Overview

The central goal of this book is to present new techniques for studying the behavior of mean-field systems with disordered interactions. We will focus in particular on certain problems of statistical inference, and on spin glasses. We will also consider simpler models of statistical mechanics for which the interactions are not disordered, as a useful training ground. We will mostly be interested in computing the asymptotic behavior of a fundamental quantity called the free energy, in the limit of large system size. In the context of statistical inference, the free energy is essentially the mutual information between the observations and the signal that we wish to recover; knowing its limit behavior allows us to identify how much of the signal can be recovered from the observations. The purpose of this work is to present a way to approach this class of problems using techniques from partial differential equations, and specifically, the theory of Hamilton–Jacobi equations. We strove to make this book self-contained, and in particular, no prior knowledge of Hamilton–Jacobi equations or other partial differential equations is assumed.

To get a sense of what spin glasses are, let us start by describing the simplest such model, namely, the Sherrington–Kirkpatrick (SK) model [237]. This model is often motivated by considering the problem of splitting N individuals into two groups. An assignment of individuals into two groups can be encoded by a vector $\sigma \in \Sigma_N := \{-1, +1\}^N$, with the understanding that σ_i represents the group to which the individual indexed by i is assigned. For each pair of individuals (i, j), we are given a number g_{ij} that encodes the quality of the interaction between individuals i and j, with g_{ij} being large if i and j get along very well, and g_{ij} being very negative if they cannot stand each other. We would like to find an assignment into groups which maximizes the sum total of the interactions; that is, we want to maximize, over the set Σ_N, the "comfort function"

$$c_N(\sigma) := \sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j. \tag{0.1}$$

Given the problem we are trying to encode, it would probably have felt more natural to write $\mathbf{1}_{\{\sigma_i=\sigma_j\}}$ in place of $\sigma_i\sigma_j$ above. However, the form $\sigma_i\sigma_j$ is more standard, partly for historical reasons having to do with the modeling of magnetic materials. Since $\sigma_i\sigma_j = 2\mathbf{1}_{\{\sigma_i=\sigma_j\}} - 1$, writing one or the other is inconsequential. It will be convenient not to impose any symmetry on the matrix $(g_{ij})_{1\leqslant i,j\leqslant N}$, so the interaction between i and j, with $i \neq j$, is actually encoded by $g_{ij} + g_{ji}$. It will also be convenient to include some diagonal terms $(g_{ii})_{1\leqslant i\leqslant N}$; these terms only change the comfort function by an additive constant, and their contribution will in any case be of lower order.

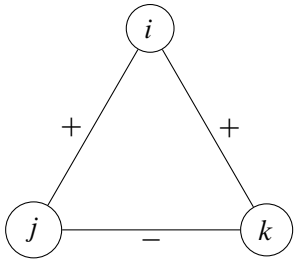

Figure 1. A simple example of frustration.

In order to gain some insight into a *typical* instance of this optimization problem, we assume a particular structure on the interaction parameters $(g_{ij})_{1\leqslant i,j\leqslant N}$. Namely, we assume that the $(g_{ij})_{1\leqslant i,j\leqslant N}$ are independent standard Gaussian random variables. By "standard Gaussian" we always mean Gaussian random variables with zero mean and unit variance.

One quickly realizes that the optimization of the comfort function (0.1) is not going to be easy because the coefficients $(g_{ij})_{1\leqslant i,j\leqslant N}$ do not have a sign. A glimpse of the difficulty can already be perceived when we consider situations in which three individuals i, j, and k are as depicted in Figure 1, that is, with $g_{ij} > 0$ and $g_{ik} > 0$, but with $g_{jk} < 0$. In view of the sign of g_{ij}, we would rather want i and j to belong to the same group, and similarly with i and k. But the sign of g_{ik} suggests to rather have i and k in different groups, and all of these "local preferences" cannot be reconciled at once. Physicists use the word *frustration* to describe situations of this nature, and a *glass* is a system that is subject to such frustrations. More generally, the presence of these frustrations suggests that it is likely to be difficult to optimize the function (0.1). For instance, a naive method that would try to optimize each *spin* σ_i one at a time in order to decrease the value of (0.1) is unlikely to reach the global optimum. The word "glass" points to the fact that this phenomenology is also present for regular window glass, as well as a number of other materials. Indeed, the making of a glass requires the very rapid cooling of an initially liquid material. A slow cooling would have allowed the material to find its preferred crystal structure, but the fast cooling has blocked the particles in a highly disordered configuration, and the complex geometric arrangement of the particles is believed to create jams that render the finding of the preferred crystalline state very difficult to achieve.

