学　位　論　文

# Replica Symmetry Breaking and Multivalley Energy Landscape in the Hopfield Model

ホップフィールド・モデルにおける
レプリカ対称性の破れと
エネルギーの多谷ランドスケープ

平成５年１２月博士(理学)申請

東京大学大学院理学系研究科

相関理化学専攻

時田恵一郎

# Replica Symmetry Breaking and Multivalley Energy Landscape in the Hopfield Model

by

Kei Tokita

Doctorate thesis, submitted in partial fulfillment of the
requirements for the degree of Doctor of Science
at the University of Tokyo.
December 1993

# Preface

The interdisciplinary field characterized by keywords like "neural networks" or "neural computation" has drawn much interest recently. Despite sharing neural systems as a common topic, the motives and perspectives of the researchers may indeed be very different: neurophysiologists are trying to construct an appropriate model which provides insights for the functions and the organization of brains; engineers and information scientists are aspiring to find alternative paradigms to the traditional one introduced by von Neumann and used as the basis of almost all machine computation to date; psychologists are utilizing neural networks for elucidating the sensory and cognitive processes.

Nowadays, it seems that neuroscience has become one of the most important disciplines with subjects of general interest. It has exerted a strong influence on various fields and physics is no exception. Although the theoretical approaches of physics to neural systems are relatively new in the field, the past decade has seen surprising successes and rapid progress. There seem to be several reasons. A first one is the enormous advance of computer systems. These enable us to investigate neural network models extensively by simulation. Such *computational science* is emerging as the third pillar of science next to the theoretical and experimental one. In particular, recent workstations provide us with a very convenient environment for developing sophisticated simulation programs and for visualizing results in a way in intelligible to the human mind. A second one is that many physicists in the statistical mechanics community rushed into that field being fascinated by an analogy between neural networks and the so-called "spin glasses (SG)". These are systems which have randomly distributed ferromagnetic and antiferromagnetic interactions. The mean field theory of SG has yielded many surprising results. It seems that many physicists have been trying to climb the double-spiral staircase of neural networks and SG to reach a sanctuary for learning of "complex systems", even though it is not guaranteed that the staircase is really connected to such a sanctuary... Wherever it may lead though, I do not expect it to finish in a dead end, which, of course, doesn't mean that there wouldn't be many obstacles. In other words, the understanding and even the focal points of attention may change a lot, but most likely some of the basic ideas and analogies inspired by SG will remain.

In the present situation where different methodologies coexist, how can the physical sciences contribute constructively to this field? In short, I believe it lies in discovering universalities in artificial neural network models. Although it goes without saying that all neural network models studied in physics are extremely simplified and cannot directly be compared with the neural systems of organisms, such simplified models do have the mighty advantage of allowing for systematic analysis. Moreover they may shed light on the principles underlying "computation" in biological systems, and on how those principles differ from the ones that we have so successfully applied in digital computers. Neural network models are also very rich sources of inspiration for the design of artificial computing networks, and the theoretical results may very well find their way into practical application.

To return to the broad context of complex systems (theory) I would like to point out the recent attention which is given to the emergence of collective behavior in groups of simple elements. I feel that the terms "Connectionism" and "Collectionism" symbolize the standpoint that complex behavior, not only in brains but also in complex systems, can emerge from the collection and cooperation of numerous "rather simple" elements. Here I used the term "simple" as an adjective for the basic units such as formal neurons in neural network models, amino acid residues in proteins[41, 15, 100, 129], sand particles in dunes[88], model agents in economic systems[130] and so on. On the other hand, the term "rather" indicates that such units should remain "reasonably" complex, like, e.g., logistic maps in coupled map lattices[63, 128]. From a theoretical viewpoint, understanding the dynamical properties of large, coupled systems is a challenging problem in its own right.

*About the thesis*

This thesis focuses on the Hopfield model of neural networks, which heralded the arrival of SG ideas into the arena of neural computation. The Hopfield model was/is extremely successful within the limits of the replica symmetry (RS) approximation, which was originally introduced in the Sherrington-Kirkpatrick (SK) model of SG[106]. However, several open questions remained.

1. How is a series of results in the RS discussion modified and affected by replica symmetry breaking (RSB) ?



   RSB was first introduced by Parisi[94, 95] to overcome some limitations of the RS approximation and to complete the framework of the replica method for an appropriate estimation of physical quantities.

2. How is a multivalley energy landscape organized? It is an origin of or an obstacle to an associative memory function.

3. How is the Hopfield model related to the current SG model?

4. What can be known about several features which are hard to detect by analytical approaches?

In the work reported here, several answers to these questions are supplied for the first time. This was achieved by employing the full RSB formulation, large-scale numerical simulations and other methods. The problem in the first question was tackled by exploiting the fact that the RS solution is unstable in the SG phase, and even in ferromagnetic retrieval (FMR) phase (where the system exhibits an associative memory) below a specific temperature. In the low temperature limit, at any rate, only the full RSB solution can provide a proper estimation of the order parameters and other physical quantities. The full RSB formulation also answers the second question by providing us with parameter characterizing the structure of multivalley energy landscape. It turns out that the valley structure continuously changes from a simple one with two basins to a complicated one. The theory developed here reveals nontrivial asymptotic shifts of the system to the Sherringron-Kirkpatrick (SK) model of SG, and the limit established thus is introduced as the important concept of the "SK limit". Consequently, the third question is simultaneously solved: in the "SK limit" the mean-field equations of the Hopfield model are formally transformed into the SK model of SG shining light on an interesting facet of the relationship beween these models. The last question is treated by a series of numerical analyses using finite size scaling, especially for the basins of attraction and generalized remanent magnetization. The numerical analysis also supports the presented notion of the "SK limit" by showing that the remanent magnetization in that limit attains same value as the one known in the SG theory.

In this thesis, I have also tried to provide a clear account of the statistical mechanical idea underlying neural networks. I hope to provide valuable information for both novices and experts alike. Furthermore, although the

discussions are rather concentrated on the mean-field theory of the Hopfield model and its relationship to spin glasses, it is my hope that this thesis can contribute to the furtherance of "complex systems" in a broader sense.

The main contents of this work are as follows.

Chapter 1: This part is a review of the material most relevant to this thesis as prior knowledge. First, the history of the discipline is briefly outlined, and the formal neuron and its time evolution are introduced. Short notes on the notions of spin glasses and the rugged multivalley landscape are also provided.

Chapter 2: The Hopfield model is introduced, and its physical and biological relevance are stressed. In particular, the mean field theory by Amit, Gutfreund and Sompolinsky (AGS) for the Hopfield model is traced in detail.

Chapter 3: The full RSB solution of the Hopfield model at finite temperature is formulated using Parisi's RSB scheme in order to investigate the multivalley structure of the rugged free energy. It is found that the resulting variational equations are equivalent to those for the SK model as a limit, i.e., the "SK limit" of the Hopfield model was established. This had previously been suggested several times but never been proven explicitly[8, 52, 45]. Numerical solutions are obtained for the SG phase. These provide us with a weight distribution of the valleys and its dependence on a parameter $\alpha \equiv p/N$ for the rate of memory-loading, where $p$ is the number of random patterns for memories and $N$ the system size. It turns out that the multivalley structure continuously changes from a simple one (corresponding to the Mattis state[75]) to a complicated one characterizing the coexistence of an infinite number of metastable states (corresponding to the SK model) as $\alpha$ gets larger and goes to the SK limit.

Chapter 4: The full RSB solution of the Hopfield model at zero temperature ($T = 0$) is investigated. By using an RSB scheme by de Dominicis, Gabay and Orland, a free energy functional in the so-called *Sompolinsky gauge* and variational equations are formulated. The resulting equations are conveniently defined for numerical analysis since the singularity at $T = 0$ is formally avoided. Elaborate numerical analyses have revealed that the critical storage capacity for the memory patterns at $T = 0$ has to be corrected to $\alpha_c = 0.155 \pm 0.002$ which is larger than the corresponding values obtained from both the RS (0.138) and the 1-step RSB (0.144) discussions. Furthermore, the order parameter functions and frozen field distributions (FFD) both



in the SG phase and the FMR phase are determined. The FFD in the SG phase shows a nontrivial double-peaked form, and how it depends on $\alpha$. On the other hand, the FFD in the FMR phase reveals its characteristic broken-symmetry. It turns out that the FFD in the FMR phase is slightly but nevertheless clearly different from a Gauss distribution assumed in the RS discussion.

Chapter 5: Remanent overlaps, the generalized remanent magnetization, and their basins of attraction are numerically studied for the Hopfield model with zero temperature sequential dynamics. Relationships between remanent overlaps and initial overlaps are obtained for relatively large $\alpha$'s. The asymptotic dependence of remanent overlaps on $\alpha$ is also shown. Remanent overlap in the SK limit gives the same value of the SK model[66].

Chapter 6: This chapter is devoted to a summary of this thesis and some concluding remarks.

Some formal derivations that are somewhat complicated but important are provided in the Appendices.

# Acknowledgements

It would be impossible to name all the people who have directly or indirectly helped me to bring this research to a successful conclusion. I would like to thank explicitly however: my supervisor Prof. K. Kaneko for his trust, continuous support and encouragement, and for allowing me to work freely in an open atmosphere; Prof. Y. Aizawa for introducing me to nonlinear science, and for setting an example as an earnest researcher; Dr. K. Nemoto, to whom I am greatly indebted for advice in choosing the topic of my thesis, for many illuminating discussions, and for making available his highly sophisticated program for a numerical analysis.

A special thanks also goes to Dr. F. H. Willeboordse who not only helped me in solving numerous computer-related problems, but also for conscientiously reading many parts of the papers on which this thesis is based, as well as for always finding the time for illuminating discussions.

I shall further never forget hospitality by all the other staff with the continuous support, the students of the University of Tokyo and many people participated in a series of the "Shuuhen" seminar held at Komaba.

Last but not least, I have no words for expressing my gratitude to my wife Yumiko whose support and patience kept me motivated even during the inevitable setbacks.

List of papers submitted for the requirement of the degree:

1. Kei Tokita
   The Replica Symmetry Breaking Solution and the "SK limit" of the Hopfield Model.
   Journal of Physics A: Mathematical and General, **26**, pp6915-6925 (1993)
   (Chapter 3.)

2. Kei Tokita
   The Replica Symmetry Breaking Solution of the Hopfield Model at Zero Temperature : Critical Storage Capacity and Frozen Field Distribution.
   Submitted to Journal of Physics A: Mathematical and General.
   (Chapter 4.)

3. Kei Tokita
   Remanent overlaps in the Hopfield model with zero temperature sequential dynamics.
   Progress of Theoretical Physics, **90**, pp329-335 (1993)
   (Chapter 5.)

List of papers added for references:

1. Y. Aizawa and K. Tokita
Zipf Law in the Non-linear Dynamics and Neural Nets.
Proceedings of the Intenational Conference on Fuzzy Logic & Neural Networks IIZUKA '90.
pp.637-640. (1990)

2. K. Tokita and Y. Aizawa
Statistical Mechanics of Neural Nets.
Proceedings of the First NOLTA Workshop on Nonlinear Dynamical Theory of Adaptively Learning Systems. pp.13-20. (1991)

# Contents

# Chapter 1

# Introduction

## 1.1 History

Although it is not the aim of this chapter to closely trace the entire history of theoretical approaches for studying brains or neural networks, let us briefly touch several significant contributions (for details, see text books like[10, 48, 57, 112]).

No one will contradict that the study of neural network models begun with the work of McCulloch and Pitts[76] in 1943, who introduced the notion of the formal neuron as a two state threshold element and showned that networks of such elements can implement any logical function. In other words, they proved that a synchronous assembly of such formal neurons is capable in principle of universal computation for suitably chosen connections between neurons. By representing the state $S_i$ of neuron $i$ as firing ($S_i = 1$) or not firing ($S_i = 0$) respectively, the time evolution can be represented as

$$S_i(t+1) = \theta\left(\sum_j J_{ij} S_j(t) - u_i\right), \tag{1.1}$$

where the time $t$ is discrete and $\theta(x)$ is the unit step function (Heaviside function). The connection matrix $\mathbf{J} = (J_{ij})$ represents the strength of the synapse connecting neuron $j$ to neuron $i$. A positive or negative value of $J_{ij}$ corresponds to an excitatory or inhibitory synapse, respectively. The absence of a synapse can be represented by $J_{ij} = 0$. The cell specific parameter $u_i$ refers to the threshold value for neuron $i$; the weighted sum of inputs must reach or exceed $u_i$ for the neuron $i$ to fire.

Next, a hypothesis made by Hebb[53] about the way in which synaptic strengths in the brain change in response to experience should be noted. Hebb suggested changes proportional to the correlation between the firing of the pre- and post-synaptic neurons.

The significant notion of a formal neuron by McCulloch and Pitts, and the variable synaptic connection by Hebb were followed by the study of a variety of models for associative memory, pattern recognition, various classification tasks and even for combinatorial optimization problems.

The first noticeable wave of activity in this field occurred around 1960 due to work by Rosenblatt[97] which focused on the problem of how to find an appropriate synaptic connection for a particular computational task. He concentrated on networks called *perceptrons*, in which the units were organized as layers with feed-forward connections between one layer and the next. Rosenblatt was able to prove the convergence of a learning algorithm of a "simple perceptron"; literally the simplest class of perceptrons without any intermediate layers. That is to say, a general method was established to change the synaptic connection iteratively so that a desired computation was performed. Unfortunately after the appearance of the famous book *Perceptron* by Minsky and Papert[82] which pointed out limitations in the ability of perceptrons for learning a task, like e.g. the exclusive or (XOR) problem, the computer science community left the neural network paradigm untouched for almost 20 years. It should be noted though that other such adaptive devices were suggested independently by Amari[4, 5], Caianiello[23], Grossberg[51], Kohonen[68] and Marr[73, 74].

Let us concentrate on the roles of physics as the latest partner in neuroscience. The first thread of development can be traced back to the analogy between the activity of a neural network and the collective states of coupled magnetic dipoles made by Cragg and Temperley[25, 26] in 1954. Twenty years later, the same theme was taken up by Little[69, 70, 71] who made the analogy between synaptic noise and temperature, and suggested that persistently firing states of neural networks appear just like the ordered phases in magnetic systems.

The first major impact on physics was provided by Hopfield's work[58] on a neural network with symmetric connections. He added some helpful physical insight by introducing an energy function, and by emphasizing the notion of memories as dynamically stable attractors and the equilibrium thermodynamical properties of random magnetic systems similar to spin glasses. His papers triggered an explosive growth of the field, particularly in the statistical physics community, leading to a whole series of dramatic advances in the understanding of symmetric networks and their properties. This was especially the case because neural networks may be viewed as distributed optimization problems, e.g., the Traveling Salesman Problem (TSP)[59]. The Hopfield model was subsequently solved analytically by Amit, Gutfreund and Sompolinsky (AGS)[8, 9]. This was the first successful, and non-trivial application of statistical mechanics to neural networks, giving rise to a variety of new and surprising results. I would like to deal with this model in detail by using a formal theory of the replica method. The definition of the Hopfield model and the results from statistical mechanics will be given in the next chapter.



The second influential development during this decade is the Backpropagation (BP) algorithm[99] which works quite well for adjusting the weights connecting units in successive layers of multi-layer perceptrons. Actually, this had already been discovered in the mid 70's[127], but not received much attention at that time. Due to the great ability of the BP model in overcoming nonlinear separation problems such as the XOR problem, ref. [99] triggered a massive explosion of work on trainable neural networks which continues to this day. The statistical mechanical approaches for the multi-layer perceptrons[17, 101, 126] seem to be a brand-new theme in the field of neural networks.

## 1.2 Spin Glasses

Slightly before the explosion of work on neural networks, spin glasses (SG) emerged as a fascinating new topic in statistical physics around the mid 1970's. An SG is a collection of spins (i.e., magnetic moments) whose low-temperature state is a frozen disordered one and is completely different from the kind of uniform or periodic pattern we are accustomed to finding in conventional magnets. It is found that such a state originates from two essential ingredients. First, there must be some sort of competition among the different interactions between the spins, in the sense that no single configuration of spins is uniquely favored by all the interactions (this is commonly called "frustration"). Second, these interactions must be at least partially random. These facts suggest that the SG state is intrinsically different from conventional forms of order

and requires new formal concepts to describe it. This challenge has been the fundamental motivation for theorists in this field. Many of the marvelous results and notions have been reported in several excellent review papers and books[36, 124, 37, 19, 125, 24, 81, 113, 114, 38].

Here, let us only briefly survey the history of the mean field theory of SG along the replica method which gives important conceptual backgrounds for this thesis. The replica method was introduced by Edwards and Anderson[35] to analyze the SG model and has been subsequently applied to a mean field model of SG, the Sherrington-Kirkpatrick (SK) model[106] with infinite-range interactions. The replica symmetric (RS) solutions of the SK model clarified important properties of SG, but, at the same time, also a fatal theoretical defect became apparent: negative entropy at low temperature. The immediate question was of course whether this was caused by the replica method itself or a consequence of the way it was applied. Fortunately, the answer was that it was NOT a failure of the Replica Method as such but that are need to use the so-called replica-symmetry-BREAKING (RSB) solution instead of the RS solution, as was suggested by Almeida and Thouless (AT)[28]. In the end, the RSB solution proposed by Parisi[94, 95] turned out to satisfy all the conditions requested by AT. Moreover, Parisi's solution provided us with a definite physical interpretation and estimated values for the physical quantities. It also solved the negative-entropy problem. At the same time, the number of metastable states was obtained as a function of temperature by Bray and Moore[21]. It is obtained by counting the number of solutions of the SG model proposed by Thouless, Anderson and Palmer (TAP)[117]. The brand-new SG pictures suggested by the RSB theory was verified by several investigations like Monte Carlo simulations[131] and direct numerical solutions of the TAP equations[84, 85]. The subsequent research on the RSB solution by Mézard, et al. revealed a surprising novel property of the metastable states: the ultrametric organization of pure states. Although this property of the SG phase was thought to be rather specific to the SK model, several investigations have shown that this is not the case and that it is shared by other infinite-ranged SG models[22, 115, 89, 90]. These abovementioned works form the background of the present studies of the Hopfield model.

Taking up new arms, statistical physics recently focused with great success on the emergence of collective behavior in large assemblies of elements. This naturally was closely reflected in the advance of the theory on phase transitions.



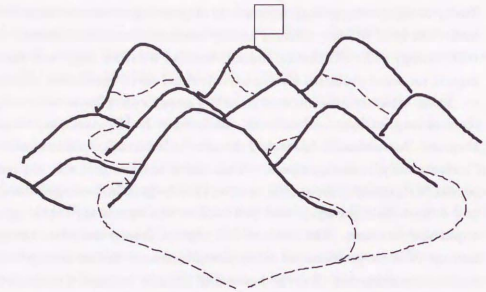
Today nothing is surprising any more in attempting to connect ideas developed in the study of SG to problems arising from or associated (however vaguely) with biology. The SG theory, indeed, has had a rather large and unexpected impact on some problems far from SG themselves in condensed matter.

Then, what *are* the notions from SG physics that have been used in biological applications and others? To borrow P. W. Anderson's words[14], they are "frustration", "quenched disorder", "replicas", "ultrametricity", and "metastability", among others. While these notions probably do not apply to real SG[1], many other problems have effectively infinite-ranged interactions, and a mean field theory (sometimes with replica symmetry breaking) may be applicable to them. The fields of SG physics proper and that applying the notions of it to problems in other disciplines have rather diverged. Several works have suggested that the notions of SG can be applied to some biological systems such as neural networks, prebiotic or adaptive evolution, random polymers, protein foldings, and so on[112], even though the discussions are rather conceptual in each limited situation. The SG theory was also applied to some combinatorial optimization problems and even information coding theory[111, 98, 87]

In particular, the neural network models may turn out to have more relevance to their interpretation of "long-ranged" interaction than the mean field models of SG themselves for real SG materials. This is because, e.g., axon of an neuron can grow over a large area of the cortex, implying that the interactions are essentially long-ranged, while a notion such as "ultrametricity" from the mean field theory is never found in the more realistic short-ranged SG. Furthermore, another important point is that, to good approximation in many kinds of neurons, every firing of a given neuron is identical with every other firing. That is, a cell has effectively just two meaningful states, firing and not firing. It can therefore be described by a binary variable, opening the way for adopting a spin analogy.

Progress in these areas and in biology-related investigations are strongly coupled, and we can reasonably expect the flow of ideas to travel in both directions. In other words, we hope that understanding spin glasses could be a key that unlocks the secrets of many other "complex systems". Physics has been an important factor in making spin glass theory such an active field in

[1]Theories for more realistic short-ranged models suggest that, unlike the infinite-ranged model, no more than two pure states can exist for any temperature and field[61, 39].



Which is the shortest route?

the last decade.

## 1.3 The Notion of Multivalley Energy (Fitness) Landscapes

Last but not least, the idea introduced in this section is another central topic of this thesis: the rugged or multivalley energy landscape. It may be one of the most important notions of SG and has had many repercussions in other fields.

Let us illustrate the problem of a rugged landscape by considering a person who lost his way in a wooded and "rugged" mountain. Although almost everybody would go downhill, such a strategy would be inadequate from the viewpoint of a professional alpinist; for a craggy place around a mountain stream may often prevent him from going down the rest of the way. Alpinists would aim at a ridge with a fine view. At first their strategy seems to be disadvantageous because they would go higher up, however, in the end they would be able to find their way successfully. On the other hand, it is much



easier in the case of bald and gently-sloping hills like the ones of a golf course where one would be able to reach the highest or the lowest place because one has a distant view. Such a fine view, therefore, eliminates a hit-or-miss search. That is, a sightless or random search should be an origin of difficulty for the people who lost their way. Likewise, the reason that the Labyrinth of Knossos functioned as a prison is that a maze is difficult for a human being to deal with (even though the Minotaur was not a human!). It is because in the maze there is only a limited view, and decisions for searching a proper way have to be made "locally". These things are not only related to mountain climbing or a labyrinth, but also to more general combinatorial optimization problems which organisms often encounter through evolution. In fact the notion of the "rugged landscape" can be applied to studying the problem of prebiotic[13] and adaptive[64, 65] evolution with several successful results, even though we should consider variable and adaptive landscapes[62, 65] if we want to address evolution properly.

In the language of spin glasses, a search for a ground state corresponds to nothing but an optimization in a multivalley energy landscape. This is also known as one of the "NP-complete" problems. Furthermore, *the simulated annealing* algorithm[67] has turned out to be very efficient for such optimization problems. Analyses, using the replica method, again reveal the connections between that algorithm and combinatorial optimization problems such as the Travelling Salesman Problem (TSP) and the matching problem[79, 42, 80].

The part of the title "multivalley energy landscape" of this thesis is motivated by the topics mentioned above. It may safely be said that the history of evolution is an iterative adaptation both of a species and the environment. Moreover, an adaptive evolution may promote a more complex environment[65] which gives rise to a rugged "multivalley landscape". Such a situation can be latent also in neural networks like the Hopfield model. One of the most plausible environmental adaptations in brains could be revealed as a morphological variation of the network. A novelty of the Hopfield model in this sense is that the structure of the multivalley energy landscape continuously varies as a parameter changes. This is also effectively related to an ability and a mechanism of memory association. Corresponding topics will be discussed in Chapter 3.

# Chapter 2

# Mean Field Theory for the Hopfield Model

## 2.1 Introduction

The formal neurons and their dynamics defined as

$$S_i(t+1) = \theta\left(\sum_j J_{ij} S_j(t) - u_i\right), \tag{2.1}$$

introduced by McCulloch and Pitts[76] seemingly without a physical perspective at the start, can be reinterpreted as a globally coupled Ising system. It exactly corresponds to the zero-temperature limit of Monte Carlo dynamics[77, 18]. $J_{ij}$ is an exchange interaction and $-u_i$ an external field on spin $i$. The fact that both positive and negative $J_{ij}$ may occur even hints at a possible analogy with spin glasses, with ferromagnetic and antiferromagnetic interactions.

Furthermore, the analogy with spin systems can be extended to nonzero temperature ($T \neq 0$). This also has a biological basis in real neural systems: the synaptic transmission is a "noisy" process and the potential on the postsynaptic membrane is not determined precisely by the values $h_i \equiv \sum_j J_{ij} S_j - u_i$ but may fluctuate. We can therefore introduce a stochastic dynamics in which $S_i(t+1) = 1$ with probability

$$P_+(h_i) = \frac{1}{2}\left[1 + \tanh(\beta h_i)\right], \tag{2.2}$$

where $\beta$ ($\equiv 1/T$) denotes a "pseudo inverse temperature". Note that $T$ is not a physical temperature and nothing but a parameter which controls the steepness of the sigmoid $P_+(h_i)$. Any sigmoidal function of $h_i$ with limiting values of 0 and 1 at $-\infty$ and $\infty$ will lead to similar behavior[1], but this choice is particularly convenient because it corresponds exactly to finite-$T$ Glauber dynamics[49]. For symmetric $J_{ij}$, then, the system is guaranteed to obey equilibrium statistical mechanics: it has a stationary distribution. In the context of artificial computing networks it can actually be advantageous to introduce this kind of noise, as we will see below.

Next, in the context of our spin system, the Hebb hypothesis mentioned in the previous chapter can be stated as follows: the connection between two cells is strengthened when the firing of one cell succeeds in causing the other to fire, while if it fails to do so, the synapse is weakened. This idea can be represented as a symmetric *correlation matrix* the so-called Hebb rule :

$$\frac{\partial J_{ij}}{\partial t} \propto S_i(t)S_j(t). \tag{2.3}$$

While the real details have not been sorted out yet, there is experimental evidence for Hebb's hypothesis, and most neuroscientists seem to accept it as the only plausible working hypothesis, namely, (long-term) memory is stored in such a way in the synaptic connection strengths[2].

