag{7.115}$$

$$\widehat{T}(\widehat{\eta}) = \int \mathrm{d}\widehat{\mathcal{P}}^{1RSB}(\widehat{P}) \widehat{P}(\widehat{\eta}) \ , \tag{7.116}$$

These two quantities are solutions of

$$T(\eta) = \sum_{d=0}^{\infty} p_d \int \left(\prod_{i=1}^{d} \mathrm{d}\widehat{T}(\widehat{\eta}_i) \right) \delta(\eta - f(\widehat{\eta}_1, \ldots, \widehat{\eta}_d)) \frac{z_{0,\Theta}(\widehat{\eta}_1, \ldots, \widehat{\eta}_d)}{z_{0,\Theta}(\widehat{\eta}_*, \ldots, \widehat{\eta}_*)} \ , \tag{7.117}$$

$$\widehat{T}(\widehat{\eta}) = \int \left(\prod_{i=1}^{k-1} \mathrm{d}T(\eta_i) \right) \delta(\widehat{\eta} - g(\eta_1, \ldots, \eta_{k-1})) \frac{\widehat{z}_{0,\Theta}(\eta_1, \ldots, \eta_{k-1})}{\widehat{z}_{0,\Theta}(\eta_*, \ldots, \eta_*)} \ . \tag{7.118}$$

These equations are definitely simpler than the full 1RSB equations, as they bear on probability distributions instead of distributions of distributions; they have however one inconvenient feature, in particular for their numerical resolution, namely the reweighting terms $z_{0,\Theta}$ and $\widehat{z}_{0,\Theta}$ which prevents their direct interpretation as recursive distributional equations. To get around this difficulty we shall define the distributions

$$P_x(\eta) = \frac{\eta(x)}{\eta_*(x)} T(\eta), \qquad \widehat{P}_x(\widehat{\eta}) = \frac{\widehat{\eta}(x)}{\widehat{\eta}_*(x)} \widehat{T}(\widehat{\eta}) \tag{7.119}$$

where x is a generic variable ($x = \sigma$ for μ_{Θ_0}, and $x = v$ for μ_{Θ_1}). From the equality in distribution $\overline{\eta}(P) \stackrel{d}{=} \overline{\eta}$ where P is drawn from $\mathcal{P}^{1RSB}$ and $\overline{\eta}$ from $\mathcal{P}^{RS}$, and since $\mathcal{P}^{RS}(\overline{\eta}) = \delta[\overline{\eta} - \eta_*]$, we have actually the identity:

$$\int \mathrm{d}T(\eta) \eta = \int \mathrm{d}\mathcal{P}^{1RSB}(P) \overline{\eta}(P) = \int \mathrm{d}\widehat{P}^{RS}(\overline{\eta}) \overline{\eta} = \eta_* \tag{7.120}$$

Similarly one has

$$\int \mathrm{d}\widehat{T}(\widehat{\eta}) \widehat{\eta} = \widehat{\eta}_* \tag{7.121}$$

Thus one can check that the distributions $P_x, \widehat{P}_x$ are well-normalized, and that

$$T(\eta) = \sum_x \eta_*(x) P_x(\eta), \qquad \widehat{T}(\widehat{\eta}) = \sum_x \widehat{\eta}_*(x) \widehat{P}_x(\widehat{\eta}) \tag{7.122}$$

Inserting the definition of $P_x, \widehat{P}_x$ in (7.117,7.118) one obtain distributional equations for $P_x, \widehat{P}_x$. The particular form of these equation depends on the choice of the biased measure chosen. When one specifies to the biased measure μ_{Θ_0}, one recovers the equations (7.53,7.54). When one specifies to the biased measure μ_{Θ_1}, one recovers the equations (7.67-7.69). Therefore the 1RSB equations at $\mathcal{X} = 1$ coincide with the recursive distributional equations found for the tree-reconstruction problem.

7.6.3 Complexity $\Sigma(\mathcal{X}=1)$ and condensation threshold

In this section we give the expression of the complexity $\Sigma(\mathcal{X}=1)$ for the measure μ_{Θ_0}. First, one can check that $\Phi_1(\mathcal{X}=1)=\Phi^{RS}$ as given in (7.32). To compute the complexity at $\mathcal{X}=1$ from (7.100) we need to take the derivative with respect to $\mathcal{X}$ of Φ_1 from (7.98). Because of its variational character ((7.98) is stationary with respect to variations of $\mathcal{P}^{1RSB}$ and $\widehat{\mathcal{P}}^{1RSB}$ as long as the 1RSB cavity equations (7.104) are fulfilled) only the explicit dependency on $\mathcal{X}$ has to be differentiated. Doing the simplification at $\mathcal{X}=1$ yields then an expression in terms of P_+ and $\widehat{P}_+$:

$$\left.\frac{\mathrm{d}}{\mathrm{d}\mathcal{X}}\Phi_1(\mathcal{X})\right|_{\mathcal{X}=1} = \sum_{d=0}^{\infty} p_d \int \left(\prod_{i=1}^{d} \mathrm{d}u_i \widehat{P}_+(u_i)\right) \ln \mathcal{Z}_0^{\mathrm{v}}(u_1,\dots,u_d) \tag{7.123}$$

$$+\alpha \sum_{\sigma_1,\dots,\sigma_k} p(\sigma_1,\dots,\sigma_k) \int \left(\prod_{i=1}^{k} \mathrm{d}h_i P_+(h_i)\right) \ln \mathcal{Z}_0^{\mathrm{c}}(\sigma_1 h_1,\dots,\sigma_k h_k) \tag{7.124}$$

$$-\alpha k \int \mathrm{d}h\mathrm{d}u\, P_+(h)\widehat{P}_+(u) \ln \mathcal{Z}_0^{\mathrm{e}}(h,u)\ ,$$

with

$$p(\sigma_1,\dots,\sigma_k) = \frac{\omega(\sigma_1,\dots,\sigma_k)}{\sum_{\sigma'_1,\dots,\sigma'_k} \omega(\sigma'_1,\dots,\sigma'_k)}\ . \tag{7.125}$$

We recall that the condensation threshold α_c is the value above which the complexity at $\mathcal{X}=1$ becomes negative.

7.7 Kesten-Stigum bound

We shall discuss here a bound on α_d, known as the Kesten-Stigum [101, 13] transition in the context of the tree reconstruction problem, or as the de Almeida-Thouless [102] transition for mean-field spin-glasses, that is tight for continuous bifurcations and that in any case provide an easy to compute analytical upper bound on α_d (besides the bound $\alpha_d < \alpha^{s=0}$ and $\alpha_d \leq \alpha_r$ we already discussed). The computation we will present have been explained in details in [103] (in App. B) and in [104].

Let us recall that the 1RSB equations (7.104) always admit as a solution the RS solution. This solution is $\mathcal{P}^{1RSB}(P)=\delta[P-P_{\mathrm{triv}}]$, with $P_{\mathrm{triv}}(h)=\delta(h)$ for the uniform measure and for the measure μ_{Θ_0} with intra-clause bias (and similarly for $\widehat{\mathcal{P}}^{1RSB}(\widehat{P})$). For the biased measure μ_{Θ_1}, one has $P_{\mathrm{triv}}(\eta)=\delta(\eta-\eta_*)$ with $\eta_*, \widehat{\eta}_*$ are solutions of the equations (7.36-7.38). One way to test the existence of a non-trivial solution of the 1RSB equations is to investigate the local stability of the RS solution, that we describe for both cases.

7.7.1 Intra-clause bias

Suppose indeed that the distributions in the support of $\mathcal{P}^{1RSB}$ are close to P_{triv}, i.e. that they are supported on small values of h. One can then expand (7.102,7.102) and study the evolution of their average moments under the iterations of (7.104). More precisely one looks at the mean and variance:

$$M = \int \mathrm{d}P(h)h, \quad V = \int \mathrm{d}P(h)h^2 \tag{7.126}$$

with $P(h)$ a distribution in the support of $\mathcal{P}^{1RSB}$ that is close to $\delta(h)$. The global spin-flip symmetry imposes that the mean M remains zero. The first non-trivial moment is thus the variance. Expanding linearly the BP equations, one obtain at the first non-trivial order a relation between variances of the form

$$V = \theta^2 \sum_{i=1}^{d} \sum_{j=1}^{k-1} V_{ij} \tag{7.127}$$

that is obtained by innjecting the 1RSB equations (7.102,7.103) in the expression of the variance V. θ is the derivative of $\widehat{f}_{\Theta_0}(u_1, \ldots, u_{k-1})$ with respect to one of its arguments, evaluated on the trivial fixed-point. This relation is linear in the variances, one can take its average with respect to $\mathcal{P}^{1RSB}$, and obtain that the averaged variance will grow if and only if

$$\left(\sum_{d=0}^{\infty} r_d d \right) (k-1)\theta^2 > 1. \tag{7.128}$$

with $\sum_{d=0}^{\infty} r_d d = \alpha k$ for the ER ensemble. We thus obtain the Kesten-Stigum threshold α_{KS} above which the trival solution of the 1RSB equations is unstable (and there must then exist a non-trivial solution) as

$$\alpha_{KS} = \frac{1}{k(k-1)\theta^2} \quad \text{with} \tag{7.129}$$

$$\theta = \frac{\sum\limits_{\sigma_1,\ldots,\sigma_k} \omega(\sigma_1, \ldots, \sigma_k)\sigma_1\sigma_2}{\sum\limits_{\sigma_1,\ldots,\sigma_k} \omega(\sigma_1, \ldots, \sigma_k)} = \frac{\sum\limits_{p=0}^{k-2} \binom{k-2}{p} (\omega_p - 2\omega_{p+1} + \omega_{p+2})}{\sum\limits_{p=0}^{k} \binom{k}{p} \omega_p} \,. \tag{7.130}$$

7.7.2 Bias with interactions at distance 1

In this case, one defines the mean and variance as follows

$$M(v) = \int \mathrm{d}P(\eta) \left(\eta(v) - \eta_*(w)\right) \tag{7.131}$$

$$V(v, v') = \int \mathrm{d}P(\eta) \left(\eta(v) - \eta_*(w)\right) \left(\eta(v') - \eta_*(w')\right) \tag{7.132}$$

with P a distribution in the support of $\mathcal{P}^{1RSB}$ that is close to the RS distribution $\delta[\eta - \eta_*]$. As before, under symmetry considerations the mean $M(v)$ is zero. One can show that the variance obeys a similar linear relation

$$V(v_1, v_2) = \sum_{v_1', v_2'} J_{v_1, v_1'} J_{v_2, v_2'} \sum_{i=1}^{l} \sum_{j=1}^{k-1} V_{ij}, \tag{7.133}$$

where J is a 4×4 matrix that can be written as the product $J = M\widehat{M}$, with M and $\widehat{M}$ the Jacobian matrices obtained from the relations (7.18) and (7.19) evaluated at the RS solution:

$$M_{v,v'} = \left.\frac{\partial \eta(v)}{\widehat{\eta}_1(v')}\right|_*, \quad \widehat{M}_{v,v'} = \left.\frac{\partial \widehat{\eta}(v)}{\eta_1(v')}\right|_* . \tag{7.134}$$

The equation (7.133) can be averaged with respect to $\mathcal{P}^{1RSB}$ to obtain the stability criterion

$$l(k-1)\theta^2 < 1 \tag{7.135}$$

where θ is the largest eigenvalue of $J_{v,v'}$. We define l_{KS} through the relation

$$l_{\mathrm{KS}}(k-1)\theta^2 = 1. \tag{7.136}$$

The matrices M and $\widehat{M}$ take the explicit form:

$$M_{v,v'} = \frac{1}{\widehat{\eta}_*(0) D} \left(\mathbb{I}[\sigma = \sigma'] A(w + w') - \frac{\left(A(w) + \frac{A(w+1)}{\widehat{y}} \right) \left(A(w') + A(w'+1) \right)}{D} \right)$$

where $A(w)$ and D have the following expressions:

$$\begin{aligned} A(w) &= \sum_{p=0}^{l-1} \binom{l-1}{p} \psi(w+p) \widehat{y}^{-p} \\ D &= 2 \left(A(0) + A(1) + \frac{A(1) + A(2)}{\widehat{y}} \right) \end{aligned} \tag{7.137}$$

and

$$\begin{aligned} \widehat{M}_{(\sigma,0),(\sigma',0)} &= \frac{1}{\widehat{D}\eta_*(0)} \left(\widehat{A}\mathbb{I}[\sigma = \sigma'] + \widehat{B}\mathbb{I}[\sigma = -\sigma'] - \frac{\widehat{C}}{\widehat{y}\widehat{D}} \right) \\ \widehat{M}_{(\sigma,0),(\sigma',1)} &= \frac{1}{\widehat{D}\eta_*(0)} \left(\mathbb{I}[\sigma = -\sigma'] - \frac{1}{\widehat{y}\widehat{D}} \right) \\ \widehat{M}_{(\sigma,1),(\sigma',0)} &= \frac{1}{\widehat{D}\eta_*(0)} \left(\mathbb{I}[\sigma = -\sigma'] - \frac{\widehat{C}}{\widehat{D}} \right) \\ \widehat{M}_{(\sigma,1),(\sigma',1)} &= -\frac{1}{\widehat{D}^2\eta_*(0)} \end{aligned} \tag{7.138}$$

where $\widehat{A}, \widehat{B}, \widehat{C}, \widehat{D}$ have the following expressions:

$$\begin{aligned}
\widehat{A} &= 2^{k-2} - k + 1 + (k-2)y^{-1} \\
\widehat{B} &= 2^{k-2} - 2 \\
\widehat{C} &= 2^{k-1} - k + (k-2)y^{-1} \\
\widehat{D} &= 2(2^{k-1} - k + (k-1)y^{-1})
\end{aligned} \tag{7.139}$$

Chapter 8

Finite k results

In this chapter we report the numerical study of the clustering transition for the bicoloring on k-hypergraphs, for small values of k ($k = 4, 5, 6$). In Section 8.1 we describe how to solve numerically the recursive distributional equations obtained in chapter 7 (see (7.53,7.54) for μ_{Θ_0}, and (7.67-7.68) for μ_{Θ_1}) using population dynamics. Solving these equations allows us to determine the clustering transition. In the section 8.2 we give a review and extension of the numerical procedures to determine accurately the clustering threshold. In section 8.3 we present the phase diagrams predicted by the cavity method for the biased measure μ_{Θ_0} that introduces a bias that factorises on the clauses. We show that this biased measure allows us to increase the dynamic threshold with respect to the uniform one. In section 8.4 we present numerical simulations on finite size samples via Simulated Annealing. We show that with well-chosen parameters the biased measure μ_{Θ_0} allows us to improve the performances of Simulated Annealing. These three sections present the results obtained in [31]. In section 8.5 we present some numerical results obtained for the biased measure μ_{Θ_1} that introduces interactions between variables at distance 1, we show that this biased measure allows us to increase further the dynamical threshold compared to the simpler biased measure μ_{Θ_0}. This last section presents some of the results obtained in [33].

8.1 Numerical resolution for finite k

We give an explicit description of the numerical resolution of the distributional equations found in the previous chapter (see (7.53,7.54) for μ_{Θ_0}, and (7.67-7.69) for μ_{Θ_1}). For a given choice of the parameters (k, α) and of the bias function (ω or ψ), the model is either in a non-reconstructible, Replica Symmetric (RS) phase if the point-to-set correlation function C_n decays to 0 as $n \to \infty$, or in a reconstructible, Replica Symmetry Breaking (RSB) phase if C_n remains strictly positive in this large distance limit. We defined α_d as the threshold value above which one enters the RSB phase. For the k-uniform $l+1$-regular graph

ensemble $l_{\rm d}$ is the smallest integer value of l which leads to RSB. We recall that in this ensemble the density of constraints α and the degree l satisfy the relation $\alpha = (l+1)/k$. There are several upper bounds on α_d that can be computed analytically: if the entropy ((7.34) for μ_{Θ_0}, (7.42) for μ_{Θ_1}) computed in the RS ansatz is negative this is certainly an evidence for the RSB phenomenon. The existence of hard fields, i.e. the possibility of naive reconstruction (see section 7.4), implies the reconstructibility, hence the rigidity bound $\alpha_d \leq \alpha_r$. Finally the Kesten-Stigum bound $\alpha_d \leq \alpha_{\rm KS}$ follows from the instability of the trivial fixed point of the reconstruction recursive equations (see section 7.7). Nevertheless there are no lower bounds on α_d that are simple to compute, hence an explicit determination of this threshold requires a numerical resolution of the equations (7.53,7.54) and (7.67-7.69). This type of Recursive Distributional Equation (RDE) admits a natural numerical procedure to solve them, called population dynamics algorithm [105, 15], in which a probability distribution is approximated by the empirical distribution of a large sample of representative elements.

We explain it for the set of equations (7.53,7.54), the procedure is the same for the equations (7.67-7.69).

Suppose that at some step n one has an approximation of $P_n(h)$ written as

$$P_n(h) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \delta(h - h_i^{(n)}) \ , \tag{8.1}$$

where $\mathcal{N} \gg 1$ is the size of the population, that controls the numerical accuracy of the procedure (in the limit $\mathcal{N} \to \infty$ empirical distributions converge to the exact ones). One can then insert this form in the r.h.s. of (7.54) to obtain an approximation of $\widehat{P}_n(u)$ as

$$\widehat{P}_n(u) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \delta(u - u_i^{(n)}) \ , \tag{8.2}$$

where each of the representants $u_i^{(n)}$ has been constructed independently by drawing $\sigma_1, \dots, \sigma_{k-1}$ according to the law $\widetilde{p}(\sigma_1, \dots, \sigma_{k-1}|+)$ (defined in (7.44)), then $k-1$ indices $i_1, \dots, i_{k-1}$ uniformly at random in $\{1, \dots, \mathcal{N}\}$, and setting:

$$u_i^{(n)} = \widehat{f}_{\Theta_0}\left(\sigma_1 h_{i_1}^{(n)}, \dots, \sigma_{k-1} h_{i_{k-1}}^{(n)}\right) \ .$$

Similarly, $P_{n+1}(h)$ can be approximated by an empirical distribution of the form (8.1), where according to (7.53) each of the representants $h_i^{(n+1)}$ is obtained by drawing an integer l from the Poisson law $r_l = e^{-\alpha k} \frac{(\alpha k)^l}{l!}$, then l indices $i_1, \dots, i_l$ uniformly at random in $\{1, \dots, \mathcal{N}\}$, and taking

$$h_i^{(n+1)} = f_{\Theta_0}\left(u_{i_1}^{(n)}, \dots, u_{i_l}^{(n)}\right) \ .$$

The initial condition $P_0(h) = \delta(h-1)$ (corresponding to a perfect observation of the variables at distance $n = 0$ from the root) can obviously be represented

by a sample with all representants $h_i^{(0)}$ equal to 1, and at every iteration step observables can be estimated as empirical averages,

$$\int \mathrm{d}P_n(h)F(h) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} F(h_i^{(n)}) \tag{8.3}$$

for an arbitrary function F.

Iterating these two steps many times one converges to a fixed point solution of (7.53,7.54), which can either be the trivial one $P(h) = \delta(h)$, $\widehat{P}(u) = \delta(u)$, or a non-trivial solution. Figure 8.1 presents some numerical results obtained in [31] with this procedure, for the measure μ_{Θ_0}. The top panel displays the decay of the correlation function $C_n(\alpha, k, \{w_p\})$ as a function of the distance n for different values of α, for $k = 6$, and with the choice of parameters $\omega_0 = \omega_k = 0.005$, $\omega_1 = \omega_{k-1} = 0.92$, $w_2 = \cdots = \omega_{k-2} = 1$. When α is small this function decays to 0 as $n \to \infty$, signalling the impossibility of reconstruction; when α increases the decay gets slower and proceeds in two steps, with a longer and longer plateau at a strictly positive value developping as α gets closer to the transition; finally for $\alpha > \alpha_d$ the plateau lasts forever, the large n limit $C(\alpha, k, \{w_p\})$ is strictly positive. This quantity, that jumps discontinuously from 0 to a strictly positive value when α crosses α_d, is plotted as a function of α in the bottom panel of Fig. 8.1. To increase the numerical accuracy we estimated the limit by averaging the value of C_n when n is large enough to have reached its plateau behavior.

8.2 On the numerical determination of the dynamic transition

As explained in the previous section, the dynamic threshold α_d is the smallest value of α such that the 1RSB equations at $\mathcal{X} = 1$ admit a non-trivial solution (besides the RS trivial one in which all fields are equal to zero). We have seen how to solve numerically these equations, we now explain how to improve the accuracy when measuring α_d. Depending on the models the appearance of a non-trivial solution can occur either in a continuous or a discontinuous way. In the former case one has $\alpha_d = \alpha_{KS}$, the bifurcation occurs via the local instability of the trivial fixed point studied in Sec. 7.7, and α_d is thus known analytically. In the latter case $\alpha_d < \alpha_{KS}$, the birth of the non-trivial solution occurs non-perturbatively and cannot be detected from the properties of the trivial fixed point. The accurate numerical determination of α_d when the transition is discontinuous is a rather difficult task. It corresponds to study a bifurcation for a fixed-point equation of the form $P = F(P, \alpha)$, where P is a probability distribution and F a functional on this space, depending on the parameter α. We shall discuss in the parts 8.2.2 and 8.2.3 the different numerical strategies that can be followed to determine α_d, in particular one that, to the best of our knowledge, is new in this context. To explain these different methods

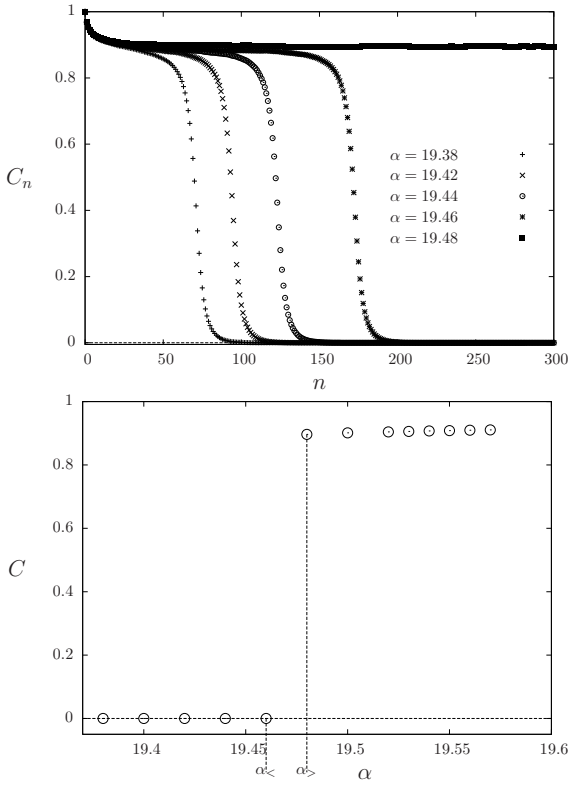

Figure 8.1: Exemple of a discontinuous dynamic transition for $k = 6$, $\omega_0 = \omega_k = 0.005$, $\omega_1 = \omega_{k-1} = 0.92$, and $\omega_2 = \cdots = \omega_{k-2} = 1$. Top: $C_n(\alpha, k, \{\omega_p\})$ versus the number of iterations n for different values of α. Bottom: $C(\alpha, k, \{\omega_p\})$ averaged over n after equilibration. The size of the population used is 10^6.

it is instructive to study first a much simpler case, in which the unknown P is replaced by a real number, that we shall instead denote x for clarity.

8.2.1 Scalar bifurcations

Let us consider a function $f(x, \alpha)$, smooth in its two real arguments, and the associated discrete dynamical system $x_{n+1} = f(x_n, \alpha)$, parametrized by α. We recall some basic facts in this setting: the stationary configurations of the dynamical system are the solutions $x_*(\alpha)$ of the fixed point equation $x = f(x, \alpha)$. Their (linear) stability is determined by the coefficient $\lambda(\alpha) = (\partial_x f)(x_*(\alpha), \alpha)$ (here and below we denote $(\partial_x f)$, $(\partial_\alpha f)$ and so on the partial derivatives of the function f); a fixed point $x_*(\alpha)$ is indeed stable under iterations if $|\lambda(\alpha)| < 1$, and unstable if $|\lambda(\alpha)| > 1$. We also recall the implicit function theorem: if (x_0, α_0) is a solution of $f(x_0, \alpha_0) = x_0$, and if $(\partial_x f)(x_0, \alpha_0) \neq 1$, then there is

a unique smooth function $x_*(\alpha)$ with $f(x_*(\alpha), \alpha) = x_*(\alpha)$ in a neighborhood of α_0, with $x_*(\alpha_0) = x_0$. Hence the bifurcations of the fixed point equation, i.e. the modifications in the number of solutions, or the singularities of these solutions, are associated to points where $(\partial_x f)(x(\alpha_0), \alpha_0) = 1$, in order for the implicit function theorem to be unapplicable. At these points the stability parameter $\lambda(\alpha)$ reaches its critical value 1.

To be more concrete we shall make the additional hypotheses that the order parameter x is restricted to non-negative values ($x \geq 0$), and that $f(0, \alpha) = 0$ for all α. Let us assume that this trivial fixed point, that exists for all α, is the unique solution for small enough values of α, and becomes non-unique when α exceeds a threshold α_d. The two simplest ways to implement these hypotheses are sketched in Fig. 8.2, corresponding to a continuous bifurcation on the top panel, a discontinuous one on the bottom. Let us state a series of simple facts on these two types of phase transitions, that will be enlightening when we turn to the functional case.

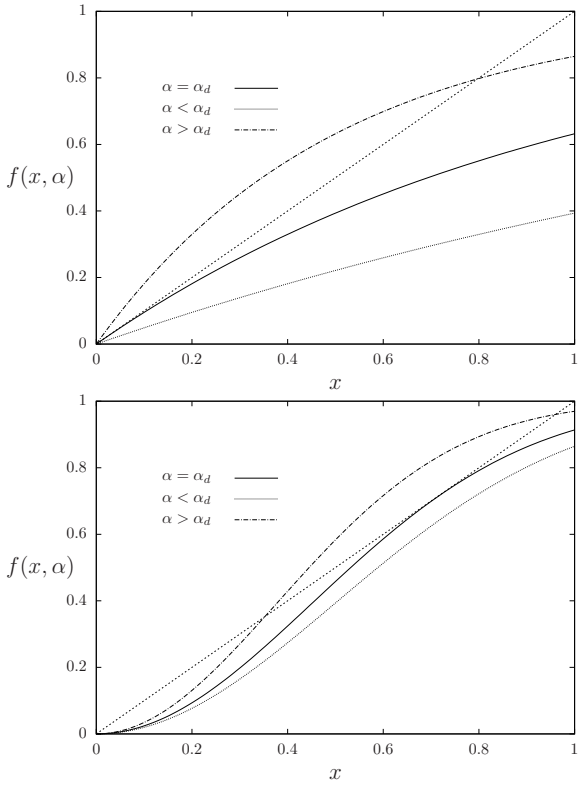

Figure 8.2: Top panel: example of a continuous bifurcation with $f(x, \alpha) = 1 - e^{-\alpha x}$, for which $\alpha_d = 1$. Bottom panel: example of a discontinuous bifurcation with $f(x, \alpha) = 1 - e^{-\alpha x^2}$, for which $\alpha_d = 2.45541$.

Consider first the continuous case illustrated on the top panel of Fig. 8.2. The bifurcation occurs at the critical parameter value α_d defined by $(\partial_x f)(0, \alpha_d) = 1$, the trivial fixed point being stable (resp. unstable) for $\alpha < \alpha_d$ (resp. $\alpha > \alpha_d$). For $\alpha > \alpha_d$ there exists a non-trivial branch of stable fixed points $x_*(\alpha) > 0$; in the neighborhood of the bifurcation the latter behaves as

$$x_*(\alpha) = K(\alpha - \alpha_d) + o((\alpha - \alpha_d)) \qquad \text{when} \quad \alpha \to \alpha_d^+ \ , \tag{8.4}$$

with $K = -2(\partial_{x\alpha} f)/(\partial_{xx} f)$, the derivatives being computed in $(0, \alpha_d)$. The stability parameter of the non-trivial solution, $\lambda(\alpha) = (\partial_x f)(x_*(\alpha), \alpha)$, reaches its marginal value 1 at the bifurcation as

$$\lambda(\alpha) = 1 - K'(\alpha - \alpha_d) + o(\alpha - \alpha_d) \qquad \text{when} \quad \alpha \to \alpha_d^+ \ , \tag{8.5}$$

with $K' = (\partial_{x\alpha} f)$.

Let us now turn to the discontinuous case (cf. the bottom panel of Fig. 8.2), and emphasize the main properties of the critical behavior of the bifurcation. The trivial fixed point is stable for all values of the parameter α; the bifurcation occurs at α_d with the abrupt appearance of a solution $x_d > 0$. These two quantities can be determined by solving the system of equations

$$\begin{cases} x_d = f(x_d, \alpha_d) \ , \\ 1 = (\partial_x f)(x_d, \alpha_d) \ . \end{cases} \tag{8.6}$$

For $\alpha > \alpha_d$ there are two branches of non-trivial solutions $x_-(\alpha) < x_d < x_+(\alpha)$ that emerge from x_d (see the bottom panel of Fig. 8.3); in the neighborhood of α_d they behave as

$$x_\pm(\alpha) = x_d \pm K\sqrt{\alpha - \alpha_d} + o(\sqrt{\alpha - \alpha_d}) \qquad \text{when} \quad \alpha \to \alpha_d^+ \ , \tag{8.7}$$

where the coefficient K can be computed from the expansion of f around the bifurcation point (explicitly, $K = \sqrt{-2(\partial_\alpha f)/(\partial_{xx} f)}$, the derivatives being computed in (x_d, α_d)). For $\alpha > \alpha_d$ $x_+(\alpha)$ (resp. $x_-(\alpha)$) is linearly stable (resp. unstable); the stability parameter $\lambda(\alpha) = (\partial_x f)(x_+(\alpha), \alpha)$ of the stable non-trivial branch reaches its critical value 1 at the bifurcation, with a critical exponent 1/2:

$$\lambda(\alpha) = 1 - K'\sqrt{\alpha - \alpha_d} + o(\sqrt{\alpha - \alpha_d}) \qquad \text{when} \quad \alpha \to \alpha_d^+ \ , \tag{8.8}$$

with $K' = \sqrt{-2(\partial_\alpha f)(\partial_{xx} f)}$. We present in the top panel of Fig. 8.3 the iterates $x_{n+1} = f(x_n, \alpha)$, for a few values of $\alpha < \alpha_d$, starting from an initial condition $x_{n=0} > x_d$. Their long time limit is of course 0, the only fixed point in this phase, but when $\alpha \to \alpha_d^-$ the decay is slower and slower, with a large number of iterations spent around a plateau value at x_d. More quantitatively one can define $n_*(\alpha)$ as the minimal n such that $x_n \leq x_d - \epsilon$, and obtain that

$$n_*(\alpha) \sim K''(\alpha_d - \alpha)^{-1/2} \qquad \text{when} \quad \alpha \to \alpha_d^- \ , \tag{8.9}$$

with $K'' = 2\pi/K'$, independently of $x^{(0)}$ and ϵ (as long as $0 < \epsilon < x_d$). Actually a whole scaling function describing the evolution of x_n around the plateau can be derived, see [63] for more details.

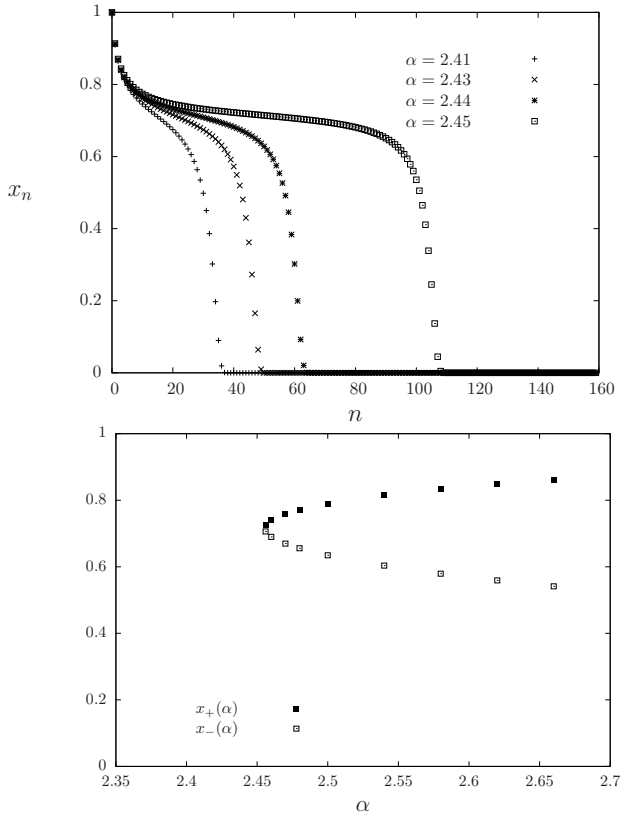

Figure 8.3: Properties of the iterates and fixed points for the discontinuous bifurcation undergone by the function $f(x, \alpha) = 1 - e^{-\alpha x^2}$. Top panel: x_n as a function of n for a few values of α close but strictly less than α_d. Bottom panel: the non-trivial solutions $x_\pm(\alpha)$ for $\alpha \geq \alpha_d$.

8.2.2 Discontinuous functional bifurcations

Let us now come back to our original goal, namely the determination of the dynamic threshold α_d above which appears a non-trivial solution of the 1RSB equations at $\mathcal{X} = 1$. As in the scalar case this transition can occur either in a continuous or in a discontinuous way; the former case was analytically dealt with in Sec. 7.7, we shall hence concentrate now on the discontinuous transitions. The numerical methods that we describe in this section are applicable to the uniform measure and the biased measures μ_{Θ_0} and μ_{Θ_1}. In practice we only implemented them on the biased measure μ_{Θ_0} during this Ph.D., but it could be interesting to adapt the method to the biased measure μ_{Θ_1} in order to obtain a numerical prediction of the clustering threshold with better accuracy.

The 1RSB equations (7.53,7.54) can be written abstractly as a functional

fixed point equation $P = F(P, \alpha)$; at variance with the scalar toy model discussed above they can only be solved approximately, for instance by the population dynamics numerical algorithm explained in Sec. 8.1. Some examples of typical numerical results that can be obtained in this way are presented in Fig. 8.1. On the top panel we plot the value of the point-to-set correlation C_n as a function of the number n of iterations, for a few values of α. One sees on this plot, reminiscent of the top panel of Fig. 8.3, the discontinuous birth of a non-trivial fixed point at α_d, with a longer and longer plateau in the low α phase as a precursor of the transition. On the bottom panel of Fig. 8.1 we present the asymptotic value of C reached for large n, for different values of α around the dynamic transition (corresponding to the bottom panel of Fig. 8.3), that jumps discontinuously from 0 when α crosses α_d.

It is not completely obvious how to extract a precise estimate of α_d from this kind of data. The simplest approach amounts to determine the curves C_n for several closely spaced values of α, and assess that $\alpha_d \in [\alpha_<, \alpha_>]$, where $\alpha_<$ is the largest value for which C_n drops to 0 at large n, $\alpha_>$ the smallest value for which a stable plateau is encountered. This determination suffers however from inaccuracies due to the finite number of α values one can investigate, the finite number of iterations one can perform (leading to an underestimation of $\alpha_>$) and to the finite size of the population that approximate the distribution P ($\alpha_<$ can thus be overestimated, finite size fluctuations having a destabilizing effect).

One can try to circumvent these difficulties by getting some inspiration from the much simpler scalar bifurcation studied above. We recall that the criticality at α_d showed up in three different ways: (i) $x_+(\alpha)$ exhibits a square root singularity when $\alpha \to \alpha_d^+$, see Eq. (8.7); (ii) the length of the plateau diverges when $\alpha \to \alpha_d^-$ with a critical exponent $-1/2$, cf. Eq. (8.9); (iii) the stability parameter $\lambda(\alpha)$ reaches 1 with a square root singularity when $\alpha \to \alpha_d^+$, as written in Eq. (8.8).

Assuming the same critical behavior to occur in the discontinuous functional bifurcation case (even if much more complicated behaviors could occur in infinite dimensional spaces) one can try to exploit these scaling laws in order to obtain more precise estimates of α_d. Point (i) translates into a square root singularity of the large n limit of C in the limit $\alpha \to \alpha_d^+$; this does not seem very useful to us, as it would involve a fit of $C(\alpha)$ in which both α_d and $C(\alpha_d)$ are unknowns. On the contrary points (ii) and (iii) yield simpler fits for the determination of α_d. The aspect (ii) is very easy to exploit: from the curves of the top panel of Fig. 8.1 one can deduce immediately a value $n_*(\alpha)$ for the number of iterations necessary to fall below the plateau (as in the scalar case one can define $n_*(\alpha)$ with any threshold strictly between 0 and the plateau value). According to (8.9) $n_*(\alpha)^{-2}$ should vanish linearly at α_d; this is indeed what we obtain with a rather good accuracy, see the top panel of Fig. 8.4. However one cannot reach in this way a very large number of iterations, the numerical rounding errors and finite population size fluctuations having the tendency to accumulate over time; this cutoff on n thus limits the accuracy of this determination of α_d. We have thus turned to the functional generalization of point (iii) above, namely

the computation of a stability parameter $\lambda(\alpha)$ for the stable non-trivial branch $\alpha > \alpha_d$, and the determination of α_d as the parameter for which λ reaches 1. This extrapolation is done using the scaling anticipated in the scalar case in (8.8), and is illustrated in the bottom panel of Fig. 8.4. The functional nature of the unknown in the fixed point equation makes the definition of λ more complicated than in the scalar case, where it was simply $\partial_x f$; we give detailed explainations on the numerical computation of $\lambda(\alpha)$ in the functional case in the next section. Before that let us emphasize that the square root behavior of λ around α_d, guessed from the scalar bifurcation, is in very good agreement with the numerical results obtained in the functional case when the dynamic transition is discontinuous (see the bottom panel of Fig. 8.4).

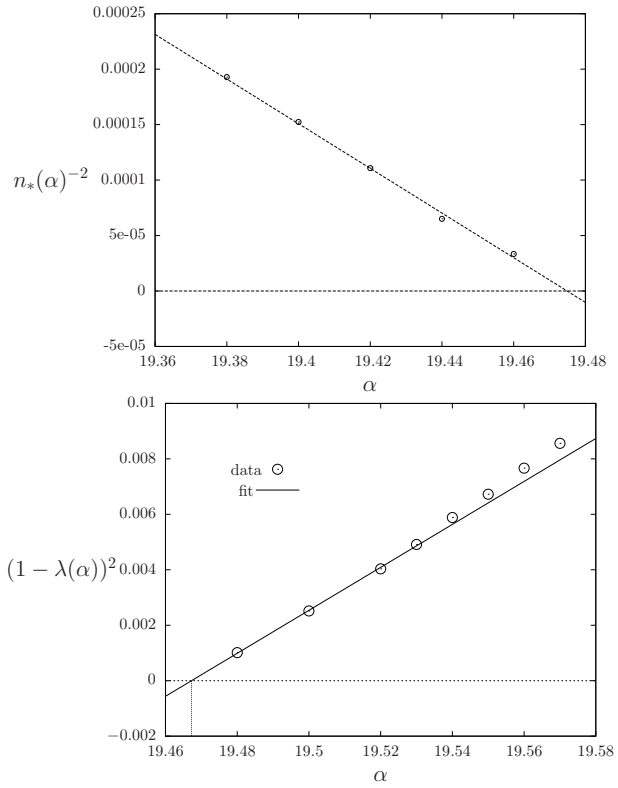

Figure 8.4: Study of the discontinuous dynamic transition encountered as a function of α for the choice of parameters $k = 6$, $\omega_0 = \omega_k = 0.005$, $\omega_1 = \omega_{k-1} = 0.92$, $\omega_2 = \cdots = \omega_{k-2} = 1$. Top panel: determination of α_d from the study of the decorrelation time $n_*(\alpha)$ for $\alpha < \alpha_d$. The plot displays $n_*(\alpha)^{-2}$ versus α, where one has defined $n_*(\alpha)$ as the first iteration time for which the overlap (or point-to-set correlation) drops below the value $C = 0.4$. The line is a fit of the data of the form $n_*(\alpha)^{-2} = A\,(\alpha_d - \alpha)$, with fitting parameters A and α_d. The linear behavior confirms the divergence of n_* with a scaling exponent $-1/2$, as in the scalar case (8.9), the fit gives the estimation $\alpha_d = 19.47$. Bottom panel: determination of α_d from the study of the stability parameter $\lambda(\alpha)$ for $\alpha > \alpha_d$. The plot displays $(1 - \lambda(\alpha))^2$ versus α, the linear fits reproduces the scaling behavior (8.8) of the scalar case, and the fit yields $\alpha_d = 19.467$.