In order to gain insight into the problem of optimizing the comfort function (0.1), it is useful to first focus on the evaluation of the asymptotic behavior of the maximum itself, and more generally of quantities of the form

$$\frac{1}{N}\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\left(\frac{\beta}{\sqrt{N}}\sum_{i,j=1}^{N}g_{ij}\,\sigma_i\sigma_j\right), \tag{0.2}$$

where $\beta \geqslant 0$ is a tunable parameter, and $\mathbb{E}$ denotes the expectation with respect to

the randomness of the (g_{ij}). The quantity in (0.2) is called the *free energy*, and it is closely related to the random probability measure, called the *Gibbs measure*, that assigns to each configuration $\sigma \in \Sigma_N$ a probability proportional to

$$\exp\left(\frac{\beta}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij}\, \sigma_i \sigma_j\right).$$

To understand the presence of the factor of $N^{-1/2}$ in the exponential above, one needs to realize that the maximum of the comfort function (0.1) will typically be of order $N^{3/2}$; see Exercise 6.1. With the factor of $N^{-1/2}$ in the exponential in (0.2), we thus ensure that the terms that have the largest contribution to the sum in (0.2) are exponential in N. Since we are then summing over 2^N terms, this allows us to interpolate between a situation in which the entropy dominates, for small β, where the Gibbs measure resembles the uniform law over Σ_N, and a situation in which the energy dominates, for large β, where the Gibbs measure concentrates on the configurations that essentially realize the maximal value of the comfort function (0.1). In particular, a good approximation of $\beta N^{-3/2}$ times the maximum of the comfort function (0.1) is obtained by choosing β sufficiently large, not depending on N, in (0.2). This is made precise in Exercise 6.3.

The problem of identifying the large-N limit of the free energy (0.2) turns out to be surprisingly difficult. Starting in the late 1970s, Giorgio Parisi and collaborators [177–179, 217–221] proposed a way to solve this problem using sophisticated non-rigorous techniques, and the limit of (0.2) is now known as the *Parisi formula*. A rich phenomenology progressively emerged concerning the structure of the associated Gibbs measure, which turns out to organize itself along an ultrametric structure. Some key elements of this picture that were uncovered in the physics literature were then progressively put on a rigorous mathematical footing [132, 210, 211, 250, 253, 254]. Although many interesting questions remain, the mathematical understanding of the SK model, and of some generalizations thereof, is by now very substantial [41, 48, 61, 134, 210, 211, 213–215, 252].

There are however many seemingly innocent generalizations of the SK model that mostly remain mathematically mysterious. Motivated in part by considerations that relate to artificial neural networks (see e.g. [257]), we would like for instance to consider the generalizations of the SK model in which the spins are organized over two (or more) layers. To encode this precisely, we can represent a configuration as a pair $\sigma = (\sigma_1, \sigma_2) \in \Sigma_N^2$, where $\sigma_1 = (\sigma_{1,i})_{1 \leqslant i \leqslant N}$ and $\sigma_2 = (\sigma_{2,i})_{1 \leqslant i \leqslant N}$ represent the spins in each of the two layers, and where we choose the layers to be of the same size for convenience of notation. The energy function can then be written as

$$H_N(\sigma) := \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij} \sigma_{1,i} \sigma_{2,j}. \tag{0.3}$$

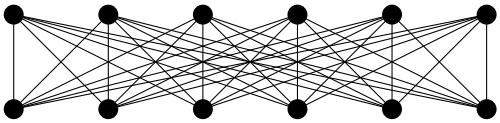

Figure 2. Illustration of the bipartite model for $N = 6$. Elementary units are organized in two layers and only interact across layers.

We will refer to this as the *bipartite model*. An illustration of the graph of interactions between the coordinates of σ is in Figure 2. Perhaps surprisingly, this model is much less understood than the SK model. In particular, the limit free energy of this model has not yet been identified rigorously.