Here we note that Hebb rule (2.3) introduces symmetric interactions $J_{ij} = J_{ji}$ insidiously, which may be suitable and convenient for the formalism of statistical mechanics, but is biologically not very realistic. Because the channels of communication in real neural systems are unidirectional, there is no reason that neuron $i$ will be affected by neuron $j$ in the same way as neuron $j$ by neuron $i$. In fact, the existence of a connection from neuron $i$ to neuron $j$ does not imply the existence of a connection in the opposite direction. Thus, the assumption of synaptic symmetry is a step backward from the point of view of biological plausibility and was frequently criticized by neurobiologists, casting doubt on the entire approach. In retrospect, the Hopfield model turned out to be profitable both for statistical physics and for artificial neural network theory. It has also provided us with an intuitive perspective of a neural network's

[1] A new technique of analysis[83, 46, 91] for non-monotonic functions of $h_i$ has revealed much better memory retrieval circumventing spurious states.

[2] Recently several new paradigms for information coding in the cortex were proposed[72, 50, 34, 107, 2, 123] which suggests the so-called *first synapse* or *dynamical cell assembly*.



behavior (e.g., the critical storage capacity), which is important in applications and desired in engineering.

Although there are, in fact, several studies for models with asymmetric interactions[55, 56, 102, 103, 104, 43], I will follow what has turned out to be a rather successful strategy, and concentrate on the symmetric case.

## 2.2 The Hopfield Model

Partially following the review by Gutfreund and Toulouse[52], let us list several interesting features of the Hopfield model from the point of view of statistical mechanics.

- It is another solvable and rich model of a random long-range system.
- It spans a whole range of intermediate models of behavior, between the infinite range Ising ferromagnet where $p = 1$ (the Mattis model [75]) and the SK spin glass model[106] where $\sqrt{\alpha} \to \infty$ (the "SK limit" of the Hopfield model[120], see Chapter 3).
- It essentially has a mixed phase where the ferromagnetic retrieval (FMR) state and the SG state coexist. The FMR state refers to a state which has a macroscopic overlap with one of the memory patterns. Furthermore, in more detail, the Almeida-Thouless (AT) transition (the RS-RSB transition) occurs on different lines for the FMR phase and the SG phase, respectively.
- The existence of macroscopic free-energy valleys which are not ground states of the system, at least between $T_C$ the real first phase transition line and $T_M$ the transition line on which the metastable states for the FMR solution disappear, is a novel property of this model. (See Fig. 2.3)

The major contribution of the Hopfield model to neural network theory was to open a large galley of concepts, techniques and analogies, and to direct the effort in a way which avoided the initial obstacles. The analysis of the model leads to results and insights which go beyond the constraint of synaptic symmetry. It turned out to be a useful starting point for a variety of modifications[20, 1, 11, 12, 110] which removed some of the constraints and drawbacks of the original formulation of the model.

The Hopfield model is inherently connected with a self-coupled network architecture. It is worth mentioning that a layered-network model, where energy can be defined properly, can be analyzed by methods similar to the ones of statistical mechanics[17, 101, 126]. It turns out that learning, i.e., the dynamics in the space of interaction, can have feature similar to the ones seen in neurodynamics.

In the next section, I will give an outline of the formal mean field theory of the Hopfield model by Amit, Gutfreund and Sompolinsky (AGS)[8, 9]. In particular I will concentrate on the nontrivial case where the number of learning patterns is in the order of the system size $N$.

## 2.3 The AGS Theory

The Hamiltonian for the Hopfield model is given by,

$$H = -\frac{1}{2}\sum_{i,j=1}^{N} J_{ij}S_iS_j - \sum_{\nu=1}^{s} h^\nu \sum_{i=1}^{N} \xi_i^\nu S_i \qquad (S_i = \pm 1), \tag{2.4}$$

where $p$ is a number of random learning patterns and $h^\nu$ is a field conjugate to one of a finite numbers ($s << p$) of "condensed patterns"$\{\xi_i^\nu\}$. $J_{ij}$ is an interaction constructed from $p$ random patterns ($\xi_i^\mu = \pm 1$, $\mu = 1, \ldots, p$) as

$$J_{ij} \equiv \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu, \qquad J_{ii} = 0, \tag{2.5}$$

which can be obtained by Hebb rule if $p$ random patterns are given as firing patterns of $S_i(t) = \xi_i^\mu$ in eq. (2.3) with the same probability of firing for a sufficiently long time (even though a proper normalization may be needed to avoid divergence). We treat the $J_{ij}$'s as *frozen* or *quenched*, so the influence of the recall process on the memory is ignored. In other words, really in the same way as a bond randomness was introduced in spin glasses where such a randomness is expected not to vary significantly in short time scales of a quenched process of spins, we assume that the synaptic connections change much more slowly than the states of the neurons. That is, the Hopfield model treats only the dynamical aspects of neurons, not the learning process at all. Rather, it is concerned with a system in which the learning has been completed and the synaptic connections $J_{ij}$ are given in (2.5).



Now, we are interested in the average

$$\langle\langle \ln Z \rangle\rangle \equiv \int dJ_{ij} P(\{J_{ij}\}) \ln [\mathrm{Tr}_S \exp(-\beta H)], \tag{2.6}$$

where the symbol $\mathrm{Tr}_S$ refers to the trace taken over $N$ Ising spins as

$$\mathrm{Tr}_S E(\{S_i\}) \equiv \prod_{i=1}^{N} \sum_{S_i=\pm 1} E(\{S_i\}). \tag{2.7}$$

The average $\langle\langle \ln Z \rangle\rangle$ taken over the distribution $P(\{J_{ij}\})$ of all random binary patterns $\{\xi_i^\mu\}$ gives the free energy averaged over *quenched* disorder. By estimating the average (2.6) properly, one can obtain the average quantities we want to know. Unfortunately it is very hard to calculate this average directly, and it is not identical to $\log \langle\langle Z \rangle\rangle$, which corresponds to *annealing* and could be obtained much easier. To get meaningful results we must average the relevant quantity, which is $\ln Z$, not $Z$.

Fortunately there is a technique, called the replica method, that enables us to calculate the average of $\ln Z$. Since the average of a power of $Z$ is more easily obtained, we can use the limit

$$\ln Z = \lim_{n\to 0} \frac{Z^n - 1}{n}, \tag{2.8}$$

which gives the free energy per spin averaged over the quenched patterns in the form

$$f = \lim_{n\to 0} \lim_{N\to\infty} \left(\frac{-1}{\beta N}\right) \frac{\langle\langle Z^n \rangle\rangle - 1}{n}. \tag{2.9}$$

Therefore a central discussion is estimation of the average of $n$-replicated partition function $Z^n$ and this is the basic idea of the replica method. The formal derivation and limiting process of above free energy density are left to Appendix A. Let us go on to the essential discussion.

AGS introduced the free energy of the Hamiltonian(2.4) as follows,

$$\begin{aligned} f_n &= \frac{\alpha}{2} + \frac{1}{2n}\sum_{\mu=1}^{s}\sum_{a=1}^{n}(m_a^\mu)^2 + \frac{\alpha}{2\beta n}\mathrm{Tr}\ln[(1-\beta)\mathbf{I} - \beta\mathbf{Q}] \\ &\quad + \frac{\alpha\beta}{2n}\sum_{a\neq b}^{n} r_{ab}q_{ab} - \frac{1}{n\beta}\langle\langle \ln Z_0 \rangle\rangle_{\{\xi^\nu\}} \qquad (\text{as} \quad n \to 0), \end{aligned} \tag{2.10}$$

where

$$Z_0 = \mathrm{Tr}_{\{n\}} \exp\left(\frac{\alpha\beta^2}{2}\sum_{a\neq b} r_{ab}S^aS^b + \beta\sum_a z_aS^a\right) \tag{2.11}$$

$$z_a = \sum_\nu (m_a^\nu + h^\nu)\xi^\nu. \tag{2.12}$$

Here we note that the average

$$\langle\langle O\{\xi^\nu\}\rangle\rangle_{\{\xi^\nu\}} \equiv \frac{1}{2^s}\sum_{\nu=1}^{s}\sum_{\xi^\nu=\pm 1} O\{\xi^\nu\} \tag{2.13}$$

denotes the random average for ($s \ll p$) "condensed patterns". $\mathbf{I}$ is a unit matrix with $n \times n$ elements while $\mathbf{Q}$ is a replica matrix which gives the order parameter $q_{ab}$, where $n$ is a total number of replicas, and $a$ (or $b$) denotes an index for the $a$-th replica. By $\mathrm{Tr}_{\{n\}}$ we denote explicitly that the trace is taken over $n$-replicated binomial spins. The parameters in (2.10) are defined as follows,

$$m_a^\nu = \langle\langle \frac{1}{N}\sum_i \xi_i^\nu \langle S_i^a\rangle\rangle\rangle, \qquad (\nu = 1 \ldots s) \tag{2.14}$$

$$q_{ab} = \langle\langle \frac{1}{N}\sum_i \langle S_i^a\rangle\langle S_i^b\rangle\rangle\rangle, \tag{2.15}$$

$$r_{ab} = \frac{1}{\alpha}\sum_{\mu>s}^{p}\langle\langle m_a^\mu m_b^\mu\rangle\rangle, \tag{2.16}$$

where $\langle\cdots\rangle$ denotes a thermal average, i.e., an average over that part of the space of network states which the dynamics allows for. The average can be regarded either as a time average or as an ensemble average. Here we note that we used a special normalization of $r_{ab}$. Since each of the $m_a^\mu$'s for $\mu > s$ (for *uncondensed* patterns) is $O(1/\sqrt{N})$, the sum is $O(p/N)$ and the coefficient makes it of order unity, even if $p$ increases linearly with $N$. By eqs. (2.14), (2.15) and (2.16), the macroscopic order parameters are given as

$$m^\nu = \lim_{n\to 0}\frac{1}{n}\sum_a m_a^\nu \tag{2.17}$$

$$q = \lim_{n\to 0}\frac{1}{n(n-1)}\sum_{a\neq b} q_{ab} \tag{2.18}$$

$$r = \lim_{n\to 0}\frac{1}{n(n-1)}\sum_{a\neq b} r_{ab}. \tag{2.19}$$

$m^\nu$ denotes the mean overlaps between the states of the network visited by the dynamics and the $\nu$-th memorized pattern. Retrieval is identified by a large time-averaged overlap with consecutive single states. $q$ corresponds to SG order parameter discriminating between SG freezing and paramagnetism. The order parameter $r$ describes the noise due to the *uncondensed* patterns. In other words, if the network is in a state with large (macroscopic) overlaps with a few of the memorized (*condensed*) patterns, the accumulation of the random overlaps with all the other patterns creates a significant amount of noise.

Here we stress that the parameter $\alpha \equiv p/N$ refers to a storage level or rate of memory loading, which changes the phase of the system.

With these three sets of parameters, the equations for the stable states of the network can be written in the limit when the system size $N$ becomes infinite. These are the mean field equations which are exact for a fully connected network.



### 2.3.1 Replica Symmetric Solutions

Most of AGS's discussion have been carried out within the replica symmetric (RS) theory, which assumes that each order parameter is symmetric under a permutation of replica indices as

$$m_a^\nu = m^\mu \tag{2.20}$$

$$q_{ab} = q, \qquad a \neq b \tag{2.21}$$

$$r_{ab} = r, \qquad a \neq b \tag{2.22}$$

By noting that the matrix $(1-\beta)\mathbf{I} - \beta\mathbf{Q}$, with the replica matrix $\mathbf{Q}$ given by (2.21), has a nondegenerate eigenvalue $1-\beta-(n-1)\beta q$ and $(n-1)$-fold degenerate eigenvalues $1-\beta(1-q)$, the free energy density can be represented using the above RS order parameters after limiting processes of $n \to 0$ properly as

$$\begin{aligned} f_{RS} &= \frac{\alpha}{2} + \frac{\alpha\beta r(1-q)}{2} + \frac{1}{2}\sum_\nu (m^\nu)^2 \\ &\quad + \frac{\alpha}{2\beta}\left[\ln(1-\beta(1-q)) - \frac{\beta q}{1-\beta(1-q)}\right] \\ &\quad - \frac{1}{\beta}\int dP_{(1)}(\tilde{z})\langle\langle \ln 2\cosh\beta\left[\sqrt{\alpha r}\tilde{z} + \sum_\nu (m^\nu + h^\nu)\xi^\nu\right]\rangle\rangle, \end{aligned} \tag{2.23}$$

where the essence of the derivation of the last term is the application of the Hubbard-Stratonovitch identity (A.2) to the quadratic term $(\sum_a S^a)^2$ which appears in (2.11). In eq. (2.23), we used simplified expressions for Gaussian integrals

$$\int dP_{(a)}(x) \to \int_{-\infty}^{\infty}\frac{dx}{\sqrt{2\pi a}}\exp\left(-\frac{x^2}{2a}\right) \tag{2.24}$$

Variation of $f_{RS}$ with respect to $m^\nu$, $q$ and $r$, leads to the equations for the stationary states. All the solutions of $f_{RS}$, including local minima, are stationary states of the dynamical process, with barriers of $O(N)$. The equations are

$$m^\nu = \left\langle\!\!\left\langle \tanh\beta\left[\sqrt{\alpha r}\tilde{z} + \sum_\nu (m^\nu + h^\nu)\xi^\nu\right]\right\rangle\!\!\right\rangle_{\tilde{z}} \tag{2.25}$$

$$q = \left\langle\!\!\left\langle \tanh^2\beta\left[\sqrt{\alpha r}\tilde{z} + \sum_\nu (m^\nu + h^\nu)\xi^\nu\right]\right\rangle\!\!\right\rangle_{\tilde{z}} \tag{2.26}$$

$$r = \frac{q}{[1-\beta(1-q)]^2} \tag{2.27}$$

where the average $\langle\langle\cdots\rangle\rangle_{\tilde{z}}$ refers to the combined average over the $\xi^\nu$'s in eq. (2.13) and over the gaussian noise by the integral operator $\int dP_{(1)}(\tilde{z})$.

Here we note that at the saddle points the values of the parameters given by eqs. (2.25), (2.26) and (2.27) have just the physical meanings defined in eqs. (2.14), (2.15) and (2.16), respectively.

### 2.3.2 Solutions at $T = 0$

Now let us go on to the core of the AGS theory at $T = 0$. First we deal with the solution for the so called "ferromagnetic retrieval (FMR) phase ", corresponding to the case $s = 1$ and $m^1 \sim 1$ (we will represent it as $m$). We use the identity in the limit $T \to 0$ or $\beta \to \infty$:

$$\begin{aligned}\int dP(\tilde{z})_{(1)} \tanh\beta(A\tilde{z} + x) &\to \sqrt{\frac{2}{\pi}}\int_0^{x/A} d\tilde{z}\exp(-\tilde{z}^2/2) \\ &\equiv \mathrm{erf}(x/\sqrt{2}A)\end{aligned} \tag{2.28}$$

Appling this to eq. (2.25), with $h^1 = 0$, one can find that

$$m = \mathrm{erf}(m/\sqrt{2\alpha r}). \tag{2.29}$$

Next, the right hand side of eq. (2.26) has the limit unity as $\beta \to \infty$. But, as $\beta \to \infty$, the appearance of the term $C \equiv \beta(1-q)$ in the denominator in eq. (2.27) requires the $\beta \to \infty$ limit of this expression, which involves the term of $O(T)$ in $q$. Following similar calculations in eqs. (2.28), $C$ can be represented as

$$C = \lim_{\beta\to\infty} \beta(1-q) = \sqrt{\frac{2}{\pi\alpha r}}\exp\left(-\frac{m^2}{2\alpha r}\right). \tag{2.30}$$



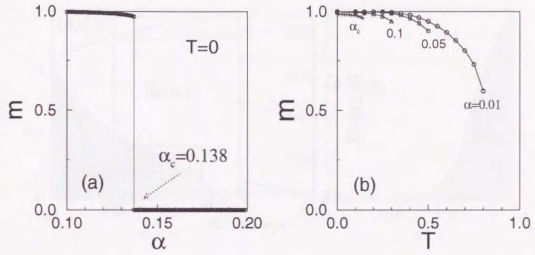



Figure 2.1: Pattern overlap $m$ (a) for $\alpha$'s at $T = 0$ (b) for $T$ at several values of $\alpha$ [9].

The value of $r$ in eq. (2.27), therefore, can be written as

$$r = (1-C)^{-2} \tag{2.31}$$

It is found that equations (2.29), (2.30) and (2.31) reduce to a single equation for the variable $y \equiv m/\sqrt{\alpha r}$, namely

$$y = \frac{\mathrm{erf}(y)}{\sqrt{2\alpha} + (2/\sqrt{\pi})\exp(-y^2)}. \tag{2.32}$$

This equation provides the dependence of the order parameter $m$ on $\alpha$ at $T = 0$, which is depicted in Fig. 2.1(a). The FMR solutions at $\alpha < \alpha_c \sim 0.138$ have a macroscopic projection on a given pattern. The solutions will have vanishingly small overlaps with each of the other patterns. Those random overlaps are of $O(1/\sqrt{N})$. Fig. 2.1(a) tells us that the transition $m \sim 1$ to $m = 0$ is abrupt: at $\alpha_c$ the overlaps drop suddenly.

Here we note that the derivative of $m^\nu$ with respect to $h^\nu$ gives the linear susceptibility $\chi_0$ as

$$\begin{aligned}\chi_0 &\equiv \left.\frac{\partial m^\nu}{\partial h^\nu}\right|_{h^\nu\to 0} \\ &= \int dP_{(1)}(\tilde{z})\beta\cosh^{-2}[\beta(\sqrt{\alpha r}\tilde{z} + \sum_\nu m^\nu)] \\ &= \beta(1-q).\end{aligned} \tag{2.33}$$

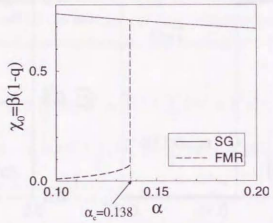



Figure 2.2: Linear susceptibility $\chi_0 = \beta(1-q)$ for SG and FMR phase.

Thus $C$ turns out to be nothing but $\chi_0$. Fig. 2.2 shows the dependence of $\chi_0$ on $\alpha$ for both the FMR and the SG phase. The transition between SG and FMR at $\alpha_c$ can be seen as a bifurcation of $\chi_0$. The line for the SG phase is given by $1/(1+\sqrt{\alpha})$[9], which is independent of the temperature.

### 2.3.3 Solutions at finite $T$

If one focuses on eqs. (2.25) and (2.26) at $s = 1$ and $h^1 = 0$, then one can derive the reduced equations in the similar way as in the case $T = 0$. They are given by

$$m = \int dP_{(1)}(\tilde{z}) \tanh\left[\beta(\sqrt{\alpha r}\tilde{z} + m)\right] \tag{2.34}$$

$$q = \int dP_{(1)}(\tilde{z}) \tanh^2\left[\beta(\sqrt{\alpha r}\tilde{z} + m)\right], \tag{2.35}$$

where the expression of $r$ is the same as in eq. (2.27).

These equations are solved numerically and yield the phase diagram shown in Fig. 2.3(a). Above $T_g = 1 + \sqrt{\alpha}$, in the paramagnetic phase region (A), there are neither FMR ($m \neq 0, q \neq 0$) nor SG ($m = 0, q \neq 0$) solutions for any values of $\alpha$, allowing only for the solution $q = m = 0$. Below $T_g$, there is an SG phase characterized by the solution $q \simeq T_g - T > 0$ and $m = 0$ (region (B)). Below the line $T_M$ (shaded region (C)), the FMR solutions $q \neq 0$ and $m \simeq 1$ appear as locally stable states. In this region, the SG solution coexists



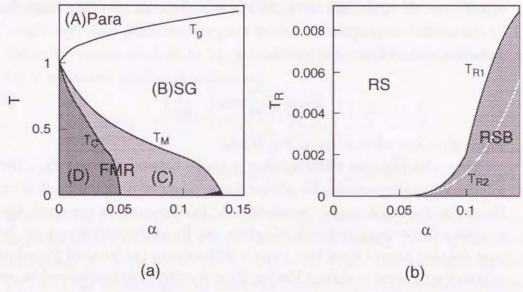



Figure 2.3: (a) The AGS phase diagram. (b) Generalized Almeida-Thouless line for the Hopfield model[9]

with the FMR one, but the free energy of the SG state is lower than the one of the FMR state. In the darkly shaded region (D) below the line $T_C$, the FMR free energy becomes the most stable state. The line $T_C$ therefore gives the first order transition except at the point $\alpha = 0, T = 1$ where the transition is second order[7]. The small black triangle near $\alpha_c$ is enlarged in Fig. 2.3(b) to show the RSB region for the FMR phase.

The dependence of the FMR solution $m$ on $T$ is also depicted in Fig. 2.1(b). The points where $m$ drastically falls down to zero gives the line $T_M$.

### 2.3.4 Generalized Almeida-Thouless (AT) line

Here we should note that the SG solution in the RS discussion above is unstable for $T < T_g$. The FMR solution, furthermore, becomes unstable to replica symmetry breaking (RSB) below a temperature $T_R$ shown as the shaded area in Fig. 2.3(b). The curves refer to the generalized Almeida-Thouless (AT) line[28, 9] originally discussed for the Sherrington-Kirkpatrick (SK) model[106] about the onset of the RSB instability where the sign of the "replicon" eigenvalue becomes negative. This condition leads, in terms of $m$ and $h$, to

$$\int dP_{(1)}(\tilde{z})\alpha\beta^2 \cosh^{-4}\beta(\sqrt{\alpha r}\tilde{z} + m + h) = [1 - \beta(1-q)]^2. \tag{2.36}$$

which gives the upper curve $T_{R1}$ in Fig. 2.3(b) for an arbitrary temperature by the numerical analysis. In the low temperature limit, $\cosh^{-4}(\cdots)$ goes to a delta-function and eq. (2.36) yields

$$T_{R2} \simeq \sqrt{\frac{8\alpha}{9\pi}} \exp\left(-\frac{1}{2\alpha}\right), \tag{2.37}$$

which gives the white curve in Fig. 2.3(b).

Thus, although the FMR solution is stable between $T_M$ and $T_{R1}$, the RS solutions become unstable for all the finite values of $\alpha$ in the limit $\beta \to \infty$. Hence the full RSB should be considered. Furthermore, in principle, the RS solution cannot characterize the SG phase and the multivalley structure. In the next chapter we will show how Parisi's RSB scheme can be used for arbitrary values of $\alpha$ in order to extract the variation of valley structure as $\alpha$ changes. In particular, the RSB discussion is important for understanding the equivalence between the SK model and the Hopfield model with large $\alpha$.

Here we note that in this thesis "the Hopfield model" denotes to the system with Ising variables in rather narrow sense. But, in fact, Hopfield and Tank[59] has proposed another version with analogue neurons with continuous value. This model has been analyzed within the replica symmetric discussions[105, 44, 45] in the similar way as the naive mean field model of SG[22, 115, 89, 90] and the result reveals that the metastable "spurious" states of analogue model is considerably suppressed compared with the Ising version. The phase diagram in the same way as AGS also has been given and the critical storage capacity gets larger than the one in the Ising version.