As a consistency check we also present in Fig. 8.5 a similar study in the case of a continuous transition. We see that the stability parameter λ computed on the non-trivial solution, i.e. for $\alpha > \alpha_d$, reaches 1 with a linear behavior (as in the scalar case, see Eq. (8.5)), and that its extrapolation is in good agreement with the analytically computed value of α_{KS} from Eq. (7.130). Moreover the numerical computation of the stability parameter of the trivial fixed-point

coincides for $\alpha < \alpha_d$ with the analytical one, $\lambda^{KS} = \alpha k(k-1)\theta^2$.

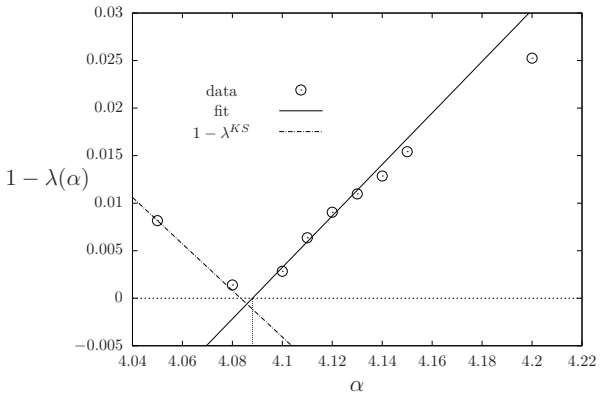

Figure 8.5: The stability parameter $\lambda(\alpha)$ in the case of a continuous transition (here for $k = 4$, $\omega_0 = \omega_k = 0$, $\omega_1 = \cdots = \omega_{k-1} = 1$). From the fit one obtains the result $\alpha = 4.088$, wich is in good agreement with the analytical one $\alpha_{KS} = 4.083$.

8.2.3 The stability parameter λ in the functional case

As an intermediate step in the generalization from the scalar to the functional case let us consider a fixed point equation of the form $\vec{x} = f(\vec{x}, \alpha)$, where the unknown $\vec{x}$ is a finite-dimensional real vector. The stability of a branch of solutions $\vec{x}(\alpha)$ can be determined by considering the Jacobian matrix J of the first derivatives of f computed at the fixed point, that can be defined through the linearization

$$f(\vec{x}(\alpha) + \vec{\varepsilon}, \alpha) = \vec{x}(\alpha) + J\,\vec{\varepsilon} + o(\|\vec{\varepsilon}\|)\ , \tag{8.10}$$

where $\vec{\varepsilon}$ is a small perturbation around the fixed point. The stability parameter $\lambda(\alpha)$ can then be defined as the spectral radius of J, i.e. the largest absolute value of the elements of its spectrum. This spectral radius can be expressed in terms of successive applications of J on a perturbation $\vec{\varepsilon}$ as

$$\lambda(\alpha) = \lim_{m\to\infty} \left(\frac{\|J^m \vec{\varepsilon}\|}{\|\vec{\varepsilon}\|} \right)^{\frac{1}{m}}\ , \tag{8.11}$$

where we assume that $\vec{\varepsilon}$ has a non-vanishing projection on the eigenspace associated to the relevant eigenvalue, and where $\|\bullet\|$ can be any norm. For future use let us define $\vec{\varepsilon}_m = J^m \vec{\varepsilon}$ and rewrite this expression as

$$\lambda(\alpha) = \lim_{m\to\infty} \left(\frac{\|\vec{\varepsilon}_m\|}{\|\vec{\varepsilon}_{m-1}\|} \frac{\|\vec{\varepsilon}_{m-1}\|}{\|\vec{\varepsilon}_{m-2}\|} \cdots \frac{\|\vec{\varepsilon}_1\|}{\|\vec{\varepsilon}\|} \right)^{\frac{1}{m}} \tag{8.12}$$

We would like now to extend the computation of a stability parameter to the 1RSB equations (7.53,7.54) that can be rewritten as $P = F(P, \alpha)$ by grouping the two lines together. P being a probability distribution the Jacobian of F is now an infinite-dimensional operator, which makes the study of its spectrum rather difficult. Even worse, we do not have at our disposal an exact description of the fixed point P around which we would like to expand F: we only have a sequence of approximations of P_n by the population representation written in Eq. (8.1). The individual elements of these representations still evolve at each iteration step, even when the observables computed as averages of P_n have reached convergence (within the numerical accuracy fixed by the population size $\mathcal{N}$). To circumvent these difficulties we have followed a strategy inspired by the expression (8.12): we consider P_n and a slight perturbation of it, $P_n + \delta P_n$, and assess the rate of growth of the perturbation along the iterations by the functional F. In order to implement this idea in practice one needs to choose a specific form for the perturbation; given that P_n is represented as a sum of Dirac deltas we perturb it by giving an infinitesimal width to each of the peaks, that we replace by Gaussian distributions with a small variance. We thus define

$$P_n(h) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \delta(h - h_i^{(n)}) \ , \qquad (P_n + \delta P_n)(h) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mathcal{G}(h; h_i^{(n)}, \varepsilon_i^{(n)}) \ , \tag{8.13}$$

$$\widehat{P}_n(u) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \delta(u - u_i^{(n)}) \ , \qquad (\widehat{P}_n + \delta \widehat{P}_n)(u) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mathcal{G}(u; u_i^{(n)}, \widehat{\varepsilon}_i^{(n)}) \ , \tag{8.14}$$

where $\mathcal{G}(\cdot; a, b)$ denotes the density of a Gaussian random variable of average a and variance b. Consider now the insertion of the form (8.14) in the right hand side of (7.53); the choice of the d peaks indexed by $i_1, \dots, i_d$ produces a random variable h equal in distribution to $f_{\Theta_0}(u_{i_1} + \sqrt{\widehat{\varepsilon}_{i_1}} z_1, \dots, u_{i_d} + \sqrt{\widehat{\varepsilon}_{i_d}} z_d)$, where $z_1, \dots, z_d$ are independent standard Gaussians (of zero mean and unit variance). As the $\widehat{\varepsilon}$ are infinitesimally small one can linearize f_{Θ_0} to compute the mean and variance of this random variable.

In summary, the determination of $\lambda(\alpha)$ is done by tracking the evolution of P, $\widehat{P}$ and their perturbed versions with populations of couples of real numbers, (h_i, ε_i) and $(u_i, \widehat{\varepsilon}_i)$, that evolve in time according to the following generalization of the update rules given in Sec. 8.1. To obtain $(h_i^{(n+1)}, \varepsilon_i^{(n+1)})$ one repeats, independently for $i = 1, \dots, \mathcal{N}$, these steps:

- draw an integer d from the law p_d
- draw d indices $i_1, \dots, i_d$ uniformly at random in $\{1, \dots, \mathcal{N}\}$
- set $h_i^{(n+1)} = f_{\Theta_0}(u_{i_1}^{(n)}, \dots, u_{i_d}^{(n)})$ and $\varepsilon_i^{(n+1)} = \sum_{j=1}^{d} (\partial_j f_{\Theta_0})^2 \widehat{\varepsilon}_{i_j}^{(n)}$, where $\partial_j f_{\Theta_0}$ denotes the derivative of f_{Θ_0} with respect to its j-th argument, computed in $(u_{i_1}^{(n)}, \dots, u_{i_d}^{(n)})$

Similarly the population $(u_i^{(n)}, \widehat{\varepsilon}_i^{(n)})$ is generated according to, again independently for $i = 1, \ldots, \mathcal{N}$:

- draw $\sigma_1, \ldots, \sigma_{k-1}$ from the probability law $\widetilde{p}(\sigma_1, \ldots, \sigma_{k-1}|+)$
- draw $k-1$ indices $i_1, \ldots, i_{k-1}$ uniformly at random in $\{1, \ldots, \mathcal{N}\}$
- set $u_i^{(n)} = \widehat{f}_{\Theta_0}(\sigma_1 h_{i_1}^{(n)}, \ldots, \sigma_{k-1} h_{i_{k-1}}^{(n)})$ and $\widehat{\varepsilon}_i^{(n)} = \sum_{j=1}^{k-1} (\partial_j \widehat{f}_{\Theta_0})^2 \varepsilon_{i_j}^{(n)}$

The rate of growth of the perturbation during the iteration $n \to n+1$ is estimated as the ratio of the L_1 norms of the perturbation parameters,

$$\lambda_n = \frac{\sum_{i=1}^{\mathcal{N}} \varepsilon_i^{(n+1)}}{\sum_{i=1}^{\mathcal{N}} \varepsilon_i^{(n)}} \ , \tag{8.15}$$

and the stability parameter is finally computed as

$$\lambda(\alpha) = (\lambda_{n_0} \lambda_{n_0+1} \ldots \lambda_{n_0+m-1})^{\frac{1}{m}} \ . \tag{8.16}$$

Indeed the first n_0 iterations are done with the usual population dynamics algorithm, evolving only the h_i's and u_i's, in order to reach an approximate convergence in distribution of the populations to their fixed points, and the perturbation is then initialized with $\varepsilon_i^{(n_0)} = 1$. A large number m of additional iterations during which the growth rates are recorded are then performed, and averaged geometrically as in (8.12); in the large m limit the value of $\lambda(\alpha)$ should be independent of the norm used to define λ_n. In practice we divide the $\varepsilon_i^{(n+1)}$ by λ_n after each iteration in order to keep the norm constant and avoid numerical underflows.

This method is similar to the one presented in [106, 107] to determine the location of a continuous RSB transition from a non-trivial RS solution.

8.3 Results of the cavity method

8.3.1 The existence of a RS phase for $\alpha > \alpha_{d,u}$

We shall study the evolution of the dynamic phase transition when the measure $\mu_{\Theta_0}(\underline{\sigma})$ over the proper bicolorings of a typical Erdős-Rényi random hypergraph is not uniform anymore. In the setting considered in equation (6.7,6.8) this corresponds to take the parameters $\{\omega_p\}$ of the interaction function different from the uniform choice $\omega_0 = \omega_k = 0$, $\omega_1 = \cdots = \omega_{k-1} = 1$.

We will concentrate first on the "zero-temperature" case, i.e. on the measures that give a non-zero weight to proper bicolorings only, which implies $\omega_0 = \omega_k = 0$. The choice of the other parameters is constrained by the global spin-flip symmetry that we want to preserve, hence $\omega_p = \omega_{k-p}$; as it is obvious from (6.7), multiplying all the ω_p by a common constant does not change the properties of the model. One realizes that for $k = 3$ there is no free parameter left, we will thus concentrate on the cases $k \geq 4$ from now on. For arbitrary large values of

k there will be of the order of $k/2$ free parameters in the ω_p; we will however make the following choice for the zero-temperature measure:

$$\omega_0 = \omega_k = 0 \ , \qquad \omega_1 = \omega_{k-1} = 1 - \epsilon \ , \qquad \omega_2 = \cdots = \omega_{k-2} = 1 \ , \tag{8.17}$$

where ϵ is the sole parameter that quantifies the deviation from the uniform measure (that is recovered for $\epsilon = 0$). This slight loss in generality is made for the sake of simplicity, and motivated by considerations on the large k limit presented in chapter 10 (section 10.1). The parameter ϵ controls the relative weight given to the "almost monochromatic" constraints that contain a single vertex of one of the possible colors (positive values of ϵ disfavoring them); as discussed in Sec. 7.4 these are precisely those responsible for the existence of frozen variables, one of the mechanism of RSB.

We present in Fig. 8.6 phase diagrams in the (α, ϵ) plane for $k = 4$, $k = 5$ and $k = 6$. The three lines in these plots correspond to the thresholds defined in (7.35) from the vanishing of the RS entropy, in (7.130) from the instability of the RS solution (Kesten-Stigum threshold), and in (7.82) from the appearance of hard fields in the solution of the 1RSB equations at $\mathcal{X} = 1$ (rigidity threshold); specializing these three expression with the choice of parameters (8.17) yields

$$\alpha^{s=0}(k,\epsilon) = \frac{\ln 2}{\frac{k(1-\epsilon)\ln(1-\epsilon)}{2^{k-1}-1-k\epsilon} - \ln\left(1 - \frac{1+k\epsilon}{2^{k-1}}\right)} \ , \tag{8.18}$$

$$\alpha_{KS}(k,\epsilon) = \frac{1}{k(k-1)} \left(\frac{2^{k-1} - 1 - k\epsilon}{1 + (k-4)\epsilon} \right)^2 \ , \tag{8.19}$$

$$\alpha_r(k,\epsilon) = \frac{1}{k} \Gamma_r(k) \frac{2^{k-1} - 1 - k\epsilon}{1 - \epsilon} \ . \tag{8.20}$$

In addition the black squares in Fig. 8.6 signal a discontinuous appearance of a non-trivial solution of the 1RSB equations at $\mathcal{X} = 1$ upon increasing α, that we located by a numerical resolution of these equations following the methods explained in Sec. 8.2. One can see on these plots that for all values of ϵ there is a critical density of constraints, $\alpha_d(\epsilon)$, such that a non-trivial solution of the 1RSB equations at $\mathcal{X} = 1$ exist if and only if $\alpha > \alpha_d(\epsilon)$. To make this separation more visible the area on the left of $\alpha_d(\epsilon)$, i.e. the RS phase of the model, has been painted in gray in Fig. 8.6. Let us call $(\alpha_{\rm opt}, \epsilon_{\rm opt})$ the coordinates of the point on the line $\alpha_d(\epsilon)$ which maximizes the density α of constraints, $\alpha_{\rm opt} = \max_\epsilon \alpha_d(\epsilon)$, that corresponds to an optimal choice of the bias parameter. The numerical values of these optimal parameters can be found in Table 8.1 for $k = 4$, 5 and 6. By definition $\alpha_{\rm opt} \geq \alpha_{d,u} = \alpha_d(\epsilon = 0)$, the dynamic transition of the usual model, with the uniform measure over the proper bicolorings; the non-trivial result here is that the inequality is strict, i.e. that a well chosen value of the biasing parameter ϵ is able to turn the clustered uniform measure into an unclustered biased one (for $\alpha \in [\alpha_{d,u}, \alpha_{\rm opt}]$).

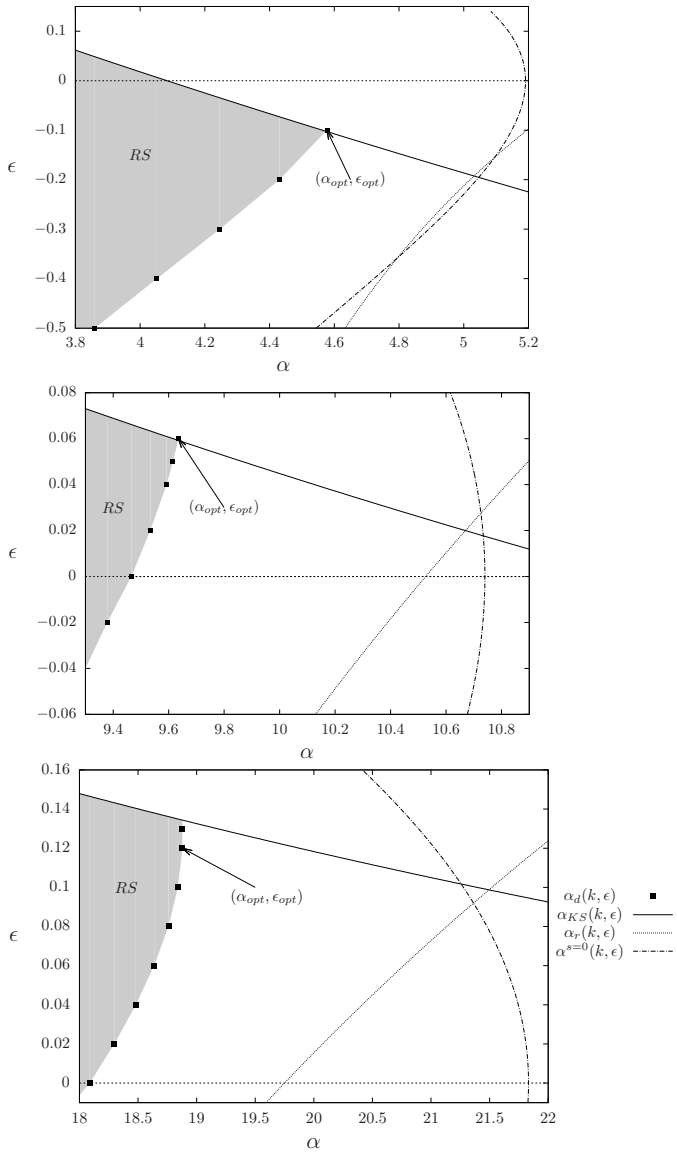

Figure 8.6: Phase diagram for $k = 4$, $k = 5$ and $k = 6$ (from top to bottom), in the plane (ϵ, α), at zero temperature $\omega_0 = 0$. The RS phase, painted in gray, is on the left of $\alpha_d(\epsilon)$, the latter corresponds either to a continuous transition with $\alpha_d(\epsilon) = \alpha_{KS}(\epsilon)$ (solid line, see (8.19)) or to a discontinous transition (black squares). The dashed horizontal line corresponds to $\epsilon = 0$, the uniform measure, which intersects α_d at $\alpha_{d,u}$. The arrow points to the optimal point that maximizes α_d. The dotted line is the rigidity threshold α_r from (8.20), the dot-dashed line marks the vanishing of the RS entropy (see Eq. (8.18)).

k	$\alpha_{d,u}$	$\alpha_{\rm opt}$	$\epsilon_{\rm opt}$
4	4.083	4.578	-0.10
5	9.465	9.636	0.06
6	18.088	18.879	0.12

Table 8.1: Dynamic threshold for the uniform measure ($\alpha_{d,u} = \alpha_d(\epsilon = 0)$), and largest α reachable in the RS phase, this optimal point having coordinates $(\alpha_{\rm opt}, \epsilon_{\rm opt})$.

A further scrutiny of the phase diagrams reveals different scenarios depending on the value of k. For $k = 4$ the nature of the bifurcation on the line $\alpha_d(\epsilon)$ changes precisely at $\epsilon_{\rm opt}$: for $\epsilon > \epsilon_{\rm opt}$ the transition is continuous and thus α_d coincides with the Kesten-Stigum line α_{KS}, while it is discontinuous for $\epsilon < \epsilon_{\rm opt}$ and there is a cusp at the optimal point (we shall come back on this point later on). It turns out that for $k = 4$, $\epsilon_{\rm opt} < 0$: this is rather counterintuitive at first sight, as it means that favoring the almost violated configurations of variables actually makes the measure less frustrated. This peculiarity can be explained by noticing that for $k = 4$ the dynamic transition of the uniform measure ($\epsilon = 0$) is continuous and that α_{KS} decreases with ϵ. As the dynamic transition of the uniform measure is discontinuous for $k \geq 5$ [103] this peculiarity is restricted to $k = 4$, and one has $\epsilon_{\rm opt}(k \geq 5) > 0$. Turning now to the phase diagram for $k = 5$ in Fig. 8.6 one observes similarly a cusp in $\alpha_d(\epsilon)$ at $\epsilon_{\rm opt}$, that separates a continuous and discontinuous branch of the dynamic transition line, but with now $\epsilon_{\rm opt} > 0$. Finally for $k = 6$ the optimal point is on the discontinuous branch of $\alpha_d(\epsilon)$; increasing further ϵ one encounters a cusp for at some value of $\epsilon > \epsilon_{\rm opt}$ and then a continuous branch $\alpha_d(\epsilon) = \alpha_{KS}(\epsilon)$. The large k behavior of the model will be further discussed in chapter 10; we can nevertheless anticipate that for large enough k the Kesten-Stigum threshold becomes irrelevant, as it happens in the negative RS entropy region (compare the leading orders of Eqs. (8.18,8.19)). In this case the whole line $\alpha_d(\epsilon)$ corresponds to a discontinuous bifurcation. As a last remark on the phase diagrams of Fig. 8.6 let us emphasize that for all ϵ one has necessarily $\alpha_d(\epsilon) \leq \min(\alpha_{KS}(\epsilon), \alpha_r(\epsilon), \alpha^{s=0}(\epsilon))$, these three thresholds implying a mechanism of failure for the hypotheses underlying a purely RS phase. This should easily convince the reader of the necessity of discontinuous branch of $\alpha_d(\epsilon)$ in some parts of the phase diagrams. For instance when $k = 4$ and $\epsilon \leq -0.3$ the rigidity and negative entropy bounds imply $\alpha_d(\epsilon) < \alpha_{KS}(\epsilon)$, in other words the dynamic transition must be discontinuous.

We have motivated earlier our study of the boundaries of the RS phase by algorithmic considerations, Monte Carlo Markov Chains being expected to equilibrate rapidly inside such a phase. However in a practical simulation one cannot assume that the initial configuration belongs to the support of a zero-temperature measure (otherwise the problem of finding a solution of the CSP would be already solved), it is thus necessary to make an annealing in temperature for a random initial condition to be allowed. For this reason we have also studied the evolution of the phase diagrams at positive temperature, modifying

the parameters (8.17) with $\omega_0 = \omega_k > 0$, see the results in Fig. 8.7. These plots show the absence of "reentrance" in temperature, in the sense that the lines $\alpha_d(\epsilon)$ move towards higher density of constraints when ω_0 is increased. Hence in principle a simulated annealing procedure with parameters (α, ϵ) in the zero temperature RS domain, progressively decreasing ω_0, should be able to remain equilibrated on polynomial time scales, hence finding solutions for $\alpha < \alpha_{\rm opt}$ if the appropriate bias is used. A numerical test of this conjecture is presented in Sec. 8.4.

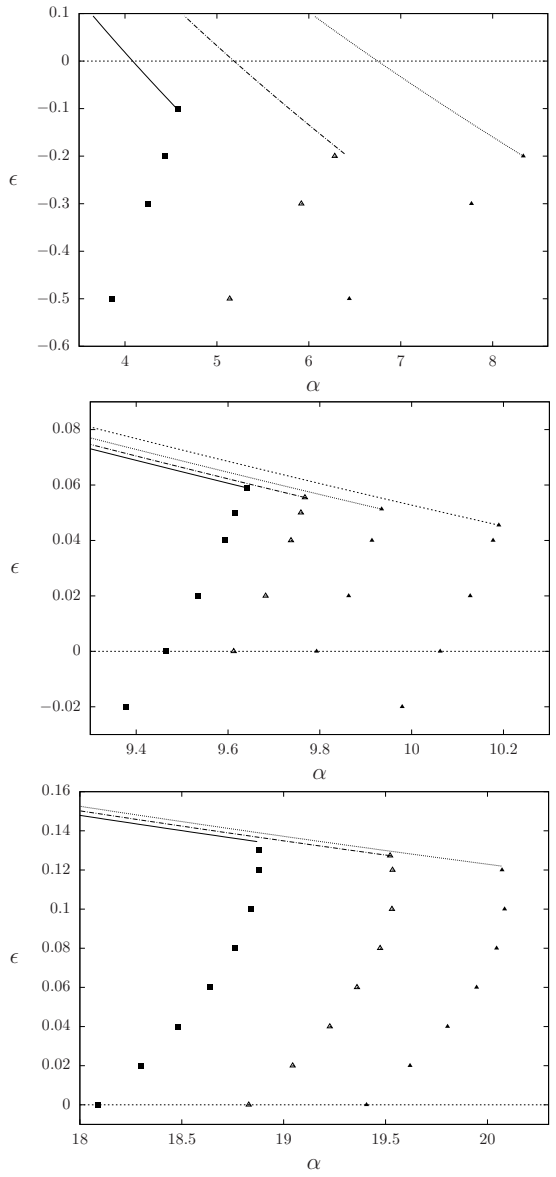

Figure 8.7: Phase diagrams for $k = 4$, $k = 5$ and $k = 6$ (from top to bottom) in the plane (α, ϵ), giving the RS phase delimitated by the KS bound and a dynamic line where the transition toward a non-trivial solution is discontinuous, for different temperatures. Top ($k = 4$): the dynamic line is given from left to right for $\omega = 0$ (filled square and solid line), $\omega = 0.1$ (empty triangle and dashed line), $\omega = 0.2$ (filled triangle and dotted line). Middle ($k = 5$): from left to right $\omega = 0$, $\omega = 0.002$, $\omega = 0.005$, $\omega = 0.01$. Bottom ($k = 6$): from left to right $\omega = 0$, $\omega = 0.005$, $\omega = 0.01$.

8.3.2 More detailed zero temperature phase diagrams

The extent of the RS domain in the (α, ϵ) phase diagram presented in Fig. 8.6 was the most interesting information to extract from the cavity formalism in the perspective of this Ph.D.. For the sake of completeness we shall nevertheless discuss with slightly more details some properties of the RSB phase, and present another version of the phase diagrams for $k = 4$ and $k = 5$ in Fig. 8.8.

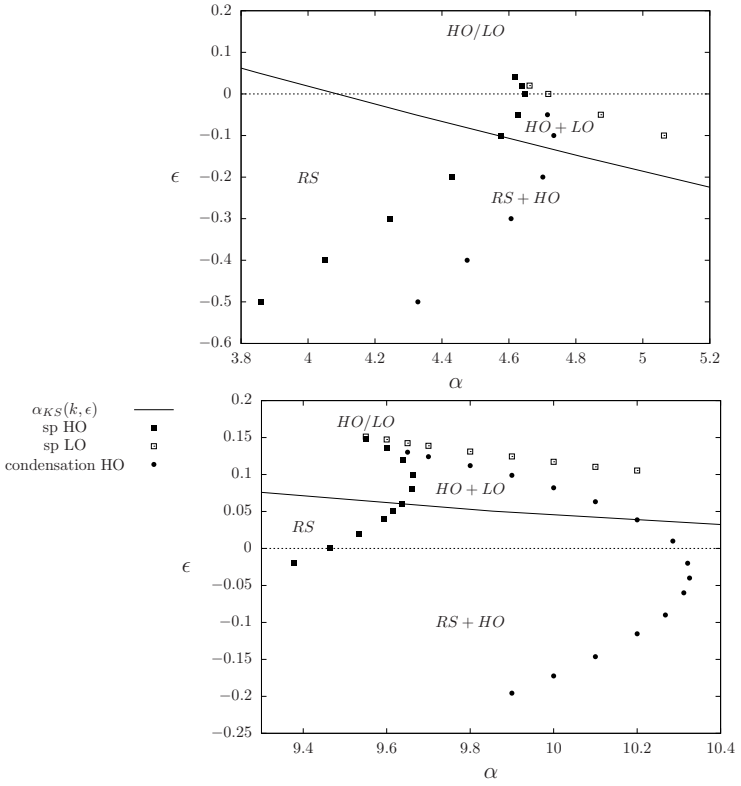

Figure 8.8: Phase diagram for $k = 4$ (top), and $k = 5$ (bottom), in the plane (α, ϵ), at zero temperature ($\omega_0 = 0$): the solid line is the Kesten-Stigum (KS) transition where a non-trivial solution of the 1RSB equations emerge continuously from the trivial one upon increasing α, the filled (resp. empty) squares corresponds to the spinodal (sp) of the HO (resp. LO) branch that appears discontinuously when α is increased (resp. decreased). The filled circles are defined by the vanishing of the complexity of the HO branch.

The most important additional feature unveiled by these phase diagrams is that for some values of the parameters k, α, ϵ, there exits (at least) two different non-trivial solutions of the 1RSB equations at $\mathcal{X} = 1$ (7.53,7.54). This type of behavior was described in [103] for a family of random CSPs generalizing the

hypergraph bicoloring, and its consequences for inference problems (or planted CSPs) have been discussed in [104]. In order to reach numerically these different solutions we used the population dynamics algorithm explained in Sec. 8.1 with an initial condition generalizing $P_{n=0}(h) = \delta(h-1)$ into

$$P_{n=0}(h) = (1-\varepsilon)\,\delta(h) + \varepsilon\,\delta(h-1)\ . \tag{8.21}$$

For each choice of the parameters we ran twice the population dynamics algorithm, once with $\varepsilon = 1$ and once with a small value of $\varepsilon > 0$ (in practice we used $\varepsilon = 0.01$); in the tree reconstruction interpretation the latter correspond to a variant known as robust tree reconstruction [108], in which only a fraction ε of the variables at large distance from the root are revealed to the observer. We will call HO, for high overlap, the initialization with $\varepsilon = 1$, and LO (low overlap) the small ε one. Depending on the parameters these two procedures can produce different solutions of the 1RSB equations, or not. More precisely, the different phases located in Fig. 8.8 are defined as follows:

- RS: both HO and LO initial conditions lead to the trivial solution.
- RS+HO: LO initial condition leads to the trivial solution, whereas HO initial condition leads to a non trivial solution.
- LO+HO: LO initial condition leads to a non trivial solution, HO initial condition leads to a non trivial solution with a higher overlap.
- HO/LO: both HO and LO initial conditions lead to the same non-trivial solution.

We present in the left panels of Fig. 8.9 the evolution of the overlap as a function of ϵ for different fixed α at $k = 5$, which should help to grasp the meaning and succession of the different phases. We recall that α_{KS} is the limit of stability of the trivial fixed point, that undergoes a bifurcation at the Kesten-Stigum transition. The two lines denoted sp (for spinodal) in Fig. 8.8 correspond to the limit of existence of the two non-trivial branches of solution of the 1RSB equations.

This more complete study of the number and domain of existence of solutions of the 1RSB equations should clarify the cusp at $\epsilon_{\rm opt}$ of the line $\alpha_d(\epsilon)$ found for $k = 4$ and $k = 5$ in Fig. 8.6: a first look at these figures could suggest that the two parts of the $\alpha_d(\epsilon)$ line join at a tri-critical point, in the sense that the discontinous transition becomes less and less discontinuous before crossing over to a continuous transition. However from Fig. 8.8 one sees that this is not the case, the discontinuous branch of $\alpha_d(\epsilon)$ extends to the RSB phase as a spinodal unrelated to the continuous transition, which can only make sense in the context of coexistence of two non-trivial solutions. A tri-critical point does exist in these phase diagrams, but it is located strictly inside the RSB phase, not at the cusp, and corresponds to the merging of the two spinodals.

Finally we have also indicated on Fig. 8.8 the threshold for the cancellation of the complexity of the HO solution (see in addition the right panels in Fig. 8.9);

this corresponds to the condensation transition of the model in the $\alpha < \alpha_{KS}$ part of the phase diagram (for $\alpha > \alpha_{KS}$ the LO solution has a negative complexity hence the problem is condensed, see [103] for a discussion of this point).

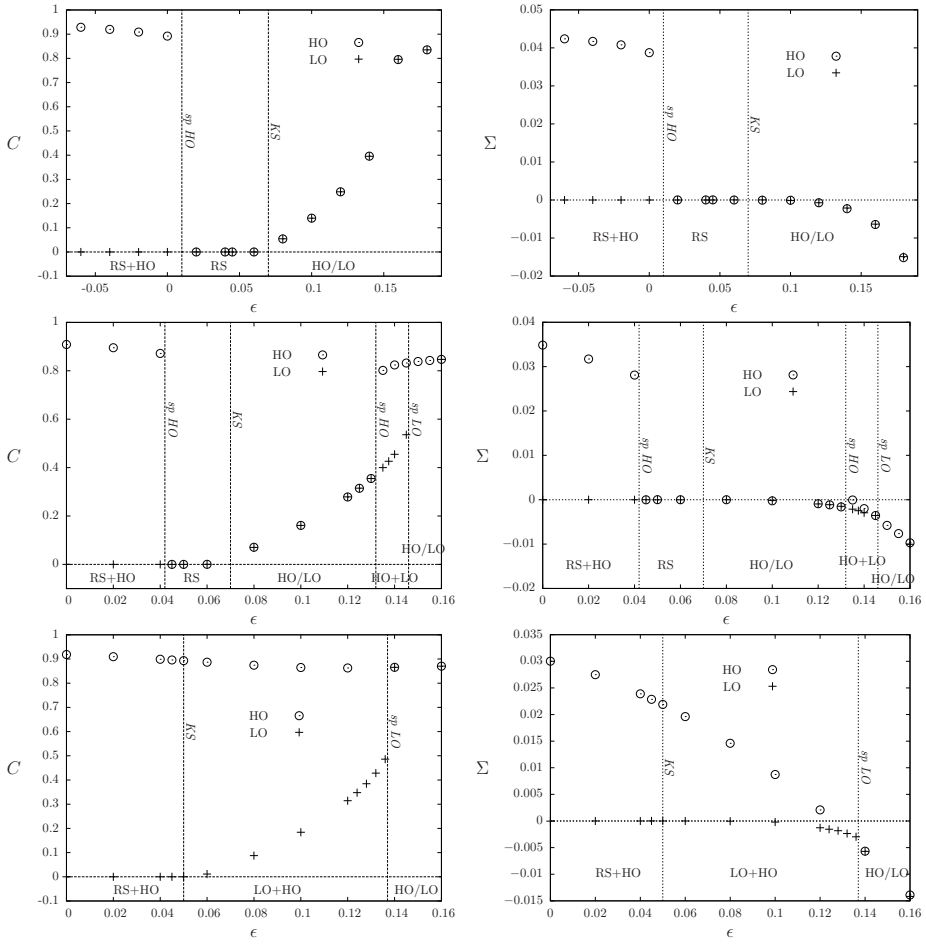

Figure 8.9: Overlaps C (left column) and complexities Σ (right column) versus ϵ at $k = 5$, for $\alpha = 9.5$ (top), $\alpha = 9.6$ (middle), and $\alpha = 9.7$ (bottom).

8.4 Results of Simulated Annealing

In this Section we present the results of extensive simulations, where we have used the Simulated Annealing (SA) algorithm [25] to find solutions of the hypergraph bicoloring problem. Our main aim is to show that SA finds solutions more easily if the biased measure μ_{Θ_0} (see equations (6.7) and the specialization

(8.17)) is used: although the uniform measure ($\epsilon = 0$) has a larger entropy, the biased one is more concentrated on solutions that can be reached in an easier way by SA and thus the SA algorithmic threshold improves if $\epsilon \neq 0$ is used.

In this setting, we consider Erdős-Rényi random hypergraphs with $k = 4, 5, 6$ and sizes ranging from $N = 10^4$ to $N = 10^6$. The parameters α and ϵ are taken in the relevant region where we expect an algorithmic phase transition to take place, that is around α_d. Let us rewrite the biased measure that we are willing to sample via the SA at a generic finite temperature $T = 1/\beta$ as

$$\mu(\underline{\sigma}) = \frac{1}{Z(G)} e^{-\beta U(\underline{\sigma})} (1-\epsilon)^{F(\underline{\sigma})} \tag{8.22}$$

where $U(\underline{\sigma})$ is the number of unsatisfied constraints (i.e. monochromatic hyperedges) and $F(\underline{\sigma})$ is the number of freezing clauses (i.e. hyperedges with exactly $k-1$ variables of the same color). This corresponds indeed to the measure (6.7) with parameters

$$\omega_0 = \omega_k = e^{-\beta} \ , \qquad \omega_1 = \omega_{k-1} = 1 - \epsilon \ , \qquad \omega_2 = \cdots = \omega_{k-2} = 1 \ . \tag{8.23}$$

The solutions of the CSP have $U = 0$, and non-uniform weights if $\epsilon \neq 0$. Our SA implementation uses the Metropolis algorithm with single-spin flip dynamics: at each time step one considers a configuration $\underline{\sigma}'$ that differs from the current configuration $\underline{\sigma}$ by the reversal of an uniformly chosen spin. The move $\underline{\sigma} \to \underline{\sigma}'$ is accepted with the probability

$$\min\left(\frac{\mu(\underline{\sigma}')}{\mu(\underline{\sigma})}, 1\right) = \min\left(e^{-\beta \Delta U}(1-\epsilon)^{\Delta F}, 1\right) \ , \tag{8.24}$$

where $\Delta U = U(\underline{\sigma}') - U(\underline{\sigma})$ and $\Delta F = F(\underline{\sigma}') - F(\underline{\sigma})$, in such a way that the detailed balance (reversibility) condition with respect to the measure (8.22) is ensured. We store the quantity $\sum_{i=1}^{k} \sigma_i$ for each clause, which allows a fast computation of the changes ΔU and ΔF when a spin is flipped.

We run SA with a very simple piecewise constant, uniformly spaced, temperature scheduling: the first Monte Carlo Sweep (MCS, i.e. N elementary steps described above) is performed with T fixed to $T_{\max}$ (we used $T_{\max} = 0.5$ in all our simulations), then T is reduced by $\Delta T = T_{\max}/\tau$ and a new MCS is performed, T is again reduced by ΔT, and so on and so forth. We perform in this way $\tau + 1$ MCS, the last one being at zero temperature, the running time of the algorithm thus scales as $N\tau$ elementary steps.

The lowest value of $U(\underline{\sigma})$ is always reached at the end of each run, when the annealing has reached zero temperature. So we present results only for the quantity $U_0 = U(T = 0)$, that is the smallest number of violated clauses that the SA is able to reach in a running time of τ MCS. SA is successful as a solver if and only if $U_0 = 0$, but we will be interested in estimating the lowest energy reachable by SA even in the regime where it is not successful. In particular we are going to study the lowest intensive energy reached by SA, $u_0 = U_0/N$, in the large size limit where it becomes independent on the problem size N.

8.4.1 Estimating the algorithmic threshold for Simulated Annealing

We shall first discuss the problem of the estimation of the algorithmic threshold for a stochastic algorithm like SA, concentrating for simplicity on the unbiased ($\epsilon = 0$) case, the extension to $\epsilon \neq 0$ will be considered later on.

The behavior of the algorithm depends on the density of constraints α, the annealing time τ, and the size of the problem N; it can be described in terms of the average energy density $u_0(\alpha, \tau, N)$ reached at the end of the run, or in terms of the probability (with respect to the random instance generation and the stochasticity of the algorithm) $p_{\rm succ}(\alpha, \tau, N)$ that the algorithm discovers a solution of the instance. It is clear that the energy (resp. success probability) reached by SA is a decreasing (resp. increasing) function of the running time τ. We are interested in the limit of large times but sub-exponential with respect to the problem size N (on exponentially large timescales any Monte Carlo simulation of a finite size system is ergodic and $u_0 = 0$ as long as $\alpha < \alpha_{sat}$, but this is not the regime we are interested in). An idealized definition of the algorithmic threshold $\alpha_{\rm algo}$ would be the smallest density of constraints such that

$$\lim_{N\to\infty} u_0(\alpha, \tau = N^c, N) > 0 \qquad \text{or} \qquad \lim_{N\to\infty} p_{\rm succ}(\alpha, \tau = N^c, N) = 0 \ , \quad (8.25)$$

for any fixed exponent c, corresponding to polynomial time algorithms. Of course time and space requirements impose strong constraints on the values of τ and N that can be used in practice. The limit above must thus be performed by an extrapolation from finite N results, and if c is free any running time could be considered as "polynomial" as long as N is finite. To resolve this ambiguity we shall restrict our study to linear times (this time scale is the only one practically accessible on very large problems), i.e. consider τ fixed (but arbitrary large) in the thermodynamic limit $N \to \infty$.

Even with this restriction the numerical extrapolation necessary to estimate $\alpha_{\rm algo}$ is far from being an easy task. The definition given above relies on the behavior of the asymptotic intensive energy u_0 as a function of α. We plot the corresponding data in Figure 8.10 for $k = 4, 5$: these data have been obtained for the unbiased measure ($\epsilon = 0$) and different problem sizes ($10^4 \leq N \leq 10^6$); note that data points with different N values are very close, i.e. the size dependence is very weak, and the values for $N = 10^5$ and $N = 10^6$ always coincide within errorebars. Unfortunately the asymptotic energy u_0 is strongly dependent on the running time τ and it is thus very difficult to extract from this figure the algorithmic threshold, i.e. the value of α where u_0 becomes positive in the large τ limit.