Inspired by [42, 43, 64, 131, 200], we propose to approach this problem using a point of view based on partial differential equations [194, 196]. It turns out that one can identify the limit of the free energy (0.2) in the SK model as

$$\lim_{N\to+\infty} -\frac{1}{N}\mathbb{E}\log\sum_{\sigma\Sigma_N}\exp\left(\sqrt{\frac{2t}{N}}\sum_{i,j=1}^{N} g_{ij}\,\sigma_i\sigma_j - Nt\right) = f(t,0),$$

where $f = f(t,\mathsf{q}) : \mathbb{R}_{\geqslant 0}\times \mathsf{Q}_2(\mathbb{R}_{\geqslant 0}) \to \mathbb{R}$ is the solution to the infinite-dimensional Hamilton–Jacobi equation

$$\partial_t f(t,\mathsf{q}) - \int_0^1 \partial_{\mathsf{q}} f(t,\mathsf{q},u)^2\,\mathrm{d}u = 0 \quad \text{on } \mathbb{R}_{>0}\times \mathsf{Q}_2(\mathbb{R}_{\geqslant 0}),$$

with $\mathsf{Q}_2(\mathbb{R}_{\geqslant 0})$ being the space of square-integrable non-decreasing paths from $[0,1)$ to $\mathbb{R}_{\geqslant 0}$, and $\partial_{\mathsf{q}} f(t,\mathsf{q},\cdot)$ denoting the Gateaux derivative of f at (t,q); see (6.54). The initial condition $f(0,\cdot)$ to this equation is described by a functional transform of the Bernoulli measure $\delta_{-1}+\delta_{+1}$ encoding the "reference" law of one spin. Theorems 6.7 and 6.8 state this result more precisely. One of the most interesting aspects of this statement is that it suggests a very natural candidate for the limit free energy of the bipartite model and its generalizations, also phrased in the language of Hamilton–Jacobi equations; see Questions 6.9 and 6.11 for precise statements. An inequality between the limit free energy and this candidate limit is proved in [195, 197], but the converse bound remains an open problem.

The same approach has already led to more complete results concerning certain problems of statistical inference, as will be explained in detail in this book. For clarity of exposition, we will focus on the following relatively simple situation. We observe a noisy version of a rank-one matrix of the form $\bar{x}\,\bar{x}^*$, where $\bar{x} = (\bar{x}_1,\dots,\bar{x}_N)$ is a vector of independent and identically distributed (i.i.d.) random variables, and the superscript $*$ denotes the transposition operator. Precisely, we assume that we observe the matrix

$$Y := \sqrt{\frac{2t}{N}}\,\bar{x}\,\bar{x}^* + W, \tag{0.4}$$

where $t \geqslant 0$ is a free parameter that allows us to vary the signal-to-noise ratio, and $W = (W_{ij})_{1 \leqslant i,j \leqslant N}$ is a matrix of independent standard Gaussian random variables. We will see in Chapter 4 that this problem is closely related to that of community detection in random networks of Erdős-Rényi type with diverging average degree. We wish to answer the following question: given the observation of Y, can we recover meaningful information about the signal $\bar{x}\,\bar{x}^*$? One way to assess this is to monitor the minimal mean-square error

$$\mathsf{mmse}_N(t) := \frac{1}{N^2} \inf_g \mathbb{E}\big|\bar{x}\,\bar{x}^* - g(Y)\big|^2 = \frac{1}{N^2}\mathbb{E}\big|\bar{x}\,\bar{x}^* - \mathbb{E}[\bar{x}\,\bar{x}^* \mid Y]\big|^2 \tag{0.5}$$

between the signal $\bar{x}\,\bar{x}^*$ and its noisy observation Y. Here, the infimum is taken over the set of measurable functions g, and for a matrix a, we write $|a| := \sqrt{\mathrm{tr}(aa^*)}$, with tr denoting the trace operator. Assuming that $\mathbb{E}\bar{x}_1 = 0$ for simplicity, we can compare this minimal mean-square error with the error one would make by simply using the null estimator, defining

$$\mathsf{var}_N(t) := \frac{1}{N^2}\mathbb{E}|\bar{x}\,\bar{x}^*|^2. \tag{0.6}$$