## Appendix A: Mean Field Theory for $\alpha > 0$

Using the replica method, let us estimate the following random average of $\xi$'s for $n$ replicated partition function as

$$\langle\langle Z^n \rangle\rangle = \left\langle\!\!\left\langle \mathrm{Tr}_S \exp\left(\frac{N\beta}{2}\sum_{\mu=1}^{p}\sum_{\rho=1}^{n}\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu S_i^\rho\right)^2 - \frac{1}{2}\beta pn + \beta\sum_{\nu=1}^{s}h^\nu\sum_{i=1}^{N}\sum_{\rho=1}^{n}\xi_i^\nu S_i^\rho\right)\right\rangle\!\!\right\rangle, \tag{A.1}$$

where by $\mathrm{Tr}_S$ we explicitly denote that the spin trace is taken over $n$-replicated binomial spins with system size $N$. The random average $\langle\langle\cdots\rangle\rangle$ is taken over $p$ patterns $\{\xi_i^\mu = \pm 1\}$ $(i = 1, \cdots, N;\ \mu = 1, \cdots, p)$. The quadratic terms in eq. (A.1) can be decoupled by means of the Hubbard-Stratonovitch identity

$$\exp(\lambda a^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx \exp(-\frac{x^2}{2} + \sqrt{2\lambda}ax) \tag{A.2}$$

and $\langle\langle Z^n \rangle\rangle$ becomes

$$\begin{aligned}\langle\langle Z^n \rangle\rangle = {} & e^{-\beta pn/2}\left\langle\!\!\left\langle \mathrm{Tr}_S \int\prod_{\mu\rho}\frac{dm_\rho^\mu}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\sum_{\mu=s+1}^{p}\sum_{\rho=1}^{n}(m_\rho^\mu)^2\right)\right.\right. \\ & \times\exp\left(\sqrt{N\beta}\sum_{\mu=s+1}^{p}\sum_{\rho=1}^{n}m_\rho^\mu\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu S_i^\rho\right) \\ & \left.\left.\times\exp\left(-\frac{1}{2}\sum_{\nu=1}^{s}\sum_{\rho=1}^{n}(m_\rho^\nu)^2 + \sum_{\nu=1}^{s}\sum_{\rho=1}^{n}(\sqrt{N\beta}\,m_\rho^\nu + \beta N h^\nu)\frac{1}{N}\sum_{i=1}^{N}\xi_i^\nu S_i^\rho\right)\right\rangle\!\!\right\rangle.\end{aligned} \tag{A.3}$$

Here let us concentrate on estimating the term ($\equiv L_1$) on the second line in eq. (A.3) which corresponds to the random average for "high" patterns $\{\xi_i^\mu\}$ $(\mu = s+1, \cdots, p)$ following as

$$\begin{aligned} L_1 &= \left\langle\!\!\left\langle \prod_{\mu=s+1}^{p}\prod_{i}^{N}\exp\left(\sqrt{\frac{\beta}{N}}\sum_{\rho=1}^{n}m_\rho^\mu S_i^\rho\xi_i^\mu\right)\right\rangle\!\!\right\rangle \\ &= \prod_{\mu=s+1}^{p}\prod_{i}^{N}\cosh\left(\sqrt{\frac{\beta}{N}}\sum_{\rho=1}^{n}m_\rho^\mu S_i^\rho\right) \\ &= \exp\left[\sum_{\mu=s+1}^{s}\sum_{i=1}^{N}\ln\cosh\left(\sqrt{\frac{\beta}{N}}\sum_{\rho=1}^{n}m_\rho^\mu S_i^\rho\right)\right], \end{aligned} \tag{A.4}$$

where we have used $\xi_i^\mu$ gets $\pm 1$ with probability 1/2. Since $\ln\cosh(x) \approx x^2/2$ for $x \ll 1$, $L_1$ reduces to

$$\begin{aligned} L_1 &\approx \exp\left[\sum_{\mu=s+1}^{p}\sum_{i=1}^{N}\frac{\beta}{2N}(\sum_{\rho=1}^{n} m_\rho^\mu S_i^\rho)^2\right] \\ &= \exp\left[\sum_{\mu=s+1}^{p}\sum_{i=1}^{N}\frac{\beta}{N}\sum_{\rho=1}^{n}\sum_{\sigma=1}^{n} m_\rho^\mu m_\sigma^\mu S_i^\rho S_i^\sigma\right] \end{aligned} \tag{A.5}$$

By substituting (A.5) into (A.3), we obtain

$$\begin{aligned} \langle\langle Z^n\rangle\rangle &= e^{-\beta p n/2}\,\mathrm{Tr}_S \int(\prod_{\mu\rho}\frac{dm_\rho^\mu}{\sqrt{2\pi}}) \\ &\times\exp\left[-\frac{1}{2}\sum_{\mu=s+1}^{p}\sum_{\rho=1}^{n}(m_\rho^\mu)^2+\frac{\beta}{2N}\sum_{\mu=s+1}^{p}\sum_{i=1}^{N}\sum_{\rho=1}^{n}\sum_{\sigma=1}^{n} m_\rho^\mu m_\sigma^\mu S_i^\rho S_i^\sigma\right] \\ &\times\left\langle\!\!\left\langle \exp\left[-\frac{1}{2}\sum_{\nu=1}^{s}\sum_{\rho=1}^{n}(m_\rho^\nu)^2+\sum_{\nu=1}^{s}\sum_{\rho=1}^{n}(\sqrt{N\beta}\; m_\rho^\nu+\beta N h^\nu)\frac{1}{N}\sum_{i=1}^{N}\xi_i^\nu S_i^\rho\right]\right\rangle\!\!\right\rangle_{\{\xi_i^\nu\}} \end{aligned} \tag{A.6}$$

where $\langle\langle\cdots\rangle\rangle_{\{\xi_i^\nu\}}$ denotes the random average for $s$ of "condensed patterns" $\{\xi_i^\nu\}$. By integrating of quadratic terms in the first $\exp(\cdots)$ ($\equiv L_2$) in (A.6) and one can find out that

$$\begin{aligned} \int(\prod_{\mu\rho}\frac{dm_\rho^\mu}{\sqrt{2\pi}})L_2 &= \int\prod_{\mu=s+1}^{p}\prod_{\rho=1}^{n}\frac{dm_\rho^\mu}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\sum_{\mu,\rho}(m_\rho^\mu)^2+\frac{\beta}{2N}\sum_{i\mu\rho\sigma} m_\rho^\mu m_\sigma^\mu S_i^\rho S_i^\sigma\right] \\ &= \int\prod_{\mu,\rho}\frac{dm_\rho^\mu}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\sum_{\mu\rho\sigma}K_{\rho\sigma}m_\rho^\mu m_\sigma^\mu\right] \\ &= [\det K]^{-(p-s)/2} \\ &\approx \exp\left[-\frac{p}{2}\mathrm{Tr}\ln K\right], \end{aligned} \tag{A.7}$$

where $K_{\rho\sigma} \equiv \delta_{\rho\sigma} - \beta/N\sum_i S_i^\rho S_i^\sigma$ and it is used that $p - s \approx p$ for large $p$ of order $N$. In the above caluculation, by $\mathrm{Tr}\ln K$, we explicitly denote that

$$\mathrm{Tr}\ln K = \sum_{\rho=1}^{n}\ln\tilde{\lambda}_\rho, \tag{A.8}$$

where $\tilde{\lambda}_\rho$ are the eigenvalues of $K$. Here we need a complicated procedure to diagonalize a matrix $K$; in fact it will be found out that $K$ is one of the replica



matrices and its diagonalization needs some concentration, in particular, if one addresses oneself to the replica symmetry breaking discussions.

By using a property of Dirac delta function, we obtain

$$\begin{aligned} \exp\left[-\frac{p}{2}\mathrm{Tr}\ln K\right] &= \int\prod_{\rho\neq\sigma}dq_{\rho\sigma}\exp\left[-\frac{p}{2}\mathrm{Tr}\ln((1-\beta)\mathbf{I}-\beta\mathbf{Q})\right] \\ &\times\prod_{\rho\neq\sigma}\delta(q_{\rho\sigma}-\frac{1}{N}\sum_{i=1}^{N}S_i^\rho S_i^\sigma), \end{aligned} \tag{A.9}$$

where $\mathbf{I}$ denotes a $n$ dimensional unit matrix. A new set of auxiliary variables $\mathbf{Q} \equiv [q_{\rho\sigma}]$ is introduced, which should satisfy

$$q_{\rho\sigma} = \begin{cases} N^{-1}\sum_{i=1}^{N} S_i^\rho S_i^\sigma & \text{for } \rho \neq \sigma\,; \\ 0 & \text{otherwise.} \end{cases} \tag{A.10}$$

Furthermore, by introducing a formal Fourier integral representation of the delta-function:

$$\delta(x-a) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{r(x-a)}dr \tag{A.11}$$

for $n(n-1)$ $\{q_{\rho\sigma}\}$, it is found as

$$\begin{aligned} \exp\left[-\frac{p}{2}\mathrm{Tr}\ln K\right] &\propto \prod_{\rho\neq\sigma}\int dr_{\rho\sigma}\int dq_{\rho\sigma}\exp\left[-\frac{p}{2}\mathrm{Tr}\ln\left((1-\beta)\mathbf{I}-\beta\mathbf{Q}\right)\right] \\ &\exp\left[-\frac{N\alpha\beta^2}{2}\sum_{\sigma\neq\rho}^{n} r_{\rho\sigma}\left(q_{\rho\sigma}-\frac{1}{N}\sum_{i=1}^{N}S_i^\rho S_i^\sigma\right)\right] \end{aligned} \tag{A.12}$$

where we have left out unimportant prefactors and scaled the $r_{\rho\sigma}$ by a factor of $N\alpha\beta^2$ for later convenience.

Thus, after rescaling the $m_\rho^\nu$ variables as $\sqrt{N\beta}\; m_\rho^\nu$ , by which variables $m_\rho^\nu$ become order 1, we can write our full expression (A.6) for $\langle\langle Z^n\rangle\rangle$ as

$$\begin{aligned} \langle\langle Z^n\rangle\rangle &\propto e^{-N\alpha\beta n}(N\beta)^{ns/2}\int\prod_{\nu=1}^{s}\prod_{\rho=1}^{n}\frac{dm_\rho^\nu}{\sqrt{2\pi}}\int\prod_{\rho\neq\sigma}dr_{\rho\sigma}\int\prod_{\rho\neq\sigma}q_{\rho\sigma} \\ &\times\exp N\left[-\frac{\alpha}{2}\mathrm{Tr}\ln\left((1-\beta)\mathbf{I}-\beta\mathbf{Q}\right)-\frac{\alpha\beta^2}{2}\sum_{\sigma\neq\rho}^{n} r_{\rho\sigma}q_{\rho\sigma}-\frac{\beta}{2}\sum_{\nu=1}^{s}\sum_{\rho=1}^{n}(m_\rho^\nu)^2\right] \\ &\times\left\langle\!\!\left\langle \mathrm{Tr}_S\exp\left[\frac{\alpha\beta^2}{2}\sum_{\sigma\neq\rho}^{n} r_{\rho\sigma}\sum_{i=1}^{N}S_i^\rho S_i^\sigma+\beta\sum_{\nu=1}^{s}\sum_{\rho=1}^{n}(m_\rho^\nu+h^\nu)\sum_{i=1}^{N}\xi_i^\nu S_i^\rho\right]\right\rangle\!\!\right\rangle_{\{\xi_i^\nu\}} \end{aligned} \tag{A.13}$$

The last line of (A.13) ($\equiv L_3$) can be represented as a function of random patterns $\{\xi_i^\nu\}$ and $n$-replicated binomial spins $\{S_i^\rho\}$, and can be further written generally as

$$\begin{aligned} L_3 &= \left\langle\!\!\left\langle \mathrm{Tr}_S \exp\left(\sum_{i=1}^{N} F\{\xi_i^\nu, S_i^\rho\}\right)\right\rangle\!\!\right\rangle_{\{\xi_i^\nu\}} \\ &= \left\langle\!\!\left\langle \prod_{i=1}^{N} \mathrm{Tr}_S \exp\left(F\{\xi_i^\nu, S^\rho\}\right)\right\rangle\!\!\right\rangle_{\{\xi_i^\nu\}} \\ &= \left\langle\!\!\left\langle \exp N\left(\frac{1}{N}\sum_{i=1}^{N} \ln \mathrm{Tr}_{S^\rho} \exp F\{\xi_i^\nu, S^\rho\}\right)\right\rangle\!\!\right\rangle_{\{\xi_i^\nu\}}, \end{aligned} \tag{A.14}$$

where $\mathrm{Tr}_{S^\rho}$ indicates trace over $n$ replicated binomial spins $\{S^\rho\ (= \pm 1)\}$. Here let us note that $s \ll N(\to \infty)$ guarantees $L_3$ to be self-averaging and it can be represented generally as

$$\frac{1}{N}\sum_{i=1}^{N} G(\xi_i^\nu) \Rightarrow \frac{1}{2^s}\sum_{\xi^\nu=\pm 1} G(\xi^\nu) \equiv \langle\!\langle G(\xi^\nu)\rangle\!\rangle_{\{\xi^\nu\}} \tag{A.15}$$

where we have succeeded in getting rid of the $i$ dependence of the expression as

$$L_3 = \exp\left(N\langle\!\langle \ln \mathrm{Tr}_S \exp F\{\xi^\nu, S^\rho\}\rangle\!\rangle_{\{\xi^\nu\}}\right). \tag{A.16}$$

Here we note that in the end we do not need the average $\langle\!\langle \cdots \rangle\!\rangle_{\{\xi_i^\nu\}}$ in (A.14), because the self-averaging of the $i$ sum in (A.16) has already carried out all of the pattern averaging.

Now we can write the whole expression for $\langle\!\langle Z^n \rangle\!\rangle$ as an integral of the exponential of something proportional to $N$:

$$\langle\!\langle Z^n \rangle\!\rangle \propto \int \left(\sqrt{\frac{N\beta}{2\pi}}\prod_{\nu\rho} dm_\rho^\nu\right) \int \left(\prod_{\rho\neq\sigma} dq_{\rho\sigma} dr_{\rho\sigma}\right) e^{-N\beta f\{m,q,r\}} \tag{A.17}$$

where

$$\begin{aligned} f\{m,q,r\} &= \frac{\alpha n}{2} + \frac{1}{2}\sum_{\nu=1}^{s}\sum_{\rho=1}^{n}(m_\rho^\nu)^2 + \frac{\alpha}{2\beta}\mathrm{Tr}\ln\left((1-\beta)\mathbf{I} - \beta\mathbf{Q}\right) \\ &\quad + \frac{\alpha\beta}{2}\sum_{\sigma\neq\sigma}^{n} r_{\rho\sigma}q_{\rho\sigma} - \frac{1}{\beta}\langle\!\langle \ln Z_0\rangle\!\rangle_{\{\xi^\nu\}} \end{aligned} \tag{A.18}$$

and

$$Z_0 = \mathrm{Tr}_{S^\rho} \exp\left(\frac{\alpha\beta^2}{2}\sum_{\sigma\neq\sigma}^{n} r_{\rho\sigma}S^\rho S^\sigma + \beta\sum_{\nu=1}^{s}\sum_{\rho=1}^{n}(m_\rho^\nu + h^\nu)\xi^\nu S^\rho\right). \tag{A.19}$$

# Chapter 3

# Multivalley Structures of the Rugged Energy Landscape with Replica Symmetry Breaking Discussion

## 3.1 Introduction

Recently the words "spin glasses (SG)" can be seen in diverse fields [81, 112]. They also have appeared in many papers for neural network models. If we consider the Hopfield model[58, 8] as a globally and not uniformly connected spin system, various SG like properties emerge. Its most significant characteristics are the existence of a large number of metastable states and multivalley structures in phase space which are also typical examples of "broken ergodicity"[93]. These properties were originally derived from an interpretation of the replica-symmetry-breaking (RSB) solution given by Parisi[94] for the Sherrington and Kirkpatrick (SK) model[106] of SG, and were thought to be rather specific to the SK model. More recent studies, however, have shown that this is not the case and that they are shared by other infinite-range SG models[90]. In the present work, we will show that the Hopfield model too can be regarded as such a model.

Very useful for obtaining some insight into the complexity of a valley structure is the basin of attraction which gives a quantitative measure for the "spread" of valley. Therefore, in Chpter 5, we will numerically study

relationships between remanent overlaps and initial overlaps in the Hopfield model using Monte Carlo simulations and finite size scaling at zero temperature ($T = 0$)[119]. Such relationships provides an intuitive understanding for the valley structure in *overlap space*. In this chapter, we consider the valleys in *spin configuration space* by obtaining the full RSB solution of the Hopfield model. It turns out that the formulation obtained is equivalent to the one for the original SK model in the limit $\sqrt{\alpha} \to \infty$, where $\alpha(\equiv p/N)$ is the rate of memory-loading, $p$ the number of random patterns for memories and $N$ the system size. In this way we have also proved conjectures by several authors[8, 52, 45] regarding the SG limit of the Hopfield model. Furthermore the full RSB solutions are numerically obtained for the SG phase.

In the previous chapter, the replica symmetric (RS) solutions for the Hopfield model was reviewed at a run. One of the main results is that the RS solutions are stable in almost all regions that belong to the FMR phase where the system behaves as an associative memory. We derived the mean field equations and critical storage capacity $\alpha_c \sim 0.138$ at $T = 0$ by extrapolating the RS solutions to the RSB region. The RS solutions for the SG states, however, are unstable in the entire region $T < T_g$ (the SG transition temperature) and the ones for the retrieval states become unstable at $T < T_R$, the generalized Almeida-Thouless line[28, 8] for the Hopfield model. To overcome this difficulty, Crisanti et al. have concentrated on estimating the critical storage capacity by considering the first step RSB scheme (which effectively means taking $K = 1$ in Appendix B) and they obtained the value $\alpha_c \sim 0.145$[27]. On the other hand, in this paper, the full RSB scheme will be employed to extract several pieces of information regarding the multivalley structure of the free energy of the Hopfield model, which interpolates between the Mattis model and the SK model. Furthermore, we will also reveal non-trivial asymptotic shifts of the system to the SK model, i.e. the "SK limit"[120], using the notion of the rugged free energy landscape.

In the next section, the weights of the valleys will be defined first. Next, the statistical mechanical interpretation of the RSB solution and the relationships between the order parameter function and the weight distribution of valleys will be given. Next, Parisi's RSB scheme[94, 95] is applied to the Hopfield model. It turns out that the obtained RSB formulation is equivalent to that for the original SK model as a limit. We numerically solve the resulting variational equations for the SG phase with rather large $\alpha$'s and some values of the



temperature. These solutions provide us with the weight distribution of the valleys.

## 3.2 Valleys

First let us define the weights of the valleys. When the phase space is divided into a number of valleys indexed by $k$, the weight of $k$-th valley $W_k$ is defined as the probability with which a randomly chosen initial state is in the $k$-th valley. If the system is at equilibrium, $W_k$ can be represented as

$$W_k = \exp(-\beta f_k)/\sum_j \exp(-\beta f_j) \qquad (\sum_k W_k = 1), \tag{3.1}$$

where $f_k$ refers to the free energy of the $k$-th valley and $\beta = 1/T$. In connection with $W_k$, Mézard et al[78] have defined a quantity

$$y = \langle \sum_k W_k^2 \rangle_J, \tag{3.2}$$

where $\langle \cdots \rangle_J$ denotes the sample average. The quantity $y$ roughly shows the number and the width of the valleys and gives the relative distribution of the weights of the valleys. For example, there is only one large valley for $y = 1$ while there are many valleys with small weights for $y \simeq 0$.

Second, let $m_{i,\rho} \equiv \langle S_i \rangle_\rho$ be the magnetic moment of the $i$th spin of the $\rho$th pure state. Overlaps of magnetization between two pure states are defined as follows,

$$q_{\rho\sigma} \equiv \frac{1}{N} \sum_{i=1}^{N} m_{i,\rho} m_{i,\sigma}. \tag{3.3}$$

It is noted that the self-overlap $q_{\rho\rho}(\equiv q^{EA})$ denotes the Edwards-Anderson order parameter. Using Boltzmann weights $P_\rho, P_\sigma$, the distribution of the overlap $q$ is represented as

$$P(q) \equiv \langle \sum_{\rho,\sigma} P_\rho P_\sigma \delta(q - q_{\rho\sigma}) \rangle_J. \tag{3.4}$$

The cumulative distribution function is defined as,

$$Y(q) \equiv \int_q^1 dq' P(q'). \tag{3.5}$$

$Y(q)$ also gives the density of pure states with an overlap larger than $q$. If we now consider the function

$$x(q) \equiv 1 - Y(q) = \int_{-1}^{q} dq' P(q'), \tag{3.6}$$

we see that its inverse $q(x)$ exactly coincides with the RSB solution of the SG order parameter, which is why it is called the physical interpretation of the RSB solution[94, 31]. The case $Y(q^{EA})$ refers to the self-overlap of a pure state by definition and is equivalent to $y$ in eq. (3.2) since $W_k = P_p$ (one pure state for one valley). Furthermore, $Y(q^{EA})$ is given by the length of the plateau of $q(x)$ whose functional form depends on $\alpha$. Hence, by obtaining the RSB solutions for various values of $\alpha$, the variation of $y$ can be estimated.

## 3.3 The RSB scheme for the Hopfield model

### 3.3.1 The full RSB scheme

Following the formalism by AGS and Parisi's recipe of taking a continuum limit of partitions of replica matrices (*Parisi gauge*), the free energy for the Hopfield model can be represented as the functional,

$$\begin{aligned} f_p &= \frac{\alpha}{2} + \frac{1}{2}\sum_{\nu}^{s}(m^\nu)^2 + \frac{\alpha\beta}{2}\left(r(1) - \int_0^1 r(x)q(x)dx\right) \\ &\quad - \int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi r(0)}}\exp\left(\frac{-z^2}{2r(0)}\right)\langle\langle g(0, h+z)\rangle\rangle_{\{\xi^\nu\}} \\ &\quad + \frac{\alpha}{2\beta}\left\{\frac{-\beta q(0)}{1-\chi(0)} + \ln[1-\chi(1)] + \int_0^1 \frac{dx}{x}\frac{\dot{\chi}}{1-\chi(x)}\right\}, \end{aligned} \tag{3.7}$$

which is maximized by the order parameter functions $q(x), r(x)$ and minimized by $m^\nu$. Here $q(x)$ and $r(x)$ correspond to the order parameters (2.15),(2.16) in the continuum limit, i.e., the SG order parameter function and the order parameter function describing the noise due to the *uncondensed* patterns, respectively. $m^\nu$ denotes the average overlap of the states and the $\nu$-th memorized pattern (2.14). The terms including $\chi(x)$ correspond to the third term in eq. (2.10) and they are obtained by taking the continuum limits of the eigenvalues of the replica matrix $\mathbf{Q}$ in eq. (2.10) since it has a recursive structure. The details of the limiting process are left to the Appendix B.



The average $\langle\langle\cdots\rangle\rangle_{\{\xi^\nu\}}$ has the same meaning as eq. (2.13). Furthermore, $g(x,z)(0 \le x \le 1, -\infty < z < \infty)$ is a solution of the following partial differential equation:

$$\dot{g} = -\frac{\dot{r}}{2}\left(g'' + \beta x g'^2\right) \qquad \text{(Parisi equation)} \left(\dot{A} \equiv \frac{\partial A}{\partial x} \quad A' \equiv \frac{\partial A}{\partial z}\right) \tag{3.8}$$

$$g(1,z) = \beta^{-1}\ln 2\cosh\sqrt{\alpha}\beta z. \tag{3.9}$$

The derivations of the terms containing $g(x,z)$ are similar to the case for the SK model[94, 90, 33] and are left to Appendix C for details. $\chi(x)$ and $h$ correspond to the local susceptibilities at scale $x$ and the effective field, respectively ($h^\nu$ denotes the external field, conjugate with the $\nu$-th condensed pattern) as

$$\chi(x) = \beta\left(1 - xq(x) - \int_x^1 q(\tilde{x})d\tilde{x}\right), \tag{3.10}$$

$$h = \frac{1}{\sqrt{\alpha}}\sum_{\nu=1}^{s}(m^\nu + h^\nu)\xi^\nu. \tag{3.11}$$

Equation (3.7) cannot directly be maximized numerically since $g(0,z)$ depends implicitly on $r(x)$ through (3.8). Therefore we apply the schemes by Sommers et al.[108] and Nemoto[85] to obtain the order parameter functions $q(x), r(x)$. They make $g(x,z)$ independent of $r(x)$ by introducing a Lagrange multiplier function $P(x,z)$. The new functional to be maximized is then defined as

$$\begin{aligned} f_s[r,q,g,P] &= f_p[r,q,g] + \int_{-\infty}^{\infty}dz P(1,z)\left\{g(1,z) - \frac{1}{\beta}\ln 2\cosh\sqrt{\alpha}\beta z\right\} \\ &\quad - \int_0^1 dx\int_{-\infty}^{\infty}dz P(x,z)\left\{\dot{g} + \frac{\dot{r}}{2}\left(g'' + \beta x g'^2\right)\right\}. \end{aligned} \tag{3.12}$$

The equations to be solved can be obtained by taking functional derivatives of (3.12) with respect to $q, r, g, P$ and $m^\nu$:

$$\dot{M} = -\frac{\dot{r}}{2}(M'' + 2\tilde{\beta}xMM') \qquad (M \equiv \frac{g'}{\sqrt{\alpha}}) \tag{3.13}$$

$$M(1,z) = \tanh\tilde{\beta}z \tag{3.14}$$

$$\dot{P} = \frac{\dot{r}}{2}(P'' - 2\tilde{\beta}x(PM)') \tag{3.15}$$

$$P(0,z) = \left\langle\left\langle\frac{1}{\sqrt{2\pi r(0)}}\exp\left(-\frac{(z-h)^2}{2r(0)}\right)\right\rangle\right\rangle_{\{\xi^\nu\}} \tag{3.16}$$

$$q(x) = \int_{-\infty}^{\infty} dz P(x,z) M(x,z)^2 \tag{3.17}$$

$$\chi(x) = \frac{\tilde{\beta}}{\sqrt{\alpha}} \left(1 - xq(x) - \int_x^1 q(\tilde{x}) d\tilde{x}\right) \tag{3.18}$$

$$r(x) = \frac{q(0)}{[1-\chi(0)]^2} + \int_0^x d\tilde{x} \frac{\dot{q}(\tilde{x})}{[1-\chi(\tilde{x})]^2} \tag{3.19}$$

$$m^\nu = \int_{-\infty}^{\infty} dz M(0,z) \left\langle\!\!\left\langle \xi^\nu \frac{1}{\sqrt{2\pi r(0)}} \exp\left(-\frac{(z-h)^2}{2r(0)}\right) \right\rangle\!\!\right\rangle_{\{\xi^\nu\}} \tag{3.20}$$

$$\tilde{\beta} \equiv \sqrt{\alpha}\beta. \tag{3.21}$$

In the above formalism, $P(1,z)$ gives the internal field distribution. Furthermore if we consider only one condensed pattern($s=1$) and $h^1 = 0$, we can estimate the sample average $\langle\!\langle \cdots \rangle\!\rangle$ in (3.16), (3.20), and obtain the following equations:

$$P(0,z) = \frac{1}{\sqrt{2\pi r(0)}} \exp\left(-\frac{(z - m/\sqrt{\alpha})^2}{2r(0)}\right), \tag{3.22}$$

$$m = \int_{-\infty}^{\infty} dz M(1,z) P(1,z), \tag{3.23}$$

where we have re-written $m^1$ as $m$ and have used $m = \int_{-\infty}^{\infty} dz M(x,z)P(x,z) = const.$ It is noted that (3.22) and (3.23) coincide with the corresponding equations for the SK model in a magnetic field. Consequently, in the same way as for the SK model, the field distribution can be corrected to a non-trivial function $P(x,z)$ from the Gaussian distribution by the RS solution.

Equations (3.13) $\sim$ (3.17) are similar to those for the SK model in which $\beta$ is scaled by $\sqrt{\alpha}$ while (3.18) and (3.19) are specific to the present analysis. When we take the limit $\sqrt{\alpha} \to \infty$ while keeping $\tilde{\beta}$ constant, we obtain that $\chi(x) \to 0$, $r(x) = q(x)$ for arbitrary $x$ and that $m = 0$. We thus find that eqs. (3.13) $\sim$ (3.23) formally coincide with those of the SK model and obtain the first result of this chapter. We have established the "SK limit" of the Hopfield model. In this way we furthermore proved the similar conjectures for the "SK limit" of the Hopfield model[8, 52, 45].

Here it is noted that the conditions $\dot{q} = \dot{r} = 0$ again yield the mean-field equations obtained from the RS discussion. Furthermore, the differentiation of (3.17) with respect to $x$ gives for $\dot{q} \neq 0$ the equation

$$\frac{1}{[1-\chi(x)]^2} \int_{-\infty}^{\infty} dz P(x,z) M'(x,z)^2 = 1. \tag{3.24}$$



This equation is equivalent to the condition for marginal stability in replica space[8, 120]. If we apply the RS solutions to eq. (3.24), one can easily obtain the equation for the generalized Almeida-Thouless line of the Hopfield model (i.e. eq. (2.36))[8] as,

$$\alpha\beta^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz e^{(-z^2/2)} \mathrm{sech}^4 \left[\beta\left(\sqrt{\alpha r} z + (m + h^1)\right)\right] = [1 - \beta(1-q)]^2, \tag{3.25}$$

where $q$ and $r$ correspond to the RS solutions of $q(x)$ and $r(x)$, respectively. $h^1$ denotes the external field, conjugate to the 1st condensed pattern.