A much more convenient way of analyzing the same data is presented in Figure 8.11, where for each value of α we study the dependence of u_0 on τ. The α values shown are such that the relative difference between the smallest and the largest α values is around 5%. Again the size dependence is weak and we can mostly ignore it. Error bars have been computed only from sample to sample fluctuations. The main observation now — note the log-log scale in the plots

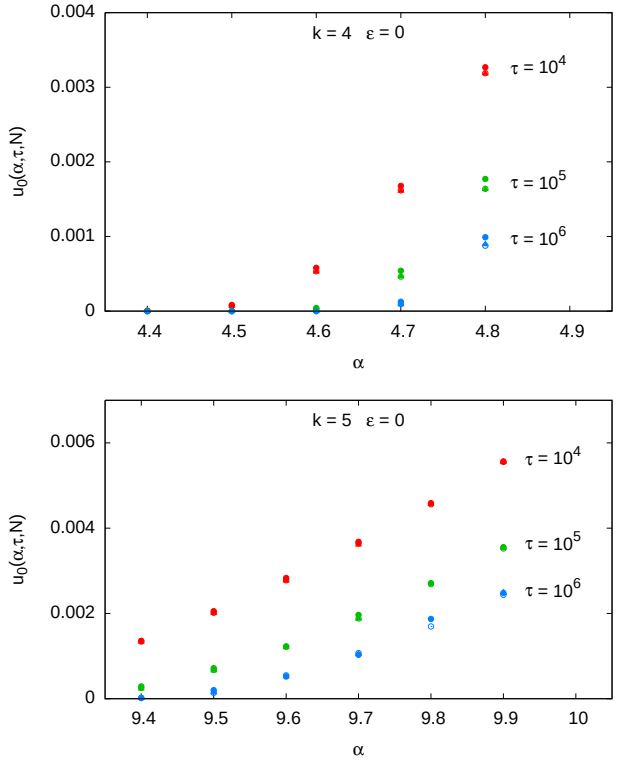

Figure 8.10: Lowest intensive energy reached by SA for $k = 4$ (top) and $k = 5$ (bottom) as a function of α for different cooling times τ. For each cooling time τ we show results for 3 problem sizes: $N = 10^4$ with filled circles, $N = 10^5$ with empty circles and $N = 10^6$ with triangles. The latter two values do always coincide (except for $k = 5$ and $\alpha = 9.8$, where the $N = 10^6$ datapoint is missing). Estimating the algorithmic threshold from these plots is very difficult due to the strong τ dependency.

— is that for the smallest α values shown in the plots the asymptotic energy is decreasing very fast with τ, faster than a power law (data not shown have $u_0 \simeq 0$); on the contrary, for the largest α values, u_0 decreases slower than a power law. In the latter case we even observe an upwards curvature, suggesting a non-zero value for u_0 in the $\tau \to \infty$ limit.

In practice, our best estimate for the SA algorithmic threshold is given by the α value such that u_0 decays as an inverse power law of τ, thus separating the regimes where u_0 decays faster and slower than a power law in τ. For $\epsilon = 0$, we find the following approximate values $\alpha_{alg}(k = 4) \approx 4.7$, $\alpha_{alg}(k = 5) \approx 9.6$ and $\alpha_{alg}(k = 6) \approx 18.5$. We notice that all these algorithmic thresholds are larger than the threshold $\alpha_{d,u}$ listed in Table 8.1, where the "dynamic" phase

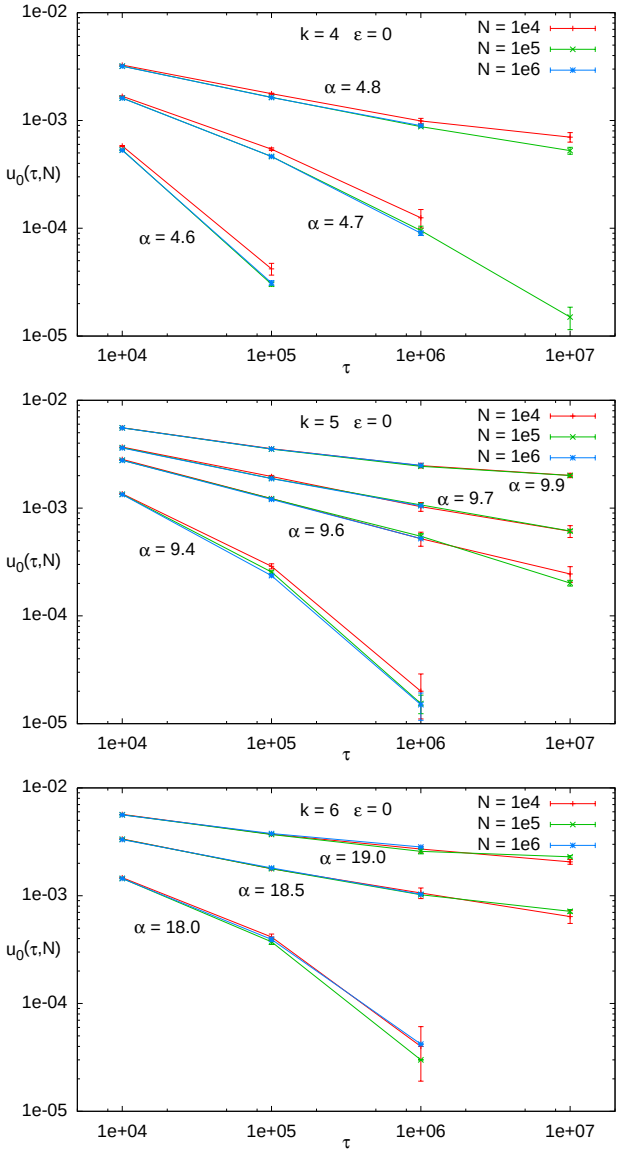

Figure 8.11: Plotting the lowest intensive energy reached by SA with no bias ($\epsilon = 0$) as a function of τ provides a better way to estimate the algorithmic threshold α_{alg}: the decay is faster (resp. slower) than a power law if $\alpha < \alpha_{alg}$ (resp. $\alpha > \alpha_{alg}$).

transition, defined as the appearance of a solution of the 1RSB equations, takes place. This observation is consistent with the idea that sampling solutions

uniformly is more difficult than just finding one or few solutions. Indeed, while a MCMC is expected to sample uniformly the solutions efficiently only for $\alpha < \alpha_{d,u}$, SA can find a solution in linear time until α_{alg}, which is greater than $\alpha_{d,u}$. In other words, in the range $[\alpha_{d,u}, \alpha_{alg}]$ the SA algorithm does not thermalize at the lowest temperatures explored during the annealing, but it can be seen as an efficient out of equilibrium process converging in linear time to a solution, as discussed for instance in [98].

In particular for $k = 4$ the model has a continuous phase transition and the SA algorithm seems to be very efficient in this case: the algorithmic threshold $\alpha_{alg}(k = 4) \approx 4.7$ is well beyond the dynamic threshold $\alpha_{d,u} = 4.083$ and not far from the 1RSB estimate of the satisfiability threshold $\alpha_{sat}(k = 4) \approx 4.9$ [103] (this is only expected to be an upper bound on the true satisfiability threshold due to an instability towards higher levels of RSB). On the contrary for $k \geq 5$ the phase transition taking place at $\alpha_{d,u}$ is of the random first order type (discontinuous) and this seems to have a dramatic effect on the performance of SA, which is able to find solutions only slightly beyond $\alpha_{d,u}$, stopping far from the α_{sat} threshold. For example, for $k = 5$ we have $\alpha_{d,u} = 9.465$, $\alpha_{alg} \approx 9.6$ and $\alpha_{sat} = 10.46$ [103].

8.4.2 Performances of Simulated Annealing with optimal RS parameters

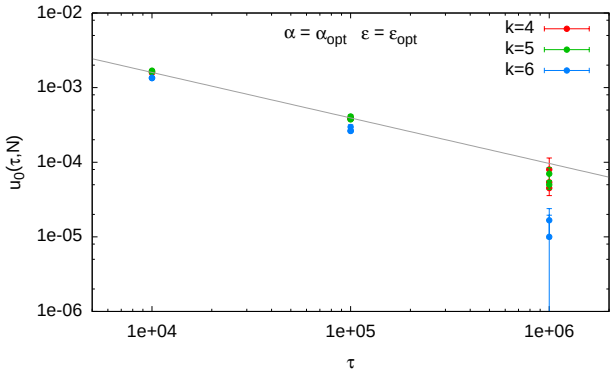

Figure 8.12: The fast decays of u_0 as a function of τ for $\alpha = \alpha_{\rm opt}$ and $\epsilon = \epsilon_{\rm opt}$ confirms that for these "optimal" parameters SA is effective in finding the ground state in linear time.

As shown in Sec. 8.3.1 we can extend the RS phase in the region $\alpha > \alpha_{d,u}$ by tuning appropriately the bias ϵ. In Figure 8.12 we show the asymptotic energy as a function of the running time τ for the parameters $\alpha = \alpha_{\rm opt}$ and $\epsilon = \epsilon_{\rm opt}$ given in Table 8.1, that are optimal from the point of view of extending the RS phase to the largest α possible. For each value of τ we report the results

obtained with 3 sizes $N = 10^4, 10^5, 10^6$ although the different data points are hardly visible due to their strong overlap (for $\tau = 10^6$ fewer sizes are shown). Errors are computed from sample to sample fluctuations.

For all the values of $k = 4, 5, 6$ the behavior of the asymptotic energy is compatible with a power law decay or even faster than that (the straight line is just a guide to the eye with a slope -0.61). So, as expected, SA seems to be an efficient algorithm to find solutions in the RS phase, even when this phase extends beyond $\alpha_{d,u}$ via the optimization of the bias ϵ.

8.4.3 Performances of Simulated Annealing with the biased measure

In Section 8.3 we have shown how the phase diagram and the corresponding thresholds change in presence of a non-zero bias ($\epsilon \neq 0$). The suggestion we get from this analytical study is that a non-zero bias should make easier for the SA algorithm to find solutions at large α values. However the connection between the phase diagram and the behavior of the SA is not obvious, as already shown in Section 8.4.1 for the $\epsilon = 0$ case.

The aim of the present section is to show the results of extensive numerical simulations running SA with the biased measure in order to gather evidence that a non-zero bias is in general beneficial for the performances of SA in finding a solution to the random hypergraph bicoloring problem.

We have already shown that finite size effects are very small and slightly visible only for $N = 10^4$. So in the following we present uniquely data obtained with size $N = 10^5$. We have checked these are practically indistinguishable from the results with $N = 10^6$ on the time scales reachable in the latter case.

In Figure 8.13 we show the data collected at the three largest α values for $k = 4$ (upper row), $k = 5$ (middle row) and $k = 6$ (lower row). In each panel we plot u_0 as a function of ϵ for three different cooling rates $\tau = 10^4, 10^5, 10^6$ (from top to bottom in each panel). The plots provides a clear evidence that reweighting solutions with a bias $\epsilon > 0$ enhances the probability that SA reaches lower energies.

Already a simple qualitative analysis reveals the advantage of using $\epsilon > 0$. In every panel we see that u_0 reaches a minimum for a strictly positive value of ϵ. The value $\epsilon_{\rm SA}$ that minimizes u_0 is only weakly dependent on the SA cooling time τ, so it is likely to assume that $\lim_{\tau\to\infty} \epsilon_{\rm SA} > 0$ and the bias is effective even in the limit of large times.

The data in Figure 8.13 suggest that the SA algorithmic threshold may grow for moderately small values of ϵ with respect to its $\epsilon = 0$ value. For example for $k = 6$ the SA algorithmic threshold for $\epsilon = 0$ was estimated around $\alpha_{alg} \approx 18.5$, but looking at the plots in the lower row it is evident that at least for $\alpha = 19$ and $\epsilon \simeq 0.15$ SA reaches the ground state $u_0 = 0$ and for $\alpha = 19.5$ and $\epsilon \simeq 0.15$ the convergence to $u_0 = 0$ is very fast in τ.

We have done a more quantitative analysis for the largest α values, shown in the three panels on the right. We have interpolated the data of u_0 with a quadratic function of ϵ, the interpolating parabolas are shown in the right panels

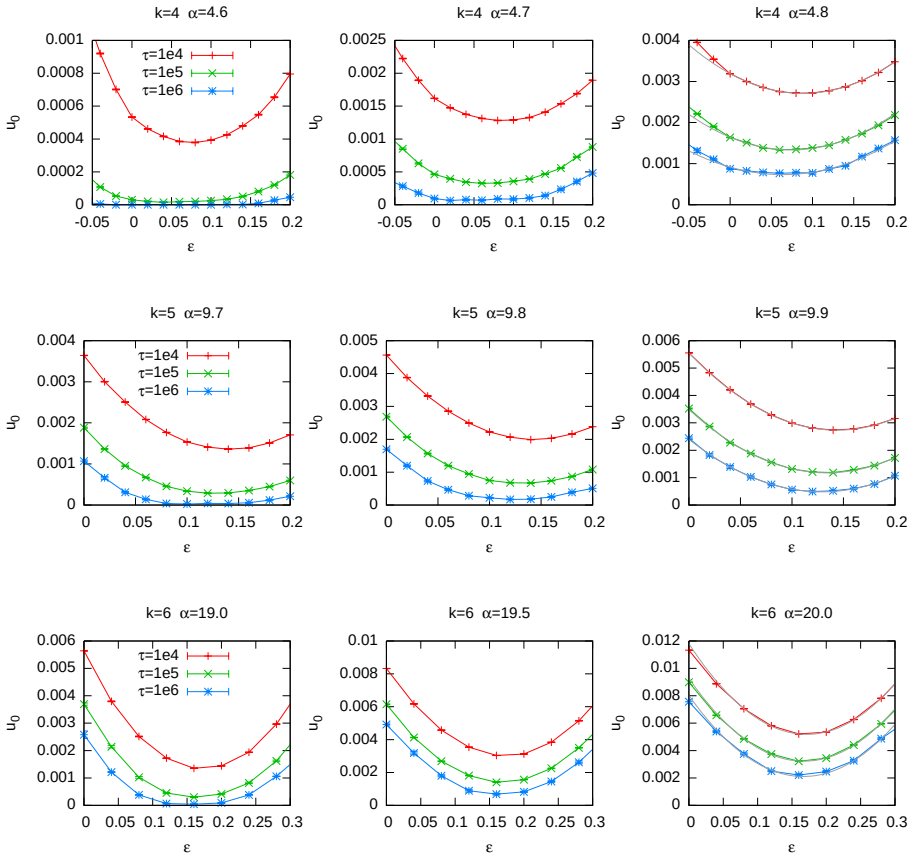

Figure 8.13: The lowest intensive energy reached by SA reaches its minimum for a positive bias parameter ϵ.

in Figure 8.13. Fitting the minimum of the parabola $u_0(\epsilon_{\rm SA})$ as a power law in τ we find the results shown in Figure 8.14. For $(k = 5, \alpha = 9.9)$ the behavior is faster than a power law and thus we expect the SA algorithmic threshold with the biased measure to be slightly greater than $\alpha = 9.9$. On the contrary for both $(k = 4, \alpha = 4.8)$ and $(k = 6, \alpha = 20)$ the behavior looks slightly slower than a power law and thus we are tempted to believe $\lim_{\tau\to\infty} u_0(\epsilon_{\rm SA}) > 0$ in these cases and the SA algorithmic threshold is slightly below.

Unfortunately the quantitative analysis cannot be made more robust, due to the strong τ dependence observed. Nevertheless we believe that the evidence that $\epsilon > 0$ makes ground states more accessible to Simulated Annealing is strong enough, both for finite τ values and in the large τ limit.

Let us finally compare the optimal value $\epsilon_{\rm SA}$ of the bias that improves most the performances of SA with the optimal value $\epsilon_{\rm opt}$ found in Section 8.3.1 that

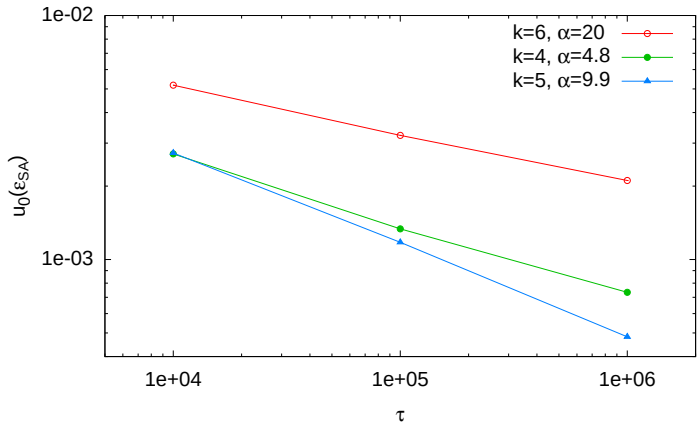

Figure 8.14: The lowest energy reached with the optimal bias $\epsilon_{\rm SA}$ for the largest α value simulated as a function of τ.

increases most the extent of the RS phase. We notice that $\epsilon_{\rm SA}$ is always larger than $\epsilon_{\rm opt}$. Considering that, for the small values of k studied in the simulations, it is approximately true that for $\epsilon > \epsilon_{\rm opt}$ the phase transition increasing α is continuous, while for $\epsilon < \epsilon_{\rm opt}$ the model undergoes a random first order transition, we believe the most natural explanation for the observation $\epsilon_{\rm SA} > \epsilon_{\rm opt}$ is the following. The ergodicity breaking taking place at a discontinuous (i.e. random first order) transition is much more severe than the one taking place at a continuous phase transition. In the case of a discontinuous phase transition, the SA algorithm can find solutions only slightly above the dynamic threshold α_d, while in the continuous case SA remain efficient in finding solutions even well above α_{KS}. The analysis supporting this scenario has been presented in Section 8.4.1. Thus it is natural to expect that SA presents its best performances for $\epsilon > \epsilon_{\rm opt}$ where the phase transition is continuous and the ergodicity breaking not too severe. The finding that $\epsilon_{\rm SA}(k=4) > 0$ also resolves the rather counterintuitive result $\epsilon_{\rm opt}(k=4) < 0$. So even for $k = 4$ the SA algorithm finds the ground state more easily if frozen variables are avoided.

8.5 Comparison with the biased measure with interactions at distance 1

In this section we present the study of the clustering threshold for the biased measure μ_{Θ_1} defined in equation (6.12), and compare it to the simpler measure μ_{Θ_0}, in the k-uniform $l+1$-regular random ensemble. We apply the numerical procedure described in Section 8.1 to solve the recursive distributional equations (7.67-7.69). The distributions $P_{w,n}$ are approximated by the empirical

distribution over a sample of $\mathcal{N}$ representants as:

$$P_{w,n}(\eta) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \delta(\eta - \eta_i) \ . \tag{8.26}$$

With $2\mathcal{N}$ fields η_i one can encode the distributions $P_{w,n}$ at distance n for $w = 0, 1$. The $2\mathcal{N}$ fields representing the two distributions $\widehat{P}_{w,n+1}$ are generated stochastically according to (7.68,7.69), and in turn the populations representing $P_{w,n+1}$ can be obtained from (7.67). At each step of this iterative procedure one computes the correlation function C_n from (7.70), interpreting the average over $\widehat{P}_{w,n}$ as an uniform sampling of an element of the corresponding population. An example of the results thus obtained is presented in Fig. 8.15, where one sees, depending on the choices of parameters, RS cases with C_n vanishing at large n, and RSB situations where C_n remains positive. The results presented in the rest of this section have been obtained with populations of size $\mathcal{N} = 10^6$; we considered that $C_n \to 0$ whenever the average value of C_n, for large enough values of n such that stationarity was reached within our numerical accuracy, dropped below a small threshold value (we used 0.005 in figures below).

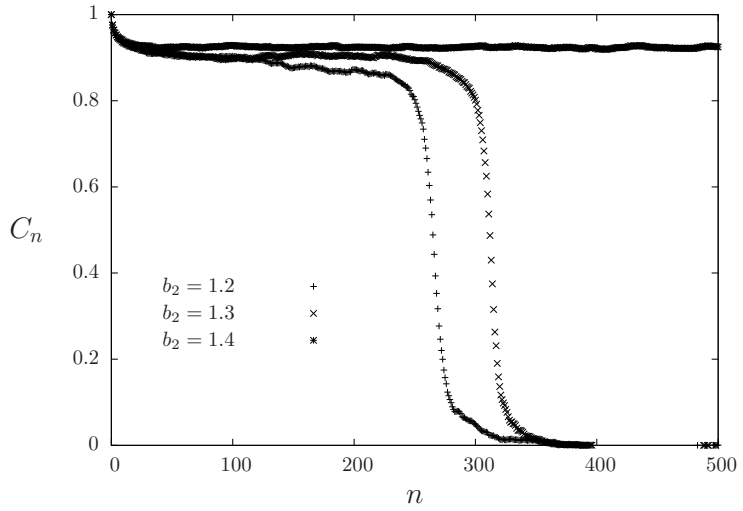

Figure 8.15: An example of the shape of the correlation function C_n as a function of n, here for $k = 6$, $l = 114$ and the bias function $\psi(p)$ defined in Eq. (8.27), with $b_1 = 1.6$, $\epsilon = 0.095$, and three values of b_2.

The function $\psi(p)$ contains a large number $(l + 1)$ of free parameters, some choices must hence be made on its specific form. We first considered the case where $\psi(p) = (1 - \epsilon)^p$, with a single free parameter $\epsilon < 1$. As explained in Sec. 6.1.3 this corresponds the biased measure μ_{Θ_0} with the choice of parameters (8.17), i.e. to a bias that factorizes over the hyperedges of the bicoloring

problem. We have presented in the section 8.3 the study made in [31] for Erdős-Rényi (ER) random hypergraphs in which the degree of a vertex has a Poisson distribution of average αk. The phase diagrams obtained for the regular case (see [33]) are presented in Fig. 8.16. They show the clustering threshold l_d (points), as well as the upper bounds corresponding to the rigidity, the Kesten-Stigum bound, and the vanishing of the RS entropy, in the plane (l, ϵ), for $k = 5$ and $k = 6$. They are qualitatively similar to the results obtained in the section 8.3, figure 8.6 (see [31]) for the ER case, and quantitatively close with the correspondence $\alpha k = l + 1$ between the average degree of the ER ensemble and the one fixed here. The important point we want to emphasize here is the fact that a suitable choice of ϵ allows us to increase $l_{\rm d}$ with respect to its value for the uniform measure ($\epsilon = 0$). For instance for $k = 5$ and $l = 47$, the RSB phase at $\epsilon = 0$ is turned into a RS phase when $\epsilon = 0.04$. Similarly for $k = 6$ the dynamic transition $l_{\rm d} = 108$ of the uniform measure can be pushed to $l_{\rm d} = 113$ for a well-chosen value of ϵ.

The natural question that arises at this point is whether the more generic bias introduced in this manuscript, i.e. the additional degrees of freedom in the choice of $\psi(p)$, allows us to further increase the dynamic transition threshold $l_{\rm d}$. To investigate this point without introducing too large a space of parameters, that would be impossible to explore systematically, we considered the following function ψ:

$$\psi(0) = 1 \ , \quad \psi(1) = b_1 \ , \quad \psi(p) = b_2(1-\epsilon)^p \quad \text{for} \quad p \geq 2 \ , \tag{8.27}$$

with the three free parameters (b_1, b_2, ϵ). We then solved numerically the RDEs with parameters close to the optimal values found previously in the restricted case with $b_1 = 1 - \epsilon$ and $b_2 = 1$. The results are shown in Fig. 8.17 in the parameter plane (b_1, b_2), for fixed values of k, l and ϵ, with squares (resp. crosses) marking RS (resp. RSB) phases. The top panel shows the existence of a RS phase at $k = 5$, $l = 48$, whereas all values of ϵ led to RSB at this value of l for the factorized bias $\psi(p) = (1-\epsilon)^p$. We did not find any choice of parameters (b_1, b_2, ϵ) with a RS phase for $l = 49$. Similarly the bottom panel shows, for $k = 6$, the largest value of l, $l = 114$, for which we found a RS phase for well-chosen parameters (see also the drop of C_n to 0 in Fig. 8.15 for $\epsilon = 0.095$, $b_1 = 1.6$, $b_2 = 1.2$ and 1.3).

We summarized the main results of this Section in the Table 8.2. The first column gives the value of $l_{\rm d}$ of the uniform measure, the second column gives the result obtained with a bias of the form μ_{Θ_0} (6.7) with a choice of parameter (8.17), when optimizing on the choice of ϵ to increase as much as possible $l_{\rm d}$. The third column gives the result obtained with a biased measure of the form μ_{Θ_1} (6.12), for well-chosen values of (b_1, b_2, ϵ). We can see that we were able to further improve the value of $l_{\rm d}$ for $k = 5$ and $k = 6$ by using a bias that introduces interactions between variables of different hyperedges, with respect to the factorized bias. An even further improvement might be achieved by a more systematic exploration of the parameter space (b_1, b_2, ϵ), or by using even more general bias functions $\psi(p)$, at the price of a large computational cost

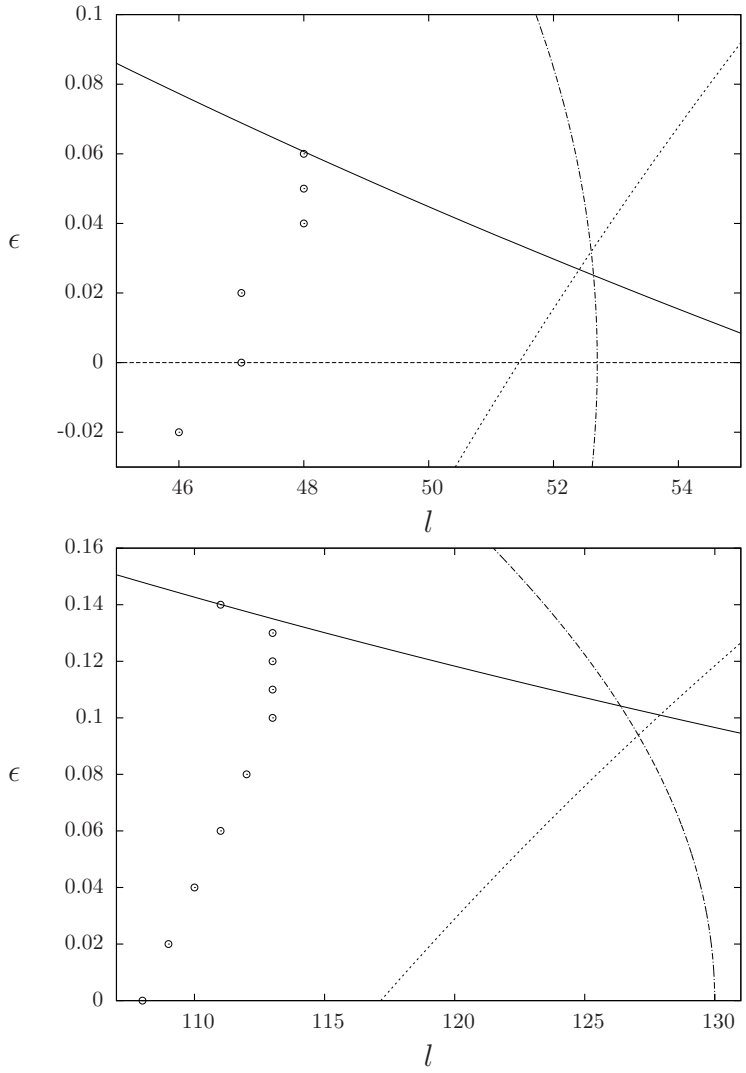

Figure 8.16: Phase diagrams in the (l, ϵ) plane for the bias function $\psi(p) = (1-\epsilon)^p$, for $k = 5$ (top panel) and $k = 6$ (bottom panel). The points are the clustering threshold $l_{\rm d}(\epsilon)$: for a given value of ϵ the RS phase corresponds to $l < l_{\rm d}$, the RSB phase to $l \geq l_{\rm d}$. The three continuous lines are upper bounds of $l_{\rm d}$, the area on their right is necessarily in a RSB phase; the solid one is the Kesten-Stigum bound explained in Sec. 7.7, the dashed line marks the rigidity threshold $l_{\rm r}$ defined in Sec. 7.4, and the dot-dashed line corresponds to the vanishing of the RS entropy (7.42).

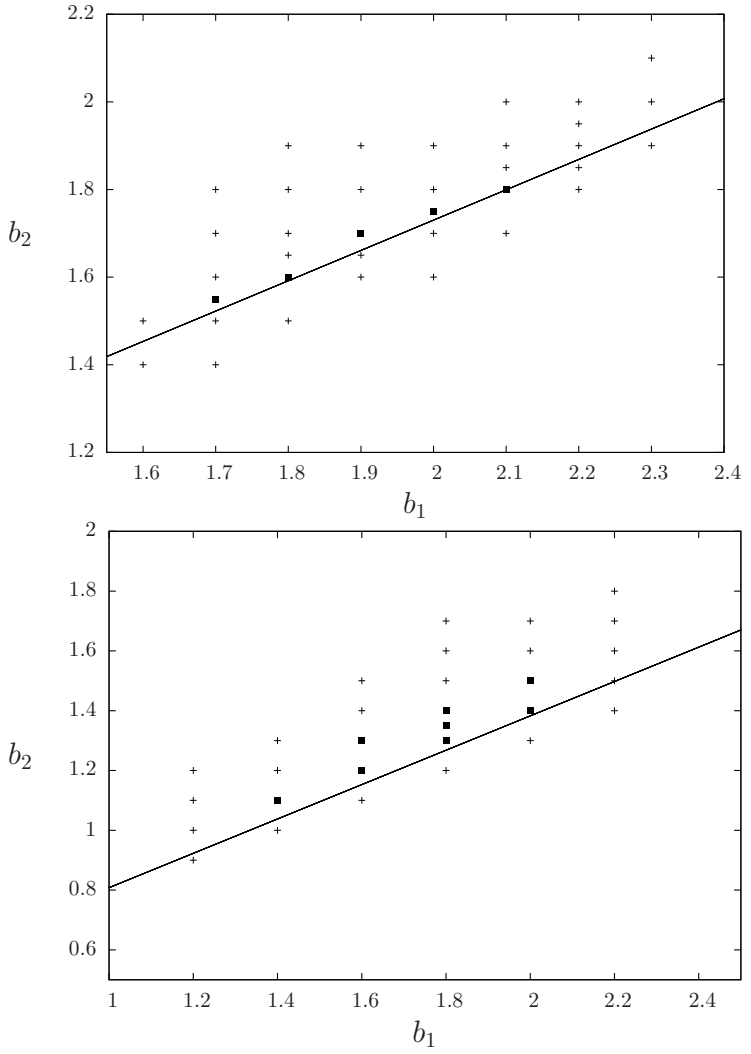

Figure 8.17: The phase diagrams in the (b_1, b_2) parameter plane for the bias function of Eq. (8.27). The top panel is for $k = 5$, $l = 48$ and $\epsilon = 0.06$, the bottom panel for $k = 6$, $l = 114$, $\epsilon = 0.095$. The points marked with squares corresponds to a RS phase ($l < l_{\rm d}(k, b_1, b_2, \epsilon)$), the crosses to a RSB phase ($l \geq l_{\rm d}(k, b_1, b_2, \epsilon)$). The line is the Kesten-Stigum bound, the area below it is RSB.

due to the increased dimensionality of the parameter space. For comparison we give in the last column the satisfiability threshold, i.e. the smallest value of l such that the typical hypergraphs have no proper bicolorings, computed within

the 1RSB ansatz (see [26] for details), that is obviously an upper bound for $l_{\rm d}$, independently of the bias.

k	uniform	ϵ	(b_1, b_2, ϵ)	$l_{\rm sat}$
5	47	48	49	52
6	108	113	115	129

Table 8.2: The values of $l_{\rm d}$ for the uniform measure, for the bias of equation (6.7) that factorizes on the hyperedges, with the optimal value of ϵ, and for the bias of equation (6.12) for well-chosen parameters (b_1, b_2, ϵ). The last column is the satisfiability threshold.

Chapter 9

The asymptotics of the clustering transition for the uniform measure

In the previous chapter we studied the possibility of tuning the value of the clustering threshold $\alpha_d(k)$ by considering a non-uniform (biased) measure on the set of solutions of CSPs. For finite values of k we have seen that well-chosen biases allow indeed to increase $\alpha_d(k)$, thus improving the range of α in which solutions of random CSPs can be found by efficient algorithms. It remains to see wether this improvement survives the large k limit, which requires being able to compute the asymptotics of $\alpha_d(k)$ both in the unbiased and biased cases. The former is the object of this chapter, and the latter will be dealt in chapter 10.

The two examples of CSPs that will appear in this chapter are the graph coloring and the hypergraph bicoloring problems. Let us recall their definition (given in Section 1.1.4). In the first one the variables are placed on the vertices of a graph, they have q possible values, to be interpreted as colors, and each edge of the graph enforce the constraint that the two vertices at its ends take different colors. The hypergraph bicoloring problem (that was studied in the previous chapter) is similarly defined on a hypergraph, with hyperedges linking subsets of k (instead of two for a graph) vertices; the variables on the vertices can take two colors, and the constraint associated to each hyperedge is that both colors are present among its k adjacent vertices. The value of α_d depends of course on the CSP under consideration and of its parameters (k, q in the two examples defined above). We have seen in chapter 7 that it can be estimated numerically by solving a functional equation (see equations (7.537.54) for the k-hypergraph bicoloring problem) that arises from the 1RSB formalism or its tree reconstruction interpretation (see for instance [61] for numerical values in the coloring case, and [85, 103] for the hypergraph bicoloring). There is in general no explicit analytical expression for α_d, but bounds on its value [13, 16] and asymptotic expansions for large k, q [38, 39, 40] complement its numerical deter-

mination (that becomes very difficult when k, q grow). In Section 7.4 we have presented a relatively simple upper bound on α_d can be obtained by analyzing the so-called naive reconstruction procedure [62, 38], where one asks whether the configuration of the variables at a large distance from a root variable implies unambiguously the value of the latter (instead of merely bringing some information on it). This property undergoes a phase transition at the "rigidity" threshold α_r, with obviously $\alpha_d \leq \alpha_r$, which can be alternatively interpreted as the appearance of "hard fields" in the solution of the 1RSB/reconstruction equations, or frozen variables taking the same value in all the solutions of a cluster. The analytic determination of α_r is much simpler than the one of α_d, because it corresponds to the bifurcation of a scalar (instead of functional) fixed-point equation (see equation (7.79) for the k-hypergraph bicoloring problem). In particular, for the hypergraph bicoloring and coloring problems, its asymptotic expansion at large k, q shows that the relevant scale of constraint densities (for the graph coloring problem we use the more natural average degree $c = 2\alpha$) is

$$\alpha = \alpha(k,\gamma) = \frac{2^{k-1}}{k}(\ln k + \ln\ln k + \gamma) \ , \qquad c = c(q,\gamma) = q(\ln q + \ln\ln q + \gamma) \ , \tag{9.1}$$

with γ a finite constant parameter, the rigidity transition α_r occuring on this scale when γ crosses the critical value $\gamma_r = 1$ (for both models). It turns out that the asymptotic behavior of α_d is also on the scale of Eq. (9.1), with another constant $\gamma_d \leq \gamma_r$. This statement follows from a series of rigorous works on this problem: for the coloring problem [38] proved that $1 - \ln 2 \leq \gamma_d \leq \gamma_r = 1$, and the strict inequality $\gamma_d < 1$ was later obtained in [40], implying the asymptotic existence of a regime $[\gamma_d, \gamma_r]$ where reconstruction is possible but naive reconstruction is not. A large family of models, including the two discussed here, was also adressed in [39] which proved $\alpha_d(k) \geq \frac{2^{k-1}}{k} \ln k$, $c_d(q) \geq q \ln q$, confirming the leading term in the scaling (9.1).

We report in this chapter a study of the asymptotic expansion of α_d for the graph coloring and hypergraph bicoloring problems that have been performed in [32], for the uniform measure. We will show that the transition indeed happens on the scale of constraint densities defined in (9.1), with the same value of γ_d for both models, which is to some extent surprising given their different microscopic nature. We characterize γ_d in terms of the behavior of a functional equation, whose numerical study yields the estimate $\gamma_d \approx 0.871$, while an analytical treatment provides the lower bound $\gamma_{\rm d} \geq 1 + \ln(2(\sqrt{2}-1)) \approx 0.812$. Even with the less demanding level of rigor of theoretical physics, to which we stick here, the computation is relatively involved because of the asymptotic proximity of the thresholds α_d and α_r, at the origin of a quite singular behavior of the probability distributions describing the intermediate regime of reconstruction without naive reconstruction.

Chapter is organized as follows. In Sec. 9.1 we perform the large k limit on the hypergraph bicoloring model and reduce the determination of γ_d to the study of a reduced set of equations, in which k does not appear anymore. This

reduced problem takes the form of recursive equations on a sequence of probability distributions, determining a correlation function at distance n. The study of the large n limit of this reduced problem, from which γ_d can be finally deduced, is performed in Sec. 9.2; this limit requires some additional reparametrizations in the interesting regime. For simplicity in the main part of the text we concentrate on the hypergraph bicoloring case, and devote the Appendix B to the graph coloring problem: we show that the large q limit yields exactly the same reduced problem than the large k limit of the hypergraph bicoloring, hence the study of Sec. 9.2 and the determination of γ_d is common to both.

9.1 The large k limit for a finite distance n

9.1.1 Evolution of the hard fields

Before studying the clustering threshold α_d that corresponds to the bifurcation of the functional bifurcation equations (7.537.54) (with $\widetilde{p}(\sigma_1, \dots, \sigma_{k-1}|+) = (1 - \mathbb{I}[\sigma_1 = \cdots = \sigma_{k-1} = +1])/(2^{k-1} - 1)$), it is much easier to consider the rigidity transition $\alpha_r(k)$, that corresponds to the bifurcation of the scalar recursion equation (7.80). We have given in (7.81) the expressions that determine analytically $\alpha_r(k)$ for all k. The value of $\Gamma_{\rm r}(k)$ cannot be given explicitly when k is finite, but it is easy to perform its asymptotic expansion at large k, which yields

$$\alpha_r(k) = \frac{2^{k-1}}{k}(\ln k + \ln\ln k + \gamma_r + o(1)) \ , \qquad \text{with} \ \ \gamma_r = 1 \ . \tag{9.2}$$

As explained above this is an upper bound on the sought for dynamic transition $\alpha_d(k)$, and the rigorous results of [39] suggest (or prove for the coloring problem [38, 40]) that the asymptotic expansion of $\alpha_d(k)$ has the same form as the one of $\alpha_r(k)$, with a different constant term γ_d.

We shall thus study the large k limit taking α to diverge simultaneously with k according to the function $\alpha(k, \gamma)$ defined in Eq. (9.1), with γ a finite constant that becomes our control parameter for the density of constraints in this limit. It is instructive to investigate the behavior of the hard field weights H_n, solution of the recursion equation (7.80), for a finite distance n. With this choice for the scaling of α with k one easily obtains by induction on n that the leading behavior of H_n reads

$$H_n = 1 - \frac{x_n}{k \ln k} + o\left(\frac{1}{k \ln k}\right) \ , \tag{9.3}$$

where x_n is a γ-dependent sequence, defined recursively by

$$x_0 = 0 \ , \qquad x_{n+1} = e^{-\gamma + x_n} \ . \tag{9.4}$$

This recursion function is illustrated on Fig. 9.1. Its study is very simple, and unveils the following behavior for x_n: if $\gamma \geq 1$ then x_n converges at large n to

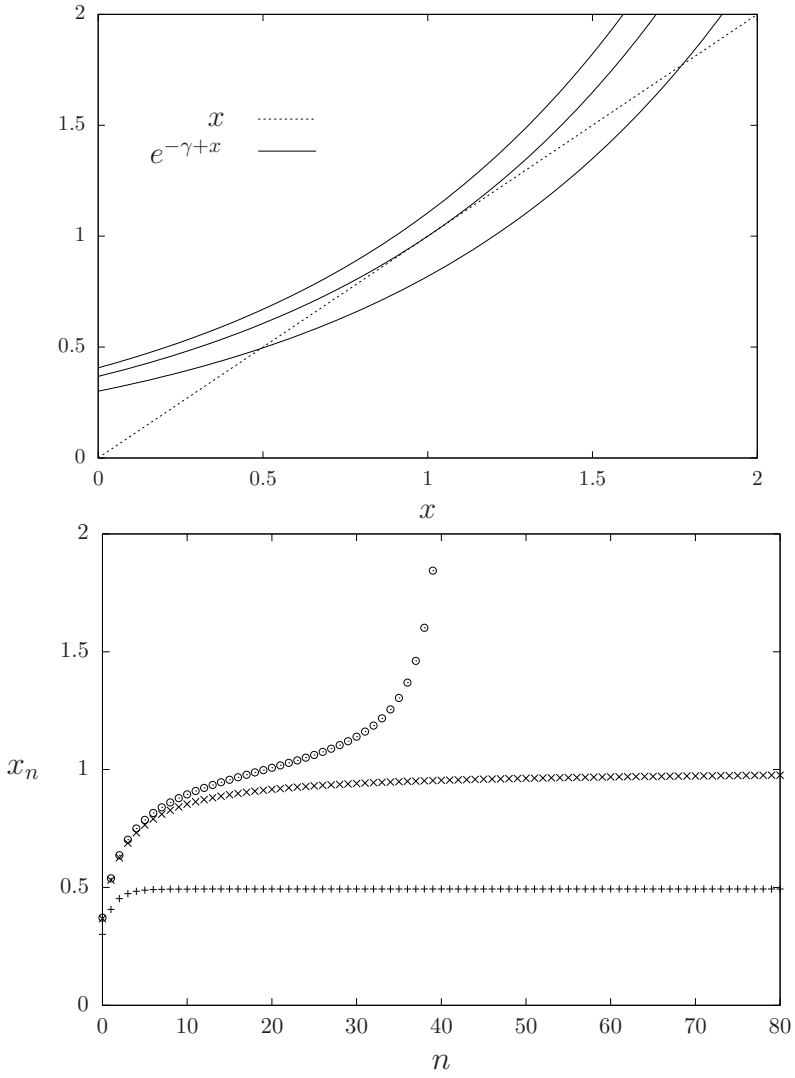

Figure 9.1: Top: the recursion function $e^{-\gamma+x}$ (solid lines) for three values of γ (from top to bottom $\gamma = 0.9$, $\gamma = 1 = \gamma_r$, $\gamma = 1.2$) compared to x (dashed line). Bottom: the iterates x_n defined by $x_0 = 0$, $x_{n+1} = e^{-\gamma+x_n}$ for three values of γ (from top to bottom $\gamma = 0.99$, $\gamma = 1 = \gamma_r$, $\gamma = 1.2$).

a finite fixed point (depicted on the bottom panel of Fig. 9.3) (upper curve), while for $\gamma < 1$ the sequence x_n diverges as $n \to \infty$, very quickly (as iterated exponentials). The transition between these two behaviors thus occurs at the critical value of the parameter $\gamma = \gamma_r = 1$, which corresponds to the expansion of $\alpha_r(k)$ stated above. Note that this coincidence is not as trivial as it may look at first sight: the expansion of Eq. (9.2) was obtained by first taking the limit $n \to \infty$, then $k \to \infty$, while in (9.4) we have taken the limit $k \to \infty$, and only later studied the behavior of x_n for large n. The commutativity of these two limits can in this case be traced back to the simple shape of the recursion equation (7.80): either it has no non-trivial fixed point, or if it has one in the large k limit it must be of order $1 - O(1)/(k \ln k)$, and hence captured by the scaling (9.3).