It turns out that there exists a critical parameter $t_c \in (0, +\infty)$ such that the following holds: for $t < t_c$, essentially no information can be obtained about $\bar{x}\,\bar{x}^*$, in the sense that the difference between $\mathsf{mmse}_N(t)$ and $\mathsf{var}_N(t)$ becomes vanishingly small as N becomes large, while for $t > t_c$, this difference remains bounded away from zero as N tends to infinity. In fact, we will be able to characterize exactly the large-N limit of $\mathsf{mmse}_N(t)$. This result can be obtained using a variety of methods [32–34, 96, 109, 161, 162]. We will use a Hamilton–Jacobi approach developed in [69, 70, 72, 73, 193, 194] and identify the large-N limit of the minimal mean-square error (0.5) to be

$$\lim_{N \to +\infty} \mathsf{mmse}_N(t) = \big(\mathbb{E}|\bar{x}_1|^2\big)^2 - \partial_t f(t, 0),$$

where $f : \mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ is the solution to the Hamilton–Jacobi equation

$$\partial_t f(t, h) - \big(\partial_h f(t, h)\big)^2 = 0 \quad \text{on } \mathbb{R}_{>0} \times \mathbb{R}_{>0},$$

subject to some explicit initial condition. This Hamilton–Jacobi approach has been extended to a very large class of other models of statistical inference in [69] (alternative approaches have not yet reached this level of generality).

Before going into this, we will train our skills on the analysis of the much simpler Curie–Weiss model and its generalizations. The Curie–Weiss model is the non-disordered version of the SK model: its energy function is obtained by setting all the interaction parameters g_{ij} to be equal to 1 in (0.1). The identification of the limit free energy of this model is a straightforward consequence of Cramér's theorem on large deviations for the sums of i.i.d. random variables. This shortcut through large deviations does not seem to be available for problems in statistical inference or spin glasses.

Our first goal will be to recover the limit free energy of the Curie–Weiss model and its variants using the Hamilton–Jacobi approach. This will give us the opportunity to develop this approach and its associated toolbox in the simplest possible context. This toolbox will then be ready for us to use, with no further addition necessary, when we turn to the problems of statistical inference.

Organization of the book

This book is organized as follows. We start in Chapter 1 with an introduction to the basics of statistical mechanics, where we motivate the notion of a Gibbs measure on physical grounds, and introduce the Curie–Weiss model as well as its generalizations. In Chapter 2, we develop the fundamentals of convex analysis and large deviation principles and use these to compute the limit of the free energy in the Curie–Weiss model and its generalizations. This large-deviation approach is not applicable to the disordered models we want to consider next, so we need to develop an alternative approach.

In Chapter 3, we define the notion of viscosity solution to a Hamilton–Jacobi equation and use it to recover the limit free energy of the Curie–Weiss model. We discover technical challenges to applying the same method to generalized versions of the Curie–Weiss model and develop a new "selection principle" based on convexity to overcome these. We then turn to statistical inference in Chapter 4, focusing on the problem of recovering a large symmetric rank-one matrix from a noisy observation. We discover that the tools developed in the previous chapter apply to this setting as well and allow us to give a closed-form description of its phase transitions. Chapter 5 is preparatory work for a discussion of the more challenging case of spin glasses. The first half of this chapter is a self-contained introduction to Poisson point processes, including limit theorems on extreme values of independent and identically distributed random variables, which we believe to be of wide interest. We finally turn to the setting of spin glasses in Chapter 6. For the Sherrington–Kirkpatrick model, we show how to relate the Parisi formula with the Hamilton–Jacobi approach. We conclude with a more informal discussion on the status of current research for more challenging models. Appendix A is a self-contained presentation of many of the basic results in real analysis and probability theory that are used throughout the book. Solutions to the exercises are also provided.

Related works

Our main goal with this book is to provide the reader with an inviting presentation of some results and open problems in mean-field disordered systems which we find fascinating, following the common thread of the Hamilton–Jacobi approach. We wish to stress however that we do not claim this to be the one-and-only definitive way to approach these problems.