### 3.3.2 Numerical analysis

We have solved (3.13)$\sim$(3.23) numerically for some parameters. Analytical results are known only near $T_g$, the SG transition temperature. Since we can interpet the linear terms of (3.13), (3.15) as diffusion equations, by introducing the Green function, the nonlinear partial differential equations (3.13), (3.16) can be transformed to the following integral equations:

$$G(x,z;x',z') = \frac{1}{\sqrt{2\pi(r(x') - r(x))}} \exp\left(-\frac{(z-z')^2}{2(r(x') - r(x))}\right), \tag{3.26}$$

$$M(x,z) = \int_{-\infty}^{\infty} dz' G(x,z;1,z') \tanh\tilde{\beta} z' + \int_x^1 dx' \tilde{\beta} \dot{r}(x') x' \times \int_{-\infty}^{\infty} dz' G(x,z;x',z') M(x',z') M'(x',z'), \tag{3.27}$$

$$P(x,z) = \frac{1}{\sqrt{2\pi r(x)}} \exp\left(-\frac{(y - m/\sqrt{\alpha})^2}{2r(x)}\right) - \int_0^x dx' \tilde{\beta} \dot{r}(x') x' \times \int_{-\infty}^{\infty} dz' G(x',z';x,z) \left(M(x',z') P(x',z')\right)'. \tag{3.28}$$

This enables us to obtain $M$, $P$, $m$, $r$, $q$ and $\chi$ by means of an iterative procedure in the order (3.27) $\to$ (3.22) $\to$ (3.23) $\to$ (3.28) $\to$ (3.17) $\to$ (3.18) $\to$ (3.19) $\to$ (3.27) $\cdots$. For numerical integration and differentiation we discretized the variables $x$ and $z$, dividing the intervals [0,1] and [-7.5,7.5] into 20 and 100 pieces, respectively, while we used a cubic spline for interpolation. We iterated the above procedure until the maximum variance of all the variables $P, M, r, q$ and $\chi$ became less than $10^{-6}$, which occurred typically in $700 - 900$ iteration (dependent on $\alpha$ and $\beta$).

Some typical order parameter functions are shown in Figs. 3.1, 3.2 and 3.3 for a fixed value of the temperature ($T(=1/\beta)=0.5$) and for $\alpha=0.5, 1.0$ and 2.0 (i.e., in the SG phase). We see that the larger $\alpha$, the smaller $r(x)$ and $\chi(x)$, but the larger $q(x)$. Consequently in the SK limit $q(x)$ and $r(x)$ are naively expected to be identical for arbitrary values of $x$. The internal field distribution $P(1,z)$ is represented in Fig. 3.4 for several $\alpha$'s and $T$'s. It is expected that the smaller $\alpha$, the nearer the system approaches the RS region at the temperature considered. Then we can see that a gaussian form will be recovered for smaller $\alpha$ or higher $T$ while a double-peaked form of $P(1,z)$ emerges for larger $\alpha$ or lower $T$. Figure 3.5 concerns the variation of $y$ for $\alpha$ at $T=0.5$. We estimate the values of $y$, the length of the plateau, by the point where $dq(x)/dx$ is less than $10^{-3}$. We see that $y$ can be scaled as $y \sim \alpha^{-\gamma}$ ($\gamma = 0.5 \pm 0.02$). Consequently, as $\alpha$ gets smaller the valleys with a large weight increase their relative frequency proportionally to $1/\sqrt{\alpha}$. Moreover, the larger $\alpha$, the more the number of relatively small basins increases. Here we would like to stress that an exponent of $\gamma \approx 0.5$ implies that for large $\alpha$, $y$ scales according to the SK limit ($\gamma = 0.5$) and not according to the SG limit ($\gamma = 1.0$) as thus far thought[8, 52, 45]. Finally the $\alpha$ dependence of $\chi(0)$ is given as circles in Fig. 3.6, which shows good agreement with the line $\chi(0) = 1/(1+\sqrt{\alpha})$ obtained analytically in ref. [8]. It also indicates that the accuracy of our numerical analysis is sufficient. Moreover, maximum errors of the compatibility condition (3.24) at $\alpha = 0.5, 1.0, 2.0$ and $T = 0.5$ were less than about 4%.

## 3.4 Summary

We have formulated the RSB solution for the Hopfield model, obtained the variational equations for the order parameter functions and found that they are identical with equations for the SK model as a limit (*the SK limit of the Hopfield model*). The order parameter functions are also numerically estimated for the SG phase and from their functional form the asymptotic dependence of the parameter $y$, characterizing the valley structure, on $\alpha$ is obtained.

Let us conclude this chapter by noting that the free energy (3.7) is invariant for gauge transformation of '$x$'. This enables us formalize a free energy with the so called *Sompolinsky gauge*[109, 29]. In the next chapter, using the new variational equations of the formalism, I will report important results for the



FMR and SG phase including an estimate for the critical storage capacity within the full RSB theory[121].

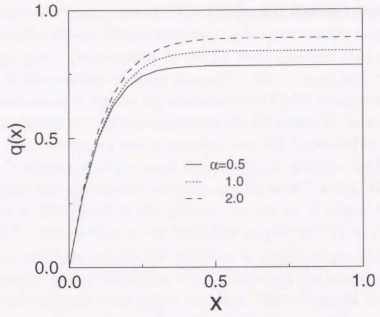



Figure 3.1: Order parameter function : $q(x)$

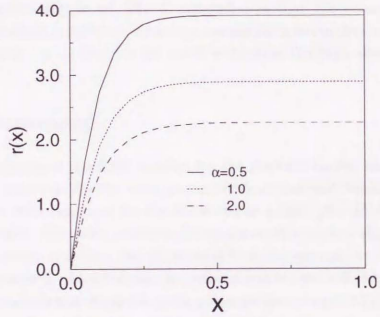



Figure 3.2: Order parameter function : $r(x)$



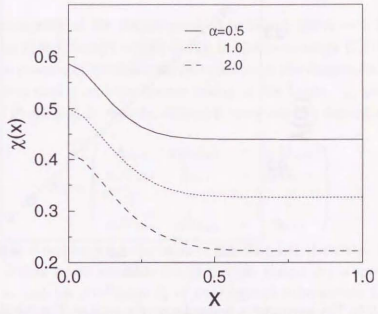



Figure 3.3: Order parameter function : $\chi(x)$

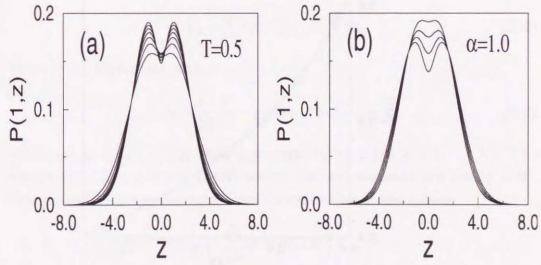



Figure 3.4: The internal field distribution $P(1,z)$ at (a)$T = 0.5$ for $\alpha = 0.5, 0.75, 1.0, 1.25, 1.5, 1.75$ and $2.0$ (bottom to top) and at (b)$T = 0.4, 0.5, 0.6$ and $0.7$ (bottom to top) for $\alpha = 1$.

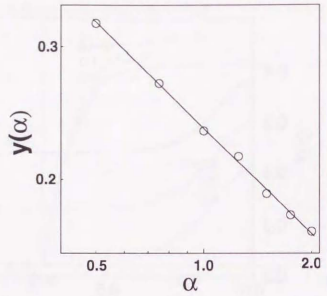



Figure 3.5: The asymptotic dependence of $y$ on $\alpha$ at $T = 0.5$ (log-log plotted).

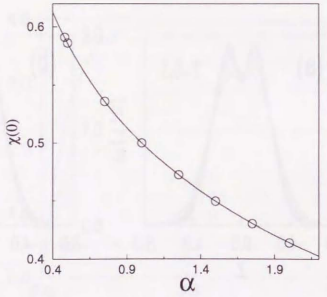



Figure 3.6: $\chi(0)(\circ)$ versus $\alpha$ for $T = 0.5$. The curve represents $\chi(0) = 1/T_g = 1/(1+\sqrt{\alpha})$.



## Appendix B: Diagonalization of Parisi's matrix

We summarize some of the details involved in taking the $n \to 0$ limit of the term $L \equiv (\alpha/(2\beta n))\mathrm{Tr}\ln[(1-\beta)\mathbf{I} - \beta\mathbf{Q}]$ in the free energy (2.10). Here we note that the essence of the following calculations is the diagonalization of the Parisi's replica matrix and the proper taking of the limits. In general, an $n$ dimensional Parisi matrix $\mathbf{A} = \mathbf{A}_0$ of level $K$ is recursively defined as the $K$-th iterate of

$$\mathbf{A}_k = \begin{pmatrix} \mathbf{A}_{k+1} & a_k\mathbf{U}_{k+1} & \cdots & a_k\mathbf{U}_{k+1} \\ a_k\mathbf{U}_{k+1} & \mathbf{A}_{k+1} & \cdots & a_k\mathbf{U}_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_k\mathbf{U}_{k+1} & a_k\mathbf{U}_{k+1} & \cdots & \mathbf{A}_{k+1} \end{pmatrix} \tag{B.1}$$

with $\mathbf{A}_{K+1} = \tilde{a}$ representing the value of the diagonal elements. $\mathbf{U}_k$ denotes the $p_k \times p_k$ matrix whose elements are all 1. Each matrix $\mathbf{A}_k$ is specified by its dimension $p_k$ and the coefficient $a_k$ of nondiagonal submatrices $\mathbf{U}_{k+1}$. Thus $\mathbf{A}_0$ is determined by the series of integers $n = p_0 > p_1 > \cdots > p_K > p_{K+1} = 1$, where naturally $p_k$ needs to divide $p_{k-1}$ in order for successive submatrices to fit correctly, and the series of coefficients $\{a_0, a_1, \ldots, a_K, a_{K+1} = \tilde{a}\}$.

By the recursive definition, the eigenvalues $\lambda$ of $\mathbf{A}$ are obtained[92] as

$$\lambda_{p_0} = \sum_{i=0}^{K}(p_i - p_{i+1})a_i + \tilde{a}, \tag{B.2}$$

which is nondegenerate, and as

$$\lambda_k = \sum_{i=k}^{K}(p_i - p_{i+1})a_i - p_k a_k + \tilde{a}, \tag{B.3}$$

which are $p_0(1/p_{k+1} - 1/p_k)$-fold degenerate for each $k(k = 0, 1, \cdots, K)$. The matrix $\mathbf{Q}$ in $L$ is a Parisi matrix whose diagonal elements are zero($\tilde{q} = 0$). Consequently, we use the eigenvalues for evaluating $L$ and obtain

$$\begin{aligned} L &= \frac{\alpha}{2\beta n}\ln\left[1 - \beta\left(1 + \sum_{i=0}^{K}(p_i - p_{i+1})q_i\right)\right] \\ &\quad + \frac{\alpha}{2\beta}\sum_{k=0}^{K}\left(\frac{1}{p_{k+1}} - \frac{1}{p_k}\right)\ln\left[1 - \beta\left(1 + \sum_{i=k}^{K}(p_i - p_{i+1})q_i - p_k q_k\right)\right]. \end{aligned} \tag{B.4}$$

Let us write the first and the second term by $\alpha/(2\beta)L_1, \alpha/(2\beta)L_2$, respectively. If we change the order of the summation of $k$(like in a partial integration), $L_2$

can be rewritten as

$$
\begin{aligned}
L_2 &= -\frac{1}{p_0}\ln\left[1-\beta(1+\sum_{i=0}^{K}(p_i-p_{i+1})q_i-p_0q_0)\right] \\
&\quad +\sum_{k=0}^{K-1}\frac{1}{p_k}\left\{\ln[1-\beta(1+\sum_{i=k-1}^{K}(p_i-p_{i+1})q_i-p_{k-1}q_{k-1})]\right. \\
&\quad \left. -\ln[1-\beta(1+\sum_{i=k}^{K}(p_i-p_{i+1})q_i-p_kq_k)]\right\} \\
&\quad +\ln[1-\beta(1-q_K)]. \qquad \text{(B.5)}
\end{aligned}
$$

Now, while denoting the terms in $L_2$ by $L_{21}, L_{22}, L_{23}$, we obtain after taking the continuum limit $K\to\infty$ and the limit $p_0(=n)\to 0$,

$$
\begin{aligned}
L_{21} &= -\frac{1}{n}\ln\left[1-\beta(1+\sum_{i=0}^{K}(p_i-p_{i+1})q_i)\right]-\frac{\beta q(0)}{1-\beta(1-\int_0^1 dx q(x))} \\
L_{22} &= \sum_{k=0}^{K-1}\left(\frac{1}{p_k}\right)\ln\left[1+\frac{\beta p_k(q_{k-1}-q_k)}{1-\beta(1+\sum_{i=k}^{K}(p_i-p_{i+1})q_i-p_kq_k)}\right] \\
&\simeq \sum_{k=0}^{K-1}\left(\frac{1}{p_k}\right)\frac{\beta p_k(q_{k-1}-q_k)}{1-\beta(1+\sum_{i=k}^{K}(p_i-p_{i+1})q_i-p_kq_k)} \\
&= -\int_0^1 dx\frac{\beta\dot{q}(x)}{1-\beta[1-\int_x^1 d\tilde{x}q(\tilde{x})-xq(x)]}, \qquad \text{(B.6)} \\
L_{23} &= \ln[1-\beta(1-q(1))], \qquad \text{(B.7)}
\end{aligned}
$$

where we set $q_0=q(0)$, $q_K=q(1)$, $p_i-p_{i+1}=dx$ and $q_{j-1}-q_j=\dot{q}(x)dx$. The first term of $L_{21}$ is canceled out by $L_1$ yielding $L'_{21}$. Finally, $L$ can then be represented as $\alpha/2\beta\,(L'_{21}+L_{22}+L_{23})$.



## Appendix C: Derivation of the Parisi equation

Here we briefly trace a derivation of Parisi equation (3.8) following schemes[94, 33, 29, 90] for taking a continuum limit of the partition function (2.11) defined in the replica space. It is convenient to write the partition function in the matrix form as

$$
Z_0(z)=\mathrm{Tr}_{\{n\}}\exp\left(\frac{1}{2}\boldsymbol{\sigma}^T\mathbf{R}\boldsymbol{\sigma}+z\boldsymbol{u}^T\boldsymbol{\sigma}\right) \qquad \text{(C.1)}
$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{u}$ are the $n\,(=p_0)$-dimensional column vectors whose elements are $\sigma_a$ and 1, respectvely. By applying the parametrization of the Parisi matrix $\mathbf{A}_0$ in Appendix B to $\mathbf{R}$ which is divided into $p_0/p_1\times p_0/p_1$ blocks, we can rewrite (C.1) as

$$
\begin{aligned}
Z_0(z) &\equiv \mathrm{Tr}_{\{n\}}\exp\left[\frac{1}{2}\sum_k^{p_0/p_1}\boldsymbol{\sigma}_{1,k}^T\mathbf{R}_1\boldsymbol{\sigma}_{1,k}+\frac{r_0}{2}\sum_{k\neq k'}^{p_0/p_1}\boldsymbol{\sigma}_{1,k}^T\mathbf{U}_1\boldsymbol{\sigma}_{1,k'}+z\sum_k^{p_0/p_1}\boldsymbol{u}_1^T\boldsymbol{\sigma}_{1,k}\right] \\
&= \mathrm{Tr}_{\{n\}}\exp\left[\frac{1}{2}\sum_k^{p_0/p_1}\boldsymbol{\sigma}_{1,k}^T(\mathbf{R}_1-r_0\mathbf{U}_1)\boldsymbol{\sigma}_{1,k}\right. \\
&\quad \left.+\frac{r_0}{2}(\sum_k^{p_0/p_1}\boldsymbol{u}_1^T\boldsymbol{\sigma}_{1,k})^2+z\sum_k^{p_0/p_1}\boldsymbol{u}_1^T\boldsymbol{\sigma}_{1,k}\right] \\
&= \int\frac{d\tilde{z}}{\sqrt{2\pi r_0}}\exp(-\frac{\tilde{z}^2}{2r_0}) \\
&\quad \times\mathrm{Tr}_{\{n\}}\exp\left[\frac{1}{2}\sum_k^{p_0/p_1}\boldsymbol{\sigma}_{1,k}^T(\mathbf{R}_1-r_0\mathbf{U}_1)\boldsymbol{\sigma}_{1,k}+(\tilde{z}+z)\sum_k^{p_0/p_1}\boldsymbol{u}_1^T\boldsymbol{\sigma}_{1,k}\right] \\
&= \int\frac{d\tilde{z}}{\sqrt{2\pi r_0}}\exp(-\frac{\tilde{z}^2}{2r_0}) \\
&\quad \times\left[\mathrm{Tr}_{\{p_1\}}\exp\left[\frac{1}{2}\boldsymbol{\sigma}_1^T(\mathbf{R}_1-r_0\mathbf{U}_1)\boldsymbol{\sigma}_1+(\tilde{z}+z)\boldsymbol{u}_1^T\boldsymbol{\sigma}_1\right]\right]^{p_0/p_1} \\
&\equiv \int\frac{d\tilde{z}}{\sqrt{2\pi r_0}}\exp(-\frac{\tilde{z}^2}{2r_0})\,[Z_1(\tilde{z}+z)]^{p_0/p_1} \qquad \text{(C.2)}
\end{aligned}
$$

where we have used the Hubbard-Stratonobitch identity (A.2) for the quadratic term $(\sum_k^{p_0/p_1}\boldsymbol{u}_1^T\boldsymbol{\sigma}_{1,k})^2$. In the above calculation, $\boldsymbol{\sigma}_{1,k}$ denotes the $k$-th $p_1$-dimensional subvector of $p_0$-dimensional vector $\boldsymbol{\sigma}_0\equiv\boldsymbol{\sigma}$ and there are $p_0/p_1$ such subvectors in the way as

$$
\boldsymbol{\sigma}_0=\left[\boldsymbol{\sigma}_{1,1},\boldsymbol{\sigma}_{1,2},\cdots,\boldsymbol{\sigma}_{1,p_0/p_1}\right]^T. \qquad \text{(C.3)}
$$

At this point in time, we can succeed in representing $Z_0(z)$ by $Z_1(\tilde{z}+z)$ , the one-step replica-symmetry-breaking (RSB) partition function, with $p_0/p_1 \times p_0/p_1$ submatrices $\mathbf{R}_1$ and $\mathbf{U}_1$, and $p_1$-dimensional subvector $\boldsymbol{\sigma}_1$ and $\boldsymbol{u}_1$.

Similarly one can construct an $l$-step RSB partition function generally by iterative equation as

$$\begin{aligned}
Z_l(z) &\equiv \mathrm{Tr}_{\{p_l\}} \exp\left[\frac{1}{2}\boldsymbol{\sigma}_l^T(\mathbf{R}_l - r_{l-1}\mathbf{U}_l)\boldsymbol{\sigma}_l + z\boldsymbol{u}_l^T\boldsymbol{\sigma}_l\right] \\
&= \mathrm{Tr}_{\{p_l\}} \exp\left[\frac{1}{2}\sum_k^{p_l/p_{l+1}} \boldsymbol{\sigma}_{l+1,k}^T(\mathbf{R}_{l+1} - r_{l-1}\mathbf{U}_{l+1})\boldsymbol{\sigma}_{l+1,k}\right. \\
&\quad \left. + \left(\frac{r_l - r_{l-1}}{2}\right)\sum_{k\neq k'}^{p_l/p_{l+1}} \boldsymbol{\sigma}_{l+1,k}^T\mathbf{U}_{l+1}\boldsymbol{\sigma}_{l+1,k'} + z\sum_k^{p_l/p_{l+1}} \boldsymbol{u}_{l+1}^T\boldsymbol{\sigma}_{l+1,k}\right] \\
&= \mathrm{Tr}_{\{p_l\}} \exp\left[\frac{1}{2}\sum_k^{p_l/p_{l+1}} \boldsymbol{\sigma}_{l+1,k}^T(\mathbf{R}_{l+1} - r_l\mathbf{U}_{l+1})\boldsymbol{\sigma}_{l+1,k}\right. \\
&\quad \left. + \left(\frac{r_l - r_{l-1}}{2}\right)(\sum_k^{p_l/p_{l+1}} \boldsymbol{u}_k^T\boldsymbol{\sigma}_{l+1,k})^2 + z\sum_k^{p_l/p_{l+1}} \boldsymbol{u}_{l+1}^T\boldsymbol{\sigma}_{l+1,k}\right] \\
&= \int \frac{d\tilde{z}}{\sqrt{2\pi(r_l - r_{l-1})}} \exp(-\frac{\tilde{z}^2}{2(r_l - r_{l-1})}) \\
&\quad \times \mathrm{Tr}_{\{p_l\}} \exp\left[\frac{1}{2}\sum_k^{p_l/p_{l+1}} \boldsymbol{\sigma}_{l+1,k}^T(\mathbf{R}_{l+1} - r_l\mathbf{U}_{l+1})\boldsymbol{\sigma}_{l+1,k} + (\tilde{z}+z)\sum_k^{p_l/p_{l+1}} \boldsymbol{u}_{l+1}^T\boldsymbol{\sigma}_{l+1,k}\right] \\
&= \int \frac{d\tilde{z}}{\sqrt{2\pi(r_l - r_{l-1})}} \exp(-\frac{\tilde{z}^2}{2(r_l - r_{l-1})}) \left[Z_{l+1}(\tilde{z}+z)\right]^{p_l/p_{l+1}} \qquad \text{(C.4)}
\end{aligned}$$

where $\mathbf{R}_{l+1}$ and $\mathbf{U}_{l+1}$ refer to the $p_{l+1}$-dimensional diagonal and nondiagonal submatrices of $p_l$-dimensional Parisi matrix $\mathbf{R}_l$, respectively. $\boldsymbol{\sigma}_{l+1,k}$ denotes the $k$-th $p_{l+1}$-dimensional subvector of $p_l$-dimensional column vector $\boldsymbol{\sigma}_l$ as

$$\boldsymbol{\sigma}_l = \left[\boldsymbol{\sigma}_{l+1,1}, \boldsymbol{\sigma}_{l+1,2}, \cdots, \boldsymbol{\sigma}_{l+1,p_l/p_{l+1}}\right]^T. \qquad \text{(C.5)}$$

The above procedure ends at $k = K+1$ as

$$\begin{aligned}
Z_{K+1}(z) &= Tr_{\{1\}} \exp\left[\frac{1}{2}(\mathbf{R}_{K+1} - q_K\mathbf{U}_{K+1})\boldsymbol{\sigma}_{K+1} + z\boldsymbol{\sigma}_{K+1}\right] \\
&= \exp(-\frac{q_K}{2})2\cosh z, \qquad \text{(C.6)}
\end{aligned}$$

where we have used $\mathbf{R}_{K+1} = 0$, $\mathbf{U}_{K+1} = 1$ and $\boldsymbol{\sigma}_{K+1} = 1$. To get the free energy per component Ising spin, we introduce



$$G_l(z) \equiv \frac{1}{p_l}\ln Z_l(z). \qquad \text{(C.7)}$$

Concerning this variable, we can get the following recursive formula:

$$\begin{aligned}
G_l(z) &= \frac{1}{p_l}\ln\int \frac{d\tilde{z}}{\sqrt{2\pi(r_l - r_{l-1})}} \exp(-\frac{\tilde{z}^2}{2(r_l - r_{l-1})}) \\
&\quad \times \left[\exp\left[p_{l+1}G_{l+1}(\tilde{z}+z)\right]\right]^{(p_l/p_{l+1})}, \qquad \text{(C.8)}
\end{aligned}$$

where

$$G_{K+1}(z) = -\frac{1}{2}q_K + \ln 2\cosh z \qquad \text{(C.9)}$$

Furthermore, by rescaling the $\tilde{z}/\sqrt{r_l - r_{l-1}}$ variable as $\tilde{z}$, we can write as

$$\begin{aligned}
G_l(z) &= \frac{1}{p_l}\ln\int \frac{d\tilde{z}}{\sqrt{2\pi}} \exp(-\frac{\tilde{z}^2}{2}) \exp\left[p_l G_{l+1}(z + \sqrt{r_l - r_{l-1}}\tilde{z})\right] \\
&= \frac{1}{p_l}\ln\int \frac{d\tilde{z}}{\sqrt{2\pi}} \exp(-\frac{\tilde{z}^2}{2}) \\
&\quad \times \exp\left[p_l(G_{l+1}(z) + \tilde{z}\sqrt{r_l - r_{l-1}}G'_{l+1} + \frac{1}{2}\tilde{z}^2(r_l - r_{l-1})G''_{l+1} + \cdots)\right] \\
&= G_{l+1}(z) + \frac{1}{p_l}\ln\int \frac{d\tilde{z}}{\sqrt{2\pi}} \exp(-\frac{\tilde{z}^2}{2}) \\
&\quad \times \left(1 + \tilde{z}p_l\sqrt{r_l - r_{l-1}}G'_{l+1} + \frac{\tilde{z}^2p_l^2(r_l - r_{l-1})}{2}G'^2_{l+1} + \frac{\tilde{z}^2p_l(r_l - r_{l-1})}{2}G''_{l+1} + \cdots\right) \\
&\approx G_{l+1}(z) + \frac{1}{p_l}\ln\left(1 + \frac{(r_l - r_{l+1})}{2}p_l(p_lG'^2_{l+1} + G''_{l+1})\right) \\
&= G_{l+1}(z) + \left(\frac{r_l - r_{l-1}}{2}\right)(p_lG'^2_{l+1} + G''_{l+1}), \qquad \text{(C.10)}
\end{aligned}$$

where we used $\sqrt{r_l - r_{l+1}} \ll 1$. Now we can take the continuum limit where $K \to \infty$, $G_l(z) \to G(x,z)$, $p_l \to x$ and $r_l - r_{l-1} \to -\dot{r}dx$ and by transforming the variable by $g(x,z) \equiv (G(x,z) + q(1)/2)/\beta$ we can get Parisi equation as

$$\begin{aligned}
\frac{dg}{dx} &= -\frac{1}{2}\frac{dr}{dx}\left[\frac{d^2g}{dz^2} + \beta x\left(\frac{dg}{dz}\right)^2\right] \\
g(1,z) &= \frac{1}{\beta}\ln 2\cosh z. \qquad \text{(C.11)}
\end{aligned}$$

For the Hopfield model, after rescaling the variables $r(x)$ and $z$ as

$$r(x) \rightarrow \alpha\beta^2 r(x), \tag{C.12}$$

$$z \rightarrow \sqrt{\alpha}\beta z, \tag{C.13}$$

we can get expressions relating to $g$ in (3.7) and (3.8).