9.1.2 The reduced order parameter

Our determination of the asymptotic behavior of $\alpha_d(k)$ will be based on three hypotheses that we now spell out explicitly. First of all, we assume that this transition occurs on the scale (9.1), hence that the sought for constant γ_d can be defined as

$$\gamma_d = \lim_{k\to\infty} \left[\frac{k\alpha_d(k)}{2^{k-1}} - \ln k - \ln\ln k \right] \ ; \tag{9.5}$$

as mentioned above the rigorous results of [38, 39, 40] support this hypothesis. We shall furthermore assume that on this scale of α the order parameter $C(\alpha, k)$ is either equal to 0 (for $\alpha < \alpha_d(k)$) or scales with k in the same way as its lower bound provided by the hard field weight $H(\alpha, k)$ (see Eq. (9.3)), hence define a reduced order parameter as

$$\widetilde{C}(\gamma) = \lim_{k\to\infty} (1 - C(\alpha(k,\gamma), k)) k \ln k \ , \tag{9.6}$$

that should be finite for $\gamma > \gamma_d$, and diverge to $+\infty$ for $\gamma < \gamma_d$. Note that in order to have $\widetilde{C}(\gamma) \geq 0$ we introduced a minus sign in this definition, hence increasing values of $\widetilde{C}$ corresponds to smaller correlations in the original order parameter C. Finally we assume that $\widetilde{C}(\gamma)$ can be computed by reversing the order of the large n and large k limits, in other words we define for n finite $\widetilde{C}_n(\gamma)$ as

$$\widetilde{C}_n(\gamma) = \lim_{k\to\infty} (1 - C_n(\alpha(k,\gamma), k)) k \ln k \ , \tag{9.7}$$

and our third hypothesis will consist of computing $\widetilde{C}(\gamma)$ as the large n limit of $\widetilde{C}_n(\gamma)$.

To rephrase these last two hypotheses, we exclude the possibility that on the scale (9.1) the fixed point order parameter $C(\alpha, k)$ has distinct scalings with k depending on the value of γ, and similarly that the finite distance correlation function $C_n(\alpha, k)$ decays with several plateaus at heights each scaling differently with k. These hypotheses are reasonable when confronted with the finite k numerical results (for instance those of Fig. 9.2), and are somehow corroborated

by the scheme of proofs of [38, 39], that rely on the fact that once C_n is "small enough", then it must decay to 0; we shall see in the rest of the paper that they are at least self-consistent. Note also that they hint at the intrisic difficulty of the computation: in the intermediate regime $[\alpha_d(k), \alpha_r(k)]$ there are no hard fields in the fixed point distribution $P(h)$, nevertheless the soft fields become asymptotically hard in the $k \to \infty$ limit as the correlation function C tends to 1 according to (9.6).

9.1.3 Evolution of the soft fields distribution

In order to complete the computation of the reduced order parameter we need now to study the large k limit of the distributions $Q_n(h)$ of the soft fields, that obeys the recursion equations (7.91,7.93). Consider first the latter equation; the integer p that appears there is a random number drawn from the binomial distribution $\mathrm{Bin}(k-1, 1-H_n)$. In the large k limit, because of the scaling (9.3) of H_n, this distribution can be approximated by a Poisson distribution $\mathrm{Po}(x_n/\ln k)$, which hence concentrates on the smallest possible value that appears in the sum, namely $p = 1$. We thus obtain at the leading non-trivial order

$$\widehat{Q}_n(u) = \left[1 - \frac{3x_n}{2^{k-1}\ln k}\right]\delta(u) + \frac{3x_n}{2^{k-1}\ln k}\widehat{R}_n(u) \ , \tag{9.8}$$

where $\widehat{R}_n(u)$ is a normalized distribution arising from the three contributions of the terms with $p = 1$ in (7.93),

$$\widehat{R}_n(u) = \frac{1}{3}\int \mathrm{d}Q_n(h)\left[\delta(u - g_1(h)) + \delta(u - g_1(-h)) + \delta(u + g_1(-h))\right] \ , \tag{9.9}$$

where $g_1(h)$ is obtained from (7.94) as

$$g_1(h) = \frac{\frac{1+h}{2}}{2 - \frac{1+h}{2}} \ . \tag{9.10}$$

Note that this function fulfills the identity $\frac{1-g_1(h)}{1+g_1(h)} = \frac{1-h}{2}$; as a consequence one can check that the Bayes symmetry is respected by $\widehat{R}_n$, i.e. that $\widehat{R}_n(-u) = \frac{1-u}{1+u}\widehat{R}_n(u)$, if Q_n verifies this symmetry.

Let us give an intuitive explanation of the equations (9.8,9.9): in the large k limit the weight of the hard fields H_n is very close to 1, hence almost all of the $k-1$ fields h_i acting on the variables of a hyperedge are forcing them. If they all force them to take the same color -1 then to satisfy the constraint the root must take the opposite color, which leads to an hard field $u = 1$ (the $k-1$ fields cannot be all forcing in the direction $+1$, this would force the root to be -1, in contradiction with the broadcast being conditional on the root equal to $+1$). If they are all hard but with the two colors represented, then the clause is satisfied for every choice of the root, and this produces a trivial soft field $u = 0$. The least unprobable situation that can produce a non-trivial soft field ($u \neq 0, 1$) is thus when among the $k-1$ entering fields h_i exactly one, say h_1, is

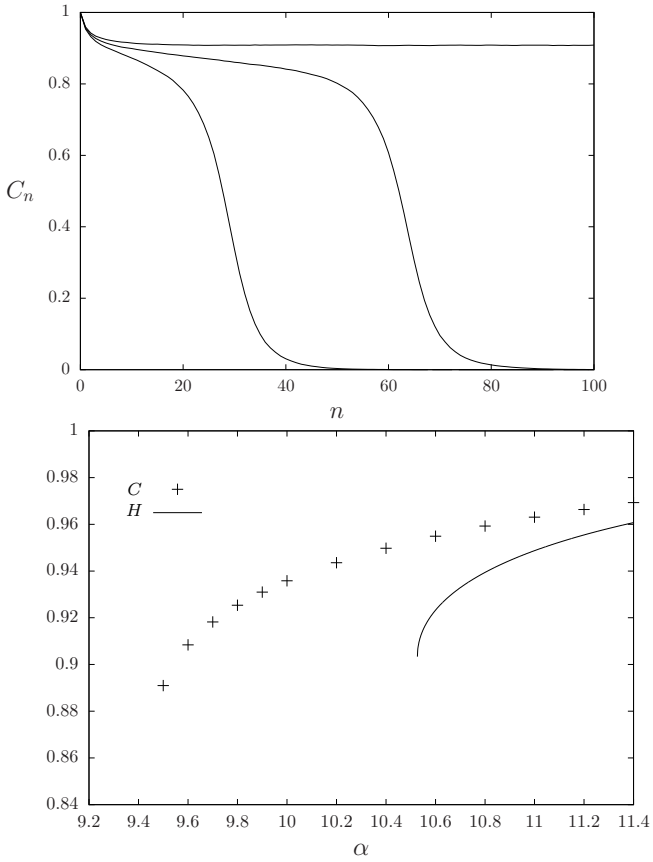

Figure 9.2: Top: the correlation function $C_n(\alpha, k)$ as a function of n for $k = 5$ and from left to right $\alpha = 9.2$, $\alpha = 9.4$, $\alpha = 9.6$. Bottom: the large n limit $C(\alpha, k) = \lim_n C_n(\alpha, k)$ as a function of α for $k = 5$, along with the contribution of the hard fields $H(\alpha, k) = \lim_n H_n(\alpha, k)$ discussed in Sec. 7.4. The first curve is non zero for $\alpha \geq \alpha_d(k = 5) \approx 9.465$, the second one for $\alpha \geq \alpha_r(k = 5) \approx 10.526$.

soft, and all the $k-2$ other are hard. If the two colors appear among the $k-2$ hard fields, again the root is unbiased, this produces a trivial soft field $u=0$. Assume thus that the $k-2$ hard fields are of the same color. If this color is $+1$, then necessarily $\sigma_1=-1$ and this yields $u=-g_1(-h_1)$. If this color is -1, then both $\sigma_1=+1$ (producing $u=g_1(-h_1)$) and $\sigma_1=-1$ (corresponding to $u=g_1(h_1)$) are possible; this concludes the interpretation of (9.8,9.9).

We can now inject the simplified form (9.8) of $\widehat{Q}_n(u)$ in the other recursion equation (7.91); as $f(u_1,\ldots,u_l,0,\ldots,0)=f(u_1,\ldots,u_l)$ the first term of $\widehat{Q}_n(u)$ does not contribute, and the number of $\widehat{R}_n(u_i)$ picked up in the r.h.s. of (7.91) is a Poisson random variable of parameter $\alpha k(1-\widehat{H}_n)\frac{3x_n}{2^{k-1}\ln k}$. This quantity converges to $3x_n$ in the large k limit for the regime we are considering, hence (7.91) becomes:

$$Q_{n+1}(h)=\sum_{l=0}^{\infty}e^{-3x_n}\frac{(3x_n)^l}{l!}\int\prod_{i=1}^{l}\mathrm{d}\widehat{R}_n(u_i)\,\delta(h-f(u_1,\ldots,u_l))\;. \tag{9.11}$$

We can finally express the reduced order parameter $\widetilde{C}_n(\gamma)$ in the large k limit: combining its definition from (9.7), the expression of C_n in terms of the soft fields distribution and of the hard fields weight given in Eq. (7.85), and the scaling of the latter quantity written in (9.3), we obtain $\widetilde{C}_n(\gamma)=x_n\int\mathrm{d}Q_n(h)(1-h)$.

9.2 The limit of large distance n

Let us summarize what we have achieved up to now; starting from the recursion equations (7.53,7.54) for the distributions $P_n(h)$ we have taken the large k limit with the density of constraints scaling as $\alpha(k,\gamma)$ of Eq. (9.1), and argued for the existence of a reduced correlation function at finite distance, $\widetilde{C}_n(\gamma)$ defined in (9.7). We have then shown that the latter could be computed by solving some recursion equations on both x_n, a scalar quantity related to the weight of the hard fields in Eq. (9.3), and on $Q_n(h)$, the distribution of the soft fields (that is a distribution supported on $[-1,1]$, with no atoms in ±1). For the sake of

readability we regroup now all the equations that will be necessary to proceed:

$$x_0 = 0 \ , \qquad x_{n+1} = e^{-\gamma + x_n} \ , \tag{9.12}$$

$$Q_{n+1}(h) = \sum_{l=0}^{\infty} e^{-3x_n} \frac{(3x_n)^l}{l!} \int \prod_{i=1}^{l} \mathrm{d}\widehat{R}_n(u_i)\, \delta(h - f(u_1, \ldots, u_l)) \ , \tag{9.13}$$

$$f(u_1, \ldots, u_l) = \frac{\prod_{i=1}^{l} \frac{1+u_i}{2} - \prod_{i=1}^{l} \frac{1-u_i}{2}}{\prod_{i=1}^{l} \frac{1+u_i}{2} + \prod_{i=1}^{l} \frac{1-u_i}{2}} \ , \qquad Q_1(h) = \delta(h) \ , \tag{9.14}$$

$$\widehat{R}_n(u) = \frac{1}{3} \int \mathrm{d}Q_n(h) \left[\delta(u - g_1(h)) + \delta(u - g_1(-h)) + \delta(u + g_1(-h))\right] \ , \tag{9.15}$$

$$g_1(h) = \frac{\frac{1+h}{2}}{2 - \frac{1+h}{2}} \ , \tag{9.16}$$

$$\widetilde{C}_n(\gamma) = x_n \int \mathrm{d}Q_n(h)(1-h) \ . \tag{9.17}$$

Let us emphasize that the parameter k has disappeared from these equations, that only depend on γ; moreover we show in the Appendix B that exactly the same equations describe the large q limit of the coloring problem. What remains to do now is to study the large n limit of these equations, as the threshold γ_d will be determined according to the behavior of $\widetilde{C}_n(\gamma)$ in this limit, namely $\widetilde{C}_n(\gamma) \to +\infty$ will signal that $\gamma < \gamma_d$, while $\widetilde{C}_n(\gamma)$ will remain finite for $n \to \infty$ if $\gamma > \gamma_d$.

9.2.1 The regime $\gamma \geq \gamma_r = 1$

Let us start by considering the case $\gamma \geq \gamma_r = 1$; then the sequence x_n remains bounded for all n (its limit $x(\gamma)$ being depicted as a function of γ on the bottom panel of Fig. 9.3), which translates the presence of hard fields in the fixed point solution of the corresponding finite k regime $\alpha \geq \alpha_r$. As a consequence the numerical resolution of (9.13) presents no difficulty and can be performed with the usual population dynamics algorithms explained in Sec. 8.1, $Q_n(h)$ being represented as an empirical distribution over a sample of h_i's. Figure 9.3 presents some numerical results obtained in this way; one can see on the top panel that $\widetilde{C}_n(\gamma)$ converges to a finite limit as n grows, this limit value being drawn as a function of γ on the bottom panel. In this regime Q_n converges (weakly) to a stationary distribution Q, solution of the fixed point equation obtained by replacing in (9.13) x_n by its limit.

The finite k inequalities $H_n \leq C_n \leq 1$ becomes in these rescaled units $0 \leq \widetilde{C}_n(\gamma) \leq x_n(\gamma)$; indeed,

$$\int \mathrm{d}Q_n(h)(1-h) = 1 - \int \mathrm{d}Q_n(h) h = 1 - \int \mathrm{d}Q_n(h) h^2 \in [0,1] \ , \tag{9.18}$$

because of the consequence stated in (7.59) of the Bayes symmetry enjoyed by Q_n. In the large n limit we thus have $\widetilde{C}(\gamma) \leq x(\gamma)$, as illustrated in the bottom panel of Fig. 9.3 which is the analog of the bottom panel of Fig. 9.2 for finite k.

Note that $\widetilde{C}(\gamma)$ seems to behave smoothly as $\gamma \to \gamma_r^+$, as expected if the transition occurs at $\gamma_d < \gamma_r$. As x has a square root singularity this means that there is also a square root singularity in $\int \mathrm{d}Q(h)(1-h)$ as a function of γ, in order for their product $\widetilde{C}(\gamma)$ to be regular.

9.2.2 The difficulties for $\gamma < 1$

The study of the large n limit is much more difficult when $\gamma < 1$: the sequence x_n solution of (9.12) is then divergent (asymptotically very strongly, as iterated exponentials). As the direct implementation of the population dynamics procedure to solve (9.13) involves the manipulation of a number of fields u_i's that is distributed as a Poisson random variable of parameter x_n, this becomes very quickly impossible to handle. A first hint of what happens for $\gamma < 1$ can nevertheless be unveiled by this numerical procedure: when γ is only slightly smaller than 1, x_n grows rather slowly at first, the range of n that can be treated in this way is then sufficiently large to gain some useful informations. A typical result obtained in this way is plotted on Fig. 9.4 for $\gamma = 0.96$: even though x_n is diverging it is tempting to conjecture from this plot that $\widetilde{C}_n$ admits a limit when $n \to \infty$. Looking at the expression (9.17) of $\widetilde{C}_n$ one realizes that this is possible provided $\int \mathrm{d}Q_n(h)(1-h)$ tends to 0 as n grows, with an order of magnitude inversely proportional to the one of x_n. This weak convergence of $Q_n(h)$ to $\delta(h-1)$ is compatible with the form of the recursion equation (9.13), the combination of a very large number of u_i's can indeed produce a field h very close to 1.

Lowering further the value of γ the speed of the divergence of x_n increases so much that an extrapolation of $\widetilde{C}_n$ to large n values becomes impossible with this numerical method, which as a consequence does not allow for an accurate estimation of γ_d. We thus need to devise a formalism that incorporates the compensation mechanism between the divergence of x_n and the concentration of Q_n around $\delta(h-1)$ at the origin of the finite value of $\widetilde{C}$. A natural idea would be to make the change of variable $h = 1 - (h'/x_n)$, with h' a random variable that should remain finite and thus obey a simplified recursion equation. Unfortunately the Bayes symmetry enjoyed by Q_n, and its consequences on the moments of h, imply that h' has higher moments that diverge with n. This strategy seems thus rather unpractical. In fact it is more probable that the value of $\widetilde{C}_n$ is determined by contributions of all values of h in $(-1,1)$, even very improbable ones where Q_n is of order $1/x_n$, while the typical values around $h = 1$ do not contribute to $\widetilde{C}_n$ because of the factor $1-h$ in the integral in (9.17). In more technical terms we conjecture that $x_n Q_n$, which is a sequence of positive measures, even if not probability ones because their total mass is the diverging sequence x_n, does converge to a positive measure η, in the sense of

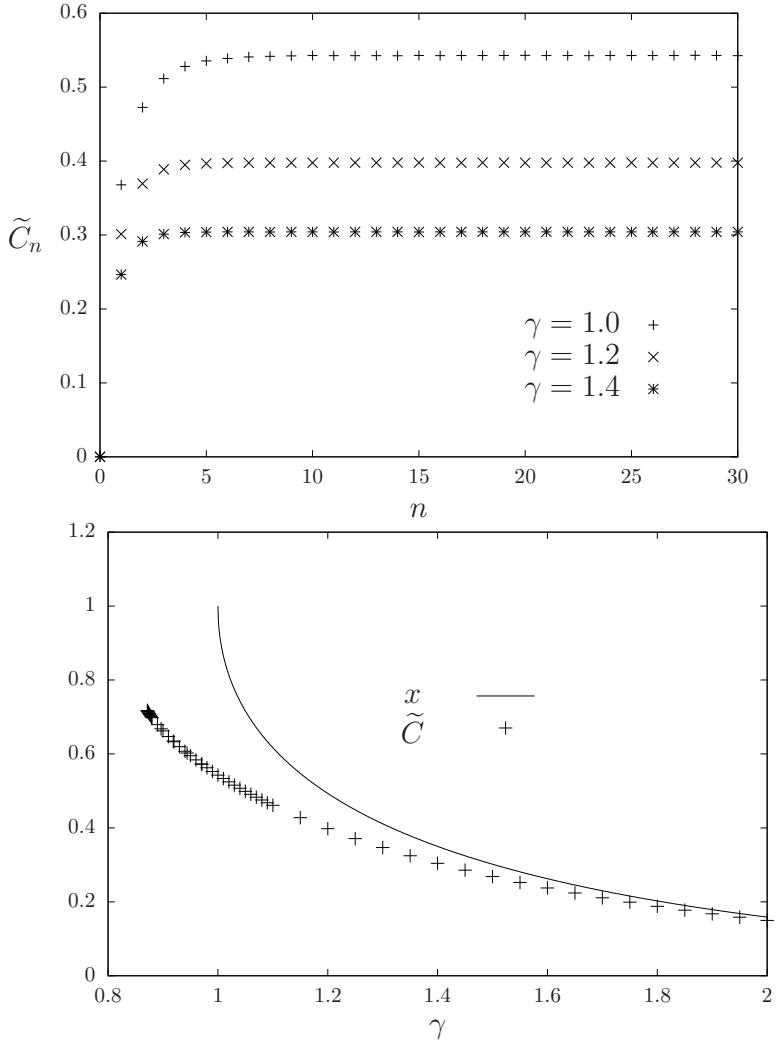

Figure 9.3: Top: the reduced correlation function $\widetilde{C}_n(\gamma)$ as a function of n for a few values of $\gamma \geq \gamma_r = 1$. Bottom: the large n limit $\widetilde{C}(\gamma) = \lim_n \widetilde{C}_n(\gamma)$ as a function of γ, along with the contribution of the hard fields $x(\gamma) = \lim_n x_n(\gamma)$. The first curve exists for $\gamma \geq \gamma_d \approx 0.871$, the second one for $\gamma \geq \gamma_r = 1$. The data for $\gamma < 1$ have been obtained through the reweighted algorithm expained in Sec. 9.2.3.

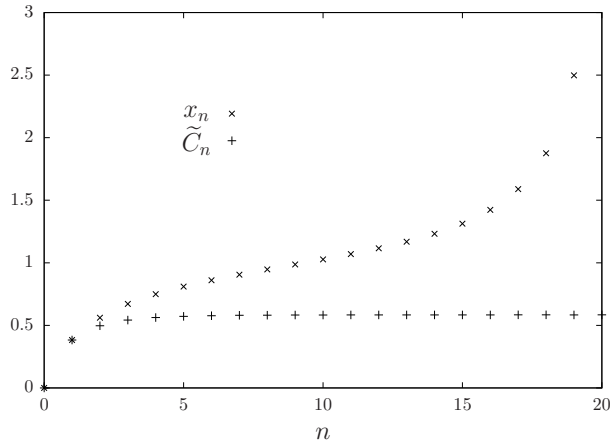

Figure 9.4: The evolution of $\widetilde{C}_n$ and x_n for $\gamma = 0.96 < \gamma_r$. The sequence x_n diverges asymptotically in a very rapid way (as iterated exponentials, the vertical range has been cut for the sake of readability) while the reduced correlation function $\widetilde{C}_n$ tends to a positive constant. This is the analog of the finite k regime $\alpha_d < \alpha < \alpha_r$ depicted in Fig. 7.4.

the vague convergence on $(-1, 1)$, i.e.

$$\lim_{n\to\infty} x_n \int_a^b \mathrm{d}Q_n(h) = \int_a^b \mathrm{d}\eta(h) \tag{9.19}$$

for all $-1 < a < b < 1$ which are continuity points of η. A direct numerical test of this conjecture, and the determination of η, seems rather difficult as η is not a probability measure and may have an infinite total mass. We have thus devised an alternative numerical procedure that allowed us to explore the regime $\gamma < 1$, as we shall now explain.

9.2.3 Reweighted probability distributions

The idea underlying this procedure is to study the evolution not of the probability measure Q_n, but of a reweighted (or tilted) version of it, that give less importance to the typical (for Q_n) values of h around $h = 1$, which do not contribute to $\widetilde{C}_n$. Let us thus consider a positive function $r_n(h)$, and define the reweighted distribution $\rho_n(h) = x_n r(h) Q_n(h)$. In more precise mathematical terms ρ_n is absolutely continuous with respect to Q_n, with relative density $x_n r$. For the reasons explained above our goal is to use a function $r(h)$ that vanishes in $h = 1$; a convenient choice is to take r of the form $r(h) = ((1-h)/(1+h))^t$, with t an exponent that is arbitrary for the moment. Indeed with this choice the reweighting will factorize in (9.13), in the sense that $r(f(u_1, \ldots, u_l)) = r(u_1) \times \cdots \times r(u_l)$. We will actually further specify the function by taking the exponent $t = 1/2$,

thus setting

$$\rho_n(h) = x_n Q_n(h) \left(\frac{1-h}{1+h}\right)^{\frac{1}{2}} \quad . \tag{9.20}$$

The reasons for this choice of the exponent are two fold: the Bayes symmetry (7.58) enjoyed by Q_n implies then that ρ_n is symmetric in the usual sense, i.e. $\rho_n(-h) = \rho_n(h)$. Furthermore (7.58) also implies the following identity (used in the proof of lemma 4.4 in [39]),

$$\int \mathrm{d}Q_n(h) \left(\frac{1-h}{1+h}\right)^{\frac{1}{2}} = \int \mathrm{d}Q_n(h) \sqrt{1-h^2} \ , \tag{9.21}$$

which can be obtained from (7.58) with the choice $b(h) = h((1-h)/(1+h))^{1/2}$. The right hand side of (9.21) is obviously between 0 and 1, thus for all finite n the measure ρ_n has a finite total mass, smaller than x_n. We shall therefore define a probability measure ν_n by dividing ρ_n by its mass m_n:

$$m_n = \int \mathrm{d}\rho_n(h) = x_n \int \mathrm{d}Q_n(h) \left(\frac{1-h}{1+h}\right)^{\frac{1}{2}} \ , \qquad \nu_n(h) = \frac{1}{m_n}\rho_n(h) \ . \tag{9.22}$$

Note that the three descriptions in terms of the (finite but not normalized to 1) positive measure ρ_n, or the pair (m_n, ν_n), or the original one in terms of (x_n, Q_n), are strictly equivalent for all finite n; in particular one can compute the reduced correlation function as

$$\widetilde{C}_n(\gamma) = \int \mathrm{d}\rho_n(h) \sqrt{1-h^2} = m_n \int \mathrm{d}\nu_n(h) \sqrt{1-h^2} \ . \tag{9.23}$$

Besides its equivalence with the original quantities at finite n the advantage of the reweighted description in terms of ρ_n is a much nicer behavior when n grows, as we shall see.

Let us now derive the recursion equations that governs the evolution of ρ_n (or equivalently of the pair (m_n, ν_n)). It will be more convenient to express some quantities in terms of $\widetilde{h} = \mathrm{arcth}(h)$ and $\widetilde{u} = \mathrm{arcth}(u)$, that corresponds to the effective magnetic fields associated to the magnetizations h and u. In particular the operation $h = f(u_1, \ldots, u_l)$ translates in this domain to a simple addition, $\widetilde{h} = \widetilde{u}_1 + \cdots + \widetilde{u}_l$, and the reweighting $r(h)$ becomes $e^{-\widetilde{h}}$. For the sake of conciseness we shall take the liberty of using simultaneously both types of quantities in the following equations, keeping implicit the relation $\widetilde{h} = \mathrm{arcth}(h)$ between them; we will also denote with the same symbol Q_n the measure for the random variables h and $\widetilde{h}$ related in this way.

Let us first rewrite the equations (9.13,9.15) in terms of $\widetilde{h}$:

$$Q_{n+1}(\widetilde{h}) = \sum_{l=0}^{\infty} e^{-3x_n} \frac{(3x_n)^l}{l!} \int \prod_{i=1}^{l} \mathrm{d}\widehat{R}_n(\widetilde{u}_i) \, \delta(\widetilde{h} - \widetilde{u}_1 - \cdots - \widetilde{u}_l)) \ , \tag{9.24}$$

$$\widehat{R}_n(\widetilde{u}) = \frac{1}{3} \int \mathrm{d}Q_n(\widetilde{h}) \left[\delta(\widetilde{u} - a(\widetilde{h})) + \delta(\widetilde{u} - a(-\widetilde{h})) + \delta(\widetilde{u} + a(-\widetilde{h}))\right] \ , \tag{9.25}$$

where we introduced the function

$$a(\widetilde{h}) = \mathrm{arcth}(g_1(\tanh(\widetilde{h}))) = \frac{1}{2}\ln\left(1+e^{2\widetilde{h}}\right) \ . \tag{9.26}$$

Under Q_{n+1} the random variable $\widetilde{h}$ has a compound Poisson distribution, it is thus more convenient to describe it in terms of its characteristic function (i.e. Fourier transform). We define the latter for the various measures as

$$\begin{aligned} \widehat{Q}_n(z) = \int \mathrm{d}Q_n(h) e^{iz\widetilde{h}} \ , \qquad \widehat{\rho}_n(z) = \int \mathrm{d}\rho_n(h) e^{iz\widetilde{h}} \ , \\ \widehat{\nu}_n(z) = \int \mathrm{d}\nu_n(h) e^{iz\widetilde{h}} = \frac{1}{m_n}\widehat{\rho}_n(z) \ . \end{aligned} \tag{9.27}$$

With these conventions the two equations (9.24,9.25) become:

$$\widehat{Q}_{n+1}(z) = \exp\left[-3x_n + x_n \int \mathrm{d}Q_n(h) \left(e^{iza(\widetilde{h})} + e^{iza(-\widetilde{h})} + e^{-iza(-\widetilde{h})}\right)\right] \ . \tag{9.28}$$

The effect of the reweighting is easily seen to translate in terms of the characteristic functions as a shift of the argument,

$$\rho_{n+1}(h) = x_{n+1} e^{-\widetilde{h}} Q_{n+1}(h) \quad \Leftrightarrow \quad \widehat{\rho}_{n+1}(z) = x_{n+1}\widehat{Q}_{n+1}(z+i) \ . \tag{9.29}$$

Recalling that $x_{n+1} = \exp[-\gamma + x_n]$ we thus obtain

$$\begin{aligned} \widehat{\rho}_{n+1}(z) = \exp\Bigg[-\gamma + x_n \int \mathrm{d}Q_n(h) \Big(e^{-a(\widetilde{h})} e^{iza(\widetilde{h})} + e^{-a(-\widetilde{h})} e^{iza(-\widetilde{h})} \\ + e^{a(-\widetilde{h})} e^{-iza(-\widetilde{h})} - 2\Big)\Bigg] \end{aligned} \tag{9.30}$$

Replacing in this equation $x_n Q_n$ by $e^{\widetilde{h}}\rho_n$ yields

$$\begin{aligned} \widehat{\rho}_{n+1}(z) &= \exp\left[-\gamma + \int \mathrm{d}\rho_n(h) \left(e^{\widetilde{h}-a(\widetilde{h})} e^{iza(\widetilde{h})} + e^{\widetilde{h}-a(-\widetilde{h})} e^{iza(-\widetilde{h})} \right.\right. \\ &\quad \left.\left. + e^{\widetilde{h}+a(-\widetilde{h})} e^{-iza(-\widetilde{h})} - 2e^{\widetilde{h}}\right)\right] \qquad (9.31) \\ &= \exp\left[-\gamma + \int \mathrm{d}\rho_n(h) \left(\sqrt{\frac{1+h}{2}} e^{iza(\widetilde{h})} \right.\right. \\ &\quad \left.\left. + \sqrt{\frac{1+h}{1-h}} \left(\sqrt{\frac{1+h}{2}} e^{iza(-\widetilde{h})} + \sqrt{\frac{2}{1+h}} e^{-iza(-\widetilde{h})} - 2\right)\right)\right] \qquad (9.32) \end{aligned}$$

where in the second line we expressed the weight factors in terms of h instead of $\widetilde{h}$, using the expression of $a(\widetilde{h})$ given in (9.26). As $\rho_n(h) = \rho_n(-h)$ one can

symmetrize the integrand to obtain

$$\widehat{\rho}_{n+1}(z) = \exp\left[-\gamma + \int \mathrm{d}\rho_n(h) \left(\frac{1}{\sqrt{2(1+h)}} \left(e^{iza(\widetilde{h})} + e^{-iza(\widetilde{h})} \right) \right.\right.$$
$$\left.\left. + \frac{1}{\sqrt{2(1-h)}} \left(e^{iza(-\widetilde{h})} + e^{-iza(-\widetilde{h})} \right) - \frac{2}{\sqrt{1-h^2}} \right) \right]$$
$$= \exp\left[-\gamma + \int \mathrm{d}\rho_n(h) \left(\sqrt{\frac{2}{1-h}} \left(e^{iza(-\widetilde{h})} + e^{-iza(-\widetilde{h})} \right) - \frac{2}{\sqrt{1-h^2}} \right) \right] . \tag{9.33}$$

These two expressions for $\widehat{\rho}_{n+1}(z)$ are obviously invariant under $z \to -z$, which enforces the conservation of the symmetry $\rho_{n+1}(h) = \rho_{n+1}(-h)$ along the iterations. The integrand of the first one is explicitly symmetric in h, while the more compact second one has been obtained by exploiting the symmetry of ρ_n.

We have thus obtained a recursion equation for the reweighted measure ρ_n, that is equivalent to the original system on (x_n, Q_n). It will be useful to spell out its translation in terms of the mass m_n of ρ_n and its normalized version ν_n. By setting $z = 0$ in (9.33) we obtain

$$m_{n+1} = \exp\left[-\gamma + \int \mathrm{d}\rho_n(h)\beta(h)\right] , \tag{9.34}$$

$$\text{with} \quad \beta(h) = \sqrt{\frac{2}{1+h}} + \sqrt{\frac{2}{1-h}} - \frac{2}{\sqrt{1-h^2}} . \tag{9.35}$$

Note that $\beta(h) \leq 1$ for $h \in [-1, 1]$, hence the integral in (9.34) is smaller than m_n, which allows to check inductively that $m_n \leq x_n$, in agreement with the direct derivation of this bound from (9.21). Dividing the expression (9.33) of $\widehat{\rho}_{n+1}(z)$ by its total mass m_{n+1} we obtain the characteristic function $\widehat{\nu}_{n+1}(z)$ of ν_{n+1}, the normalized probability distribution. Replacing ρ_n by $m_n\nu_n$ we thus have a recursion equation on the pair (m_n, ν_n), namely

$$m_{n+1} = \exp\left[-\gamma + m_n \int \mathrm{d}\nu_n(h)\beta(h)\right] , \tag{9.36}$$

$$\widehat{\nu}_{n+1}(z) = \exp\left[m_n \int \mathrm{d}\nu_n(h) \sqrt{\frac{2}{1-h}} \left(e^{iza(-\widetilde{h})} + e^{-iza(-\widetilde{h})} - 2 \right)\right] , \tag{9.37}$$

completed by the initial condition $m_1 = e^{-\gamma}$, $\nu_1(h) = \delta(h)$, that derives directly from $x_1 = e^{-\gamma}$, $Q_1(h) = \delta(h)$.

9.2.4 Numerical resolution

There is one last difficulty to face in the derivation of a numerical procedure able to solve the equations (9.36,9.37), that we shall now explain. As ν_n is a normalized probability measure we can represent it as a large sample of fields,

according to the population dynamics strategy explained in Sec. 8.1. Suppose this has been done up to the n-th iteration, and that an estimate of the real number m_n is also known. Then m_{n+1} can be easily computed from (9.36), the integration over ν_n being estimated as an empirical average over the representative population. There remains to give an interpretation of the law ν_{n+1} whose characteristic function is given in (9.37), and to devise a sampling procedure from it in order to produce the population of i.i.d. samples that will represent ν_{n+1}. In order to do so let us introduce a positive (not normalized) measure π_n and its total mass λ_n according to

$$\pi_n(y) = m_n \int \mathrm{d}\nu_n(h) \sqrt{\frac{2}{1-h}} (\delta(y - a(-\widetilde{h})) + \delta(y + a(-\widetilde{h}))) \tag{9.38}$$

$$\lambda_n = 2m_n \int \mathrm{d}\nu_n(h) \sqrt{\frac{2}{1-h}} \ , \tag{9.39}$$

in such a way that (9.37) can be rewritten

$$\widehat{\nu}_{n+1}(z) = \exp\left[\int \mathrm{d}\pi_n(y)(e^{izy} - 1) \right] \ . \tag{9.40}$$

In more technical terms π_n is the Lévy measure associated to the infinitely divisible distribution ν_{n+1} [109], with λ_n the total mass of π_n, which is finite for a Poisson compound distribution. This form makes clear that under ν_{n+1} the random variable $\widetilde{h}$ can be described as the distributional equality $\widetilde{h} \stackrel{d}{=} \sum_{i=1}^{l} y_i$, where in the right hand side l is a Poisson random variable of mean λ_n, and the y_i's are i.i.d. copies drawn with the probability measure $\pi_n(y)/\lambda_n$. If ν_n is known as a sample of fields it is possible to sample efficiently y from π_n/λ_n, by extracting a field $h_i^{(n)}$ with a probability proportional to $(1 - h_i^{(n)})^{-1/2}$ and then setting $y = \pm a(-\widetilde{h}_i^{(n)})$, the two signs being chosen with probability 1/2. It could thus seem at first sight that the distributional interpretation given above leads to a valid numerical procedure to solve (9.36,9.37). A closer look reveals that this strategy cannot work as it is for $\gamma < 1$. Let us indeed rewrite the expression of λ_n by exploiting the symmetry of ν_n:

$$\begin{aligned} \lambda_n &= m_n \int \mathrm{d}\nu_n(h) \left(\sqrt{\frac{2}{1-h}} + \sqrt{\frac{2}{1+h}} \right) \\ &= m_n \int \mathrm{d}\nu_n(h) \frac{\sqrt{2}(\sqrt{1+h} + \sqrt{1-h})}{\sqrt{1-h^2}} \ . \end{aligned} \tag{9.41}$$

On the other hand the sequence x_n can be expressed from the definition (9.20) of ρ_n as

$$\begin{aligned} x_n &= \int \mathrm{d}\rho_n(h) \sqrt{\frac{1+h}{1-h}} = m_n \int \mathrm{d}\nu_n(h) \frac{1}{2} \left(\sqrt{\frac{1+h}{1-h}} + \sqrt{\frac{1-h}{1+h}} \right) \\ &= m_n \int \mathrm{d}\nu_n(h) \frac{1}{\sqrt{1-h^2}} \ , \end{aligned} \tag{9.42}$$

where we used the normalization of Q_n and the symmetry of ν_n. Noting that the function $\sqrt{2}(\sqrt{1+h}+\sqrt{1-h})$ is bounded between 2 and $2\sqrt{2}$ when $h \in [-1,1]$ we can conclude by comparing these two expressions of λ_n and x_n that $2x_n \leq \lambda_n \leq 2\sqrt{2}x_n$. Hence for $\gamma < 1$ the lower bound implies that λ_n diverges with n, which brings us back to the situation we are trying to avoid of a Poisson compound distribution with a diverging number of summands.

Fortunately the reweighting performed above will allow us to circumvent this difficulty. One sees indeed on the expression (9.39) that the divergent contribution to λ_n arises from the neighborhood of $h = 1$, i.e. of $\widetilde{h} = +\infty$. However the corresponding summands $\pm a(-\widetilde{h})$ are very small (recall the expression of a from Eq. (9.26)), it is thus conceivable that such a sum of a very large number of very small contributions can be handled in a simplified way. To formalize this intuition we rewrite (9.37) as

$$\widehat{\nu}_{n+1}(z) = \widehat{\nu}^{(\leq)}_{n+1}(z)\,\widehat{\nu}^{(>)}_{n+1}(z)\ , \tag{9.43}$$

$$\widehat{\nu}^{(\leq)}_{n+1}(z) = \exp\left[\int \mathrm{d}\rho_n(h)\sqrt{\frac{2}{1-h}}\left(e^{iza(-\widetilde{h})}+e^{-iza(-\widetilde{h})}-2\right)\mathbb{I}(h\leq 1-\varepsilon_n)\right]\ , \tag{9.44}$$

$$\widehat{\nu}^{(>)}_{n+1}(z) = \exp\left[\int \mathrm{d}\rho_n(h)\sqrt{\frac{2}{1-h}}\left(e^{iza(-\widetilde{h})}+e^{-iza(-\widetilde{h})}-2\right)\mathbb{I}(h> 1-\varepsilon_n)\right]\ , \tag{9.45}$$

where ε_n is for the moment an arbitrary threshold; such a decomposition is loosely inspired by an analogy with the Lévy-Itô decomposition of Lévy processes [109].