Trying to survey the range of works and topics related to those discussed in this book is a daunting task. We will attempt to fulfill it to the best of our abilities, hoping that the references listed here will at least provide the reader with good entry points to explore further according to their interests. Apologies to all those whose work is not included here.

Chapters 1 and 2 cover very classical material on statistical mechanics, convex analysis, and large deviations. References on these topics include [117], [53, 108, 138, 187, 230], and [92, 93, 258], respectively.

Concerning the theory of viscosity solutions to Hamilton–Jacobi equations discussed in Chapter 3, the most classical reference on the topic is probably [84]; we also mention [38, 39, 115] for more accessible presentations that focus on first-order equations. The notion of viscosity solution was introduced in [83, 85, 114], and the Hopf–Lax and Hopf formulas were introduced in [140,159] and proved to be viscosity solutions in [165] and [37, 166], respectively. The infinite-dimensional Hamilton–Jacobi equations discussed in Chapter 6 are developed in [74, 75, 195].

The problems of statistical inference discussed in Chapter 4 are also explored in the surveys [185, 268], along with several aspects not covered here. We also mention the monograph [175] presenting the interplay between statistical mechanics, information theory, and combinatorial optimization, with extensive coverage of a class of methods called message-passing or belief-propagation algorithms that have shown their usefulness in a wide variety of circumstances.

The calculation of the limit free energy (or mutual information) for the inference problem discussed in Chapter 4 and its generalizations has been approached using a large variety of techniques. Together with an interpolation in the spirit of [132–134], these may involve algorithmic approaches in [32, 95, 96, 148, 162], a cavity method in the spirit of [13] in [161, 163, 172, 182, 228], concentration of measure in [109], or adaptive interpolation in [33–35, 168, 169, 227]. The Hamilton–Jacobi approach presented in Chapter 4 has been developed in [69, 70, 72, 73, 193, 194]. Examples of models that are covered by this approach but currently not by other techniques are discussed in [168, Section 7].

In Chapter 4 on statistical inference, we are mostly concerned with determining whether or not one can recover meaningful information about the signal from a noisy observation, and if so how much. In practice, it is also fundamental to know whether or not one can recover this information in a reasonable amount of time. We barely scratch the surface on this point in Section 4.4. The example in Proposition 4.21 shows a situation in which for some regime of parameters, it is theoretically possible to recover meaningful information about the signal, but the standard (and rapid) PCA method completely fails. In fact, one expects that in this regime, there is no polynomial-time algorithm that recovers meaningful information about the signal. This difference between the moment when one can theoretically infer non-trivial information about the signal and the one when one can do so efficiently is often called

the statistical-to-computational gap. It is believed to exist in a wide variety of situations, and we refer to [31, 123, 125, 268] for surveys on this.

In Section 4.5, we discuss the relationship between the inference problem (0.4) and the problem of detecting communities from the observation of a random graph where the probability for an edge to be drawn between two nodes depends on the community of each node. A survey on this problem of community detection is [1]; we also refer to [91] where many predictions were put forward at the physics level of rigor. In the setting in which the average degree of the graph of connections remains bounded, the identification of the regime of parameters for which one can reconstruct non-trivial information about the community structure from the observation of the random graph was achieved in [171, 188, 191] for models with two communities. To the best of our knowledge, the identification of a closed-form description of this regime of parameters remains open in settings with more than two communities, and the difficulties seem to be of a similar nature as those encountered in the analysis of the bipartite spin-glass model (0.3). We refer in particular to [5] for positive results in this direction, and again to [91] for several predictions. A related problem concerns the determination of the asymptotic mutual information between the community structure and the graph of connections. This problem was successfully resolved in "convex" cases [4, 80] and even in some "non-convex" cases [3, 129, 149, 189, 192, 267], but a fully general solution has still not been identified [129]. (What counts as a "convex" or a "non-convex" model should hopefully become clear upon reading Chapter 6.) This problem seems very similar to those presented in Questions 6.9 and 6.11 in the context of spin glasses, as discussed further in [104, 153].

Books with a focus on spin glasses include [55, 60, 63, 67, 81, 88, 179, 203, 205, 211, 240, 253, 254]. The construction of Poisson–Dirichlet cascades in Sections 5.5 and 5.6 essentially follows [211]. We also refer to [211] for a complete proof of the Parisi formula (6.7).