# Chapter 4

# Replica Symmetry Breaking Discussion at Zero Temperature

## 4.1 Introdcution

Since the Hopfield model[58, 8] and its family of neural network models[10, 57] have been fully researched in the context of the spin glass (SG) theory[81, 112] and other new technics of analysis[46], studies on the Hopfield model in itself may seem to have a rather classical flavor nowadays. However there are still important open problems: low temperature behavior in the phase where the replica symmetric (RS) solutions are unstable and, in particular, the critical storage capacity ($\alpha_c$) in the low temperature limit ($\beta \equiv 1/T \rightarrow \infty$). This chapter discusses the RSB solution of the Hopfield model in order to consider such topics.

AGS have shown that the replica symmetric (RS) solution is stable in almost all regions which belong to the FMR phase, and by extrapolating to the RSB region they derived the mean-field equations at $T = 0$ and determined $\alpha_c = 0.138$. Crisanti, Amit and Gutfreund[27], moreover, have examined the one-step RSB solution and have corrected $\alpha_c$ from 0.138 to 0.144. However at $T = 0$ both the RS solution and the RSB solution become unstable when only a finite number of steps is used. Therefore the full (infinite steps) RSB solution should be considered, where $\alpha_c$ is expected to be corrected to a still larger value. Moreover, only the full RSB discussion provides us with an appropriate

estimation for the order parameter functions and the frozen field distribution which is expected to give some helpful information if one addresses oneself to dynamical profiles[6, 86].

In section 4.2, the RSB scheme by de Dominicis, Gabay and Orland (DGO) [30, 29, 90] is examined with the help of the so-called Sompolinsky gauge[109]. DGO's RSB scheme uses a different replica matrix (*the DGO matrix*) from the one employed in Parisi's RSB scheme[94, 95]. In particular, in subsection 4.2.1, the diagonalization of the DGO matrix is outlined because the corresponding terms in the free energy only appear in the Hopfield model, and not in the SK model[106] which was investigated in the RSB discussions of refs. [94, 30, 29]. Here we note that the resulting "gauge invariance" of the free energy functional plays an important role since it enebles us to numerically solve the variational equations avoiding a singularity originating at $T = 0$. Such a numerical solution at $T = 0$ cannot be achieved within the framework of Parisi's RSB scheme. In section 4.3, we numerically solve the variational equations at $T = 0$ for both the SG and the FMR phases for several values of $\alpha$.

## 4.2 The full RSB formulation

### 4.2.1 Diagonalization of the DGO matrix

Here let us concentrate on calculating the third and fourth terms in the free energy (2.10) with the help of DGO's RSB scheme. The essence of the following calculation is the diagonalization of the DGO matrix. In the scheme, the $n \times n$ dimensional DGO's replica matrix $\mathbf{Q}$ in (2.10) is recursively defined as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_0 - \mathbf{D}_0 & \mathbf{Q}_0 & \cdots & \mathbf{Q}_0 \\ \mathbf{Q}_0 & \mathbf{Q}_0 - \mathbf{D}_0 & \cdots & \mathbf{Q}_0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}_0 & \cdots & \cdots & \mathbf{Q}_0 - \mathbf{D}_0 \end{pmatrix}. \tag{4.1}$$

where $\mathbf{Q}_0$ and $\mathbf{D}_0$ are the $p_0 \times p_0$ Parisi matrices parameterized in the Appendix B. Here we note that the diagonal elements of $\mathbf{Q}_0$ and $\mathbf{D}_0$ are $\tilde{q} = q_K$ and $\tilde{d} = d_K + q_K$. To get a proper solution, one should take the limits $p_0 \gg p_1 \gg \cdots \gg p_K \to \infty$ (we will call this *the DGO limit*) and $K \to \infty$ before $n \to 0$.

First, let us consider the term $\Phi \equiv (1/n) \sum_{a \neq b}^{n} r_{ab} q_{ab}$ in eq. (2.10). We will denote the diagonal and off-diagonal submatrices of $\mathbf{R} = \{r_{ab}\}$ by $\mathbf{R}_0$ and $\mathbf{E}_0$



which are parameterized by $\{r_i\}$ and $\{e_i\}$, respectively, in the same way as $\mathbf{Q}_0$ and $\mathbf{D}_0$ in $\mathbf{Q}$. Consequently, $\Phi$ can be represented as

$$\begin{aligned}
\Phi &= \frac{1}{n}\mathrm{Tr}\mathbf{R}\mathbf{Q} = \frac{1}{p_0}\left[\mathrm{Tr}(\mathbf{R}_0 - \mathbf{E}_0)(\mathbf{Q}_0 - \mathbf{D}_0) + (n/p_0 - 1)\mathrm{Tr}\mathbf{R}_0\mathbf{Q}_0\right] \\
&= \frac{1}{p_0}\left[\frac{n}{p_0}\mathrm{Tr}\mathbf{Q}_0\mathbf{R}_0 - \mathrm{Tr}\mathbf{D}_0\mathbf{R}_0 - \mathrm{Tr}\mathbf{Q}_0\mathbf{E}_0 + \mathrm{Tr}\mathbf{D}_0\mathbf{E}_0\right] \\
&= \frac{n}{p_0}\left(\sum_{l=0}^{K}(q_l r_l - q_{l-1}r_{l-1})p_l - q_K r_K + \tilde{q}\,\tilde{r}\right) \\
&\quad - \left(\sum_{l=0}^{K}(d_l r_l - d_{l-1}r_{l-1})p_l - d_K r_K + \tilde{d}\,\tilde{r}\right) \\
&\quad - \left(\sum_{l=0}^{K}(q_l e_l - q_{l-1}e_{l-1})p_l - q_K e_K + \tilde{q}\,\tilde{e}\right) \\
&\quad + \sum_{l=0}^{K}(d_l e_l - d_{l-1}e_{l-1})p_l - d_K e_K + \tilde{d}\,\tilde{e} \\
&= -q_K r_K - \sum_{l=0}^{K}(\Delta_l^{(r)} q_l + \Delta_l^{(q)} r_l) \\
&= -q(1)r(1) - \frac{1}{\beta}\int_0^1 (\dot{\Delta}_r(x)q(x) + \dot{\Delta}_q(x)r(x))dx.
\end{aligned} \tag{4.2}$$

In the above calculation, by Tr we explicitly denote that the summation is taken over all the elements of the Parisi matrix as

$$\frac{1}{p_0}\mathrm{Tr}\mathbf{A}_0 = \frac{1}{p_0}\sum_{a,b}^{p_0} A_{0,ab} = \sum_{i=0}^{K}(a_i - a_{i-1})p_i - a_K + \tilde{a} \quad (a_{-1} \equiv 0), \tag{4.3}$$

where $A_{0,ab}$ denotes the $(a,b)$ element of the $p_0 \times p_0$ Parisi matrix. We furthermore used that $d_l - d_{l-1} = \Delta_l^{(q)}/p_l$ , $e_l - e_{l-1} = \Delta_l^{(r)}/p_l$, $\Delta_l^{(q)} = \dot{\Delta}_q(x)dx/\beta$ and $\Delta_l^{(r)} = \dot{\Delta}_r(x)dx/\beta$[29, 90] in the DGO limit and $K \to \infty$. $\Delta_r(x)$ and $\Delta_q(x)$ correspond to Sompolinsky's nonergodicity functions[109] determining the gauge of $x$[29, 108] together with $r(x)$ and $q(x)$ respectively as $x = -T\dot{\Delta}_q(x)/\dot{q}(x) = -T\dot{\Delta}_r(x)/\dot{r}(x)$ (Parisi's $x$). Here we note that these functions are scaled by $\beta$ since the Parisi equations (4.16) and (4.18) to appear later are scaled in the same way.

Second, let us consider the diagonalization of the DGO matrix as a preliminary of the calculation of the term $L \equiv 1/n\mathrm{Tr}\ln[(1-\beta)\mathbf{I}) - \beta\mathbf{Q}]$ in eq. (2.10). Since the Parisi matrices $\mathbf{Q}_0$ and $\mathbf{D}_0$ are commutate[95], one can write the

common eigenvector $\mathbf{u}_\lambda$ of $\mathbf{Q}_0$ and $\mathbf{D}_0$ as

$$\begin{pmatrix} \mathbf{Q}_0-\mathbf{D}_0 & \mathbf{Q}_0 & \cdots & \mathbf{Q}_0 \\ \mathbf{Q}_0 & \mathbf{Q}_0-\mathbf{D}_0 & \cdots & \mathbf{Q}_0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}_0 & \cdots & \cdots & \mathbf{Q}_0-\mathbf{D}_0 \end{pmatrix} \begin{pmatrix} a_1\mathbf{u}_\lambda \\ a_2\mathbf{u}_\lambda \\ \vdots \\ a_m\mathbf{u}_\lambda \end{pmatrix} = \epsilon_\lambda \begin{pmatrix} a_1\mathbf{u}_\lambda \\ a_2\mathbf{u}_\lambda \\ \vdots \\ a_m\mathbf{u}_\lambda \end{pmatrix}, \tag{4.4}$$

where $m = n/p_0$ and $\epsilon_\lambda$ denote the eigenvalue of $\mathbf{Q}$. Consequently, with respect to the coefficients $a_1, a_2, \cdots, a_m$, one can obtain that

$$\begin{pmatrix} \lambda_q-\lambda_d & \lambda_q & \cdots & \lambda_q \\ \lambda_q & \lambda_q-\lambda_d & \cdots & \lambda_q \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_q & \cdots & \cdots & \lambda_q-\lambda_d \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \epsilon_\lambda \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \tag{4.5}$$

where $\lambda_q$ and $\lambda_d$ denote the eigenvalues of $\mathbf{Q}_0$ and $\mathbf{D}_0$, respectively. Thus, the eigenvalue $\epsilon_\lambda$ of the DGO matrix is represented by

$$\epsilon_0 = m\lambda_q - \lambda_d \qquad \text{(nondegenerate)} \tag{4.6}$$

$$\epsilon_l = -\lambda_d \qquad (l = 1, 2, \cdots, m-1)(m-1\text{-fold degenerate}). \tag{4.7}$$

The eigenvalues of a Parisi matrix are known to be

$$\lambda_{p_0} = \sum_{i=0}^{K}(p_i - p_{i+1})a_i + \tilde{a}, \tag{4.8}$$

which is nondegenerate, and

$$\lambda_k = \sum_{i=k}^{K}(p_i - p_{i+1})a_i - p_k a_k + \tilde{a}, \tag{4.9}$$

which is $p_0(1/p_{k+1} - 1/p_k)$-fold degenerate for each $k(k = 0, 1, \cdots, K)$[120, 92]. By substituting (4.8) and (4.9) into (4.6) and (4.7), the eigenvalues of the DGO matrix therefore can obtained as

$$\epsilon_{0,p_0} = -\sum_{i=0}^{K}\Delta_i^{(q)} - q_K + \frac{n}{p_0}\left\{\sum_{j=0}^{K}(p_j - p_{j+1})q_j + q_K\right\}, \tag{4.10}$$

$$\epsilon_{0,k} = -\sum_{i=k+1}^{K}\Delta_i^{(q)} - q_K + \frac{n}{p_0}\left\{\sum_{j=k}^{K}(p_j - p_{j+1})q_j + q_K - p_k q_k\right\}, \tag{4.11}$$

$$\epsilon_{l,p_0} = -\sum_{i=0}^{K}\Delta_i^{(q)} - q_K, \tag{4.12}$$

$$\epsilon_{l,k} = -\sum_{i=k+1}^{K}\Delta_i^{(q)} - q_K, \tag{4.13}$$



where $k = 0, 1, \cdots, K$. We also used $\Delta_i^{(q)} \equiv p_i(d_i - d_{i-1})$[90]. The numbers of degeneracy of $\epsilon_{0,p_0}, \epsilon_{0,k}, \epsilon_{l,p_0}$ and $\epsilon_{l,k}$ are 1, $p_0(1/p_{k+1} - 1/p_k)$, $n/p_0 - 1$ and $(n - p_0)(1/p_{k+1} - 1/p_k)$, respectively. The total number of degeneracy is $n$ as can be easily checked. Now it is possible to obtain the term $L \equiv 1/n\mathrm{Tr}\ln[(1-\beta)\mathbf{I}) - \beta\mathbf{Q}]$ in the continuum limit by substituting the above eigenvalues in $L$ and taking the proper limits. The details of the limitation process are left to the Appendix D.

### 4.2.2 Free energy functional and the Parisi equations

After the calculations in the previous subsection and others for the Parisi equation[94], the free energy for the Hopfield model can be represented as a functional (*the Sompolinsky gauge*):

$$\begin{aligned} f_s = {} & \frac{\alpha}{2} + \frac{1}{2}\sum_{\nu}^{s}(m^\nu)^2 + \frac{\alpha\beta}{2}r(1)\,(1-q(1)) \\ & - \frac{\alpha}{2}\int_0^1 \left(\dot{\Delta}_r(x)q(x) + \dot{\Delta}_q(x)r(x)\right)dx \\ & - \int_{-\infty}^{\infty}\frac{dz}{\sqrt{2\pi r(0)}}\exp\left(\frac{-z^2}{2r(0)}\right)\langle\langle\varphi(0, z+h)\rangle\rangle_{\{\xi^\nu\}} \\ & + \frac{\alpha}{2\beta}\left\{\frac{-\beta q(0)}{1-\chi(0)} + \ln[1-\chi(1)] - \int_0^1 dx\frac{\beta\dot{q}(x)}{1-\chi(x)}\right\} \\ & + \int_{-\infty}^{\infty}dz\,P(1,z)\left\{\varphi(1,z) - \frac{1}{\beta}\ln 2\cosh\sqrt{\alpha}\beta z\right\} \\ & - \int_0^1 dx\int_{-\infty}^{\infty}dz\,P(x,z)\left\{\dot{\varphi} + \frac{\dot{r}(x)}{2}\varphi'' - \frac{\dot{\Delta}_r(x)}{2}\varphi'^2\right\}, \end{aligned} \tag{4.14}$$

which is maximized by the order parameter functions $r(x)$, $q(x)$, $\Delta_r(x)$, $\Delta_q(x)$, $\varphi(x,z)$, $P(x,z)$ and minimized by $m^\nu$. Dots and primes denote the derivatives with respect to $x$ and $z$, respectively. The average $\langle\langle\cdots\rangle\rangle_{\{\xi^\nu\}}$ has the same meaning as eq. (2.13). Here $q(x)$ and $r(x)$ correspond to the order parameters in the continuum limit, i.e., the SG order parameter function and the order parameter function describing the noise due to the *uncondensed* patterns, respectively. $m^\nu$ denotes the average overlap of the states and the $\nu$-th memorized pattern. The term involving $\chi(x)$ (defined in the Appendix) corresponds to $L \equiv 1/n\mathrm{Tr}\ln[(1-\beta)\mathbf{I}) - \beta\mathbf{Q}]$ in the continuum limit mentioned in the previous subsection. The last two terms, including the function $\varphi(x,z)$,

are introduced via a Lagrange multiplier function $P(x,z)$ in the scheme for numerical analysis by Sommers et al.[108], Nemoto[85], Takayama[116] and the author[120]. Those two terms should vanish when the maximum of the free energy is reached. The derivations are similar to the case for SG models[29, 90] and the details of derivation are left to Appendix E. The definition of the local field $h$ is given as

$$h = \frac{1}{\sqrt{\alpha}}\sum_{\nu=1}^{s}(m^\nu + h^\nu)\xi^\nu. \tag{4.15}$$

With regard to Sompolinsky's nonergodicity functions[109] $\Delta_r(x)$ and $\Delta_q(x)$, we note that the free energy (4.14) is 'gauge invariant' since eq. (4.14) and its variational equations still hold if $x$ is replaced by some *monotonic* function $u(x)$ with $u(0)=0$ and $u(1)=1$.

Now, the generalized 'Parisi equations' for the Hopfield model can be obtained by taking the functional derivatives of (4.14) with respect to the order parameter functions after the variable transformations $\sqrt{\alpha}\dot{\Delta}_r(x) \to \dot{\Delta}_r(x)$ and $\sqrt{\alpha}\dot{\Delta}_q(x) \to \dot{\Delta}_q(x)$ as,

$$\dot{M} = -\frac{\dot{r}(x)}{2}M'' + \dot{\Delta}_r(x)MM' \qquad (M \equiv \frac{\varphi'}{\sqrt{\alpha}}) \tag{4.16}$$

$$M(1,z) = \tanh\tilde{\beta}z \tag{4.17}$$

$$\dot{P} = \frac{\dot{r}(x)}{2}P'' + \dot{\Delta}_r(x)(PM)' \tag{4.18}$$

$$P(0,z) = \left\langle\!\!\left\langle \frac{1}{\sqrt{2\pi r(0)}}\exp\left(-\frac{(z-h)^2}{2r(0)}\right)\right\rangle\!\!\right\rangle_{\{\xi^\nu\}} \tag{4.19}$$

$$q(x) = \int_{-\infty}^{\infty} dz P(x,z)M(x,z)^2 \tag{4.20}$$

$$\tilde{\beta}(1-q(1)) + \Delta_q(x) - \Delta_q(1) = \int_{-\infty}^{\infty} d\tilde{z}P(x,\tilde{z})M'(x,\tilde{z}) \tag{4.21}$$

$$\chi(x) = \frac{1}{\sqrt{\alpha}}\left[\tilde{\beta}(1-q(1)) + \Delta_q(x) - \Delta_q(1)\right] \tag{4.22}$$

$$\dot{r}(x) = \frac{\dot{q}(x)}{[1-\chi(x)]^2} \tag{4.23}$$

$$\dot{\Delta}_r(x) = \frac{\dot{\Delta}_q(x)}{[1-\chi(x)]^2} \tag{4.24}$$

$$m^\nu = \int_{-\infty}^{\infty} d\tilde{z}M(0,\tilde{z})\left\langle\!\!\left\langle \xi^\nu \frac{1}{\sqrt{2\pi r(0)}}\exp\left(-\frac{(\tilde{z}-h)^2}{2r(0)}\right)\right\rangle\!\!\right\rangle_{\{\xi^\nu\}} \tag{4.25}$$



$$\tilde{\beta} \equiv \sqrt{\alpha}\beta. \tag{4.26}$$

In the above formalism $P(1,z)$ gives the internal field distribution. When we consider only one condensed pattern($s=1$) and $h^1=0$, we can estimate the sample average $\langle\!\langle\cdots\rangle\!\rangle$ in (4.19), (4.25), to obtain the following equations:

$$P(0,z) = \frac{1}{\sqrt{2\pi r(0)}}\exp\left(-\frac{(z-m/\sqrt{\alpha})^2}{2r(0)}\right), \tag{4.27}$$

$$m = \int_{-\infty}^{\infty} dz M(1,z)P(1,z), \tag{4.28}$$

where we have written $m^1$ as $m$ and have used $m = \int_{-\infty}^{\infty} dz M(x,z)P(x,z) = const.$ It is noted that (4.27) and (4.28) coincide with the corresponding equations for the SK model[106, 85] under a magnetic field. Consequently, in the same way as for the SK model, the field distribution is found to a non-trivial function $P(x,z)$ which is clearly different from the Gaussian distribution obtained by the RS solution.

Here we note that differentiation of the two functions $q(x)$ (eq. 4.20) and $\Delta_q(x)$ (eq. 4.21) with respect to $x$ gives for $\dot{\Delta}_q(x), \dot{q}(x) \neq 0$ the equation for the condition for marginal stability in replica space[8, 120]:

$$1 = \frac{1}{[1-\chi(x)]^2}\int_{-\infty}^{\infty} dz P(x,z)M'(x,z)^2. \tag{4.29}$$

By differentiating once more, one can obtain the equation

$$\begin{aligned} 0 &= \dot{q}(x)\int_{-\infty}^{\infty} dz P(x,z)M''(x,z)^2 \\ &\quad +2\dot{\Delta}_q(x)\left[\int_{-\infty}^{\infty} dz P(x,z)M'(x,z)^2 + [1-\chi(x)]^3/\sqrt{\alpha}\right] \end{aligned} \tag{4.30}$$

which implies that eqs. (4.20) and (4.21) are essentially equivalent; those two equations determine only the gauge relation between $q(x)$ and $\Delta_q(x)$ (e.g. the Parisi gauge is given by $\dot{\Delta}_q(x) = -\tilde{\beta}x\dot{q}(x)$, providing us with the variational equations with Parisi's RSB scheme[120]). This gauge relation also holds for the pair of $r(x)$ and $\dot{\Delta}_r(x)$ via eqs. (4.23) and (4.24).

In the limit $\beta \to \infty$, one can easily see that eqs. (4.17), (4.20) and (4.21) reduce to

$$M(1,z) = 2\theta(z) - 1 \tag{4.31}$$

$$M'(1,z) = 2\delta(z) \tag{4.32}$$

$$q(1) = \int_{-\infty}^{\infty} dz P(1,z) = 1 \tag{4.33}$$

Moreover, by substituting equation (4.33) into eqs. (4.21) and (4.22) and setting $x = 1$, one has

$$\sqrt{\alpha}\chi(1) = \tilde{\beta}\,[1 - q(1)] = P(1,0) = 0, \tag{4.34}$$

which is also supported by numerical studies of the SK model in an external field[16]. Thus it is found that the two functions $\chi(x)$ and $\Delta_q(x)$ are essentially equivalent

$$\Delta_q(x) = \sqrt{\alpha}\chi(x) = \int_{-\infty}^{\infty} dz P(x,z) M'(x,z), \tag{4.35}$$

where we set $\Delta_q(1) = 0$ according to definition[109].

## 4.3 Numerical analysis at $T = 0$

For several values of $\alpha$, we have solved (4.16)$\sim$(4.28) numerically both for the SG phase and the FMR phase. Since we can interpret the linear terms of (4.16) and (4.18) as diffusion equations, by introducing the Green function, the nonlinear partial differential equations (4.16), (4.18), (4.27) and (4.31) can be transformed to the following integral equations at $T = 0$:

$$G(x,z;\tilde{x},\tilde{z}) = \frac{1}{\sqrt{2\pi(r(\tilde{x}) - r(x))}} \exp\left(-\frac{(z-\tilde{z})^2}{2(r(\tilde{x}) - r(x))}\right) \tag{4.36}$$

$$M(x,z) = \mathrm{erf}\left(\frac{z}{\sqrt{2(r(1) - r(x))}}\right) - \int_x^1 d\tilde{x}\dot{\Delta}_r(\tilde{x}) \times \int_{-\infty}^{\infty} d\tilde{z} G(x,z;\tilde{x},\tilde{z}) M(\tilde{x},\tilde{z}) M'(\tilde{x},\tilde{z}) \tag{4.37}$$

$$P(x,z) = \frac{1}{\sqrt{2\pi r(x)}} \exp\left(-\frac{(y - m/\sqrt{\alpha})^2}{2r(x)}\right) + \int_0^x d\tilde{x}\dot{\Delta}_r(\tilde{x}) \times \int_{-\infty}^{\infty} d\tilde{z} G(\tilde{x},\tilde{z};x,z)\,(M(\tilde{x},\tilde{z})P(\tilde{x},\tilde{z}))' \tag{4.38}$$

where $\mathrm{erf}(x) \equiv 2/\pi \int_0^x \exp(-t^2)dt$ denotes the error function. This enables us to obtain $M$, $P$, $m$, $r$, $q$ and $\chi$ by means of an iterative procedure in the order (4.37) $\rightarrow$ ( (4.27) $\rightleftharpoons$ (4.28) ) $\rightarrow$ (4.38) $\rightarrow$ (4.20) $\rightarrow$ (4.23) $\rightarrow$ (4.37) $\cdots$. In the SG phase, $\Delta_q(x)$ ($\Delta_r(x), \chi(x)$) can be determined a priori by choosing a special gauge of Parisi's $x$. On the other hand, in the FMR phase, $\Delta_q(x)$ has to be determined through eq. (4.35) at each step of the iterative procedure. Details are explained in the following subsections for each phase.



For numerical integration and differentiation we have discretized the variable $x$, dividing the interval [0,1] into 40 $\sim$ 100 pieces. The variable $z$ is also discretized, dividing the interval [-$A$,$A$] into 100 pieces, where the value of $A$ is determined differently for $P(x,z)$ and $M(x,z)$ because $P(x,z)$ spreads over a wider range of $z$ than $M(x,z)$ which is expected to change its value drastically near $z = 0$ (i.e., very close to a step function, especially in the FMR phase), e.g. $A = 12.0$ and $A = 0.05$ for $P(x,z)$ and $M(x,z)$, respectively. Furthermore, we have used a cubic spline for interpolation. Quadratic functions and hyperbolic tangents are also used for extrapolations of $\ln(P(x,z))$ and $M(x,z)$, respectively, both defined on the entire range of $z$. We have carried out the above iterative procedure until the maximum variance of all the variables $P, M, r, q$ and $\chi$ is less than $10^{-6}$.

Once a set of solutions $(m, q, r, \chi, P, M)$ for a value of $\alpha$ is obtained, a new solution at an adiabatically shifted value of $\alpha$ is calculated (i.e., the previously obtained solutions are included as an initial condition). Thus we detect the vanishing point of the metastable states corresponding to the FMR phase and obtain $\alpha_c$.