These two contributions to $\widehat{\nu}_{n+1}(z)$ are characteristic functions of probability distributions, hence under the law ν_{n+1} the random variable $\widetilde{h}$ is the sum of independent draws from $\nu^{(\leq)}_{n+1}$ and $\nu^{(>)}_{n+1}$. The first one can be described as above, modulo the introduction of the cutoff ε_n; defining

$$\pi^{(\leq)}_n(y) = m_n \int \mathrm{d}\nu_n(h)\sqrt{\frac{2}{1-h}}(\delta(y-a(-\widetilde{h}))+\delta(y+a(-\widetilde{h})))\mathbb{I}(h\leq 1-\varepsilon_n)\ , \tag{9.46}$$

$$\lambda^{(\leq)}_n = 2m_n \int \mathrm{d}\nu_n(h)\sqrt{\frac{2}{1-h}}\,\mathbb{I}(h\leq 1-\varepsilon_n)\ , \tag{9.47}$$

one sees that under the law $\nu^{(\leq)}_{n+1}$ the random variable $\widetilde{h}$ satisfies the distributional equality $\widetilde{h} \stackrel{d}{=} \sum_{i=1}^{l} y_i$, where in the right hand side l is a Poisson random variable of mean $\lambda^{(\leq)}_n$ and the y_i are i.i.d. with the law $\pi^{(\leq)}_n/\lambda^{(\leq)}_n$. The law of $\nu^{(>)}_{n+1}$ will instead be described via its cumulants, to be denoted $c_{n+1,p}$ for the p-th one. By taking derivatives with respect to z of $\ln \widehat{\nu}^{(>)}_{n+1}(z)$ one finds easily

that

$$c_{n+1,p} = m_n \int \mathrm{d}\nu_n(h) \sqrt{\frac{2}{1-h}} (1 + (-1)^p)\, a(-\widetilde{h})^p\, \mathbb{I}(h > 1 - \varepsilon_n) \ . \tag{9.48}$$

As expected from the symmetry of the law $\nu_{n+1}^{(>)}$ all its cumulants of odd order vanish and one can write for the even ones:

$$c_{n+1,2p} = 2m_n \int \mathrm{d}\nu_n(h) \sqrt{\frac{2}{1-h}}\, a(-\widetilde{h})^{2p}\, \mathbb{I}(h > 1 - \varepsilon_n) \ . \tag{9.49}$$

The strategy we have followed in our numerical resolution is to choose ε_n large enough such that $\lambda_n^{(\leq)}$ does not grow with n, in such a way that the contribution from $\nu_{n+1}^{(\leq)}$ can be generated with a finite Poissonian number of summands, but small enough so that $\nu_{n+1}^{(>)}$ can be safely approximated as a Gaussian distribution, neglecting the cumulants $c_{n+1,2p}$ for $p > 1$ (there exist some ways to draw better approximations of infinitely divisible distributions by using more cumulants [110, 111], but we did not try to implement them). Let us give an explicit description of the algorithm. At the n-th iteration we assume to have an estimation of m_n and of ν_n as a population, namely

$$\nu_n(h) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \delta(\widetilde{h} - \widetilde{h}_i^{(n)}) \ . \tag{9.50}$$

We will assume that the population has been sorted, $\widetilde{h}_1^{(n)} \leq \widetilde{h}_2^{(n)} \leq \cdots \leq \widetilde{h}_{\mathcal{N}}^{(n)}$, and translate the cutoff ε_n by defining $\mathcal{N}_n$ in such a way that $h_{\mathcal{N}_n}^{(n)} \leq 1 - \varepsilon_n < h_{\mathcal{N}_n+1}^{(n)}$. We can thus estimate all integrals with respect to ν_n according to

$$\int \nu_n(h) F(\widetilde{h}) \approx \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} F(\widetilde{h}_i^{(n)}) \tag{9.51}$$

for all functions F, and in particular this gives us m_{n+1} from (9.36). We then compute

$$\begin{aligned} \lambda_n^{(\leq)} &\approx 2m_n \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}_n} \sqrt{\frac{2}{1-h_i^{(n)}}} \ , \\ c_{n+1,2} &\approx 2m_n \frac{1}{\mathcal{N}} \sum_{i=\mathcal{N}_n+1}^{\mathcal{N}} \sqrt{\frac{2}{1-h_i^{(n)}}}\, a(-\widetilde{h}_i^{(n)})^2 \ . \end{aligned} \tag{9.52}$$

We then construct the new population elements $\widetilde{h}_i^{(n+1)}$ by repeating $\mathcal{N}$ times, independently for $i = 1, \ldots, \mathcal{N}$:

- draw a random number l from a Poisson distribution of parameter $\lambda_n^{(\leq)}$;

- draw $i_1, \dots, i_l$ i.i.d. in $\{1, \dots, \mathcal{N}_n\}$ with probability proportional to $\frac{1}{\sqrt{1-h_i^{(n)}}}$ (this can be done efficiently by precomputing a cumulative table);
- set $\widetilde{h}_i^{(n+1)} = \delta_1 \, a(-\widetilde{h}_{i_1}^{(n)}) + \cdots + \delta_l \, a(-\widetilde{h}_{i_l}^{(n)}) + \sqrt{c_{n+1,2}} z$ where the δ_i's are ± 1 with probability $1/2$, independently from each other and from anything else, and z is a standard Gaussian random variable.

We can then compute the reduced correlation function $\widetilde{C}_{n+1}$ from (9.23), and sort the elements of the new population to be ready for the next step.

In our implementation we chose the threshold ε_n, or equivalently $\mathcal{N}_n$, in an adaptive way: at each step we take the largest value of $\mathcal{N}_n$ that maintains $\lambda_n^{(\leq)}$ computed from (9.52) smaller than a value λ fixed beforehand and kept constant along the iterations. The numerical accuracy of our procedure is thus limited by the finiteness of $\mathcal{N}$ (the representation (9.50) becomes exact only in the limit $\mathcal{N} \to \infty$) and of λ (because of the Gaussian approximation we perform for $\nu_{n+1}^{(>)}$). The numerical results presented in figures have been obtained with $\mathcal{N} = 10^7$ and $\lambda = 20$; we checked that varying λ between 10 and 30, or reducing $\mathcal{N}$ to 10^6, did not affect our conclusions in a quantitative way.

On the top panel of Fig. 9.5 we show that this new method allows us to track the evolution of $\widetilde{C}_n$ for values of γ deep in the regime $\gamma < 1$ that was unaccessible to the simplest strategy presented in Sec. 9.2.1. In particular for $\gamma \gtrsim 0.88$ there is a clear plateau in $\widetilde{C}_n$, the value of which was reported as a function of γ in the bottom panel of Fig. 9.3 along with the results previously obtained for $\gamma \geq 1$. Note that the new algorithm is also valid for $\gamma > 1$ (and in this case the cutoff ε_n can be safely put to 0), we found as expected a perfect agreement between the results of the two procedures in this case.

On the bottom panel of Fig. 9.5 we have plotted the evolution of (the inverse of) m_n as a function of n, that presents the same kind of plateau behavior as $\widetilde{C}_n$. These numerical results suggest that $\widetilde{C}_n$ diverges with n if and only if m_n diverges, and hence that γ_d can be defined as the smallest value of γ such that m_n remains bounded when $n \to \infty$. We have not been able to prove analytically this statement, that will be taken as an additional hypothesis in the rest of the paper. This conjecture could be wrong if there existed a regime of γ such that $\widetilde{C}_n$ remains finite while m_n diverges, because of the existence of another reweighting scheme that deals with the compensation between the divergence of x_n and the convergence of $Q_n(h)$ to $\delta(h-1)$ in a more efficient way than the one we introduced in (9.20). We found numerically no trace of such a phenomenon, which if it existed would only modify our estimate of γ_d by a quantitatively very small amount, as for $\gamma = 0.87$ the finite n study strongly suggests a common divergence of $\widetilde{C}_n$ and m_n (and for $\gamma = 0.86$ the divergences are obvious on Fig. 9.5).

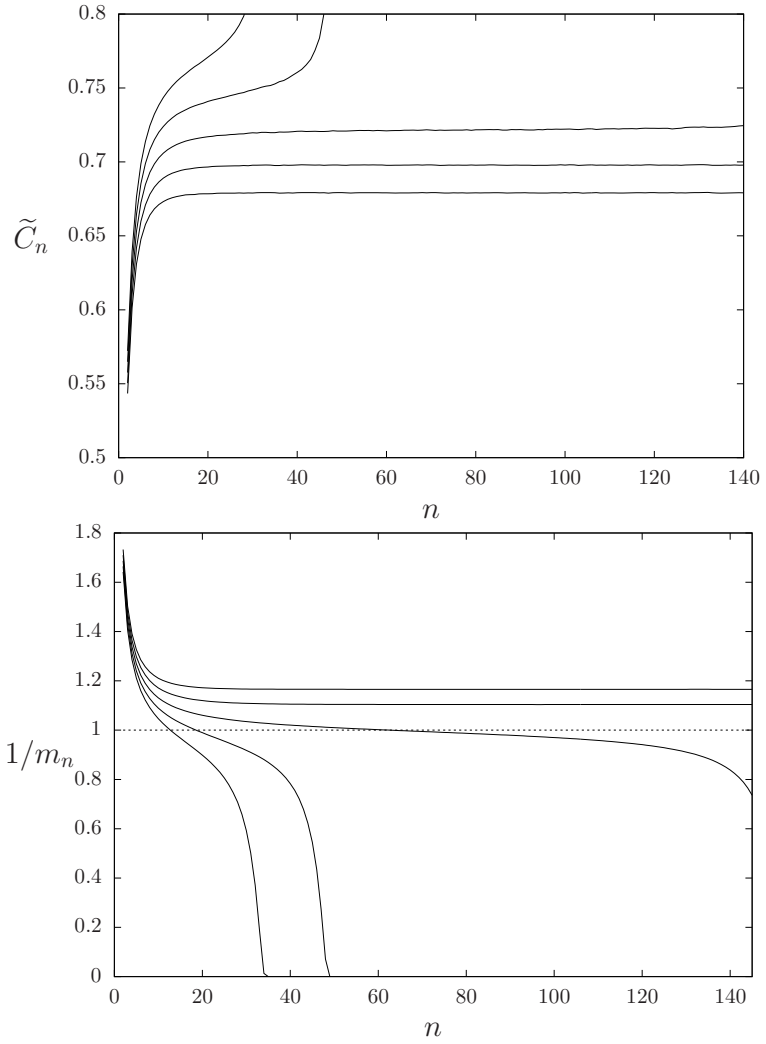

Figure 9.5: The results of the reweighted distribution method in the neighborhood of γ_d: on the top panel the reduced order parameter $\widetilde{C}_n(\gamma)$ as a function of n, on the bottom panel the inverse of the mass m_n of the reweighted measure ρ_n. On both panels the five curves correspond to $\gamma = 0.85$, $\gamma = 0.86$, $\gamma = 0.87$, $\gamma = 0.88$, $\gamma = 0.89$, from top to bottom on the top panel and from bottom to top on the bottom panel. These curves show that $0.87 < \gamma_d < 0.88$, $\widetilde{C}_n$ and m_n diverge for $\gamma = 0.87$, even if it is not completely visible on the displayed range of n.

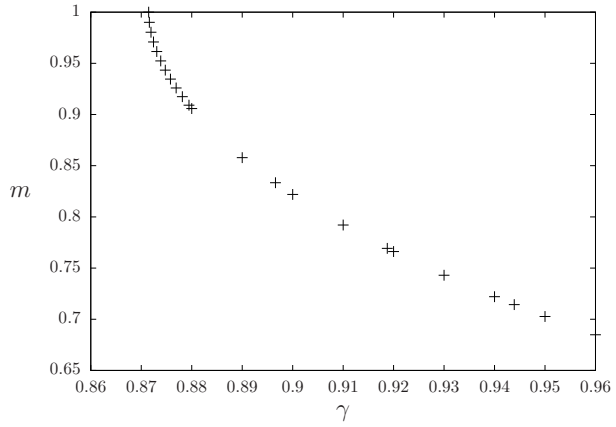

Figure 9.6: The relationship between the parameter γ and the mass m of the reweighted measure ρ in the $n \to \infty$ limit. The numerical results suggest that m reaches 1 with a square root behavior as γ tends to γ_d.

9.2.5 The fixed point equation and the determination of γ_d

The numerical results presented above suggest that for $\gamma \geq \gamma_d$ both m_n and $\widetilde{C}_n$ converge when $n \to \infty$. It is then tempting to conjecture that the probability distribution ν_n itself converges to some ν, in other words that we can remove the indices n from (9.36,9.37) and get a system of self-consistent equations on the fixed-point (m, ν) as:

$$m = \exp\left[-\gamma + m \int \mathrm{d}\nu(h)\beta(h)\right] , \tag{9.53}$$

$$\widehat{\nu}(z) = \exp\left[\int \mathrm{d}\pi(y)(e^{izy} - 1)\right] , \tag{9.54}$$

$$\pi(y) = m \int \mathrm{d}\nu(h)\sqrt{\frac{2}{1-h}}(\delta(y - a(-\widetilde{h})) + \delta(y + a(-\widetilde{h}))) . \tag{9.55}$$

Let us make a few remarks on these equations; first, one can check that if m is finite and ν is a probability measure with no atoms in $h = \pm 1$ (i.e. in $\widetilde{h} = \pm\infty$), then the measure π defined in (9.55) gives a finite mass to all sets that are bounded away from 0, does not have an atom in 0, and is such that $\int \mathrm{d}\pi(y)|y|\mathbb{I}(|y| \leq 1) < \infty$. This ensures that the integral in the definition of $\widehat{\nu}(z)$ converges, hence π is a valid Lévy measure for the infinitely divisible distribution ν. Moreover the assumption of convergence of ν_n towards ν is self-consistent in the following sense: if $\nu_n \xrightarrow{\mathrm{w}} \nu$ (i.e. in the sense of weak convergence of probability measures) then $\pi_n(f) \to \pi(f)$ for all bounded continuous functions f that vanish in a neighborhood of 0, and $\lim_{\epsilon\to 0} \limsup_{n\to\infty} \int \mathrm{d}\pi_n(y) y^2 \mathbb{I}(|y| \leq$

$\epsilon) = 0$. These two properties allow to invoke theorem 8.7 of [109] to conclude that $\nu_n \stackrel{\text{w}}{\to} \nu$, the infinitely divisible distribution with Lévy measure π. Of course this circular argument does not prove the convergence $\nu_n \stackrel{\text{w}}{\to} \nu$, but at least shows its consistency.

Note that in (9.53-9.55) we have finally achieved the main goal of the article, namely getting rid of the parameters k and n that have been sent to infinity and obtaining a system of equations that only depend on the parameter γ. Moreover γ only appears in (9.53): one can thus view (9.54,9.55) as a self-consistent equation on ν, parametrized by m, from which one deduces γ according to (9.53). We have performed this numerical resolution at fixed m, relying again on the Gaussian approximation to deal with the divergence of the total mass of π, and checked that the results were in perfect agreement with the one obtained fixing γ in the $n \to \infty$ limit. In Fig. 9.6 we have plotted the correspondance function between γ and m. The numerical resolution of the fixed point equation (9.55) led us to the following observations: (i) a solution of the equation only exists for $m \leq 1$ (ii) the tail behavior of ν is very well described numerically by

$$\nu(\{\widetilde{h} \geq x\}) \sim S(x) \frac{e^{-r(m)x}}{x^\rho} \qquad \text{as} \quad x \to +\infty \ , \tag{9.56}$$

where S is a slowly varying function, ρ is an exponent very close to 1 for all the values of m we have considered, and $r(m) \geq 0$ controls the dominant exponential decrease of the tail (iii) $r(m) > 0$ for $m < 1$, and $r(m) \to 0$ as $m \to 1$.

Some of these numerical observations can be rationalized through the following analytical considerations. If one assumes the form (9.56), the exponential behavior of ν when $\widetilde{h} \to +\infty$ (and also in $-\infty$ thanks to the symmetry of ν) allows to continue $\widehat{\nu}(z)$ from the real to the imaginary axis, i.e. to define the Laplace transform of ν as

$$L(t) = \widehat{\nu}(-it) = \int \mathrm{d}\nu(h) e^{t\widetilde{h}} = \int \mathrm{d}\nu(h) \cosh(t\widetilde{h}) \ . \tag{9.57}$$

This function L is well-defined for $t \in (-r(m), r(m))$, formally infinite for $|t| > r(m)$, with $L(0) = 1$ and L non-decreasing on $[0, r(m))$. Its behavior as $t \to r(m)$ depends on the exponent ρ that controls the algebraic decay in (9.56). If $\rho \leq 0$ then L diverges in this limit, with a dominant behavior of $(r(m) - t)^\rho$ (and a prefactor that involves the slowly varying function S, see theorem XIII.5.1 in [112] for a rigorous statement of such a Tauberian argument), while if $\rho > 0$ the limit $L(r(m))$ is finite. We claim that for the probability measure ν solution of (9.55) one should have $\rho > 0$. Let us indeed rewrite this equation in terms of the Laplace transforms as $L(t) = \exp[R(t)]$, with $R(t) = \int \mathrm{d}\pi(y)(e^{ty} - 1)$. Consider now the tail behavior of the measure π related to ν by (9.55); a simple computation, based in particular on the fact that $a(\widetilde{h}) = \widetilde{h} + O(e^{-2\widetilde{h}})$ as $\widetilde{h} \to \infty$, shows that the asymptotic expansion (9.56) is also valid for the tail of π, up to the multiplicative coefficient m: $\pi(\{y \geq x\}) \sim m\, \nu(\{\widetilde{h} \geq x\})$ as $x \to \infty$. As a consequence if $\rho < 0$ both $L(t)$ and $R(t)$ would diverge as $(r(m) - t)^\rho$ when $t \to r(m)$, which is clearly incompatible with the equation $L(t) = \exp[R(t)]$.

We must thus have $\rho > 0$, in such a way that L is finite in $r(m)$; nevertheless some derivatives of L will be divergent in $r(m)$, each derivative with respect to t in (9.57) multiplying the integrand by $\widetilde{h}$, hence reducing the exponent ρ by 1. Suppose for simplicity that $0 < \rho \leq 1$, in such a way that $L'(t)$ is divergent (higher values of ρ would lead to the same conclusion by considering higher order derivatives). Because of the proportionality between the tails of ν and π the function $R'(t)$ is also divergent, with $R'(t)/L'(t) \to m$ as $t \to r(m)$. Since $L'(t) = R'(t)\exp[R(t)] = R'(t)L(t)$, we are led to the conclusion that

$$1 = m\, L(r(m)) \ . \tag{9.58}$$

This type of reasoning can actually be put on rigorous grounds, the statement (9.58) being essentially the content of corollary 2.1 in [113]. As $L(r(m)) \geq L(0) = 1$ we can conclude that a solution of (9.55) with a tail behavior of the form (9.56) can only exist for $m \leq 1$, and that $r(m)$ must vanish when $m \to 1$, which confirms part of the numerical observations reported above. As m is a decreasing function of γ (see in particular Fig. 9.6) we are thus led by (9.53) to the following definition for our conjectured value of γ_d:

$$\gamma_d = \int \mathrm{d}\nu(h)\beta(h) \ , \tag{9.59}$$

where ν is the solution of (9.55) with $m = 1$ (and hence has a power law decay when $\widetilde{h} \to \infty$). A numerical evaluation of this quantity led us to the estimate $\gamma_d \approx 0.871$. The property $m \to 1$ as $\gamma \to \gamma_d$ was illustrated with the dotted horizontal line in the bottom panel of Fig. 9.5, that corresponds to the expected height of the plateau of the curves in the critical region.

The square root behavior of m as a function of γ when γ is close to γ_d, that is clearly visible on Fig. 9.6 and confirmed by a numerical fit of the data, is in line with the bifurcation of the functional equation (9.55). It is more surprising to observe that $\widetilde{C}(\gamma)$ seems to behave linearly in the limit $\gamma \to \gamma_d$ (see the bottom panel of Fig. 9.3), as it could suggest that $\widetilde{C}$ admits a continuation to smaller values of γ, in violation of the hypothesis made at the end of Sec. 9.2.4. There are alternative explanations to this observation: the square root contribution could be very small and not observable with our numerical accuracy, or the expected square root behavior of $C(\alpha, k)$ when $\alpha \to \alpha_d$ at finite k could be washed out in the limit $k \to \infty$, i.e. the limits $\alpha \to \alpha_d$ and $k \to \infty$ do not necessarily commute to correspond to the limit $\gamma \to \gamma_d$.

9.2.6 An analytic lower bound on γ_d

As we have seen above our conjecture for γ_d is given in terms of the solution of a functional equation that does not seem to admit a simple expression. One can however derive explicitly a quantitatively close lower bound, by observing that the function $\beta(h)$ defined in (9.35) is such that $\beta(h) \geq 2(\sqrt{2}-1)$ for all $h \in [-1, 1]$. As a consequence the sequences of masses m_n obey the inductive bounds $m_{n+1} \geq \exp[-\gamma + 2(\sqrt{2}-1)m_n]$. The function $m \to \exp[-\gamma + 2(\sqrt{2}-1)m]$ being

increasing the sequence $\widehat{m}_n$ defined by $\widehat{m}_1 = e^{-\gamma}$, $\widehat{m}_{n+1} = \exp[-\gamma + 2(\sqrt{2} - 1)\widehat{m}_n]$ can be shown by induction to lower bound the sequence m_n, namely $m_n \geq \widehat{m}_n$ for all n. It is then simple to show that for $\gamma < 1 + \ln(2(\sqrt{2} - 1)) \approx 0.812$ the sequence $\widehat{m}_n$ diverges with n, and as a consequence so does m_n. Under the assumption spelled out at the end of Sec. 9.2.4 on the equivalence of the divergence of m_n and of $\widetilde{C}_n$ this yields the lower bound $\gamma_d \geq 1 + \ln(2(\sqrt{2} - 1))$.

Chapter 10

The asymptotics of the clustering transition for biased measures

In this chapter, we investigate the large k asymptotics of the clustering transition undergone by the bicoloring of k-uniform random hypergraphs, when its solutions are weighted non-uniformly, with a soft interaction between variables belonging to distinct hyperedges. We find that $\alpha_d(k) = \frac{2^{k-1}}{k}(\ln k + \ln\ln k + \gamma_d + o(1))$, where the constant γ_d is strictly larger than for the uniform measure over solutions.

As already mentioned in chapter 5, an important problem in the field of random CSPs concerns the behavior of algorithms in the satisfiable regime, where the goal is to find a solution, as typical instances admit such configurations. In particular, one would like to determine the algorithmic thresold $\alpha_{alg}(k)$ above which no algorithm is able to find a solution with high probability (assuming P$\neq$NP). We have seen (in section 5.1) that for small values of k it is possible to design algorithms (see [6, 17, 18, 19, 20]) for which a numerical study shows that they are efficient up to densities very close to the satisfiability threshold. In the large k limit these algorithms cannot be studied numerically. It is however possible to perform the analytical study of some algorithms based on simpled heuristics. Up to know the best result is the one of [21], which provides an algorithm that provably works in polynomial time up to densities of the order of $2^{k-1}\ln k/k$, which coincides with the asymptotic scaling of α_d. We recall that the satisfiability threshold occurs at the scaling $\alpha_{sat}(k) \sim 2^{k-1}\ln 2$, therefore this leaves a wide range of α where typical instances do admit solutions, but no known algorithm is able to find them efficiently. Moreover some families of algorithms have been proven to fail [22, 23, 24].

Although the connection between the algorithmic performances and the several structural phase transitions that occur in the satisfiable phase is very delicate, the strategy adopted in this Ph.D was to study the clustering threshold

α_d. One can reasonably state that the clustering transition is a lower bound on the algorithmic one, $\alpha_d(k) \leq \alpha_{alg}(k)$. Indeed for $\alpha < \alpha_d$, simulated annealing is able to equilibrate in polynomial time down to arbitrary small temperatures [14], and hence to sample the solution set (with a biased or uniform measure). For α slightly larger than α_d one expects simulated annealing to fall out-of-equilibrium in polynomial time scales, but in many cases it should be able to find solutions, as we have seen in section 8.4 for the biased measure μ_{Θ_0}. Hence the bound $\alpha_d(k) \leq \alpha_{alg}(k)$ is not tight in general.

In chapter 8 we have studied two implementations of the bias in the measure on the set of solution of a hypergraph bicoloring instance for small values of k. For the simplest biased measure μ_{Θ_0}, the interactions can be factorized over the bicoloring constraints. The second biased measure μ_{Θ_1} induces interactions between variables at distance 1. We studied the modifications of the clustering threshold α_d induced by the non-uniformicity between the solutions. We showed, for k between 4 and 6, that with well-chosen parameters the bias allows to increase α_d, and that it can be further increased with the biased measure μ_{Θ_1} compared to the simpler implementation μ_{Θ_0}. We checked for μ_{Θ_0} that the bias could improve the performances of simulated annealing.

The large k behavior of α_d is more involved. We presented some rigorous bounds on its asymptotic expansion in Section 5.2.1 (see [38, 39, 40]). In chapter 9 (that report the results of [32]) we found that the clustering threshold occurs on the scale $\alpha \sim 2^{k-1}(\ln k + \ln\ln k + \gamma)/k$, with γ constant, and more precisely for the uniform measure $\gamma_{d,u} \approx 0.871$, which falls into the range allowed by the previous bounds [38, 39, 40]. In this chapter we adapt the large k expansion of chapter 9([32]) to the biased measures μ_{Θ_0}, μ_{Θ_1} and manage to assess the asymptotic effect of the bias on α_d. We find that the factorized bias of μ_{Θ_0} (see equation (6.7)) studied in [31] cannot improve on the constant γ_d in the asymptotic expansion with respect to $\gamma_{d,u}$, while the biased measure μ_{Θ_1} with larger interaction range allows to increase its value up to $\gamma_d \approx 0.977$. This is arguably a modest improvement, bearing on the third order of the asymptotic expansion of α_d, nevertheless it opens the possibility to study further generalizations of the bias and to bring some light on the nature of the algorithmic gap between α_{alg} and α_{sat}.

This chapter is organized as follows. In Section 10.1 we derive a first upper bound on the possible increase of the dynamic threshold for the factorized measure μ_{Θ_0}, that was obtained in [31]. Although we will see later that this bound is not tight, it already indicates that there is no possibility to improve on the first term of the asymptotic expansion of α_d. Moreover, it motivates the choice (8.17) for the bias parameters ω_p defining the bias. The analytical expansion of the dynamic transition threshold in the large k limit for the biased measure is presented in Sections 10.2-10.5. These sections present the results obtained in [33]. We will focus on the $(k, l+1)$-regular ensemble, where the probability is uniform on the set of hypergraphs for which each vertex belongs to $l+1$ hyperedges. We have seen indeed in chapter 7 that the study of the biased measure μ_{Θ_1} that introduces interaction between variables at distance 1 on the regular ensemble is simplified compared to the Erdős Rényi ensemble

$G_N(k, M)$. In Section 10.2 we give the main ideas of the computations made in the next sections, and simplify some of the equations that will be needed in the following. In Section 10.3 we perform the large k limit of the recursive distributional equations (7.67-7.69), and reduce the determination of γ_d to a set recursive distributional equations that does not depends on k. The study of the large n limit of these equations is performed in Section 10.4. In section 10.5 we give the numerical results obtained fom the resolution of these distributional equations that allows to compute the asymptotic expansion of α_d.

10.1 A first upper-bound for the intra-clause bias

The numerical resolution of the cavity equations presented in Sec. 8.3 shows that for small values of k one has $\alpha_{\rm opt}(k) > \alpha_{d,u}(k)$, in other words that distorting the measure over solutions can make it RS for larger densities of constraints than the uniform one. We want now to investigate the large k limit, for which the gap between the satisfiability threshold and the algorithmic one is most clearly demonstrated. In this section we will derive an asymptotic bound on the large k behavior of $\alpha_{\rm opt}(k)$, for the biased measure μ_{Θ_0} with the specific choice (8.17).

As explained before for any ϵ one has $\alpha_d(\epsilon) \leq \min(\alpha_{KS}(\epsilon), \alpha_r(\epsilon), \alpha^{s=0}(\epsilon))$, these three upper bounds having simple expressions given in Eqs. (8.18-8.20). At large enough k it is easy to convince oneself that the Kesten-Stigum transition occurs after $\alpha^{s=0}$, hence is completely irrelevant (the dominant term in the asymptotic expansion of α_{KS} is of the order $2^{2(k-1)}$ instead of 2^{k-1} for $\alpha^{s=0}$ and α_r). We show in Fig. 10.1 the lines $\alpha_r(\epsilon), \alpha^{s=0}(\epsilon)$ for a large value of k. We define $(\alpha_*(k), \epsilon_*(k))$ as the coordinates of the intersection of the rigidity and the zero-entropy line, in such a way that $\alpha_{\rm opt}(k) \leq \alpha_*(k)$: for larger densities either the RS entropy is negative, or there exists a 1RSB solution with hard-fields (or both), in any case no RS phase can exist for $\alpha \geq \alpha_*$. We will now derive an asymptotic expansion at large k of this upper bound $\alpha_*(k)$.

In the expression (8.20) of α_r the coefficient $\Gamma_r(k)$ is a series depending solely on k, that was defined in Eq. (7.81); in order to obtain more easily its asymptotic equation we define a series w_k by $H_r(k) = 1 - e^{-w_k}$, in such a way that w_k is solution of the implicit equation $e^w = 1 + (k-1)w$. Taking the logarithm of this equation and iterating once yields

$$w_k = \ln k + \ln\ln k + O\left(\frac{\ln\ln k}{\ln k}\right) \ . \tag{10.1}$$

One can then compute

$$\Gamma_r(k) = w_k \left(1 + \frac{1}{(k-1)w_k}\right)^{k-1} = \ln k + \ln\ln k + O(1) \ . \tag{10.2}$$

This gives immediately the expansion of the rigidity threshold for the uniform

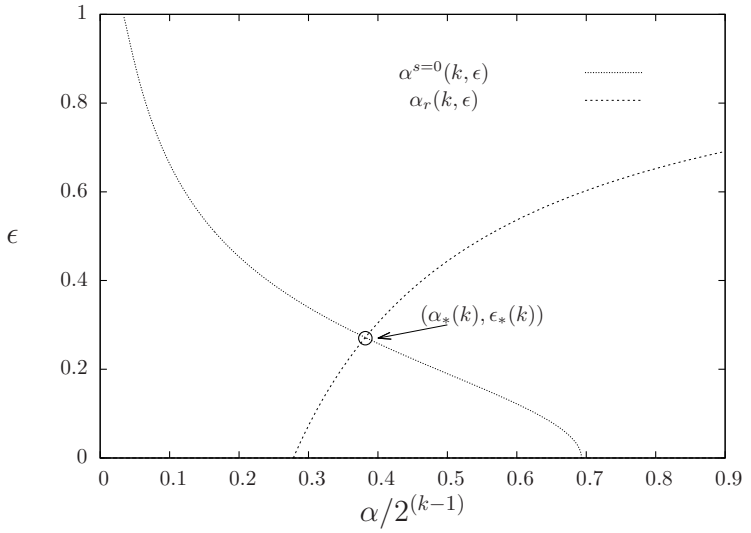

Figure 10.1: For $k = 20$, plot in the plane (α, ϵ) of the rigidity line and zero RS entropy line, that intersect at the point $(\alpha_*(20), \epsilon_*(20))$.

measure ($\epsilon = 0$),

$$\alpha_{r,u}(k) = 2^{k-1}\frac{1}{k}\left(\ln k + \ln\ln k + O(1)\right) \ . \tag{10.3}$$

that is confirmed by the asymptotic expansion (5.13) (see [62] and [38] in which the third term ($O(1)$) is also derived: $\gamma_r = 1$). We come back to the determination of $(\alpha_*(k), \epsilon_*(k))$; this intersection of the rigidity and zero-entropy line is solution of the two following equations, immediately obtained from (8.18,8.20):

$$\alpha = \frac{\ln 2}{\frac{k(1-\epsilon)\ln(1-\epsilon)}{2^{k-1}-1-k\epsilon} - \ln\left(1-\frac{1+k\epsilon}{2^{k-1}}\right)} = \frac{1}{k}\Gamma_r(k)\frac{2^{k-1}-1-k\epsilon}{1-\epsilon} \ . \tag{10.4}$$

Neglecting only exponentially small corrections we can simplify (10.4) into

$$\frac{\alpha}{2^{k-1}} = \frac{\ln 2}{k(1-\epsilon)\ln(1-\epsilon) + 1 + k\epsilon} = \frac{1}{k}\Gamma_r(k)\frac{1}{1-\epsilon} \ . \tag{10.5}$$

Without making additional approximations we see that $\epsilon_*(k)$ is solution of

$$\epsilon + (1-\epsilon)\ln(1-\epsilon) = \frac{\ln 2}{\Gamma_r(k)} - \frac{1}{k} - \epsilon\frac{\ln 2}{\Gamma_r(k)} \ . \tag{10.6}$$

Given the asymptotic behavior of $\Gamma_r(k)$ stated in (10.2) it is easy to see that $\epsilon_*(k)$ must vanish in the limit; one can thus expand the l.h.s. of (10.6) and

obtain

$$\epsilon_*(k) = \sqrt{\frac{2\ln 2}{\Gamma_r(k)}} + O\left(\frac{1}{\ln k}\right) \ . \tag{10.7}$$

Reinserting in (10.5) we obtain

$$\alpha_*(k) = 2^{k-1}\frac{1}{k}\Gamma_r(k)\left(1 + \sqrt{\frac{2\ln 2}{\Gamma_r(k)}} + O\left(\frac{1}{\ln k}\right)\right) \ . \tag{10.8}$$

Using the expansion (10.2) of $\Gamma_r(k)$ we have finally

$$\alpha_*(k) = 2^{k-1}\frac{1}{k}\left(\ln k + \sqrt{2\ln 2}\sqrt{\ln k} + O(\ln\ln k)\right) \ . \tag{10.9}$$

The comparison with the expansion of $\alpha_{r,u}(k)$ given in (10.3) shows that the leading order is not modified, the term $\ln\ln k$ in the correction being replaced by a (larger) term of order $\sqrt{\ln k}$.

In chapter 9 (corresponding to the results presented in [32]) we have found that $\alpha_{d,u}$ has the same asymptotic expansion as $\alpha_{r,u}$ (see equation (5.15)), with a strictly smaller constant $\gamma_d \approx 0.871$. We thus conclude that

$$\alpha_{d,u}(k) = 2^{k-1}\frac{1}{k}\left[\ln k + \ln\ln k + \gamma_d\right] \leq \alpha_{\text{opt}}(k) \leq \alpha_*(k) \tag{10.10}$$

$$\alpha_*(k) = 2^{k-1}\frac{1}{k}\left(\ln k + \sqrt{2\ln 2}\sqrt{\ln k} + O(\ln\ln k)\right) \ , \tag{10.11}$$

hence that the best improvement of $\alpha_{\text{opt}}(k)$ with respect to $\alpha_{d,u}(k)$ that can be hoped for with the bias considered in this paper is a replacement of $\ln\ln k$ by $\sqrt{\ln k}$ in the second order term of their asymptotic expansions. We shall actually see in the next sections that this upper bound is too optimistic, and that with the bias μ_{Θ_0} and the choice (8.17) we could not increase the asymptotic expansion of the clustering threshold.

We shall finally come back briefly on the choice of parameters we made in (8.17), where we used a single parameter ϵ for the bias instead of trying to exploit all the $k/2$ free values of ω_p. Let us define $\alpha'_{\text{opt}}(k) = \sup \alpha_d(k, \{\omega_p\})$, where the maximization is now over all possible values of the ω_p, under the conditions $\omega_0 = \omega_k = 0$ and $\omega_p = \omega_{k-p}$. We have certainly $\alpha'_{\text{opt}}(k) \geq \alpha_{\text{opt}}(k)$, and the inequality is probably strict at least for small enough values of k; however we shall now show that $\alpha'_{\text{opt}}(k) \leq \alpha_*(k)$, hence that the larger freedom of choice of generic parameters does not allow to beat the upper bound derived and discussed in the special case (8.17). To see this more easily let us exploit the invariance of the measure (6.7) under a multiplication of all ω_p by a common constant, and fix their normalizations in such a way that

$$\sum_{p=1}^{k-1}\binom{k}{p}\omega_p = 1 \ . \tag{10.12}$$

With this choice the expressions of the RS entropy (7.34) and rigidity threshold (7.82) become

$$s^{RS}(\alpha, \{\omega_p\}) = \ln 2 + \alpha \left(\sum_{p=1}^{k-1} \binom{k}{p} \omega_p \ln \omega_p - k \ln 2 \right) ,$$

$$\alpha_r(k, \{\omega_p\}) = \frac{1}{k} \Gamma_r(k) \frac{1}{2\omega_1} . \tag{10.13}$$

Consider now a given choice of the parameter ω_1, and a value of $\alpha \leq \alpha_r$; for these to allow a RS phase the corresponding entropy should be positive. Maximizing the entropy in (10.13) with respect to $\{\omega_2, \dots, \omega_{k-2}\}$, under the normalization condition (10.12) and for a fixed value of ω_1 is easily seen to yield $\omega_2 = \cdots = \omega_{k-2}$, i.e. precisely the choice of parameters (8.17). In other words this bias is the one that allows to tune the fraction of frozen variables while keeping the measure as uniform as possible, in order to minimize the entropy cost it induces.

10.2 The asymptotics of the clustering threshold

10.2.1 Setting

The rest of chapter will be devoted to an asymptotic expansion of the clustering threshold when $k \to \infty$. We will focus on the k-uniform $l+1$-regular hypergraph ensemble, and on the measure μ_{Θ_1} defined in (6.12). We have seen in the section 8.5 that for finite k this biased measure has a larger l_{d} with respect to the uniform one, and that the inclusion of interactions between variables at larger distance brings a further improvement compared to a biasing function factorized over the hyperedges (i.e. the measure μ_{Θ_0} defined in (6.7)). It is thus natural to investigate this phenomenon in the large k limit, that allows for some analytical simplifications, and where the algorithmic gap discussed in the introduction is most clearly seen. One would like in particular to understand at which order of the asymptotic expansion of l_{d} the effect of the bias does appear.

We give here, for the convenience of the reader, the main ideas and explain the organization of the forecoming computation, which is the generalization of the one we presented in [32] (see chapter 9) for the uniform measure. We will focus on the particular form of the function $\psi(p)$ defined in (6.13), with the two parameters b and ϵ that we recall here:

$$\psi(0) = 1, \quad \psi(p) = b(1-\epsilon)^p \text{ for } 1 \leq p \leq l+1 . \tag{10.14}$$

Note that the specific choice of parameters (8.17) in the measure μ_{Θ_0} corresponds in the above equation to set $b = 1$ (i.e. to take $\psi(p) = (1-\epsilon)^p$), as explained in the section 6.1.3.

We start in the section 10.2.2 by summarizing the main equations derived in chapter 7, for arbitrary k, in this special case. In order to take the $k \to \infty$

limit we must specify how the degree l and the parameters (b, ϵ) behave with k; we will set

$$l = 2^{k-1}(\ln k + \ln\ln k + \gamma) \ , \qquad \epsilon = \widetilde{\epsilon}\sqrt{\frac{2}{k \ln k}} \ , \tag{10.15}$$

where γ and $\widetilde{\epsilon}$ are constants independent of k that parametrize the degree and the bias in this limit (the factor $\sqrt{2}$ being for later notational convenience), while b will be independent of k. We will find that both $l_{\rm d}$ and the rigidity threshold $l_{\rm r}$ have asymptotic expansions of the form (10.15), our goal being to determine the corresponding rescaled thresholds $\gamma_{\rm d}$ and $\gamma_{\rm r}$, as a function of the parameters $(b, \widetilde{\epsilon})$.

To do so we shall first expand the correlation function C_n and its hard-fields contribution H_n, for a finite distance n, and find that both go to their maximal value 1, with the correction term scaling as

$$C_n = 1 - \frac{\widetilde{C}_n}{k \ln k} + o\left(\frac{1}{k \ln k}\right) \ , \qquad H_n = 1 - \frac{\widetilde{H}_n}{k \ln k} + o\left(\frac{1}{k \ln k}\right) \ , \tag{10.16}$$

where $\widetilde{C}_n$ and $\widetilde{H}_n$ are independent of k. These sequences depend on the rescaled parameters γ, b and $\widetilde{\epsilon}$, and we present in Sec. 10.3 recursion equations that allow to compute them (in Sec. 10.3.1 for $\widetilde{H}_n$ and in Sec. 10.3.2 for $\widetilde{C}_n$).

The tresholds $l_{\rm d}$ and $l_{\rm r}$ have been defined for finite k according to the positivity of the large n limit of the sequences C_n and H_n, respectively. Their asymptotic expansion should thus be performed by taking the large k limit after the large n one; however, under the natural hypothesis (that can be checked explicitly for H_n) that the large n limit of C_n and H_n is either strictly vanishing or scales with k as in (10.16), one can determine $\gamma_{\rm d}$ and $\gamma_{\rm r}$ by reversing the order of the limits and studying whether $\widetilde{C}_n$ and $\widetilde{H}_n$ remain bounded or not in the large n limit. The large n limit of $\widetilde{C}_n$ is thus discussed in Sec. 10.4. Additional difficulties need to be overcome in the intermediate regime $\gamma_{\rm d} < \gamma < \gamma_{\rm r}$ where reconstruction is possible but naive reconstruction is not: even if strictly hard fields are not present here the scaling (10.16) reveals that the soft fields are actually quasi-hard, the correlation function tending to one. We thus reformulate in Sec. 10.4.2 the recursion of 10.3.2 and put it in a form for which the large n limit can be performed in a numerically tractable way. Finally our explicit results for $\gamma_{\rm d}$ are presented in 10.5.