The idea that the limit free energy of the SK model should be described by the Parisi formula emerged in the series of works [217–219] based on the non-rigorous replica method. The physical understanding of the model progressed further in many contributions including [177–179, 220]. A mathematical proof of the Parisi formula was then obtained in [132, 250]. A more robust proof was later found in [210, 212] using the idea of ultrametricity discussed in Section 5.7, and also inspired by [11, 13, 21, 57, 128, 207–209]. We refer to [211] for a more thorough presentation of historical developments on this topic. The Parisi formula for spherical models was obtained in [77, 249] and takes a form that was predicted in [86].

Early contributions in the physics literature on the bipartite spin-glass model (0.3) include [119, 120, 152, 156]. On the mathematical side, the limit free energy of spin glasses with multiple types has been identified in a number of cases. For Hamiltonians such that (6.129) holds, the limit free energy has been identified in [41, 206, 213–215] when ξ is convex over $\mathbb{R}^{D\times D}$; see also [47, 48, 155, 216] in the spherical case. The

more general case when ξ is convex over the space of positive semi-definite matrices is obtained in [71, 75, 195, 197] using some of the ideas presented here. For fully general models in the form of (6.129), Theorem 6.12 from [71] imposes strong constraints on what the limit free energy can be; it is at present unclear whether this statement is a complete characterization of the limit free energy or not (even though there are choices of (t, q) such that (6.135) is satisfied for multiple pairs $(\mathsf{q}', \mathsf{p})$, it may still be the case that, for instance, there is only one continuous function f that satisfies the properties listed in Theorem 6.12). Spin glasses with multiple types were already present in the first proof of the Parisi formula from [250, 251, 253, 254], in the form of two copies of the original system coupled together through a constraint on their overlap.

That there exist connections between limit free energies of statistical-mechanics systems and solutions to Hamilton–Jacobi equations dates back at least to [64, 200]; see also [49] for a survey of related contemporary research topics. In the context of spin glasses, heuristic connections between limit free energies and Hamilton–Jacobi equations were first pointed out in [8, 42, 43, 131], under a replica-symmetric or one-step replica symmetry breaking assumption. The possibility to rephrase the Parisi formula in terms of a Hamilton–Jacobi equation as in Theorems 6.7 and 6.8 is from [196, 198]. Theorems 6.10 and 6.12 are from [195] (generalized in [197]) and [71], respectively. A high-temperature version of Theorem 6.12 is in [97].

For spherical models with multiple types, the limit free energy is identified in some non-convex cases in [244–247] using an approach inspired by the early work [255], and in [29] for the bipartite spherical SK model. We also mention [51, 144, 154, 174] for a geometric analysis of the energy landscape of spin glasses with multiple types; see also [23, 24, 118, 242] in the single-type case.

By taking a low-temperature limit, the determination of the limit free energy of spin glasses allows one to infer the asymptotic value of the maximum of the Hamiltonian; see Exercise 6.3 and [26, 78, 102]. In the context of the Sherrington–Kirkpatrick model, with H_N as in (6.3), this would amount to determining the asymptotic behavior of the maximum of $H_N(\sigma)/N$ over $\sigma \in \Sigma_N$; we denote the limit by OPT. One may ask whether there exists an efficient algorithm for actually finding a configuration $\sigma \in \Sigma_N$ such that $H_N(\sigma)/N$ is close to OPT. The answer turns out to depend on the specifics of the model and relates to the overlap gap property [124]. A specific value ALG described by an analogue of the Parisi formula was identified such that there exists an efficient algorithm that can identify a configuration σ with $H_N(\sigma)/N$ approximately equal to ALG with high probability [110, 184, 236, 243]. Moreover, it was argued in [142] that no algorithm within a broad class can exist which improves upon this value. Remarkably, these results on algorithmic thresholds have been extended to non-convex spherical models such as the bipartite model in (0.3), despite the fact that we do not know of a characterization of OPT or of the free

energy in this case [143, 145]. The reference [27] surveys a number of results related to optimization algorithms for spin glasses.