### 4.3.1 SG solution

Let us take a closer look at the solutions in the SG phase. Here we note that $\Delta_q(0)$ in the SG phase depends only on $\alpha$ as $\Delta_q(0) = \sqrt{\alpha}/(1 + \sqrt{\alpha})$[8], which is independent of the temperature. Therefore one can see that $\Delta_q(0) \rightarrow 1$ in the "SK limit ($\sqrt{\alpha} \rightarrow \infty$)[120]". Consequently we choose the special gauge

$$\Delta_q(x) = \sqrt{\alpha}(1 - x)/(1 + \sqrt{\alpha}). \tag{4.39}$$

For several values of $\alpha$ in the SG phase, $q(x)$ and $r(x)$ are shown in Figs. 4.1 and 4.2, respectively, where $q(0) = 0$ denotes $m = 0$. The frozen field distribution $P(1,z)$ is depicted in Fig. 4.3 in which we see the symmetry and the typical form of a double-peak.

### 4.3.2 FMR solution

In the FMR phase, since a dependency of $\chi(0)$ on $\alpha$ is not given explicitly like in the SG phase above mentioned, $\chi(0)$ has to be determined self-consistently during the numerical calculations. Therefore we apply the special gauge

$$\chi(x) = \chi(0)(1 - x), \tag{4.40}$$

where $\chi(0)$ is reset by eq. (4.35) during each step of the iterative procedure. The values for $\chi(0)$ are shown in fig. 4.6 together with the ones obtained by the RS discussion.

For several values of $\alpha$ in the FMR phase, $q(x)$ and $r(x)$ are shown in Figs. 4.4 and 4.5, respectively. We note that in Fig. 4.4(a), the $q(x)$ for $\alpha = 0.1, 0.11, 0.12$ are so close to unity that it is difficult to distinguish them. Therefore, at $\alpha < 0.13$ the RS solutions are almost recovered, although one can see that $q(0) \neq 1$ even at $\alpha = 0.1$. The frozen field distributions $P(1, z)$ for several values of $\alpha$ are given in Fig. 4.7 in which we can see the broken-symmetry. Their main peak is nearly Gaussian but they have a second peak as was also the case in Fig. 4.1.

The percentage of errors, $(1 - m)/2$, in the FMR phase is shown in Fig. 4.8 as a function of $\alpha$ at $T = 0$. For comparison, the results of the present calculations are plotted together with the predictions of the RS theory[8] and the 1-step RSB theory[27]. It clearly shows that $\alpha_c$ is somewhat higher than the values obtained by the RS (0.138) and 1-step RSB (0.144). We determined $\alpha_c$ in the following way. That is, at $\alpha = 0.153$ the RSB solution with finite $m$ was found, on the other hand, at $\alpha = 0.157$ and beyond, no RSB solution with finite $m$ was found. At $0.153 < \alpha < 0.157$, the calculation converges to an unexpected solution, therefore, it is concluded that $\alpha_c = 0.155 \pm 0.002$. These results are remarkable in the sense that RSB is directly observed, the order parameter functions and frozen field distributions are explicitly determined for $\alpha$'s.

## 4.4 Summary and Discussion

We have formulated the RSB solution of the Hopfield model with the Sompolinsky gauge at $T = 0$ and obtained the variational equations. Extensive numerical analyses were carried out, both for the SG and the FMR phases, in the most interesting region where the FMR phase disappears. The first result of this paper is that the storage capacity $\alpha_c$ at $T = 0$ is corrected to a value which is higher than the ones obtained by the RS and one step RSB discussions. Our result is the first self-consistent estimation of the critical storage capacity using the full RSB scheme. Another interesting aspect of this result is that it also indicates that RSB promotes the stability of the FMR solution against the



increase of the so-called "slow" or "stochastic synaptic" noise[10] originating in $\alpha$'s increase. The second major result of this paper is that the frozen field distribution (FFD) $P(1, z)$ is corrected to a non-Gaussian form for both the SG phase and the FMR phase. It is found that the FFD for the FMR phase is only slightly different from the Gaussian form obtained in the RS discussion. Therefore, as far as the FMR phase is concerned, the AGS theory is able to go beyond the RS approximation even at $T = 0$. This situation also explains why the dynamical evolution of $m$ in the FMR phase can approximately be described by only a few macroscopic variables[6, 86]. However, we note that the FFD for the SG phase cannot be fitted by any Gaussian form. Thus, it is quite natural that the description for the convergence to the SG phase fails if one tries to describe it in the same way as the dynamics which converges to the FMR phase. The convergence to the SG phase, if anything, can very effectively be described by assuming the phenomenological non-Gaussian form of the field distribution introduced in ref. [54].

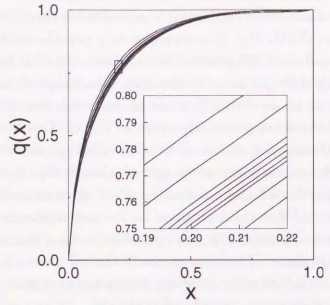



Figure 4.1: Order parameter function $q(x)$ in the SG phase for $\alpha = 0.1$, 0.12, 0.14, 0.145, 0.155, 0.16, 0.18 and 0.2 (top to bottom)

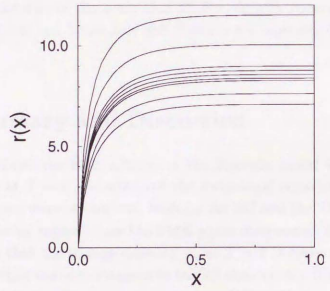



Figure 4.2: Order parameter function $r(x)$ in the SG phase for $\alpha = 0.1$, 0.12, 0.14, 0.145, 0.155, 0.16, 0.18 and 0.2 (top to bottom)



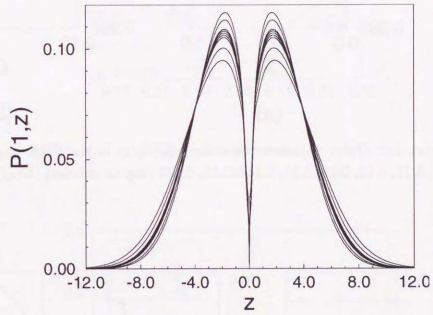



Figure 4.3: The frozen field distribution $P(1, z)$ at $\beta \to \infty$ in the SG phase for $\alpha = 0.1$, 0.12, 0.14, 0.145, 0.15, 0.155, 0.16, 0.18 and 0.2 (bottom to top).

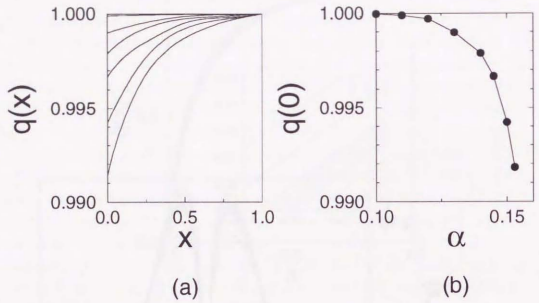



Figure 4.4: Order parameter function : (a) $q(x)$ in the FMR phase for $\alpha =$ 0.1, 0.11, 0.12, 0.13, 0.14, 0.145, 0.15, 0.153 (top to bottom) (b) q(0) versus $\alpha$

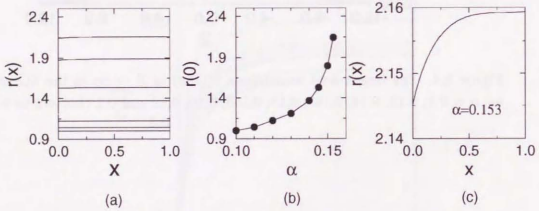



Figure 4.5: Order parameter function : (a) $r(x)$ in the FMR phase for $\alpha =$ 0.1, 0.11, 0.12, 0.13, 0.14, 0.145, 0.15, 0.153 (bottom to top) (b) r(0) versus $\alpha$ (c) $r(x)$ for $\alpha = 0.153$



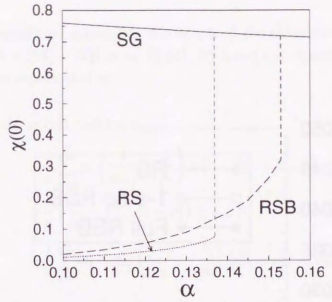



Figure 4.6: $\chi(0)$ vs. $\alpha$

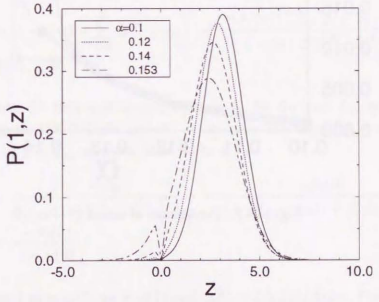



Figure 4.7: Frozen field distribution $P(1, z)$ for $\alpha = 0.1, 0.12, 0.14, 0.153$

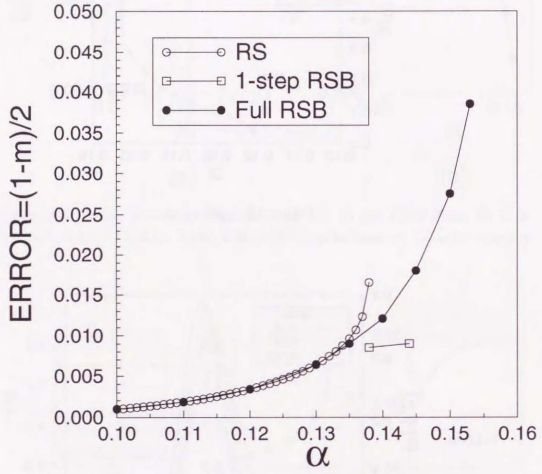



Figure 4.8: Percentage of errors $(1-m)/2$



## Appendix D: Diagonalization of the DGO matrix

In this Appendix we summarize the details of the calculation of the term $L \equiv 1/n \mathrm{Tr} \ln[(1-\beta)\mathbf{I}) - \beta\mathbf{Q}]$ in eq. (2.10). By using the eigenvalues (4.10) $\sim$ (4.13), $L$ can be represented as

$$\begin{aligned} L &= \frac{1}{n}\ln[1-\beta(1+\epsilon_{0,p_0})] \\ &+\frac{1}{n}\sum_{k=0}^{K} p_0\left(\frac{1}{p_{k+1}}-\frac{1}{p_k}\right)\ln[1-\beta(1+\epsilon_{0,k})] \\ &+\frac{1}{n}\left(\frac{n}{p_0}-1\right)\ln[1-\beta(1+\epsilon_{l,p_0})] \\ &+\frac{1}{n}\sum_{k=0}^{K}(n-p_0)\left(\frac{1}{p_{k+1}}-\frac{1}{p_k}\right)\ln[1-\beta(1+\epsilon_{l,k})]. \end{aligned} \tag{D.1}$$

Here we denote the four terms in $L$ by $L_1, L_2, L_3$ and $L_4$, respectively. First, by substituting eq. (4.10) into $L_1$, it is found that

$$\begin{aligned} L_1 &= \frac{1}{n}\ln\left[1-\beta\left(1-\sum_{i=0}^{K}\Delta_i^{(q)}-q_K\right)-\frac{n}{p_0}\beta\left\{\sum_{j=0}^{K}(p_j-p_{j+1})q_j+q_K\right\}\right] \\ &= \frac{1}{n}\ln\left[1-\beta\left(1-\sum_{i=0}^{K}\Delta_i^{(q)}-q_K\right)\right]+\frac{-(\beta/p_0)\{\sum_{j=0}^{K}(p_j-p_{j+1})q_j+q_K\}}{1-\beta(1-\sum_{k=0}^{K}\Delta_k^{(q)}-q_K)} \\ &+O(n). \end{aligned} \tag{D.2}$$

Let us denote the first and second term in $L_1$ by $L_{11}$ and $L_{12}$ respectively. After taking the DGO limit and $K \to \infty$, only the term for $j=0$ in the summation of $j$ in $L_{12}$ survives, thus we have

$$\begin{aligned} L_{12} &= \frac{-\beta q_0}{1-\beta(1-\sum_{k=0}^{K}\Delta_k^{(q)}-q_K)} \to \frac{-\beta q(0)}{1-\beta(1-q(1))+\int_0^1 dx\dot{\Delta}_q(x)} \\ &= \frac{-\beta q(0)}{1-\chi(0)}, \end{aligned} \tag{D.3}$$

where we used $q_0 = q(0)$, $q_K = q(1)$ and $\Delta_i^{(q)} \to (\dot{\Delta}_q(x)/\beta)dx$. The definition of $\chi(x)$ is given by

$$\chi(x) \equiv \beta(1-q(1))+\Delta_q(x)-\Delta_q(1). \tag{D.4}$$

Next, the logarithm in $L_2$ can be expanded with respect to $n$ and it is obtained that

$$\begin{aligned} L_2 &= \frac{p_0}{n}\sum_{k=0}^{K}\left(\frac{1}{p_{k+1}}-\frac{1}{p_k}\right)\ln\left[1-\beta(1-\sum_{i=k+1}^{K}\Delta_i^{(q)}-q_K)\right] \\ &\quad -\beta\sum_{k=0}^{K}\left(\frac{1}{p_{k+1}}-\frac{1}{p_k}\right)\frac{\sum_{i=k}^{K}(p_i-p_{i+1})q_i+q_K-p_kq_k}{1-\beta(1-\sum_{j=k+1}^{K}\Delta_j^{(q)}-q_K)} \\ &\quad +O(n). \end{aligned} \tag{D.5}$$

Here we denote the first and second term in $L_2$ by $L_{21}$ and $(-\beta)L_{22}$, respectively. After changing the order of summation (like in a partial integration), $L_{22}$ becomes

$$\begin{aligned} L_{22} &= \frac{1}{p_0}\frac{\sum_{i=0}^{K}(p_i-p_{i+1})q_i+q_K-p_0q_0}{1-\beta(1-\sum_{i=1}^{K}\Delta_i^{(q)}-q_K)} \\ &\quad +\sum_{k=0}^{K-1}\frac{1}{p_k}\left\{\frac{\sum_{i=k-1}^{K}(p_i-p_{i+1})q_i+q_K-p_{k-1}q_{k-1}}{1-\beta(1-\sum_{i=k}^{K}\Delta_i^{(q)}-q_K)}\right. \\ &\quad \left.-\frac{\sum_{i=k}^{K}(p_i-p_{i+1})q_i+q_K-p_kq_k}{1-\beta(1-\sum_{i=k+1}^{K}\Delta_i^{(q)}-q_K)}\right\} \\ &\quad -\frac{p_Kq_K-p_{K+1}q_K+q_K-p_Kq_K}{1-\beta(1-q_K)}, \end{aligned} \tag{D.6}$$

where we used $p_{K+1}=1$. The first and third term vanish in the DGO limit. If, furthermore, we use $q_{k-1}-q_k=\dot{q}(x)dx$ and take the DGO and the $K\to\infty$ limits, $L_{22}$ finally can be reduced to

$$\begin{aligned} L_{22} &= \sum_{k=0}^{K-1}\frac{-(q_{k-1}-q_k)}{1-\beta(1-\sum_{i=k}^{K}\Delta_i^{(q)}-q_K)} \\ &\simeq \int_0^1 dx\frac{\dot{q}(x)}{1-\beta(1-q(1))+\int_x^1 d\tilde{x}\dot{\Delta}_q(\tilde{x})} \\ &= \int_0^1 dx\,\frac{\dot{q}(x)}{1-\chi(x)}. \end{aligned} \tag{D.7}$$

The third term of equation (A-1) becomes

$$\begin{aligned} L_3 &= \frac{1}{p_0}\ln[1-\beta(1-\sum_{i=0}^{K}\Delta_i^{(q)}-q_K)] \\ &\quad -\frac{1}{n}\ln[1-\beta(1-\sum_{i=0}^{K}\Delta_i^{(q)}-q_K)]. \end{aligned} \tag{D.8}$$



Lastly $L_4$ can be expressed as

$$\begin{aligned} L_4 &= \sum_{k=0}^{K}\left(\frac{1}{p_{k+1}}-\frac{1}{p_k}\right)\ln\left[1-\beta(1-\sum_{i=k+1}^{K}\Delta_i^{(q)}-q_K)\right] \\ &\quad -\frac{p_0}{n}\sum_{k=0}^{K}\left(\frac{1}{p_{k+1}}-\frac{1}{p_k}\right)\ln\left[1-\beta(1-\sum_{i=k+1}^{K}\Delta_i^{(q)}-q_K)\right]. \end{aligned} \tag{D.9}$$

The first term ($L_{41}$) can be represented by changing the order of summation like in $L_{22}$ as

$$\begin{aligned} L_{41} &= \ln[1-\beta(1-q(1))] \\ &\quad +\sum_{k=0}^{K-1}\frac{1}{p_k}\left\{\ln[1-\beta(1-\sum_{i=k}^{K}\Delta_i^{(q)}-q_K)]-\ln[1-\beta(1-\sum_{i=k+1}^{K}\Delta_i^{(q)}-q_K)]\right\} \\ &\quad -\frac{1}{p_0}\ln[1-\beta(1-\sum_{i=1}^{K}\Delta_i-q_K)] \\ &= \ln[1-\beta(1-q_K)] \\ &= \ln[1-\chi(1)]. \end{aligned} \tag{D.10}$$

After summing up all the terms (and taking the DGO limit where not done yet), we finally obtain

$$L=\frac{-\beta q(0)}{1-\chi(0)}-\int_0^1 dx\frac{\beta\dot{q}(x)}{1-\chi(x)}+\ln[1-\chi(1)]. \tag{D.11}$$

## Appendix E: Derivation of the Parisi equation in the Sompolinsky gauge

Here let us concentrate on the calculation of the continuum limit of the partition fucntion (2.11) including the DGO matrix . Calculations are essentially similar to the ones in Appendix C, where the "finite-step RSB partition function" is represented in recursive formula. The DGO matrix in the replica partition function is $\mathbf{R}$, which is parametrized in the same way as the DGO matrix $\mathbf{Q}$ (4.1) as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_0 - \mathbf{E}_0 & \mathbf{R}_0 & \cdots & \mathbf{R}_0 \\ \mathbf{R}_0 & \mathbf{R}_0 - \mathbf{E}_0 & \cdots & \mathbf{R}_0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_0 & \cdots & \cdots & \mathbf{R}_0 - \mathbf{E}_0 \end{pmatrix}. \tag{E.1}$$

where $\mathbf{R}_0$ and $\mathbf{E}_0$ are the $p_0 \times p_0$ Parisi matrices parameterized in Chapter 3. Here, $p_0$ needs to divide $n$ (i.e. $n > p_0$), which is different from the case in Appendix C. Using this another type of replica matrix, let us again derive the formulation for the continuum limit of the partition function:

$$Z_0(z) = \mathrm{Tr}_{\{n\}} \exp\left(\frac{1}{2}\boldsymbol{\sigma}^T \mathbf{R}\boldsymbol{\sigma} + z\boldsymbol{u}^T\boldsymbol{\sigma}\right), \tag{E.2}$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{u}$ are the $n$-dimensional column vectors whose elements are $\sigma_a$ and 1, respectvely. In a similar way as in Appendix C, one can represent as

$$Z_0(z) = \mathcal{Z}_0(\{z_i = z\}), \tag{E.3}$$

where

$$\mathcal{Z}_0(z_i) = \mathrm{Tr}_{\{n\}} \exp\left[\underbrace{\frac{1}{2}\sum_i^{n/p_0} \boldsymbol{\sigma}_{0,i}^T(-\mathbf{E}_0)\boldsymbol{\sigma}_{0,i}}_{L_1} + \underbrace{\frac{1}{2}(\sum_i^{n/p_0} \boldsymbol{\sigma}_{0,i})^T \mathbf{R}_0 (\sum_i^{n/p_0} \boldsymbol{\sigma}_{0,i})}_{L_2} + \underbrace{\sum_i^{n/p_0} z_i \boldsymbol{u}_0^T \boldsymbol{\sigma}_{0,i}}_{L_3}\right]. \tag{E.4}$$

Here $\boldsymbol{\sigma}_{0,i}$ denotes $i$-th $p_0$-dimensional column subvectors dividing $n$-dimensional column vector $\boldsymbol{\sigma}$ to $n/p_0$ blocks as,

$$\boldsymbol{\sigma} = \left[\boldsymbol{\sigma}_{0,1}, \boldsymbol{\sigma}_{0,2}, \cdots, \boldsymbol{\sigma}_{0,n/p_0}\right]^T. \tag{E.5}$$



The first term $L_1$ in $exp(\cdots)$ in rhs of (E.4) can be represented with lower subblocks of replica matrices and subvectors as

$$\begin{aligned} L_1 &= \sum_i^{n/p_0} \left\{ -\frac{1}{2}\sum_j^{p_0/p_1} \boldsymbol{\sigma}_{1,ij}^T \mathbf{E}_1 \boldsymbol{\sigma}_{1,ij} - \frac{\epsilon_0}{2} \sum_{j\neq j'}^{p_0/p_1} \boldsymbol{\sigma}_{1,ij}^T \mathbf{U}_1 \boldsymbol{\sigma}_{1,ij'} \right\} \\ &= \sum_i^{n/p_0} \left\{ -\frac{1}{2}\sum_j^{p_0/p_1} \boldsymbol{\sigma}_{1,ij}^T (\mathbf{E}_1 - \epsilon_0 \mathbf{U}_1)\boldsymbol{\sigma}_{1,ij} - \frac{\epsilon_0}{2} \left(\sum_j^{p_0/p_1} \boldsymbol{\sigma}_{1,ij}^T \boldsymbol{u}_1\right)^2 \right\} \end{aligned} \tag{E.6}$$

where $\boldsymbol{\sigma}_{1,ij}$ denotes $i$-th $p_1$-dimensional column subvectors dividing $p_0$-dimensional column vector $\boldsymbol{\sigma}_{0,i}$ to $p_0/p_1$ blocks as

$$\boldsymbol{\sigma}_{0,i} = \left[\boldsymbol{\sigma}_{1,i1}, \boldsymbol{\sigma}_{1,i2}, \cdots, \boldsymbol{\sigma}_{1,i(n/p_0)}\right]^T. \tag{E.7}$$

We note $\mathbf{U}_1 = \boldsymbol{u}_1\boldsymbol{u}_1^T$ denotes $p_1$-dimensional matrix whose all elements are all 1.