10.2.2 A specialization of some formulas

Let us first specialize some of the formulas we wrote previously for a generic $\psi(p)$ to the case defined in equation (10.14) with the two parameters b and ϵ.

The BP equation (7.18) for the function $\eta = f_{\Theta_1}(\widehat{\eta}_1, \dots, \widehat{\eta}_l)$ becomes

$$\eta(\sigma,1) = \frac{1}{z} b(1-\epsilon) \prod_{i=1}^{l} \left(\widehat{\eta}_i(\sigma,0) + (1-\epsilon)\widehat{\eta}_i(\sigma,1)\right) \ , \tag{10.17}$$

$$\eta(\sigma,0) = \frac{1}{z}\left[(1-b)\prod_{i=1}^{l}\widehat{\eta}_i(\sigma,0) + b\prod_{i=1}^{l}\left(\widehat{\eta}_i(\sigma,0) + (1-\epsilon)\widehat{\eta}_i(\sigma,1)\right)\right] \ ,$$

for $\sigma = \pm 1$. The equation (7.40) for the factorized RS solution reads

$$y = \frac{1}{1-\epsilon}\left(1 + \frac{1-b}{b}(1 + (1-\epsilon)\widehat{y}^{-1})^{-l}\right) \ ,$$

$$\widehat{y} = 2^{k-1} - k - 1 + \frac{k-1}{y} \ . \tag{10.18}$$

The evolution equations (7.88,7.89) for the hard fields become

$$h_{0,n+1} = 1 - \frac{1-b+b(1+(1-\widehat{h}_{n+1})(1-\epsilon)\widehat{y}^{-1})^l}{1-b+b(1+(1-\epsilon)\widehat{y}^{-1})^l} \ , \tag{10.19}$$

$$h_{1,n+1} = 1 - \frac{(1+(1-\widehat{h}_{n+1})(1-\epsilon)\widehat{y}^{-1})^l}{(1+(1-\epsilon)\widehat{y}^{-1})^l} \ , \tag{10.20}$$

$$H_n = 1 - \frac{1-b+b(1+(1-\widehat{h}_n)(1-\epsilon)\widehat{y}^{-1})^{l+1}}{1-b+b(1+(1-\epsilon)\widehat{y}^{-1})^{l+1}} \ , \tag{10.21}$$

where we recall the initial condition $h_{0,n=0} = 1$ and the fact that $\widehat{h}_{n+1} = (h_{0,n})^{k-1}$. We can thus write a closed equation on $h_{0,n}$:

$$h_{0,n+1} = 1 - \frac{1-b+b(1+(1-(h_{0,n})^{k-1})(1-\epsilon)\widehat{y}^{-1})^l}{1-b+b(1+(1-\epsilon)\widehat{y}^{-1})^l} \ . \tag{10.22}$$

One can check numerically that this equation undergoes a discontinuous bifurcation when l increases above the rigidity threshold $l_{\rm r}$. Here all the formulas depend analytically on l, we can thus consider it as a real parameter, even if the original model is only defined for integer l. The fixed point $h_0 = \lim_{n\to\infty} h_{0,n}$ jumps abruptly from 0 to a strictly positive value when l is increased above $l_{\rm r}$. We can determine the location of this threshold by noting that at such a bifurcation the function that maps $h_{0,n}$ to $h_{0,n+1}$ is tangent with the diagonal, hence $l_{\rm r}$ and the bifurcating fixed point $h_{0,{\rm r}}$ are solutions of

$$h_{0,{\rm r}} = 1 - \frac{1-b+b(1+(1-(h_{0,{\rm r}})^{k-1})(1-\epsilon)\widehat{y}^{-1})^{l_{\rm r}}}{1-b+b(1+(1-\epsilon)\widehat{y}^{-1})^{l_{\rm r}}} \ , \tag{10.23}$$

$$1 = \frac{l_{\rm r}(k-1)(h_{0,{\rm r}})^{k-2}(1-\epsilon)\widehat{y}^{-1}b(1+(1-(h_{0,{\rm r}})^{k-1})(1-\epsilon)\widehat{y}^{-1})^{l_{\rm r}-1}}{1-b+b(1+(1-\epsilon)\widehat{y}^{-1})^{l_{\rm r}}} \ . \tag{10.24}$$

For a generic bias $\psi(p)$ the distribution $R_{w,n}$ of the hard fields introduced in (7.86) is a priori non-trivial, but for the particular choice of ψ defined in (10.14) it simplifies into

$$R_{w,n}(\eta) = \delta(\eta - \eta^+) \ , \quad \text{where} \quad \eta^+(\sigma,0) = \frac{1}{2-\epsilon}\delta_{\sigma,+} \ , \quad \eta^+(\sigma,1) = \frac{1-\epsilon}{2-\epsilon}\delta_{\sigma,+} \ , \tag{10.25}$$

for all w and n. We will also denote $\eta^- = (\eta^+)^f$ the message forcing to $-$. This allows to simplify the equation (7.96) on the soft fields distribution, which reads now:

$$\widehat{Q}_{1,n+1}(\widehat{\eta}) = \sum_{u=1}^{k-1} \frac{\binom{k-1}{u}(h_{0,n})^{k-1-u}(1-h_{0,n})^u}{1-(h_{0,n})^{k-1}} \int \prod_{i=1}^{u} \mathrm{d}Q_{0,n}(\eta_i) \\ \times \delta(\widehat{\eta} - \widehat{f}_{\Theta_1}(\eta_1^f, \dots, \eta_u^f, \eta^-, \dots, \eta^-)) \ . \tag{10.26}$$

It will be useful in the following to encode in a compact way the value of $\widehat{f}_{\Theta_1}(\eta_1, \dots, \eta_{k-1})$ (defined in equation (7.19)) when all, or almost all, the arguments of $\widehat{f}_{\Theta_1}$ are forcing messages. We shall hence define, for a real number α, the message $\widehat{\eta} = g_0(\alpha)$ as

$$\widehat{\eta}(\sigma, w) = \delta_{w,0} \frac{1 + \sigma \tanh(\alpha)}{2} \ ; \tag{10.27}$$

the value w is thus fixed to 0, while σ can be seen as an Ising spin submitted to an effective magnetic field α. One then founds that the values of $\widehat{f}_{\Theta_1}$ when all its arguments are forcing are:

- $\widehat{f}_{\Theta_1}(\eta^-, \dots, \eta^-) = \widehat{\eta}^+$ and $\widehat{f}_{\Theta_1}(\eta^+, \dots, \eta^+) = \widehat{\eta}^-$, the usual combination rule to obtain a forcing message $\widehat{\eta}$;
- $\widehat{f}_{\Theta_1}(\eta^+, \eta^-, \dots, \eta^-) = g_0(\epsilon')$ and $\widehat{f}_{\Theta_1}(\eta^-, \eta^+, \dots, \eta^+) = g_0(-\epsilon')$, with $\epsilon' = -\frac{1}{2}\ln(1-\epsilon)$, when all the messages except one are forcing in the same direction, the last one in the opposite direction;
- $\widehat{f}_{\Theta_1}(\eta^+, \dots, \eta^+, \eta^-, \dots, \eta^-) = g_0(0)$ when there are at least two forcing fields in each direction.

We will also introduce two functions g_+ and g_- that gives the value of $\widehat{f}_{\Theta_1}$ when all its arguments are forcing in the same direction, except one which is arbitrary, namely $g_+(\eta) = \widehat{f}_{\Theta_1}(\eta, \eta^-, \dots, \eta^-)$ and $g_-(\eta) = \widehat{f}_{\Theta_1}(\eta, \eta^+, \dots, \eta^+)$. Explicitly, $\widehat{\eta} = g_\sigma(\eta)$ means

$$\widehat{\eta}(\sigma,1) = \frac{1}{\widehat{z}}\eta(-\sigma,0) \ , \quad \widehat{\eta}(-\sigma,1) = 0 \ , \quad \widehat{\eta}(\sigma,0) = \frac{1}{\widehat{z}}\eta(\sigma,0) \ , \\ \widehat{\eta}(-\sigma,0) = \frac{1}{\widehat{z}}\eta(\sigma,1) \ , \tag{10.28}$$

with $\widehat{z}$ normalizing this distribution. Note that the two functions g_+ and g_- are linked by the spin-flip operation according to $g_+(\eta^f) = (g_-(\eta))^f$.

10.3 The large k limit for a finite distance n

10.3.1 Evolution of the hard fields

We start our large k asymptotic expansion, using the scaling of the parameters defined in (10.15), by considering the solution (10.18) of the translationally invariant RS equation; its leading order behavior is easily found to be

$$y = 1 + \widetilde{\epsilon}\sqrt{\frac{2}{k\ln k}} + o\left(\frac{1}{\sqrt{k\ln k}}\right) , \qquad \widehat{y} = 2^{k-1}\left(1 + O\left(\frac{k}{2^{k-1}}\right)\right) . \quad (10.29)$$

Turning to the sequences $h_{w,n}$ for the weights of the hard fields, solutions of the recursion equations (10.19,10.20), one realizes easily that, for n finite in the large k limit with the scaling of the parameters stated above,

$$h_{0,n} = 1 - \frac{x_{0,n}}{k\ln k} + o\left(\frac{1}{k\ln k}\right) , \quad h_{1,n} = 1 - \frac{x_{1,n}}{k\ln k} + o\left(\frac{1}{k\ln k}\right) ,$$
$$\widehat{h}_{n+1} = 1 - \frac{x_{0,n}}{\ln k} + o\left(\frac{1}{\ln k}\right) , \quad (10.30)$$

where $x_{0,n}$ and $x_{1,n}$ are independent of k and solutions of the recursion relations:

$$x_{0,n+1} = B\, e^{-\gamma} + e^{-\gamma + x_{0,n}} , \quad (10.31)$$
$$x_{1,n+1} = e^{-\gamma + x_{0,n}} = x_{0,n+1} - B\, e^{-\gamma} . \quad (10.32)$$

Here and sometimes in the following it is more convenient to use the notation

$$B = \frac{1-b}{b} \quad (10.33)$$

as a parameter equivalent to b. The recursion above is closed on $x_{0,n}$, and satisfies the initial condition $x_{0,n=0} = 0$, that follows immediately from $h_{0,n=0} = 1$. Note that for $b = 1$ (i.e. $B = 0$) one recovers the result of equation (9.4) for $x_n = x_{0,n} = x_{1,n}$, as it should in the uniform case. One also finds by expanding (10.21) that H_n, the hard fields contribution to the correlation function, is indeed given by the asymptotic expansion stated in (10.16), with $\widetilde{H}_n = x_{0,n}$.

The behavior of the sequence $x_{0,n}$ solution of (10.31) is easily determined by plotting the shape of the function $x \mapsto B\, e^{-\gamma} + e^{-\gamma + x}$, see the top panel of Fig. 10.2 for an example. For a given value of b (hence of B) there exists a critical value $\gamma_{\rm r}(b)$ such that this function remains strictly above the diagonal when $\gamma < \gamma_{\rm r}(b)$, while it intersects it for $\gamma > \gamma_{\rm r}(b)$. As a consequence in the former case the sequence $x_{0,n}$ diverges (very rapidly, as iterated exponentials) with n, whereas in the latter it converges to the smallest fixed point; these behaviors are illustrated in the bottom panel of Fig. 10.2. The divergence of $x_{0,n} = \widetilde{H}_n$ corresponds, in the large k limit, to the vanishing of H_n at finite k (recall the definition (10.16)), i.e. to the impossibility of naive reconstruction. The value of $\gamma_{\rm r}(b)$ can be obtained by noticing that at this bifurcation the function

$x \mapsto B\,e^{-\gamma} + e^{-\gamma+x}$ is tangent with the diagonal at their unique intersection point $x_{\rm r}(b)$, hence that $(x_{\rm r}(b), \gamma_{\rm r}(b))$ are solution of

$$\begin{cases} x = B\,e^{-\gamma} + e^{-\gamma+x} \\ 1 = e^{-\gamma+x} \end{cases} \quad \Rightarrow \quad \begin{cases} x = \gamma \\ \gamma = 1 + B\,e^{-\gamma} \end{cases} \quad . \tag{10.34}$$

As $b > 0$, $B > -1$, this equation admits a unique solution with $\gamma > 0$ (the sequence $x_{0,n}$ being positive this is also the case for the fixed point x, and hence also of γ at the bifurcation), which can be expressed as

$$\gamma_{\rm r}(b) = 1 + W\left(\frac{B}{e}\right) \ , \tag{10.35}$$

where $W(z)$ is the Lambert function, i.e. the principal solution of the equation $z = We^W$. Note that this result coincides with the asymptotic expansion of $l_{\rm r}$ one obtains from (10.23), which shows the commutativity of the limits $n \to \infty$ and $k \to \infty$ for the determination of the rigidity transition.

The function $\gamma_{\rm r}(b)$ is plotted in figure 10.6 (bottom panel, upper curve): it is a decreasing function of b, with $\gamma_{\rm r}(1) = 1$ for the uniform measure. An example for the values of the fixed point reached by $x_{0,n}$ for $\gamma > \gamma_{\rm r}(b)$ can be found in the bottom panel of Fig. 10.3.

10.3.2 Evolution of the soft fields distribution

We shall now study the large k limit of the soft fields distributions $Q_{w,n}, \widehat{Q}_{w,n}$. The crucial point we shall exploit to simplify them is the fact that the hard fields weights $h_{w,n}$ are very close to 1 according to the scaling (10.30), hence the dominant contributions to $\widehat{Q}_{w,n}$ will arise when the incoming messages are almost all forcing. To put this remark on a quantitative ground we start with the equation (10.26) on the distribution $\widehat{Q}_{1,n+1}$. The integer u that appears in this equation is a random number drawn from the binomial distribution $\mathrm{Bin}(k-1, 1-h_{0,n})$, conditioned to be strictly positive. In the large k limit, using the scaling behavior (10.30) of $h_{0,n}$, one sees that the average $(k-1)(1-h_{0,n})$ of the binomial distribution vanishes as $O(1/\ln k)$, hence the main contribution in (10.26) arises from the smallest value $u = 1$ appearing in the sum. We thus obtain at the leading order:

$$\widehat{Q}_{1,n+1}(\widehat{\eta}) = \int \mathrm{d}Q_{0,n}(\eta)\, \delta(\widehat{\eta} - g_+(\eta^f)) \ , \tag{10.36}$$

where the function g_+ was defined in (10.28). Consider now the equation (7.69) for $\widehat{Q}_{0,n+1} = \widehat{P}_{0,n+1}$. When all the η_i's are extracted from the hard part of $P_{0,n}$ and $P_{1,n}$ the arguments of $\widehat{f}_{\Theta_1}$ are all forcing, with the two possible directions represented; in most of these terms there are at least two messages in each direction, except for the term in the first line of (7.69), and the term with $t = k-2$ in the second line. According to the discussion in Sec. 10.2.2 this will yield contributions of the form $g_0(0)$ and $g_0(\pm\epsilon')$, where the function g_0

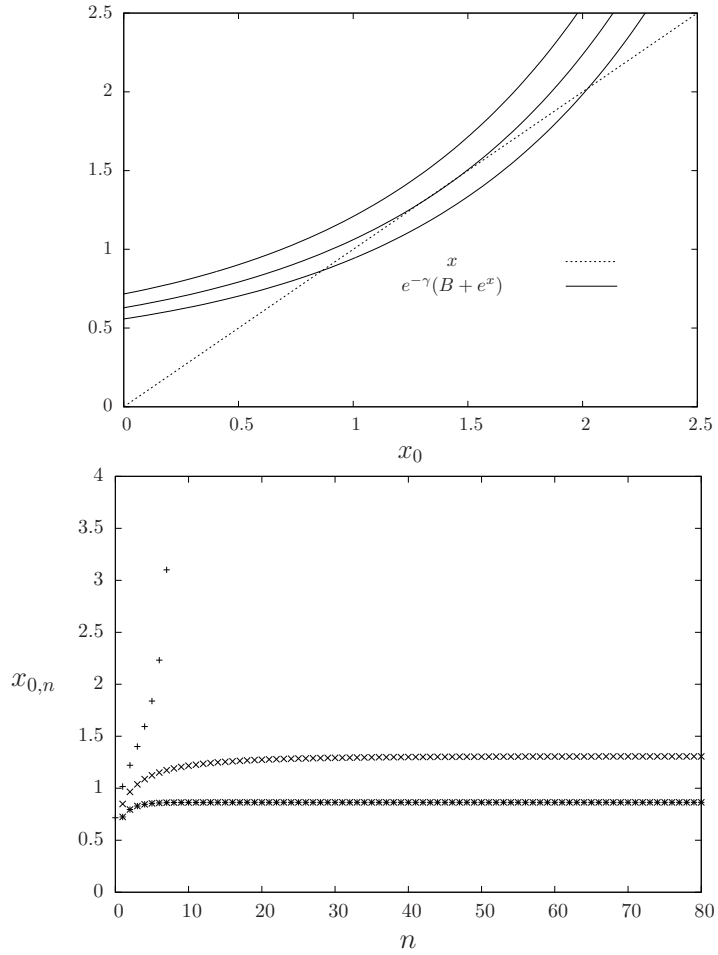

Figure 10.2: Top panel: the functions x and $B\,e^{-\gamma} + e^{-\gamma+x}$ as a function of x, with $b = 0.4$ and from top to bottom $\gamma = 1.25, 1.38, 1.50$. The bifurcation occurs when these two functions cross, which happens here at $\gamma_{\rm r}(0.4) \approx 1.378$. Bottom: $x_{0,n}$ as a function of n, for $b = 0.4$ and from top to bottom $\gamma = 1.25, 1.38, 1.50$.

was defined in (10.27). When exactly one of the η_i is a soft field, we obtain a contribution $g_-(\eta_1^f)$ from the first line if it is η_1 which is soft, a contribution $g_+(\eta_{k-1})$ from the term $t = k-2$ of the second line if η_{k-1} is the unique soft field, all other cases leading to subdominant contributions of the form $g_0(\alpha)$ with α of order ϵ'. Collecting these various contributions we thus obtain at the leading order:

$$\begin{aligned}\widehat{Q}_{0,n+1}(\widehat{\eta}) = & \frac{x_{1,n}}{2^{k-1}\ln k}\int \mathrm{d}Q_{1,n}(\eta)\,\delta(\widehat{\eta}-g_-(\eta^f)) \\ & + \frac{x_{0,n}}{2^{k-1}\ln k}\int \mathrm{d}Q_{0,n}(\eta)\,\delta(\widehat{\eta}-g_+(\eta)) \\ & + \beta_+\delta(\widehat{\eta}-g_0(\epsilon')) + \beta_-\delta(\widehat{\eta}-g_0(-\epsilon')) \\ & + \left(1-\frac{x_{0,n}+x_{1,n}}{2^{k-1}\ln k}-\beta_+-\beta_-\right)\delta(\widehat{\eta}-g_0(0)) \;, \end{aligned} \tag{10.37}$$

with

$$\beta_+ = \frac{k-1}{\widehat{y}}(h_{0,n})^{k-1} \;, \qquad \beta_- = \frac{k-1}{y\widehat{y}}h_{1,n}(h_{0,n})^{k-2} \;. \tag{10.38}$$

We turn now our attention to the equation (7.95) for $Q_{w,n+1}$; in the limit we are considering the non-vanishing contributions are found to arise only for values of p that remain finite, the law of p becoming $\mathcal{P}_{w,n}(p)$, with

$$\mathcal{P}_{1,n}(p) = \mathrm{Po}(p;x_{0,n}) \;, \quad \mathcal{P}_{0,n}(p) = \frac{1}{1-b+b\,e^{x_{0,n}}} \times \begin{cases} 1 & \text{if} \quad p=0 \\ b\frac{(x_{0,n})^p}{p!} & \text{if} \quad p>0 \end{cases} \;, \tag{10.39}$$

where we have introduced the notation $\mathrm{Po}(p;\lambda) = e^{-\lambda}\frac{\lambda^p}{p!}$ for the Poisson law of parameter λ; the $\mathcal{P}_{w,n}$ are indeed well-normalized probability distributions. In the right hand side of (7.95) a (finite) number p of messages $\widehat{\eta}_i$ are thus drawn from $\widehat{Q}_{1,n+1}$, for which we can use the limit form (10.36), while the others $l-p \sim 2^{k-1}\ln k$ are drawn from $\widehat{Q}_{0,n+1}$. Observing the form of Eq. (10.37) one realizes that the number of times the first two terms of $\widehat{Q}_{0,n+1}$ will be picked become Poissonian random variables of parameter $x_{1,n}$ and $x_{0,n}$, respectively. All the other terms are of the form $g_0(\alpha)$, which will be dealt with thanks to the simple exact identity:

$$f_{\Theta_1}(\widehat{\eta}_1,\ldots,\widehat{\eta}_s,g_0(\alpha_{s+1}),\ldots,g_0(\alpha_l)) = f_{\Theta_1}(\widehat{\eta}_1,\ldots,\widehat{\eta}_s,g_0(\alpha_{s+1}+\cdots+\alpha_l)) \;. \tag{10.40}$$

The sum α of the arguments of g_0 is thus $\alpha = \epsilon'(a_+ - a_-)$, where a_+, a_- are a pair of integers drawn from the multinomial distribution of parameters $(l;\beta_+,\beta_-)$. One can thus compute the first two cumulants of α as

$$\mathbb{E}[\alpha] = \epsilon' l(\beta_+-\beta_-) \;, \qquad \mathrm{Var}[\alpha] = (\epsilon')^2 l(\beta_++\beta_--(\beta_+-\beta_-)^2) \;. \tag{10.41}$$

In the limit we are considering one finds that these two quantities converge to $\widetilde{\epsilon}^2$, while the cumulants of higher order vanish, which show that α tends to a Gaussian distributed random variable with mean and variance both equal to $\widetilde{\epsilon}^2$; we

will denote the corresponding probability density as $D\underset{\epsilon}{\sim}\alpha = \frac{\mathrm{d}\alpha}{\sqrt{2\pi\widetilde{\epsilon^2}}} e^{-\frac{1}{2\widetilde{\epsilon^2}}(\alpha-\widetilde{\epsilon^2})^2}$. Collecting all these facts yields

$$
\begin{aligned}
Q_{w,n+1}(\eta) = \sum_{p,q,r=0}^{\infty} \mathcal{P}_{w,n}(p)\mathrm{Po}(q;x_{1,n})\mathrm{Po}(r;x_{0,n}) \int D\underset{\epsilon}{\sim}\alpha \prod_{i=1}^{p} \mathrm{d}Q_{0,n}(\eta_i) \\
\times \prod_{i=p+1}^{p+q} \mathrm{d}Q_{1,n}(\eta_i) \prod_{i=p+q+1}^{p+q+r} \mathrm{d}Q_{0,n}(\eta_i) \qquad (10.42) \\
\times \delta(\eta - f_{\Theta_1}(g_0(\alpha), g_+(\eta_1^f), \dots, g_+(\eta_p^f), g_+(\eta_{p+1})^f, \dots, g_+(\eta_{p+q})^f, g_+(\eta_{p+q+1}), \dots, g_+(\eta_{p+q+r}))) \ ,
\end{aligned}
$$

where we used the identity $g_-(\eta^f) = g_+(\eta)^f$ to transform q of the arguments of f_{Θ_1}. In this equation the function f_{Θ_1} is the one defined in Eq. (10.17), in which one can take $\epsilon = 0$ at this leading order; explicitly, $\eta = f_{\Theta_1}(\widehat{\eta}_1, \widehat{\eta}_2, \dots)$ means

$$
\begin{aligned}
\eta(\sigma,1) &= \frac{1}{z} b \prod_i \left(\widehat{\eta}_i(\sigma,0) + \widehat{\eta}_i(\sigma,1)\right) \ , \qquad (10.43) \\
\eta(\sigma,0) &= \frac{1}{z}\left[(1-b)\prod_i \widehat{\eta}_i(\sigma,0) + b\prod_i \left(\widehat{\eta}_i(\sigma,0) + \widehat{\eta}_i(\sigma,1)\right)\right] \ ,
\end{aligned}
$$

with $\sigma = \pm 1$.

The initial condition on $Q_{w,n=1}$ can be deduced after a short computation from the one on $\widehat{P}_{v,0}$ given in (7.62):

$$
Q_{w,1}(\eta) = \sum_{q,r=0}^{\infty} \mathrm{Po}(q;e^{-\gamma})\mathrm{Po}\left(r;\frac{e^{-\gamma}}{b}\right) \int D\underset{\epsilon}{\sim}\alpha \, \delta\left(\eta - f_{\Theta_1}\left(g_0\left(\alpha + \frac{q-r}{2}\ln b\right)\right)\right) \ , \qquad (10.44)
$$

for both values $w = 0, 1$. The explicit value of η for a given choice of α, q and r reads

$$
\eta(+,0) = \frac{1}{z} b^q e^{\alpha} \ , \quad \eta(+,1) = \frac{1}{z} b^{q+1} e^{\alpha} \ , \quad \eta(-,0) = \frac{1}{z} b^r e^{-\alpha} \ , \quad \eta(-,1) = \frac{1}{z} b^{r+1} e^{-\alpha} \ , \qquad (10.45)
$$

with z normalizing this distribution.

The recursion relation (10.42) bears on the two sequences of distributions $Q_{0,n}$ and $Q_{1,n}$; however the two sequences are not independent, and obey some symmetry properties, that follow from the equations (7.73,7.74). In the large k limit these relations translate into

$$
\begin{aligned}
Q_{w,n}(\eta^f) &= \frac{\eta(-,w)}{\eta(+,w)} Q_{w,n}(\eta) \ , \quad \text{or equivalently} \qquad (10.46) \\
\int \mathrm{d}Q_{w,n}(\eta) A(\eta) &= \int \mathrm{d}Q_{w,n}(\eta) A(\eta^f) \frac{\eta(-,w)}{\eta(+,w)} \ ,
\end{aligned}
$$

and

$$x_{0,n}Q_{0,n}(\eta) = \frac{\eta(+,0)}{\eta(+,1)} x_{1,n}Q_{1,n}(\eta) \ , \tag{10.47}$$
$$x_{0,n}\int \mathrm{d}Q_{0,n}(\eta)A(\eta) = x_{1,n}\int \mathrm{d}Q_{1,n}(\eta)A(\eta)\frac{\eta(+,0)}{\eta(+,1)} \ ,$$

for any function A such that the integrals exist. One can check by induction on n that the sequences $Q_{0,n}$ and $Q_{1,n}$ solution of (10.42) with the initial condition (10.44) do indeed satisfy these identities.

We can finally establish the scaling stated in (10.16) for the correlation function C_n, by simplifying the expression (7.90) in the large k limit. Observing in particular that the probability law for p is the same in this equation and in (7.95) with $w = 0$ (modulo the shift $l \to l + 1$ which is irrelevant in the limit), one finds after after a short computation the expression

$$\widetilde{C}_n = x_{0,n}\int \mathrm{d}Q_{0,n}(\eta)(1 - \widetilde{m}(\eta)) \ , \tag{10.48}$$

for the reduced correlation function $\widetilde{C}_n$, where we defined

$$\widetilde{m}(\eta) = \frac{\eta(+,0) - \eta(-,0)}{\eta(+,0) + \eta(-,0)} \ . \tag{10.49}$$

Note that $\widetilde{C}_n$ satisfies the inequalities $0 \le \widetilde{C}_n \le x_{0,n} = \widetilde{H}_n$, which are immediate consequences of the bounds $H_n \le C_n \le 1$ we obtained at finite k and of the definitions in (10.16). As a consistency check one can also derive the bounds on $\widetilde{C}_n$ directly in the large k formalism; one of them is obvious from the observation that $\widetilde{m}(\eta) \le 1$ for all η, the other one follows from the identity

$$\int \mathrm{d}Q_{0,n}(\eta)\, \widetilde{m}(\eta) = \int \mathrm{d}Q_{0,n}(\eta)\, \widetilde{m}(\eta)^2 \ge 0 \ , \tag{10.50}$$

which can be proven from the Bayes symmetry expressed in (10.46), using the test function $A(\eta) = \widetilde{m}(\eta)(1 - \widetilde{m}(\eta))$.

10.4 The limit of large distance n

Let us summarize what we have just achieved and underline the main equations that will be used in the following. We have obtained recursive equations, in which the parameter k has disappeared, that allow to compute the reduced correlation function $\widetilde{C}_n$ and its hard-fields contribution $\widetilde{H}_n$ introduced in (10.16). The latter can be obtained from the scalar recursion (10.31), it depends on γ and b, and the asymptotic expansion of the rigidity threshold is of the form (10.15) with a constant $\gamma_{\mathrm{r}}(b)$ easily determined from the large n behavior of $\widetilde{H}_n$: for $\gamma < \gamma_{\mathrm{r}}(b)$ one has $\widetilde{H}_n \to \infty$ as $n \to \infty$, while $\widetilde{H}_n$ remains bounded for $\gamma \ge \gamma_{\mathrm{r}}(b)$. The computation of the reduced correlation function $\widetilde{C}_n$ requires

instead the resolution of the functional recursion equation (10.42) on the distributions of the soft-fields $Q_{w,n}$, supplemented by the initial condition (10.44), from which $\widetilde{C}_n$ is computed using the equation (10.48). The sequence $\widetilde{C}_n$ depends on the parameters γ, b and $\widetilde{\epsilon}$, and the constant $\gamma_{\rm d}(b,\widetilde{\epsilon})$ in the asymptotic expansion of the dynamic threshold $l_{\rm d}$ is deduced from the large n asymptotics of $\widetilde{C}_n$ (if $\gamma < \gamma_{\rm d}(b,\widetilde{\epsilon})$ then $\widetilde{C}_n \to \infty$, while it remains bounded for $\gamma > \gamma_{\rm d}(b,\widetilde{\epsilon})$). We shall now discuss the computation of $\widetilde{C}_n$ in the large n limit, as the final step to complete the determination of $\gamma_{\rm d}(b,\widetilde{\epsilon})$.

10.4.1 For $\gamma > \gamma_{\rm r}(b)$

The most natural way to solve numerically the functional recursion equation (10.42) on $Q_{w,n}$ is to use the population dynamics algorithm already explained in Sec. 8.1, that consists of approximating $Q_{w,n}$ by the empirical distribution over a sample of $\mathcal{N}$ representative elements $\{\eta_1,\ldots,\eta_{\mathcal{N}}\}$. An iteration step $n \to n+1$ amounts to update the populations by drawing the integers p, q, and r from their respective laws, extracting the η_i's from the current populations, and creating an η of the new population according to the argument of the Dirac delta in (10.42). When $\gamma > \gamma_{\rm r}(b)$ this procedure can be performed without difficulty for arbitrarily large distances n, as the sequences $x_{0,n}, x_{1,n}$ remain bounded for all n.

Figure 10.3 presents numerical results obtained in this way for $b = 0.4$ and $\widetilde{\epsilon} = 0$. We have plotted on the top panel $\widetilde{C}_n(\gamma,b,0)$ as a function of n for some values of γ above the rigidity threshold $\gamma_{\rm r}(b) \approx 1.378$. One can see that $\widetilde{C}_n(\gamma,b,0)$ converges at large n to a finite limit $\widetilde{C}(\gamma,b,0)$, that we have plotted as a function of γ in the bottom panel, along with the limit $\widetilde{H}$ of $\widetilde{H}_n = x_{0,n}$. As we mentioned before the reduced overlap satisfy the bounds $0 \le \widetilde{C}_n \le x_{0,n}$, hence in the large n limit one has $0 \le \widetilde{C} \le \widetilde{H}$, that is indeed verified in the bottom panel of figure 10.3. This implies that $\widetilde{C}_n$ remains bounded for $\gamma > \gamma_{\rm r}(b)$, hence the expected inequality $\gamma_{\rm d}(b,\widetilde{\epsilon}) \le \gamma_{\rm r}(b)$. The observation of the bottom panel of figure 10.3 suggests the less obvious fact that this inequality is strict; indeed $\widetilde{H}$ has a square root singularity when $\gamma \to \gamma_{\rm r}^+$, as a consequence of the bifurcation it undergoes, while $\widetilde{C}$ seems pefectly smooth in this limit, suggesting that it remains finite down to a critical value $\gamma_{\rm d} < \gamma_{\rm r}$.

Unfortunately the most interesting regime $\gamma_{\rm d} < \gamma < \gamma_{\rm r}$ cannot be studied with the simple numerical procedure we just described: when $\gamma < \gamma_{\rm r}(b)$ the sequences $x_{0,n}$ and $x_{1,n}$ diverge, hence the random numbers p,q,r of fields η_i that must be manipulated to implement (10.42) become very quickly too large for any practical purpose. We shall thus devise in the next subsection an alternative formulation to circumvent this difficulty, that was used in particular to obtain the points of the curve $\widetilde{C}$ below the rigidity threshold in the bottom panel of figure 10.3.

In order to give an intuition on how this reformulation should be performed we first present in figure 10.4 the results of the simple procedure for γ slightly below $\gamma_{\rm r}$, and distances n not too large. One sees clearly in this plot that

$\widetilde{H}_n = x_{0,n}$ diverges, while $\widetilde{C}_n$ seems to remain bounded; the expression (10.48) of $\widetilde{C}_n$ reveals that such a situation is possible if $Q_{0,n}$ concentrates on fields η with $\widetilde{m}(\eta)$ very close to 1. By inspection of the populations in our numerical simulations we have checked that this is indeed the case, and more precisely that both $Q_{0,n}$ and $Q_{1,n}$ tend to a Dirac peak on the hard-field η^+. The finite value of $\widetilde{C}_n$ in the large n limit of the intermediate regime $\gamma_{\rm d}(b,\widetilde{\epsilon}) < \gamma < \gamma_{\rm r}(b)$ which is reconstructible without strictly hard-fields arises thus from a delicate compensation in the multiplication of the diverging factor $x_{0,n}$ and of the vanishing integral $\int \mathrm{d}Q_{0,n}(\eta)(1-\widetilde{m}(\eta))$. The relevant contribution of the latter arises from atypical values of η for which $Q_{0,n}$ is of order $1/x_{0,n}$, the typical values of $\eta \approx \eta^+$ having $1-\widetilde{m}(\eta) \approx 0$.

10.4.2 A reweighting scheme

To handle the difficulty that arises in the intermediate regime $\gamma_{\rm d} < \gamma < \gamma_{\rm r}$ we will adapt the approach we developed in chapter 9 (see [32]) for the uniform measure, introducing a reweighted probability distribution $\rho_n(\eta)$ that gives less importance to the typical quasi-hard fields that do not contribute to $\widetilde{C}_n$. We define it as

$$\rho_n(\eta) = x_{1,n} Q_{1,n}(\eta) \sqrt{\frac{\eta(-,1)}{\eta(+,1)}} \ , \tag{10.51}$$

the reweighting factor proportional to $\sqrt{\frac{\eta(-,1)}{\eta(+,1)}}$ indeed vanishes when η is a hard-field η^+. This choice also ensures the invariance of ρ_n under a spin-flip transformation, $\rho_n(\eta^f) = \rho_n(\eta)$, as can be easily seen from the first equality in (10.46) with $w = 1$. Note that ρ_n is a positive measure, but not a normalized probability measure anymore. One can nevertheless check that its total mass, that we shall denote m_n, is finite for all finite n. One has indeed, inverting the relation (10.51) and exploiting the normalization of $Q_{1,n}$,

$$\begin{aligned} x_{1,n} &= \int \mathrm{d}\rho_n(\eta) \sqrt{\frac{\eta(+,1)}{\eta(-,1)}} = \int \mathrm{d}\rho_n(\eta) \, \frac{1}{2} \left(\sqrt{\frac{\eta(+,1)}{\eta(-,1)}} + \sqrt{\frac{\eta(-,1)}{\eta(+,1)}} \right) \\ &= \int \mathrm{d}\rho_n(\eta) \, \frac{\eta(+,1)+\eta(-,1)}{2\sqrt{\eta(+,1)\eta(-,1)}} \ , \end{aligned} \tag{10.52}$$

where we used the invariance $\rho_n(\eta^f) = \rho_n(\eta)$ to symmetrize the integrand. Thanks to the inequality between arithmetic and geometric means of positive numbers the last integral is larger than m_n, which implies finally $m_n \leq x_{1,n}$. We can thus define a probability distribution ν_n by dividing ρ_n by its total mass, $\nu_n(\eta) = \rho_n(\eta)/m_n$. The problem at hand is equivalently described in terms of the $\{x_{w,n}, Q_{w,n}\}$, of ρ_n, or of the pair (ν_n, m_n); for instance the reduced overlap can be expressed as

$$\widetilde{C}_n = m_n \int \mathrm{d}\nu_n(\eta) C(\eta) \ , \quad C(\eta) = \frac{1}{\sqrt{\eta(+,1)\eta(-,1)}} \frac{2\eta(+,0)\eta(-,0)}{\eta(+,0)+\eta(-,0)} \ , \tag{10.53}$$

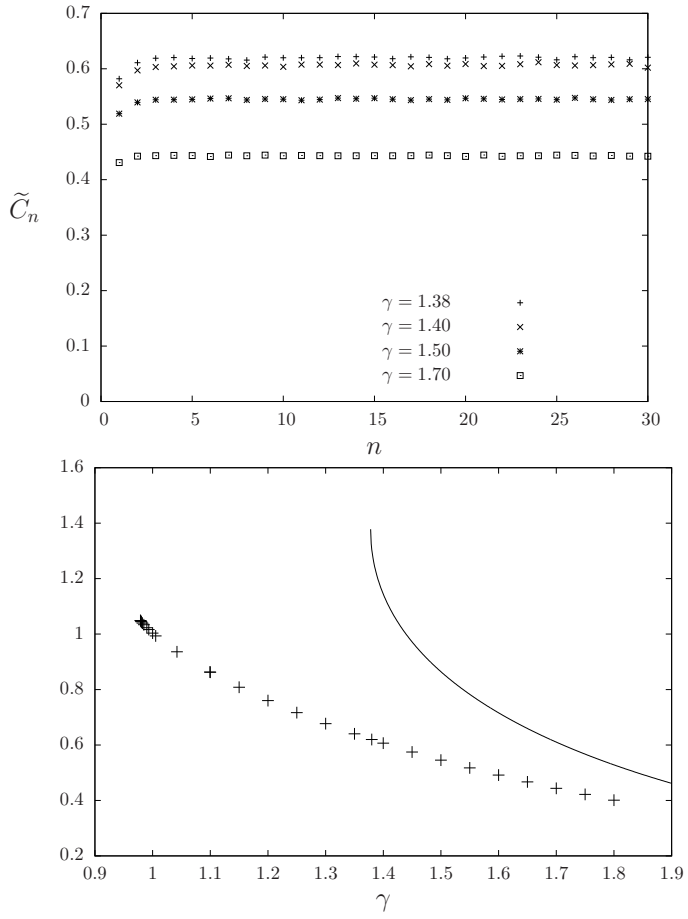

Figure 10.3: Top: the reduced correlation function $\widetilde{C}_n$ as a function of the distance n for $b = 0.4$, $\widetilde{\epsilon} = 0$, and several values of γ larger than the rigidity threshold $\gamma_{\rm r}(0.4) \approx 1.378$. Bottom: the large distance limit $\widetilde{C}$ (points) and its hard-field contribution $\widetilde{H}$ (solid line) as a function of γ for $b = 0.4$ and $\widetilde{\epsilon} = 0$. The points of $\widetilde{C}$ below the rigidity threshold have been obtained with the reweighted algorithm presented in Sec. 10.4.2.

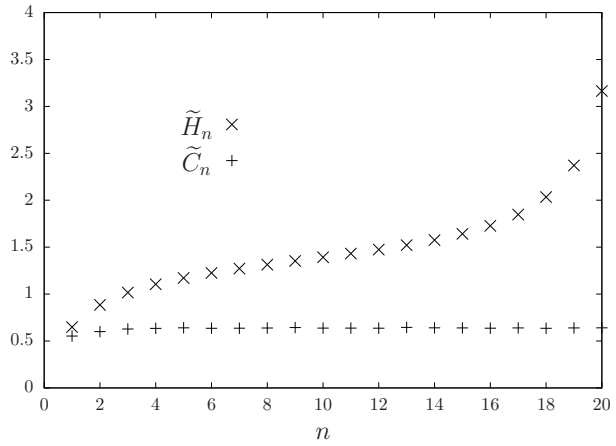

Figure 10.4: The reduced correlation function $\widetilde{C}_n$ and its hard-field contribution $\widetilde{H}_n$ as a function of the distance n for $b = 0.4$, $\widetilde{\epsilon} = 0$ and $\gamma = 1.35$, slightly below the rigidity threshold.

where we used (10.47) to transform the integral over $Q_{0,n}$ as one over $Q_{1,n}$. As we shall see the reweighted formulation is however much more convenient to study the large n limit, as it avoids the direct manipulation of the diverging quantities $x_{w,n}$.