The study of mean-field spin glasses has inspired developments in many other contexts. Works that explore connections between spin glasses and neural networks include [9, 40, 44–46, 89, 122, 136, 170, 257]; see [7, 18, 90, 113, 121, 137, 226, 256] for overviews. For a suitable choice of the reference measure denoted by P_N in Chapter 6, the law of the spins of a given type in the bipartite model (0.3) is the same as the law of the spins in the Hopfield model [40]. Some versions of the perceptron model can be obtained similarly. The Hopfield and perceptron models were introduced in [17, 141, 167, 173, 181] as toy models for memory storage and retrieval or classification tasks. Early works on the statistical mechanics of these models include [19, 20, 126, 127, 222–224]; recent rigorous works include [56, 101, 254, 265]. Among many other topics related to spin glasses, we mention random constraint satisfaction problems [98, 100, 158, 175, 180, 183], the random assignment and traveling salesman problems [14, 15, 176], error correcting codes in information theory [229], and combinatorial problems such as graph coloring [79, 99, 199]. A recent book on "spin glass theory and far beyond" is [67].

Acknowledgments

This book grew out of lectures given by JCM at the 2021 CRM-PIMS summer school in Montreal, at ETH Zurich in 2022, and to graduate students at New York University and ENS Lyon. The two of us (TD and JCM) met on the occasion of the summer school in Montreal and have been working together since then. We are extremely grateful to the organizers of these events for their support, including financial support to TD from ETH Zurich to take part in the writing of this book. We also very warmly thank the participants of these two events and of the graduate courses for their interest and very useful feedback.

Course outline

We describe here how the series of lectures in Zurich were organized. We do not believe that this is the optimal organization, and some of the results discussed in the book were not known at the time (in particular Theorem 6.12), but we hope that it can still be of some interest. This series of lectures comprised 12 sessions of 2×45 min. The first session was an overview of motivations and a description of the plan for the course. The second session started with a refresher on large deviations concluding with Theorem 2.16, followed by the contents of Section 1.1. The third session covered the rest of Chapter 1 and finished with Theorem 2.19. The fourth session covered Sections 3.1 and 3.2. The fifth session contained the proof of the comparison principle in Theorem 3.5 in the simpler case where the space domain $\mathbb{R}^d$ is replaced by

the torus, the statements of the variational formulas in Theorems 3.8 and 3.13 without proofs, and concluded with Corollary 2.20 recovered using the Hamilton–Jacobi approach. The plan for the sixth session was to cover Section 3.6, with a direct proof of the convex selection principle in Theorem 3.21 that bypasses Lemma 3.23, and also to present some material from Section 2.1.3 on subdifferentials. This actually spilled over to the seventh session. The rest of this session was spent discussing the setup of the statistical inference problem in Chapter 4, and motivating it informally in relation with the problem of community detection discussed in Section 4.5. The eighth session covered Section 4.1 and Section 4.3 up to (4.50), mostly skipping Section 4.2 except for the simple Gaussian integration by parts in (4.19) and some simple generalization of it. The ninth session was meant to cover the rest of Section 4.3, taking the concentration of the free energy for granted (any estimate stating that (4.67) tends to zero with N will do, and one can show an upper bound of the order of $N^{-1/3}$ on this quantity using simple arguments based on the Efron–Stein and Gaussian Poincaré inequalities in Exercises 4.7 and 4.8). This actually spilled over a bit to the tenth session. Also, the proof of Theorem 4.9 was only obtained under the additional assumption that $\mathbb{E}\bar{x}_1 = 0$, since the convex selection principle was only shown in the form of Theorem 3.21, but the more refined Lemma 3.22 is needed to conclude in general here. The major part of the tenth session was spent covering Section 6.2 and its analogue for the bipartite model, and discussing the new difficulties that show up. The eleventh session started with a discussion of the random energy model from Section 6.3 and then presented the main results of Section 5, in particular Proposition 5.13, as well as the elementary properties of the Poisson–Dirichlet process from Section 5.5. In the last session, the Poisson–Dirichlet cascades from Section 5.6 were defined, and the contents of Section 6.4 were covered without proving everything and with some hand-waving for the remainder of Chapter 6. A recording of these lectures can be found at https://tinyurl.com/HJ-ETHZ.

Tomas Dominguez
University of Toronto

Jean-Christophe Mourrat
ENS Lyon and CNRS