The second term $L_2$ can be represented similarly with lower subblocks of replica matrices as

$$\begin{aligned} L_2 &= \frac{1}{2}\sum_{i,i'}^{n/p_0} \boldsymbol{\sigma}_{0,i}^T \mathbf{R}_0 \boldsymbol{\sigma}_{0,i'} \\ &= \frac{1}{2}\sum_{i,i'}^{n/p_0} \left\{ \sum_j^{p_0/p_1} \boldsymbol{\sigma}_{1,ij}^T \mathbf{R}_1 \boldsymbol{\sigma}_{1,i'j} + r_0 \sum_{j\neq j'}^{p_0/p_1} \boldsymbol{\sigma}_{1,ij}^T \mathbf{U}_1 \boldsymbol{\sigma}_{1,i'j'} \right\} \\ &= \frac{1}{2}\sum_{i,i'}^{n/p_0} \left\{ \sum_j^{p_0/p_1} \boldsymbol{\sigma}_{1,ij}^T (\mathbf{R}_1 - r_0\mathbf{U}_0)\boldsymbol{\sigma}_{1,i'j} + r_0 \sum_{j,j'}^{p_0/p_1} \boldsymbol{\sigma}_{1,ij}^T \mathbf{U}_1 \boldsymbol{\sigma}_{1,i'j'} \right\} \\ &= \frac{1}{2}\sum_j^{p_0/p_1} (\sum_i^{n/p_0} \boldsymbol{\sigma}_{1,ij})^T (\mathbf{R}_1 - r_0\mathbf{U}_1)(\sum_i^{n/p_0} \boldsymbol{\sigma}_{1,ij}) + \frac{r_0}{2}\left(\sum_i^{n/p_0}\sum_j^{p_0/p_1} \boldsymbol{u}_1^T \boldsymbol{\sigma}_{1,ij}\right)^2 \end{aligned} \tag{E.8}$$

Finally, the third term $L_3$ in rhs of (E.4) can be represented as

$$L_3 = \sum_i^{n/p_0} z_i \sum_j^{p_0/p_1} \boldsymbol{u}_1^T \boldsymbol{\sigma}_{1,ij}. \tag{E.9}$$

By substituting (E.6), (E.8) and (E.9) into (E.4) and using again the Hubbard-Stratonovitch identity (A.2), one can represent $\mathcal{Z}_0(z_i)$ as a one-step

RSB partition function as

$$\begin{aligned}
\mathcal{Z}_0(z_i) &= \int dP_{(r_0)}(w)\left[\prod_i^{n/p_0}\int dP_{(-\epsilon_0)}(\eta_i)\right]\mathrm{Tr}_{\{n\}}\exp\left[\sum_j^{p_0/p_1}\Big\{\right.\\
&\quad -\frac{1}{2}\sum_i^{n/p_0}\boldsymbol{\sigma}_{1,ij}^T(\mathbf{E}_1-e_0\mathbf{U}_1)\boldsymbol{\sigma}_{1,ij}\\
&\quad +\frac{1}{2}(\sum_i^{n/p_0}\boldsymbol{\sigma}_{1,ij}^T)(\mathbf{R}_1-r_0\mathbf{U}_1)(\sum_i^{n/p_0}\boldsymbol{\sigma}_{1,ij})\\
&\quad \left.\left.+\sum_i^{n/p_0}(z_i+w+\eta_i)\boldsymbol{u}_1^T\boldsymbol{\sigma}_{1,ij}\Big\}\right]\right]\\
&= \int dP_{(r_0)}(w)\left[\prod_i^{n/p_0}\int dP_{(-\epsilon_0)}(\eta_i)\right]\left[\mathrm{Tr}_{\{np_1/p_0\}}e^{(\text{for any } j)}\right]^{p_0/p_1}\\
&\equiv \int dP_{(r_0)}(w)\left[\prod_i^{n/p_0}\int dP_{(-\epsilon_0)}(\eta_i)\right][\mathcal{Z}_1(w+z_i+\eta_i)]_i^{p_0/p_1} \qquad \text{(E.10)}
\end{aligned}$$

where a short notation for Gaussian integrals is used as

$$\int dP_{(a)}(x)\equiv\int_{-\infty}^{\infty}\frac{dx}{\sqrt{2\pi a}}\exp\left(-\frac{x^2}{2a}\right). \qquad \text{(E.11)}$$

Generally, we define "$k$-steps RSB partition function" and represent it by $k+1$-steps one as

$$\begin{aligned}
\mathcal{Z}_k(\{z_i\}) &\equiv \mathrm{Tr}_{\{np_0/p_k\}}\exp\left\{-\frac{1}{2}\sum_i^{n/p_0}\boldsymbol{\sigma}_{k,i}^T(\mathbf{E}_k-e_{k-1}\mathbf{U}_k)\boldsymbol{\sigma}_{k,i}\right.\\
&\quad \left.+\frac{1}{2}(\sum_i^{n/p_0}\boldsymbol{\sigma}_{k,i}^T)(\mathbf{R}_k-r_{k-1}\mathbf{U}_k)(\sum_i^{n/p_0}\boldsymbol{\sigma}_{k,i})+\sum_i^{n/p_0}z_i\boldsymbol{u}_k^T\boldsymbol{\sigma}_{k,i}\right\} \qquad \text{(E.12)}\\
&= \mathrm{Tr}_{\{np_k/p_0\}}\exp\left\{-\frac{1}{2}\sum_i^{n/p_0}\left[\sum_j^{p_k/p_{k+1}}\boldsymbol{\sigma}_{k+1,ij}^T(\mathbf{E}_{k+1}-e_k\mathbf{U}_{k+1})\boldsymbol{\sigma}_{k+1,ij}\right.\right.\\
&\qquad \left.-\delta e_k\left(\sum_j^{p_k/p_{k+1}}\boldsymbol{\sigma}_{k+1,ij}^T\boldsymbol{u}_1\right)^2\right]\\
&\qquad -\frac{1}{2}\left[\sum_j^{p_k/p_{k+1}}(\sum_i^{n/p_0}\boldsymbol{\sigma}_{k+1,ij})^T(\mathbf{R}_{k+1}-r_k\mathbf{U}_{k+1})(\sum_i^{n/p_0}\boldsymbol{\sigma}_{k+1,ij})\right.\\
&\qquad \left.\left.+\delta r_k\left(\sum_i^{n/p_0}\sum_j^{p_k/p_{k+1}}\boldsymbol{u}_{k+1}^T\boldsymbol{\sigma}_{k+1,ij}\right)^2\right]\right\}.
\end{aligned}$$



$$\begin{aligned}
&\quad +\sum_i^{n/p_0}\sum_j^{p_k/p_{k+1}}z_i\boldsymbol{u}_{k+1}^T\boldsymbol{\sigma}_{k+1,ij}\Big\}\\
&= \int dP_{(\delta r_k)}(w)\left[\prod_i^{n/p_0}\int dP_{(-\delta\epsilon_k)}(\eta_i)\right]\mathrm{Tr}_{np_k/p_0}\exp\left[\sum_j^{p_k/p_{k+1}}\Big\{\right.\\
&\quad -\frac{1}{2}\sum_i^{n/p_0}\boldsymbol{\sigma}_{k+1,ij}^T(\mathbf{E}_{k+1}-e_k\mathbf{U}_1)\boldsymbol{\sigma}_{k+1,ij}\\
&\quad +\frac{1}{2}(\sum_i^{n/p_0}\boldsymbol{\sigma}_{k+1,ij}^T)(\mathbf{R}_{k+1}-r_k\mathbf{U}_{k+1})(\sum_i^{n/p_0}\boldsymbol{\sigma}_{k+1,ij})\\
&\quad \left.\left.+\sum_i^{n/p_0}(z_i+w+\eta_i)\boldsymbol{u}_{k+1}^T\boldsymbol{\sigma}_{k+1,ij}\Big\}\right]\right]\\
&\equiv \int dP_{(\delta r_k)}(w)\left[\prod_i^{n/p_0}\int dP_{(-\delta\epsilon_k)}(\eta_i)\right][\mathcal{Z}_{k+1}(w+z_i+\eta_i)]_i^{p_k/p_{k+1}} \qquad \text{(E.13)}
\end{aligned}$$

where $\delta r_k = r_k - r_{k-1}$ and $\delta e_k = e_k - e_{k-1}$.

Here let us introduce the free energy defined at each step of RSB as

$$\mathcal{G}_k(\{z_i\}) \equiv \frac{p_0}{np_k}\ln\mathcal{Z}_k(\{z_i\}) \qquad \text{(E.14)}$$

$$G_k(z) = \mathcal{G}_z(\{z_i=z\}) \qquad \text{(E.15)}$$

By substituting (E.14) into (E.13) and noting $\Delta_k^{(r)} = \delta\epsilon_k p_k$, we get the following recursion formula:

$$\begin{aligned}
\mathcal{G}_k(\{z_i\}) &= \frac{np_0}{p_k}\ln\int dP_{(\delta r_k)}(w)\left[\prod_i^{n/p_0}\int dP_{(-\delta\epsilon_k)}(\eta_i)\right]\exp\left(\frac{np_0}{p_k}\mathcal{G}_{k+1}(w+z_i+\eta_i)\right)\\
&= \frac{np_0}{p_k}\ln\int dP_{(\delta r_k)}(w)\left[\prod_i^{n/p_0}\int\sqrt{\frac{p_k}{2\pi(-\Delta_k^{(r)})}}d\eta_i\right]\\
&\quad \times\exp\left\{p_k\left(\underbrace{-\frac{1}{2(-\Delta_k^{(r)})}\sum_i^{n/p_0}\eta_i^2+\frac{n}{p_0}\mathcal{G}_{k+1}(w+z_i+\eta_i)}_{\mathcal{F}}\right)\right\} \qquad \text{(E.16)}
\end{aligned}$$

When we consider the DGO limit ($p_k\to\infty$), we can estimate the integration with respect to $\eta_i$ by the saddle point method. By differentiating $\mathcal{F}$ by $\eta_i$, the saddle point $\eta_i^c$ can be given as

$$\left.\frac{d\mathcal{F}}{d\eta_i}\right|_{\eta_i=\eta_i^c}=0 \quad\Longleftrightarrow\quad \eta_k^c = (-\Delta_k^{(r)})\frac{n}{p_0}\left(\frac{\partial\mathcal{G}_{k+1}}{\partial\eta_i}\right)_{\{\eta_i=\eta_k^c\}}$$

$$= (-\Delta_k^{(r)})\frac{dG(w+z+\eta_k^c)}{dz}, \qquad \text{(E.17)}$$

where we have used $f = f(x_1, x_2, \cdots, x_N)$ and $\tilde{f} = f|_{x_i=x}$ give $\frac{d\tilde{f}}{dx} = N\frac{df}{dx_i}$. Thus the integration of rhs in (E.16) can be replaced by its integrand at the saddle point as

$$\begin{aligned}
G_k(z) &\simeq \left(\frac{p_0}{np_k}\right)\ln\int dP_{(\delta r_k)}(w)\exp\left[\frac{np_k}{p_0}\left\{-\frac{1}{2(-\Delta_k^{(r)})}(\eta_k^c)^2+\frac{n}{p_0}G_{k+1}(w+z+\eta_k^c)\right\}\right]\\
&\simeq \left(\frac{p_0}{np_k}\right)\ln\left[1+\int dP_{(\delta r_k)}(w)\left(\frac{np_k}{p_0}\right)\left\{-\frac{(\eta_k^c)^2}{2(-\Delta_k^{(r)})}+G_{k+1}(w+z+\eta_k^c)\right\}\right]\\
&\simeq \int dP_{(\delta r_k)}(w)\left\{-\frac{(\eta_k^c)^2}{2(-\Delta_k^{(r)})}+G_{k+1}(w+z+\eta_k^c)\right\}\\
&= \int dP_{\{1\}}(\tilde{w})\left\{-\frac{(\eta_k^c)^2}{2(-\Delta_k^{(r)})}+G_{k+1}(\sqrt{\delta r_k}\tilde{w}+z+\eta_k^c)\right\}\\
&\simeq -\frac{(\eta_k^c)^2}{2(-\Delta_k^{(r)})}+\int dP_{(1)}(\tilde{w})\left\{G_{K+1}(z+\eta_k^c)+z\sqrt{\delta r_k}G'_{k+1}(z+\eta_k^c)\right.\\
&\qquad \left.+\frac{1}{2}z^2\delta r_k G''_{k+1}(z+\eta_k^c)\right\}\\
&= -\frac{(\eta_k^c)^2}{2(-\Delta_k^{(r)})}+G_{K+1}(z+\eta_k^c)+\frac{1}{2}\delta r_k G''_{k+1}(z+\eta_k^c), \qquad \text{(E.18)}
\end{aligned}$$

where we have used the transformed variable $\tilde{w} = w/\sqrt{\delta r_k}$ for the Gaussian integration and also $\sqrt{\delta r_k} << 1$.

Here let us consider the boundary condition where the above iterative procedure by substituting $k = K+1$ to (E.12) as

$$\begin{aligned}
\mathcal{Z}_{K+1}(z) &= \mathrm{Tr}_{np_{K+1}/p_0}\exp\left[-\frac{1}{2}\sum_i^{n/p_0}\sigma_i{}^2(\tilde{e}-e_K)+\frac{1}{2}\sum_i^{n/p_0}(\tilde{r}-r_K)\sigma_i{}^2+z\sum_i^{n/p_0}\sigma_i\right]\\
&= \left[\mathrm{Tr}_{\{1\}}\exp\left\{-\frac{1}{2}r_K\sigma^2+z\sigma\right\}\right]^{n/p_0} \qquad \text{(E.19)}
\end{aligned}$$

where we have used $p_{K+1} = 1$. We have also used definitions of the diagonal element of Parisi matrix as $\tilde{e} \equiv e_K + r_K$, $\tilde{r} \equiv r_K$. Thus the boundary condition of $G(z)$ can be given as

$$\begin{aligned}
G_{K+1}(z) &= \left(\frac{p_0}{np_{K+1}}\right)\ln\mathcal{Z}_{K+1}(z)\\
&= \ln\left[\mathrm{Tr}_{\{1\}}\exp\left\{-\frac{1}{2}r_K\sigma^2+z\sigma\right\}\right]\\
&= -\frac{r_K}{2}+\ln 2\cosh z \qquad \text{(E.20)}
\end{aligned}$$



After taking the continuum limit:

$$\begin{aligned}
&K\to\infty, \qquad \frac{k}{K}\to x, \qquad \Delta_k^{(r)}\to\dot{\Delta}_r(x)dx,\\
&\delta r_k\to -\dot{r}(x)dx,\\
&\eta_k^c(z) = -\Delta_k^{(r)}\frac{dG}{dz}\to -\dot{\Delta}_r(x)G'dx\\
&G_k(z)\to G(x,z), \qquad \text{(E.21)}
\end{aligned}$$

the eqs. (E.18) and (E.20) reduce to

$$\frac{dG(x,y)}{dx} = -\frac{dr}{dx}\frac{d^2G}{dz^2}+\frac{\Delta_r}{2}\left(\frac{dG}{dz}\right)^2 \qquad \text{(E.22)}$$

$$G(1,z) = -\frac{r(1)}{2}+\ln 2\cosh z. \qquad \text{(E.23)}$$

By taking the transformation

$$\varphi(x,z) = (G(x,z)+r(1)/2)/\beta, \qquad \text{(E.24)}$$

the generalized Parisi equation in the Sompolinsky gauge can be obtained as

$$\dot{\varphi} = -\frac{\dot{r}}{2}\varphi''+\frac{\beta}{2}\dot{\Delta}_r\varphi'^2 \qquad \text{(E.25)}$$

$$\varphi(1,z) = \frac{1}{\beta}\ln 2\cosh z \qquad \text{(E.26)}$$

where dots and primes denote the detivatives with respect to $x$ and $z$, respectively. Here we note that contribution $\tilde{f}$ for the free energy via. $\varphi(x,z)$ can be given in the limit $n\to 0$ as

$$\begin{aligned}
\tilde{f} &= \frac{1}{\beta}G_0 z = \int dP_{(\delta r_0)}(w)\left\{-\frac{(\eta_0^c)^2}{2\Delta_0}+G_1(z+w+\eta_0^c)\right\}\\
&= \int dP_{(r(0))}(w)G_1(w+z)\\
&\sim \int dP_{(r(0))}(w)G(0,w+z) \qquad \text{(E.27)}
\end{aligned}$$

where one cannot neglect $\delta r_0$ bacause $r_{-1} = 0$. For the Hopfield model, after rescaling variables as

$$r(x) \Longrightarrow \alpha\beta^2 r(x), \tag{E.28}$$
$$z \Longrightarrow \sqrt{\alpha}\beta z, \tag{E.29}$$
$$\Delta_r(x) \Longrightarrow \alpha\beta^2 \Delta_r(x), \tag{E.30}$$

we can obtain the terms contributing to the free energy functional (4.14).

# Chapter 5

# Basins of Attraction

## 5.1 Introduction

The most significant characteristic of globally coupled systems such as the Hopfield model is the existence of a large number of (meta)stable states and multivalley structures in phase space. Indeed one expects that an initial state evolves until the bottom of a valley is reached where the system is permanently trapped.

In Chapter 3, we showed that the multivalley structure can be measured by its valleys' "depth" given by energy through the RSB discussion. In this chapter, we focus on the basin of attraction which is another quantitative measure: the "spread" of the valley.

In fact, many theoretical studies have reported relationships between remanent overlaps and initial overlaps in the Hopfield model with sequential[9, 40, 60] or synchronous[47, 6, 54] dynamics. Although these contain some numerical analysis, the system size and the number of samples were rather small. Moreover, a systematic numerical analysis which includes finite size effects on the zero temperature($T = 0$) sequential dynamics has never been performed yet.

This chapter reports remanent overlaps and the basins of attraction for the Hopfield model with zero temperature sequential dynamics, in order to extract some information for the valley structure. Here the remanent overlap defined later refers to a macroscopic order parameter, which corresponds either to the overlap with a memorized pattern or to a generalized remanent magnetization.

According to the finite size scaling analysis, the distribution of the remanent

overlap $m_s$ approaches a delta function as the system size becomes larger. This implies that at $T = 0$ and $N \to \infty$ the value of $m_s$ is determined by the initial overlap and the parameter $\alpha(\equiv p/N)$, the rate of memory-loading, where $p$ is the number of random patterns for memories and $N$ the system size. Hence the dependence of the value of $m_s$ on the initial overlap and on a parameter $\alpha$ is obtained. Here we concentrate on relatively large values of $\alpha$ where the system is in the spin glass phase. The above dependence of $m_s$ enables us to get the relationship between remanent overlaps and initial overlaps, which leads to a distribution function of the remanent overlap. We have also studied the asymptotic dependence of $m_s$ on $\alpha$ and have obtained the power-law decay of $m_s$ to the value corresponding to the remanent magnetization obtained in the SK model. This supports the equivalence between the Hopfield model in the "SK limit($\sqrt{\alpha} \to \infty$)" and the SK model.

The system under consideration has the Ising spins $\{S_i\}$ and the Hamiltonian (2.4) with interaction (2.5). Here we concentrate on the zero temperature sequential dynamics,

$$S_i(t+1) = \theta(\sum_j J_{ij}\, S_j(t)\,) \tag{5.1}$$

where $\theta(x)$ is the step function ($\theta(x) = 1$ for $x > 0$ ; $-1$ for $x \leq 0$). The time $t$ is given in units of Monte Carlo steps(MCS). Thus we will deal with a dynamics which is deterministic except for a random order of spin flips. At $T = 0$, the model is expected to converge rapidly to a final state which is stable with regard to one spin flip.

Let us now consider the time-dependent macroscopic order parameter, called the *pattern overlap*,

$$m^\mu(t) = \frac{1}{N}\sum_i \xi_i^\mu S_i(t). \tag{5.2}$$

Here it should be noted that the overlap represents a generalized magnetization, i.e., $m^\mu$ becomes the usual magnetization by the transformation $\widetilde{S}_i = \xi_i^\mu S_i$. We expect that the overlap $m^\mu(t \to \infty)$ does not vanish on a considerably wide range of $\alpha$ at $T = 0$. We will call this nonzero value the "remanent overlap($m_s^\mu \equiv m^\mu(\infty)$)".



## 5.2 Finite Size Scaling Analysis

Considering finite size effects and fluctuations between samples, we have investigated the remanent overlap at $T = 0$ from various initial states at various values of $\alpha$ with the use of Monte-Carlo simulations and the finite size scaling analysis.

We take sample averages in several steps, as follows:

(a) Generate a set of random patterns $\{\xi^\mu\}$, to form one interaction matrix $J = \{J_{ij}\}$ from eq. (2.5).

(b) Prepare $p$ sets of initial states $\{S_i(t = 0)\}$ which have the same initial overlap $m^\mu(0)$ with each pattern $\xi^\mu$.

(c) Execute the Monte Carlo dynamics (5.1) for each $\mu$ until the system converges to a stable state (empirically, at $T = 0$, the system rapidly evolves towards a stable state in about 20 MCS/S). Therefore $p$ sets of $m_s^\mu$ are obtained simultaneously for each $J$. This is advantageous because one can reduce the number of samples $\{J\}$.

(d) Evaluate the $m_s^\mu$'s, which are regarded as one set of one $m_s$, and then update the histogram of $m_s$ accordingly.

(e) Return to (a).

The above procedure is executed $N_\epsilon$ times,

$$N_\epsilon = N_{max}/p = N_{max}/(\alpha N) \qquad (N_{max} = 2 \times 10^4). \tag{5.3}$$

The total number of runs $N_{max}$ was determined by the requirement that the resulting distribution of remanent overlaps does not change significantly any more under increasing $N_{max}$. In comparison with previous similar studies[40, 60, 47, 6], we would like to note that our value of $N_{max}$ can be considered as rather large.

Finally histograms for certain values of $m(0)$ and $\alpha$ are obtained for various system sizes ($N = 200, 500, 1000, 2000$). Figure 5.1 is an example of the histogram of $m_s$ at $m(0) = 0.5$ and $\alpha = 0.5$. Dependence of average values $\langle m_s \rangle$ and variance $\langle (m_s - \langle m_s \rangle)^2 \rangle$ on $N$ is obtained by varying the system size $N$. The average $\langle m_s \rangle$ can be extrapolated to a non-zero value depending on $\alpha$

and $m(0)$ as,

$$\langle m_s(N; m(0), \alpha)\rangle \sim m_s(\infty; m(0), \alpha) + \frac{\text{const.}}{\sqrt{N}}, \tag{5.4}$$

while the variance is scaled to zero as,

$$\langle (m_s - \langle m_s\rangle)^2\rangle \sim \frac{1}{N}. \tag{5.5}$$

Figure 5.1 also represents the basins of attraction of the remanent overlap $m_s$ from the specific value of the initial overlap $m(0) = 0.5$ for different system sizes at $\alpha = 0.5$. With the increase of size $N$, the distribution $P_\alpha(m_s; m(0))$ of $m_s$ from the initial overlap $m(0) = 0.5$ approaches a delta function as follows,

$$P_\alpha(m_s; m(0)) \sim \delta(m_s - F_\alpha(m(0))) \quad \text{as} \quad N \to \infty, \tag{5.6}$$

where $F_\alpha(m(0))$ denotes a function of initial overlap $m(0)$ with a parameter $\alpha$, which is mentioned later.

After taking into account these scaling relationships, we get the remanent overlaps at $N \to \infty$. The remanent overlaps depend both on $m(0)$ and $\alpha$ as given in Fig. 5.2. The result explicitly shows that the final attracting state (remanent overlap) can be given from an initial state (initial overlaps) if $\alpha$ is given.

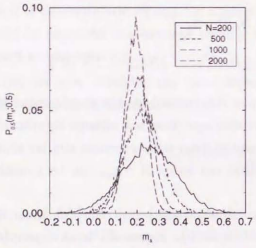



Figure 5.1: The basins of attraction for different system sizes.

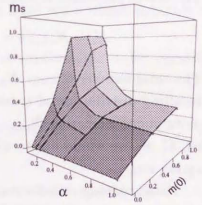



Figure 5.2: Dependence of remanent overlap $m_s$ on two parameters, $\alpha$ and $m(0)$.



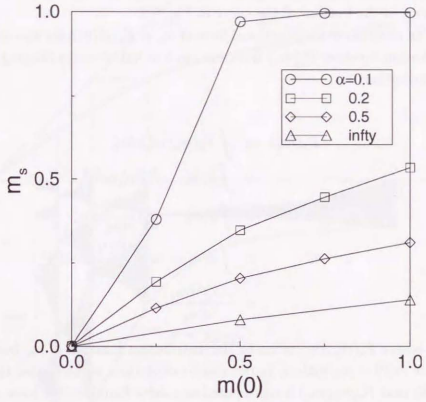



Figure 5.3: Dependence of remanent overlap $m_s$ on $m(0)$ for various values of $\alpha$.

Our results also give the characteristic dependence of stable states on initial states even in the parameter range corresponding to the SK model ($\alpha \to \infty$).

Here let us see the dependence of $m_s$ on $m(0)$ with a certain value of $\alpha$. As shown in Fig. 5.3, $m_s$ is represented as a monotonic increasing function of $m(0)$ with the parameter $\alpha$ : $m_s \equiv F_\alpha(m(0))$. At a rather small value of $\alpha(= 0.1)$ below the critical storage capacity $\alpha_c$ ($\sim 0.155$) (see Chapter 4), $F_{0.1}(m(0))$ is steep at a rather small region of $m(0)$ and saturates to unity at the high $m(0)$ region. This indicates that the system has rather large basins for the corresponding pattern; almost all initial states tend to go to the corresponding patterns. At an intermediate region($\alpha < 1$), $F_\alpha(m(0))$ increases slowly. At $\alpha = 0.2$ we get

$$m_s = F_{0.2}(m(0)) \sim m(0)^\gamma, \quad \gamma \sim 0.73 \tag{5.7}$$

in the rather high $m(0)$ region. The limiting values of $m_s$'s with $\alpha \to \infty$ are shown in the bottom of the curve in Fig. 5.3.

In addition to the functional form of $m_s = F_\alpha(m(0))$, we also obtain the distribution function $P_\alpha(m_s)$ from the random initial states obeying the uniform distribution:

$$\begin{aligned} P_\alpha(m_s) &= \int P_\alpha(m_s; m_0')dm_0' \\ &= \int \delta(m_s - F_\alpha(m_0'))dm_0' \\ &= \int \delta(m_s - m_s'))\frac{dm_0'}{dm_s'}dm_s' \\ &= \int \delta(m_s - m_s')\frac{dF_\alpha^{-1}(m_s')}{dm_s'}dm_s' \\ &= \frac{dF_\alpha^{-1}(m_s)}{dm_s}, \end{aligned} \tag{5.8}$$

where $P_\alpha(m_s; m_0)$ refers to the distribution function of $m_s$ from the initial state $m(0) = m_0$ with $\alpha$. In the above calculations we have used the first result (5.6) that $P_\alpha(m_s; m_0)$ is represented as a delta function. We have also used the existence of the inverse function $F_\alpha^{-1}$ since $F_\alpha$ is a monotonically increasing function.

Equations (5.7) and (5.8) imply that the distribution function $P_\alpha(m_s)$ for the remanent overlap also has a power-law dependence on $m_s$. Thus stable states are distributed continuously in the space of overlap and the measure of basins of attraction follows the power law in the space of overlaps. This result is consistent with the power law distribution of remanent overlaps as obtained numerically in ref. [3].

To summarize our results so far, flows in the overlap space are schematically shown in Fig. 5.4. In the figure, upper column represents the map of $m_s$ from $m(0)$. Lower column represents flows on the space of overlap and spin configuration space schematically. The lower vertical axis shows the "spread of configuration space". The dotted curve $N(m)$ schematically represents the upper limit of the number of states in the configuration space with overlap $m$, e.g. $N(1) = 1$. In the lower part, solid arrows at the shaded area represent flows from initial states with $m(0)$ to stable states with $m_s$. For example, an initial state with $m(0) = 1$ is attracted to the fixed point(A) with remanent



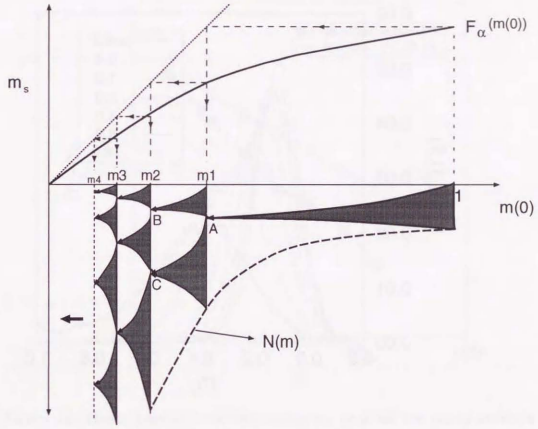



Figure 5.4: The schematic flows on the space of overlaps.

overlap $m_1$. However almost all states with $m(0) = m_1$ except the point A are attracted to a state with $m_s = m_2$(point B or C). These sequences are iterated infinitely as $m_2 \to m_3$, $m_3 \to m_4$ and so on. Actually the states attracting initial states(like points A, B and C) exist infinitely in the configuration space, while there are much more remaining states with the same overlap attracted onto states with lower overlaps. In other words, the fixed points like A, B or C have no measure in the configuration space on the set of states with the same overlap. Since there exist infinite stable states like points A, B or C, the stable states are distributed on a Cantor set on the hyper plane with the same overlap. Our results support a similar conjecture by Amari and Maginu[6] and provides us with a more intuitive understanding of the structure of the stable states in the overlap space.