We will now derive recursion relations for (ν_n, m_n); to do so it will be convenient to first introduce a different parametrization of the messages η. These are normalized probability distributions on a space of four states (σ, w), they can be thus encoded with three real numbers, that we shall choose as

$$u^{(1)} = \sqrt{\frac{\eta(+,1)}{\eta(-,1)}}\ , \quad u^{(2)} = \sqrt{\frac{\eta(+,0)-\eta(+,1)}{\eta(-,0)-\eta(-,1)}}\ ,$$
$$u^{(3)} = \sqrt{\frac{(\eta(+,0)-\eta(+,1))(\eta(-,0)-\eta(-,1))}{B^2\,\eta(+,1)\eta(-,1)}}\ . \tag{10.54}$$

We will group them as a row vector with three columns, $u = (u^{(1)}, u^{(2)}, u^{(3)})$, and define for later use the associated canonical basis $e^{(1)} = (1,0,0)$, $e^{(2)} = (0,1,0)$, $e^{(3)} = (0,0,1)$. Consider now the BP equation $\eta = f_{\Theta_1}(\widehat{\eta}_1, \widehat{\eta}_2, \dots)$ defined in (10.43); it becomes in terms of this parametrization

$$u^{(1)} = \prod_i \sqrt{\frac{\widehat{\eta}_i(+,0)+\widehat{\eta}_i(+,1)}{\widehat{\eta}_i(-,0)+\widehat{\eta}_i(-,1)}}\ , \qquad u^{(2)} = \prod_i \sqrt{\frac{\widehat{\eta}_i(+,0)}{\widehat{\eta}_i(-,0)}}\ , \tag{10.55}$$
$$u^{(3)} = \prod_i \sqrt{\frac{\widehat{\eta}_i(+,0)\widehat{\eta}_i(-,0)}{(\widehat{\eta}_i(+,0)+\widehat{\eta}_i(+,1))(\widehat{\eta}_i(-,0)+\widehat{\eta}_i(-,1))}}\ .$$

This shows that the arguments of the square roots in (10.54) are non-negative numbers, as they should for the definition of u to be meaningful. Moreover this expression reveals the motivation for this peculiar choice of parametrization: the BP equation $\eta = f_{\Theta_1}(\widehat{\eta}_1, \widehat{\eta}_2, \dots)$ becomes multiplicative with respect to its arguments when η is expressed in terms of u. We will also use the notation $\widetilde{u} = (\widetilde{u}^{(1)}, \widetilde{u}^{(2)}, \widetilde{u}^{(3)})$ with $u = (e^{\widetilde{u}^{(1)}}, e^{\widetilde{u}^{(2)}}, e^{\widetilde{u}^{(3)}})$; as the components of u are positive those of $\widetilde{u}$ are real numbers, and the BP equation becomes additive in terms of $\widetilde{u}$. It will also be useful to define the spin-flip operation on the triplets u and $\widetilde{u}$; as $\eta^f(\sigma, w) = \eta(-\sigma, w)$ one deduces easily from (10.54) the corresponding transformations:

$$u^f = \left(\frac{1}{u^{(1)}}, \frac{1}{u^{(2)}}, u^{(3)}\right) , \qquad \widetilde{u}^f = (-\widetilde{u}^{(1)}, -\widetilde{u}^{(2)}, \widetilde{u}^{(3)}) . \tag{10.56}$$

In the following we will take the liberty to use the three equivalent parametrizations η, u and $\widetilde{u}$ according to which one is the most convenient, keeping implicit the relationships between them that we have just defined.

Let us now rewrite (10.42) by translating the image η of the function f_{Θ_1} in the $\widetilde{u}$ parametrization:

$$\begin{aligned} Q_{w,n+1}(\eta) = \sum_{p,q,r=0}^{\infty} \mathcal{P}_{w,n}(p) \mathrm{Po}(q; x_{1,n}) \mathrm{Po}(r; x_{0,n}) \int D\underset{\epsilon}{\sim}\alpha \prod_{i=1}^{p} \mathrm{d}Q_{0,n}(\eta_i) \\ \times \prod_{i=p+1}^{p+q} \mathrm{d}Q_{1,n}(\eta_i) \prod_{i=p+q+1}^{p+q+r} \mathrm{d}Q_{0,n}(\eta_i) \\ \times \delta\left(\widetilde{u} - V_0(\alpha) - \sum_{i=1}^{p} V_+(\eta_i^f) - \sum_{i=p+1}^{p+q} V_-(\eta_i) - \sum_{i=p+q+1}^{p+q+r} V_+(\eta_i)\right) , \end{aligned} \tag{10.57}$$

where we defined

$$V_0(\alpha) = (\alpha, \alpha, 0) , \tag{10.58}$$

$$\begin{aligned} V_+(\eta) = \Bigg(\frac{1}{2}\ln\left(\frac{\eta(+,0)+\eta(-,0)}{\eta(+,1)}\right) , \\ \frac{1}{2}\ln\left(\frac{\eta(+,0)}{\eta(+,1)}\right) , \frac{1}{2}\ln\left(\frac{\eta(+,0)}{\eta(+,0)+\eta(-,0)}\right)\Bigg) , \end{aligned} \tag{10.59}$$

$$\begin{aligned} V_-(\eta) = V_+(\eta)^f = \Bigg(-\frac{1}{2}\ln\left(\frac{\eta(+,0)+\eta(-,0)}{\eta(+,1)}\right) , \\ -\frac{1}{2}\ln\left(\frac{\eta(+,0)}{\eta(+,1)}\right) , \frac{1}{2}\ln\left(\frac{\eta(+,0)}{\eta(+,0)+\eta(-,0)}\right)\Bigg) . \end{aligned} \tag{10.60}$$

For completeness we also state the expression of V_+ with its argument translated

in the u parametrization, namely

$$V_+^{(1)}(\eta) = \frac{1}{2}\ln\left(1 + B\frac{u^{(2)}u^{(3)}}{u^{(1)}} + \frac{1}{(u^{(1)})^2} + B\frac{u^{(3)}}{u^{(1)}u^{(2)}}\right) , \tag{10.61}$$

$$V_+^{(2)}(\eta) = \frac{1}{2}\ln\left(1 + B\frac{u^{(2)}u^{(3)}}{u^{(1)}}\right) , \tag{10.62}$$

$$V_+^{(3)}(\eta) = V_+^{(2)}(\eta) - V_+^{(1)}(\eta) , \tag{10.63}$$

Because of the additivity property of the parametrizations in terms of $\widetilde{u}$ it is easier to describe $Q_{w,n}$ in terms of its characteristic function, that we define as

$$\Xi_{w,n}(z) = \int \mathrm{d}Q_{w,n}(\eta)\, e^{iz\cdot\widetilde{u}} , \tag{10.64}$$

where $z = (z^{(1)}, z^{(2)}, z^{(3)})$ and we denoted the standard scalar product $z\cdot\widetilde{u} = z^{(1)}\widetilde{u}^{(1)} + z^{(2)}\widetilde{u}^{(2)} + z^{(3)}\widetilde{u}^{(3)}$. Indeed the equation (10.57) translates into

$$\begin{aligned}\Xi_{w,n+1}(z) = e^{\widetilde{\epsilon}^2(i(z^{(1)}+z^{(2)})-\frac{1}{2}(z^{(1)}+z^{(2)})^2)} \sum_{p,q,r=0}^{\infty} \mathcal{P}_{w,n}(p)\mathrm{Po}(q;x_{1,n}) \\ \times \mathrm{Po}(r;x_{0,n}) \left(\int \mathrm{d}Q_{0,n}(\eta) e^{iz\cdot V_+(\eta^f)}\right)^p \\ \times \left(\int \mathrm{d}Q_{1,n}(\eta) e^{iz\cdot V_-(\eta)}\right)^q \left(\int \mathrm{d}Q_{0,n}(\eta) e^{iz\cdot V_+(\eta)}\right)^r ,\end{aligned} \tag{10.65}$$

where the first factor comes from the Gaussian integration on α. For $w = 1$ the three integers p, q, r have Poisson distributions, the sums can then be easily performed to obtain

$$\begin{aligned}\Xi_{1,n+1}(z) = \exp\Big[\widetilde{\epsilon}^2(i(z^{(1)}+z^{(2)}) - \frac{1}{2}(z^{(1)}+z^{(2)})^2) - x_{1,n} - 2x_{0,n} \\ + x_{1,n}\int \mathrm{d}Q_{1,n}(\eta) e^{iz\cdot V_-(\eta)} \\ + x_{0,n}\int \mathrm{d}Q_{0,n}(\eta)\left(e^{iz\cdot V_+(\eta)} + e^{iz\cdot V_+(\eta^f)}\right)\Big] .\end{aligned} \tag{10.66}$$

We can now come back to the reweighted measure ρ_n we introduced in (10.51), and its normalized version ν_n, for which we define the characteristic functions similarly

$$\widehat{\rho}_n(z) = \int \mathrm{d}\rho_n(\eta)\, e^{iz\cdot\widetilde{u}} , \quad \widehat{\nu}_n(z) = \int \mathrm{d}\nu_n(\eta)\, e^{iz\cdot\widetilde{u}} = \frac{1}{m_n}\widehat{\rho}_n(z) . \tag{10.67}$$

The reweighting factor $\sqrt{\frac{\eta(-,1)}{\eta(+,1)}}$ between ρ_n and $Q_{1,n}$ can be expressed as $e^{-\widetilde{u}^{(1)}}$, the characteristic functions of these two measures are thus linked by a simple shift of their arguments:

$$\rho_n(\eta) = x_{1,n}Q_{1,n}(\eta)e^{-\widetilde{u}^{(1)}} \iff \widehat{\rho}_n(z) = x_{1,n}\Xi_{1,n}(z + ie^{(1)}) . \tag{10.68}$$

Using this shift of argument in (10.66), and recalling from (10.32) that $x_{1,n+1} = e^{-\gamma+x_{0,n}}$ we obtain:

$$\begin{aligned}\widehat{\rho}_{n+1}(z) = \exp\Bigg[&-\gamma - \frac{\widetilde{\epsilon}^2}{2} - \frac{\widetilde{\epsilon}^2}{2}(z^{(1)}+z^{(2)})^2 - x_{0,n} - x_{1,n} \\ &+x_{1,n}\int \mathrm{d}Q_{1,n}(\eta)\sqrt{\frac{\eta(+,0)+\eta(-,0)}{\eta(+,1)}}e^{iz\cdot V_-(\eta)} \\ &+x_{0,n}\int \mathrm{d}Q_{0,n}(\eta)\left(\sqrt{\frac{\eta(+,1)}{\eta(+,0)+\eta(-,0)}}e^{iz\cdot V_+(\eta)}\right. \\ &\left.+\sqrt{\frac{\eta(-,1)}{\eta(+,0)+\eta(-,0)}}e^{iz\cdot V_+(\eta^f)}\right)\Bigg] \ . \end{aligned} \tag{10.69}$$

We will now trade the integrations over $Q_{0,n}$ and $Q_{1,n}$ for integrals over ρ_n, thanks to the change of densities expressed in (10.47) and (10.51). We will also write $x_{0,n} + x_{1,n} = 2x_{1,n} + Be^{-\gamma}$ according to (10.32), and write $2x_{1,n}$ as an integral over ρ_n following (10.52). This yields

$$\begin{aligned}\widehat{\rho}_{n+1}(z) = \exp\Bigg[&-\gamma - B\,e^{-\gamma} - \frac{\widetilde{\epsilon}^2}{2} - \frac{\widetilde{\epsilon}^2}{2}(z^{(1)}+z^{(2)})^2 \\ &+\int \mathrm{d}\rho_n(\eta)\sqrt{\frac{\eta(+,0)+\eta(-,0)}{\eta(-,1)}}\,e^{iz\cdot V_-(\eta)} \\ &+\int \mathrm{d}\rho_n(\eta)\left(\frac{\eta(+,0)}{\sqrt{\eta(-,1)(\eta(+,0)+\eta(-,0))}}\,e^{iz\cdot V_+(\eta)}\right. \\ &\left.+\frac{\eta(+,0)}{\sqrt{\eta(+,1)(\eta(+,0)+\eta(-,0))}}\,e^{iz\cdot V_+(\eta^f)}\right) \\ &-\int \mathrm{d}\rho_n(\eta)\frac{\eta(+,1)+\eta(-,1)}{\sqrt{\eta(+,1)\eta(-,1)}}\Bigg] \ . \end{aligned} \tag{10.70}$$

Using the invariance under spin-flip of ρ_n one can regroup the two terms in the second line of this equation; simplifying the prefactors one obtains

$$\begin{aligned}\widehat{\rho}_{n+1}(z) = \exp\Bigg[&-\gamma - B\,e^{-\gamma} - \frac{\widetilde{\epsilon}^2}{2} - \frac{\widetilde{\epsilon}^2}{2}(z^{(1)}+z^{(2)})^2 \\ &+\int \mathrm{d}\rho_n(\eta)\left(\sqrt{\frac{\eta(+,0)+\eta(-,0)}{\eta(-,1)}}\left(e^{iz\cdot V_+(\eta)}+e^{iz\cdot V_-(\eta)}\right)\right. \\ &\left.-\frac{\eta(+,1)+\eta(-,1)}{\sqrt{\eta(+,1)\eta(-,1)}}\right)\Bigg] \ . \end{aligned} \tag{10.71}$$

This is a recursion equation for the reweighted measure ρ_n (and its characteristic function $\widehat{\rho}_n$). It will be more convenient in the following to work with the pair

(ν_n, m_n); the mass m_n of ρ_n can be expressed as $\widehat{\rho}_n(0)$, we thus obtain

$$m_{n+1} = \exp\left[-\gamma - B\,e^{-\gamma} - \frac{\widetilde{\epsilon}^2}{2} + m_n \int \mathrm{d}\nu_n(\eta) M(\eta)\right] , \tag{10.72}$$

$$\widehat{\nu}_{n+1}(z) = \exp\left[-\frac{\widetilde{\epsilon}^2}{2}(z^{(1)} + z^{(2)})^2 + m_n \int \mathrm{d}\nu_n(\eta) L(\eta) \left(e^{iz\cdot V_+(\eta)} + e^{iz\cdot V_-(\eta)} - 2\right)\right] , \tag{10.73}$$

where we introduced the functions

$$M(\eta) = \frac{\sqrt{\eta(+,0)+\eta(-,0)}(\sqrt{\eta(+,1)}+\sqrt{\eta(-,1)}) - \eta(+,1) - \eta(-,1)}{\sqrt{\eta(+,1)\eta(-,1)}} , \tag{10.74}$$

$$L(\eta) = \sqrt{\frac{\eta(+,0)+\eta(-,0)}{\eta(-,1)}} . \tag{10.75}$$

The initial condition for the recursion on (ν_n, m_n) is obtained from the one on $Q_{1,1}$ given in (10.44):

$$m_1 = \exp\left[-\gamma + e^{-\gamma}\left(\frac{2}{\sqrt{b}} - \frac{1}{b} - 1\right) - \frac{\widetilde{\epsilon}^2}{2}\right] , \tag{10.76}$$

$$\nu_1(\eta) = \sum_{q,r=0}^{\infty} \mathrm{Po}\left(q; \frac{e^{-\gamma}}{\sqrt{b}}\right) \mathrm{Po}\left(r; \frac{e^{-\gamma}}{\sqrt{b}}\right) \int \frac{\mathrm{d}\alpha}{\sqrt{2\pi\widetilde{\epsilon}^2}} e^{-\frac{\alpha^2}{\widetilde{\epsilon}^2}} \delta\left(\widetilde{u} - V_0\left(\alpha + \frac{q-r}{2}\ln b\right)\right) . \tag{10.77}$$

For completeness we give here the expressions of the functions we introduced in terms of the u-parametrization:

$$L(\eta) = \sqrt{1 + B\frac{u^{(1)}u^{(3)}}{u^{(2)}} + (u^{(1)})^2 + Bu^{(1)}u^{(2)}u^{(3)}} , \tag{10.78}$$

$$M(\eta) = \sqrt{1 + B\frac{u^{(2)}u^{(3)}}{u^{(1)}} + \frac{1}{(u^{(1)})^2} + B\frac{u^{(3)}}{u^{(1)}u^{(2)}}} - \frac{1}{u^{(1)}} \tag{10.79}$$

$$+ \sqrt{1 + B\frac{u^{(1)}u^{(3)}}{u^{(2)}} + (u^{(1)})^2 + Bu^{(1)}u^{(2)}u^{(3)}} - u^{(1)} , \tag{10.80}$$

$$C(\eta) = 2\frac{\left(1 + B\frac{u^{(2)}u^{(3)}}{u^{(1)}}\right)\left(1 + B\frac{u^{(1)}u^{(3)}}{u^{(2)}}\right)}{u^{(1)} + Bu^{(2)}u^{(3)} + \frac{1}{u^{(1)}} + B\frac{u^{(3)}}{u^{(2)}}} . \tag{10.81}$$

10.4.3 A Gaussian approximation for the quasi-hard fields

We have obtained above the recursion equations (10.72,10.73) for the scalar m_n and the probability distribution ν_n, complemented by the initial conditions (10.76,10.77). We will now discuss the possibility to solve numerically this recursion with a population representation of ν_n, and its advantages with respect

to the direct resolution in terms of $Q_{w,n}$. To do so let us first rewrite the recursion equation (10.73) on ν_n as

$$\widehat{\nu}_{n+1}(z) = \exp\left[-\frac{\widetilde{\epsilon}^2}{2}(z^{(1)}+z^{(2)})^2\right]\exp\left[\int \mathrm{d}\pi_n(\widetilde{u})(e^{iz\cdot\widetilde{u}}-1)\right] \ , \tag{10.82}$$

where we have introduced a measure π_n of total mass we shall denote λ_n, according to

$$\pi_n(\widetilde{u}) = m_n \int \mathrm{d}\nu_n(\eta)L(\eta)\left(\delta(\widetilde{u}-V_+(\eta))+\delta(\widetilde{u}-V_-(\eta))\right) \ , \tag{10.83}$$

$$\begin{aligned}\lambda_n &= \int \mathrm{d}\pi_n(\widetilde{u}) = 2m_n\int \mathrm{d}\nu_n(\eta)L(\eta) \\ &= m_n\int \mathrm{d}\nu_n(\eta)\sqrt{\frac{\eta(+,0)+\eta(-,0)}{\eta(+,1)\eta(-,1)}}\left(\sqrt{\eta(+,1)}+\sqrt{\eta(-,1)}\right) \ ,\end{aligned} \tag{10.84}$$

where the last expression of λ_n has been obtained by symmetrizing the integrand.

According to the equation (10.82) a random variable $\widetilde{u}$ drawn from ν_{n+1} can be decomposed as the sum of two random variables, one Gaussian distributed and the other with a compound Poisson distribution. More explicitly one has the following equality in distribution, $\widetilde{u} \stackrel{d}{=} \alpha(e^{(1)}+e^{(2)}) + \sum_{i=1}^{p}\widetilde{u}_i$ where α is a Gaussian with zero mean and variance $\widetilde{\epsilon}^2$, p is extracted from a Poisson law of mean λ_n, and the $\widetilde{u}_i$'s are i.i.d. copies extracted from π_n/λ_n. If ν_n is known as an empirical distribution over a sample then it is possible to draw $\widetilde{u}$ from the probability law π_n/λ_n by extracting a field η in the population representing ν_n with a probability proportional to $L(\eta)$, and then set $\widetilde{u} = V_\delta(\eta)$ with $\delta = \pm$ with equal probability $1/2$. It seems then possible to use this distributional interpretation to solve numerically the recursion on (m_n, ν_n). However this is doable in practice only if λ_n remains bounded when n grows, otherwise one falls back on the problem we wanted to avoid of having to manipulate a diverging number of summands. As a matter of fact the reweighting has not offered a free lunch from this point of view: it turns out that λ_n diverges if and only if $x_{1,n}$ does, in other words if and only if $\gamma < \gamma_{\mathrm{r}}(b)$. This statement is a consequence of the bounds $c_-(b)x_{1,n} \leq \lambda_n \leq c_+(b)x_{1,n}$, where $c_\pm(b)$ are positive constants, the proof of which we defer to the Appendix C for the sake of readability.

Fortunately the reweighting procedure we followed will help us to handle the divergence of λ_n more easily than the one of $x_{1,n}$ in the direct recursion. Indeed the divergence of λ_n comes from the contributions of fields for which $L(\eta)$ becomes very large; the crucial point is that these η yield very small values of $V_\pm(\eta)$, we can thus make a Gaussian approximation for this sum of a very large number of very small random variables. To put this idea at work we rewrite (10.82) by decomposing it as

$$\widehat{\nu}_{n+1}(z) = \exp\left[-\frac{\widetilde{\epsilon}^2}{2}(z^{(1)}+z^{(2)})^2\right]\widehat{\nu}^{(\leq)}_{n+1}(z)\,\widehat{\nu}^{(>)}_{n+1}(z) \ , \tag{10.85}$$

with

$$\widehat{\nu}_{n+1}^{(\leq)}(z) = \exp\left[m_n \int \mathrm{d}\nu_n(\eta) L(\eta) \left(e^{iz\cdot V_+(\eta)} + e^{iz\cdot V_-(\eta)} - 2\right) \mathbb{I}\left[L(\eta) \leq \xi_n\right]\right] , \tag{10.86}$$

$$\widehat{\nu}_{n+1}^{(>)}(z) = \exp\left[m_n \int \mathrm{d}\nu_n(\eta) L(\eta) \left(e^{iz\cdot V_+(\eta)} + e^{iz\cdot V_-(\eta)} - 2\right) \mathbb{I}\left[L(\eta) > \xi_n\right]\right] , \tag{10.87}$$

where ξ_n is a threshold that is arbitrary for the moment, we shall specify it later on. The decomposition (10.85) means that under the law ν_{n+1} the random variable $\widetilde{u}$ is the sum of the Gaussian random variable described previously and of two random variables, one with the law $\nu_{n+1}^{(\leq)}$, the other with the law $\nu_{n+1}^{(>)}$.

We describe the distribution $\nu_{n+1}^{(\leq)}$ using the distributional interpretation explained above, defining

$$\pi_n^{(\leq)}(\widetilde{u}) = m_n \int \mathrm{d}\nu_n(\eta) L(\eta) \left(\delta(\widetilde{u} - V_+(\eta)) + \delta(\widetilde{u} - V_-(\eta))\right) \mathbb{I}\left[L(\eta) \leq \xi_n\right] , \tag{10.88}$$

$$\lambda_n^{(\leq)} = \int \mathrm{d}\pi_n^{(\leq)}(\widetilde{u}) = 2m_n \int \mathrm{d}\nu_n(\eta) L(\eta) \mathbb{I}\left[L(\eta) \leq \xi_n\right] . \tag{10.89}$$

Under the law $\nu_{n+1}^{(\leq)}$ the variable $\widetilde{u}$ obeys the distributional equality $\widetilde{u} = \sum_{i=1}^{p} \widetilde{u}_i$ where p is a Poisson variable of mean $\lambda_n^{(\leq)}$, and the $\widetilde{u}_i$'s are i.i.d copies extracted from $\pi_n^{(\leq)}/\lambda_n^{(\leq)}$.

The contribution $\nu_{n+1}^{(>)}$ is instead approximated by a multivariate Gaussian $\Gamma(\overline{V}_n, \Sigma_n)$ with $\overline{V}_n$ and Σ_n the mean and the covariance matrix of $\nu_{n+1}^{(>)}$, computed by taking derivatives of $\ln \widehat{\nu}_{n+1}^{(>)}$ with respect to z:

$$\overline{V}_n = m_n \int \mathrm{d}\nu_n(\eta) L(\eta) \mathbb{I}\left[L(\eta) > \xi_n\right] \left(V_+(\eta) + V_-(\eta)\right) , \tag{10.90}$$

$$\Sigma_n^{(a),(b)} = m_n \int \mathrm{d}\nu_n(\eta) L(\eta) \mathbb{I}\left[L(\eta) > \xi_n\right] \left(V_+(\eta)^{(a)} V_+(\eta)^{(b)} + V_-(\eta)^{(a)} V_-(\eta)^{(b)}\right) , \tag{10.91}$$

for $a, b \in \{1, 2, 3\}$. As $V_-(\eta) = (V_+(\eta))^f$ several components of $\overline{V}_n$ and Σ_n vanish, namely $\overline{V}_n^{(1)} = \overline{V}_n^{(2)} = \Sigma_n^{(1),(3)} = \Sigma_n^{(2),(3)} = 0$.

Replacing $\nu_{n+1}^{(>)}$ by a Gaussian is an approximation, that amounts to neglect the cumulants of order larger than 2, the accuracy of which is controlled by the cutoff ξ_n. The larger is ξ_n the better the truncation is, because a smaller part of the full law ν_{n+1} is treated approximatively, but the price to pay is a simultaneous increase of $\lambda_n^{(\leq)}$, the average number of fields that must be summed in desribing $\nu_{n+1}^{(\leq)}$. A compromise needs thus to be found between these two effects, we explain below how we fixed ξ_n in practice.

10.4.4 Algorithmic implementation

We now give an explicit description of the algorithm we implemented to solve the recursion equations (10.72,10.73) for m_n and ν_n. Suppose that at the n-th step of the iteration we have an estimation of m_n and of ν_n, with ν_n represented as a population of fields:

$$\nu_n(\eta) \simeq \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \delta(\eta - \eta_i) \ . \tag{10.92}$$

One can evaluate the average of an arbitrary function A with respect to ν_n as

$$\int \mathrm{d}\nu_n(\eta) A(\eta) \simeq \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} A(\eta_i) \ , \tag{10.93}$$

and in particular compute in this way m_{n+1} from (10.72). We further assume that the fields η_i have been sorted by increasing values of $L(\eta)$, and translate the cutoff ξ_n by defining the index $\mathcal{N}_n$ such that $L(\eta_{\mathcal{N}_n}) \leq \xi_n < L(\eta_{\mathcal{N}_n+1})$. The integrals where the indicator function $\mathbb{I}[L(\eta) \leq \xi_n]$ (resp. $\mathbb{I}[L(\eta) > \xi_n]$) can thus be translated as sums over the population elements from 1 to $\mathcal{N}_n$ (resp. from $\mathcal{N}_n + 1$ to $\mathcal{N}$), which allows to compute easily $\lambda_n^{(\leq)}$ from (10.89), and $\overline{V}_n$ and Σ_n from (10.90) and (10.91).

Each of the $\mathcal{N}$ elements of the new population representing ν_{n+1} is then generated independently of the others, by translating equation (10.85) as follows:

- draw an integer p from the Poisson law of mean $\lambda_n^{(\leq)}$.
- extract $i_1, \ldots, i_p$ independently in $\{1, \ldots, \mathcal{N}_n\}$ with probability proportional to $L(\eta_i)$ (this can be done efficiently by precomputing a cumulative table).
- insert in the new population $\widetilde{u} = V_{\delta_1}(\eta_{i_1}) + \cdots + V_{\delta_p}(\eta_{i_p}) + \alpha(e^{(1)} + e^{(2)}) + \overline{V}_n + g$ where the δ_i's are $\pm$ with equal probability, α is a centered Gaussian random variable of variance $\widetilde{\epsilon}^2$, and g is a centered three-dimensional Gaussian vector with covariance matrix Σ_n.

We can then compute the reduced overlap $\widetilde{C}_{n+1}$ from (10.53), and sort the elements of the new population according to their values of $L(\eta)$.

In practice we chose the threshold ξ_n (or equivalently $\mathcal{N}_n$) in an adaptive way: for each iteration step we took the largest $\mathcal{N}_n \leq \mathcal{N}$ that gave $\lambda_n^{(\leq)} \leq \lambda$, where λ is a parameter fixed beforehand. The accuracy of this numerical procedure is thus controlled by $\mathcal{N}$, the approximation in (10.92) being better and better as $\mathcal{N}$ grows, and by λ, the Gaussian truncation being more precise when λ is larger. Obviously the memory and time requirements of the procedure also increase with $\mathcal{N}$ and λ; the numerical results presented below have been obtained with population sizes between 10^6 and 10^7, and λ around 20, we checked that the conclusions were not modified, within our numerical accuracy, by modifying these values within a reasonable range.

10.5 Results

In Fig. 10.5 we complete our study of the case $b = 0.4$, $\widetilde{\epsilon} = 0$ that was started in Figs. 10.3 and 10.4. The reweighting procedure allows now to investigate the regime $\gamma < \gamma_{\rm r} \approx 1.378$; as displayed on the top panel of Fig. 10.5 the large distance limit $\widetilde{C}$ remains bounded for values of γ down to 0.98, these results being reported as a function of γ in the bottom panel of Fig. 10.3. Further simulations allowed us to pinpoint more precisely $\gamma_{\rm d}(b = 0.4, \widetilde{\epsilon} = 0) \approx 0.977$, as the largest value of γ for which $\widetilde{C}_n$ diverges. As shown in the bottom panel of Fig. 10.5 the condition of divergence of $\widetilde{C}_n$ coincides, within our numerical accuracy, with the divergence of m_n. The value of the large n limit of m_n is seen to be close to 1 when γ reaches $\gamma_{\rm d}$ from above (see the plateau in the bottom panel of Fig. 10.5), an observation that we also made for the other values of b we investigated. We have given analytical arguments in [32] that indeed the plateau value of m_n is exactly equal to 1 at $\gamma_{\rm d}$ for the uniform measure, our numerical results suggest that this remains true when $(b, \widetilde{\epsilon}) \neq (1, 0)$, even if we do not have analytical support for this assumption in the general case.

We have repeated this procedure of determination of $\gamma_{\rm d}(b, \widetilde{\epsilon})$ for various values of b and $\widetilde{\epsilon}$, and we present now the phase diagrams obtained in this way. Consider first the top panel of Fig. 10.6, which deals with the case $b = 1$, i.e. the bias factorized over the hyperedges considered in [31]. We see that for all values of $\widetilde{\epsilon} \neq 0$ one has $\gamma_{\rm d}(1, \widetilde{\epsilon}) < \gamma_{\rm d}(1, 0) = \gamma_{\rm d,u}$, i.e. this bias has, in the large k limit, a detrimental effect on the dynamic phase transition that is pushed to lower values with respect to the one of the uniform measure. In the bottom panel of Fig. 10.6 we have plotted instead the threshold $\gamma_{\rm d}$ as a function of b for $\widetilde{\epsilon} = 0$; one sees now that decreasing b below 1 (that corresponds to the uniform measure and is marked as an horizontal dashed line on figure) has a beneficial effect with an increase of $\gamma_{\rm d}$. The largest value we could reach was for $b = 0.4$, decreasing b further below reduces again $\gamma_{\rm d}$. The lowest value of b we could investigate was $b = 0.3$, for $b < 0.3$ we encountered numerical accuracy problems, the distribution ν_n exhibiting strong fluctuations that prevented an accurate representation as a population. Finally in Fig. 10.7 we have checked that the parameter $\widetilde{\epsilon}$ has a detrimental effect also for values of $b \neq 1$, we found indeed that $\gamma_{\rm d}(b, \widetilde{\epsilon}) < \gamma_{\rm d}(b, 0)$ when $\widetilde{\epsilon} \neq 0$, for all the values of b we considered. This leads us to the conclusion that, within the biasing strategy we considered in the large k limit, the optimal choice of parameters is $\widetilde{\epsilon} = 0$ and $b \approx 0.4$, yielding a constant $\gamma_{\rm d} \approx 0.977$, strictly larger than the one of the uniform case, $\gamma_{\rm d,u} \approx 0.871$.

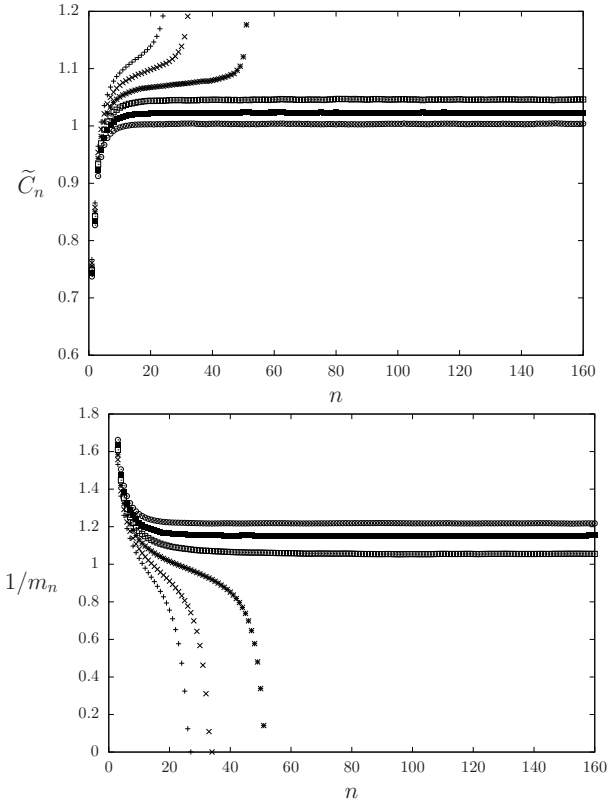

Figure 10.5: The reduced correlation function $\widetilde{C}_n$ (top panel) and the inverse of the mass m_n of ρ_n (bottom panel) for $b = 0.4$, $\widetilde{\epsilon} = 0$, and from left to right in both panels, $\gamma = 0.95, 0.96, 0.97, 0.98, 0.99, 1$.

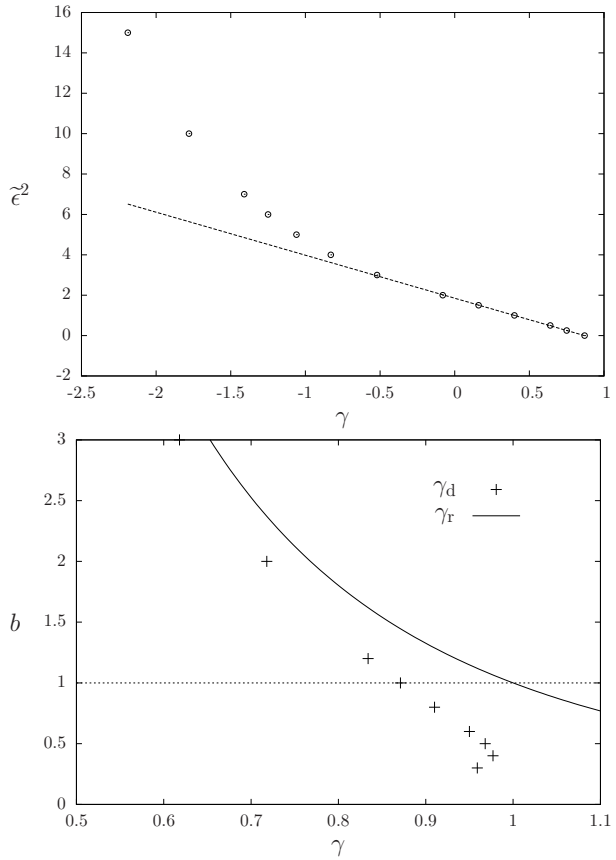

Figure 10.6: Top panel: $\gamma_{\rm d}$ as a function of $\widetilde{\epsilon}^2$ for $b = 1$, the line being a guide to the eye. Bottom panel: $\gamma_{\rm d}$ as a function of b for $\widetilde{\epsilon} = 0$, the solid line corresponding to the rigidity upper bound $\gamma_{\rm r}(b)$.

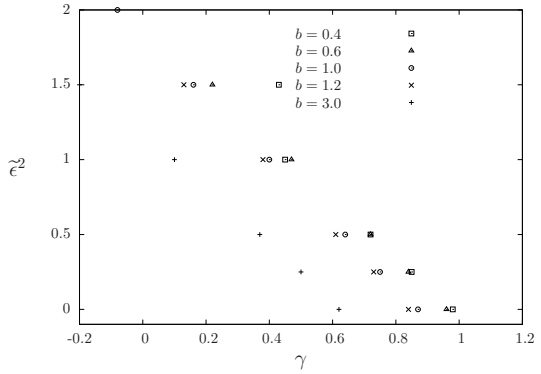

Figure 10.7: $\gamma_{\rm d}$ as a function of $\widetilde{\epsilon}$ for various b.

Conclusion

In this Ph.D. we have studied the typical difficulty of Constraint Satisfaction Problems, via the introduction of random ensemble of instances, focusing on the bicoloring problem on k-hypergraphs. Following the studies [26, 27, 28, 29, 30], we have introduced a biased probability measure over the set of solutions of a random CSP. Our goal was to study the modifications of the clustering threshold α_d induced by the non-uniformity between solutions. We have focused on two particular implementations of the bias. In the first one, the interactions induced by the bias can be factorized over the bicoloring constraints. In the second biased measure we incorporate interactions between variables belonging to different clauses (in practice we have considered interactions between variables at distance 1 in the k-hypergraph).

Specifying further the choice of the bias, we have weighted the solutions according to the presence of forcing clauses, where one says that a clause is forcing a variable if the clause becomes violated when changing the value of the variable. Forcing clauses are related to the presence of frozen variables, which are one of the mechanism for the RSB phenomenon that arises at the clustering transition. This choice can be also motivated by the observation that algorithms usually reach solutions without frozen variables, hence favoring solutions without forcing clauses would bias the measure toward solutions that are easier to find for algorithms.

We have determined the clustering transition for the two different implementations of the bias mentioned above. For finite k values, α_d can be computed via the numerical resolution of the 1RSB equations. In the paper [31] written in collaboration with Federico Ricci-Tersenghi, that is presented in chapter 8 (sections 8.2-8.4), we have shown that for well-chosen parameters defining the biased measure factorizing on the clauses, it is possible to delay the clustering threshold to higher density of constraints. We have checked that this strategy has a positive impact on the performances of Simulated Annealing algorithms. Introducing a bias in the measure reduces its entropy, which is maximal for the uniform measure. Although this effect represents a limitation for the possible choices of the bias, our results show that reducing the weight of some of the solutions can help in the search of the remaining ones.

From a practical point of view, we have also developed a numerical method to determine accurately the dynamic threshold, that corresponds to discontinuous bifurcation in a functional equation, by studying a stability parameter close to

the bifurcation (see section 8.2).

In the paper [33], that is presented in section 8.5 (for the finite k results) and in chapter 10 (for the large k analysis), we have studied the biased measure with a larger range of interaction. We have shown that for finite k values, increasing the range of interaction allows us to delay further the clustering threshold.

In this Ph.D. we have also determined the dynamic transition in the large k limit through a partly analytic asymptotic expansion. In the paper [32], that is presented in chapter 9, we have obtained for the bicoloring problem on k-hypergraph the following asymptotic expansion of the clustering threshold:

$$\alpha_d(k) = \frac{2^{k-1}}{k} \left(\ln k + \ln \ln k + \gamma_d + o(1) \right) \ ,$$

for the uniform measure, with $\gamma_{d,u} \approx 0.871$. We have shown that the q-coloring problem admits a similar asymptotic expansion with the same constant γ_d: $\alpha_d(q) = (q/2)(\ln q + \ln \ln q + \gamma_d + o(1))$.

In the paper [33] we adapt this computation to the two biased measures studied in the Ph.D., for the bicoloring problem on k-hypergraphs. This large k limit computation is presented in chapter 10. We show that the clustering threshold for our implementation of the bias arises at the same scale as in the uniform case, but with a constant γ_d that depends on the rescaled parameters describing the bias. We find that the factorized bias of [32] cannot improve on the constant γ_d with respect to the value $\gamma_{d,u} \approx 0.871$ obtained in the uniform case. We find instead that the bias with larger interaction range allows us to increase its value up to $\gamma_d \approx 0.977$.

The improvement in the asymptotic behavior of the dynamic transition is modest, as it occurs in the third order of the asymptotic expansion. Note also that the constant γ_{d} remains itself smaller than the rigidity threshold $\gamma_{\mathrm{r,u}} = 1$ of the uniform measure. However, the conceptual link between this transition and the important algorithmic gap observed in this limit justifies for us the efforts devoted to achieve this improvement, and calls for further investigations.

The natural direction to continue this work would be indeed to study a more general form of the bias, and to try to determine which choice would lead to the optimal scaling in the large k limit. In particular, is it possible to design a bias that improves on the second or first order in the asymptotic expansion compared to the asymptotic expansion of the uniform measure, or can we show that for some family of biased measures the only improvement bears on the third term of the asymptotic expansion ? The next step could be for instance to consider a more general form for the function $\psi(p)$ defined in equation (10.14), retaining more information on the local configuration than just the number of forcing clauses around one variable. One could also increase the range of interaction to distance 2 or more. Both positive or negative results, i.e. the impossibility to increase α_{d} beyond the third order term of the asymptotic expansion, would shed light on the intrisic difficulty of random CSPs.