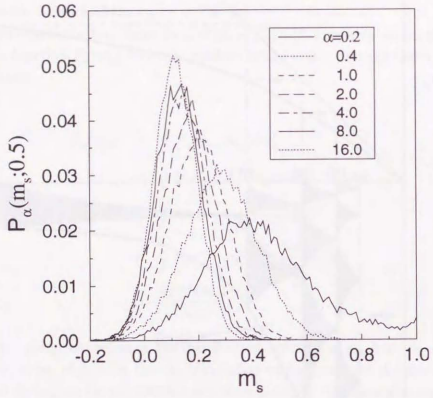



Figure 5.5: The basins of attraction for different values of $\alpha$: distributions of remanent overlaps $m_s$ from initial states with $m(0) = 0.5$ at $N = 200$ for several values of $\alpha$.

The second result concerns the dependence of $m_s$ on $\alpha$. Distribution of $m_s$ from the initial overlap $m(0) = 0.5$ at $N = 200$ for various values of $\alpha$ are given in Fig. 5.5. Figure 5.5 also represents the dependence of the basins of attraction of $m_s$ on the value of $\alpha$. Accordingly we study the asymptotic dependence of $m_s$ on $\alpha$. As shown in Fig. 5.6, $m_s \equiv G_{m(0)}(\alpha)$ monotonically decreases for a high $\alpha$ region as,

$$m_s = G_{m(0)}(\alpha) = \widetilde{m} + A\,\alpha^{-\beta}, \qquad \beta > 0 \tag{5.9}$$

where $\widetilde{m} \sim 0.08\,, \beta \sim 0.5$ for $m(0) = 0.5$ and $\widetilde{m} \sim 0.14\,, \beta \sim 0.3$ for $m(0) = 1$.

Indeed $m_s \sim 0.14$ agrees very well with the remanent magnetization of the SK model reported by Kinzel[66]. On the other hand the value $m_s \sim 0.08$ from $m(0) = 0.5$ probably coincides with the remanent overlap when we quench the



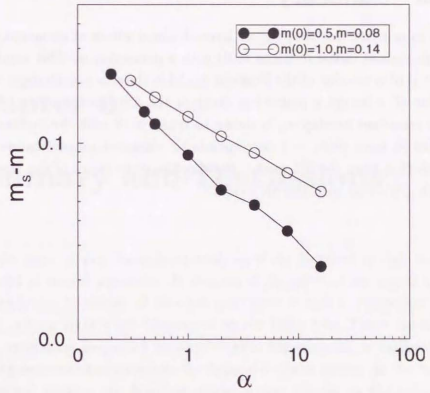



Figure 5.6: Dependence of remanent overlap $m_s$ on $\alpha$ for two initial overlaps at $m(0) = 1$ and 0.5. The initial state $m(0) = 1$ corresponds to a ferromagnetically ordered configuration while $m(0) = 0.5$ to a state with 75% of spins 1.

system from a specific initial state where 75% of spins get 1 and others $-1$. We note that there remains a finite remanent overlap(magnetization) $m_s$ even for an initial state with $m(0) \neq 1$.

Figures 6 also suggests that the Hopfield model gradually approaches the SK model according to a power law with the increase of $\alpha$, which implies that the transition occurs very slowly from the Hopfield model to the SK model. Hence we conclude that there are more intermediate models worthy of study, which differ both from the Hopfield model and the fully investigated SK model. These intermediate models also have rich and complex structures of the basins of attraction that interpolate the above two models.

## 5.3 Summary

To summarize this chapter, we have obtained values of remanent overlaps $m_s$ from various initial overlaps $m(0)$ with a parameter $\alpha$. This result shows the basins of attraction of the Hopfield model in the space of overlaps. At a certain value of $\alpha$ we get a power-law decay of the distribution of $m_s$. Furthermore the remanent overlap $m_s$ is shown to decay to $\overline{m}$ with the increase of $\alpha$. The value $\overline{m}$ from $m(0) = 1$ corresponds to remanent magnetization and agrees with that from the SK model, showing the equivalence of the Hopfield model with $\sqrt{\alpha} \to \infty$ with the SK model.

# Chapter 6

# Summary and Discussions

The main topic of this thesis is an analysis of the Hopfield model; the physical model of neural networks. It consists of formal neurons represented by two-state binary variables. It also has symmetric synaptic-connections (interactions), whose weights are determined by the Hebb rule. The noisy response of each neuron, as suggested by physiological experiments, is realized by introducing some stochasticity into the dynamics of the model. In the language of statistical physics, the Hopfield model is very similar to the spin glasses (SG), and has become one of the most celebrated paradigms in both statistical physics and neural networks.

The most important point I would like to stress is that the Hopfield model has enabled us to carry out a clear and rich quantitative analysis. The analysis was possible because the model is defined as an energy system via its Hamiltonian. It therefore allows us to use traditional methods of statistical mechanics, including techniques such as the replica method.

I would now like to summarize the main findings of this thesis. First, it was shown that the Hopfield model shares a significant feature of the SG phase in having a multivalley energy landscape (Chapter 3). This is highly non-trivial and could never have been clarified without the help of the RSB scheme. It was found that such a multivalley energy landscape changes its structure as the parameter $\alpha$ varies. The complexity of such landscapes has been measured by using a quantity which characterizes their valley structure, and which was defined in eq. (3.2) as

$$y \equiv \langle\langle \sum_k W_k^2 \rangle\rangle, \qquad (W_k : \text{weight of the } k\text{-th valley}). \tag{6.1}$$

At the same time, $y$ also corresponds to the inverse of the partition ratio. Let us briefly illustrate the meaning of $y$: if, e.g., $y \sim 1$, there is only on simple valley, while an infinite number of small valleys coexist if $y \sim 0$. With the help of Parisi's physical interpretation of the RSB solution[96], it was then shown in Chapter 3 that $y$ can be obtained via order parameter function $q(x)$. In fact, the dependence of $y$ on $\alpha$ is obtained as

$$y \sim \alpha^{-\gamma} \qquad (\gamma \sim 0.5), \tag{6.2}$$

which means that $y$ vanishes in the limit $\sqrt{\alpha} \to \infty$. Therefore, in this limit, the multivalley energy landscape of the Hopfield model is composed of an infinite number of small valleys and attains the most complex structure. This structure corresponds to the one discovered in the naive mean field model of SG by quenched dynamics[89]. Here, at first glance, the result "$y \to 0$ in the limit $\sqrt{\alpha} \to \infty$" seems to contradict a prediction for the SK model[78] which states that $y \sim 1$ However the dependence of $y$ on $\alpha$ in eq. (6.2) is obtained without keeping $\tilde{\beta}$ ($\equiv \sqrt{\alpha}\beta$) constant, where $\beta$ denotes the usual inverse temperature ($1/T$). Accordingly, the "effective inverse temperature $\tilde{\beta}$ " becomes infinitely large in the limit $\sqrt{\alpha} \to \infty$, which may be interpreted as the result of a quenched dynamics.

By solving the Parisi equations numerically for several values of $\alpha$ and at several temperature, the order parameter functions (Figs. 3.1, 3.2 and 3.3) and the internal field distributions $P(1, z)$ of the SG phase are determined for the first time in the present work. In this way, the non-trivial dependence of $P(1, z)$ on $\alpha$ and the temperature is clarified (Fig. 3.4). It is found that the typical double-peaked form of $P(1, z)$ becomes sharper as the temperature gets lower and $\alpha$ gets larger just like SG. Hence it can be said that $\alpha$ plays the role of an inverse temperature.

Furthermore, it is found that the Parisi equations of the Hopfield model formally coincide with those obtained for the SK model[85] in the limit $\sqrt{\alpha} \to \infty$ while keeping $\tilde{\beta}$ constant. We call this the "SK limit" of the Hopfield model. The SK limit characterizes an asymptotic dependence of the Hopfield model on $\alpha$. Consequently the Hopfield model approaches the SK model as $\alpha$ gets larger.

On the topics above-mentioned, we have restricted our attention mainly to the SG phase at a rather large value of $\alpha$ where the FMR phase does not appear. Now let us summarize the topics developed in Chapter 4 which



shed light on the region where the RSB solutions of the SG and the FMR phase coexist. This low temperature region below the AT line (Fig. 2.3(b)) had been unclear so far. In such a region the RS solution for both the SG and FMR phases are unstable and one should consider the RSB. Therefore the full RSB with the so-called Sompolinsky gauge is formulated. The focal attention is on the numerical estimation of the order parameters and other physical quantities at zero temperature ($T = 0$). When we address the zero temperature case, the RSB scheme with the Sompolinsky gauge is defined conveniently for numerical analysis, while the RSB scheme with Parisi's is inconvenient since it cannot overcome the singularity at $T = 0$. Extensive numerical analyses were carried out, both for the SG and FMR phases, in the most interesting region where the FMR phase disappears. Consequently, it is shown that the critical storage capacity at $T = 0$ needs to be corrected to $\alpha_c = 0.155 \pm 0.002$ which is larger than the capacities previously obtained by both the RS (0.138)[8, 9] and the 1-step RSB (0.144)[27] discussions. Our result is the first self-consistent estimation of the critical storage capacity using the RSB scheme. Another interesting aspect of our results is that it also indicates that RSB promotes the stability of the FMR solution against the increase of the so-called "slow" or "stochastic synaptic" noise[10] originating in $\alpha$'s increase. At the same time, the order parameter functions and the frozen field distributions at zero temperature are determined. It is clarified for the SG phase that the order parameter functions (Figs. 4.1 and 4.2) and frozen field distributions (Fig. 4.3) do not vary drastically around $\alpha_c$. Hence the disappearance of the metastable FMR states does not affect the SG phase. On the other hand, the order parameter function $q(x)$ in the FMR phase shows a typical dependence on $\alpha$. In particular, $q(0)$ decreases as $\alpha \to \alpha_c$, showing that the variety of overlaps between two pure states increases abruptly near $\alpha_c$. Consequently the transition where the FMR states (dis)appear is accompanied with an abrupt (dis)appearance of many pure states which have different patterns of spin configurations but the same macroscopic overlaps with one of the learning patterns.

Here we note that $P(1, z)$ in the SG phase has a non-trivial form and cannot be approximated by any Gaussian distribution at all. This is the main reason why one cannot describe the dynamics of the Hopfield model by a finite number of macroscopic time-dependent order parameters (it should be a function as $q(x)$ !) in such a way that it converges to the to the spurious state or the SG

state, while one can approximately describe a dynamics which converges to the FMR states[6, 47]. Due to the lack of an explicit description of the dynamics which leads to the spurious state, in Chapter 5, the Monte Carlo simulation and the finite size scaling analysis is carried out in order to investigate the behavior of macroscopic order parameters, such as the *remanent overlap*. The analyses also clarified the dynamics in the complex energy valley lead to a final state, which is at most a point attractor. This yields a definite value for the overlap because only the zero temperature sequential dynamics was considered. The main result is that the value of the remanent overlap is determined by the initial overlap and is only influenced by $\alpha$ if $N \to \infty$. This has enabled us to establish a mapping of initial overlaps to remanent overlaps (Fig. 5.4). Such a mapping is further useful for imagining the composition of the spurious (meta)stable states. It is found that the (meta)stable states form a Cantor set on the hyper plane with the same overlap. Here we note that the remanent overlap from a certain value of initial overlap in the limit $\sqrt{\alpha} \to \infty$ coincides with the remanent magnetization previously studied for the SK model. This can be considered as indirect evidence for the SK limit of the Hopfield model.

# Bibliography


[1] Abeles, M., E. Vaadia and H. Bergman (1990). *Networks* **1**, 13

[2] Ad. M. H. J. Aertsen and G. L. Gerstein (1991). In *Neuronal Cooperativity* ed. by J. Krüger (Berlin: Springer)

[3] Aizawa, Y. and K. Tokita (1989). *Proceedings on Conf. of Fuzzy Logic & Neural Networks* (Kyushu Inst. of Tech., Iizuka JAPAN, 1989), 637

[4] Amari, S. (1977). *Biol. Cybern.* **27**, 77

[5] Amari, S. (1980). *Bulletin of Mathematical Biology* **42**, 339

[6] Amari, S. and K. Maginu (1988). *Neural Networks* **1**, 63

[7] Amit, D. J., H. Gutfreund and H. Sompolinsky (1985). *Phys. Rev* **A32**, 1007

[8] Amit, D. J., H. Gutfreund and H. Sompolinsky (1985). *Phys. Rev. Lett.* **55**, 1930

[9] Amit, D. J., H. Gutfreund and H. Sompolinsky (1987). *Ann. Phys., NY* **173**, 30

[10] Amit, D. J. (1989). *Modeling Brain Function* (Cambridge)

[11] Amit, D. J. and A. Treves (1990). *Proc. Natl. Acad. Sci. USA* **86**, 465

[12] Amit, D. J., M. Evans and M. Abeles (1990). *Network* **1**

[13] Anderson, P. W. (1983). *Proc. Natl. Acad. Sci. USA* **80**, 3386

[14] Anderson, P. W. (1992). In *Spin Glasses and Biology*, ed. by D. Stein (Singapore: World Scientific)

[15] Austin, R. H. and C. M. Chen (1992). In *Spin Glasses and Biology* ed D Stein (Singapore: World Scientific)

[16] Bantilan, F. T. and R. G. Palmer (1981). *J. Phys. F* **11**, 261

[17] Barkai, E., D. Hansel and H. Sompolinsky (1992) *Phys. Rev. A* **45**, 4146

[18] Binder, K and D. W. Heerman (1988). *Monte Carlo Simulation in Statistical Mechanics* (Berlin: Springer-Verlag)

[19] Binder, K. and A. P. Young (1986). *Rev. Mod. Phys.* **58**, 801

[20] Buhmann, J. (1989). *Phys. Rev. A* **40**, 4145

[21] Bray, A. J. and M. A. Moore (1980). *J. Phys. C: Solid State Phys.* **13**, L469

[22] Bray, A. J., H. Sompolinsky and C. Yu (1986). *J. Phys. C: Solid State Phys.* **19**, 6389

[23] Caianiello, E. R. (1961). *J. Theor. Biol.* **1**, 204

[24] Chowdhury, D. (1986). *Spin Glasses and Other Frustrated Systems* (Singapore: World Scientific)

[25] Cragg, B. G. and H. N. V. Temperley (1954). *EEG and Clinical Neurophysiology* **6**, 86

[26] Cragg, B. G. and H. N. V. Temperley (1955). *Brain* **78**, *II*, 304

[27] Crisanti, A., D. J. Amit and H. Gutfreund (1986). *Europhys. Lett.* **2**, 337

[28] De Almeida J. R. L. and D. J. Thouless (1978). *J. Phys. A: Math. Gen.* **11**, 983

[29] De Dominicis, C., M. Gabay and B. Duplantier (1982). *J. Phys. A: Math. Gen.* **15**, L47

[30] De Dominicis, C., M. Gabay and H. Orland (1981). *J. Physique Lett.* **42**, L523

[31] De Dominicis, C. and A. P. Young (1983). *J. Phys. A: Math. Gen.* **16**, 2063



[32] Derrida, B (1988). In *Nonlinear Evolution and Chaotic Phenomena*, ed. by G. Gallavotti and P. F. Zweifel (Plenum)

[33] Duplantier, B. (1981). *J. Phys. A: Math. Gen.* **14**, 283

[34] Eckhorn R. (1991). In *Neuronal Cooperativity* ed. by J. Krüger (Berlin: Springer)

[35] Edwards, S. F. and P. W. Anderson (1975). Theory of Spin Glasses. *J. Phys.* **F5**, 965

[36] Fischer, K. H. (1983). *Phys. Status Solidi. B* **116**, 357

[37] Fischer, K. H. (1985). *Phys. Status Solidi. B* **130**, 13

[38] Fischer, K. H. and J. A. Hertz (1991) *Spin Glasses* (Cambridge)

[39] Fisher D. S. and D. A. Huse (1987). *J. Phys. A: Math. Gen.* **20**, L1005

[40] Fontanari, J. F. and R. Köberle (1988). *J. Phys. A: Math. Gen.* **21**, L667

[41] Frauenfelder, H. (1987). In *Protein Structure: Molecular and Electronic Reactivity* (New York: Springer)

[42] Fu, Y. T. and P. W. Anderson (1986). *J. Phys. A: Math. Gen.* **19**, 1605

[43] Fukai, T. and M. Shiino (1990). *Phys. Rev. Lett.* **64**, 1465

[44] Fukai, T. and M. Shiino (1990). *Phys. Rev. A* **42**, 7459

[45] Fukai, T. and M. Shiino (1992). *J. Phys. A: Math. Gen.* **25**, 2873

[46] Fukai, T. and M. Shiino (1993) *Phys. Rev. E* **48** 867

[47] Gardner, E., B. Derrida and P. Mattishaw (1987). *J. Physique* **48**, 741

[48] Geszti, T. (1990). *Physical Models of Neural Networks* (Singapore: World Scientific)

[49] Glauber, R. J. (1963). Time-dependent statistics of the Ising model. *J. Math. Phys.* **4**, 294

[50] Gray, C. M., P. König, A. K. Engel and W. Singer (1989). *Nature* **338**, 334

[51] Grossberg, S. (1988). *Neural Networks* **1**, 17 and references therein.

[52] Gutfreund, H. and G. Toulouse (1992). In *Spin Glasses and Biology* ed D Stein (Singapore: World Scientific) p26

[53] Hebb, D. O. (1949). *The Organization of Behavior* (New York: Wiley)

[54] Henkel, R. D. and M. Opper (1991). *J. Phys. A: Math. Gen.* **24**, 2201

[55] Hertz, J. A., G. Grinstein and S. A. Solla (1986). In *Neural Networks for Computing* AIP Conf. Proc. **151** ed. by J. S. Denker

[56] Hertz, J. A., G. Grinstein and S. A. Solla (1986). In *Heidelberg colloquium on Glassy Dynamics*, Lecture Notes in Physics **275**, ed. by J. L. van Hemmen and I. Morgenstern (Berlin: Springer)

[57] Hertz, J., A. Krogh and R. G. Palmer (1991). *Introduction to the Theory of Neural Computation* (Santa Fe Institute Studies in the Sciences of Complexity: Addison-Wesley)

[58] Hopfield, J. (1982). *Proc. Natl. Acad. Sci. USA* **79**, 2554

[59] Hopfield, J .J . and D. W. Tank (1985). *Biol. Cybern.* **52**, 141

[60] Horner, H., D. Bormann, M. Frick, H. Kinzelbach and A. Schmidt (1989). *Z. Phys. B Condensed Matter* **76**, 381

[61] Huse D. A. and D. S. Fisher (1987). *J. Phys. A: Math. Gen.* **20**, L997

[62] Ikegami, T. and K. Kaneko (1992). *Chaos* **2**, 397

[63] Kaneko, K. (1993). *Theory and applications of coupled map lattices* (Chichester: Willey), and references therein.

[64] Kauffman, S. A. (1992). In *Spin Glasses and Biology*, ed. by D. Stein (Singapore: World Scientific)

[65] Kauffman, S. A. (1993). *The Origins of Order* (New York: Oxford)

[66] Kinzel, W (1986). *Phys. Rev. B* **33**, 5086

[67] Kirkpatrick S., C. D. Gelatt, Jr. and M. P. Vecchi (1983). *Science* **220**, 671



[68] Kohonen, T. (1989). *Self-Organization and Associative Memory* (3rd ed.) (Berlin: Springer-Verlag) and references therin.

[69] Little, W. A. (1974). *Mathematical Biosciences* **19**, 101

[70] Little, W. A. and G. L. Shaw (1975). *Behavioral Biology* **14**, 115

[71] Little, W. A. and G. L. Shaw (1978). *Mathematical Biosciences* **39**, 281

[72] von der Malsburg, C. and E. Bienenstock (1987). *Europhys. Lett.* **3**, 1243

[73] Marr, D. (1969). *Journal of Physiology* **202**, 437

[74] Marr, D. (1970). *Proceedings of the Royal Society of London B* **176**, 161

[75] Mattis, D. C. (1976). *Phys. Lett. A* **56**, 421

[76] McCulloch, W. S. and W. Pitts (1943). *Bulletin of Mathematical Biophysics* **5**, 115

[77] Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller (1953). *J. Chem. Phys.* **21**, 1087

[78] Mézard, M., G. Parisi, N. Sourlas, G. Toulouse and M. Virasoro (1984). *J. Physique* **45**, 84

[79] Mézard, M. and G. Parisi (1985). *J. Physique Lett* **46**, L771

[80] Mézard, M. and G. Parisi (1986). *J. Physique* **47**, 1286

[81] Mézard, M., G. Parisi and M .Virasoro (1987). *Spin Glass Theory and Beyond* (Singapore: World Scientific)

[82] Minsky, M. L. and S. A. Papert (1969). *Perceptrons* (MIT Press)

[83] Morita, M., S. Yoshizawa and K. Nakano (1990). *Trans. IEICE* **J73-DII**, 232

[84] Nemoto, K and H. Takayama (1985). *J. Phys. C: Solid State Phys.* **18**, L529

[85] Nemoto, K. (1987). *J. Phys. C: Solid State Phys.* **20**, 1325

[86] Nishimori, H. and T. Ozeki (1993). *J. Phys. A: Math. Gen.* **26**, 859

[87] Nishimori, H. (1993). *J. Phys. Soc. Jpn* **62**, 2973

[88] Nishimori, H. and N. Ouchi (1993). *Phys. Rev. Lett.* **71**, 197

[89] Nishimura K., K. Nemoto and H. Takayama (1990). *J. Phys. A: Math. Gen.* **23**, 5915

[90] Nishimura, K. and K. Nemoto (1992). *J. Phys. A: Math. Gen.* **4**, 5561

[91] Nozawa, H. (1992). *Chaos* **2**, 377

[92] Ogielski, A. T. and D. L. Stein (1985). *Phys. Rev. Lett.* **55**, 1634

[93] Palmer, R. G. (1982). *Advances in Phys.* **31**, 669

[94] Parisi, G. (1980). *J. Phys. A: Math. Gen.* **13**, L115

[95] Parisi, G. (1980). *J. Phys. A: Math. Gen.* **13**, 1887

[96] Parisi, G. (1983). *Phys. Rev. Lett.* **50**, 1946

[97] Rosenblatt, F. (1962). *Principles of Neurodynamics* (New York: Spartan)

[98] Ruján, P. (1993). *Phys. Rev. Lett.* **70**, 2968

[99] Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). *Parallel Distributed Processing* (MIT Press)

[100] Sasai, M. and P. G. Wolynes (1990). *Phys. Rev. Lett.* **65**, 2740

[101] Seung, H. S., H. Sompolinsky and N. Tishby (1992). *Phys. Rev. A* **45**, 6056

[102] Shiino M., H. Nishimori and M. Ono (1989). *J. Phys. Soc. Jpn* **8**, 763

[103] Shiino, M. (1990). *J. Stat. Phys* **59**, 1051

[104] Shiino M., and T. Fukai (1990). *J. Phys. Soc. Jpn* **59**, 1529

[105] Shiino M., and T. Fukai (1990). *J. Phys. A: Math. and Gen.* **23**, L1009

[106] Sherrington, D. and S. Kirkpatrick (1975). *Phys. Rev. Lett.* **35**, 1792



[107] Singer, W. (1991) In *Neuronal Cooperativity*, ed. by J. Krüger (Berlin: Springer)

[108] Sommers, H-J. and W. Dupont (1984). *J. Phys. C: Solid State Phys.* **17**, 5785

[109] Sompolinsky, H. (1981). *Phys. Rev. Lett.* **47**, 935

[110] Sompolinsky, H., D. Golomb and D. Kleinfeld (1990). *Proc. Natl. Acad. Sci. USA* **87**, 7200

[111] Sourlas, N. (1989). *Nature* **339**, 693

[112] Stein, D. L. (ed) (1992). *Spin Glasses and Biology* (Singapore: World Scientific)

[113] Takayama, H. (1989). *Cooperative Dynamics in Complex Physical Systems* (Berlin: Springer Sereies in Synergetics, **43**)

[114] Takayama, H. (1990). *Spin Glass* (Tokyo: Maruzen)

[115] Takayama, H. and K. Nemoto (1990). *J. Phys. C: Solid State Phys.* **2**, 1997

[116] Takayama H. (1993). *J. Phys. Soc. Jpn* **61**, 2512

[117] Thouless, D. J., P. W. Anderson and R. G. Palmer (1977). *Phill. Mag.* **35**, 593

[118] Tokita, K. (1992). In *Proceedings of the conference "complexity"*, Kyoto Jul. 8 - 10. in *Bussei Kenkyuu* **59**, 322

[119] Tokita, K. (1993). *Prog. Theor. Phys.* **90**, 329

[120] Tokita, K. (1993). *J. Phys. A: Math. Gen.* **26**, 6915

[121] Tokita, K. (1993). *The replica-symmetry-breaking solution of the Hopfield model at zero temperature : critical storage capacity and frozen field distribution*, preprint, submitted to *J. Phys. A: Math. Gen.*

[122] Tokita, K. (1993). In *Proceedings of the conference "complexity 2"*, Kyoto Aug. 6 - 8. to appear in *Bussei Kenkyuu*

[123] Vaadia E., E. Ahissar, H. Bergman and Y. Lavner (1991) In *Neuronal Cooperativity*, ed. by J. Krüger (Berlin: Springer)

[124] van Hemmen, J. L. and I. Morgenstern (eds) (1983). *Heidelberg colloquium on Spin Glasses*, Lecture Notes in Physics **192** (Berlin: Springer)

[125] van Hemmen, J. L. and I. Morgenstern (eds) (1986). *Heidelberg colloquium on Glassy Dynamics*, Lecture Notes in Physics **275** (Berlin: Springer)

[126] Watkin, T. L. K. and A. Rau (1993). *Rev. Mod. Phys.* **65**, 499

[127] Werbos, P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* Ph.D. Thesis, Harvard University.

[128] Willeboordse, F. H. (1992). *Order and Chaos in Coupled Logistic Maps* Ph.D. Theiss, University of Tsukuba, and references therein.

[129] Wolynes, P. G. (1992). In *Spin Glasses and Biology* ed D Stein (Singapore: World Scientific)

[130] Yasutomi, A. (1994). *Suuri-Kagaku*, **368**, 48

[131] Young, A. P. (1983). *Phys. Rev. Lett.* **51**, 1206