Another interesting question to investigate in the future is the effect of the decimation on the biased measure, which can be analyzed along the lines of [37]. Can we improve the performances of the Belief Propagation guided decimation

(or the Survey Propagation guided decimation) algorithms, by using a biased measure on which one estimates the marginals ?

Finally it would also be interesting to extend this study to other CSPs, in particular to the k-satisfiability and the q-coloring problems. We have indeed focused our study on the k-hypergraph bicoloring problem because the computations are simpler for this model than for k-SAT or q-COL. Indeed the degree of freedom are binary, and the replica symmetric solution is trivial. At the same time this model exhibits the same phase transitions than more complicated random CSPs. We believe that the strategy developed in this Ph.D. can be extended to these models, and we expect a similar increase on the clustering threshold for an appropriate choice of the bias.

Bibliography

[1] M.R. Garey and D.S. Johnson. *Computers and intractability: A guide to the theory of NP-completeness.* Freeman, San Francisco, 1979.

[2] C. H. Papadimitriou. *Computational complexity.* Addison-Wesley, 1994.

[3] C. Moore and S. Mertens. *The Nature of Computation.* Oxford University Press, 2011.

[4] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky. 2+p-sat: Relation of typical-case complexity to the nature of the phase transition. *Random Structures and Algorithms*, 15:414, 1999.

[5] G. Biroli, R. Monasson, and M. Weigt. A variational description of the ground state structure in random satisfiability problems. *Eur. Phys. J. B*, 14:551, 2000.

[6] M. Mézard, G. Parisi, and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297:812–815, 2002.

[7] Stephan Mertens, Marc Mézard, and Riccardo Zecchina. Threshold values of random k-sat from the cavity method. *Random Struct. Algorithms*, 28(3):340–373, 2006.

[8] Florent Krzakala, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborova. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

[9] Dimitris Achlioptas and Federico Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In *Proc. of 38th STOC*, pages 130–139, New York, NY, USA, 2006. ACM.

[10] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proceedings of FOCS 2008*, page 793, 2008.

[11] M. Molloy. The freezing threshold for k-colourings of a random graph. In *Proceedings of the 44th symposium on Theory of Computing*, page 921. ACM, 2012.

[12] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 59–68, 2015.

[13] Elchanan Mossel and Yuval Peres. Information flow on trees. *Ann. Appl. Probab.*, 13(3):817–844, 08 2003.

[14] A. Montanari and G. Semerjian. Rigorous inequalities between length and time scales in glassy systems. *J. Stat. Phys.*, 125:23, 2006.

[15] M. Mézard and G. Parisi. The bethe lattice spin glass revisited. *Eur. Phys. J. B*, 20:217, 2001.

[16] Marc Mézard and Andrea Montanari. Reconstruction on trees and spin glass transition. *J. Stat. Phys.*, 124:1317–1350, september 2006.

[17] Bart Selman, Henry A. Kautz, and Bram Cohen. Noise strategies for improving local search. In *Proc. 12th AAAI*, pages 337–343, Menlo Park, CA, USA, 1994. AAAI Press.

[18] John Ardelius and Erik Aurell. Behavior of heuristics on large and hard satisfiability problems. *Phys. Rev. E*, 74:037702, 2006.

[19] Mikko Alava, John Ardelius, Erik Aurell, Petteri Kaski, Supriya Krishnamurthy, Pekka Orponen, and Sakari Seitz. Circumspect descent prevails in solving random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 105(40):15253–15257, 2008.

[20] Raffaele Marino, Giorgio Parisi, and Federico Ricci-Tersenghi. The backtracking survey propagation algorithm for solving random k-sat problems. *Nature Communications*, 7:12996, 2016.

[21] Amin Coja-Oghlan. A better algorithm for random k-sat. *SIAM Journal on Computing*, 39(7):2823–2864, 2010.

[22] D. Gamarnik and M. Sudan. Performance of sequential local algorithms for the random nae-k-sat problem. *SIAM Journal on Computing*, 46(2):590–619, 2017.

[23] A. Coja-Oghlan, A. Haqshenas, and S. Hetterich. Walksat stalls well below satisfiability. *SIAM Journal on Discrete Mathematics*, 31(2):1160–1173, 2017.

[24] Samuel Hetterich. Analysing survey propagation guided decimation on random formulas. *arXiv preprint arXiv:1602.08519*, 2016.

[25] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[26] Alfredo Braunstein, Luca Dall'Asta, Guilhem Semerjian, and Lenka Zdeborova. The large deviations of the whitening process in random constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(5):053401, 2016.

[27] Carlo Baldassi, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina. Local entropy as a measure for sampling solutions in constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(2):023301, 2016.

[28] Carlo Baldassi, Christian Borgs, Jennifer T. Chayes, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina. Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes. *Proceedings of the National Academy of Sciences*, 113(48):E7655–E7662, 2016.

[29] Thibaud Maimbourg, Mauro Sellitto, Guilhem Semerjian, and Francesco Zamponi. Generating dense packings of hard spheres by soft interaction design. *SciPost Phys.*, 4:39, 2018.

[30] Han Zhao and Hai-Jun Zhou. Maximally flexible solutions of a random k-satisfiability formula. *arXiv preprint arXiv:2006.07023*, 2020.

[31] Louise Budzynski, Federico Ricci-Tersenghi, and Guilhem Semerjian. Biased landscapes for random constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(2):023302, 2019.

[32] Louise Budzynski and Guilhem Semerjian. The asymptotics of the clustering transition for random constraint satisfaction problems. *arXiv preprint arXiv:1911.09377v2*, 2019.

[33] Louise Budzynski and Guilhem Semerjian. Biased measures for random constraint satisfaction problems: larger interaction range and asymptotic expansion. *arXiv preprint arXiv:2007.10303*, 2020.

[34] Guilhem Semerjian and Rémi Monasson. Relaxation and metastability in a local search procedure for the random satisfiability problem. *Phys. Rev. E*, 67(6):066103, Jun 2003.

[35] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: An algorithm for satisfiability. *Random Struct. Algorithms*, 27(2):201–226, 2005.

[36] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Solving constraint satisfaction problems through belief propagation-guided decimation. 2007. `arXiv:0709.1667`, Proceedings of the 45th Allerton Conference.

[37] Federico Ricci-Tersenghi and Guilhem Semerjian. On the cavity method for decimated random constraint satisfaction problems and the analysis of belief propagation guided decimation algorithms. *Journal of Statistical Mechanics: Theory and Experiment*, page P09001, 2009.

[38] Allan Sly. Reconstruction of random colourings. *Communications in Mathematical Physics*, 288(3):943–961, 2009.

[39] Andrea Montanari, Ricardo Restrepo, and Prasad Tetali. Reconstruction and clustering in random constraint satisfaction problems. *SIAM Journal on Discrete Mathematics*, 25(2):771–808, 2011.

[40] Allan Sly and Yumeng Zhang. Reconstruction of colourings without freezing. *arXiv preprint arXiv:1610.02770*, 2016.

[41] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proc. 3rd STOC*, pages 151–158, New York, NY, USA, 1971. ACM.

[42] M.R. Garey, D.S. Johnson, and L. Stockmeyer. Some simplified np-complete graph problems. *Theoretical Computer Sciences*, 1:237–267, 1976.

[43] Martin Davis, George Logemann, and Donald Loveland. A machine program for theorem-proving. *Commun. ACM*, 5(7):394–397, 1962.

[44] M. Mézard and A. Montanari. *Physics, Information, Computation.* Oxford Press, Oxford, 2009.

[45] S. Kirkpatrick and B. Selman. Critical behavior in the satisfiability of random boolean expression. *Science*, 264:1297–1301, 1994.

[46] David G. Mitchell, Bart Selman, and Hector J. Levesque. Hard and easy distributions for SAT problems. In *Proc. 10th AAAI*, pages 459–465, Menlo Park, California, 1992. AAAI Press.

[47] Paul Beame, Richard Karp, Toniann Pitassi, and Michael Saks. On the complexity of unsatisfiability proofs for random k-cnf formulas. *Conference Proceedings of the Annual ACM Symposium on Theory of Computing*, 04 1998.

[48] Alan Frieze and Stephen Suen. Analysis of two simple heuristics on a random instance of k-sat. *J. Algorithms*, 20(2):312–355, 1996.

[49] Simona Cocco and Rémi Monasson. Trajectories in phase diagrams, growth processes, and computational complexity: How search algorithms solve the 3-satisfiability problem. *Phys. Rev. Lett.*, 86(8):1654–1657, Feb 2001.

[50] D. Aldous and M. Steele. The objective method: Probabilistic combinatorial optimization and local weak convergence. In H. Kesten, editor, *Encyclopedia of Mathematical Sciences*, volume 110, pages 1–72. Springer, 2003.

[51] Vasek Chvatal and Bruce Reed. Mick gets some (the odds are on his side) (satisfiability). In *Proceedings., 33rd Annual Synopsium on Foundations of Computer Science*, pages 620–627, Pittsburgh, PA, USA, 1992.

[52] W. Fernandez de la Vega. On random 2-sat. Unpublished manuscript, 1992.

[53] Andreas Goerdt. A threshold for unsatisfiability. *Journal of Computer and System Sciences*, 53:469–486, 1996.

[54] E. Friedgut. Sharp thresholds of graph proprties, and the k-sat problem. *Journal of the AMS*, 12:1017–1054, 1999.

[55] J. Franco and M. Paull. Probabilistic analysis of the davis-putnam procedure for solving satisfiability. *Discrete Applied Mathematics*, 5:77–87, 1983.

[56] Lefteris M. Kirousis, Evangelos Kranakis, Danny Krizanc, and Yannis C. Stamatiou. Approximating the unsatisfiability threshold of random formulas. *Random Structures and Algorithms*, 12(3):253–269, 1998.

[57] Dimitris Achlioptas and Cristopher Moore. Two moments suffice to cross a sharp threshold. *SIAM Journal on Computing*, 36:740–762, 2006.

[58] J. P. Bouchaud and G. Biroli. On the Adam-Gibbs-Kirkpatrick-Thirumalai-Wolynes scenario for the viscosity increase of classes. *J. Chem. Phys.*, 121:7347–7354, 2004.

[59] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Clusters of solutions and replica symmetry breaking in random k-satisfiability. *J. Stat. Mech.*, P04004, 2008.

[60] R. Monasson. Structural glass transition and the entropy of the metastable states. *Phys. Rev. Lett.*, 75:2847, 1995.

[61] L. Zdeborová and F. Krzakala. Phase transitions in the coloring of random graphs. *Phys. Rev. E*, 76:031131, 2007.

[62] Guilhem Semerjian. On the freezing of variables in random constraint satisfaction problems. *J. Stat. Phys.*, 130:251, 2008.

[63] A. Montanari and G. Semerjian. On the dynamics of the glass transition on bethe lattices. *J. Stat. Phys.*, 124:103–189, 2006.

[64] Christos H. Papadimitriou. On selecting a satisfying truth assignment (extended abstract). In *Proceedings of the 32nd annual symposium on Foundations of computer science*, pages 163–169, Los Alamitos, CA, USA, 1991. IEEE Computer Society Press.

[65] Uwe Schöning. A probabilistic algorithm for k-sat based on limited local search and restart. *Algorithmica*, 32:615–623, 2002.

[66] Bart Selman, Henry A. Kautz, and Bram Cohen. Local search strategies for satisfiability testing. In Michael Trick and David Stifler Johnson, editors, *Proceedings of the Second DIMACS Challange on Cliques, Coloring, and Satisfiability*, Providence RI, 1996.

[67] Sakari Seitz, Mikko Alava, and Pekka Orponen. Focused local search for random 3-satisfiability. *J. Stat. Mech.*, page P06006, 2005.

[68] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[69] Robert G. Gallager. Low-density parity check codes. *IEEE Trans. Inform. Theory*, 8:21–28, 1962.

[70] F. R. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory*, 47(2):498–519, 2001.

[71] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In *Exploring Artificial Intelligence in the New Millennium*, pages 239–236. Science & Technology Books, 2003.

[72] H. A. Bethe. Statistical physics of superlattices. *Proc. Roy. Soc. London A*, 150:552–575, 1935.

[73] J. S. Yedidia, W. F. Freeman, and Y. Weiss. Constructing free energy approximations and generalized belief propagation algorithms. *technical report TR-2002-35, Mitsubishi Electrical Research Laboratories*, 2002. available at http://www.merl.com.

[74] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Solving constraint satisfaction problems through belief propagation-guided decimation. To appear in Proceedings of Allerton 2007. Preprint: arXiv:0709.1667v1 [cs.AI], 2007.

[75] M. Mézard and R. Zecchina. Random k-satisfiability problem: From an analytic solution to an efficient algorithm. *Phys. Rev. E*, 66:056126, 2002.

[76] Bart Selman, Hector J. Levesque, and D. Mitchell. A new method for solving hard satisfiability problems. In Paul Rosenbloom and Peter Szolovits, editors, *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 440–446, Menlo Park, California, 1992. AAAI Press.

[77] Wolfgang Barthel, Alexander K. Hartmann, and Martin Weigt. Solving satisfiability problems by fluctuations: The dynamics of stochastic local search algorithms. *Phys. Rev. E*, 67:066104, Jun 2003.

[78] D. Achlioptas and C. Moore. Almost all graphs with average degree 4 are 3-colorable. *J. Comput. Syst. Sci.*, 67:441, 2003.

[79] P. Orponen S. S. Seitz. An efficient local search method for random 3-satisfiability. *Electronic Notes in Discrete Mathematics*, 16:71–79, 2003.

[80] G. Parisi. Some remarks on the survey decimation algorithm for k-satisfiability. arXiv:cs/0301015, 2003.

[81] A. Kaporis, L. Kirousis, and E. Lalas. Selecting complementary pairs of literals. In *Proc. LICS'03 Workshop on Typical Case Complexity and Phase Transitions*, 2003.

[82] M. Hajiaghayi and G. B. Sorkin. The Satisfiability Threshold of Random 3-SAT Is at Least 3.52. arXiv: math/0310193, 2003.

[83] Elitza N. Maneva, Elchanan Mossel, and Martin J. Wainwright. A new look at survey propagation and its generalizations. *J. ACM*, 54(4), 2007.

[84] A. Braunstein and R. Zecchina. Survey propagation as local equilibrium equations. *J. Stat. Mech.*, P06007, 2004.

[85] L. Dall'Asta, A. Ramezanpour, and R. Zecchina. Entropy landscape and non-gibbs solutions in constraint satisfaction problems. *Phys. Rev. E*, 77:031118, 2008.

[86] A. Coja-Oghlan. On belief propagation guided decimation for random k-sat. In *Proceedings of the 2011 Annual ACM-SIAM Symposium on Discrete Algorithms*, page 10. SIAM, 2011.

[87] D. Achlioptas, A. Naor, and Y. Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435:759–764, 2005.

[88] Moore C. Achlioptas D. On the 2-colorability of random hypergraphs. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 78–90. Springer, Berlin, Heidelberg, 2002.

[89] Dimitris Achlioptas and Cristopher Moore. The Chromatic Number of Random Regular Graphs. *arXiv e-prints*, pages cond–mat/0407278, July 2004.

[90] Ming-Te Chao and John Franco. Probabilistic analysis of a generalization of the unit-clause literal selection heuristics for the k satisfiability problem. *Inf. Sci.*, 51(3):289–314, 1990.

[91] D. Achlioptas. Lower bounds for random 3-SAT via differential equations. *Theor. Comput. Sci.*, 265(1-2):159–185, 2001.

[92] A. Coja-Oghlan and A. Frieze. Analyzing walksat on random formulas. *SIAM Journal on Computing*, 43(4):1456–1485, 2014.

[93] Jian Ding, Allan Sly, and Nike Sun. Satisfiability threshold for random regular nae-sat. *Communications in Mathematical Physics*, 341(2):435–489, Jan 2016.

[94] A. Braunstein, M. Mézard, M. Weigt, and R. Zecchina. Constraint satisfaction by survey propagation. In Allon Percus, Gabriel Istrate, and Cristopher Moore, editors, *Computational Complexity and Statistical Physics*, page 107. Oxford University Press, 2003.

[95] G. Parisi. On local equilibrium equations for clustering states. arXiv:cs.CC/0212047, 2002.

[96] Mauro Sellitto and Francesco Zamponi. A thermodynamic description of colloidal glasses. *EPL (Europhysics Letters)*, 103(4):46005, aug 2013.

[97] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin-Glass Theory and Beyond*, volume 9 of *Lecture Notes in Physics*. World Scientific, Singapore, 1987.

[98] Lenka Zdeborová and Florent Krzakala. Generalization of the cavity method for adiabatic evolution of gibbs states. *Phys. Rev. B*, 81:224205, 2010.

[99] T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2007.

[100] H. Nishimori. *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Oxford University Press, Oxford, UK, 2001.

[101] H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional galton-watson processes. *The Annals of Mathematical Statistics*, 37:1463, 1966.

[102] J. R. L. de Almeida and D. J. Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin-glass model. *J. Phys. A*, 11:983–990, 1978.

[103] Marylou Gabrié, Varsha Dani, Guilhem Semerjian, and Lenka Zdeborová. Phase transitions in the q-coloring of random hypergraphs. *Journal of Physics A: Mathematical and Theoretical*, 50(50):505002, 2017.

[104] Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Typology of phase transitions in bayesian inference problems. *arXiv preprint arXiv:1806.11013*, 2018.

[105] R Abou-Chacra, D J Thouless, and P W Anderson. A selfconsistent theory of localization. *Journal of Physics C: Solid State Physics*, 6(10):1734, 1973.

[106] G Parisi, F Ricci-Tersenghi, and T Rizzo. Diluted mean-field spin-glass models at criticality. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(4):P04013, 2014.

[107] G Parisi, F Ricci-Tersenghi, and T Rizzo. Erratum: Diluted mean-field spin-glass models at criticality. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(3):E03001, 2015.

[108] Svante Janson and Elchanan Mossel. Robust reconstruction on trees is determined by the second eigenvalue. *Ann. Probab.*, 32:2630–2649, 2004.

[109] Ken-Iti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 2013.

[110] Soren Asmussen and Jan Rosinski. Approximations of small jumps of lévy processes with a view towards simulation. *Journal of Applied Probability*, 38(2):482–493, 2001.

[111] Zhiyi Chi. Nonnormal small jump approximation of infinitely divisible distributions. *Advances in Applied Probability*, 46(4):963–984, 2014.

[112] William Feller. *An introduction to probability theory and its applications, Vol II, 2nd ed.* Wiley New York, 1971.

[113] Toshiro Watanabe and Kouji Yamamuro. Ratio of the tail of an infinitely divisible distribution on the line to that of its lévy measure. *Electron. J. Probab.*, 15:44–74, 2010.

Appendix A

Existence and uniqueness of the RS solution

In this appendix we shall show that the translationally invariant RS equation (7.40) admits a unique solution for all choices of the bias function ψ that is strictly positive, $\psi(p) > 0 \ \forall p \in \{0, \dots l+1\}$.

We first remark that in the uniform case, where $\psi(p)$ is a positive constant independent of p, the equation (7.40) obviously admits a unique solution (with $y = 1$, $\widehat{y} = 2^{k-1} - 2$). We will now show that the number of solutions cannot change when ψ varies in its allowed domain. To achieve this we first rewrite (7.40) in the equivalent form

$$G(x;\psi) = \sum_{p=0}^{l+1} \psi(p) X_p(x) = 0 \ , \tag{A.1}$$

where for simplicity we denoted $x = 1/\widehat{y}$ and where the coefficients $X_p(x)$ are:

$$X_p(x) = x^p \left[(2^{k-1} - k - 1) \binom{l}{p} x + (k-1) \binom{l}{p-1} - \binom{l}{p} \right] \tag{A.2}$$

$$= \frac{x^p}{l+1} \binom{l+1}{p} \left[(2^{k-1} - k - 1)(l+1-p)x - (l+1-kp) \right] \ ; \tag{A.3}$$

in the first line we used the convention $\binom{l}{l+1} = \binom{l}{-1} = 0$.

The function $G(x;\psi)$ introduced in (A.1) depends smoothly on its two arguments (polynomially in x, and linearly in ψ), the number of solutions $x(\psi)$ of the equation $G = 0$ can thus only change at a bifurcation point, i.e. a pair $(x;\psi)$ such that $G(x;\psi) = \partial_x G(x;\psi) = 0$, otherwise the implicit function theorem allows to smoothly continue any branch of solution. As we remarked above the solution is unique when ψ is independent of p, the uniqueness for all ψ will

then follow if we show the absence of solution to the bifurcation equation:

$$\sum_{p=0}^{l+1} \psi(p) X_p(x) = 0 \ , \quad \sum_{p=0}^{l+1} \psi(p) Y_p(x) = 0 \ , \quad \psi(p) \geq 0 \quad \forall p \ , \tag{A.4}$$

where

$$Y_p(x) = \frac{\partial X_p}{\partial x} = \frac{x^{p-1}}{l+1} \binom{l+1}{p} \left[(2^{k-1} - k - 1)(l+1-p)(p+1)x - (l+1-kp)p \right] \ . \tag{A.5}$$

This equation being invariant under the multiplication of ψ by a positive constant we can further assume the normalization condition

$$\sum_{p=0}^{l+1} \psi(p) = 1 \ . \tag{A.6}$$

For a given value of x, the existence of a ψ satisfying (A.4,A.6) is equivalent to the origin of $\mathbb{R}^2$ being in the convex hull of the $l+2$ points of coordinates $\begin{pmatrix} X_p(x) \\ Y_p(x) \end{pmatrix}$ for $p = 0, \ldots, l+1$. We can then invoke the Caratheodory theorem that states that any point of the convex hull of a set $A \subset \mathbb{R}^d$ can be written as the convex combination of $d+1$ points of A. Here $d = 2$, so the absence of solutions of (A.4,A.6) follows from the impossibility to satisfy, for any $p, q, r \in \{0, \ldots, l+1\}$, the system

$$\alpha_p X_p(x) + \alpha_q X_q(x) + \alpha_r X_r(x) = 0 \ , \tag{A.7}$$

$$\alpha_p Y_p(x) + \alpha_q Y_q(x) + \alpha_r Y_r(x) = 0 \ , \tag{A.8}$$

$$\alpha_p \geq 0, \quad \alpha_q \geq 0, \quad \alpha_r \geq 0, \quad \alpha_p + \alpha_q + \alpha_r > 0 \ . \tag{A.9}$$

This is equivalent to the three quantities $X_p(x)Y_q(x) - X_q(x)Y_p(x)$, $X_q(x)Y_r(x) - X_r(x)Y_q(x)$ and $X_r(x)Y_p(x) - X_p(x)Y_r(x)$ being of the same sign; using the expressions (A.3,A.5) of X and Y we have checked the impossibility of this condition, for all $x > 0$ and all triplets p, q, r, which concludes the reasoning.

Appendix B

The graph coloring case

This Appendix is devoted to the graph coloring problem. We shall present computations that are the counterparts of the ones explained in Sec. 7.3 and 9.1 of the main text for the hypergraph bicoloring problem, namely the definition of the reconstruction problem and its study in the limit where the number of colors q and the degree c diverge simultaneously according to (9.1), and show that the very same equations summarized at the beginning of Sec. 9.2 arise in this limit. As the computations are quite similar we will be more succint than in the main text and concentrate on the specificities of the coloring problem.

Let us consider a rooted tree with spins σ_i placed on its vertices. These spins can take q values, interpreted as colors, $\sigma_i \in \{1, \dots, q\}$. A proper coloring of the tree is a configuration of the spins such that no edge is monochromatic (i.e. no pair of adjacent vertices are given the same color). A uniform proper coloring can be drawn in a broadcast fashion, by choosing the color σ of the root uniformly at random among the q possible ones, then each descendent of the root is assigned a color uniformly at random among the $q-1$ colors distinct from σ, and this is repeated recursively down to the n-th generation of the tree. In the reconstruction problem an observer is then provided with the colors on the vertices of the n-th generation of the tree only, and asked to guess the color of the root. The optimal strategy is to compute η, the posterior probability of the root given the observations, which is a distribution over $\{1, \dots, q\}$. A moment of thought reveals that the probability distribution of η, with respect to a broadcast process conditioned on the root value σ, and with respect to a random choice of the tree as a Galton-Watson branching process with Poisson offspring distribution of mean c, is a measure $P_{\sigma,n}(\eta)$ that can be determined recursively through the induction relation:

$$P_{\sigma,n+1}(\eta) = \sum_{l=0}^{\infty} e^{-c} \frac{c^l}{l!} \frac{1}{(q-1)^l} \sum_{\sigma_1,\dots,\sigma_l \neq \sigma} \int \prod_{i=1}^{l} \mathrm{d}P_{\sigma_i,n}(\eta_i)\, \delta(\eta - f_{\mathrm{c}}(\eta_1, \dots, \eta_l)) \ , \tag{B.1}$$

where the Belief Propagation recursion function f_{c} is defined here in such a way

that $\eta = f_{\rm c}(\eta_1, \ldots, \eta_l)$ means

$$\eta(\tau) = \frac{\prod_{i=1}^{l} (1 - \eta_i(\tau))}{\sum_{\tau'} \prod_{i=1}^{l} (1 - \eta_i(\tau'))} \ . \tag{B.2}$$

The initial condition of this recursive computation is given by $P_{\sigma,0} = \delta(\eta - \delta_\sigma)$, where δ_σ is the measure concentrated on the color σ, i.e. $\delta_\sigma(\tau) = \mathbb{I}(\sigma = \tau)$, which corresponds to the colors being revealed on the leaves of the tree. These equations correspond to (7.7,7.51,7.52) (with $\omega(\sigma_1, \ldots, \sigma_k) = 1 - \mathbb{I}[\sigma_1, \sigma_k \ a.e]$) for the hypergraph bicoloring problem.

We separate now the contribution of "hard fields", or frozen variables, namely the configurations of boundary variables that determine unambiguously the root in the naive reconstruction procedure. This forced value can only be the correct one the root had in the broadcast process, hence we shall write :

$$P_{\sigma,n}(\eta) = H_n \, \delta(\eta - \delta_\sigma) + (1 - H_n) \, Q_{\sigma,n}(\eta) \ , \tag{B.3}$$

where $Q_{\sigma,n}$ has no atom on δ_σ; this mimicks the decomposition (7.78) of the main text. Plugging this decomposition in (B.1) yields the evolution equation for the weight of the hard fields:

$$H_{n+1} = \left(1 - e^{-\frac{c H_n}{q-1}}\right)^{q-1} \ , \qquad \text{with} \ \ H_0 = 1 \ . \tag{B.4}$$

Indeed $f_{\rm c}(\eta_1, \ldots, \eta_l) = \delta_\sigma$ if and only if for each color $\sigma' \neq \sigma$ at least one of the arguments η_i is equal to $\delta_{\sigma'}$, thus forbidding all colors except σ. The number of hard fields of the color $\sigma' \neq \sigma$ is easily seen from (B.1) to be Poisson distributed with average $\frac{c H_n}{q-1}$, independently from one color σ' to another, from which (B.4) follows.

The recursion (B.4) has a bifurcation at $c_{\rm r}(q)$, in the sense that $H_n \to 0$ as $n \to \infty$ if and only if $c < c_{\rm r}(q)$. Writing down the equations fixing $c_{\rm r}$ and $H_{\rm r}$ at the bifurcation, which are similar to (7.81), and then expanding them for large q one finds the asymptotic expansion $c_{\rm r} = q(\ln q + \ln\ln q + 1 + o(1))$. We shall thus study the large q limit with c getting also large, on the scale $c = q(\ln q + \ln\ln q + \gamma)$ with γ finite. In this limit one finds that for n finite

$$H_n = 1 - \frac{x_n}{\ln q} + o\left(\frac{1}{\ln q}\right) \ , \tag{B.5}$$

where x_n is of order 1 and obeys exactly the same recursion as in the bicoloring case, namely $x_0 = 0$ and $x_{n+1} = e^{-\gamma + x_n}$.

Let us now simplify the evolution equation for the distributions $Q_{\sigma,n}$ of the soft fields. First of all we insert the decomposition (B.3) in the right hand side

of (B.1) and obtain, without any approximation,

$$P_{\sigma,n+1}(\eta) = \sum_{l=0}^{\infty} e^{-c(1-H_n)} \frac{(c(1-H_n))^l}{l!} \frac{1}{(q-1)^l} \sum_{\sigma_1,\dots,\sigma_l \neq \sigma} \sum_{\{p_{\sigma'}=0,1\}_{\sigma'\neq\sigma}} \prod_{\sigma'\neq\sigma} \left(e^{-\frac{cH_n}{q-1}}\right)^{p_{\sigma'}} \left(1-e^{-\frac{cH_n}{q-1}}\right)^{1-p_{\sigma'}} \int \prod_{i=1}^{l} \mathrm{d}Q_{\sigma_i,n}(\eta_i)\, \delta(\eta - \widetilde{f}_{\mathrm{c}}(\sigma, \{p_{\sigma'}\}_{\sigma'\neq\sigma}; \eta_1, \dots, \eta_l)) \ , \tag{B.6}$$

where l is the number of neighbors of the root that receive a soft field, $\sigma_1, \dots, \sigma_l$ the colors these vertices have in the broadcast, and the indicator variables $p_{\sigma'}$ are equal to 1 if and only if no neighbor of the root assigned the color σ' in the broadcast is perfectly recovered (i.e. receives a hard field). Hence the colors $\sigma' \neq \sigma$ with $p_{\sigma'} = 0$ are precisely the ones forbidden for the root, as at least one of its neighbors is forced to this value. The relation $\eta = \widetilde{f}_{\mathrm{c}}(\sigma, \{p_{\sigma'}\}_{\sigma'\neq\sigma}; \eta_1, \dots, \eta_l)$ is obtained by specializing f_{c} of (B.2) to this pattern for the presence of hard fields in its arguments, and thus reads

$$\eta(\tau) = \frac{p_\tau \prod_{i=1}^{l} (1-\eta_i(\tau))}{\sum_{\tau'} p_{\tau'} \prod_{i=1}^{l} (1-\eta_i(\tau'))} \ , \tag{B.7}$$

with the convention $p_\sigma = 1$.

Let us call $p = \sum_{\sigma'\neq\sigma} p_{\sigma'}$ the number of colors that satisfy the condition explained above; in the equation (B.6) it corresponds to a random variable with a binomial distribution of parameters $(q-1, e^{-\frac{cH_n}{q-1}})$. According to (B.5) the product of these parameters go to zero as $1/\ln q$ in the limit we are considering, we shall thus truncate (B.6) on the smallest possible values of p. As $p = 0$ yields a hard field in the left hand side of (B.6), the distribution of the soft fields is dominated in this limit by the case $p = 1$. As a consequence the fields η in the support of $Q_{\sigma,n}$ have non-zero values on two colors only, σ and another one σ' uniformly distributed on the $q-1$ possibilities. Let us parametrize this type of distributions via a distribution $Q_n(h)$ on real random variables $h \in [-1, 1]$,

$$Q_{\sigma,n}(\eta) = \int \mathrm{d}Q_n(h) \frac{1}{q-1} \sum_{\sigma'\neq\sigma} \delta(\eta - s(\sigma,\sigma';h)), \tag{B.8}$$

$$\text{with} \quad s(\sigma,\sigma';h)(\tau) = \begin{cases} \frac{1+h}{2} & \text{if } \tau = \sigma \\ \frac{1-h}{2} & \text{if } \tau = \sigma' \\ 0 & \text{otherwise} \end{cases} .$$

Let us also denote $\widehat{f}_{\mathrm{c}}(\sigma, \sigma'; \eta_1, \dots, \eta_l)$ the function $\widetilde{f}_{\mathrm{c}}(\sigma, \{p_{\sigma''}\}_{\sigma''\neq\sigma}; \eta_1, \dots, \eta_l)$ with $p_{\sigma''} = \delta_{\sigma'',\sigma'}$, in such a way that the only two non-zero components of

$\widehat{f}_c(\sigma,\sigma';\eta_1,\dots,\eta_l)$ correspond to the colors σ and σ'. Using this notation, and the parametrization (B.8), one can deduce from (B.6) the evolution equation for $Q_{\sigma,n}(\eta)$ at lowest order:

$$Q_{\sigma,n+1}(\eta) = \frac{1}{q-1}\sum_{\sigma'\neq\sigma}\sum_{l=0}^{\infty} e^{-c(1-H_n)}\frac{(c(1-H_n))^l}{l!}\prod_{i=1}^{l}\left(\frac{1}{(q-1)^2}\sum_{\substack{\sigma_i\neq\sigma\\ \sigma_i'\neq\sigma_i}}\int \mathrm{d}Q_n(h_i)\right) \tag{B.9}$$
$$\delta(\eta - \widehat{f}_c(\sigma,\sigma';s(\sigma_1,\sigma_1';h_1),\dots,s(\sigma_l,\sigma_l';h_l)))$$

To put this expression under the form (B.8), and hence close the recursion on $Q_n(h)$, it remains to notice that

$$\widehat{f}_c(\sigma,\sigma';\eta_1,\dots,\eta_l) = s(\sigma,\sigma',h) \quad \text{with} \quad h = f(u_1,\dots,u_l)\ , \quad u_i = \widehat{u}(\sigma,\sigma';\eta_i)\ , \tag{B.10}$$

where $f = f_{\Theta_0}$ is the function defined for Ising spins in (7.8), and

$$\widehat{u}(\sigma,\sigma';\eta) = \frac{\eta(\sigma') - \eta(\sigma)}{2-\eta(\sigma)-\eta(\sigma')}\ . \tag{B.11}$$

For a fixed choice of σ and $\sigma' \neq \sigma$, one can see that $\widehat{u}(\sigma,\sigma';s(\sigma_i,\sigma_i';h_i))$ is a random variable with respect to the uniform choices of $\sigma_i \neq \sigma$ and $\sigma_i' \neq \sigma_i$, that takes the following values (recall the definition of the function g_1 from (9.10)):

$$\begin{cases} g_1(h_i) & \text{with probability } \frac{q-2}{(q-1)^2}\ \text{, when } \sigma_i = \sigma'\ ,\ \sigma_i' \neq \sigma\ ,\\ g_1(-h_i) & \text{with probability } \frac{q-2}{(q-1)^2}\ \text{, when } \sigma_i' = \sigma'\ ,\ \sigma_i \neq \sigma\ ,\\ -g_1(-h_i) & \text{with probability } \frac{q-2}{(q-1)^2}\ \text{, when } \sigma_i' = \sigma\ ,\ \sigma_i \neq \sigma'\ ,\\ h_i & \text{with probability } \frac{1}{(q-1)^2}\ \text{, when } \sigma_i' = \sigma\ ,\ \sigma_i = \sigma'\ ,\\ 0 & \text{otherwise .} \end{cases} \tag{B.12}$$

As $c(1-H_n)\frac{q-2}{(q-1)^2} \to x_n$ in the regime we are considering, the number of occurences of the first three cases in (B.9) is Poissonian of mean x_n; on the other hand the number of times the fourth case happens vanishes when q diverges (as $1/q$), while the fifth does not contribute to (B.9), because $f(u_1,\dots,u_l,0,\dots,0) = f(u_1,\dots,u_l)$. We thus see by comparison with (9.13,9.15) that the probability distribution $Q_n(h)$ defined in (B.8) obeys exactly the same recursion equations as the one derived in the main text for the hypergraph bicoloring model, the initial condition $Q_1(h) = \delta(h)$ being also valid here.

Appendix C

An Inequality

We provide in this Appendix a proof of the bounds $c_-(b)x_{1,n} \leq \lambda_n \leq c_+(b)x_{1,n}$ that we used in Sec. 10.4.3. We start by stating some inequalities that are fulfilled by the messages η in the support of ν_n, and that are consequences of the BP equation (10.17). They are more compactly stated in terms of the u-parametrization; from (10.55) one obtains indeed

$$\frac{u^{(2)}u^{(3)}}{u^{(1)}} = \prod_i \frac{\widehat{\eta}_i(+,0)}{\widehat{\eta}_i(+,0)+\widehat{\eta}_i(+,1)}, \quad \frac{u^{(1)}u^{(3)}}{u^{(2)}} = \prod_i \frac{\widehat{\eta}_i(-,0)}{\widehat{\eta}_i(-,0)+\widehat{\eta}_i(-,1)}, \quad \text{(C.1)}$$

which allows to conclude that $\frac{u^{(2)}u^{(3)}}{u^{(1)}} \leq 1$ and $\frac{u^{(1)}u^{(3)}}{u^{(2)}} \leq 1$, for all the η's in the support of ν_n.

Consider now the expressions (10.52) for $x_{1,n}$, and (10.84) for λ_n; the ratio of the integrands in these two equations reads

$$\begin{aligned} &2\sqrt{\eta(+,0)+\eta(-,0)}\,\frac{\sqrt{\eta(+,1)}+\sqrt{\eta(-,1)}}{\eta(+,1)+\eta(-,1)} \\ &= \sqrt{\frac{\eta(+,0)+\eta(-,0)}{\eta(+,1)+\eta(-,1)}}\left(2\sqrt{\frac{\eta(+,1)}{\eta(+,1)+\eta(-,1)}}+2\sqrt{\frac{\eta(-,1)}{\eta(+,1)+\eta(-,1)}}\right) \end{aligned}$$

The parenthesis in the right hand side of this eqation is of the form $2(\sqrt{x}+\sqrt{1-x})$ for some $x \in [0,1]$, which is necessarily in the interval $[2, 2\sqrt{2}]$. The prefactor in front of the parenthesis can be written, in terms of th u-parametrization,

$$\sqrt{\frac{1}{1+(u^{(1)})^2}\left(1+Bu^{(1)}u^{(2)}u^{(3)}+(u^{(1)})^2+B\frac{u^{(1)}u^{(3)}}{u^{(2)}}\right)} \quad \text{(C.2)}$$

Consider first the case $b \leq 1$, i.e. $B \geq 0$; as the components of u are non-negative the expression in (C.2) is certainly lower bounded by 1. Moreover the bounds $\frac{u^{(2)}u^{(3)}}{u^{(1)}} \leq 1$ and $\frac{u^{(1)}u^{(3)}}{u^{(2)}} \leq 1$ imply that it is upper bounded by $\sqrt{1+B} = 1/\sqrt{b}$. The case $b \geq 1$, $B \leq 0$ can be treated similarly, with now the expression in (C.2)

being in the interval $[1/\sqrt{b}, 1]$. Combining these observations we obtain finally $c_-(b)x_{1,n} \leq \lambda_n \leq c_+(b)x_{1,n}$, with for $b \leq 1$ $c_-(b) = 2$, $c_+(b) = 2\sqrt{2/b}$, and for $b \geq 1$ $c_-(b) = 2/\sqrt{b}$, $c_+(b) = 2\sqrt{2}$.

RÉSUMÉ

La complexité typique des Problèmes de Satisfaction de Contraintes (CSP) peut être étudiée à l'aide d'ensembles aléatoires de contraintes. On observe un phénomène de seuil quand la densité de contraintes augmente. En particulier à la transition de clustering, l'ensemble des solutions typiques se fracture en groupes de solutions séparés les uns des autres. Dans cette thèse nous introduisons un biais qui brise l'uniformité entre les solutions d'une instance de CSP, et nous étudions son effet sur la valeur du seuil de clustering. Nous étudions en particulier le problème de bicoloriage de k-hypergraphes. Pour de petites valeurs de k, nous montrons que ce biais peut augmenter la valeur du seuil de clustering, et que cela a un effet positif sur les performances de l'algorithme de Simulated Annealing pour la recherche de solutions d'une instance du problème de bicoloriage. Dans la limite où k tend vers l'infini, nous calculons le développement asymptotique du seuil de clustering pour la mesure uniforme et pour une mesure biasée. Nous évaluons le gain obtenu avec cette implémentation du biais.

MOTS CLÉS

Barrières algorithmiques, Problèmes de satisfaction de contraintes, Aléatoire.

ABSTRACT

The typical complexity of Constraint Satisfaction Problems (CSP) can be studied using random ensembles of instances. One observes threshold phenomena when the density of constraints increases, in particular a clustering phase transition at which typical solutions shatter into disconnected components. In this Ph.D., we introduce a bias that breaks the uniformity among solutions of a given instance of CSP, and look at the evolution of the clustering threshold under this bias, focusing on the bicoloring of k-uniform random hypergraphs. For small values of k, we show that this bias can delay the clustering transition to higher densities of constraints, and that it has a positive impact on the performances of Simulated Annealing algorithm to find a solution for a given instance of the bicoloring problem. In the large k limit, we compute the asymptotic expansion of the clustering threshold for the uniform and the biased measure, and characterize the gain obtained with our implementation of the bias.

KEYWORDS

Algorithmic barriers, Constraint satisfaction problems, Random.